On Stability under Aggregation - Draft Version.

Albrecht Iseler

Date and time of present version: 23.11.99/08.00

Table of contents at the end of this document.

Please send e-mail to iseler@zedat-fu-berlin.de, if you want to be notified about later versions, mentioning the above date and time of present version.

If there should be **poorly readable mathematical characters** in a *.pdf-version of this document, please visit
http://userpage.fu-berlin.de/~iseler/papers/aggrstab.htm

Preliminary Remark

The present document is a draft of parts of a manuscript, which is planned to be submitted for publication. It is deposited in the Internet as a substitute for the final document, which is quoted in other publications of the author.

Due to this status of a draft, the degree of elaboration varies between sections. In particular, the character sequence ### marks some (but not all) places, where text or cross references have to be added. Furthermore, due to software problems, second degree subscipts and exponents are marked by parentheses instead of a vertical displacement. For instance, λ with subscript q_j is written as $\lambda_{q(j)}$.

Quotation and critique in publications should duely mention the draft status of the present document.

Comments to the following adress are welcome. Since formal bugs (including minute details) are detected easier in the first reading of a text than in repetitive proof reading, notes about such bugs are explicitly wanted. Please mention the above date and time of the present version.

Author's adress:
Albrecht Iseler
Freie Universität Berlin
Habelschwerdter Allee 45
D-14195 Berlin.
E-mail: iseler@zedat.fu-berlin.de
WWW: http://userpage.fu-berlin.de/~iseler

Abstract

In the derivation of predictions for empirical research, psychologists have frequently asked whether a property of interest is stable under aggregation in the sense that the presence of this property in (almost) all elements entering a process of aggregation grants the same property for the result of this process. Answers to such questions are supported by two tools: By a definition of aggregation stability based on a set theoretical relational structure, and by an application of the theory of real vector spaces (in particular function spaces) to mathematical objects (e.g. probability distributions or other functions), which are used to characterise the aggregated elements and the aggregate as a whole. Some results, which are immediate consequences of the formal definition of stability under aggregation, allow to derive the aggregation stability of complex properties from the stability of simpler ones. In real vector spaces, the convexity of the set of all elements with a property under study is frequently (but not always) a sufficient or necessary condition of aggregation stability.

Introduction

Several times in the history of psychology, proponents of strict derivation of predictions for empirical research have been concerned under varying motivations with a type of questions with a common formal structure. In processes of aggregation (e.g., in the computation of an 'average learning curve' from 'individual learning curves'), there are some properties, whose presence in all aggregated elements grants the same property for the aggregate. But other properties of interest may be lacking in the aggregate even if they are shared by all aggregated elements. In other words, some properties are stable under aggregation, and others are not. So the problem arises whether there are sufficient or necessary conditions for this stability or whether it must be checked seperately for each property of interest.

In the following article, a formal explication of this concept of aggregation stability will be proposed, and from this formalisation there will be derived some rather general consequences, which may be useful tools to study the aggregation stability of properties claimed by hypotheses in concrete research situations. In Section ###....(### preview to be added!)

More subtle problems will arise for situations with an infinite domain underlying the aggregation. E.g., if the data of an experiment are considered as a sample from an infinite domain of 'persons in situations', then the said criterion of aggregation stability may fail to be sufficient, and additional assumptions will have to be introduced. Since an explicit statements of these problems requires some basic formalisations and results presented in Section ###, an overview of the remaining sections is postponed until Subsection ###.

At this point, the reader may expect a review of some typical examples from the pertinent literature. However, the classification of examples as well as the appreciation of their contributions will profit from a formalisation of the central concept of aggregation stability. Hence, this formalisation will be introduced first, and the reader is asked to rest content for the moment with the choice of an example from classical psychology of learning - a field of research, which has, indeed, been very stimulating for the study of stability under aggregation. In particular, the analysis of this example will immediately lead to a simple criterion of aggregation stability covering most results reported in the pertinent literature.

1 Basic Formalisations and Results for Stability under Averaging

In this section, we will demonstrate by a concrete example an approach, which will be generalised in subsequent sections. In the example, aggregation consists in the computation of average learning curves for samples of individual learning curves. The example will be treated in a way providing three types of generalisable results. The first (and most important) one is a formalised concept of aggregation stability based on a set theoretical relational structure. Furthermore, we will demonstrate the application of some basic results of the algebraic theory of real vector spaces (a generalisation of the well known spaces consisting of n-dimensional vectors of real numbers). Finally, we will derive a necessary and sufficient criterion of aggregation stability for situations, where aggregation can be conceived as averaging of vectors in such spaces.

1.1 An Example

Example 1.1: Imagine a tracking task[1], where the subject has to keep a pointing device on a moving target, and assume that the learning time is subdivided into 10 intervals of equal length ('trials'), the performance in these trials being represented by learning curves with the percentage of the time on target as dependent variable. For this situation, consider the following properties P_1, P_2 and P_3, which could be claimed for such learning curves by various research hypothesess:

P_1: The learning curve is strictly increasing.

P_2: The percentage of time on target is a strictly increasing, negatively accelerated quadratic function of the trial number.

P_3: The learning curve starts with a strictly increasing part followed by a strictly decreasing section and another strictly increasing one covering the rest of the curve.

For each one of these properties, the question for its stability under aggregation can be asked in the following format: Does the presence of property P_j in the 'average learning curve' of a sample follow from the premissa that all individual learning curves have property P_j?

The reader is invited to answer this question for each one of the three properties in the frame of her or his own mathematical background; but before we derive such answers from a fromalised concept of aggregation stability, we should pay some attention to two questions, which demonstrate the methodological context of aggregation stability. First, one could ask why we bother at all to consider an average learning curve, if the property claimed by a research hypothesis can be tested separately for each individual learning curve. Undoubtedly, the approach suggested by this question is superior to the statistical analysis of the average learning curve, if the hypothesis claims one of the above properties for the curves made up by the observed scores of individual subjects. However, a

[1] The choice of a tracking experiment for the central example of Section ### should be understood as a reverence for the late Hans Jürgen Eysenck, whom the authors owes much more than the acquaintance with the tracking paradigm (Eysenck, Iseler, Star & Willett, 1969).

typical motivation to analyse average curves is based on the idea that 'experimental error' may hide a property which would be present in 'error free' individual curves.Under this view, it may be sensible to test the prediction that an average curve doesn't significantly deviate from a property, which is claimed for hypothetical error free individual curves.[2] In later sections, we will also consider situations, where values of a function cannot be obtained for individual subjects, but only for an aggregate.

A second critical argument referring to the question formulated in our example may be motivated by a comparison with a stronger concept of stability under aggregation. E.g., in the process of mixing cocktails, the property of being non-alcoholic is stable in a sense, which can be expressed as an equivalence: The aggregate (i.e., the cocktail) is non-alcoholic, if and only if (!) all ingredients are non-alcoholic.[3] It could be argued that only an equivalential aggregation stability of this kind would allow to make the usual inferences from the presence of a property in an average curve. Indeed, it would be seriously premature to infer a property of interest for all elemts of a sample or a population from its presence in an everage learning curve, even if the implicative aggregation stability formulated in our example can be proved for the property under consideration. But on the other side, it can be considered as a widely accepted result of the philosophy of science that it is hopeless in most situations to strive for empirically testable predictions, whose fulfillment is logically equivalent with the validity of a theory. Furthermore, the subsequent formalisation of aggregation stability will allow to derive (in Section ###) a strategy of hypothesis testing complying much better than conventional practice with requirements of strictness.

Certainly, the question for the aggregation stability of properties P_1, P_2 and P_3 in Example ### can be answered without this formalisation. Subsequent references to this example are motivated by a well approved practice in the presentation of software packages. E.g., we don't need a computer to solve the equation $2 \cdot x + 3 = 17$; but as a first step in the process of becoming familiar with a new software, it may be helpful to let it solve this simple equation.

---

[2] More explicitly, this issue is discussed by Estes (1956). Note that his arguments imply that the concept of 'error' mustn't necessarily be the one of classical test theory or of analysis of variance. The author explicitly gives room to the possibility that the mean of the errors systematically differs from 0. Furthermore, in the discussion of non-linear (e.g. logarithmic) transformations applied to the measurements, he implicitely assumes that the 'error free' value of the transformed measurements is obtained by applying the transformation to the error free value of the untransformed measurements. But then the interpretation of an ideal 'error free curve' as a curve of 'true scores' in the understanding of classical test theory is prevented by the inequality $E\log(X) < \log(EX)$, which applies to every strictly positive random variable X with non-zero variance. (###: Wieso?)

[3] Of course, this equivalence is valid only if we confine the claim to cocktail receipts where alcohol isn't destroyed or generated by chemical reactions. Generalisation: Aggregation stability of a property has to be stated relative to a set of aggregates under consideration.

1.2 A formalised Concept of Stability under Aggregation.

The following conceptualisation of aggregation stability will be guided by the goal to give a precise meaning to a general format of assertions claiming this stability. We will start with the format of the question asked for Example ###, transform it into a more precise format of an assertion, work out a set theoretical relational structure underlying a general concept of aggregation stability, and then summarise in a definition the structure and the concept.

In Example ###, the question for the aggregation stability of a property P_j was asked in the following format: Does the presence of property P_j in the 'average learning curve' of a sample follow from the premissa that all individual learning curves have property P_j? To prepare a more precise assertional format, observe first that this question asks, whether an implication is valid for all samples consisting of individual learning curves in the outlined experiment. Although this reference to the universe of all samples isn't formulated explicitly, the question is clear enough to require a negative answer, if there should be a conceivable sample, where property P_j holds for all individual learning curves, but not for the average learning curve.

Before we give a more precise definition of this universe, whe should formally introduce the learning curves. With the denotation Q for the set of trial numbers 1 through 10, let V be the set of all maps of Q into the interval [0, 100] of real numbers. Then this set is the set of all (potential) learning curves.[4] Subsequently, this set will also be called the 'vocabulary set', since its elements form a vocabulary used to qualify individuals (by individual learning curves) and samples (by average learning curves). A vocabulary set with typicall denotation V will also be the most basic set in our general conceptualisation of aggregation stability. In many examples, there will be a set Q such that the vocabulary set consists of maps $Q \to \mathbb{R}$, but to keep the conceptualisation open for other vocabularies, the set Q isn't introduced as a constituent component.

Returning to the tracking task, we can represent 'samples' by a set Π consisting of all finite sequences[5] of leaning curves - i.e., of elements of V. The denotation Π for this set alludes mnemonically to the fact that the formal role of this set will frequently be taken a set of probability measures. But our example shows that this property of the set Π isn't constitutive.

It will turn out helpful to represent properties, whose stability under aggregation is studied, by the subset of V consisting of all elements of V with the respective property. E.g., the properties P_1, P_2 and P_3 of learning curves in the tracking task are represented by sets S_1, S_2 and S_3, each one consisting of all maps $Q \to [0, 100]$ with the respective property. This set representation of properties will open a convenient interface to some useful mathematical results referring to a special class of

---

[4] One could argue that (due to rounding of time measurement) only maps of Q into the set of rational numbers can occur as empirical learning curves. So we stipulate that the question for aggregation stability in Example ### refers also to the averaging of ideal error free learning curves, where the percentage of time on target may be every real number in the interval [0, 100].

[5] Since the learning curves of some subjects may be identical, it would be imprecise to speak of finite subsets of V: In such subsets, no learning curve can appear more than once, but it may very well in a finite sequence. If we would have to list all elements of Π, then it would undoubtedly be a luxury to conceive sequences π_1 and π_2 as different elements of Π even if π_2 is only the result of reindexing π_1. But this concepualisation supports simplicity of proofs.

sets: Convex subsets of real vector spaces. (This concept will be introduced in Section ###.) Furthermore, the stability of the property of 'being an element of S' can be claimed or disclaimed for every subset S of the vocabulary set V without any intensional definition of a property represented by S. For the tracking experiment, aggregation stability of property P_j can be asserted in a proposition with the following format: For every sample of individual learning curves (i.e., for every $\pi \in \Pi$), the average learning curve is an element of S_j, if all individual learning curves are elements of S_j.

To prepare a generalisation of this propositional format, we conceive the computation of average learning curves as a map $\Phi:\Pi \to V$. Then $\Phi(\pi)$ is the element of V which is used to qualify the aggregate π. In the tracking example, $\Phi(\pi)$ denotes the average learning curve of sample π.

To represent the premissa (i.e., the if-clause) in the above format of an assertion of aggregation stability, let H be the set of all ordered pairs (S, π), where S is a subset of V and π a finite sequence of individual learning curves (i.e., an element of Π) such that all these individual learning curves are elements of the set S. Obviously, this set H is a subset of the cartesian product $\mathsf{P}V \times \Pi$ (where $\mathsf{P}V$ denotes the power set of V); i.e., H represents a relation. With these definitions, the format of assertions claiming aggregation stability for the property of 'being an element of S' can be written as

$$\forall \pi \in \Pi: ((S, \pi) \in H \Rightarrow \Phi(\pi) \in S). \qquad (\#\#\#)$$

To complete the relational structure, we introduce a set system T consisting of all subsets S of the vocabulary set V with the property expressed by Formula (###). Then we can summarise in the following definition the set theoretical structure and some assumptions ('axioms') making up a general conceptualisation of stability under aggregation.

Definition 1.2: A *structure of stability under aggregation* (abbr.: *SSA*) is an ordered quintuple (V, Π, Φ, H, T) consisting of non-empty sets V and Π, a map $\Phi:\Pi \to V$, a relation $H \subseteq \mathsf{P}V \times \Pi$, and a set system $T \subseteq \mathsf{P}V$ such that the following properties hold for every $\pi \in \Pi$ and all subsets S' and S'' of V:
- (i) $(\varnothing, \pi) \notin H$.
- (ii) $(V, \pi) \in H$.
- (iii) $(S' \subseteq S'' \wedge (S', \pi) \in H) \Rightarrow (S'', \pi) \in H$.
- (iv) $T = \{S \subseteq V: \forall \pi \in \Pi: ((S, \pi) \in H \Rightarrow \Phi(\pi) \in S)\}$.

Since this definition is rather formal and abstract, some comments may support the understanding of its meaning and the relationship to applications.

a. Whereas the components of the quintuple (V, Π, Φ, H, T) have been previously introduced with a concrete interpretation for the tracking experiment, only the properties formulated in Definition 1.2 are constitutive for the general concept of a structure of stability under aggregation (not withstanding a subsequent reintroduction of verbal circumscriptions).
b. Assertions (i), (ii), (iii) and (iv) will subsequently be called SSA-Axioms. Additional axioms for a special class of SSAs will be introduced by Definition 4.1 in Section 4.1.
c. Typically, the elements of the 'vocabulary set' V will be mathematical objects available for the

result of measurement in a general sense.[6] But the set V can also consist of other elements, e.g. everyday language words or identifiers for units and aggregates.

d. The set Π (the set of 'aggregates' under consideration) will usually have a higher order relationship to the vocabulary set V, which may be verbalised as 'being composed of elements of V' in some way. In particular, this 'being composed' may be direct (e.g. in a sample of learning curves), but a view with a more indirect relationship is also possible (e.g. for a sample of persons, who are qualified by elements of the vocabulary set). However, for purposes of generality, this property of the set Π is introduced only indirectly via the SSA-Axioms (i), (ii) and (iii).
e. The map Φ introduces the assumption that each aggregate (i.e., every $\pi \in \Pi$) is characterised by a unique element of V, which will subsequently be denoted as $\Phi(\pi)$ or (more conveniently) as Φ_π. Since Φ is something like a rule for the assessment of this unique element of V, we will also speak of the 'aggregation rule' Φ. So the SSA-formalisation cannot be applied immediately to situations where not a unique element of V, but a collection of them (say, v_1, v_2 ...) is used to characterise an aggregate π. However, such situations may be brought into the assumed format by a transition to a set Π' containing ordered pairs like (π, v_1), (π, v_2) etc. as separate elements.[7]
f. The relation H can be verbalised as being 'sheer'.[8] Hence we will say that an aggregate π is 'S-sheer' iff $(S, \pi) \in H$. For convenience, we will also use the propositional notation $H(S, \pi)$ instead of $(S, \pi) \in H$, and $\neg H(S, \pi)$ for $(S, \pi) \notin H$. (### nötig?)
g. The understanding of SSA-axioms (i), (ii) and (iii) may be supported by the following verbalisations: No aggregate is sheerly composed of elements of the empty set, and each aggregate is sheerly composed of elements of V. Finally, if S' is a subset of S'' and an aggregate π is sheerly composed of elements of S', then π is also sheerly composed of elements of S''. In other word, the axioms define the class of predicates, which may replace the phrase 'sheerly composed of ...' .
h. This implies that other properties, which would also be plausible under this verbalisation, are unwanted overinterpretations. Consider e.g. the following assumption: If $\{S_i\}_{i \in I}$ is a family of subsets of V, and π an element of Π such that $(S_i, \pi) \in H$ for every $i \in I$, then $(\bigcap_{i \in I} S_i, \pi) \in H$. Indeed, if the claim 'aggregate π is sheerly composed of elements of S_i' holds for every $i \in I$, then it would be plausible to stipulate that π is sheerly composed of elements of the intersection. But it will turn out in Section 4.3 that an axiom of this kind can conflict with other axioms of a class of SSAs, which can be considered to form a useful compromise between generality and specifity. Furthermore, various weaker versions of the intersection property will be sufficient for conclusions to be presented in Section 3. In particular, the most important results of that section will not need any assumption of this kind. For these reasons, the above intersection property is not given the status of an axiom, and substitutes with locally specified index sets I will be

---

[6] See e.g. Krantz...### for a general concept of measurement, where mathematical objects used as 'measurement values' must not necessarily be real numbers.

[7] See Section 6.2 in the Appendix for an example.

[8] Mnemonic: The letters H and T for the last two components of an SSA refer to the second letters in the words 'sheer' and 'stability'. Certainly, the mnemonic function would be supported better by the choice of first letters; but the letter S occurs twice and is reserved for 'conditions' and 'sets'.

introduced, where they enable additional conclusions.
Another assumption would also be plausible under the above introduced verbal interpretation of the relation H: If S' and S" are disjoint subsets of the vocabulary set V, then an aggregate cannot simultaneously be sheerly composed of elements of S' and sheerly composed of elements of S". In other words, the properties (S', π) $\in$ H and (S", π) $\in$ H are mutually exclusive for the same aggregate π and disjoint subsets S' and S" of the vocabulary set. But again, this property isn't given the status of an axiom, since the main results of the general theory to be derived from Definition 1.2 do not need an assumption of this kind.

i. It is easily verified that SSA-Axioms (i), (ii) and (iii) hold under the interpretation of the relation H for the tracking experiment. In Sections ### and ###, we will introduce other interpretations of the relation H by 'almost sure' properties in probability spaces.

j. The set system T and SSA-Axiom (iv) form the top of the relational structure. We will say that a property is 'stable under aggregation' with respect to the SSA (V, Π, Φ, H, T) iff the set of all elements of V with the considered property is an element of the set system T. For convenience, we will also say that a set S is stable under aggregation iff it is an element of T. Furthermore, the introduction of the set system T allows convenient notations, e.g. C $\subseteq$ T for the claim that all subsets of V belonging to a class C are stable under aggregation.

k. Whereas the admissible definitions of the relation H are specified by SSA-axioms (i), (ii) and (iii), the fruitfulness of a definition for a specific application is particularly determined by SSA-Axiom (iv). So it will be helpful to derive this interpretation from a verbal formulation of the implication, which must hold for every $\pi \in \Pi$ to make a set S stable under the considered kind of aggregation. Then the if-clause of the implication will yield an appropriate interpretation of the relation H. Note that this approach has already been applied for the tracking experiment. We started with a proposition, which must hold for every aggregate π to grant stability under averaging for a subset S of the vocabulary set V: If all individual learning curves are elements of S, then the average learning curve is also contained in S. Since this is the implication of interest, a definition of the relation H, which formalises the if-clause in this implication, is appropriate for the analysis of stability under averaging.

As in almost every relational structure, one could consider to add more components to the definition of an SSA. So we could introduce a system of subsets of V representing the stronger concept of 'equivalential' aggregation stability mentioned in Subsection 1.1. The restriction to five constituent components in Definition 1.2 has been motivated by the consideration to list only those necessary for a definition of implicative stability under aggregation. Formally, this is reflected in SSA-axiom (iv), where all five components turn up. This implies that claims of aggregation stability - i.e., of membership in the set system T - have to be understood relative to the other components. Indeed, in the hitherto elaborated SSA for the tracking experiment, some properties are stable under aggregation, but would loose this qualification under a different aggregation rule (e.g. if aggregates would be characterised by the curve of sample medians instead of the average learning curves). Similarly, the presence or absence of aggregation stability would change for some subsets of V, if the set Π would be confined to a proper subset (e.g. sequences of a length not greater than the number of living human beings) or if the definition of 'sheerness' (i.e., of the relation H) would be altered. To express this frame of reference, the term 'aggregation' in the phrase 'stability under aggregation' may be replaced by a more specific concept for the kind of aggregation. So we will also speak of 'stability under averaging' in contexts like average learning curves.

Although equivalential aggregation stability is not represented by a constituent component of an SSA, it will sometimes be useful as a tool in the derivation of implicative stability. So we introduce the notational convention that adding the subscript e to the last component of an SSA (leading e.g. to T_e) refers to the system of all subsets of V, where the implication in SSA-axiom (iv) can be sharpened to an equivalence; i.e.,[9]

$$T_e := \{S \subseteq V: \forall \pi \in \Pi: ((S, \pi) \in H \Leftrightarrow \Phi(\pi) \in S)\}. \tag{1.1}$$

Some elementary properties of the set systems T and T_e, which follow immediately from SSA-Axiom (iv) and Equation (1.1), are stated in the following Lemma, which is proved in Section 6.3.

Lemma 1.3: In an SSA (V, Π, Φ, H, T), the set system T and the set system T_e given by Equation (1.1) have the following properties.

(i) $T_e \subseteq T$.
(ii) The empty set and the set V are contained in T and in T_e.
(iii) T contains all subsets S of the vocabulary set V with $\Phi(\Pi) \subseteq S$.
(iv) A subset S of V which is contained in T is also contained in T_e iff the implication $\Phi_\pi \in S \Rightarrow (S, \pi) \in H$ holds for every $\pi \in \Pi$.
(v) If the relation H is replaced by a relation $H' \subseteq PV \times \Pi$ with $H \subseteq H'$, then the set system T' in an SSA (V, Π, Φ, H', T') is included in T (i.e., $T' \subseteq T$)

Further consequences of Definition 1.2, which can be used as tools for the study of aggregation stability, will be presented in Section 3.

Although Definition 1.2 has been tailored for analytical purposes, it may also be applied to empirical aggregation. E.g., we could ask whether the behaviour of a group will be 'clever' (or 'silly' or 'non-aggressive'), if all its members have the respective property. Then V may be a set of classes for the variable under study, and Π a set of groups. Of course, we would need a suitable empirical procedure to assess whether the behaviour of an entire group has the property of interest, and whether the group is entirely composed of members with that property. Hence, the map Φ and the relation H would be empirical in this situation. The question whether our formal concepts and results may be helpful in such fields is left to researchers interested in this kind of empirical aggregation stability.

1.3 Convex Sets and Stability under Averaging

In this subsection, we will introduce a notation for average learning curves as *convex linear combinations* of individual curves and introduce *real function spaces* $\mathbb{R}^Q$ as well as some basic concepts related to such spaces. In particular, the convexity of subsets of a vocabulary set will turn

[9] For references to Equation (1.1) in later sections, where other denotations are used for the components of an SSA, these denotations will have to be adapted suitably. In particular, the left hand side of the equation will results from the addition of the subscript e to the denotation for the last component of the SSA under consideration.

out to be a sufficient criterion of stability under averaging, and a slightly weaker criterion will be necessary and sufficient. On this background, a criterion of aggregation stability presented by Estes (1956) can be generalised, and many examples of lacking aggregation stability reported in the pertinent literature can be classified as instances of non-convex subsets of a vocabulary set.

In the exposition of Example 1.1, the 'average learning curve' of a sample was introduced unexplained, but the reader will have understood it correctly as a curve representing the sample averages of individual measurement values for each trial. At first glance, it is only a minor reformulation of this description, if we say that the individual curves are averaged 'pointwise'; but in this restatement, the averaging is considered as an operation applied to entire curves. There is an obvious resemblance to wellknown operations with n-dimensional vectors of real numbers: Their 'elementwise' addition and multiplication by a scalar. Indeed, these operations can be easily generalised to maps of a set Q into the set of real numbers. If Q is an arbitrary non-empty set, then the *scalar multiplication* of a real number λ and a map $f:Q \to \mathbb{R}$ results in the map $g:Q \to \mathbb{R}$ with $g(q) = \lambda \cdot f(q)$ for every $q \in Q$. Similarly, the sum of two maps $f:Q \to \mathbb{R}$ and $g:Q \to \mathbb{R}$ is the map $h:Q \to \mathbb{R}$, where the equation $h(q) = f(q) + g(q)$ holds for every element q of the element Q. Observe that the usual notation $g = \lambda \cdot f$ resp. $h = f + g$ isn't just a shorthand. It expresses that the operands - the maps $Q \to \mathbb{R}$ - are considered as wholes.

Based on these operations, the average learning curve $Z\pi$ of sample π can be written as

$$\Phi_\pi = \sum_{i=1..n} \lambda_i \cdot v_i, \tag{1.2}$$

where n is the number of pairwise different learning curves v_i occurring with relative frequency λ_i in sample π.

It is more than an analogy, if we apply the well known concept of a *linear combination* to the sum on the right hand side of Equation (1.2). Since the mathematical concept of function spaces underlying this generalisation may be unfamiliar to some readers, we will briefly introduce these spaces and some related concepts. (Other readers can click here ### to skip to the next paragraph.) Observe first that the vocabulary set V of maps $Q \to [0, 100]$ isn't closed under the above defined operations of multiplication by a scalar and addition: The maps $\lambda \cdot f$ or $f + g$ may be maps outside the set V, even if the operands f and g are elements of V. However, if we introduce the denotation E for the set of *all* maps of our set Q into $\mathbb{R}$, then it is easily verified that this set E is closed under the above operations. Moreover, addition is commutative ($f + g = g + f$) and associative ($(f + g) + h = f + (g + h)$), and the distributive properties $\lambda \cdot (f + g) = \lambda \cdot f + \lambda \cdot g$ and $(\lambda + \mu) \cdot f = \lambda \cdot f + \mu \cdot f$ hold for the product of a scalar and a map. These properties, which are well known from n-dimensional vectors of real numbers, are some of the axioms of a geralised concept of *real vector spaces*, and since the remaining axioms (listed in ###) apply as well, the set E - endowed with pointwise addition and multiplication by scalars - is a real vector space. In a space of this kind, all real numbers are *scalars*, and the elements of the set E are *vectors*. Of course, these conclusions are valid not only for the set Q of trial numbers 1 through 10 in the tracking experiment, but for every (finite or infinite) non-empty set Q: With the commonly used denotation $\mathbb{R}^Q$ for the set of all maps $Q \to \mathbb{R}$, we can summarise our results in the conclusion that a set $\mathbb{R}^Q$ of this kind is a real vector space, if it is endowed with pointwise addition and multiplication by scalars. More specifically, $\mathbb{R}^Q$ becomes a *real function space* by this endowment.

Returning to Equation (1.2), observe that the relative frequencies λ_i are non-negative real numbers summing up to $\sum_{i=1..n} \lambda_i = 1$. So the right hand side of the equation belongs to the special class of *convex linear combinations*. (Terminology: In a real vector space, a *linear combination* is a (finite!) sum of the form $\sum_{i=1..n} \lambda_i \cdot x_i$ with scalars λ_i and vectors x_i. A sum of this kind is a 'convex linear combination (of the vectors x_i)' iff the scalars λ_i are non-negative and sum up to 1.)

The term 'convex' reflects a close relationship of these linear combinations to a class of subsets of real vector spaces, which will turn out crucial for aggregation stability: The *convex subsets*. In fact, almost all examples of lacking aggregation stability, which are discussed in the pertinent literature, can be conceived as instances of lacking convexity. Briefly speaking, a subset S of a real vector space is convex iff all linear combinations $\lambda \cdot x_1 + (1-\lambda) \cdot x_2$ with $\lambda \in [0, 1]$, $x_1 \in S$ and $x_2 \in S$ result in an element of S.[10] The following geometric interpretation may support the understanding of this concept: If x_1 and x_2 are points in a 2-dimensional or 3-dimensional space, then the line segment connecting these points consists of the results of all linear combinations $\lambda \cdot x_1 + (1-\lambda) \cdot x_2$ with $\lambda \in [0, 1]$. So we can also say that a set S of points in these spaces is convex iff every line segment connecting two points of S is entirely contained in S. The reader is invited to visualise positive and negative examples and to verify (by the algebraic definition) that the vocabulary set V introduced for the tracking task is convex.

Now it can be easily derived from the above definitions that every convex linear combination of elements of a convex set S results in an element of S.[11] Applying this result to Equation (1.2), we can conclude for the tracking experiment that every convex subset S of the vocabulary set V is stable under aggregation: If all individual learning curves in a sample π are elements of S (i.e., if $(S, \pi) \in H$), then the average learning curve Φ_π is a convex linear combination of elements of S. Hence Φ_π is an element of S, if S is convex.

This result opens a way to show that the properties P_1 and P_2 of Example 1.1 are stable under averaging: It is sufficient to verify the convexity of the sets S_1 and S_2 of elements of V with the respective properties. Since the convexity of subsets of the vocabulary set will turn out to be a sufficient condition of aggregation stability in many situations, it may be helpful to demonstrate some useful schemes for proofs of convexity, even if slightly shorter proofs would be possible for the sets S_1 and S_2 of the tracking experiment. Furthermore, the approach to the set S_2 will prepare the treatment of a well known result of Estes (1956).

To start with S_1 (i.e., the set of strictly increasing learning curves), observe first that this set can

---

[10] Note that definitions of convex sets in the pertinent literature disagree in their treatment of the empty set. In accordance with some authors (e.g., ###), the above definition implies that this set is convex: Since there are no linear combinations $\lambda \cdot x_1 + (1-\lambda) \cdot x_2$ with $x_1 \in \varnothing$ and $x_2 \in \varnothing$, all linear combinations of this kind result in an element of $\varnothing$. (A universal proposition referring to an empty universe is always true!) But other definitions (e.g., ###) exclude the empty set from the concept of convexity. Since the empty set is stable under aggregation according to Definition ###, its exclusion from convexity would complicate later discussions of convexity as a necessary criterion of aggregation stability.

[11] See Section 6.4 for a proof.

be represented as an intersection of sets with a more simple definition. For the trial numbers[12] $q = 1..9$, let the sets S^*_q be given by the definition

$$S^*_q := \{f \in \mathbb{R}^Q\colon f(q+1) - f(q) > 0\}. \tag{1.3}$$

Then the set S_1 can be represented as an intersection:

$$S_1 = (\bigcap_{q=1..9} S^*_q) \cap V. \tag{1.4}$$

Note that the sets S^*_q contain not only learning curves: Other maps $f{:}Q \to \mathbb{R}$ (i.e., other elements of $\mathbb{R}^Q$) will also be elements of S^*_q. But they are removed in Equation (1.4) by the intersection with V.

Now it is a well known (and easily verified) fact that the intersection of convex sets is a convex set.[13] Since the convexity of V has been established earlier, it suffices to show that the sets S^*_q are convex. So let the maps $f{:}Q \to \mathbb{R}$ and $g{:}Q \to \mathbb{R}$ be arbitrary elements of S^*_q, and λ an arbitrary real number such that $0 < \lambda < 1$. To show that the map $h{:}Q \to \mathbb{R}$ given by

$$h := \lambda \cdot f + (1-\lambda) \cdot g \tag{1.5}$$

is an element of S^*_q, we can write

$$\begin{aligned} h(q+1) - h(q) &= \lambda \cdot f(q+1) + (1-\lambda) \cdot g(q+1) - \lambda \cdot f(q) - (1-\lambda) \cdot g(q) \\ &= \lambda \cdot (f(q+1) - f(q)) + (1-\lambda) \cdot (g(q+1) - g(q)) \\ &> 0. \end{aligned} \tag{1.6}$$

The first equality results from Equation (1.5) by pointwise linear combination, and the second one from a rearrangement of terms. For the final inequality, the properties $f(q+1) - f(q) > 0$ and $g(q+1) - g(q) > 0$ follow immediately from the assumption that f and g are elements of S^*_q. Furthermore, the assumption $0 < \lambda < 1$ grants that the coefficients λ and $1 - \lambda$ are greater than 0. But the inequality $h(q+1) - h(q) > 0$ implies that h is an element of S^*_q. Hence S^*_q is convex.

Before we proceed to the set S_2, we should note two generalisable tools for the proof of the convexity of a set S. First, the set S can be represented as an intersection of sets, whose convexity can be proved more easily. The second tool has been applied tacidly in the proof for the sets S^*_q. Although the definition of convexity requires that the vector $\lambda \cdot x_1 + (1-\lambda)\, x_2$ with $x_1 \in S$ and

---

[12] The subscript q denoting trial numbers may seem strange; but in the context of function spaces $\mathbb{R}^Q$, we will consistently use this letter for elements of Q.

[13] For a real vector space E, let C be an arbitrary collection of convex subsets of E, and S the intersection of these sets. (I.e., S consists of those vectors, which are elements of all members of the collection C.) To verify $x \in S$ for the vector $x := \lambda\, x_1 + (1-\lambda)\, x_2$ (with $x_1 \in S$, $x_2 \in S$, and $\lambda \in [0, 1]$), we can derive an immediate consequence from the definition of S: The vectors x_1 and x_2 are elements of every member of C, if they are elements of S. But since all members of C are assumed to be convex, they must also contain the vector x. So $x \in S$ follows from another reference to the definition of S.

$x_2 \in S$ is contained in S for $0 \leq \lambda \leq 1$, the assumption $x_1 \in S$ grants this property for $\lambda = 1$, and for $\lambda = 0$, it follows from $x_2 \in S$. So it suffices to verify the containment for $0 < \lambda < 1$.

Property P_2 is slightly more complex than P_1, but the convexity of the set S_2 can be proved in a similar way: We will represent S_2 as an intersection of simpler subsets of $\mathbb{R}^Q$ and prove their convexity. To prepare this approach, we consider three subsets of $\mathbb{R}^Q$ representing properties, whose conjunction will specify the maps $f:Q \rightarrow \mathbb{R}$ which are elements of S_2. Let S' be the set of all maps $f:Q \rightarrow \mathbb{R}$, which are quadratic; i.e, functions which can be written as

$$f(q) = \alpha + \beta \cdot q + \gamma \cdot q2$$

with 'parameters' (real numbers) α, β and γ. In particular, note that functions with parameter $\gamma = 0$ are also 'quadratic' in the sense of the above definition. But these functions as well as the positively accelerated ones will be excluded, if the difference $f(2) - f(1)$ is greater than $f(3) - f(2)$. A minor reformulation of this property in the subsequent definition of a set S'' will support the proof of its convexity:

$$S'' := \{f \in \mathbb{R}^Q: 2\, f(2) - f(1) - f(3) > 0\}.$$

Finally, to be an element of S_2, the function f must be strictly increasing and and element of the vocabulary set V. Both properties can be combined in the requirement that f must be an element of the set S_1. So we cant represent the set S_2 as an intersection:

$$S_2 = S' \cap S'' \cap S_1.$$

Since the convexity of S_1 has already been verified, it is left to show this property for S' and S''. So we assume first that the maps $f:Q \rightarrow \mathbb{R}$ and $g:Q \rightarrow \mathbb{R}$ are elements of the set S'. This means that they can be written in the format of Equation (###); but since the parameters α, β and γ may be different for both functions, we denote the parameters for function f as α_f, β_f and γ_f, whereas α_g, β_g and γ_g are the parameters for function g. Now let λ be a real number such that $0 < \lambda < 1$, and let the map h again be given by Equation (1.5). Then it is easily verfied that the equation

$$h(q) = \alpha_h + \beta_h \cdot q + \gamma_h \cdot q^2$$

holds for everys q with parameters $\alpha_h := \lambda \cdot \alpha_f + (1-\lambda) \cdot \alpha_g$ etc.. So h is an element of S'; i.e., S' is convex.

Since the set S'' is defined in a formally similar way as the sets S^*_q defined above for the treatment of S_1, it can be left to the reader to transfer the proof of convexity.

The reader will also have found that Property 3 in Example 1.1 isn't stable under aggregation, which can be proved by a single example. The subsequently listed learning curves f and g have Property P_3, but it is lacking in the average learning curve h.

f	12	21	28	39	47	42	36	57	65	78
g	24	37	42	35	31	48	58	63	71	76
h	18	29	35	37	39	45	47	59	68	77

In addition to being an example of the lacking aggregation stability of Property 3, the learning curves f, g and h also show that the subset S_3 of V consisting of learning curves with Property 3 isn't convex. Since f and g are elements of S_3, convexity of S_3 would imply $h \in S_3$, since the averaging of f and g mounts up to an application of Equation (1.5) with $\lambda = ½$. So S_3 is non-convex, since h isn't element of S_3.

Note, however, that the lacking aggregation stability of Property 3 cannot be derived from the non-convexity of S_3: In our example, convexity is a sufficient condition of aggregation stability, but not a necessary one. Consider, e.g., the set S_r consisting of learning curves $f:Q \to [0, 100]$ with rational function value f(q) for every $q \in Q$. Since the average of rational numbers is a rational number, the property defining this set is stable under averaging. However, if we take two non-identical elements f and g of S_r and introduce a map h by Equation (1.5) with irrational λ, then h will not be an element of S_r. So S_r is non-convex, but stable under averaging; i.e., convexity isn't necessary for stability under averaging.

A weaker property than convexity, which is necessary and sufficient for stability under averaging, can be based on a slight modification of the definition of convexity of a set S. Whereas linear combinations of the form $\lambda \cdot x_1 + (1-\lambda) \cdot x_2$ with $x_1 \in S$ and $x_2 \in S$ must result in an element of S for *every* $\lambda \in [0, 1]$ to make the set S convex, Assertion (ii) of the subsequent lemma confines this requirement to rational numbers λ. Furthermore, the lemma (which is proved in Section 6.5) generalises the SSA for stability under averaging, which was introduced for the tracking experiment.

Lemma 1.4: Let (V, Π, Φ, H, T) be an SSA such that V is a subset of a real vector space, and Π the set of all finite sequences of elements of V. Furthermore, let the map $\Phi:\Pi \to V$ and the relation $H \subseteq PV \times \Pi$ be given as follows: For every element $\pi = \{v_i\}_{i=1..n(\pi)}$ of Π, where n(π) is the length of the sequence and the components v_i are elements of V, the function value Φ_π is given by

$$\Phi_\pi = n(\pi)^{-1} \cdot \sum_{i=1..n(\pi)} v_i, \tag{1.7}$$

and an ordered pair (S, π) with $S \subseteq V$ is an element of H iff $v_i \in S$ for i = 1..n(π).

Then the following assertions are equivalent for every subset S of V:

(i) $S \in T$.

(ii) For all elements v_1 and v_2 of S and every rational number λ with $0 \le \lambda \le 1$, the result of the linear combination $\lambda \cdot v_1 + (1-\lambda) \cdot v_2$ is an element of S.

(iii) For every natural number n, every sequence $\{v_i\}_{i=1..n}$ of elements of S, and every sequence $\{\lambda_i\}_{i=1..n}$ of non-negative rational numbers λ_i with $\sum_{i=1..n} \lambda_i = 1$, the result of the linear combination $\sum_{i=1..n} \lambda_i \cdot v_i$ is an element of S.

Note that the equivalence of Assertions (i) and (ii) in Lemma 1.4 means that Assertion (ii) formulates a necessary and sufficient condition of stability under averaging, whereas convexity turned out to be sufficient, but not necessary. Furthermore, this equivalence may facilitate the analysis of stability under averaging for properties of interest: It suffices to examine whether Assertion (ii) holds for the set of all elements of the vocabulary set with that property. More generally, we will subsequently strive for such easily examined conditions of aggregation stability, which hold for large classes of SSAs. Certainly, necessary and sufficient conditions will be most satisfactory. But sufficient and unnecessary conditions will do for the proof of aggregation stability of a property of interest, whereas necessary, but insufficient conditions can be helpful to show lacking aggregation stability.

Nevertheless, one can very well ask whether Lemma 1.4 is more than an instance of the well known class of theoretical 'results', whose triviality is misted by an unnecessarily blown up formalism. To some degree, the answer to this question will depend upon the mathematical background of the reader. In the view of the author, the lemma demonstrates that stability under averaging becomes trivial, indeed, by a combination of two approaches: By a precise definition of aggregation stability, and by an application of the theory of real vector spaces to the mathematical objects which are used to characterise individuals and aggregates. Moreover, the first one of these approaches hasn't yet been given an opportunity to demonstrate its efficacy. Lemma 1.4 could be reformulated without any reference to an SSA. As a substitute for Assertion (i), consider the following Assertion (i'): For all finite sequences $\{v_i\}_{i=1..n}$ of elements of S, the vector $n^{-1} \cdot \sum_{i=1..n} v_i$ is an element of S. Then a revised Lemma 1.4' could claim that for every subset S of real vector space, the new Assertion (i') is equivalent with Assertions (ii) and (iii) of Lemma 1.4. Even the representation of 'properties' by subsets could be disposed of by Assertions (i"), (ii") and (iii"), where membership in S is replaced by the property under consideration. So the reader is asked to recall (from Section ###) the analogy with the demonstration of a new software by examples, which could very well be handled without the software.

Before we derive some less trivial consequences of Definition 1.2, we will now bring on the postponed review of examples from the pertinent literature.

2 Positive and Negative Examples of Aggregation Stability

Undoubtedly, experts in the history of psychology will be able to show that more or less informal discussions of stability under aggregation are as old as the test of psychological hypotheses referring to individuals by predictions about sample statistics. However, the following presentation of examples doesn't aspire to be historically exhaustive. Rather, it will be one of our objectives to demonstrate by prototypical examples a rather wide scope of problems, which have emerged in concrete situations of psychological research and can be subsumed under the conceptualisation of Section ###. Some examples will also be taken from fields outside psychology, since they illustrate interesting facts, whose potential application in psychological theories will be outlined.

However, the subsequent treatment of examples will go beyond their mere subsumption under the formal concept of aggregation stability. We will also note down some observations for generalisations in later sections. In particular, we will show that a reformulation of historical results in the framework of function spaces $\mathbb{R}^Q$ makes available as useful tools some well known results of the mathematical analysis of such spaces. In some examples, a vocabulary set V consisting of maps $Q \to \mathbb{R}$ - i.e., elements of $\mathbb{R}^Q$ for some non-empty set Q - is given directly by the historical formulation of problems, and in other instances it will turn out easy to define a suitable set Q enabling this approach.

Although the discussion of stability under aggregation in the pertinent literature started with the averaging of functions like learning or other growth curves, an adequate review of some examples will require a more general type of aggregation than averaging in real vector spaces. It is introduced in the following subsection.

2.1 Aggregation by Convex Linear Combinations

In the analysis of average learning curves for the tracking task, the average curve Φ_π was represented in Equation (1.2) as the result of a convex linear combination $\sum_{i=1..n} \lambda_i v_i$, where the coefficients λ_i were the relative frequencies of n pairwise different curves v_i. In convex linear combinations with this interpretation, only rational numbers can occur as coefficients λ_i. This restriction had a rather unsatisfactory consequence: Some odd sets like the set S_r of all learning curves with rational values v(q) are stable under aggregation in the same way as those representing properties, which may be of psychological interest, like the sets of learning curves with properties P_1 resp. P_2 in Example 1.1. In fact, the aggregation stability of sets like S_r is no problem, if we study only the aggregation stability of a specific set of greater psychological relevance. But if we strive for sufficient *and* necessary conditions of aggregation stability, then it is awkward, if a condition is prevented from being necessary only by such odd sets. E.g., convexity of a subset S of the vocabulary set V in the tracking task isn't necessary, since only convex linear combinations with rational coefficients λ_i can occur in a representation of averaging by Equation (1.2).

In the examples from the literature, which will be reviewed in the following subsections, the problem can be solved, if we remove the restriction to convex linear combinations with rational coefficients and apply a concept of aggregation, where convex linear combinations with irrational coefficients can also occur. In his study of parametric families of functions, Estes (1956) indicated a suitable approach. He moved forward the interpretation of aggregation as pointwise averaging of functions to pointwise expectations. According to modern standards of probability theory, the real valued random variables, whose expectations are considered, must be embedded into suitable probability spaces. In later sections, we will introduce a class of SSAs, which will be designed to comply with these standards; but the following review of examples would be overcharged by this approach, and a lightened version will be entirely sufficient. Let D be a finite set of n units (e.g., persons), and Q an arbitrary non-empty set such that every unit i is characterised by an 'individual map' v_i:Q→$\mathbb{R}$. Assume furthermore that one of the units is selected by a procedure, where unit i is selected with probability λ_i (these probabilities being not necessarily equal for all units). With the denotation U for the selected unit, each function value $v_U(q)$ for some $q \in Q$ can be considered as a real valued random variable,[14] whose expectation is $\sum_{i=1..n} \lambda_i v_i(q)$. Collecting these expectations in a map Q→$\mathbb{R}$ and using the denotation Φ_π for this map, Equation (1.2) applies again; but now without the confinement to rational coefficients λ_i. Such processes of random selection and observation of units will subsequently be called RSO-processes, and stability under this kind of aggregation will be called stability under RSO-processes or just RSO-stability. Furthermore, the

[14] In the framework of probability measures, consider a probability space (D, PD, P) with $P(A) = \sum_{i\in A} \lambda_i$ for every subset A of D. Then the random variables, whose expectations are considered, can be based on maps Y_q:D→$\mathbb{R}$ given by $Y_q(i) = v_i(q)$ for each $q \in Q$. Certainly, these maps are measuruble, since the underlying σ-algebra in D is its power set PD.

selection probabilities λ_i will also be called to form the 'selection distribution', which is the probability distribution of the random variable U.

It should be mentioned that there are two formal approaches to such RSO-processes, which can be characterised as 'top down' and 'bottom up'. Their difference may be demontrated by the example of maps $Q \to \mathbb{R}$, which are learning curves as in Example 1.1. Following the reformulation of Classical Test Theory in terms of conditional expectations (Zimmermann, ###), the top down approach typically starts with a probability space for the entire RSO-process and considers random variables Y_q for every $q \in Q$. Then a map $v_i: Q \to \mathbb{R}$ characterising the i^{th} unit is defined such that $v_i(q)$ is the conditional expectation of Y_q under the condition that the respective unit is selected. Similarly, a map $Q \to \mathbb{R}$ for the entire RSO-process is made up by the unconditional expectations of Y_q. With the denotation Φ_π for this map, Equation (1.2) can be derived from well known properties of conditional expectations. Conversely, the bottom up approach starts with individual maps $v_i: Q \to \mathbb{R}$ and assumes an aggregation of the kind specified by Equation (1.2). Since this aggregation rule is common to both approaches, we can postpone a comparative discusssion (until Section ###) and introduce the aggregation rule like an axiom, whose applicability to a concrete process of aggregation is assumed for the rest of Section 2.

To formalise stability under this kind of aggregation, we have to introduce a suitable relation $H \subseteq \mathsf{P}V \times \Pi$. As in the tracking experiment, we start with a verbal formulation of the implication, which must hold for every RSO-process to make a subset S of the vocabulary set V stable under the considered kind of aggregation: If the element v_i characterising unit i is contained in S for every unit, which is given a non-zero probability of being selected, then the map $\Phi_\pi: Q \to \mathbb{R}$ characterising the RSO-process will also be an element of S. The if-clause in this implication is formalised by Assertion (iv) of the subsequent lemma, which will be basical for the following subsections. The lemma is proved in Section 6.6, and it abstracts away some concrete aspects of the hitherto assumed situation. (Similarly, the concrete properties of learning curves in the tracking experiment could be neglected in Lemma 1.4.)

Lemma 2.1: Let (V, Π, Φ, H, T) be an SSA with the following properties:

(i) V is a subset of a real vector space.

(ii) Π is the set of all finite sequences $\pi = \{(\lambda_i, v_i)\}_{i=1..n(\pi)}$ of ordered pairs (λ_i, v_i), where each v_i is an element of V, and every λ_i is a nonnegative real number such that $\sum_{i=1..n(\pi)} \lambda_i = 1$.

(iii) For every such π, Equation (1.2) holds with $n := n(\pi)$.

(iv) An ordered pair (S, π) with $S \subseteq V$ and $\pi \in \Pi$ is an element of H iff the implication $\lambda_i > 0 \Rightarrow v_i \in S$ holds for $i = 1..n(\pi)$.

Then T consists of all convex subsets of V.[15]

Certainly, the lemma cannot be considered as a noteworthy result: It is not more than a reformulation of the well known fact that a subset of a real vector space is closed under all convex linear combinations of its elements if and only if it is convex. But this very reformulation may be a sufficient motivation to speak of *stability under convex linear combinations* as another form of aggregation stability.

In fact, we will show this association of convexity and stability under aggregation in all results to

---

[15] Recall that the concept of convex sets has been defined such that the empty set is convex.

be reviewed from the pertinent literature. But it would be premature to conjecture a general equivalence of both properties in processes of aggregation, where the aggregation rule Φ can be interpreted as a generalisation of expectations to random variables in real vector spaces. We have already seen that some non-convex sets are stable under averaging. Conversely, Sections ### will present examples of convex sets, which are not stable under the pointwise forming of expectations.[16]

2.2 Parametric Families of Functions

A series of Psychological Bulletin articles by Sidman (1952), Bakan (1954) and Estes (1956) may be said to form the best known analytical study of aggregation stability in psychology. But it isn't only historical venerableness of the results of these authors, which deserves a detailed review. We will also show that an application of some basic and simple tools of modern mathematical analysis allows to dispense of some more complicated older ones. A reformulation of the results in this language will prepare generalisations in later sections.

2.2.1 Constributions by Sidman, Bakan and Estes

The historical series of studies by Sidman (1952), Bakan (1954) and Estes (1956) started with an observation of the first one of these authors, which was motivated by mathematical models of learning and other growth processes: An average curve is not necessarily of the same shape as the individual curves. For the notion of 'shape' underlying this statement and the closely related concept of a parametric family of functions, we will introduce formal definitions in Section 2.2.2. For a first approach, we may follow Sidman and exemplify them by the family of growth curves of the form

$$v(q) = M - M\,e^{-kq}, \tag{2.1}$$

where q is the value of a non-negative numerical independent variable (typically time or trial number), $M > 0$ the asymptote of the growth, and $k > 0$ determines the rate of approach to M. To prepare the definition of a vocabulary set, let Q be a set of more than three non-negative real numbers including 0.[17] Then the vocabulary set V consists of maps $v{:}Q\to\mathbb{R}$. Furthermore, the

[16] At first glance, the existence of convex sets, which are not stable under the pointwise forming of expectations, may seem to contradict Lemma 2.1. Note, however, that this lemma covers only expectations based on convex linear combinations.

[17] Sidman (1952) doesn't explicitly specify the set of values of the independent variable; but a figure suggests that he has in mind an interval starting at 0. Hower, his results hold also for every set Q of more than three non-negative real numbers including 0. For Q consisting of three numbers including 0, every strictly increasing map $v{:}Q\to\mathbb{R}$ with $v(0) = 0$ can be brought into the form of Equation (2.1) by a suitable choice of parameters M and k: Applied to the funtion values v(q) for the
(continued...)

property (i.e., the 'shape'), whose aggregation stability is studied, is represented by the set S of those maps $v{:}Q{\to}\mathbb{R}$, where Equation (2.1) is valid for all $q \in Q$ after a suitable parametrisation with $M > 0$ and $k > 0$. Then Sidman's main result referring to this example can be rewritten in the following statement: If $\{v_i\}_{i=1..n(\pi)}$ is a finite sequence of elements of S, then the 'average growth curve' (i.e., the map $\Phi_\pi{:}Q{\to}\mathbb{R}$ given by Equation (1.7)) is an element of S if and only if the k-parameters of all individual curves are identical.

We should note that this result has a logical form, which hasn't been provided for directly in the definition of an SSA: The set S is stable under aggregation only under an additional condition. Generally, 'conditional aggregation stability' will be introduced in Section 3.7. However, in the present example we can reformulate the result to fit into the system. of Definition 1.2. For every $\gamma > 0$, let S_γ be the set of all maps $v{:}Q{\to}\mathbb{R}$, where Equation (2.1) holds for every $q \in Q$ with $k = \gamma$ and an arbitrary, but constant parameter $M > 0$. Then Sidman' result can be rewritten: Every such set S_γ is stable under aggregation, but unions of several such sets are not.

To examine the association with convexity, consider first the positive example of a set S_γ with arbitrary $\gamma > 0$. For elements v' and v'' of S_γ with growth asymptotes M' and M'', and every real number $\lambda \in\ ]0, 1[$, the map

$$v := \lambda\, v' + (1-\lambda)\, v'' \tag{2.2}$$

fulfills Equation (2.1) for $M := \lambda\, M' + (1-\lambda)\, M''$ and $k = \gamma$; i.e., $v \in S_\gamma$. But if this holds for all elements v' and v'' of S_γ and every $\lambda \in\ ]0, 1[$, then S_γ is convex.

For the union of several such sets S_γ, let γ' and γ'' be different real numbers greater than 0, and let v' and v'' be elements of V such that $v' \in S_{\gamma'}$ and $v'' \in S_{\gamma''}$. Then Sidman's result for average curves with varying k-parameters implies that the map v given by Equation (2.2) with $\lambda := 0.5$ cannot be represented in the form of Equation (2.1). So every union of several sets S_γ is non-convex.

Sidman (1952) also analysed the stability under averaging of another 'shape' of growth curves: With parameters $m > 0$ and $x > 0$, the funtion value is $m\,q$ for $q \le x$, amd $m\,x$ for $q \ge x$. Again, this shape turned out to be stable under averaging only in samples with identical values of the parameter x. It can be left to the reader to verify the association with convexity for this example.

For genralisations of Sidman's results, Bakan (1954) and Estes (1956) considered functions $y = f(x, a, b, c, ...)$, where the dependent variable y as well as the 'parameters' a, b, c etc. are real numbers. The assumptions of both authors about the 'independent variable' x were different due to their different treatment of the function f. Bakan considered derivatives of f with respect to x, and for this approach x must be a real number. Since Estes analysed only partial derivatives of f for the parameter values, he didn't need any assumption about the values of the variable x. This will turn out very useful in generalisations to be presented later. Furthermore, Estes seems to have been the first one (in psychology) to consider stability under processes of aggregation based on expectations instead of averages. But since his main result consists in the identification of classes of parametric families granting aggregation stability, it suffices for the moment to introduce these classes as

[17] (...continued)
non-zero elements of Q, Equation (2.1) yields a system of two equations, which may be solved for M and k.

instances of stability under convex linear combinations and to postpone (until Section ###) a generalisation to SSAs based on other expectations. Finally, the results of Estes (1956) imply those of Bakan (1954) as special cases. For all these reasons, the results of Estes deserve a detailed review.

The central one refers to situations, where all second order partial derivatives of the function f(x, a, b, c, ...) for parameters (a, b, c etc.) are zero. In Section 2.2.3, we will give examples as well as another equivalent description of this class of functions, which doesn't need differential calculus, but for the moment we will only review the central result presented by Estes (1956) for functions of this class. It consists in the equation

$$\mathsf{E}(y) = f(x, \mathsf{E}(a), \mathsf{E}(b), \mathsf{E}(c), ...), \tag{2.3}$$

'where $\mathsf{E}()$ represents the mean, or expected values of the term in parentheses' (p. 137).[18] Furthermore, Estes analysed reparametrisations and transformations of the dependent variable y, which can bring a given function f(x, a, b, c, ...) into this class. Since his results can be extended after a reformulation in the theory of function spaces, we will first present a conceptual framework for this reformulation.

2.2.2 Conceptual Framework for Reformulations and Extensions

The subsequently presented conceptual framework for a reformulation of the problems discussed by Sidman (1952), Bakan (1954) and Estes (1956) may be motivated by an analogy with the application of vector and matrix algebra in multivariate statistics. Although it is possible to formulate this discipline entirely in terms of scalar algebra, it is helpful to combine real numbers in higher order objects forming just one of the magical 7 ± 2 chunks in human information processing (###Lit.). The resemblance is obvious, if we combine the parameters a, b, c etc. in an m-dimensional vector θ of real numbers, where m is the number of parameters. We will write $\theta(k)$ for the k^{th} element of a vectorial parameter θ.[19] Since the function f may imply operations, which are not defined for all real

---

[18] This explication (which is given only for a special example by Estes, 1956) deserves a direct quotation, since it seems to be the first time in psychology that the general discussion of stability under aggregation was extended from averages to expectations.

[19] Undoubtedly, the notation $\theta(k)$ for the k^{th} component of θ is somewhat unusual, and θ_k would be more familiar to most readers. But subscripts will be reserved to specify whole vectors in $\mathbb{R}^m$. So θ_i may denote the parameter vector underlying the particular map $v_i{:}Q\to\mathbb{R}$ of the i^{th} element of a sample, and then $\theta_i(k)$ will be the k^{th} component of θ_i. This notation isn't quite arbitrary. If we conceive maps $v{:}Q\to\mathbb{R}$ as elements of the vector space $\mathbb{R}^Q$, then the similarities between $\mathbb{R}^Q$ and $\mathbb{R}^m$ are emphasised, if we identify the vector space $\mathbb{R}^m$ with a function space $\mathbb{R}^K$, where K is the set of integers k = 1..m. Then an m-dimensional parameter vector θ is a map $\theta{:}K\to\mathbb{R}$, and its k^{th} component is the functional value $\theta(k)$, which is assigned to $k \in K$ by this map. In this view, it is only natural to write $v_i(q)$ for a function value assigned to $q \in Q$ by a particular map $v_i{:}Q\to\mathbb{R}$, and

(continued...)

numbers, the admissible range of the parameter θ, which will subsequently called the *parameter space* Θ, may be a proper subset of the vector space $\mathbb{R}^m$. In fact, an inclusion of Θ in a vector space $\mathbb{R}^m$ isn't necessary for all classes of functions to be analysed. So we start with a more general assumption.

A parametric family of maps (or functions) is a collection of maps of the same non-empty set Q into the set $\mathbb{R}$ of real numbers (or into another set Y[20]), the collection being related to another non-empty set - the parameter space Θ - by a relation with the following property: For every function $g:Q\to\mathbb{R}$ contained in the collection, there is at least one element θ of the parameter space Θ such that g is the (unique) function with parameter θ. This relation can be formalised equivalently by either one two functions:

- A map $f:Q\times\Theta\to\mathbb{R}$, where $f(q, \theta)$ is the function value which is assigned to the element q of the set Q by the function with parameter $\theta \in \Theta$, or
- a map $\psi:\Theta\to\mathbb{R}^Q$ such that $\psi(\theta)$ is the function $Q\to\mathbb{R}$ with parameter θ.

Undoubtedly, the representation of a parametric family of functions by a function $f:Q\times\Theta\to\mathbb{R}$ is more familiar, and for this reason, this function will be called the *representation function* of the family. E.g., the family of density functions of normal distributions is usually represented by a function $f(q, \theta)$, where q is a scale value, and θ an element of the parameter space

$$\Theta = \{(\mu, \sigma) \in \mathbb{R}^2: \sigma > 0\}, \tag{2.4}$$

whose components represent the expectation and the standard deviation of a member of the family. But sometimes the denotation $N(\mu, \sigma)$ is used to refer to the entire density function of a normal distribution with expectation μ and standard deviation σ, and then the letter N corresponds to the symbol ψ in the above formalisation, whereas the ordered pair $(\mu, \sigma) \in \Theta$ is the parameter θ. Indeed, this is usually regarded only as a notational convention and not as a map $N:\Theta\to\mathbb{R}^{\mathbb{R}}$; but since the view on ψ as a mapping of the parameter space Θ into the function space $\mathbb{R}^Q$ will turn out helpful in our later analyses, we introduce a verbal denotation: The map $\psi:\Theta\to\mathbb{R}^Q$, where $\psi(\theta)$ is the member with parameter θ of a considered family of maps $Q\to\mathbb{R}$, will subsequently be called the *parametrisation map* of the family, since it reflects the dependence of the members the family upon the parameter θ.

In fact, the two functions $f:Q\times\Theta\to\mathbb{R}$ and $\psi:\Theta\to\mathbb{R}^Q$ are only two ways of looking at the the same interrelationships. Before we formalise this issue, we should demonstrate it by an example, which is also analysed by Estes (1956). In Figure 2.1.A, the shaded surface is the graph of the function $f(q, \theta) = \log(\theta\, q)$ with $\Theta = [e^{-1}, e]$ and $Q =]0, 2]$. The curved lines on the surface for constant values of θ (i.e., those running roughly from left to right) represent members of the parametric family of maps $Q\to\mathbb{R}$, which is based on the function f. The above terminology is demonstrated by

---

[19] (...continued)
$\theta_i(k)$ for the k^{th} component of a particular vector θ_i in $\mathbb{R}^m$.

[20] The subsequent formalisation can very well be generalised to situations, where the role of the set $\mathbb{R}$ is replaced by an arbitrary non-empty set Y. But since later applications will typically refer to situations with $Y = \mathbb{R}$, a formalisation restricted to such situations may be easier to follow up.

the statement that the curve for a constant value of θ plots the member $\psi(\theta)$ of a parametric family of maps $Q \to \mathbb{R}$ with representation function f. There is still a third way of looking at the same situation, which is represented by the curves for constant values of q: They reflect the dependence of the function value for a given element q of Q upon the value of the parameter θ.

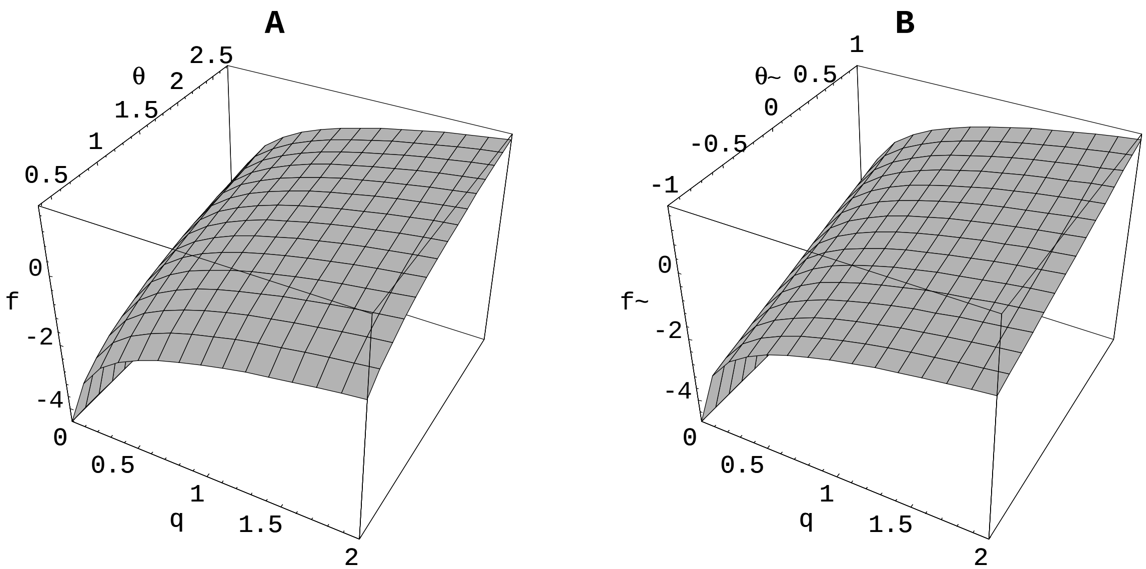

Figure 2.1: Graphs for $f(q, \theta) = \log(\theta\ q)$ and $f^{\sim}(q, \theta^{\sim}) = \theta^{\sim} + \log(q)$. The lines on the shaded surfaces represent partial maps.

The understanding of the three views of the situation may be supported by another analogy to matrix algebra. A matrix can be considered as a two-dimensional array of real numbers, but also as an array of column vectors or of row vectors. Similarly, the shaded surface in Figure 2.1.A can be viewed as an arrangement of points (one for each point (q, θ) of the plain $Q \times \Theta$), or as a family of curves for maps $Q \to \mathbb{R}$ (one for each $\theta \in \Theta$) or as a family of curves for maps $\Theta \to \mathbb{R}$ (one for each $q \in Q$). A generalisation of 'row vectors' and 'column vectors', which can be applied to such situations, is provided by the concept of *partial maps*.[21] A curve for a constant parameter θ represents a partial map, which is denoted as $f(. , \theta)$, whereas the curve for a constant value of q represents the partical map $f(q, .)$. The underlying notational rule is obvious: In the expression $f(q, \theta)$, a dot replaces the independent variable of the partial map under consideration, whereas the symbol for the fixed value of the other argument of f is maintained.

The difference between the two kinds of partial maps is well known (although with different notations) from the area of parametric families of probability distribution functions, where $f(q, \theta)$ may denote the probability (or probability density) for data q under parameter θ. Then the distribution function with paramter θ is given by the partial map $f(. , \theta)$, whereas $f(q, .)$ is the likelihood function for data q.

[21] See e.g. Dieudonnée (1985, p. 21) for the concept of partial maps and the dot notation. More generally, for a map $f: (X \times Y) \to \Phi$ and a fixed $x \in X$, the partial map $f(x, .)$ is the map $Y \to Z$ assigning the function value $f(x, y)$ to every $y \in Y$. Similarly for fixed $y \in Y$, the partial map $f(. , y)$ is the map $X \to Z$ assigning the function value $f(x, y)$ to every $x \in X$.

Up to this point, the treatment of the sets Q and Θ has been entirely symmetric; but it becomes asymmetric if we recall that the function f has been introduced to represent parametric families of maps $Q \to \mathbb{R}$ and not $\Theta \to \mathbb{R}$.[22] Whereas Estes (1956) as well as his predecessors Sidman (1952) and Bakan (1954) treated the values of the dependent variable of a function f as single real numbers, it will turn out useful to consider the members of the family as whole objects, namely as elements of the function space $\mathbb{R}^Q$. Then the parametrisation map $\psi:\Theta \to \mathbb{R}^Q$ reflects the dependence of the entire functions upon the value of the parameter. E.g., in the example plotted in Figure 2.1.A, the parametrisation map $\psi:\Theta \to \mathbb{R}^Q$ represents changes of the curves for partial maps $f(. ,\theta)$, which occur, if we step along the θ-axis.

The claim that the representation function and the parametrisation map are two equivalent formalisations of the same relation becomes more explicit by two equations, which allow to derive one of the functions, if the other one is given. For a representation function $f:Q\times\Theta \to \mathbb{R}$, the correspnding parametrisation map ψ is given by the equation

$$\psi(\theta) = f(. , \theta) \tag{2.5}$$

for every $\theta \in \Theta$. Conversely, if a parametrisation map $\psi:\Theta \to \mathbb{R}^Q$ is given, then

$$f(q, \theta) = \psi(\theta)(q) \tag{2.6}$$

must hold for every $(q, \theta) \in Q\times\Theta$, where $\psi(\theta)(q)$ denotes the real number, which is assigned to the element q of the set Q by the member $\psi(\theta)$ of the considered family.

Fortunately, it isn't necessarys to choose between the two views and to declare one of the functions as fundamental and the other one as derived. For reasons of historical continuity, we will follows Sidman (1952), Bakan (1954) and Estes (1956) and introduce some classes of parametric families of functions by properties of the representation function f; but in most cases it will turn out useful to reformulate them by equivalent properties of the parametrisation map.

Concretely, the gain of this approach will become apparent in its application, but we can already give a general idea of it. At first glance, it may seem to be only a notatational convenience that the

---

[22] Readers working with programing languages derived from Pascal may ask whether this asymmetry wouldn't be reflected better, if the order of the arguments in the representation function would be reversed. Indeed, an array [1..n, 1..m] of integer can be reconceived in these languages as an array[1..n] of array[1..m] of integer, but not as an array[1..m] of array[1..n] of integer. Furthermore, it could be considered after this change to write $f(\theta)$ instead of $\psi(\theta)$, and the equation $f(\theta, q) = f(\theta)(q)$ would emphasise again the analogy with two-dimensional arrays, where the expressions x[i, j] and x[i][j] refer to the same element. But the separate symbol ψ for the parametrisation map will allow simple formulations like 'ψ is linear' instead of 'f - considered as a map $\Theta \to \mathbb{R}^Q$ - is linear'. If the distinction of both views by different symbols f and ψ is maintained, then the common practice to write the argument (like q) of functions belonging to a family before the parameter θ seems to be a stronger argument than conformity with notational practice in programming languages.

introduction of the parametrisation map ψ enables the application of a commonly used convention[23]: For every subset S of the parameter space Θ, we can write $\psi(S)$ for the image of S under the map ψ, i.e, for the set of all maps $v:Q\rightarrow\mathbb{R}$, where the equation $v = \psi(\theta)$ holds for some $\theta \in S$. In particular, $\psi(\Theta)$ is the set of all members of the parametric family under consideration. But the view on members of the parametric family as results of a map $\psi:\Theta\rightarrow\mathbb{R}^Q$ will also be helpful, if we analyse the stability under aggregation of the set $\psi(S)$ for situations, where S is a proper subset of the parameter space Θ. Readers with an elementary familiarity with the theory of real vector spaces may anticipate applications of basic results of this discipline, which are enabled by the view upon a set of maps $Q\rightarrow\mathbb{R}$ as an image of the parameter space Θ or a subset of it under the parametrisation map $\psi:\Theta\rightarrow\mathbb{R}^Q$. E.g., since the image of a convex set under a linear or affine map is convex, the image of a convex subset of the parameter space under a linear or affine parametrisation map will be convex, which implies its stability under convex linear combinations. (See Lemma 2.1 for this conclusion.)

The entire approach can be summarised in the following definition:

Definition 2.2: If Q and Θ are non-empty sets, then a set S of maps $Q\rightarrow\mathbb{R}$ is a *parametric family of maps* $Q\rightarrow\mathbb{R}$ with *parameter space* Θ, *representation function* f and *parametrisation map* ψ, iff f is a function $f:Q\times\Theta\rightarrow\mathbb{R}$, and ψ a map $\psi:\Theta\rightarrow\mathbb{R}^Q$ such that $S = \psi(\Theta)$, and

$$f(q, \theta) = \psi(\theta)(q) \tag{2.7}$$

for every $(q, \theta) \in Q$.

Note that Equation (2.5) follows immediately from this definition, since it is nothing more than a summary (for every $q \in Q$ and fixed $\theta \in \Theta$) of the claim of Equation (2.7).

It is left to explicate in this terminology the word *shape* in the initial statement of Sidman (1952), who observed that an average curve is not necessarily of the same shape as the individual curves. It refers to the common geometric property of all members of a parametric family, i.e., of all maps $\psi(\theta)$ resp. $f(. , \theta)$. In Figure 2.1.A, it can be described as being identical with $f(1, q) = \log(q)$ up to a vertical translation by an amount in the interval $[-1, 1]$. More generally, a shape associated with a parameteric family of maps may frequently be specified by a description of the geometrical transformations which may be applied to the graph of some prototypical member of the family to obtain the graphs of other members.[24]

2.2.3 The Basic Result of Estes: Reformulation and Extensions

We are now prepared to reformulate the main results of Estes (1956) in the conceptual framework of Section 2.2.2. This author distinguished three classes of functions, which can be reconceived as maps $f:Q\times\Theta\rightarrow\mathbb{R}$ with an arbitrary non-empty set Q and a parameter space Θ, which

---

[23] Generally, for a map $f:X\rightarrow Y$ and a subset A of X, the notation f(A) refers to the set $\{f(x): x \in A\}$, i.e., the set of those elements of Y which are equal to f(x) for some $x \in A$. This set is called the image of the set A under the map f.

[24] Note, however, that these transformations may go beyond those commonly denoted as 'shape-conserving' (like translation, rotation or proportional stretching).

is typically (but not always necessarily) a subset of $\mathbb{R}^m$ for some natural number m.

Class A is the most basic one, and results for other classes are side results ('corollaries') based on transformations of parameters or dependent variables. In Class A, the parameter space Θ is identical with the vector space $\mathbb{R}^m$ for some natural number m. Estes (1956) described this class in two ways, whose relationship will be discussed later. The following redefinition (whose equivalence with the description in Section 2.2.1 will be stated in Theorem 2.4) opens a more direct interface to the intended reformulation of Estes' results:

Definition 2.3: A function $f:Q\times\mathbb{R}^m\to\mathbb{R}$ (with a non-empty set Q and a natural number m) belongs to Class A iff there is a family $\{\gamma_k\}_{k=0..m}$ of maps $\gamma_k:Q\to\mathbb{R}$ such that the equation

$$f(q, \theta) = \gamma_0(q) + \sum_{k=1..m} \theta(k) \cdot \gamma_k(q) \tag{2.8}$$

holds for every $(q, \theta) \in Q\times\mathbb{R}^m$.

The requirement imposed upon the representation function f by this definition can be reformulated in terms of the parametrisation map $\psi:\Theta\to\mathbb{R}^Q$ derived from f by Equation (2.5): There must be a family $\{\gamma_k\}_{k=0..m}$ of maps $\gamma_k:Q\to\mathbb{R}$ such that the equation

$$\psi(\theta) = \gamma_0 + \sum_{k=1..m} \theta(k)\ \gamma_k \tag{2.9}$$

holds for every $\theta \in \mathbb{R}^m$.

The main result of Estes (1956) for parametric families with a representation function f of this class has been reported by Equation (2.3) in terms of single real numbers. More explicitly, that equation holds for every conceivable value x of the independent variable. The following reformulation opens an interface to the theory of function spaces: If $f:Q\times\mathbb{R}^m\to\mathbb{R}$ is a function of Class A and $\psi:\Theta\to\mathbb{R}^Q$ the respective parametrisation map given by Equation (2.5), then the result of a convex linear combination $\sum_{i=1..n} \lambda_i\ \psi(\theta_i)$ is identical with the parametrisation map $\psi(\theta^*)$, where θ^* is the result of a linear combination of the parameters with identical coefficients λ_i; i.e.:

$$\sum_{i=1..n} \lambda_i\ \psi(\theta_i) = \psi(\sum_{i=1..n} \lambda_i\ \theta_i) \tag{2.10}$$

The most important aspect of the result for our concept of aggregation stability (Definition 1.2) is the existence of *some* element θ^* of $\mathbb{R}^m$ such that the result of the convex linear combination on the left hand side of Equation (2.10) is identical with $\psi(\theta^*)$ - i.e., a member of the family $\psi(\mathbb{R}^m)$. So the family as a whole is stable under convex linear combinations. In fact, this result could have been derived directly from Lemma 2.1 (with $\psi(\mathbb{R}^m) \subseteq V \subseteq \mathbb{R}^Q$), since it is easily verified that the set $\psi(\mathbb{R}^m)$ is convex, if the underlying representation function f is of Class A.[25] But beyond the mere existence of a suitable parameter θ^* (which would be sufficient for the aggregation stability of the parametric family), this parameter value can also be specified as the result of an anologous linear

[25] Proof: For elements v_1 and v_2 of $\psi(\mathbb{R}^m)$ with parameters θ_1 and θ_2, and a real number $\lambda \in]0, 1[$, Equation (2.9) can be used to derive the second equality in

$$\lambda\ v_1 + (1-\lambda)\ v_2 = \lambda\ \psi(\theta_1) + (1-\lambda)\ \psi(\theta_2) = \psi(\lambda\ \theta_1 + (1-\lambda)\ \theta_2).$$

Obviously, $\psi(\lambda\ \theta_1 + (1-\lambda)\ \theta_2)$ is an element of $\psi(\mathbb{R}^m)$.

combination of the individual parameters θ_i. This result implies that Definition 1.2 abstracts away an aspect of aggregation, which may be helpful in further applications: An aggregation rule can imply the applicability of the underlying process of aggregation (e.g., convex linear combinations) to mathematical objects, which are not mentioned explicitly in the aggregation rule (e.g., to the parameters θ_i). Since Definition 1.2 puts up with this price of generality, we should note down to pay attention to such implications. In Sections 3.5, 4.5 and 4.7, they will turn out to be noteworthy not only for their own sake, but also as useful tools in the analysis of aggregation stability. We will also demonstrate this tool function in generalisations of results reported by Estes (1956) for other classes of representation functions underlying parametric families.

Up to this point, we have followed Estes (1956): The description of Class A by a function $f{:}Q\times\mathbb{R}^m\to\mathbb{R}$ with the structure of Equation (2.8) has been introduced only as a sufficient condition for the validity of Equation (2.10) with suitable coefficients λ_i. More precisely, Estes gave two descriptions of the class, which are not entirely equivalent. Their relationship will be cleared in comments to Assertions (i) and (ii) of the following Theorem 2.4, whose subsequent denotation as the Estes-Theorem has to be undertstood as an acknowledgement of the fact that its core is a reformulation of Estes' most basic result in the framework of function spaces. Assertions (iii), (iv), (v) and (vi) of the theorem are further equivalent descriptions of Class A. A proof of the theorem is given in Section 6.7.

Estes-Theorem 2.4: Let Q be a nonempty set, m a natural number, and $f{:}Q\times\mathbb{R}^m\to\mathbb{R}$ an arbitrary map. Furthermore, let a map $\psi{:}\mathbb{R}^m\to\mathbb{R}^Q$ be given such that Equation (2.5) holds for every $\theta \in \mathbb{R}^m$. Finally, let vectors $\{\eta_k\}_{k=0..m}$ in $\mathbb{R}^m$ be defined such that the j^{th} component of η_k is 1 for j = k, and 0 for $j \neq k$, and let γ_z be the zero-element of $\mathbb{R}^Q$.[26] Then the following assertions (i) through (vi) are equivalent:

(i) For every $q \in Q$ and $\theta \in \mathbb{R}^m$, the equation
$$(\partial^2 / \partial\theta(k)\, \partial\theta(k'))\, f(q, \theta) = 0 \tag{2.11}$$
holds for the second order partial derivative[27] of f with respect to the components $\theta(k)$ and $\theta(k')$ of θ.

(ii) There is a family $\{\gamma_k\}_{k=0..m}$ of maps $\gamma_k{:}Q\to\mathbb{R}$ such that Equation (2.8) holds for every $(q, \theta) \in Q\times\mathbb{R}^m$.

(iii) The map ψ is affine.

(iv) The equation
$$\psi(\textstyle\sum_{i=1..n} \lambda_i\, \theta_i) = (1 - \sum_{i=1..n} \lambda_i)\, \psi(\eta_0) + \sum_{i=1..n} \lambda_i\, \psi(\theta_i) \tag{2.12}$$
holds for all linear combinations of elements θ_i of $\mathbb{R}^m$ with coefficients λ_i.

(v) Equation (2.10) holds for all linear combinations of maps $\psi(\theta_i)$ with $\theta_i \in \mathbb{R}^m$ and non-

[26] In other words, η_0 is the zero-vector of $\mathbb{R}^m$, and for k =1..m, the vector η_k is identical with the k^{th} column of an m×m identity matrix.

[27] It goes almost without saying that Equation (2.11) implies the existence of the second order partial derivate on its left hand side.

negative[28] coefficients λ_i such that $\sum_{i=1..n} \lambda_i = 1$.

(vi) The equation

$$\psi(\lambda\, \theta_1 + (1-\lambda)\, \theta_2) = \lambda\, \psi(\theta_1) + (1-\lambda)\, \psi(\theta_2) \tag{2.13}$$

holds for all elements θ_1 and θ_2 of $\mathbb{R}^m$ and every real number $\lambda \in [0, 1]$.

If these properties are given, then the maps γ_k in Assertion (ii) are uniquely specified for given maps f and ψ by

$$\gamma_0 = \psi(\eta_0), \tag{2.14}$$

and

$$\gamma_k = \psi(\eta_k) - \psi(\eta_0) \tag{2.15}$$

for k = 1..m. Conversely, for given maps γ_k, the equation

$$\psi(\theta) = \gamma_0 + \sum_{k=1..m} \theta(k)\, \gamma_k \tag{2.16}$$

holds for every $\theta \in \mathbb{R}^m$.

Finally, the map ψ is linear iff $\gamma_0 = \gamma_z$, and it is injective iff the vectors $\{\gamma_k\}_{k=1..m}$ are linearly independent[29] vectors in $\mathbb{R}^Q$.

The equivalences which are claimed by the theorem are worth some comments. Assertions (i) and (ii) are two specifications of the most basic class of functions $f{:}Q\times\mathbb{R}^m\to\mathbb{R}$ analysed by Estes (1956). In his Mathematical Note 1, he derived his basic result for Class A (our Equation (2.10)) for functions with the property described in Assertion (i). But in the main text of his article, Estes gave a description and examples for his Class A, where the map γ_0 is missing. Of course, such functions can be subsumed under the description of Class A by Assertion (ii) as special cases with $\gamma_0 = \gamma_z$ (i.e., $\gamma_0(q) = 0$ for every $q \in Q$). We can leave it open whether Estes overlooked that the premissa of his Mathematical Note 1 covers more functions than the description in the main text, or whether he deliberately confined the presentation in the main text to a subclass. Indeed, functions belonging to this subclass have a convenient property, which is stated at the end of the theorem: The linearity of the map ψ. Furthermore, we will show that every map $f{:}Q\times\mathbb{R}^m\to\mathbb{R}$ with the properties stated in Assertion (ii) can be brought into this convenient subclass by a suitable reparametrisation.

Although linearity of the map ψ is a convenient property, many mathematical results for such maps can be used to derive (as corollaries) similar properties of affine maps. So the equivalence of the original descriptions of Class A by Assertions (i) and (ii) with affinity of the map ψ (Assertion

[28] The premissa of non-negative coefficients λ_i in Assertion (v) of Theorem 2.4 could be canceled, since the factor $1 - \sum_{i=1..n} \lambda_i$ in Equation (2.12) is zero for all linear combinations with coefficients λ_i summing up to 1. But the description of Class A by Assertion (v) has been added to the theorem to show that membership in Class A is a necessary condition for the central result of Estes (1956) for this class.

[29] The concepts of linear dependence and independence in real vector spaces are the same as in the well known $\mathbb{R}^n$. A vector x is linearly dependent upon the elements of a subset S of a real vector space iff it can be represented as a linear combination of elements of S. In particular, the zero-element of a vector space in linearly dependent upon the elements of every (empty or non-empty) subset of the space. The elements of a subset S of a real vector space are linearly dependent iff some $x \in S$ is linear dependent upon the elements of $S\backslash\{x\}$, and otherwise the elements of S are linearly independent.

(iii)) is noteworthy in itself. As an example of a corollary for affine maps derived from properties of linear maps, consider Equation (2.12). For linear maps, the term $(1 - \sum_{i=1..n} \lambda_i)\ \psi(\eta_0)$ can be neglected, since $\psi(\eta_0)$ is the zero-vector of $\mathbb{R}^Q$. (A linear map E→F between arbitrary real vector spaces E and F maps the zero-element of E into the zero-element of F, and η_0 is the zero-element of $\mathbb{R}^m$.)

Assertion (v) resumes our reformulation of the main result of Estes (1956) by Equation (2.10). The equivalence of this property with a description of Class A in Assertion (ii) is noteworthy for several reasons. Since this equivalence is proved in Section 6.7 without any reference to differential calculus, it demonstrates the effectiveness of an application of the theory of real vector spaces in this realm. In particular, this equivalence is a stronger result than a mere implication. (Estes (1956) proved only that membership in Class A is a sufficient condition for this result.) At first glance, one could be even tempted to question whether this equivalence doesn't solve the entire problem of stability under an aggregation, which is based on pointwise expectations of maps $Q\to\mathbb{R}$. Indeed, this conjecture is well-aimed, if the subset of $\mathbb{R}^Q$, whose aggregation stability is studied, is a convex subset of a finite dimensional subspace of $\mathbb{R}^Q$. (See ###.) Without going into details, the relevance of finite dimension can be derived from the fact that the image of a finite dimensional vector space (like $\mathbb{R}^m$) under an affine (resp. linear) map ψ is a finite dimensional affine (resp. linear) space.

Finally, the description of Class A by Assertion (vi) has been added to have an easily checked criterion of membership in this class for situations, where the representation function $f:Q\times\mathbb{R}^m\to\mathbb{R}$ underlying a parametric family of maps $Q\to\mathbb{R}$ isn't fully specified. It may be useful, if not one such function, but an entire class is analysed.

It should also be noted, that membership of the function f in Class A is necessary only for the applicability of Equation (2.10) to *all* linear combinations with non-negative coefficients λ_i summing up to 1. If aggregation is confined to averaging, the coefficients λ_i will always be rational numbers. But even then, membership in Class A is necessary for the applicability of Equation (2.10) under a weak additional assumption, which is given in most parametric models applied in psychology. If we consider the function value $f(q, \theta)$ for a fixed $q \in Q$ and vary only one component $\theta(k)$ of the parameter θ, then $f(q, \theta)$ is usually a continuous function of $\theta(k)$. In such situations, the validity of Assertion (vi) for all rational numbers $\lambda \in [0, 1]$ implies that it holds also for all real numbers $\lambda \in [0, 1]$.[30] So we obtain an additional result: For functions $f:Q\times\mathbb{R}^m\to\mathbb{R}$, where $f(q, \theta)$ is a continuous function of $\theta(k)$ for every $q \in Q$ and $k = 1..m$, the membership of f in Class A is sufficient *and* necessary for the central result of Estes (1956), even if aggregation is confined to averaging.

Before we comment the claims of Estes-Theorem 2.4 following its Assertion (vi), we will extend the implications for stability under aggregation, which can be drawn from Equation (2.10). If S is a convex subset of $\mathbb{R}^m$ and all parameters θ_i in Equation (2.10) are elements of S, then the same holds for the convex linear combination $\sum_{i=1..n} \lambda_i\ \theta_i$. With the denotation $\psi(S)$ for the image of S under the map ψ, we can also say: The result of a convex linear combination $\sum_{i=1..n} \lambda_i\ v_i$ of maps $v_i:Q\to\mathbb{R}$,

[30] Proof: For given elements θ_1 and θ_2 of $\mathbb{R}^m$ and $q \in Q$, define a map $h:[0, 1]\to\mathbb{R}$ by $h(\lambda) := f(\lambda\ \theta_1 + (1-\lambda)\ \theta_2, q)$. If $f(q, \theta)$ is a continuous function of every component $\theta(k)$, then $h(\lambda)$ is a continuous function of λ. This implies: If the equation $h(\lambda) = \lambda\ f(\theta_1, q) + (1\text{-}\lambda)\ f(\theta_2, q)$ holds for every rational $\lambda \in [0, 1]$, then it can be extended to every real $\lambda \in [0, 1]$.

which are elements of $\psi(S)$, is an element of $\psi(S)$. In other words, not only the entire family $\psi(\mathbb{R}^m)$ is stable under convex linear combinations, but every set $\psi(S)$, where S is a convex subset of $\mathbb{R}^m$.

Two applications of this extension of the result reported by Estes (1956) are worth mentioning. Estes didn't pay much attention to the parameter space Θ and identified it (implicitly) with the entire vector space $\mathbb{R}^m$. But for some applications, only a proper subset of $\mathbb{R}^m$ is available for the parameter θ of meaningful maps $\psi(\theta)$. E.g., if partial maps are probability distribution functions, the maps $\psi(\theta)$ for some $\theta \in \mathbb{R}^m$ will have meaningless negative function values. So the parameter space must be restricted, and then it is helpful to identify the set Θ of meaningful parameter values with the set S in the above result. (See Section 2.3 for applications.)

Particularly in empirical research, the set S may also be a proper subset of the parameter space Θ. E.g., if Θ contains some vectors θ with $\theta(1) \geq \theta(2)$, an empirical hypothesis may nevertheless claim $\theta(1) < \theta(2)$ for all empirical individual maps v_i:Q→$\mathbb{R}$.[31] So S may be the set of all elements of Θ with the property $\theta(1) < \theta(2)$, and it is easily verified that this set is convex.

In fact, we have already studied another example of a set $\psi(S)$, where S is a proper subset of $\mathbb{R}^m$. For the tracking experiment, we considered the set S_2 of strictly increasing, negatively accelerated quadratic maps v:Q→[0, 100], where Q was the set of trial numbers 1 through 10. The property of being quadratic can be based on a function $f:Q\times\mathbb{R}^3\to\mathbb{R}$ given by $f(q, \theta) = \theta(1) + \theta(2)\, q + \theta(3)\, q^2$. In other words, Equation (2.8) applies with m = 3, $\gamma_0(q) = 0$, $\gamma_1(q) = 1$, $\gamma_2(q) = q$, and $\gamma_3(q) = q^2$. However, the set S_2 contains only maps with additional properties, which can be expressed in terms of components of the parameter θ by the inequalities

$$\theta(3) < 0,$$

$$\theta(1) + 9\, \theta(2) + 9^2\, \theta(3) < \theta(1) + 10\, \theta(2) + 10^2\, \theta(3) \leq 100,$$

and

$$\theta(1) + \theta(2) + \theta(3) \geq 0.$$

The first inequality selects the negatively accelerated quadratic curves. Given this property, a curve is strictly increasing iff the function value is smaller for q = 9 than for q = 10. Finally, if the curve is strictly increasing, then the function values for all trials are contained in the interval [0, 100] iff the last one isn't greater than 100 and the first one isn't negative. Using the denotation S for the set of $\theta \in \mathbb{R}^3$ fulfilling these inequalities, our set S_2 is identical with $\psi(S)$, and its stability under convex linear combinations follows from the convexity of S, which can be checked by the reader.

A more direct way to the stability under convex linear combination of a set $\psi(S)$ with convex S may demonstrate the effectiveness of the vector space approach to problems of this kind. It is an elementary result of this theory that the image of a convex set under a linear or affine map is convex and hence closed under convex linear combinations of its elements. This property can be applied immediately to our situation, since Estes-Theorem 2.4 tells that the representation map $\psi:\mathbb{R}^m\to\mathbb{R}^Q$

---

[31] E.g., if the parameters $\theta(1)$ and $\theta(2)$ represent performance levels in early and late phases of a process, then the empirical hypothesis $\theta(1) < \theta(2)$ means that performance improves during the process.

is affine iff the representation function $f:Q\times\mathbb{R}^m\to\mathbb{R}$ is a member of Class A.

One may be tempted to suppose that convexity of S is also a necessary condition of the convexity of $\psi(S)$; but this holds only if the map ψ is injective. Since this issue will be of some relevance in later applications, we should analyse an example. For $Q := \mathbb{R}$, let a map $f:Q\times\mathbb{R}^3\to\mathbb{R}$ of Class A be given by

$$f(q, \theta) = \theta(1)\, q + \theta(2)\, q^2 + \theta(3)\,(2\, q + q^2) + q^3, \tag{2.17}$$

and let a subset S of $\mathbb{R}^3$ consist of all vectors in $\mathbb{R}^3$ with $\theta(1) \neq 0$. This set is non-convex: The vectors $\theta_1 := (-1, 2, 5)$ and $\theta_2 := (1, 0, 1)$ are elements of S, but the vector θ_3 given by a convex linear combination of θ_1 and θ_2 as

$$\theta_3 := 0.5\, \theta_1 + 0.5\, \theta_2 = (0, 1, 3) \tag{2.18}$$

is not contained in S, since $\theta_3(1) = 0$. Now let maps $v_i:Q\to\mathbb{R}$ for i = 1..3 be given as $v_i := \psi(\theta_i)$; i.e.:

$$v_1(q) = (-1)\, q + 2\, q^2 + 5\,(2\, q + q^2) + q^3 = 9\, q + 7\, q^2 + q^3,$$

$$v_2(q) = \quad 1\, q + 0\, q^2 + 1\,(2\, q + q^2) + q^3 = 3\, q + 1\, q^2 + q^3,$$

and

$$v_3(q) = \quad 0\, q + 1\, q^2 + 3\,(2\, q + q^2) + q^3 = 6\, q + 4\, q^2 + q^3.$$

Obviously Equation (2.10) holds for this situation, since $v_3 = 0.5\, v_1 + 0.5\, v_2$. But although the map $v_3:Q\to\mathbb{R}$ has been derived from the vector θ_3, which is not contained in S, this map v_3 *is* an element of $\psi(S)$, since there are elements θ of S such that $v_3 = \psi(\theta)$. Take e.g. the vector $\theta_4 := (4, 3, 1)$ and consider the map $v_4 := \psi(\theta_4)$, i.e.,

$$v_4(q) = \quad 4\, q + 3\, q^2 + 1\,(2\, q + q^2) + q^3 = 6\, q + 4\, q^2 + q^3.$$

Although v_3 and v_4 are derived from different vectors θ_3 and θ_4, they are not two different maps, but one and the same, since $v_3(q) = v_4(q)$ for every $q \in Q$.[32] So this map *is* an element of $\psi(S)$. (Recall that $\psi(S)$ is the set of all maps $v:Q\to\mathbb{R}$, where the equation $v = \psi(\theta)$ holds for *some* element θ of S.) We can generalise the result in a statement, which can be verified by the reader: For every vector θ in $\mathbb{R}^3$ with $\theta(1) = 0$ (i.e., $\theta \notin S$) and every non-zero real number ξ, the vector θ^* with components $\theta^*(1) = \theta(1) + 2\,\xi$, $\theta^*(2) = \theta(2) + \xi$, and $\theta^*(3) = \theta(3) - \xi$ is an element of S such that $\psi(\theta^*) = \psi(\theta)$.

Problems of this kind can arise, if and only if the map ψ is non-injective; i.e., if there are non-

[32] For the interpretation of this identity, note that a map isn't identical with a formula describing it. Two maps are identical iff they assign the same values of the dependent variable to the values of the independent variable, even if formulas used to describe or to derive them are more dissimilar than in the example.

identical elements θ_1 and θ_2 of $\mathbb{R}^m$ such that $\psi(\theta_1) = \psi(\theta_2)$. Then there are convex sets S such that $\theta_1 \notin S$, $\theta_2 \in S$, leading to $\psi(\theta_1) \in \psi(S)$ due to the identity of the maps $\psi(\theta_1)$ and $\psi(\theta_2)$.

It may be argued that this non-injectivity is highly untypical for the parametric families of functions, which are commonly used in psychology. Indeed, the example has been introduced mainly as a starting point for a reparametrisation leading to an injective map $\psi^{\sim}$. A simple rearrangement of terms in Equation (2.17) leads to

$$f(q, \theta) = (\theta(1) + 2\,\theta(3))\, q + (\theta(2) + \theta(3))\, q^2 + q_3. \tag{2.19}$$

We can reformulate this equation in terms of the partial maps $\psi(\theta)$: For parameter vectors θ_1 and θ_2 with $\theta_1(1) + 2\,\theta_1(3) = \theta_2(1) + 2\,\theta_2(3)$ and $\theta_1(2) + \theta_1(3) = \theta_2(2) + \theta_2(3)$, the partial maps $\psi(\theta_1)$ and $\psi(\theta_2)$ are identical. But if the partial map $\psi(\theta)$ is completely specified by the numbers $\theta(1) + 2\,\theta(3)$ and $\theta(2) + \theta(3)$, we can take these numbers as components of a new parameter vector $\theta^{\sim}$ in $\mathbb{R}^2$ and consider a map $f^{\sim}: Q\times\mathbb{R}^2\to\mathbb{R}$ given by

$$f^{\sim}(q, \theta^{\sim}) = \theta^{\sim}(1)\, q + \theta^{\sim}(2)\, q^2 + q^3 \tag{2.20}$$

as the representation function of a family of maps $Q\to\mathbb{R}$, which are partial maps $f^{\sim}(. , \theta^{\sim})$. Again it is useful to represent the dependence of these partial maps upon the parameter $\theta^{\sim}$ by a map $\psi^{\sim}:\mathbb{R}^2\to\mathbb{R}^Q$, which is defined by

$$\psi^{\sim}(\theta^{\sim}) := f^{\sim}(. , \theta^{\sim}). \tag{2.21}$$

Then the relationship between the maps ψ and $\psi^{\sim}$ can be written as

$$\psi(\theta) = \psi^{\sim}(t(\theta)) \tag{2.22}$$

for every $\theta \in \mathbb{R}^3$, where $t:\mathbb{R}^3\to\mathbb{R}^2$ is a map representing the parameter transformation; i.e., $t(\theta)$ is the vector $\theta^{\sim}$ in $\mathbb{R}^2$ with components $\theta^{\sim}(1) = \theta(1) + 2\,\theta(3)$, and $\theta^{\sim}(2) = \theta(2) + \theta(3)$. In other words, the reparametrisation is based on a representation of the map ψ as a concatenation $\psi^{\sim} \circ t$.

Since reparametrisations will also be applied in other situations, it is useful to generalise the logic underlying Equation (2.22) in the following lemma, which is proved in Section 6.8. Because of its general relevance for reparametrisations, it is subsequently called the Reaparametrisation-Lemma.

Lemma 2.5: Let non-nonempty sets Q, Θ and $\Theta^{\sim}$ as well as maps $t:\Theta\to\Theta^{\sim}$, $f:Q\times\Theta\to\mathbb{R}$ and $f^{\sim}:Q\times\Theta^{\sim}\to\mathbb{R}$ be given such that the equation

$$f(q, \theta) = f^{\sim}(q, t(\theta)) \tag{2.23}$$

holds for every $(q, \theta) \in Q\times\Theta$. Furthermore, let maps $\psi:\Theta\to\mathbb{R}^Q$ and $\psi^{\sim}:\Theta^{\sim}\to\mathbb{R}^Q$ be defined by Equations (2.5) and (2.21). Then Equation (2.22) holds for every $\theta \in \Theta$; i.e.,

$$\psi = \psi^{\sim} \circ t, \tag{2.24}$$

and

$$\psi(\Theta) = \psi^{\sim}(t(\Theta)) \tag{2.25}$$

With $Y := \mathbb{R}$, $\Theta := \mathbb{R}^3$, and $\Theta^{\sim} := \mathbb{R}^2$, we obtain the situation of the preceeding example. Now recall that the reparametrisation has been applied to get rid of some unconvenient properties of a non-injective map ψ. The reader is invited to verify on her or his own mathematical background that the map $\psi^{\sim}$ resulting from the reparametrisation is injective. It should also be noted that the last sentence of Estes-Theorem 2.4 contains a necessary and sufficient criterion of injectivity: The linear independence of the maps $\{\gamma_k\}_{k=1..m}$ in a representation of the function f by Equation (2.8). To write the function f given by Equation (2.17) in this way, we use maps γ_k given by $\gamma_0(q) = q^3$, $\gamma_1(q) = q$, $\gamma_2(q) = q^2$, and $\gamma_3(q) = 2\ q + q^2$. The linear dependence is obvious, since $\gamma_3 = 2\ \gamma_1 + 1\ \gamma_2$. But after the reparametrisation, the map γ_3 has disappeared in Equation (2.20), and the remaining maps γ_1 and γ_2 are linearly independent.[33] So the map $\psi^{\sim}$ is injective, indeed.

Up to this point, we have demonstrated the reparametrisation leading to an injective map $\psi^{\sim}$ for a specific example. The following lemma, which is proved in Section 6.9, generalises the result. It states the existence of a suitable reparametrisition for all non-trivial[34] functions of Class A and outlines an approach to its construction.

Lemma 2.6: Let Q be a non-emty set, m a natural number, $\{\gamma_k\}_{k=0..m}$ a family of maps $\gamma_k:Q\to\mathbb{R}$, and $f:Q\times\mathbb{R}^m\to\mathbb{R}$ a map such that Equation (2.8) holds for every $(q, \theta) \in Q\times\mathbb{R}^m$, and $f(q, \theta) \neq \gamma_0(q)$ for some $(q, \theta) \in Q\times\mathbb{R}^m$. Furthermore, let F be the smallest linear subspace of $\mathbb{R}^Q$ containing all maps γ_k for k = 1..m. Let n be the dimension of F, and $\{k_j\}$j=1..n a sequence of pairwise different integers with $1 \leq k_j \leq m$ such that the maps $\gamma_{k(j)}$ with j = 1..n form a basis of F.[35] With the definition $\gamma^{\sim}_j := \gamma_{k(j)}$ for j = 1..n and $\gamma^{\sim}_0 := \gamma_0$, let a map $f^{\sim}:Q\times\mathbb{R}^n\to\mathbb{R}$ given by the equation

$$f^{\sim}(q, \theta^{\sim}) = \gamma^{\sim}_0(q) + \sum_{j=1..n} \theta^{\sim}(j) \cdot \gamma^{\sim}_j(q). \tag{2.26}$$

Finally, let λ_{jk} with j = 1..n and k = 1..m be the unique real numbers fulfilling the equation

$$\gamma_k = \sum_{j=1..n} \lambda_{jk}\ \gamma^{\sim}_j \tag{2.27}$$

for k = 1..m, and let the (linear) map $t:\mathbb{R}^m\to\mathbb{R}^n$ be defined by the equation

$$t(\theta)(j) := \sum_{k=1..m} \lambda_{jk} \cdot \theta(k) \tag{2.28}$$

for j = 1..n, where $t(\theta)(j)$ is the j^{th} component of $t(\theta)$ for $\theta \in \mathbb{R}^m$.

Then the maps $\{\gamma^{\sim}_j\}_{j=1..n}$ are linearly independent vectors in $\mathbb{R}^Q$, and Equation (2.23) holds for every $(q, \theta) \in Q\times\mathbb{R}^m$.

The main results are stated in the last sentence of the lemma. The validity of Equation (2.23) for the map $f^{\sim}:Q\times\mathbb{R}^n\to\mathbb{R}$ given by Equation (2.26) opens an interface to the general treatment of reparametrisations in Lemma 2.5. Furthermore, we can apply the last sentence of Estes-Theorem 2.4 to the map $\psi^{\sim}:\mathbb{R}^n\to\mathbb{R}^Q$, which is derived from $f^{\sim}$ by Equation (2.21): Since the maps $\{\gamma^{\sim}_j\}_{j=1..n}$ are

[33] Note that only the linear independence of the maps $\{\gamma_k\}_{k=1..m}$ is required in Estes-Theorem 2.4 for an injective map ψ, whereas γ_0 is irrelevant for this issue.

[34] Functions $f:Q\times\mathbb{R}^m\to\mathbb{R}$ with $f(q, \theta) = \gamma_0(q)$ for all $(q, \theta) \in Q\times\mathbb{R}^m$ cannot be reparametrised in the way described in Lemma 2.6, since $f(q, \theta) \neq \gamma_0(q)$ for some (q, θ) is assumed. But although such functions are formally a member of Class A, their triviality is obvious: The map $\gamma_0:Q\to\mathbb{R}$ is the only member of a parametric family of maps $Q\to\mathbb{R}$, whose representation function has this property.

[35] More explicitly, the maps $\{\gamma_{k(j)}\}_{j=1..n}$ must be linearly independent vectors in $\mathbb{R}^Q$, and every element of F must be the result of some linear combination of these maps.

linearly independent vectors in $\mathbb{R}^Q$, the map $\psi^\sim$ is injective.

It has been mentioned in in comments to Theorem 2.4 that Estes (1956) specified his Class A of functions $f{:}Q\times\mathbb{R}^m\to\mathbb{R}$ in the main text of his article such that it contains only cases, where the map $\gamma_0{:}Q\to\mathbb{R}$ in Equation (2.8) is missing (or implicitly identical with the zero-element of $\mathbb{R}^Q$). Such functions have some convenient properties: The map $\psi{:}\mathbb{R}^m\to\mathbb{R}^Q$ is linear, and Equation (2.12) reduces to the simpler Equation (2.10) for all linear combinations and not only those with coefficientes λ_i summing up to 1. It may be noteworthy that every function $f{:}Q\times\mathbb{R}^m\to\mathbb{R}$ of Class A allows a simple reparametrisation. We can replace the term $\gamma_0(q)$ in Equation (2.8) by $1 \cdot \gamma_0(q)$ without changing the result. So define a map $t{:}\mathbb{R}^m\to\mathbb{R}^{m+1}$ such that the first m components of $t(\theta)$ and θ are identical, and $t(\theta)(m+1) := 1$. With the additional definitions $\gamma^\sim{}_k = \gamma_k$ for k = 1..m, and $\gamma^\sim{}_{m+1} := \gamma_0$, the function $f^\sim{:}Q\times\mathbb{R}^{m+1}\to\mathbb{R}$ given by

$$f^\sim(q, \theta^\sim) := \sum_{k=1..m+1} \theta^\sim(k)\, \gamma^\sim{}_k(q) \tag{2.29}$$

is of the type leading to a linear map $\psi^\sim{:}\mathbb{R}^{m+1}\to\mathbb{R}^Q$. Furthermore, Equation (2.23) holds for every $(q, \theta) \in Q\times\mathbb{R}^m$, which makes available the results of Lemma 2.5.

If the maps $\{\gamma^\sim{}_k\}_{k=1..m+1}$ are linearly dependent, we can apply the reparametrisation described in Lemma 2.6 to obtain a function $Q\times\mathbb{R}^{m+1}\to\mathbb{R}$, where the parametrisation map derived from this function combines the two useful properties of injectivity and linearity. The class of such functions deserves a special denotation, which is explicated in the following definition.

Definition 2.7: For a natural number m and a non-empty set Q, a map $f{:}Q\times\mathbb{R}^m\to\mathbb{R}$ belongs to Class A_0 iff there exists a family $\{\gamma_k\}_{k=1..m}$ of linearly independent maps $\gamma_k{:}Q\to\mathbb{R}$ such that the equation

$$f(q, \theta) = \sum_{k=1..m} \theta(k)\, \gamma_k(q) \tag{2.30}$$

holds for every $(q, \theta) \in Q\times\mathbb{R}^m$.

The combination of linearity and injectivity of the parametrisation map $\psi{:}\mathbb{R}^m\to\mathbb{R}^Q$ from a representation function f of Class A_0 can be summarised in the statement that ψ is a vector space isomorphism of $\mathbb{R}^m$ onto $\psi(\mathbb{R}^m)$. This conclusion should be understood as an interface to the treatment of vector space isomorphisms in textbooks of Linear Algebra. All examples presented by Estes (1956) for his Class A are members of Class A_0. Although it is a proper subclass of Class A, the existence of suitable reparametrisations bringing elements of Class A into Class A_0 will allow to assume membership in Class A_0 without loss of generality in subsequent treatments of functions belonging to Class A.

2.2.4

2.2.4 Nonparametric Representation of Families with Affine Parametrisation Maps

The hitherto used way of looking at parametric families of map $Q\to\mathbb{R}$ with a representation function of Class A has enabled proofs of stability under convex linear combinations by rather elementary tools of linear algebra. Another approach to such families will be useful in later sections, where more general kinds of aggregation than convex linear combinations are considered. Subsequently, the approach will be demonstrated by an example, and then the results will be generalised in a lemma.

Example 2.8: Reconsider the set V of all learning curves in the tracking task (Example 1.1), i.e., the set of all maps $v:Q\to[0, 100]$, where Q is the set of natural numbers up to 10, and let S* the set of all quadratic learning curves. In the notation of the preceeding subsections, the set S* can be defined as the set of those elements v of the set V where a vector θ in $\mathbb{R}^3$ exists such that the equation

$$v(q) = \sum_{k=1..3} \theta(k) \cdot q^{k-1} \tag{2.31}$$

holds for every $q \in Q$.

There are several criteria of membership in the set S* for a given learning curve v. For instance, one can compute the 'differences of differences' $(v(q+2) - v(q+1)) - (v(q+1) - v(q))$ for q = 1..8 and check whether they are equal. This criterion has a noteworthy property: Although it is derived from a parametric model, membership in the parametric family can be described by a system of equations such that the parameters do not explicitly turn up in the equations. One could also say that these equations form a non-parametric description of membership in a parametric family of functions

The following approach is a bit more indirect, but it has the advantage of being generalisable to all parametric families with a representation function of Class A and an injective parametrisation map. We can use the first three function values v(1), v(2) and v(3) to obtain a unique element θ of $\mathbb{R}^3$ such that Equation (2.31) holds for q = 1..3. Then the map v is an element of the set S*, if and only if Equation (2.31) holds for every $q \in Q$ and for the parameter θ obtained from the first three function values.

More formally, let $\{h_k\}_{k=1..3}$ be a family of maps $h_k:V\to\mathbb{R}$ with the following property: For every element v of the set V, the function values $h_k(v)$ for k = 1..3 are the components of the unique element θ of $\mathbb{R}^3$ where Equation (2.31) holds for q = 1..3 (i.e., a hypothetical parameter value obtained from the first three function values). In other words, the numbers $h_k(v)$ are the solutions of a system of three equations

$$v(q) = \sum_{k=1..3} h_k(v) \cdot q^{k-1} \tag{2.32}$$

with q = 1..3. Then it is easily verified[36] that the solution of this system is

$$h_1(v) = \;\; 3 \;\;\; v(1) - 3\, v(2) + 1 \;\;\; v(3),$$

$$h_2(v) = -2.5\, v(1) + 4\, v(2) - 1.5\, v(3),$$

and

---

[36] Exercise! Suggestion: Both the proof of the uniqueness of the solution and the generalisation of Example 2.8 in Lemma 2.9 are supported, if the system of linear equations resulting from Equation (2.32) is represented in matrix algebra such that the coefficients q^{k-1} form a suitably arranged matrix B. Then it suffices to verify that the coefficients of the function values v(1), v(2) and v(3) in the claimed solution form the inverse of the matrix B.

$h_3(v) = \ \ 0.5\ v(1) - 1\ v(2) + 0.5\ v(3).$

Based on these maps $h_k:V\to\mathbb{R}$, the following consideration leads to a non-parametric description of the set S*. If v is an element of the set S* (i.e., if a true parameter θ exists such that Equation (2.31) holds for every $q \in Q$), then Equation (2.32) can be used to predict the function values v(q) for every $q \in Q$. So we can define a family $\{g_q\}_{q\in Q}$ of maps $g_q:V\to\mathbb{R}$ such that $g_q(v)$ is the factual function value v(q) minus the predicted one, i.e.,

$$g_q(v) := v(q) - \textstyle\sum_{k=1..3} h_k(v) \cdot q^{k-1}. \tag{2.33}$$

Then an element v of the set V is contained in S* iff the equality $g_q(v) = 0$ holds not only for q = 1..3 (which follows from the construction of the maps $h_k:V\to\mathbb{R}$), but for every $q \in Q$.

Before the example is generalised, note that the elements q = 1..3 of the set Q underlying the computation of the components $h_k(v)$ of a 'hypothetical parameter' could have been replaced by any sequence $\{q_j\}_{j=1..3}$ of pairwise different elements of the set Q. Of course, the maps $h_k:V\to\mathbb{R}$ and $g_q:V\to\mathbb{R}$ would change,[37] but the main result would be the same: A map $v:Q\to\mathbb{R}$ is an element of S* iff the equality $g_q(v) = 0$ holds for every $q \in Q$.[38]

Now observe that the set S* is a subset of a parametric family of maps based on a parametrisation map $\psi:\mathbb{R}^3\to\mathbb{R}^Q$ such that $\psi(\theta) = \gamma_0 + \sum_{k=1..3} \theta(k) \cdot \gamma_k$ with $\gamma_0(q) = 0$ and $\gamma_k(q) = q^{k-1}$ for k = 1..3. The generalisation of the main result in the following lemma is a bit more complex to cover situations, where the equality $\gamma_0(q) = 0$ may fail to hold for some elements q of the set Q. For a parametrisation map $\psi(\theta) = \gamma_0 + \sum_{k=1..m} \theta(k) \cdot \gamma_k$ and a sequence $\{q_j\}_{j=1..m}$ of elements of a set Q, the equation

$$v(q_j) - \gamma_0(q_j) = \textstyle\sum_{k=1..m} h_k(v) \cdot \gamma_k(q_j) \tag{2.34}$$

yields a system system of linear equations for the components $h_k(v)$ of a hypothetical parameter vector. Beyond being more complex than Equation (2.32) by the additional term $\gamma_0(q_j)$, Equation (2.34) shows a further problem. The solution of this system of equations is unique, if and only if an m×m-matrix B with elements $b_{jk} := \gamma_k(q_j)$ is non-singular. (In Example 2.8, this requirement was fulfilled for every sequence $\{q_j\}_{j=1..m}$ of pairwise different elements of Q.[39]) In the lemma, the inverse of this matrix turns up as a matrix Z with elements z_{kj}, whose counterparts in the example are the coefficients of v(1), v(2) and v(3) in the above solution. Furthermore, some more or less

---

[37] Exception: For $v \in S^*$, the components $h_k(v)$ of the hypothetical parameter would be identical with those of the underlying parameter θ, regardless of the elements q_j underlying the computation.

[38] It goes almost without saying that the validity of Equation (2.32) for every $q \in Q$ is trivial, if the maps h_k are changed such that Equation (2.32) is always applied with a maps h_k, where q is contained in the underlying sequence $\{q_j\}_{j=1..m}$. In other words, Equation (2.32) is sufficient for $v \in S^*$ only if the same maps $h_k:V\to\mathbb{R}$ are used for every q.

[39] Exercise!

complex sums are represented by real numbers ξ_k, τ_q and ζ_{qj}, which are only claimed to exist and to have certain properties. But these properties are sufficient for the conclusions to be drawn after the statement of the lemma. A specification of these sums is given by Equations (6.16), (6.17) and (6.18) in the proof of the lemma in Section 6.10.

Lemma 2.9: In the situation of Estes-Theorem 2.4, let $\{\gamma_k\}_{k=0..m}$ be a family of maps γ_k:Q→$\mathbb{R}$ such that Equation (2.9) holds for every $\theta \in \mathbb{R}^m$, and assume that the parametrisation map ψ:$\mathbb{R}^m$→$\mathbb{R}^Q$ of that equation is injective. Furthermore, let V be a set of maps Q→$\mathbb{R}$, and let a subset S* of V, be given by the definition

$$S^* := V \cap \psi(\mathbb{R}^m) = \{\psi(\theta) \in V\colon \theta \in \mathbb{R}^m\}. \tag{2.35}$$

Then there exist a family $\{q_j\}_{j=1..m}$ of elements of the set Q, a family $\{h_k\}_{k=1..m}$ of maps h_k:V→$\mathbb{R}$, a family $\{g_q\}_{q\in Q}$ of maps g_q:V→$\mathbb{R}$, a non-singular m×m-matrix Z with elements z_{kj}, and families $\{\xi_k\}_{k=1..m}$, $\{\tau_q\}_{q\in Q}$ and $\{\zeta_{qj}\}_{q\in Q,\, j=1..m}$ of real numbers with the following properties:

(i) $$h_k(v) = \xi_k + \textstyle\sum_{j=1..m} z_{kj} \cdot v(q_j) \tag{2.36}$$
for k = 1..m and every $v \in V$.

(ii) $$g_q(v) = \tau_q + v(q) + \textstyle\sum_{j=1..m} \zeta_{qj} \cdot v(q_j) \tag{2.37}$$
for every $q \in Q$ and every $v \in V$.

(iii) The equation
$$g_q(v) = v(q) - (\gamma_0(q) + \textstyle\sum_{k=1..m} h_k(v) \cdot \gamma_k(q)) \tag{2.38}$$
holds for every $q \in Q$ and every $v \in V$.

(iv) If v is an element of S* and θ is an element of $\mathbb{R}^m$ with $v = \psi(\theta)$, then $\theta(k) = h_k(v)$ for k = 1..m.

(v) S* is the set of those elements of V where the equality $g_q(v) = 0$ holds for every $q \in Q$.

Before parallels to Example 2.8 are pointed out, a minor difference should be noted: The addition of the terms ξ_k, τ_q and $\gamma_0(q)$ in Equations (2.36), (2.37) and (2.38) has no counterpart in the example, since the equality $\gamma_0(q) = 0$ for every $q \in Q$ enabled a simpler computation of the components $h_k(v)$ of a 'hypothetical parameter': It could be based on Equation (2.32) instead of the more complex Equation (2.34).[40] But with the exception of these additive components, the properties claimed by the lemma are generalisations of Example 2.8. According to Equation (2.36), the components $h_k(v)$ of the hypothetical parameter for a map v are based on a linear combination of the function values $v(q_j)$ (plus the number ξ_k). The interpretation of the function values $h_k(v)$ as components of a hypothetical parameter is supported by Assertion (iv): If a true parameter θ exists, then $h_k(v)$ is the k^{th} component of θ. Under this interpretation, the expression $\gamma_0(q) + \sum_{k=1..m} h_k(v) \cdot \gamma_k(q)$ in Equation (2.38) is the function value v(q), which would have to be predicted under the assumption that $h_k(v)$ is the k^{th} component of a parameter θ underlying a map v:Q→$\mathbb{R}$. So Equation (2.38) tells that $g_q(v)$ is again the deviation of the factual function value from the predicted one. Finally, Assertion (v) is a description of the set S* by a system of equations, where the parameter θ doesn't turn up explicitly, although the equations are derived from a parametric description of the set S*.

---

[40] More explicitly, the specification of the numbers ξ_k and τ_q by Equations (6.16) and (6.17) in the proof shows that these numbers are zero, if $\gamma_k(q) = 0$ for every $q \in Q$.

Under the additional ussumption that the set V is convex, the non-parametric description of the set S* enables a very straightforward approach to its convexity. For every $q \in Q$, let S_q be the set of all elements v of V with $g_q(v) = 0$. Then Lemma 2.9.(v) can be summarised in the equation $S^* = \bigcap_{q \in Q} S_q$, and it suffices to verify the convexity of the sets S_q. So let q be an arbitrary element of Q, and consider a linear combination $v = \lambda\, v' + (1-\lambda)\, v''$ with elements v' and v'' of the set S_q, and $0 \le \lambda \le 1$. To derive the property $v \in S_q$, note that the first equality in the equation

$$g_q(v) = \lambda \cdot g_q(v') + (1-\lambda) \cdot g_q(v'') = \lambda \cdot 0 + (1-\lambda) \cdot 0 = 0 \tag{2.39}$$

for every $q \in Q$ is obtained from Equation (2.37), and the second one from the premissa that v' and v'' are elements of the set S_q.

Furthermore, recall that the basic result of Estes (1956) for his Class A went beyond the mere aggregation stability of parametric families of funtions: A convex linear combination of elements of the family results in another element of the family, whose parameter can be obtained by an application of the same convex linear combination to the individual parameters. Now an application of Equation (2.36) to a convex linear combination $v = \sum_{i=1..n} \lambda_i \cdot v_i$ of elements of the set Vyields

$$h_k(v) = \sum_{i=1..n} \lambda_i\, h_k(v_i) \tag{2.40}$$

for k = 1..m. If the maps v_i are elements of S*, then the convexity of the set S* implies $v \in S^*$, and then Lemma 2.9.(iv) allows to interpret $h_k(v)$ and $h_k(v_i)$ as the the k^{th} component of parameters θ resp. θ_i underlyung v and v_i, and then Equation (2.40) is a reformulation of Estes' above quoted result for his Class A.

Finally, Equation (2.40) can also be used in the study of aggregation stability of subsets of S*, e.g., the set S of those elements of S* where the inequality $\theta(1) < \theta(2)$ holds for the components of the underlying parameter. If the vectors v_i in a convex linear combination are elements of S, then Lemma 2.9.(iv) allows to rewrite this property as $h_1(v_i) < h_2(v_i)$ for every v_i, and then the inequality $h_1(v) < h_2(v)$ follows from Equation (2.40).

Although the restatement and the extension of Estes' results for his Class A become simple by the non-parametric representation of a parametric family, the gain achieved by the somewhat laborious reformulation may be questioned. Indeed, if aggregation is confined to convex linear combinations, then the parametric approach of Section 2.2.3 is sufficient. But in later sections, a generalisation of the results to other kinds of aggregation will be problematic under the parametric approach, whereas it is supported by the non-parametric reformulation in Lemma 2.9. (See Examples 4.18, ? and 4.40). For the moment, it may suffice to recall that the reformulation of Lemma 2.9.(v) by the equality $S^* = \bigcap_{q \in Q} S_q$ allowed to confine the study of convexity to the simple sets S_q.

Alternative approach by the following lemma:

Lemma 2.10: Let Q be a non-empty set, F a linear subspace of $\mathbb{R}^Q$ of finite dimension m, and S a subset of F. Furthermore, let $\{\gamma_k\}_{k=1..m}$ be a family of linearly independent elements of F, define a map $\psi:\mathbb{R}^m \to \mathbb{R}^Q$ by $\psi(\theta) := \sum_{k=1..m} \theta(k) \cdot \gamma_k$ for every $\theta \in \mathbb{R}^m$, and a subset S* of $\mathbb{R}^m$ by $S^* := \psi^{-1}(S)$.

Then the map ψ is injective and linear, and $\psi(\mathbb{R}^m) = F$. Furthermore, there exist linear and surjective maps $g:\mathbb{R}^Q \to \mathbb{R}^m$ and $h:\mathbb{R}^Q \to F$ with the following properties:

(i) The equality

$$g(\psi(\theta)) = \theta \tag{2.41}$$

holds for every $\theta \in \mathbb{R}^m$.

(ii) There exists a family $\{q_j\}_{j=1..m}$ of element of Q and a non-singular m×m-matrix Z with elements z_{kj} such that the equality

$$g(v)(k) = \sum_{j=1..m} z_{kj} \cdot v(q_j) \tag{2.42}$$

holds for every map $v:Q \to \mathbb{R}$ and for k = 1..m.

(iii) The equality

$$h(v)(q) = \sum_{k=1..m} g(v)(k) \cdot \gamma_k(q) \tag{2.43}$$

holds for every map $v:Q \to \mathbb{R}$ and every $q \in Q$.

(iv) F is the set of all maps $v:Q \to \mathbb{R}$ with $h(v) = v$.

(v) S is the set of all maps $v:Q \to \mathbb{R}$ with $h(v) = v$ and $g(v) \in S^*$.

2.2.5 Reparametrisations Leading to Affine Parametrisation Maps

Having reviewed and supplemented the results of Estes (1956) for his Class A of functions $f:Q\times\mathbb{R}^m \to \mathbb{R}$, we can now turn to Class B. It consists of functions, which are not members of Class A, but can be brought into this class by a suitable reparametrisation. We have already introduced one of his examples for this class in Figure 2.1.A, which should now be seen as a partial graph[41] of the map $f:Q\times\Theta \to \mathbb{R}$ given by

$$f(q, \theta) = \log(\theta\, q), \tag{2.44}$$

with $Q = \Theta =]0, +\infty[$. Now consider the map $f^{\sim}:Q\times\mathbb{R} \to \mathbb{R}$ with

$$f^{\sim}(q, \theta^{\sim}) = \theta^{\sim} + \log(q). \tag{2.45}$$

If we identify the parameter space $\mathbb{R}$ with $\mathbb{R}^1$, we have an instance of Equation (2.8) with m = 1, $\gamma_0(q) = \log(q)$, and $\gamma_1(q) = 1$. So $f^{\sim}$ is a function of Class A. Taking both equations together, we get

$$f(q, \theta) = f^{\sim}(q, \log(\theta)) \tag{2.46}$$

for every $(q, \theta) \in Q\times\Theta$. This equation can be rewritten in terms of partial maps as

$$f(\,.\,, \theta) = f^{\sim}(\,.\,, \log(\theta)) \tag{2.47}$$

for every $\theta \in \Theta$.

Implications of the identity of the partial maps $f(\,.\,, \theta)$ and $f^{\sim}(\,.\,, \log(\theta))$ can be studied in Figure

---

[41] The denotation of Figure 2.1.A as a partial graph refers to the fact that the function f is plottet only for subsets of the sets Q and Θ.

2.1, whose part B is a partial graph of the function $f^{\sim}$. The alterations resulting from the reparametrisation are striking for the partial maps f(q, .) resp. $f^{\sim}$(q, .): They change from vertically translated logarithmic curves to straight lines. But the partial maps f(. , θ) of the left graph reappear in the right one as partial maps $f^{\sim}(. , \theta^{\sim})$, suitably displaced only in the parameter dimension at $\theta^{\sim} = \log(\theta)$. In particular, the set of *all* partial maps for constant parameter values is identical for both functions.[42] So the stability under aggregation (by convex linear combinations) of the set of all partial maps $f^{\sim}(. , \theta^{\sim})$, which is granted by the membership of $f^{\sim}$ in Class A, is tantamount with aggregation stability of the set of all partial maps f(. , θ).

The same result can be obtained analytically. Since Equation (2.46) is an instance of Equation (2.23) with t(θ) = log(θ), Equations (2.22) and (2.10) can be combined for the treatment of convex linear combinations of partial maps $\psi(\theta_i)$:

$$\begin{aligned} \textstyle\sum_{i=1..n} \lambda_i\, \psi(\theta_i) &= \textstyle\sum_{i=1..n} \lambda_i\, \psi^{\sim}(t(\theta_i)) \\ &= \psi^{\sim}(\textstyle\sum_{i=1..n} \lambda_i\, t(\theta_i)) \\ &= \psi(\theta^*), \end{aligned} \tag{2.48}$$

where θ^* is a suitable element of Θ, which will be specified.immediately.

Equation (2.48) is the typical result of Estes (1956) for his Class B of functions $f{:}Q\times\Theta\to\mathbb{R}$: The mere existence of a suitable parameter θ^* grants that the set ψ(Θ) of all maps ψ(θ) with $\theta \in \Theta$ is stable under convex linear combinations; but - different from Class A - the parameter θ^* of the map $\psi(\theta^*)$ resulting from a convex linear combination can generally not be obtained from a convex linear combination of the parameters θ_i with the given coefficients λ_i.

But we can go a step further than Estes (1956) and generalise a specification of the parameter θ^*, which he gave only for a single example. For the function f specified by Equation (2.44), a logarithmic transformation $t{:}\Theta\to\mathbb{R}$, and simple averaging of maps (i.e., $\lambda_i = 1/n$ for i = 1..n), the parameter θ^* is the solution for θ^* of the equation $\log(\theta^*) = \sum_{i=1..n} \log(\theta_i)$, i.e., the geometric mean of the parameters $\{\theta_i\}_{i=1..n}$. More generally, a parameter θ^* is suitable for the last step in Equation (2.48), if

$$t(\theta^*) = \textstyle\sum_{i=1..n} \lambda_i\, t(\theta_i). \tag{2.49}$$

Then this last step is legitimated by Equations (2.22). If the underlying transformation map $t{:}\Theta\to\mathbb{R}^m$ is injective, we can also write $\theta^* = t^{-1}(\sum_{i=1..n} \lambda_i\, t(\theta_i))$.

Since the results for Class B have been obtained rather informally, they are stated more systematically in the following lemma, which is proved in Section 6.11.[43]

---

[42] Note, however, that the curves in both graphs representing partial maps f(. , θ) and $f^{\sim}(. , \theta^{\sim})$ plot only a sample of all partial maps. There is no one-to-one-correspondens between the sampled parameter values θ and $\theta^{\sim}$.

[43] Note that Lemma 2.11 applies not only to Class B in its definition by Estes (1956), where it
(continued...)

Lemma 2.11: Let nonempty sets Q and Θ and a map $f:Q\times\Theta\to\mathbb{R}$ be given with the following property: There is a map $\tilde{f}:Q\times\mathbb{R}^m\to\mathbb{R}$ of Class A (with $m\in\mathbb{N}$) and a map $t:\Theta\to\mathbb{R}^m$ such that the set $t(\Theta)$ is convex and Equation (2.23) holds for every $(q, \theta)\in Q\times\Theta$. Finally, let a map $\psi:\Theta\to\mathbb{R}^Q$ be given by Equation (2.5). Then the result of every convex linear combination $\sum_{i=1..n}\lambda_i\ \psi(\theta_i)$ is an element of $\psi(\Theta)$, and for an injective[44] map t, it is given by the equation

$$\sum_{i=1..n}\lambda_i\ \psi(\theta_i) = \psi(t^{-1}(\sum_{i=1..n}\lambda_i\ t(\theta_i))) \qquad (2.50)$$

Typically, the set Θ in Lemma 2.11 will be the parameter space for a considered parametric family of maps $Q\to\mathbb{R}$; but as in Class A, it can also be replaced by a proper subset S of the parameter space representing a hypothesis in empirical research.

For Class A, we noted down to pay attention to situations, where an aggregation rule implies the applicability of the same process of aggregation (e.g., pointwise averaging) to other objects than those treated in the aggregation rule. The above supplement to the results of Estes (1956) for his Class B shows that this note should be generalised: An aggregation rule can also imply the applicability of a different, but related aggregation rule to other objects. E.g., convex linear combinations of partial maps $\psi(\theta_i)$ based on a function f of Class B implies the indirect aggregation of the parameters θ_i indicated by Equation (2.50): The convex linear combination is applied to the transformed parameters $t(\theta_i)$, and the result is transformed back. Recall the special case of the geometric mean: The averaging is applied to the logarithmically transformed parameter values, and the result of averaging is transformed back.

We will now show that membership in one of the Classes A and B is not only a sufficient condition of stability under convex linear combinations under the assumptions of Lemma 2.11, but also a necessary one, if the parameter space Θ is a subset of a finite dimensional real vector space. So assume that Q is a nonempty set, Θ a subset of a finite dimensional real vector space, and $f:Q\times\Theta\to\mathbb{R}$ the representation function of a parametric family of maps $Q\to\mathbb{R}$ such that the set $\psi(\Theta)$ is stable under convex linear combinations for the parametrisation map $\psi:\Theta\to\mathbb{R}^Q$ given by Equation (2.5). By Lemma 2.1, the assumed stability under convex linear combinations of the set $\psi(\Theta)$ implies its convexity, and this property allows to apply a basic result of the analysis of real vector spaces,

---

[43] (...continued)
contains only functions, which are not members of Class A. If the set Θ in Lemma 2.11 is a subset of a real vector space and the parameter transformation map $t:\Theta\to\mathbb{R}^m$ is affine or can be extended to such a map, then the parametrisation map $\psi:\Theta\to\mathbb{R}^m$ - being the concatenation of two affine maps by Equation (2.24) - shares this property. (See Estes-Theorem 2.4 for the affinity of the parametrisation map $\tilde{\psi}:\mathbb{R}^m\to\mathbb{R}^Q$, which is derived from the representation map $\tilde{f}$, and note that membership of $\tilde{f}$ in Class A is assumed by the lemma.) So the underlying representation map f is of class A by Estes-Theorem 2.4; but Lemma 2.11 applies nevertheless, since it doesn't exclude an affine map $t:\Theta\to\mathbb{R}^m$.

[44] For a non-injective map $t:\Theta\to\mathbb{R}$, Equation (2.50) applies under a suitable interpretation of the term $t^{-1}(\sum_{i=1..n}\lambda_i\ t(\theta_i))$: It may be understood as the set of $\theta\in\Theta$ fulfilling the equation $t(\theta) = \sum_{i=1..n}\lambda_i\ t(\theta_i)$. (See Footnote 56 for the underlying definition of inverse maps.) If θ_1 and θ_2 are two elements of Θ with this property, then $\psi(t(\theta_1)) = \psi(t(\theta_2))$ follows immediately. So Equation (2.50) specifies a unique result of the convex linear combination on its left hand side.

which is proved in Section 6.12:

Lemma 2.12: Let Θ be a non-empty subset of a real vector space E with finite dimension n, F another real vector space, and $\psi:\Theta\to F$ a map such that the set $\psi(\Theta)$ is convex and contains a non-zero element of F. Then $\psi(\Theta)$ contains linearly independent vectors $\{y_k\}_{k=1..m}$ with $m \le n+1$ such that every element of $\psi(\Theta)$ has a unique representation by a linear combination of the vectors y_k. I.e., the vectors y_k form a basis of the linear subspace of F which is spanned by $\psi(\Theta)$.

Transferring the role of the vector space F of the lemma to our function space $\mathbb{R}^Q$, let a suitable sequence $\{\gamma_k\}_{k=1..m}$ of elements of $\psi(\Theta)$ be given, and define a map $t:\Theta\to\mathbb{R}^m$ such that $t(\theta)$ with $\theta \in \Theta$ is the unique vector $\theta^\sim$ in $\mathbb{R}^m$ with

$$\psi(\theta) = \textstyle\sum_{k=1..m} \theta^\sim(k)\ \gamma_k. \tag{2.51}$$

Furthermore, define a map $f^\sim:Q\times\mathbb{R}^m\to\mathbb{R}$ by $f^\sim(q, \theta^\sim) := \sum_{k=1..m} \theta^\sim(k)\ \gamma_k(q)$. Then the function $f^\sim$ is a member of Class A (even A_0), and Equation (2.23) holds for every $(q, \theta) \in Q\times\Theta$. So everything is prepared to apply the result for Class B, which is summarised in Lemma 2.11. Observe also that the map $t:\Theta\to\mathbb{R}^m$ is injective if and only if the map $\psi:\Theta\to\mathbb{R}^Q$ has this property. If this property is given, then the indirect aggregation by Equation (2.50) can also be applied.

Finally, we can combine Lemmas 2.1 and 2.12 to give a precise specification of all subsets S of $\mathbb{R}^Q$, which are stable under convex linear combinations and can be considered as parametric families of maps $Q\to\mathbb{R}$ with a parameter space $\Theta \subseteq \mathbb{R}^m$. The specification is given by the subsequent corollary, which is proved in Section 6.13.

Corollary 2.13: Let Q be a non-empty set, and S a non-empty subset of $\mathbb{R}^Q$. Then the following assertions are eqivalent:

(i) S is a convex subset of a finite dimensional linear subspace of $\mathbb{R}^Q$.

(ii) S is stable under convex linear combinations, and for some natural number m, some subset Θ of $\mathbb{R}^m$ and some function $f:Q\times\Theta\to\mathbb{R}$, the set S consists of all members of a parametric family of maps $Q\to\mathbb{R}$ with parameter space Θ and representation function f.

2.2.6 Stability under Indirect Aggregation

A similar kind of indirect aggregation as the one specified by Equation (2.50) is also suggested by Estes (1956) for certain members of his Class C. This class consists of functions $f:Q\times\Theta\to\mathbb{R}$, which are neither members of Class A nor of Class B. Certainly, the set $\psi(\Theta)$ of all members of a parametric family with a representation function of Class C isn't stable under convex linear combinationce, since we have verified that membership in one of these classes is necessary for this stability. But Estes pointed out that this stability may be obtained, if the aggregation is performed after a suitable transformation of the dependent variable. E.g., for the family of growth curves based on the function $f(q, \theta) = 1 - e^{-\theta\cdot q}$ (with $\Theta = Q =]0, +\infty[$), the set $\psi(\Theta)$ isn't stable under convex linear combinations. But if the transformation $\tau(y) = -\log(1-y)$ (with $0 \le y < 1$) is applied to the dependent variable, the resulting representation function $\tau \circ f$ is given by $\tau(f(q, \theta)) = \theta\ q$, and this is the restriction to Θ of a function of Class A_0.

Again, we can obtain more general and more detailed results than Estes, if we analyse not only single function values, but consider partial maps $\psi(\theta)$ as whole objects. Then a transformation may be based not only upon a pointwise application of the same transformation to all values of the dependent variable, but on a map of a vocabulary set V into a set of maps $Q \to \mathbb{R}$, where Q may be an entirely new set.

This approach will subsequently be demonstrated by an example, which is based upon an article by Thomas and Ross (1980). These authors studied aggregation stability in cumulative distribution functions for continuous real valued random variables; but different from our hitherto presented analyses, the aggregation analysed by these authors isn't based on convex linear combinations of such distribution functions, but upon an approach, which has been introduced (in particular for the study of reaction times) by Ratcliff (1979) under the name of Vincentising. A central notion in this approach is the *p^{th} order quantile* (with $p \in]0, 1[$) of the continuous distribution of a real valued random variable, and this quantile is the real number with cumulative probability p. (A well known example with p = 0.5 is the median.) To grant the existence of a unique real number with this property for ervery $p \in]0, 1[$, a suitable vocabulary set V of cumulative distribution functions consists of all non-decreasing maps $v:\mathbb{R} \to [0, 1]$ with the following property: For every $p \in [0, 1]$, there exists a unique real number ξ with $v(\xi) = p$.[45]

Whereas our previous analyses were based on pointwise expectations of the function values v(q) for every $q \in Q$, the aggregation by Vincentising uses expectations of the p^{th} order quantiles. In Section 4.6, the technique will be defined in terms of general expectations; but for the moment, it can be considered as an application of convex linear combinations to the p^{th} order quantiles: For a family $\{v_i\}_{i=1..n}$ of elements of the above set V and non-negative coefficients λ_i summing up to 1, the result of Vincentising is a map $v^*:\mathbb{R} \to [0,1]$, whose p^{th} order quantiles result from linear combinations (with coefficients λ_i) of the p^{th} order quantiles obtained from the individual maps v_i. (It can be left to the reader to verify, that this definition leads to a unique element v* of V.[46])

Now we can reconceive this Vincentising as another instance of the indirect aggregation, which has been applied to parameters θ_i in Equation (2.50). For this purpose, define $Q :=]0, 1[$, and let V' be the set of all continuous, strictly increasing maps $v':Q \to \mathbb{R}$. Furthermore, let a map $g:V \to V'$ be given such that g(v) is a map $v':Q \to \mathbb{R}$, where v'(q) is the q^{th} order quantile of v. Since the map g is obviously bijective,[47] the above description of a result v* of Vincentising can be written as

$$v^* = g^{-1}(\textstyle\sum_{i=1..n} \lambda_i\, g(v_i)). \tag{2.52}$$

Thomas and Ross (1980) considered stability under Vincentising for parametric families of distribution functions, where the p^{th} order quantile is given by a function f belonging to one of the Classes A and B of Estes (1956), the number $p \in]0, 1[$ taking the role of the argument q in Equation

---

[45] See Section 6.34 for some properties resulting from the definition of the vocabulary set V.

[46] Section 6.34 contains a proof; but it is based on the generalisation of Vincentising to be presented in Section 4.6.

[47] See Section 6.34 for a proof.

(2.8).[48] Independent of the class of f, we have the following situation: There is a parameter space Θ and a function $f: Q \times \Theta \to \mathbb{R}$ (with $Q =]0, 1[$) such that $f(q, \theta)$ is the q^{th} order quantile of a probability distribution with parameter θ. With the usual definition of a map $\psi: \Theta \to \mathbb{R}^Q$ by $\psi(\theta) := f(\, . \, , \theta)$ for every $\theta \in \Theta$, we can interpret $\psi(\theta)$ as a map $Q \to \mathbb{R}$ giving the q^{th} order quantile of a probability distribution with parameter θ.

Note, however, that the maps f and ψ are not representation functions resp. parametrisation maps in the understanding of Definition 2.2. To clarify this issue, let a map $d: \Theta \to V$ (with the above specified vocabulary set V) be given such that $d(\theta)$ is the cumulative probability distribution function with parameter θ. Then $f(q, \theta)$ is not a function value of the map $d(\theta)$, but the q^{th} order quantile. Similarly, a parametrisation map should fulfill the equation $\psi(\theta) = d(\theta)$; but we have $\psi(\theta) = g(d(\theta))$. In other words, the indirect aggregation specified by Equation (2.52) can be covered by an application of the results of Estes (1956) to the maps g(v) and not to the elements of the vocabulary set V.

Now let $t: \Theta \to \mathbb{R}^m$ be an injective map with the properties assumed for the treatment of Class B in Lemma 2.11, including the convexity of the set $t(\Theta)$. (If f is of Class A, let t be the identity map in Θ, i.e., $t(\theta) := \theta$). Then an application of Equation (2.52) to the Vincentising of cumulative distribution functions $\{d(\theta_i)\}_{i=1..n}$ leads to

$$g^{-1}(\textstyle\sum_{i=1..n} \lambda_i \, g(d(\theta_i))) = d(t^{-1}(\textstyle\sum_{i=1..n} \lambda_i \, t(\theta_i))). \tag{2.53}$$

To derive this result from Equation (2.50) (for a map f of Class B) resp. from Equation (2.10) (for Class A), recall that the map g is bijective.

2.2.7 Summary of Results for Parametric Families of Functions

In summary, all hitherto encountered forms of stability under (direct or indirect) aggregation by convex linear combinations are variations of the basic result of Estes (1956) for his Class A, which is reformulated in Equation (2.10) and supplemented in Estes-Theorem 2.4: For maps $\psi(\theta_i) := f(\, . \, , \theta_i)$, which are based on a map $f: Q \times \mathbb{R}^m \to \mathbb{R}$ of Class A, the result of a convex linear combination (with coefficients λ_i) can be obtained by an application of a linear combination with the same coefficients λ_i to the parameters θ_i. In Class B, the aggregation of parameters in Equation (2.50) is more indirect: The convex linear combination is applied to transformed parameters $t(\theta_i)$, and the result is transformed back. Finally, a similar indirect aggregation may be applied to the aggregated objects, e.g. probability distributions $d(\theta_i)$ in Vincentising: The convex linear combination is applied to transformed objects $g(d(\theta_i))$, and the result is transformed back.

[48] For the sake of quotational correctness, it should be recalled that our presentation isn't claimed to be a simple review of Thomas and Ross (1980), but a reconceptualisation of their results. These authors defined classes of parametric families of elements of our our vocabulary set V and wondered afterwards (!) that their classes turned out to be very similar to those of Estes (1956). The reconceptualisation of their approach is motivated by the view on entire maps $Q \to \mathbb{R}$ as mathematical objects (i.e., elements of the function space $\mathbb{R}^Q$).

2.3 Parametric Families of Distribution Functions

It is a well known, but frequently neglected fact that averages or convex linear combinations of probability distributions, which are members of the same parametric family (binomial or normal distributions etc.) are usually not members of that family. E.g., if maps $v_1:\mathbb{R}\rightarrow\mathbb{R}$ and $v_2:\mathbb{R}\rightarrow\mathbb{R}$ are probability density functions of normal distibutions with expectations μ_1 = ### and μ_2 = ###, and standard deviations σ_1 = ### and σ_2 = ###, then the map $\lambda_1 v_1 + \lambda_2 v_2$ with $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$ is a probability density function, but not of a normal distribution, which can be seen from its plot in Figure ###.

Convex linear combinations of probability distributions are usually called mixture distributions. We will use the term 'mixture' for the aggregation of maps $Q\rightarrow\mathbb{R}$, if their function values can be conceived as probabilities or probability densities and the aggregation rule is based on pointwise expectations.[49] Stability under this kind of aggregation will be called *stability under mixtures*. For the following review and reformulation of results from the literature, it suffices again to consider expectations based on convex linear combinations of maps $Q\rightarrow\mathbb{R}$, generalisations to other expectations being postponed until Section ###.

There are several interpretations of such mixtures. Most concretely, the coefficients λ_i of a convex linear combination can be relative sizes of subpopulations or subsamples with distribution functions v_i. Then the result of the convex linear combination $\sum_{i=1..n} \lambda_i v_i$ is the distribution function for the pooled population resp. sample. For another interpretation, we can resume the understanding of aggregation, which has been presented in Section 2.1 to motivate the SSA described in Lemma 2.1: In a process of randoms seletion and observation of units, the coefficients λ_i can be selection probabilities of units (e.g. persons) belonging to a finite domain D. A more detailed discussion of the applicability to such processes of an aggregation rule based on convex linear combinations will follow in Section ###. For the moment, we can formulate the problem of the present subsection in a way saving this discussion: Does a convex linear linear combination of members of a considered parametric family of probability distribution functions result in another member of the same family? In other words: Is membership in a considered family of probability distribution functions stable under mixtures?

Undoubtedly, warnings against premature assumptions about stability under mixtures are justified, if they refer to the commonly known parametric families of distribution functions. In particular, Sixtl (####) derived radical consequences and postulated an entirely new methodology of empirical research in psychology, which should be based on the mixture distribution approach. The present author agrees with Sixtl about the fruitfulness of this approach in the derivation of predictions for empirical research. But he disagrees about consequences, which have to be drawn from the lacking stability under mixtures of the commonly known parametric families of distribution functions. Later subsections will prove the stability under mixtures for properties, which are more immediately related to most substantial hypotheses in psychology than membership in a parametric

---

[49] Indeed, if we mix liquids to cocktails and conceive the weight proportions of components (like water, sugar, alcohol etc.) as probabilities, then the composition of the cocktail is given by pointwise expectations of the compositions of the mixed liquids, the coefficient ('mixing probability') of each liquid representing its relative weight.

family, and consequences for the derivation of testable predictions in empirical research will be drawn in Section 5. Nevertheless, the rest of the present subsection will demonstrate that other parametric families of distribution functions than the commonly known ones are very well stable under convex linear combination. This result will also disprove the claim of Sixtl (###, p. ###) that the Bernoulli distributions form the only parametric family with this stability, and we will replace this claim by a more general description of such families.

Basically, this description will mount up to a straightforward application of Corollary 2.13 under the additional assumption that all maps $Q\rightarrow\mathbb{R}$, which are elements of the set S of the corollary, must be probability distribution functions. To prepare the application of the corollary, we have to explicate this premissa. The requirement that the function values of a map $Q\rightarrow\mathbb{R}$ are probabilities, is necessary, but not sufficient to call this map a distribution function. In later sections, we will consider maps $Q\rightarrow\mathbb{R}$, which cannot be called distribution functions, although the function values are probabilities. But we have to face the fact that there is no commonly accepted terminology in this field. For a set Q of real numbers and a random variable Y with 'range' Q and 'distribution function' $v:Q\rightarrow\mathbb{R}$, a function value v(q) may be the probability of one of the events $Y = q$, $Y < q$ or $Y \leq q$, but also the probability density in q. In a still broader understanding, the set Q may also be a σ-algebra or a set of events generating a σ-algebra, and then a distribution function $v:Q\rightarrow\mathbb{R}$ may be (or specify) a probability measure on the σ-algebra. The subsequent considerations apply (with due changes in details) to all these kinds of distribution functions. In an abstract way, they can be summarised in the statement that Assertions (i) and (ii) of Corollary 2.13 are equivalent for a set S of maps $Q\rightarrow\mathbb{R}$, which are probability distributions under the same understanding of this concept.

To draw concrete concrete conclusions from this equivalence, it seems advisable to demonstrate them under one of the understandings of probability distributions, and cumulative distribution functions (CDFs) seem to form an optimal compromise of simplicity, concreteness and generality.[50] A generalisation to other understandaings of probability distributions (including probability measures) will be outlined at the end of the present subsection.

For purposes of mathematical exactnes, the central concept is defined formally in terms of the theory of probability measures, and a more intuitive interpretation will be given after the definition.
(ii)

Definition 2.14: For a non-empty set Q of real numbers, a map $v:Q\rightarrow\mathbb{R}$ is a *cumulative distibution function (CDF) with domain* Q iff there exists a probability meaure P on $\mathcal{B}$ (the σ-algebra of Borel sets in $\mathbb{R}$) with the following properties:

(i) The equation

[50] More explicitly, the preference for a protypical CDF-interpretation of maps $v:Q\rightarrow\mathbb{R}$ representing probability distributions is motivated not only by arguments of mathematical convenience (saving separate treatment of discrete and continuous distributions). The CDF-approach to probability distributions can also be considerd as an optimal compromise between the theory of probability spaces and the more familiar probability distribution functions. Furthermore, probability distributions of random variables are *defined* in the theory of probability spaces as probability measures on a σ-algebra. But since it is frequently hard to follow up with the details of an entire σ-algebra, concrete distributions on σ-algebras or families of such distributions are commonly *specified* either by the probabilities of a simpler system of events generating the σ-algebra, or by density functions, which imply probabilities of events like $Y \leq \xi$ by integration.

$$v(q) = P(]-\infty, q]) \tag{2.54}$$

holds for every $q \in Q$.

(ii) There exists a subset A of Q such that $A \in B$ and $P(A) = 1$.

The requirements made upon a map $v:Q\to\mathbb{R}$ by Definition 2.14 can be restated more intuitively: It must must be consistent with the axioms of probability theory to interpret $v(q)$ as the probability of the event $Y \leq q$ for a random variable Y, whose values are elements of Q.[51] At first glance, one could be tempted to suppose that this is equivalent with the existence of an extension of the map $v:Q\to\mathbb{R}$ to a map $F:\mathbb{R}\to\mathbb{R}$, which is a legitimate CDF of a real valued random variable[52], but an example shows that this conjecture fails. Let Q be the set of all rational numbers in the interval]0, 1[, and consider the map $v:Q\to\mathbb{R}$ given by $v(q) = q$. This map has a unique extension to a CDF $F:\mathbb{R}\to\mathbb{R}$, namely $F(\xi) = 0$ for $\xi \leq 0$, $F(\xi) = \xi$ for $0 < \xi < 1$, and $F(\xi) = 1$ for $\xi \geq 1$. However, the probability distribution specified by this CDF is a rectangular distribution in the interval]0, 1[. Now the value of a random variable with this distribution is contained in Q with probability 0, and the same holds for every subset A of Q.[53] So Assertion (ii) of Definition 2.14 is violated, and the considered map $v:Q\to\mathbb{R}$ isn't a CDF with domain Q. In terms of the informal interpretation of the definition, the interpretation of $v(q)$ as probability of the event $Y \leq q$ for a random variable Y with range Q doesn't comply with the axioms of probability theory. Generalisations of the example including a necessary and sufficient criterion are discussed in Iseler (###).

In our discussion of parametric families of distribution functions, the set Q in Definition 2.14 will typically be a common range of all random variables with a distrubution belonging to some

---

[51] In some mathematical textbooks (e.g., Bauer, 1992, p. 38), the cumulative distribution function F of a real valued random variable Y is defined such that $F(\xi)$ is the probability of the event $Y < \xi$. Certainly, this approach has a very immediate relationship to the general definition of Borel sets as the coarsest σ-algebra containing all open sets in a topological space. However, to be consistent with the definition of cumulative probabilities in most psychology oriented textbooks, we use the term CDF for a function, where $F(\xi)$ is the probability of the event $Y \leq \xi$. (Note that the σ-algebra B of Borel sets in $\mathbb{R}$ is also generated by the system of intervals $]-\infty, \xi]$ with $\xi \in \mathbb{R}$.) As a consequence, these CDFs are right-continuous ($\lim_{\zeta\downarrow\xi} F(\zeta) = F(\xi)$) and not necessarily left-continuous ($\lim_{\zeta\uparrow\xi} F(\zeta) = F(\xi)$) as those treated by Bauer. (In these formulas, the expressions $\zeta\downarrow\xi$ and $\zeta\uparrow\xi$ stand for 'ζ opproaches ξ from above' resp. '... from below'.) Indeed, if $\{\xi_n\}_{n=1..\infty}$ is a non-increasing sequence of real numbers with $\lim_{n\to\infty} \xi_n = \xi$, then $\lim_{n\to\infty} F(\xi_n) = F(\xi)$ follows for our CDFs from well known properties of measures (see e.g. Bauer, 1992, p. 13, Proposition 3.2.(c)).

[52] See e.g. Bauer (1992, p. 38, Proposition 6.6) for the properties of CDFs $F:\mathbb{R}\to\mathbb{R}$ representing the distribution of a real valued random variable Y, and recall (from Footnote 51) that our interpretation of $F(\xi)$ as probability of the event $Y \leq \xi$ implies that F is right-continuous, but not necessarily left-continuous.

[53] More precisely, the probability measure P on the σ-algebra B of Borel sets in $\mathbb{R}$ is specified by the following property: For every $A \in B$, the probability P(A) is the Lebesgue-Borel measure of the set $A \cap]0, 1[$. But since Q (the set of all rational numbers in the interval]0, 1[) is countable, the Lebesgue-Borel measure of $Q \cap]0, 1[$ is zero.

considered parametric family. E.g., for the family of normal distributions, the set Q is the entire set of real numbers, and for Poisson distributions, Q is the set of non-negative integers. Some parametric families, where the range of random variables depends on the parameter as in binomial distributions, may seem to be excluded by the assumed common range Q of random variables under consideration; but the example of binomial distribution may also be used to demonstrate a solution: The set Q can be assumed to consist of all non-negative integers, and for a binomial distribution based on n 'trials', the assignment of cumulative probabilities to integers ξ with $0 \le \xi \le n$ is extended by the assignment of a cumulative probability of 1 to integers greater than n. More generally, if Q_θ is the range of random variables, whose probability distribution is the member with parameter θ in a parametric family, then the set $Q := \bigcup_{\theta\in\Theta} Q_\theta$ can be used in Definition 2.14.

Certainly, the assumption $Q \subseteq \mathbb{R}$ is a limitation for the application of results, which will be derived from Definition 2.14; but a generalisation will be outlined at the end of the present subsection.

Since the set of all VDFs with domain Q can take the role of a vocabulary set in subsequent considerations of aggregation stability, it will be denoted as V_Q. Some immediate consequences of Definition 2.14 are summarised in the following lemma, which is proved in Section 6.14.

Lemma 2.15: Let Q be a non-empty set of real numbers, and let the subset V_Q of $\mathbb{R}^Q$ be the set of all CDFs with domain Q. Then the set V_Q has the following properties:

(i) For every element v of V, there exists a unique probability measure P on the σ-algebra B of Borel sets in $\mathbb{R}$ with the properties specified by Assertions (i) and (ii) of Definition 2.14, and this probability measure is given by the equation

$$P(]-\infty, \xi]) = \sup v(Q \cap]-\infty, \xi]), \tag{2.55}$$

(ii) V_Q is convex.

(iii) For every natural number m, a family $\{v_k\}_{k=1..m}$ of linearly independent elements of V_Q exists, if and only if m isn't greater than the number of elements of Q.

(iv) If a linear combination $\sum_{k=1..m} \lambda_k v_k$ of elements v_k of V_Q results in an element of V_Q, then $\sum_{k=1..m} \lambda_k = 1$.

Under the intuitive interpretation of Definition 2.14, the claim of Lemma 2.15.(i) isn't surprising: A map $v:Q\to\mathbb{R}$, which is a CDF with domain Q, specifies a unique probability distribution, and for a random variable Y with that destribution, the probability of the event $Y \le \xi$ (with an arbitrary real number ξ) is the smallest upper bound of the function values v(q) for $q \le \xi$.

Before we use the remaining assertions of the lemma to derive general conclusions about parametric families of CDFs with domain Q, the underlying principle is demonstated by the following example.

Example 2.16: Let Q be the set of the integers from 1 through 5, and - as in Lemma 2.15 - let V_Q be the set of all CDFs with domain Q. Furthermore, let maps $\gamma_k:Q\to\mathbb{R}$ for k = 1..3 be given as

$$\gamma_k(q) = (q/5)^k, \tag{2.56}$$

and note that these maps are CDFs with domain Q.[54] Finally, let a map $\psi:\mathbb{R}^3 \to \mathbb{R}^Q$ be derived by Equation (2.5) from the map $f:Q\times\mathbb{R}^3\to\mathbb{R}$ resulting from Equation (2.30) with m = 3, and let Θ be the set of all vectors in $\mathbb{R}^3$ with non-negative components summing up to 1. Then $\psi(\theta)$ with $\theta \in \Theta$ is the map $v:Q\to\mathbb{R}$ given by

$$v(q) = \sum_{k=1..3} \theta(k) \cdot (q/5)^k. \quad (2.57)$$

Now the definition of Θ implies that v is a convex linear combination of elements of V_Q, and since this set is convex (Lemma 2.15.(ii)), $\psi(\theta)$ is a CDF with domain Θ.
In summary, the set $\psi(\Theta)$ is a parametric family of CDFs with domain Q, whose stability under mixtures follows from the results of Section 2.2.3, since the set Θ is obviously convex.

It should be noted that the family of distribution functions described in this example can be reparametrized such that a 2-dimensional parameter is sufficient. Since the components of elements of Θ must sum up to 1, we can replace $\theta(3)$ by $1 - \theta(1) - \theta(2)$. Applying this replacement in Equation (2.57) and rearranging terms, we obtain the same CDF with a 2-dimensional parameter $\theta^{\sim}$ and the equation

$$v(q) = (q/5)^3 + \sum_{k=1..2} \theta^{\sim}(k)\, ((q/5)^k - (q/5)^3). \quad (2.58)$$

for the CDF with parameter $\theta^{\sim}$ contained in a parameter space $\Theta^{\sim}$, which consists of all elements $\theta^{\sim}$ of $\mathbb{R}^2$, whose components are non-negative and sum up to a number not greater than 1.

The example can be generalised in the following procedure of constructing parametric families of CDFs with a given domain $Q \subseteq \mathbb{R}$: For a natural number m > 1, we can take an arbitrary sequence $\{\gamma_k\}_{k=1..m}$ of linearly independent CDFs with domain Q, define a function $f:Q\times\mathbb{R}^m\to\mathbb{R}$ of Class A_0 by Equation (2.30), derive a map $\psi:\mathbb{R}^m\to\mathbb{R}^Q$ by Equation (2.5), and let S be an arbitrary non-empty convex subset of the set $\psi(\mathbb{R}^m) \cap V_Q$, where V_Q is again the set of all CDFs with domain Q.[55] Then S is a parametric family of CDFs with domain Q, parameter space[56] $\Theta := \psi^{-1}(S)$ and representation

---

[54] With the definition of maps $f_k:Q\to\mathbb{R}$ by $f_k(1) = \gamma_k(1)$, and $f_k(q) = \gamma_k(q) - \gamma_k(q-1)$, a probability measure P_k fulfilling Definition 2.14 for γ_k is given by $P_k(A) = \sum_{q \in A\cap Q} f_k(q)$ for every $A \in \mathsf{B}$.

[55] Whereas $\psi(\mathbb{R}^m)$ is the set of all results of linear combinations of the vectors $\{\gamma_k\}_{k=1..m}$, the intersection $\psi(\mathbb{R}^m) \cap V_Q$ is the set of those results of such linear combinations, which are CDFs with domain Q. Since the sets $\psi(\mathbb{R}^m)$ and V_Q are convex, their intersection has the same property. Now this intersection contains the verctors γ_k. Hence it is non-empty, and an infinite number of convex subsets S of the intersection exists.

[56] According to a commonly accepted notational convention (see e.g. Dieudonné, 1985, p. 20), $\psi^{-1}(S)$ is the set of all elements θ of $\mathbb{R}^m$ with $\psi(\theta) \in S$. A general definition for sets X, Y and Φ with $\Phi \subseteq Y$ and a map $f:X\to Y$, is

$$f^{-1}(\Phi) := \{x \in X: f(x) \in \Phi\}.$$

Furthermore, if the set Φ contains only one element y of Y, we can also write $f^{-1}(y)$ instead of $f^{-1}(\{y\})$.

(continued...)

function f, the parametrisation map being the restriction of ψ to Θ. Finally, the convexity of the set S grants its stability under convex linear combinations.

Assertions (iii) and (iv) of Lemma 2.15 imply limitations of this approach. Assertion (iii) specifies the maximal length of the sequence $\{\gamma_k\}_{k=1..m}$ underlying the construction. Furthermore, since the coefficients in linear combinations of elements of V_Q, which result in an element of V_Q, sum up to 1 (Assertion (iv) of the lemma), we have $\sum_{k=1..m} \theta(k) = 1$ for every element θ of the parameter space Θ resulting from a construction of the above kind. In other words, the parametric family $\psi(\Theta)$ of CDFs with domain Q resulting from the construction is overparametrised in the way demonstrated for Example 2.16. But the reparametrisation applied in the example can be easily generalised: For every $\theta \in \Theta$, let $t(\theta)$ be the element of $\mathbb{R}^{m-1}$, whose components are identical with the first $m-1$ components of θ, and let $\Theta^{\sim}$ be the set of all results of this transformation. Then the CDF with parameter θ in the original parametrisation can also be written as a CDF with parameter $\theta^{\sim} = t(\theta)$ based on the follwing generalisation of Equation (2.58):

$$v(q) = \gamma_m(q) + \sum_{k=1..m-1} \theta^{\sim}(k)\,(\gamma_k(q) - \gamma_m(q)). \tag{2.59}$$

Note, however, that the parametric family S will still be overparametrised after this reparametrisation, if the set S is chosen such that the dimension of the smallest linear subspace of $\mathbb{R}^Q$ including S is smaller than m. But then an approach, which will be outlined immediately, can be used to obtain an entirely new parametrisation of S.

Before we discuss further limitations of the procedure, we will show that the family of Bernoulli distributions, which is erroneously claimed by Sixtl (###, p. ###) to be the only parametric family of probability distribution functions with stability under mixtures, is in fact the simplest case of a family constructed by the above procedure, including the removal of overparametrisation by Equation (2.59). Since a natural number $m > 1$ of linearly independent elements $\{\gamma_k\}_{k=1..m}$ of V_Q is required for a non-trivial result,[57] the set Q must have at least 2 elements (see Assertion (iii) of Lemma 2.15 for this conclusion). So let the set Q consist of the numbers 0 and 1, take $m = 2$, and the maps $\{\gamma_k\}_{k=1..m}$ given by $\gamma_1(0) = 0$ and $\gamma_1(1) = \gamma_2(0) = \gamma_2(1) = 1$. It can be verified by the reader that the construction of a parametric family of CDFs by the outlined procedure (with $S = \psi(\mathbb{R}^m) \cap V_Q$) and a reparametrisation by Equation (2.59) leads to $\Theta^{\sim} = [0, 1]$, and to $v(0) = 1 - \theta^{\sim}(1)$ and $v(1) = 1$ for a CDF $v{:}Q\to\mathbb{R}$ with domain Q and parameter $\theta^{\sim} \in \Theta^{\sim}$. But this is exactly the CDF of a Bernoulli distribution with probability $\theta^{\sim}(1)$ of obtaining 1.

Other limitations of the construction procedure are more indirect: The resulting representation function is always the restriction (to $Q\times\Theta$ resp. to $Q\times\Theta^{\sim}$) of a function of Class A, and a subset S of V_Q can be the result of the construction only if it is included in a finite dimensional linear subspace

---

[56] (...continued)

For later applications, it should be noted that the frequently used equality $f(S) = f^{-1}(f(S))$ is not tautological for a non-injective map $g{:}X\to Y$ and $S \subseteq X$: If x_1 and x_2 are two elements of X with $x_1 \in S$ and $f(x_2) = f(x_1)$, then x_2 is contained in $f^{-1}(f(S))$, even if it is not contained in S.

[57] It can be verified by the reader that the outlined construction of a parametric family of CDFs from a sequence $\{\gamma_k\}_{k=1..m}$ with $m = 1$ would necessarily lead to $S = \{\gamma_1\}$. But a family containing only one member is trivial, indeed.

of $\mathbb{R}^Q$.[58] It may be questioned whether these limitations apply to all parametric families of CDFs with domain Q, which are stable under convex linear combinations. With respect to the membership of the representation function in Class A, the answer to this question is negative. Instead of the reparametrisation, which has been used to remove the overparametrisation, we can take an arbitrary surjective map $t:\Theta^{\sim}\rightarrow\Theta$, where Θ is the parameter space resulting from the original construction before the reparametrisation, and $\Theta^{\sim}$ may be any non-empty set which is large enough to be mapped surjectively onto Θ. Then a representation function $f^{\sim}:Q\times\Theta^{\sim}\rightarrow\mathbb{R}$ defined by $f^{\sim}(q, \theta^{\sim}) := f(q, t(\theta^{\sim}))$ will not be the restriction to $Q\times\Theta^{\sim}$ of a function of Class A, unless there is a natural number n such that $\Theta^{\sim} \subseteq \mathbb{R}^n$ and the map t can be extended to an affine map $\mathbb{R}^n\rightarrow\mathbb{R}^m$.[59] So the limitation to representation functions, which are restrictions of functions of Class A, can be overcome by reparametrisations.

But for the inclusion of a parametric family of CDFs with range Q in a finite dimensional subspace of $\mathbb{R}^Q$, Corollary 2.13 shows that this property as well as the convexity of the set of all members of the family is necessary for stability under mixtures, if the parameter space is a subset of $\mathbb{R}^m$ for some natural number m. Given this limitation, we can also modify the construction procedure and start with an arbitrary set S of CDFs with range Q, which will be the set of all members of the parametric family to be constructed. If S is not convex, then it isn't stable under mixtures. If it is convex, but not included in a finite dimensional subspace of $\mathbb{R}^Q$, then there will be no parametrisation map $\psi:\Theta\rightarrow\mathbb{R}^Q$ with $S = \psi(\Theta)$ and $\Theta \subseteq \mathbb{R}^m$ for any natural number m. But if S is convex and contained in a finite dimensional subspace of $\mathbb{R}^Q$, we can take an arbitrary sequence $\{\gamma_k\}_{1..m}$ of elements of S forming a basis of the smallest subspace of $\mathbb{R}^Q$ including S,[60] and the previously outlined procedure can be used to parametrise the set S.

The main results for parametric families of CDFs with domain $Q \subseteq \mathbb{R}$ can be summarised in the following statements: Every convex set S of CDFs with domain Q, which is included in a finite dimensional linear subspace of $\mathbb{R}^Q$, is stable under mixtures and can be parametrised with a parameter space $\Theta \subseteq \mathbb{R}^m$ for some natural number m. Conversely, every parametric family of CDFs with domain Q is a convex subset of a finite dimensional linear subspace of $\mathbb{R}^Q$, if the family is stable under mixtures and if its parameter space is a subset of $\mathbb{R}^m$ for some natural number m.

It is left to generalise these results to other kinds of probability distribution functions. For

---

[58] More precisely, the set $\psi(\mathbb{R}^m)$, which includes S by the construction principle, is an m-diemsionensional linear subspace of $\mathbb{R}^Q$. Proof: The function $f:Q\times\mathbb{R}^m\rightarrow\mathbb{R}$, which is used for the construction, is of Class A_0; hence the map $\psi:\mathbb{R}^m\rightarrow\mathbb{R}^Q$, which is derived from f by Equation (2.5), is injective and linear (Estes-Theorem 2.4): But the image of an m-dimensional real vector space ($\mathbb{R}^m$) under an injective and linear map is an m-dimensional real vector space.

[59] An application of Lemma 2.5 (with interchanged roles of f and $f^{\sim}$) leads to $\psi^{\sim} = \psi \circ t$ for the new parametrisation map. Furthermore, it follows from Estes-Theorem 2.4 that $\psi^{\sim}$ is extendable to an affine map, if and only if $f^{\sim}$ is the restriction to $Q\times\Theta^{\sim}$ of a function of Class A. Finally, the linearity of the original map ψ is granted by the membership of f in Class A_0. So it follows from well known concatenation properties of linear and affine maps that $f^{\sim}$ is the restriction to $Q\times\Theta^{\sim}$ of a function of Class A, if and only if the map t can be extended to an affine one.

[60] See Lemma 2.12 for the existence of suitable vectors.

distribution functions $v:Q\to\mathbb{R}$ with $Q \subseteq \mathbb{R}$, where the function values v(q) represent probabilities of events Y = q or Y < q, as well as for probability density functions, the above considerations (including the construction of parametric families, which are stable under mixtures) can be transferred almost verbatim after a due alteration of Equation (2.54) and reinterpretation of the set V_Q. To overcome the limitation to probability distributions of real valued random variables, let Ω be an arbitrary non-empty set, and Q a (non-empty) system of subsets of Ω, which is closed under finite intersections.[61] E.g., Q can - but must not necessarily - be a σ-algebra in Ω. Then the extension of a map $Q\to\mathbb{R}$ to a probability measure P on the σ-algebra generated by Q is unique, if it exists. So let V_Q be the set of all maps $Q\to\mathbb{R}$ with this property. Then V_Q is again a convex subset of the function space $\mathbb{R}^Q$ of all maps $Q\to\mathbb{R}$. So Assertions (i) and (ii) of Corollary 2.13 are equivalent for every nonempty subset S of $\mathbb{R}^Q$, and particularly for subsets of V_Q.

2.4 Individual and Aggregated Pairs of Probability Distributions

Allgemein wird Expl-Deduc in Section 5.4 behandelt! Hier kürzen!!!

The frequently criticised 'loose' derivation of statistical predictions from psychological hypotheses can sometimes be coped with by a subdivision of the derivation into two steps: An *explication* of the psychological hypothesis in the framework of probability theory and statistics, which enables the strict *deduction* of a testable statistical prediction. An example (whose mathematical claim will be proved later) may demonstrate the issue.

Example 2.17: Assume that a scientific topic is presented in two textbook-like documents a and b, and that a psychological hypothesis claims the superiority of version b with respect to the achieved understanding of the topic, which is measured by an ad-hoc constructed test. The commonly applied loose derivation could lead to the prediction that the average testscores in randomised independent samples will be significantly lower for version a than for version b. This derivation can be subdivided into the following steps:

- It can be considered to explicate the psychological hypothesis[62] in the following way: For every person ('unit') u belonging to a domain D, there are (hypothetical) probability distributions of test scores (random variable Y) for both experimental conditions, and the expectation of the respective distribution (i.e., the 'true score') will be lower under condition

[61] A system Q of subsets of a set Ω is finite under closed intersections iff the intersection of two elements of Q is also contained in Q. Then the property $\bigcap_{i=1..n} A_i \in Q$ follows by induction over n for all finite sequences with $A_i \in Q$. See Bauer (###, p. ###) for the uniqueness of the extension of a map $Q\to\mathbb{R}$ to a probability measure on the σ-algebra generated by a set system Q with this property.

[62] More precisely, the object of an explication is a concept and not a proposition like a hypothesis. In the example, the predicate '... has a lower expectation of Y under condition a than under condition b' explicates the superiority of text version b, which is claimed by the psychological hypothesis.

a than under condition b.

If this explication is accepted, then strict mathematical deduction allows a prediction referring to RSO-processes of the kind described in Section 2.1. Let the random variable Y be the result (i.e., the testscore) in the process of selecting a unit at random, presenting one of the two texts, and administering the test. Since RSO-processes can be performed with different selection distributions,[63] the probability distribution of the random variable Y will depend not only on the presented version of the text, but also upon the selection distribution. Nevertheless, the following prediction follows from the above explication of the psychological hypothesis:

- For a given selection distribution, the expectation of Y is smaller under condation a than under condition b.

This prediction is testable, if we consider the testscores of two randomised samples as realisations of the two distributions of Y for RSO-processes with identical selection distributions.[64]

The advantage of partitioning the derivation into an explicative and a deductive step rests upon the criteria of adequateness, which are different for both steps, but considerably stricter for each one of them than those of the commonly applied derivation in one step. Briefly speaking, the derivation of a statistical prediction from a psychological is considered adequate in this approach, if it can be reconstructed as a concatenation of an adequate explication and a strict deduction.

Since aspects of stability under aggregation play only a subordinate role in the first step,[65] a discussion of the adequateness of the explication is shortcircuited for a first approach to the example by the clause 'if this explication is accepted'. But the deduction of a testable prediction is based on the stability under aggregation of a property, which is claimed for units in the explicated hypothesis, and for RSO-processes in the prediction: The expectation of Y is lower under condition a than under condition b. The main objective of the rest of the present subsection will consist in the application of an SSA, which can be used to subsume the aggregation stability of this property and similar ones under Lemma 2.1.

Before we present the SSA, it should be mentioned that a formally simpler explication of the psychological hypothesis is possible as long as we are interested only in expectations. Whereas the 'true scores' of individuals under the two conditions have been introduced above as expectations of hypothetical probability distributions characterising units for given experimental conditions, these true scores could also be identified more directly with conditional expectations of the random variable Y (the test scores) in a probability space representing the entire RSO-process. Under this approach, a suitable vocabulary set V would consist of ordered pairs of real numbers such that an individual is characterised by the true scores under both conditions, and the entire RSO-process by expectations, which are conditioned only upon the applied condition. Compared with the simplicity of this vocabulary set, it may look like an unnecessary roundabout way to start with hypothetical probability distributions for individuals. But since we will also discuss other explications of the

[63] Recall (from Section 2.1) that the selection distribution is formed by the selection probabilities λ_i.

[64] See Section ### for this issue.

[65] In empirical sciences, an adequate explication has to enable testable predictions.

psychological hypothesis, which will not be based on true scores, the assumed probability distributions will be a convenient common basis for all explications to be considered.

For the same reasons as in the study of parametric families of distribution functions (Section 2.3), probability distributions of the dependent variable Y will subsequently be represented by cumulative probabilities. Given this approach, there are two almost equivalent ways to define a vocabulary set. Maintaining the denotation V_Q for the set of all CDFs with domain Q (where Q is the set of all possible values of the dependent variable Y), we can characterise units as well as RSO-processes for given selection distributions by ordered pairs (v_a, v_b) of elements of V_Q representing the respective probability distributions under the conditions a and b. Then a vocabulary set of an SSA would be the set of all such ordered pairs - i.e., the Cartesian product $V_Q \times V_Q$.

Although this approach may seem natural for the special situation of Example 2.17, later generalisations (e.g. to situations with more than two conditions) will be facilitated by a slight reformulation. With the denotations C for the set of all experimental conditions under consideration (i.e., C := {a, b} for Example 2.17), and W for the set of possible values of the dependent variable Y, a suitable vocabulary set V consists of all maps $v{:}C \times W \to \mathbb{R}$, where the partial map v(c, .) for each $c \in C$ is a CDF with domain W. This specification can be reformulated more intuitively in a similar way as the definition of a CDF. The vocabulary set V consinsts of all maps $v{:}C \times W \to \mathbb{R}$, where the following interpretation is consistent with the axioms of probability theory: If a given unit (resp. an RSO-process with a given selection distribution) is characterised by a map $v{:}C \times W \to \mathbb{R}$ contained in the vocabularity set V, then the function values v(c, w) is the probability of the event $Y \le w$ under the assumption that condition c is applied to the given (resp. to the randomly selected) unit. Beyond enabling easy generalisations to situations with more conditions, this vocabulary set has another convenient property: With the definition Q := C×W, it is a subset of the function space $\mathbb{R}^Q$ of all maps $Q \to \mathbb{R}$, and this inclusion will allow the immediate application of general results referring to subsets of such spaces, which will be presented in later sections.

Although the aggregation rule underlying Lemma 2.1 has been introduced as an axiom for Section 2, its application to a vocabulary set with this interpretation should be made explicit. The map $\Phi_\pi{:}C \times W \to \mathbb{R}$ resulting from Equation (1.2) characterises an RSO-process with selection probabilities λ_i for units, whose behavioural dispositions are specified by individual maps $v_i{:}C \times W \to \mathbb{R}$. Applying Equation (1.2) to a specific element of the set Q (i.e., to an ordered pair $(c, w) \in C \times W$), we obtain

$$\Phi_\pi(c, w) = \textstyle\sum_{i=1..n} \lambda_i \cdot v_i(c, w). \qquad (2.60)$$

Under the intuitive reinterpretation of the set V, this means: If a unit is selected at random and tested under condition c, then the probability of the event $Y \le w$ results from the respective 'individual probabilities' $v_i(c, w)$ by a convex linear combination, whose coefficients λ_i are the selection probabilities. If the function values $v_i(c, w)$ are conceived as conditional probabilities (given the selection of unit i), then Equation (2.60) follows from well known properties of conditional probabilities. Note, however, that the interpretation of $v_i(c, w)$ as a conditional probability isn't introduced as a constituent assumption in our analysis, but only as a conceivable way of giving a meaning to the assumptions of Lemma 2.1.

Having introduced the vocabulary set V, we turn to the aggregation stability of the property, which is claimed for all individuals by the explicated hypothesis in Example 2.17: The true score (i.e., the expectation of the dependent variable Y) is assumed to be lower under condition a than under condition b. Observe first that this claim rests upon the tacid assumption that the expectations exist. Since generalisations will be discussed in later sections, we can save problems with the existence of expectations by the additional assumption that the set W of possible values of the dependent variable is bounded by a real number ξ such that $|w| \leq \xi$ for every $w \in W$. Given this property, we can introduce the notation $\mu'_c(v)$ (with $c \in C$ and $v \in V$) for the expectation of a W-valued random variable, whose CDF is given by the partial map v(c, .).[66] In the intuitive interpretation of the vocabulary set V, $\mu'_c(v)$ is the expectation of a random variable Y, where v(c, w) is the probability of the event $Y \leq w$. With this definition, the set

$$S_E := \{v \in V: \mu'_a(v) < \mu'_b(v)\} \tag{2.61}$$

consists of all elements of the vocabulary set with the property, which is claimed for all individuals by the explicated hypothesis in Example 2.17.

According to Lemma 2.1, the aggregation stability of this set can be established by a proof of its convexity.[67] But to prepare generalisations in later sections, we resume an observation, which was made in the review of the results of Estes (1956). There we noted down to pay attention to situations, where an aggregation rule implies the applicability of a similar process of aggregation to mathematical objects not mentioned explicitly in the aggregation rule. Indeed, although the expectations $\mu'_c(v)$ are not mentioned explicitly in the aggregation rule (Equation (2.60)), this rule implies the equation

$$\mu'_c(\Phi_\pi) = \sum_{i=1..n} \lambda_i \, \mu'_c(v_i) \tag{2.62}$$

for every $c \in C$.[68] Now assume that (S_E, π) with $\pi := \{(v_i, \lambda_i)\}_{i=1..n}$ is contained in the relation H of Lemma 2.1, which means that the implication $\lambda_i > 0 \Rightarrow v_i \in S_E$ holds for i = 1..n. Recalling that the coefficients λ_i are non-negative numbers summing up to 1 (Lemma 2.1.(ii)), we obtain the inequality

---

[66] Recall from Lemma 2.15.(i) that a CDF with domain W specifies a unique probability measure on the σ-algebra $\mathcal{B}$ of Borel sets in $\mathbb{R}$, which can be interpreted as the distribution of a random variable, whose values are almost surely elements of W.

[67] A proof of the convexity of the set S_E can be based on Inequality (2.63) with n = 2, where v_1 and v_2 are elements of S_E, $\lambda_1 \in]0, 1[$, and $\lambda_2 = 1 - \lambda_1$.

[68] A proof of Equation (2.62) can be based on a repeated application of the following basic property of Lebesgue integrals: Let μ, μ_1 and μ_2 be finite measures on a measurable space $(\Omega, \mathcal{A})$ such that $\mu = \lambda_1 \mu_1 + \lambda_2 \mu_2$ with non-negative real numbers λ_1 and λ_2. (More explicitly, this means $\mu(A) = \lambda_1 \mu_1(A) + \lambda_2 \mu_2(A)$ for every $A \in \mathcal{A}$.) Then the equation

$$\int X \, d\mu = \lambda_1 \int X \, d\mu_1 + \lambda_2 \int X \, d\mu_2$$

holds for every measurable map $X:\Omega \to \mathbb{R}$, where the integrals on the right hand side of the equation are finite. Although this property isn't mentioned explicitly in standard textbooks of integration theory, it follows immediately from the definition of Lebesgue integrals. A proof of a more gerneral property, which implies Equation (2.62) as a special case, is given in Section ###.

$$\mu'_b(\Phi_\pi) - \mu'_a(\Phi_\pi) = \sum_{i=1..n} \lambda_i\ (\mu'_b(v_i) - \mu'_a(v_i)) > 0 \tag{2.63}$$

from Equation (2.62), and this means $\Phi_\pi \in S_E$. But if $(S_E, \pi) \in H$ implies $\Phi_\pi \in S_E$, then the set S_E is stable under the considered kind of aggregation.

The consequence of this stability for Example 2.17 is obvious: The testable prediction of a differerence in the expectations of the dependent variable in RSO-processes with identical selection distribution can be derived by strict deduction from the explicated psychological hypothesis. It could be argued that this very result rehabilitates the commonly applied intuitive derivation of the same prediction in one step. Indeed, this argument would question the usefulness of subdividing the derivation into an explicative and a deductive step, if similar results could be obtained for explications of the psychological hypothesis in terms of other numbers (e.g. medians), which are commonly considered to be legitimate substitutes for expections in their function as measures of central tendeny. But for medians, it has been pointed out by Iseler (1996) that an order, which is present in all aggregated units, may be reversed as a result of aggregation.

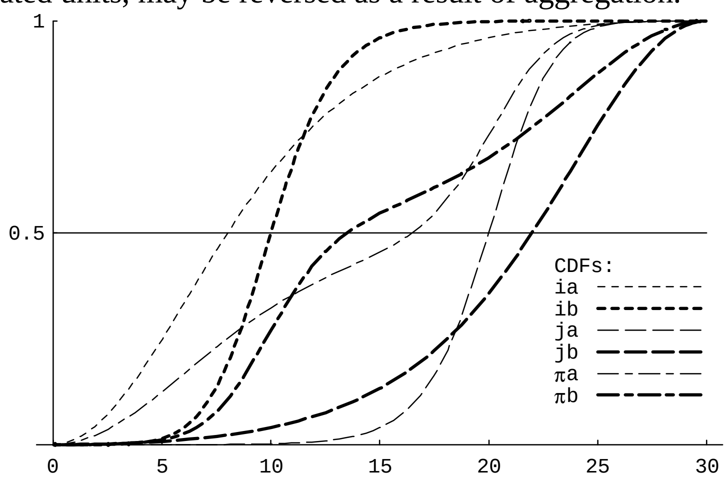

Figure 2.2 Demonstration of the median paradox by CDFs. For unit *i* (short dashes) and unit *j* (long dashes), the median is smaller under condition *a* (thin plot) than under condition *b*; but for RSO-processes with selection probabilities $\lambda_i = \lambda_j = 0.5$ (alternating short and long dashes), the median is greater under condition *a*.

For an example of this phenomenon, which is called the median paradox, consider Figure 2.2, where CDFs are plotted for hypothetical distributions of a dependent variable Y under two conditions a (thin plot) and b, these CDFs referring to two units i (short dashes) and j (long dashes) and to RSO-processes with selection probabilities 0.5 for both units (alternating short and long dashes). Since the respective medians can be read as the abscissas of the intersections between the CDFs and the horizontal line with an ordinate of 0.5, the plots show that the median is smaller under condition a than under condition b for the two units, but greater under condition a for the RSO-processes.[69]

To apply the formalisation introduced for Example 2.17, assume that the dependent variable Y represents response times in a task, which is solved almost surely within 30 seconds.[70] With $C := \{a, b\}$ and $W := [0, 30]$, let the vocabulary set V consist of all maps $v: C \times W \to \mathbb{R}$, where the partial maps v(c, .) are strictly increasing, continuous CDFs with domain W.[71] Furthermore, let S_M be the subset of those elements of V, where the median is smaller in the CDF for condition a than for condition b. Then the example plotted in Figure 2.2 shows that this set isn't convex, and this excludes stability under convex linear combinations.

Although it is common practice to explicate hypotheses about a 'positive' effect of a condition b (relative to condition a) upon a dependent variable in terms of characteristics of central tendency like expectations or medians, a drawback of such explications can also be demonstrated by Figure 2.2. For unit i (short dashes), the median as well as the expectation[72] is lower under condition a than under condition b. But above the intersection of the CDFs, there is a clear tendency for higher values under condition a, and the same occurs below the intersection of the CDFS for unit j. One could also say that the scale of the dependent variable is partitioned for both units at the repsective intersection points of the CDFs such that the tendency for higher values under condition b applies to unit i only in the lower scale partition and to unit j only in the upper one. The explication of causal effects by the order of expectations or medians implies different rules of balancing such opposite effects; but it may be questioned whether at all a balancing of this kind is intended by a psychological hypothesis. To clear this issue is one of the objectives of a careful explication.

An explication of a positive effect of condition b vs. a without such balancing can be based upon concepts of *stochastic order*. For a vocabulary set V consisting of all maps $v: C \times W \to \mathbb{R}$, where the partial maps v(c, .) are CDFs with domain $W \subseteq \mathbb{R}$, the weakest form of stochastic order is represented by the set

$$S_O := \{v \in V: \forall\, w \in W: v(a, w) \geq v(b, w)\}. \tag{2.64}$$

But since elements of V with identical CDFs v(a, .) and v(b, .) are contained in this set, some additional requirement is necessary to obtain a positive effect of condition b. There are several options for such additional requirements. The weakest one is the existence of some element w of W

[69] See Iseler (1996) for a more detailed description of the plotted curves and for numerical computations of the medians.

[70] Think, e.g., of a visual recognition task, where the stimulus is presented vaguely at t = 0 and sharpened continuously within 30 seconds, unless it is recognised earlier by the subject. The conditions a and b could reflect variations in the learning phase.

[71] The assumption of strictly increasing, continuous CDFs may be weakened, if the existence of well defined medians and the convexity of the vocabulary set are maintained.

[72] The expectations under conditions a and b are 8.9377 resp. 10.1598 for unit i, and 19.8402 resp. 21.0623 for unit j. These values are results of numerical integration based on the functions specified by Iseler (1996).

with v(a, w) > v(b, w). The resulting *strict stochastic order* is represented by the set

$$S_S := \{v \in S_O \colon \exists\, w \in W\colon v(a, w) > v(b, w)\}. \tag{2.65}$$

The convexity of both sets, which implies their stability under convex lienear combinations, is easily verified. For given elements v_1 and v_2 of the set S_O and a real number $\lambda \in\,]0, 1[$, the definition

$$v := \lambda\, v_1 + (1-\lambda)\, v_2 \tag{2.66}$$

leads to

$$v(a, w) - v(b, w) = \lambda\, (v_1(a, w) - v_1(b, w)) + (1-\lambda)\, (v_2(a, w) - v_2(b, w)) \geq 0 \tag{2.67}$$

for every $w \in W$, which makes v an element of S_O. Furthermore, if v_1 and v_2 are also elements of S_S, then take any element w of W with $v_1(a, w) > v_1(b, w)$, and the assumption $\lambda \in\,]0, 1[$ allows to sharpen Inequality (2.67) into $v(a, w) - v(b, w) > 0$. So v is also contained in S_S. Again, the consequence for the deduction of predictions from an explicated psychological hypothesis is obvious: If strict stochastic order (i.e., $v_i \in S_S$) is claimed for every unit, then this property follows for every map $C \times W \to \mathbb{R}$ characterising RSO-processes with identical selection distributions.

It has already been mentioned that the requirement v(a, w) > v(b, w) for some element w of W can be strengthened. E.g., it could also be considered to postulate this inequality for every non-maximal element w of W.[73] If v_1 and v_2 are elements of V with this property, then it is again easy to derive it for the map v (Equation (2.66)) by a suitable sharpening of Inequality (2.67).

Before we finish the analysis of stability under convex linear combinations for the various types of stochastic order, we should return to the set S_S and consider an aspect, which will be a challenge to our general conceptualisation of stability under aggregation. Let S_{st} be the set of all elements v of the vocabulary set, where the strict order v(a, w) > v(b, w) holds for some element w of W. In other words, the set S_{st} represents the additional property, which is required beyond membership in the set S_O for strict stochastic order (see Equation (2.65)). It is easily verified that this property isn't generally stable under convex linear combinations.[74] However, under the additional assumption of the weakest stochastic order, which is represented by the set S_O, the set S_{st} becomes stable under convex linear combinations. Indeed, the sharpening of Inequality (2.67) under the assumption $v_1(a, w) > v_1(b, w)$ was possible only under the additional condition $v_2(a, w) \geq v_2(b, w)$. So we have another example of the conditional aggregation stability, which has already been encountered in the results of Sidman (1952) in Section 2.2.1. Since this conditional aggregation stability is not formalised explicitly in the definition of an SSA, its implementation is noted down as a task for the general analysis of SSAs in Section 3.

---

[73] If a maximal element w of W exists, then the definition of the set V forces cumulative probabilities v(a, w) and v(b, w) of 1 for that element, and the inequality v(a, w) < v(b, w) cannot hold.

[74] It can be left to the reader to construct examples - e.g. for W = {1, 2, 3} - with $v_1 \in S_{st}$ and $v_2 \in S_{st}$ such that the map v given by Equation (2.66) is not contained in S_{st}.

2.5 Pooling of Intraindividual and Interindividual Variation

Many examples of lacking stability under aggregation can be attributed to a fact, which is well known from analysis of variance: If several groups are pooled, then the variance in the pooled sample is made up both by the variance within groups and by the variance between groups. E.g., if there exist two measurement values Y_a and Y_b for every subject and the variance is smaller for Y_a than for Y_b in each group, then the apposite can hold for the pooled group, if the variance between groups is considerably greater for Y_a than for Y_b.

The same can also happen in RSO-processes. For a demonstration, we resume the set V of the preceeding subsection, which consists of those maps $v{:}C{\times}W{\to}\mathbb{R}$ (with $C = \{a, b\}$ and a given bounded set $W \subseteq \mathbb{R}$) where the partial maps v(a, .) and v(b, .) are CDFs with domain W. Now let S be the set of those elements v of V, where the variance of the probability distribution specified by the CDF v(a, .) is smaller than the one resulting from v(b, .). To analyse these variances, we resume the denotation $\mu'_c(v)$ (with $c \in C$) for the expectation of a random variable Y, whose distribution is specified by the CDF v(c, .), and add the denotation $\mu''_c(v)$ for the expectation of Y^2, where Y is again a random variable with CDF v(c, .). Then the variance of a random variable Y with CDF v(c, .) is $\mu''_c(v) - \mu'_c(v)^2$, and the set S can be specified by the equation

$$S := \{v \in V{:}\ \mu''_a(v) - \mu'_a(v)^2 < \mu''_b(v) - \mu'_b(v)^2\}. \qquad (2.68)$$

For a convex linear combination $\Phi_\pi = \sum_{i=1..n} \lambda_i v_i$, Equation (2.62) can again be used for $\mu'_c(\Phi_\pi)$, and the equation

$$\mu''_c(\Phi_\pi) = \sum_{i=1..n} \lambda_i \mu''_c(v_i) \qquad (2.69)$$

results from similar considerations. For the variances, Equations (2.62) and (2.69) can be combined to get

$$\mu''_c(\Phi_\pi) - \mu'_c(\Phi_\pi)^2 = \sum_{i=1..n} \lambda_i \mu''_c(v_i) - (\sum_{i=1..n} \lambda_i \mu'_c(v_i))^2. \qquad (2.70)$$

Subtracting and adding the sum $\sum_{i=1..n} \lambda_i \mu'_c(v_i)^2$ on the right hand side of this equation and rearranging terms, we obtain

$$\sum_{i=1..n} \lambda_i (\mu''_c(v_i) - \mu'_c(v_i)^2) + \sum_{i=1..n} \lambda_i \mu'_c(v_i)^2 - (\sum_{i=1..n} \lambda_i \mu'_c(v_i))^2, \qquad (2.70')$$

and in this form, we see that the first sum is a convex linear combination (with coefficients λ_i) of within-variances (i.e., variances for $v_i(c, .)$), whereas the rest is a between-variance as in the analysis of variance. From this result, it can be anticipated that the set S given by Equation (2.68) is non-convex: If the between-variance is sufficiently large under condition a and small under condition b, then the variance given by Equation (2.70) may be larger under condition a than under condition b, even if all elements v_i entering the RSO-process are contained in the set S defined by Equation (2.68). It can be left to the reader to verify this conjecture by an example with $n = 2$, a large between-variance for $c = a$ and a small one for $c = b$.

The same approach can also be used to anticipate and to verify the aggregation stability of

another set. For a given real number $\xi > 0$, let S_ξ be the set consisting of those elements v of V where the variance of a random variable with CDF v(a, .) is greater than ξ. For elements v_i of this set, the first sum in (2.70') is a convex linear combination of numbers greater than ξ. So the result is greater than ξ, and this property is maintained, if the (non-negative) between-variance is added.

###

2.6 Mixtures of Hazard Gambles

###

2.7 Aggregation of Processes

###

3 General Tools for the Study of Aggregation Stability

In the analysis of stability under averaging in Section 1, the convexity of a subset S of a vocabulary set V turned out to be a sufficient condition of aggregation stability, and for stability under convex linear combinations, convexity is necessary and sufficient (Lemma 2.1). However, an example to be presented in Subsection 3.1 will show that the association between convexity and aggregation stability can be lost, if the number of units entering the aggregation is infinite. Certainly, such problems will not arise if the elements of the domain D in an RSO-process are persons, since there will always be only a finite number of persons available for selection. But as soon as we consider a domain D of 'persons in situation', where the situation is a configuration of continuous variables, then the set D will be infinite. So it becomes necessary to look for other properties than convexity, whose presence in a set of interest grants its stability under aggregation.

Whereas the presentation of such properties for a rather general class of SSAs will be the objective of Section 4, the present Section 3 will prepare some general tools, which are immediate consequences of the SSA-Axioms (Assertions (i), (ii), (iii) and (iv) of Definition 1.2). In fact, the most important results have some resemblance to tools for the analysis of convexity. E.g., it turned out helpful in preceeding sections to represent a set of interest as an intersection of other sets, whose convexity is proved more easily. Similarly, Subsection 3.6 will establish the aggregation stability for the intersection of sets, which are stable under some kind of aggregation. Another useful property for the analysis of convexity is its invariance under affine maps. (Recall from Section 2.2.3 that the image of a convex parameter space $\Theta \subseteq \mathbb{R}^m$ under an affine parametrisation map $\psi:\mathbb{R}^m \to \mathbb{R}^Q$ is convex.) A similar approach to stability under aggregation by maps between the vocabulary sets of SSAs will be presented in Subsection 3.5. Furthermore, it has been noted down on the occasion of examples in preceeding sections that two kinds of stability, which are not covered explicitly by Definition 1.2, deserve attention: Conditional and equivalential stability under aggregation. Tools for their analysis will be introduced in Subsections 3.7 and 3.8.

Although the results to be presented in this section apply to all structures of aggregation stability, the understanding of their tool function may be supported by their application to an example, which will be introduced in Subsection 3.2. Beyond enabling these applications, the example will also prepare the definition of a rather general class of SSAs in Section 4, since it is itself a member of the class.

3.1 An Example of Lacking Aggregation Stability in a Convex Set

The following subsection will present an example of an RSO-process, where a convex subset of a vocabulary set is not stable under aggregation.

The mathematical background of the example mounts up to a demonstration of the relevance of an essential feature in the definition of a linear combination: Only finite sums $\sum_i \lambda_i x_i$ with scalars λ_i and vectors x_i are linear combinations, and only under this limitation the result of a convex linear combination of elements of a convex set is contained in the set. For an example with an infinite sum, consider the function space $\mathbb{R}^Q$ of all maps $Q \to \mathbb{R}$, where Q is the set of all natural numbers, and let V be the set of all maps $v:Q \to \mathbb{R}$ with $0 \le v(q)$ for every natural number q, and $\sum_{q=1..\infty} v(q) = 1$. Furthermore, let S_f be the subset of those elements of V, where a non-zero function value v(q) occurs only for a finite set of arguments q. Now let a sequence $\{\lambda_i\}_{i=1..\infty}$ of real numbers be given by $\lambda_i = 2^{-1}$, and a sequence $\{v_i\}_{i=1..\infty}$ of elements of the set S_f by $v_i(q) = i^{-1}$ for $q \le i$, and $v_i(q) = 0$ for $q > i$.

Before the sum $\sum_{i=1..\infty} \lambda_i v_i$ is considered for this situation, it has to be mentioned that infinite sums are defined in real vector spaces only if they are endowed with a topology. Since an approach to aggregation stability, which is based on topological vector spaces, will only be outlined in Section 4.8, the following statement should be sufficient for the moment: Under a suitable topology[75], the pointwise addition of maps $Q \to \mathbb{R}$ is extended to infinite sums. So the sum $\sum_{i=1..\infty} \lambda_i v_i$ with scalars λ_i and maps $v_i:Q \to \mathbb{R}$ results in a map $v:Q \to \mathbb{R}$ with $v(q) = \sum_{i=1..\infty} \lambda_i v_i(q)$ for every $q \in Q$, if a map v with this property exists; otherwise the sum doesn't converge. For the assumed situation, this means

$$v(q) = \sum_{i=q..\infty} i^{-2}/i, \tag{3.1}$$

and the map v with this specification is not contained in the set S_f, since it assignes a non-zero function value v(q) to every natural number q. Note, however, that the set S_f is convex[76] and that the coefficients λ_i are non-zero real numbers summing up to 1 as in a convex linear combination

The main result of the example can be summarised in the following statement: If infinite sums of maps $Q \to \mathbb{R}$ are understood pointwise, then a sum $\sum_{i=1..\infty} \lambda_i v_i$ with non-zero coeffincients λ_i

---

[75] Readers, who are familiar with topological vector spaces, can think of the coursest linear toplogy in $\mathbb{R}^Q$, where all projection maps $pr_q:v \to v(q)$ are continuous.

[76] For elements v' and v'' of S_f and a scalar $\lambda \in [0, 1]$, define $v := \lambda v' + (1-\lambda)v''$. Furthermore, let B' resp. B'' resp. B be the sets of all natural numbers q with $v'(q) > 0$ resp. $v''(q) > 0$ resp. $v(q) > 0$. Then $B \subseteq B' \cup B''$. So B is finite, since B' and B'' are finite.

summing up to 1 and elements v_i of a convex subset S of $\mathbb{R}^Q$ doesn't necessarily result in an element of the set S. To demonstrate the relevance of this result for stability under aggregation, it is subsequently interpreted as an RSO-process with an infinite domain set D.

Example 3.1: Assume that D is an infinite, but countable set $\{u_1, u_2, ..., u_i, ...\}$ of units, which are random number generators producing natural numbers. With the denotation Q for the set of all natural numbers, let a suitable vocabulary set V constist of all maps $v{:}Q\to\mathbb{R}$ with $0 \le v(q)$ for every $q \in Q$ and $\sum_{q=1..\infty} v(q) = 1$ under the interpretation that v(q) is the probability of producing the number n. In particular, let each unit u_i be characterised by a map $v_i{:}Q\to\mathbb{R}$ with $v_i(q) = i^{-1}$ for $q \le i$, and $v_i(q) = 0$ for $q > i$.

Now assume that one of the units is randomly selected to produce one random number. Then the entire process can be modelled by a probality space (Ω, A, P) with $\Omega := D\times Q$ and $\mathsf{A} := \mathsf{P}\Omega$, the element (u, q) of the set Ω representing the outcome that unit u is selected and generates the number q. In this probability space, the selected unit can be represented by a random variable U defined by $U(u_i, q) := u_i$, and the generated number by a random variable Y with $Y(u_i, q) := q$.[77]

As a sencond element of an SSA, let Π be the set of all 'selection distributions' - i.e., probability measurures on $\mathsf{P}D$. Since the probability measure P in the above probability space will depend upon the selection distribution, we introduce the notation P_π for the probability measure on A resulting from selection distribution π. With the denotation $\pi(u_i)$ for the selection probability of unit u_i under selection distribution π, the interrelationship between π and P_π is given by the equations

$$\pi(u_i) := P_\pi(U = u_i) = \textstyle\sum_{q=1..\infty} P_\pi(u_i, q), \tag{3.2}$$

and

$$P_\pi(u_i, q) := P_\pi(U = u_i, Y = q) = \pi(u_i) \cdot v_i(q). \tag{3.3}$$

Note that the last equality mounts up the an axiom-like interpretation of the maps $v_i{:}Q\to\mathbb{R}$: For every $u_i \in D$ and $q \in Q$, it is assumed that the conditional probability $P_\pi(Y = q \mid U = u_i)$ is identical under all selection distributions π with $\pi(u_i) > 0$, and these conditional probabilities are collected in the map $v_i{:}Q\to\mathbb{R}$, which characterises unit u_i.

For an aggregation rule $\Phi{:}\Pi\to V$, let Φ_π be the element of V describing the distribution of the random variable Y under selection distribution π; i.e.,

$$\Phi_\pi(q) := P_\pi(Y = q) = \textstyle\sum_{i=1..\infty} P_\pi(u_i, q) \tag{3.4}$$

for every $q \in Q$.

Finally, let the relation $H \subseteq \mathsf{P}V\times\Pi$ be given by the following specification: An ordered pair (S, π) with $S \subseteq V$ and $\pi \in \Pi$ is contained in H iff the implication $\pi(u_i) > 0 \Rightarrow v_i \in S$ holds for every $u_i \in D$. Then the set system T completing the SSA (V, Π, Φ, H, T) is specified by SSA-Axiom (iv).

Again, let S_f be the set of all elements v of the vocabulary set V, where v(q) is non-zero only for a finite set of natural numbers q. Although this set is convex, it isn't stable under the considered aggregation. To establish this claim, it suffices to present a selection distribution π with $(S_f, \pi) \in H$ and $\Phi_\pi \notin S_f$. Indeed, the selection distribution π defined by $\pi(u_i) := 2^{-i}$ has both properties: Under the assumptions $v_i(q) = 1/i$ for $q \le i$ and $v_i(q) = 0$ for $q > 0$, the containment $(S_f, \pi) \in H$ follows immediately from the definition of the relation H; but Equations (3.3) and

[77] More explicitly, U is a random variable on $\mathsf{P}D$, and Y a random variable on $\mathsf{P}Q$.

(3.4) lead to $\Phi_\pi(q) = \sum_{i=q..\infty} 2^{-i}/i > 0$ for every natural number q, and this implies $\Phi_\pi \notin S_f$.[78]

As a consequence for our further considerations, convexity cannot be taken generally as a sufficient condition for stability under aggregation, e.g. in RSO-processes with an infinite domain set D. But the following subsections will present some results, which can be used as tools to derive stability under aggregation for complex properties from simpler ones.

3.2 A Stability Structure for the Study of Strict Stochastic Order

In Section 2.4, we studied the stability under convex linear combinations of various subsets of a vocabulary set V consisting of maps $v{:}C\times W\to\mathbb{R}$, where $C := \{a, b\}$ is a set of two experimental conditions, and $W \subseteq \mathbb{R}$ is the set of possible values of a the dependent variable Y in an experiment. To make a map $v{:}C\times W\to\mathbb{R}$ an element of the vocabulary set V, the partial maps v(c, .) have to be CDFs with domain W. To save the discussion of some mathematical subtleties, we add the assumption that W is a closed interval [w', w''] of real numbers with w' < w''. In this situation, the requirement that the partial maps v(c, .) must be CDFs with domain W can be restated by the inequality

$$0 \le v(c, w) \le v(c, w^*) \le v(c, w'') = 1 \tag{3.5}$$

and the equation

$$v(c, w) = \lim_{n\to\infty} v(c, w + (w''-w)/n), \tag{3.6}$$

which must hold for $c \in C$ and all elements w and w* of W with $w \le w^*$.[79]

The example will differ from the SSA of Section 2.4 by weaker premissas about the process of aggregation. We will first introduce these premissas more or less intuitively, and then we will formalise them in the frameworks of SSAs and of probability spaces. Again, the aggregation can be understood as an RSO-process, but with a potentially infinite domain set D - e.g., a set of persons in situations. As in in Section 2.4, it is assumed that every unit is charachterised by an element of the vocabulary set V, and the element of this vocabulary set, which characterises unit u, will now be denoted as $\phi(u)$ or - more conveniently - as ϕ_u. Its intuitive interpretation is the same as for the maps $v_i{:}C\times W\to\mathbb{R}$ in in Section 2.4: Each function value $\phi_u(c, w)$ with $(c, w) \in C\times W$ represents the probability of the event $Y \le w$ under the assumption that unit u is tested under condition c. More

---

[78] More generally, π may be any selection distribution with a non-zero selcetion probability $\pi(u_i)$ for an infinite number of units u_i. Given a selection distribution π with this property, the containment (S_f, π) follows again from the definition of the relation H. Now let q' be an arbitrary natural number, and i' a natural number with $q' \le i'$ and $\pi(u_{i'}) > 0$. Then the sum on the right hand side of Equation (3.4) contains the probability $P_\pi(u_{i'}, q') = 2^{-i'}/i' > 0$, leading to $\Phi_\pi(q') > 0$. So Φ_π isn't contained in S_f.

[79] See Footnote 51.

formally, $\phi(u)$ resp. ϕ_u is the function value which is assigned to the element u of D by a map ϕ:D→V.

Whereas in Section 2.4 the selection probabilities were identified with the coefficients λ_i of a convex linear combination of maps v_i:C×W→$\mathbb{R}$, it is now assumed that the random selection is governed by a probability measure π on a σ-algebra A_D in the set D. For technical reasons, it is assumed that this σ-algebra is fine enough to contain all sets A of the form

$$A = \{u \in D: \phi_u(c, w) \leq \lambda\} \tag{3.7}$$

with (c, w) ∈ C×W and $\lambda \in [0, 1]$.[80] The probability measure π on A_D governing the selection of a unit will subsequently be called the selection distribution underlying an RSO-process, and the denotation Π for the set of all probability measures on A_D anticipates its role in an SSA under construction.

To prepare an aggregation rule, we resume the denotation U for a D-valued random variable indicating the selected unit. More technically, U is a random variable on A_D, whose distribution is the selection distribution π underlying the RSO-process. Then each function value $\phi_U(c, w)$ for that unit with arbitrary (c, w) ∈ C×W can be considered as a real valued random variable. The central assumption of the aggregation rule can be summarised in the following claim: The probability of the event $Y \leq w$ in the process of selecting a unit with selection distribution π and testing it under condition c is identical with the expectation of the random variable $\phi_U(c, w)$ under the assumption that U is a random variable with distribution π. These expectations are collected in the map Φ_π:C×W→$\mathbb{R}$, which is assigned to the selection distribution π by the aggregation rule Φ.

Before we explicate this assumption in an equation, we should realise that a term like $\mathsf{E}\phi_U(c, w)$ would be a sufficient specification of an expectation only in situations, where a unique probability distribution of the random variable U is introduced (directly or indirectly) in the context. But a specification for the map Φ:Π→V by a formula for $\Phi_\pi(c, w)$ has to express the dependence of this expectation upon the distribution of π of the random variable U. To account for this situation, we introduce a notational convention: We will write $\mathsf{E}_{U\sim\pi}\, \phi_U(c, w)$ for 'expectation of $\phi_U(c, w)$ under the assumption that U is a random variable with distribution π'. The following more general definition will be helpful in later sections. To support such applications, the definition refers to a the set where $\mathbb{R}^*$ of 'extended real numbers' (i.e., $\mathbb{R}^* := \mathbb{R} \cup \{-\infty, +\infty\}$), and to the σ-algebra B^* of Borel sets in $\mathbb{R}^*$, which is the system of all subsets A of $\mathbb{R}^*$ with $(A \cap \mathbb{R}) \in \mathsf{B}$.

Definition 3.2: For a random variable X on a measurable space (Ω', A'), a probability measure P on A' and an A'-B^*-measurable map g:Ω'→$\mathbb{R}^*$, the term $\mathsf{E}_{X\sim P}\, g(X)$ is defined by the equation

$$\mathsf{E}_{X\sim P}\, g(X) := \int g(X)\, dP, \tag{3.8}$$

where $\int$ denotes the Lebesgue integral. More generally, the probability measure P may be replaced by another specification of a probability measure on A'.

[80] In other words, it is assumed that the map u→$\phi_u(c, w)$ is measurable for every (c, w) ∈ C×W. See Proposition 9.2 in Bauer (1972, p. 58) for the equivalence of this property and the assumption that A_D contains all sets A with a suitable description by Equation (3.7), and note that the equalities A = ∅ for $\lambda < 0$ and A = D for $\lambda > 1$ follow from the definition of the set V.

An example for the more general case is $\mathsf{E}_{Y\sim v(a,\,.\,)}\, g(Y)$, where v is an element of the currently considered vocabulary set V. Since the definition of this set implies that the partial map v(a, .) is a CDF with domain W, this partial map specifies a probability measure on a suitable σ-algebra. So the term $\mathsf{E}_{Y\sim v(a,\,.\,)}\, g(Y)$ can be verbalised as 'expectation of g(Y) under the assumption that Y is a random variable, whose probability distribution is specified by the CDF v(a, .).

With the notational convention of Definition 3.2, a map Φ_π:C×W→ℝ can be specified by the equation

$$\Phi_\pi(c, w) = \mathsf{E}_{U\sim\pi}\, \phi_U(c, w) \tag{3.9}$$

for every $\pi \in \Pi$ and (c, w) ∈ C×W. The membership of this map Φ_π in the vocabulary set V, which is necessary to fulfill the requirements of Definition 1.2, follows from elementary properties of integrals.[81]

It is left to introduce a suitable relation H ⊆ PW×Π for our SSA. As in previous examples, we start with a verbal formulation of the implication, which must hold (according to SSA-Axiom (iv)) for every selection distribution π to make a subset S of the vocabulary set stable under the considered kind of aggregation: If U is a random variable with distribtuion π representing the selected unit, and if the map ϕ_U:C×W→ℝ characterising this unit is almost surely contained in S, then the map Φ_π:C×W→ℝ specified by Equation (3.9) will also be an element of S. Again, the if-clause in this implication will be used to define the relation H. More explicitly, this clause assumes the existence of a subset A of the domain set D with the properties $A \in \mathsf{A}_D$, π(A) = 1, and φ(A) ⊆ S (where φ(A) denotes the set of all elements v of V with v = φ(u) for some u ∈ A).

The demonstrations of general results to be presented in the following subsections will mount up

---

[81] More explicitly, it has to be proved that the map Φ_π given by Equation (3.9) has the properties specified by Inequality (3.5) and Equation (3.6). So let π be a given element of Π, and we will write E instead of $\mathsf{E}_{U\sim\pi}$. Since ϕ_u is an element of V for every u ∈ D, Inequality (3.5) leads to

$$0 \leq_{a.s.} \phi_U(c, w) \leq_{a.s.} \phi_U(c, w^*) \leq_{a.s.} \phi_U(c, w'') =_{a.s.} 1,$$

for w' ≤ w ≤ w* ≤ w'', the subscript a.s. standing for 'almost surely'. But this implies

$$0 \leq \mathsf{E}\, \phi_U(c, w) \leq \mathsf{E}\, \phi_U(c, w^*) \leq \mathsf{E}\, \phi_U(c, w'') = 1,$$

and Equation (3.9) yields

$$0 \leq \Phi_\pi(c, w) \leq \Phi_\pi(c, w^*) \leq \Phi_\pi(c, w'') = 1.$$

For Equation (3.6), let c and w be arbitrary elements of C and W, and define maps g_n:V→ℝ for n = 1..∞ by

$$g_n(v) := v(c, w) + 1 - v(c, w+(w''-w)/n).$$

Then Inequality (3.5) implies

$$0 \leq g_n(v) \leq g_{n+1}(v) \leq 1,$$

and the monotone convergence theorem (see Bauer, 1992, p. 68, Proposition 11.4) yields

$$\mathsf{E}\lim_{n\to\infty} g_n(\phi_U) = \lim_{n\to\infty} \mathsf{E} g_n(\phi_U).$$

Now Equation (3.6) grants $\lim_{n\to\infty} g_n(\phi_U) =_{a.s.} 1$, whereas elementary transformations on the right hand side of the above lim-equation lead to

$$1 = \lim_{n\to\infty} \mathsf{E}(\phi_U(c, w) + 1 - \phi_U(c, w+(w''-w)/n)) = \Phi_\pi(c,w) + 1 - \lim_{n\to\infty} \Phi_\pi(c, w+(w''-w)/n),$$

i.e.,

$$\Phi_\pi(c, w) = \lim_{n\to\infty} \Phi_\pi(c, w+(w''-w)/n).$$

to a proof of the claim that the set S_S of maps $v:C\times W\rightarrow\mathbb{R}$ with strict stochastic order[82] is stable under the assumed kind of aggregation. The view upon these results as a tool-box for the study of aggregation stability will be supported by their complementary contributions to proofs of stability under aggregation for more and more complex subsets of the vocabulary set. These subsets are specified at the end of the following summary of the assumed SSA.

Example 3.3: Most demonstrations in the following subsections will refer to an SSA (V, Π, Φ, H, T), which is based on a set $C := \{a, b\}$ of experimental conditions, a set W, which is an interval $[w', w'']$ of real numbers with $w' < w''$, and a set D of units (e.g. persons or persons in situations). The following assumptions are made:

(i) The vocabulary set V is the set of those maps $v:C\times W\rightarrow\mathbb{R}$ where the partial maps $v(a, .)$ and $v(b, \ )$ are CDFs with domain W.

(ii) There is a map $\phi:D\rightarrow V$ and a σ-algebra A_D in D, which contains all subsets A of D fulfilling Equation (3.7) for some $(c, w) \in C\times W$ and some $\lambda \in [0, 1]$.[83]

(iii) Π is the set of all probability measures on A_D.

(iv) The aggregation rule Φ is specified by Equation (3.9) for every $\pi \in \Pi$ and $(c, w) \in C\times W$.

(v) An ordered pair (S, π) with $S \subseteq V$ and $\pi \in \Pi$ is an element of the relation H iff the σ-algebra A_D contains a subset A of D with $\pi(A) = 1$ and $\phi(A) \subseteq S$.

Stability under aggregation (i.e., membership in the set system T) will be studied for the following subsets of the vocabulary set V, where w may be any element of W:

$$S'_w := \{v \in V: v(a, w) \geq v(b, w)\}, \tag{3.10}$$

$$S''_w := \{v \in V: v(a, w) > v(b, w)\}, \tag{3.11}$$

$$S^*_w := \{v \in V: v(a, w) = v(b, w)\}, \tag{3.12}$$

$$S_O := \bigcap_{w\in W} S'_w, \tag{3.13}$$

$$S_L := \bigcap_{w\in W} S^*_w, \tag{3.14}$$

$$S_{st} := \bigcup_{w\in W} S''_w, \tag{3.15}$$

and

$$S_S := S_O \cap S_{st}. \tag{3.16}$$

For referential convenience, the SSA specified in Example 3.3 will subsequently be called the CDF-SSA.

Under a systematic view, it may seem surprising that random variables like U, Y and $\phi_U(c, w)$ as well as their distributions have been introduced above without reference to a basic probability space governing the entire RSO-process. This approach is motivated by a diversity of conceivable formal models for such processes with different probability spaces and ways of introducing the random variables. In a situation of this kind, it is not only legitimate, but even advisable to formulate a common nucleus, which is based only upon the probability distributions and the interrelationships of

---

[82] See Equations (2.64) and (2.65) for a definition of the sets S_O and S_S

[83] Section 3.9 will point out advantages of working with the coarsest σ-algebra A_D in D fulfilling the requirements of Assertion (ii). But this property isn't part of the definition, since problems resulting from finer σ-algebras can be discussed only if they are not excluded from the outset.

the involved random variables, but not upon their embedding into particular probability spaces.[84] A nucleus of this kind is formulated in the above summary of Example 3.3, and results derived from Assertions (i) through (v) will apply to all models of RSO-processes complying with these premissas.

Understood in this way, the assumptions have the formal status of axioms. This view makes it advisable to demonstrate their meaning in a prototypical class of probability spaces; but this demonstration (which makes up the rest of the present Subsection 3.2) can be skipped without momentary loss of continuity. The subseequently described probability spaces model RSO-processes by probability measures on a measurable space (Ω, A) with $\Omega := D{\times}W$. In other words, the elementary outcomes - i.e., the elements of the set Ω - are ordered pairs (u, w) representing the selected unit and the obtained value in the dependent variable. The random variables U and Y are based upon maps $U{:}\Omega{\to}D$ and $Y{:}\Omega{\to}W$ with $U(u, w) := u$ and $Y(u, w) := w$, and the measurability of these maps with respect to suitable σ-algebras A_D and A_W in D resp. W will be assumed. In particular, A_W is the σ-algebra of those Borel sets in $\mathbb{R}$ which are subsets of W.

Since the probabilities of events of interest will depend upon the selection distribution as well as upon the experimental condition applied to the selected unit, let $P_{\pi c}$ be the probability measure governing the process of selecting a unit with selection distribution π (a probability measure on A_D) and testing it under the fixed condition $c \in C$, one such probability measure $P_{\pi c}$ being assumed for every $\pi c \in \Pi{\times}C$. More explicitly, this means: For every probability distribution π on A_D and each element c of the set C, there is a unique probability measure $P_{\pi c}$ on A with $P_{\pi c}(A{\times}W) = \pi(A)$ for every element A of the σ-algebra A_D. A consistency assumption referring to these probability measures $P_{\pi c}$ will be introduced immediately; but for the moment, the assumption of their uniqueness suffices to define the map $\Phi{:}\Pi{\to}V$ by the equation

$$\Phi_\pi(c, w) := P_{\pi c}(Y \le w) = P_{\pi c}(D{\times}[w', w]) \tag{3.17}$$

for every $\pi \in \Pi$ and $(c, w) \in C{\times}W$, where the term $P_{\pi c}(Y \le w)$ refers to the probability of the event $Y \le w$ in the probability space $(\Omega, \mathsf{A}, P_{\pi c})$.

There are several approaches to an interpretation of the map $\phi{:}D{\to}V$, which will be explicated in Section ###. For an understanding of applications in the following subsections, it suffices again to describe a common nucleus of these interpretation. This nucleus becomes easy to follow up under the assumption that the σ-algebra A_D contains each set {u} consisting of just one element u of the domain set D. Under this assumption, the testing of a fixed unit u can be conceived as the limiting case of an RSO-process with a selection distribution, where unit u is selected with probability 1.[85] Then we can specify the map $\phi{:}D{\to}V$ by the following statement about the element ϕ_u of the set V (i.e., the map $\phi_u{:}C{\times}W{\to}\mathbb{R}$), which is assigned to unit u: For every $(c, w) \in C{\times}W$, the function value $\phi_u(c, w)$ is the probability of the event $Y \le w$ under a probability measure $P_{\pi c}$ on A, where unit u is selected with probability 1.

It has to be noted that the aggregation rule (Assertion (iv) in the summary of the example)

---

[84] See e.g. Bauer (1991, p. 15) for this practice.

[85] The assumption $\{u\} \in \mathsf{A}_D$ can be dispensed with, if we consider an RSO-process, whose selection distribution is the 'Dirac measure in u'. This probability measure is characterised by a probability of 1 for every $A \in \mathsf{A}_D$ with $u \in A$.

follows from this interpretation only under additional assumptions about the internal consistency of the probability measures $P_{\pi c}$. These assumptions mount up the claim that the behavioural dispositions of a unit do not depend upon the probability of its selection. (Recall that a similar assumption about random generators was made in Example 3.1.) But for the moment, a discussion of this issue can be postponed until Section ###. In particular, that section will also present a more general approach, where such consistency assumptions referring to units are unnecessary.

3.3 Redefining a Subset of the Vocabulary Set

In Section 2.2.5, we have reformulated a result of Estes (1956) concerning parametric families of maps $Q \to \mathbb{R}$ with parameter space Θ and parametrisation map $\psi:\Theta \to \mathbb{R}^Q$: If the parametrisation map ψ can be decomposed into a parameter transformation map $t:\Theta \to \mathbb{R}^m$ and an affine map $\psi^{\sim}:\mathbb{R}^m \to \mathbb{R}^Q$ such that $t(\Theta)$ is a convex subset of $\mathbb{R}^m$ and $\psi = \psi^{\sim} \circ t$, then the parametric family $\psi(\Theta)$ is stable under convex linear combinations, since it is identical with $\psi^{\sim}(t(\Theta))$, and the image of a convex set under an affine map is convex. In other words, the set $\psi(\Theta)$ was reconceived as $\psi^{\sim}(t(\Theta))$ to make available a result referring to a special class of parametrisation maps.

Frequently, such redefinitions of a set S, whose stability under some kind of aggregation is studied, will be helpful to make available results, which will be reported in the following subsections. In fact, the original definition of the set S_S by Equation (2.65) has already been reconceived in Equation (3.16) to make S_S the top of a hierarchy of simpler sets, which can be analysed using the subsequently presented tools. But whereas this redefinition was based on simple set theoretical operations, we will now demonstrate that a redefinition can also be based upon deeper mathematical properties of a vocabulary set.

For the CDF-SSA of Example 3.3, let W* be set set of all rational numbers which are contained in the set W. Under this notation, it is proved in Section 6.15 that the definitions of the sets $S_{O,}$ S_L and S_{st} in Equations (3.13), (3.14) and (3.15) can be rewritten as

$$S_O := \bigcap_{w^* \in W^*} S'_{w^*} \tag{3.18}$$

$$S_L := \bigcap_{w^* \in W^*} S^*_{w^*} \tag{3.19}$$

and

$$S_{st} := \bigcup_{w^* \in W^*} S''_{w^*}. \tag{3.20}$$

Although the gain of such redifinitions will be earned not before Equations (3.19) will be used in Subection 3.8, it can be anticipated that countable intersections and unions of sets will be easier to handle in the framework of probability theory than uncountable ones. (Certainly, the index set W in Equations (3.13), (3.14) and (3.15) is uncountable, since it is an interval [w', w''] with w' < w'', whereas W* is countable, since the set of all rational numbers is countable.)

To prepare a conclusion, which can be generalised beyond the example, it should be noted that the equivalence of the above redefinitions and the original definitions is proved in Subection 6.15 with substantial recourse upon mathematical properties of CDFs. It would be possible to analyse the aggregation stability of our set S_S in an SSA, whose vocabulary set consists of all maps $C \times W \to \mathbb{R}$.

Indeed, units and aggregates are characterised by maps of this kind; but the additional assumption that all partial maps v(c, .) are CDFs with domain W allows conclusions, which will turn out useful in later sections. Undoubtedly, it would be dubious to add further assumptions, which may be empirically false in a concrete situation. But the intended interpretation of the function values v(c, w) as probabilies of the event $Y \leq w$ under condition c implies all properties of CDFs for the partial maps v(c, .).

More generally, fruitful redifinitions of sets of interest will typically be supported by a definition of the the vocabulary of an SSA, which incorporates intended interpretation of elements of the set. In other word, the set V should be conceived as a vocabulary set in a semantic understanding of this term and not only syntactically. On the other side, a definition of a vocabulary set, which incorporates such properties, has its price beyond the well known loss of generality following from strong assumptions: It may need some effort to prove that a plausible approach to an aggregation rule like Assertion (iv) in Example 3.3 will always result in an element Φ_π of the vocabulary set.[86]

3.4 Restrictions of Basic Sets

At the end of the preceeding subsection 3.3, we pleaded for definitions of vocabulary sets enabling useful deductions. In other situations, it will be fruitful to replace the set Π of aggregates in an SSA by a smaller one. Such restrictions of the two basic sets of SSAs raise the question for implicit changes of other components of an SSA. The following lemma describes these changes for the most general case where both basic sets are replaced by a subset. Restrictions for only one of them can be subsumed as special cases, where the other set is replaced by itself.

Lemma 3.4: For an SSA (V, Π, Φ, H, T), let V' be a subset of V and Π' a non-empty subset of Π such that the properties $\Phi_\pi \in V'$ and $(V', \pi) \in H$ hold for every $\pi \in \Pi'$. Furthermore, let a map $\Phi':\Pi' \to V'$ be the restriction to Π' of the map $\Phi:\Pi \to V$, and define a relation $H' \subseteq PV' \times \Pi'$ as

$$H' := H \cap (PV' \times \Pi') = \{(S', \pi') \in PV' \times \Pi': (S', \pi') \in H\}. \tag{3.21}$$

Then there exists a unique set system $T' \subseteq PV'$ such that the ordered quintuple $(V', \Pi', \Phi', H', T')$ is an SSA. The relationship between the set systems T and T' is described by the following equations:

$$T \backslash T' = \{S \in T: S \not\subseteq V'\} \tag{3.22}$$

$$T' \backslash T = \{S \in (PV' \backslash T): \forall \pi \in \Pi: (((S, \pi) \in H \wedge \Phi_\pi \notin S) \Rightarrow \pi \in \Pi \backslash \Pi'\} \tag{3.23}$$

Similarly, the relationship between the set systems T_e and T'_e is described by the equations

$$T_e \backslash T'_e = \{S \in T_e: S \not\subseteq V'\} \tag{3.24}$$

$$T'_e \backslash T_e = \{S \in (PV' \backslash T_e): \forall \pi \in \Pi: (\neg((S, \pi) \in H \Leftrightarrow \Phi_\pi \in S) \Rightarrow \pi \in \Pi \backslash \Pi'\} \tag{3.25}$$

With the definition of set systems $T^* \subseteq PV'$ and $T^*_e \subseteq PV'$ as

$$T^* := T \cap PV' = \{S \in T: S \subseteq V'\}, \tag{3.26}$$

and

$$T^*_e := T_e \cap PV' = \{S \in T_e: S \subseteq V'\}, \tag{3.27}$$

the following properties hold:

(i) $T^* \subseteq T'$ and $T^*_e \subseteq T'_e$.

---

[86] See Footnote 81 for the CDF-SSA of Example 3.3.

(ii) If $V' = V$, then $T \subseteq T'$ and $T_e \subseteq T'_e$.
(iii) If $\Pi' = \Pi$, then $T' = T^*$ and $T'_e = T^*_e$.

Lemma 3.4 will subsequently be called the *Restriction-Lemma*. For referential convenience, the following definition introduces denotations for derived SSAs:

Definition 3.5: In the situation of Restriction-Lemma 3.4, the SSA $(V', \Pi', \Phi', H', T')$ is an *SSA-restriction* or - more specifically - the *V'-Π'-restriction* of the SSA (V, Π, Φ, H, T). In situations with $V' = V$ or $\Pi' = \Pi$, the identical set may be omitted in the denotation. So a *V'-restriction* (with $V' \subseteq V$) is a V'-Π'-restriction with $\Pi' = \Pi$, and a *Π'-restriction* (with $\Pi' \subseteq \Pi$) is a V'-Π'-restriction with $V' = V$.[87]

Instead of a formal proof of the Restriction-Lemma, we will now motivate its premissas and definitions, and then the consequences will be self-explaining. The general idea is the same as in the restriction of a map $f:X \rightarrow Y$ to a subset X' of X: To obtain a map $f':X' \rightarrow Y$, all assignments of function values to elements of X, which are not contained in X' must be dropped; but for elements of the set X', the assignment of function values is left unchanged. In the same way, the premissas and the changes of components of an SSA in Restriction-Lemma 3.4 are confined to those resulting from the formal properties of SSAs specified by Definition 1.2 under the demand to obtain an SSA, whose first two components are V' and Π'.

The assumptions $\Phi_\pi \in V'$ and $(V', \pi) \in H$ for all elements π of the set Π' are motivated by properties of the remaining components. To work with an aggregation rule $\Phi':\Pi' \rightarrow V'$, which is the restriction to Π' of the original aggregation rule $\Phi:\Pi \rightarrow V$, implies that the assignment of an element Φ_π of the vocabulary set to an aggregate π is dropped from the aggregation rule for elements of the set Π which are cancelled under the restriction, but left unchanged for elements of the set Π'. So the property $\Phi_\pi \in V'$ must hold for every $\pi \in \Pi'$, since the function values of an aggregation rule must be elements of the vocabulary set of an SSA. Similarly, the new relation H' must be a subset of $\mathsf{P}V' \times \Pi'$ which implies that it cannot contain ordered pairs (S, π) where S is not a subset of V' or π is not an element of Π'. So these pairs are dropped from the relation H by the definition of H' in Equation (3.21). The same equation conserves the membership or non-membership in the relation H for all ordered pairs (S, π) with $S \subseteq V'$ and $\pi \in \Pi'$. Note that that SSA-Axioms (i) and (iii) follow for the new relation H', if they hold for the original relation H. But to grant the validity of SSA-Axiom (ii), the property $(V', \pi) \in H$ for every $\pi \in \Pi'$ must be introduced as a premissa.

If the first four components of the new SSA $(V', \Pi', \Phi', H', T')$ are given, then the set systems T' and T'_e are completely specified by SSA-Axiom (iv) and Equation (1.1). Typically, these set systems will differ from those of the original SSA. These changes, which are the most important ones in a theory of stability under aggregation, are described as set differences in Equations (3.22) through (3.25). Since only subsets of the new vocabulary set V' can be elements of the set systems T' and T'_e, all elements of the old set systems T and T_e which are not subsets of V' are removed from the set systems (Equations (3.22) and (3.24)). Conversely, subsets of the new vocabulary set V' which were

[87] To avoy confusions by the almost identical denotations for restrictions of only one set, it must be clear from the local context whether the restricted set appearing in the denotation is a subset of V or of Π.

not members of T resp. of T_e will become members of T' resp. T'_e, if and only if all elements π of the set Π leading to a violation of the implication $(S, \pi) \in H \Rightarrow \Phi_\pi \in S$ (resp. the equivalence $(S, \pi) \in H \Leftrightarrow \Phi_\pi \in S$) are removed by the restriction of the set Π, and this means that every such π is contained in the set difference $\Pi \backslash \Pi'$ (Equations (3.23) and (3.25)).[88]

For another description of the changes in the set systems, Restriction-Lemma 3.4 introduces the set systems T* and T^*_e. They can be seen as the status of the set systems after the removal of those members of T resp. T_e which are not subsets of V', but without the addition of potential new members. So they are subsystems of T' resp. T'_e, as claimed by Assertion (i) of the lemma. Furthermore, this interpretation imples that T* and T^*_e are identical with T resp. T_e in situations where the vocabulary set is left unchanged (premissa of Assertion (ii), i.e., in Π'-restrictions), and then the inclusions claimed by Assertion (i) can be rewritten as $T \subseteq T'$ resp. $T_e \subseteq T'_e$. Conversely, if the set Π of aggregates is left unchanged (i.e., in V'-restrictions), then the cancellation of non-subsets of the new vocabulary set V' is the only alteration of the set systems T and T_e, and this means that T' and T'_e are identical with T* resp. T^*_e (Assertion (iii)).

The following corollary is follows immediately from these considerations:

Corollary 3.6: In the situation of Restriction-Lemma 3.4, the V'-restriction of the Π'-restriction of the considered SSA is identical with its V'-Π'-restriction. Furthermore, if the properties $\Phi_\pi \in V'$ and $(V', \pi) \in H$ hold for every $\pi \in \Pi$, then the Π'-restriction of the V'-restriction of the SSA is identical with its V'-Π'-restriction.

Whereas Restriction-Lemma 3.4 describes simultaneous restrictions of the basic sets V and Π, the subsequent demonstrations will be confined to restrictions of only one of these sets, which are subsumed as special cases by the lemma. For a V'-restriction, assume that the analysis of pairs of CDFs in Example 3.3 started with a vocabulary set V consisting of all maps of the set $C \times W$ into the interval [0, 1], and that the vocabulary set described in Example 3.3.(i) is introduced as a new vocabulary set V' in a V'-restriction. The (identical) sets Π and Π' are again the set of all probability measures on the σ-algebra A_D. The aggregation rule Φ and the relation H are defined as in Example 3.3, the larger vocabulary set V being referenced in these definitions. Then the assumptions $(V', \pi) \in H$ and $\Phi_\pi \in V'$ for every $\pi \in \Pi$, which are required by the Restriction-Lemma for such situations, are legitimated by the interpretation of the situation.[89] The claim of Lemma 3.4.(iii)

---

[88] For a complete proof of Equations (3.22) through (3.25), it has to be added that the membership in the set systems doesn't change for sets which don't appear on the right hand side of the respective equations. For a subset S of V', membership in T resp. in T_e means that the critical implication resp. equivalence holds for every $\pi \in \Pi$, and then it will also hold for every $\pi \in \Pi'$, since Π' is assumed to be a subset of Π. Conversely, if an element π of the set Π leading to a violation of the critical implication or equivalence is not contained in the set $\Pi \backslash \Pi'$ and hence not removed under the restriction, then the respective set S will not be contained in T' resp. in T'_e.

[89] More explicitly, it is assumed that the map $\phi_u: C \times W \rightarrow [0, 1]$ characterising individual u is contained in the new vocabulary set V' for every $u \in D$. So we have $\phi(D) \subseteq V'$, and the set D can (continued...)

referring to such V'-restrictions can be verbalised as follows: Subsets of the larger vocabulary set V which are not subsets of V' are cancelled from the set systems T and T_e; but for subsets of V', their membershisp or non-membership in the set systems T and T_e is left unchanged.

An important application of Π'-restrictions will occur in Theorem 3.18 in Subsection 3.7; but the gist of the claims of Restriction-Lemma 3.4 referring to Π'-restrictions can be seen from the following facts, which have already been showed at the end of Section 2.4 for aggregation by convex linear combination, whereas they will be reestablished later for the CDF-SSA of Example 3.3. The set S_{st} of all elements v of the vocabulary set V where the inequality $v(a, w) > v(b, w)$ holds for some $w \in W$ isn't generally stable under aggregation, since there may be elements π of the set Π with $(S_{st}, \pi) \in H$ and $\Phi_\pi \notin S_{st}$. But for every such π, the ordered pair (S_O, π) is not contained in the set S_O of all $v \in V$ with $v(a, w) \geq v(b, w)$ for every $w \in W$, and these elements of the set Π are removed in a $\Pi^\sim$-restriction with $\Pi^\sim := \{\pi \in \Pi\colon (S_O, \pi) \in H\}$. So the set S_{st} becomes stable under aggregation after this restriction.

The inclusions $T \subseteq T'$ and $T_e \subseteq T'_e$, which hold for Π'-restrictions, have another consequence for applications: If a subset S of the vocabulary set is contained in T', but not in T, then the hypothesis $\phi(D) \subseteq S$ implies the prediction $\Phi_\pi \in S$ for RSO-processes with selection distribution π as long as π is contained in Π', but not for every $\pi \in \Pi$. If the set Π' is such that we can take care to realise a selection distribution contained this set, then the hypothesis $\phi(D) \subseteq S$ implies the prediction $\Phi_\pi \in S$ for all RSO-processes whose selection distribution is an element of the set Π'

3.5 Maps between Vocabulary Sets

In some examples, which have been analysed in Section 2, it turned out that an aggregation rule implied the applicability of a similar kind of aggregation to mathematical objects, which were not mentioned explicitly in the aggregation rule. E.g., Equation (2.10) reformulated a result of Estes (1956) for certain parametric families of functions: Convex linear combinations of members of the family yield an element of the family, whose parameter can be obtained from an application of the same linear combination to the parameters of the maps entering the combination. Gerneralisations to other types of parametric families of functions are formalised by Lemma 2.11 and Equation (2.53). Equations (2.62) and (2.63) summarise similar results for the expectations of random variables, whose probability distributions are represented by CDFs entering a convex linear combination. In all these examples, the vocabulary set of an SSA was mapped into a set, which could be considered as the vocabulary set of another SSA, and this approach could be used for the study of aggregation stability of sets, which were defined in terms of the parameters resp. the expectations.

The following subsection will first generalise the underlying pattern of proof in a theorem, which is designed for the study of aggregation stability of a specific set of interest. Some applications to the CDF-SSA of Example 3.3 will motivate the extension of the theorem to corollaries, which can be used to prove simultaneously the aggregation stability for large classes of subsets of a vocabulary set. In particular, some maps between the vocabulary sets of SSAs will turn out to have useful properties of homomorphisms and isomorphism.

Since the above outline of a general situation has been rather informal, a theorem referring to

---

[89] (...continued)
takte the role of the set A in Example 3.3.(v). For $\Phi_\pi \in V'$, see Footnote 81.

such situations is subsequently prepared by another example. In the CDF-SSA of Example 3.3, let w be an arbitrary element of the set W, and consider the set S'_w consisting of all elements v of the vocabulary set V with $v(a, w) \geq v(b, w)$. This set can also be described by a map $g_w: V \to \mathbb{R}$, which is given by the equation

$$g_w(v) := v(a, w) - v(b, w) \tag{3.28}$$

for every $v \in V$. Then the set S'_w can also be described by the equation[90]

$$S'_w = \{v \in V: g_w(v) \in [0, 1]\}. \tag{3.29}$$

Now consider the set $g_w(S'_w)$, i.e., the set of all real numbers ξ, where the equation $g_w(v) = \xi$ holds for some element v of V. If w is identical with w'' (the upper boundary of the interval W = [w', w'']), then Inequality (3.5) implies $v(a, w) = v(b, w) = 1$ for every $v \in W$; i.e., the set $g_w(S'_w)$ contains only the number 0. For $w < w''$, it is easily verified that the set $g_w(S'_w)$ is identical with the interval [0, 1].[91] In both cases, the equation

$$S'_w = g_w^{-1}(g_w(S'_w)) \tag{3.30}$$

follows immediately from the definitions of the set S'_w and the map g_w,[92] and this equation will be an important premissa in later generalisations.

Another property of the map g_w follows from the aggregation rule (Assertion (iv) of Example 3.3):

$$\begin{aligned} g_w(\Phi_\pi) &= \Phi_\pi(a, w) - \Phi_\pi(b, w) = \mathsf{E}_{U\sim\pi}\, \phi_U(a, w) - \mathsf{E}\phi_U(b, w) = \mathsf{E}_{U\sim\pi}\, (\phi_U(a, w) - \phi_U(b, w)) \\ &= \mathsf{E}_{U\sim\pi}\, g_w(\phi_U). \end{aligned} \tag{3.31}$$

Now let π be an element of the set Π (i.e., a distribution of the random variable U) such that $(S'_w, \pi) \in H$. Then we obtain from Equation (3.29) that the random variable $g_w(\phi_U)$ is almost surely

---

[90] Note that the difference $v(a, w) - v(b, w)$ cannot be greater that 1, since $v(a, w)$ and $v(b, w)$ are probabilities.

[91] For $w < w''$ and every $\xi \in [0, 1]$, an element v of the set S'_w with $g_w(v) = \xi$ is given by $v(a, w^*) = \xi$ and $v(b, w^*) = 0$ for $w' \leq w^* < w''$, and $v(a, w'') = v(b, w'') = 1$. Certainly, the description of elements of V by Inequality (3.5) and Equation (3.6) holds for that v, and also the inequality $v(a, w) \geq v(b, w)$, which is required for membership in S'_w.

[92] Recall (from the general notational convention in Footnote 56) that $g_w^{-1}(g_w(S'_w))$ is the set of all elements v of the V with $g_w(v) \in g_w(S'_w)$, and that the equality $S = f^{-1}(f(S))$ is not tautological for a non-injective map $f: X \to Y$ and $S \subseteq X$. But in our situation, Equation (3.30) follows from the definitions of S'_w and g_w.

contained in the interval [0, 1],[93] and this implies that the expectation of the random variable must also be contained in the interval. Since Equation (3.31) tells that this expectation is equal to $g_w(\Phi_\pi)$, the property $\Phi_\pi \in S'_w$ follows from another reference to Equation (3.29). But if this property holds for every $\pi \in \Pi$ with $(S'_w, \pi) \in H$, then the set S'_w is stable under aggregation, i.e., contained in the set system T. (See SSA-Axiom (iv) for this conclusion.)

To prepare a generalisation of the underlying pattern of proof, observe that it is based on a well known fact, which can also be conceived as an instance of stability under aggregation: If a real valued random variable Y has a finite expectation and if the value of Y is almost surely contained in an interval S of real numbers, then the expectation of Y is also contained in S. The following lemma specifies a suitable class of SSAs, and its claim is not much more than a reformulation of the said fact in the framework of Definition 1.2:

Lemma 3.7: Let V' a (finite or inifinite) non-empty interval of real numbers, T' the set of all subsets of V', which are finite or infinite intervals of real numbers (including $]v', v'[= \emptyset$ and $[v', v'] = \{v'\}$ with arbitrary $v' \in V'$), and Π' the set of those probability measures π' on the σ-algebra B of Borel sets in $\mathbb{R}$, where random variables with distribution π' are contained in V' with probability 1 and have a finite expectation. Furthermore, let a map $\Phi':\Pi' \to V'$ be given such that $\Phi'(\pi')$ with $\pi' \in \Pi'$ is the expectation of random variables with distribution π'. Finally, let a relation $H' \subseteq \mathsf{P}V' \times \Pi'$ be defined by the following specification: An ordered pair (S', π') with $S' \subseteq V'$ and $\pi' \in \Pi'$ is contained in H' iff there exists a Borel set A in $\mathbb{R}$ such that $A \subseteq S'$ and $\pi'(A) = 1$.
Then the ordered quintuple $(V', \Pi', \Phi', H', T')$ is an SSA. Furthermore, the set system T'_e consists of the empty set, the set V' and all sets $\{v'\}$, where v' is an element of V' and a (lower or upper) boundary of V'.

Again, a denotation for the class of SSAs specified in Lemma 3.7 is introduced for referential convenience: Such SSAs will subsequently be called *Interval-SSAs*. A formal proof of the lemma is unnecessary, since the SSA-Axioms can be verified easily.

Now let $(V', \Pi', \Phi', H', T')$ be the Interval-SSA with $V' = [-1, 1]$. Then the map $g_w:V \to \mathbb{R}$ given by Equation (3.28) can also be conceived as a map $g_w:V \to V'$.[94] Furthermore, let a map $f_w:\Pi \to \Pi'$ be defined by the following specification: For a random variable U with distribution $\pi \in \Pi$, the distribution of the random variable $g_w(\phi_U)$ is $f_w(\pi)$.[95] So Equation (3.31) can be rewritten as

---

[93] According to the specification of the relation H in Assertion (v) of Example 3.3, the assumption $(S'_w, \pi) \in H$ implies the existence of a subset A of D with the properties $A \in \mathsf{A}_D$, $\pi(A) = 1$, and $\phi(A) \subseteq S'_w$. In other words, we have $\phi_u \in S'_w$ for every $u \in A$, and $g_w(\phi_u) \in [0, 1]$ follows from Equation (3.29). So the random variable $g_w(\phi_U)$ is almost surely contained in the interval [0, 1].

[94] Since v(a, w) and v(b, w) are probabilities, the difference v(a, w) − v(b, w) must be contained in the interval [−1, 1]. For w = w'', this difference is 0 for every $v \in V$; but nevertheless, the respective map $g_{w''}$ can be conceived as a (non-surjective) map $g_{w''}:V \to V'$.

[95] More formally, define a family $\{d_w\}_{w \in W}$ of maps $d_w:D \to \mathbb{R}$ by

(continued...)

$$\Phi'(f_w(\pi)) = g_w(\Phi(\pi)). \tag{3.32}$$

Since a property of this kind will be of central relevance for our subsequent considerations, it is worth a verbalisation. If U is a random variable with distribution π, then Equation (3.32) tells that it doesn't matter whether the transition to the Interval-SSA is performed before or after the aggregation: On the left hand side of the equation, the distribution π of the random variable U is first transformed into the distribution $f_w(\pi)$ of the random variable $g_w(\phi_U)$, and then the aggregation rule $\Phi':\Pi'\rightarrow V'$ of the Interval-SSA is applied to $f_w(\pi)$. Conversely, on the right hand side of Equation (3.32), the aggregation rule $\Phi:\Pi\rightarrow V$ of the CDF-SSA is applied first, and then the result is transformed into the Interval-SSA by the map g_w. Note that this interchangeability of the sequence of aggregation and transition to another SSA doesn't apply generally. It is a special property of our maps g_w and f_w, which follows from Equation (3.31). But this very interchangeability will be a central premissa in the theorem which is prepared by the example.

Another premissa has already been specified by Equation (3.30). A third property, which will also be assumed in the theorem, is an implication, which has also been used in the proof of the aggregation stability of the set S'_w: If U is a random variable with a distribution π such that ϕ_U is almost surely contained in the set S'_w (i.e., if $(S'_w, \pi) \in H$), then the random variable $g_w(\phi_U)$ is almost surely contained in $g_w(S'_w)$. Since the distribution of $g_w(\phi_U)$ is $f_w(\pi)$, a reference to the definition of the relation H' in Lemma 3.7 shows that the conclusion can also be written as $(g_w(S'_w), f_w(\pi)) \in H'$. So the entire implication is summerised in the formula

$$(S'_w, \pi) \in H \Rightarrow (g_w(S'_w), f_w(\pi)) \in H', \tag{3.33}$$

which holds for every $\pi \in \Pi$. But we have already seen that the set $g_w(S'_w)$ is the interval [0, 1], and this interval is a subset of V' = [−1, 1]. So it is contained in the set system T' of our Interval-SSA, and a reference to SSA-Axiom (iv) leads to $\Phi'(f_w(\pi)) \in g_w(S'_w)$. Finally, $\Phi_\pi \in S'_w$ follows from Equations (3.30) and (3.32).

Certainly, the original proof without an Interval-SSA was much more direct, and this observation may motivate questions whether anything is gained by the roundabout way. Indeed, the description of Interval-SSAs in Lemma 3.7 is not more than a reformulation of a well known mathematical fact. As long as the vocabulary set V of a studied SSA is mapped into the vocabulary set V' of such simple SSAs, it is easier to refer directly to the underlying mathematical fact. But as soon as properties of less trivial SSAs are available, these SSAs can take the role of (V', Π', Φ', H', T') in the above proof. Then it will be helpful to know properties of maps $g:V\rightarrow V'$ and $f:\Pi\rightarrow\Pi'$ such that the aggregation stability of a set S follows from these properties by an implication, which is based only upon the SSA-Axioms. The following theorem states some implications of this kind.

---

[95] (...continued)

$$d_w(u) := \phi_u(a, w) - \phi_u(b, w)$$

for every $w \in W$ and $u \in D$. Then every such map d_w is measurable, since it is the difference of the maps $u\rightarrow\phi_u(a, w)$ and $u\rightarrow\phi_u(b, w)$, whose measurability has been verified in Footnote 80. So the map $f_w:\Pi\rightarrow\Pi'$ for a given element w of the set W is uniquely specified by the equation

$$f_w(\pi)(A) := \pi(d_w^{-1}(A))$$

for every $\pi \in \Pi$ and every $A \in \mathsf{B}$.

Theorem 3.8: For SSAs (V, Π, Φ, H, T) and $(V', \Pi', \Phi', H', T')$, maps $g:V \rightarrow V'$ and $f:\Pi \rightarrow \Pi'$, and subsets S of V and S' of V' with

$$S = g^{-1}(S'), \tag{3.34}$$

the following properties follow from the SSA-Axioms:

(i) If the implication

$$(S, \pi) \in H \Rightarrow (S', f(\pi)) \in H' \tag{3.35}$$

holds for every $\pi \in \Pi$, and if

$$\Phi'(f(\pi)) = g(\Phi(\pi)) \tag{3.36}$$

for every $\pi \in \Pi$ with $(S, \pi) \in H$, then $S' \in T'$ implies $S \in T$.

(ii) If Equation (3.36) as well as Implication (3.35) and its reversal hold for every $\pi \in \Pi$, then $S' \in T'_e$ implies $S \in T_e$.

The following properties follow under the additional assumption that the map $f:\Pi \rightarrow \Pi'$ is surjective[96]:

(iii) If the reversal of Implication (3.35) holds for every $\pi \in \Pi$ and Equation (3.36) for every $\pi \in \Pi$ with $(S', f(\pi)) \in H'$, then $S \in T$ implies $S' \in T'$.

(iv) If Equation (3.36) as well as Implication (3.35) and its reversal hold for every $\pi \in \Pi$, then the properties $S \in T$ and $S' \in T'$ are equivalent, and the properties $S \in T_e$ and $S' \in T'_e$ are also equivalent.

The above theorem, which is proved in Section 6.16, will subsequently be called the *Mapping-Theorem*. It is mainly designed for situations where aggregation stability in the SSA (V, Π, Φ, H, T) is derived (via Assertions (i) and (ii)) from a more easily analysed SSA $(V', \Pi', \Phi', H', T')$ like the Interval-SSA. In situations of this kind, the implication $S \in T \Rightarrow S' \in T'$ in Assertion (iii) can be rewritten as $S' \notin T' \Rightarrow S \notin T$, and then it can be used to disprove the aggregation stability of a set S.

The surjectivity premissa for the map $f:\Pi \rightarrow \Pi'$ in Assertions (iii) and (iv) is a limitation for their application. An example demonstrating the role of this premissa is postponed until Subsection 3.9, since its contributions to the topics of that subsection are more important. If the map Π is non-surjective, we can define $\Pi^{\sim} := f(\Pi)$ and replace the SSA $(V', \Pi', \Phi', H', T')$ by its $\Pi^{\sim}$-restriction (in the understanding of Definition 3.5) to obtain a surjective map $f:\Pi \rightarrow \Pi^{\sim}$. With the denotations $T^{\sim}$ and $T^{\sim}_e$ for the respective set systems in the $\Pi^{\sim}$-restriction, the equivalences $S \in T \Leftrightarrow S' \in T^{\sim}$ and $S \in T_e \Leftrightarrow S' \in T^{\sim}_e$ result from Theorem 3.8.(iv). Note, however. that these equivalences can be used only if we know whether the set S' is contained in $T^{\sim}$ resp. $T^{\sim}_e$. Otherwise, the inclusions $T' \subseteq T^{\sim}$ and $T'_e \subseteq T^{\sim}_e$ which can be obtained from Restriction-Lemma 3.4.(ii), allow to derive the implications $S' \in T' \Rightarrow S \in T$ and $S' \in T'_e \Rightarrow S \in T_e$ from the equivalences; but these implications are not new, since they are already stated in Theorem 3.8.(i) and (ii). However, in some situations, it may be excluded that the set S' is contained in the set difference $T^{\sim} \backslash T'$ resp. $T^{\sim}_e \backslash T'_e$,[97] and then the equivalence $S \in T \Leftrightarrow S' \in T'$ resp. $S \in T_e \Leftrightarrow S' \in T'_e$ will hold.

[96] The assumption of a surjective map $f:\Pi \rightarrow \Pi'$ can be weakened. See the note after the proof in Section 6.16.

[97] See Equations (3.23) and (3.25) for a specification of such set differences, and note the local differences in notation.

Certainly, the implication claimed by Assertion (i) is the most important one for applications and for later sections. In the above demonstration for $S = S'_w$, Equation (3.36) could be veirified (in the form of Equation (3.32)) for every $\pi \in \Pi$. So it may be questioned, why Equation (3.36) isn't introduced once and for all as a property, which must hold for every $\pi \in \Pi$. Indeed, a reformulation formulation of this kind would make the results easier to follows up; but in later applications, it will turn out helpful that Assertion (i) requires Equation (3.36) only for elements π of the set Π with $(S, \pi) \in H$.

The results stated in Assertions (ii), (iii) and (iv) will be demonstrated by their contributions to the solutions of problems in later sections. In particular, the demonstration of results referring to the set system T_e is postponed until the treatment of equivalential aggregation stability in Subsection 3.8.

Applications of Mapping-Theorem 3.8 to the sets S''_w and S^*_w of Example 3.3 are facilitated by the fact that the same maps g_w and f_w as before can take the role of the maps g and f in the theorem. But we should take the opportunity to introduce a corollary, which can be used to study the aggregation stability of larger classes of subsets of V.

Corollary 3.9: Let two SSAs (V, Π, Φ, H, T) and $(V', \Pi', \Phi', H', T')$, a set G of maps $g:V\to V'$ and a map $f:G\times\Pi\to\Pi'$ be given.
Then T contains all sets $g^{-1}(S')$ with $g \in G$ and $S' \in T'$, where the implication

$$(g^{-1}(S'), \pi) \in H \;\Rightarrow\; (S', f(g, \pi)) \in H' \tag{3.37}$$

holds for every $\pi \in \Pi$, and

$$\Phi'(f(g, \pi)) = g(\Phi(\pi)) \tag{3.38}$$

for every $\pi \in \Pi$ with $(g^{-1}(S'), \pi) \in H$.
Furthermore, T_e contains all subsets $g^{-1}(S')$ with $g \in G$ and $S' \in T'_e$, where Equation (3.38) as well as Implication (3.37) and its reversal hold for every $\pi \in \Pi$.

The above corollary will subsequently be called the *Mapping-Corollary*. Since it is not more than a simultaneous application of Mapping-Theorem 3.8 to many maps $g:V\to V'$ and many sets $g^{-1}(S')$,[98] a separate proof is unnecessary. For an application to the CDF-SSA of Example 3.3 and to the Interval-SSA of Lemma 3.7 (with $V' := \mathbb{R}$), consider the projection maps $pr_{cw}:V\to\mathbb{R}$ with $c \in C$ and $w \in W$, which are defined by

$$pr_{cw}(v) := v(c, w), \tag{3.39}$$

and let G be the set of all linear combinations of such projection maps. In other words, G is the set of those maps $g:V\to V'$ which can be written as

$$g(v) = \textstyle\sum_{i=1..n} \lambda_i\, v(c_i, w_i) \tag{3.40}$$

with suitable real numbers λ_i and ordered pairs $(c_i, w_i) \in C\times W$. Furthermore, let a map $f:G\times\Pi\to\Pi'$ be given by the following specification: For every $g \in G$ and every random variable U with

[98] For a given $g \in G$, the partial map $f(g, .\,)$ can be the map $f:\Pi\to\Pi'$ of the theorem.

distribution $\pi \in \Pi$, the distribution of the random variable $g(\phi_U)$ is $f(g, \pi)$.[99] Then Equation (3.38) and Implication (3.37) hold for every $\pi \in \Pi$,[100] and Mapping-Corollary 3.9 leads to the following result: If g:V→V' is a map, which can be written in the form of Equation (3.40), and S' is an arbitrary interval of real numbers, then the set $g^{-1}(S')$ is contained in the set system T of the CDF-SSA (Example 3.3).

On this background, it is easily proved that the sets S'_w, S''_w and S^*_w in the CDF-SSA are elements of T. Obviously, the maps g_w:V→V' given by $g_w(v) := v(a, w) - v(b, w)$, which have already been introduced previously, are contained in the above defined set G, and the equations $S'_w = g_w^{-1}([0, 1])$, $S''_w = g_w^{-1}(]0, 1])$ and $S^*_w = g_w^{-1}([0, 0])$ show that the respective sets are contained in T.

Whereas Mapping-Corollary 3.9 supplies sufficient conditions for membership in the set systems T and T_e, the following one gives a complete description of these set systems under stronger premissas.

Isomorphism-Corollary 3.10

Corollary 3.10: Let SSAs (V, Π, Φ, H, T) and (V', Π', Φ', H', T'), a bijective map g:V→V' and a surjective map f:Π→Π' begiven such that Equation (3.36) as well as Implication (3.35) (with $S := g^{-1}(S')$) and its reversal hold for every $\pi \in \Pi$ and every subset S' of V'. Then the equivalences

$$S \in T \;\Leftrightarrow\; g(S) \in T' \tag{3.41}$$

and

$$S \in T_e \;\Leftrightarrow\; g(S) \in T'_e \tag{3.42}$$

hold for every subset S of V; i.e.,

$$T = g^{-1}(T') = \{g^{-1}(S'): S' \in T'\} = \{S \subseteq V: g(S) \in T'\}, \tag{3.43}$$

and

$$T_e = g^{-1}(T'_e) = \{g^{-1}(S'): S' \in T'_e\} = \{S \subseteq V: g(S) \in T'_e\}. \tag{3.44}$$

Again, a formal proof of the corollary is unnecessary, since the assumed bijectivity of the map g implies $S = g^{-1}(g(S))$ for every subset S of V. So Equivalences (3.41) and (3.42) follow immediately from Theorem 3.8.(iv), and these equivalences are rewritten in Equations (3.43) and (3.44).

---

[99] In other words, we define $f(g, \pi)(A) := \pi(\phi^{-1}(g^{-1}(A)))$ for every $g \in G$, $\pi \in \Pi$, and every $A \in \mathsf{B}$. Note that the map $u \mapsto g(\phi_u)$ is measurable for every $g \in G$, since it is a linear combination (with coefficients λ_i) of the maps $u \mapsto \phi_u(c_i, w_i)$, whose measurability follows from Footnote 80.

[100] Writing E instead of $\mathsf{E}_{U\sim\pi}$ and referring to Equation (3.40) for a map g:V→V', we have

$$\Phi'(f(g, \pi)) = \mathsf{E}g(\phi_U) = \mathsf{E}\sum_{i=1..n} \lambda_i\, \phi_U(c_i, w_i) = \sum_{i=1..n} \lambda_i\, \mathsf{E}\phi_U(c_i, w_i) = \sum_{i=1..n} \lambda_i\, \Phi_\pi(c_i, w_i) = g(\Phi_\pi).$$

For Implication (3.37), let an element g of G be given by Equation (3.40), and let elements S' and π of T' resp. Π be given such that $(g^{-1}(S'), \pi) \in H$. According to the definition of the relation H in Example 3.3.(v), this implies the existence of an element A of A_D with $\pi(A) = 1$ and $\phi(A) \subseteq g^{-1}(S')$, i.e., $g(\phi(A)) \subseteq S'$. Since S' is an element of T', it must be an interval of real numbers, i.e., a Borel set in $\mathbb{R}$, and $f(g, \pi)(S') = 1$ follows for this situation from the definition of the map f. But then the definition of the relation H' in Lemma 3.7 implies $(f(g, \pi), S') \in H'$.

Before the corollary is demonstrated by an application, it should be noted that the bijectivity of the map g and Equivalences (3.41) and (3.42) motivate an isomorphism concept, which is explicated in the following definition.

Definition 3.11: If (V, Π, Φ, H, T) and (V', Π', Φ', H', T') are SSAs, then a bijective map $g:V \to V'$ is an *SSA-isomorphism* (from (V, Π, Φ, H, T) onto (V', Π', Φ', H', T')) iff there exists a map $f:\Pi \to \Pi'$ such that Equation (3.36), Implication (3.35) and its reversal (with S' = g(S)), and Equivalences (3.41) and (3.42) hold for every subset S of V and every $\pi \in \Pi$. Furthermore, the SSAs (V, Π, Φ, H, T) and (V', Π', Φ', H', T') are *SSA-isomorphic* iff a map g with the above properties exists.

As a reference to this isomorphism concept, Corollary 3.10 will subsequently be called the *Isomorphism-Corollary*.[101]

For an application to the CDF-SSA of Example 3.3, recall that the choice of a vocabulary set based on cumulative probabilities has been motivated in Sections 2.3 and 2.4 by arguments of convenience: Wheras it is frequently hard to follow up with the details of an entire σ-algebra, the consistency requirements upon cumulative probabilities are much more easily summarised, e.g. by Inequality (3.5) and Equation (3.6) for the CDF-SSA. Furthermore, a vocabulary set consisting of maps $v:C\times W \to \mathbb{R}$ can be subsumed under a general theory of the function space $\mathbb{R}^Q$ (with $Q := C\times W$ for the present situation). Although it is intuitively convincing that this choice of the language of cumulative probabilities should lead to the same results as a characterisation of units and RSO-processes by ordered pairs of probability measures for the dependent variable Y, it may be satisfying that this conjecture can be reconstructed formally. So let M_W be the set of all probability measures on the σ-algebra B of Borel sets in $\mathbb{R}$, where a probability of 1 is assigned to our set W = [w', w''], and define $V' := M_W \times M_W$. Then a map $g:V \to V'$ can be defined by the requirement that g(v) with $v \in V$ is an ordered pairs of elements of M_W, whose components are the probability measures specified by the partial maps v(a, .) and v(b, .). With Π' := Π, the role of the map $f:\Pi \to \Pi'$ in Definition 3.11 can be taken by the identity map (i.e., $f(\pi) := \pi$). Finally, Equation (3.36) as well as Implication (3.35) and its reversal can be used to define an aggregation rule $\Phi':\Pi' \to V'$ and a relation $H' \subseteq PV' \times \Pi'$, and set systems T' and T'_e are uniquely specified by SSA-Axiom (iv) and Equation (1.1). Then the new SSA (V', Π', Φ', H', T') is SSA-isomorphic with the CDF-SSA of Example 3.3, since Corollary 3.10 confirms the intuitive conjecture that stability under aggregation is equivalent in both SSAs.

Another application of Isomorphism-Corollary 3.10 is also interesting. It resumes the aggregation of continuous CDFs by the technique of Vincentising, which has been specified in Section 2.2.6 by Equation (2.52). The vocabulary set V will again consist of all non-decreasing maps $v:\mathbb{R} \to [0, 1]$, where a p^{th} order quantile - i.e., a unique real number ξ with v(ξ) = p - exists for every $p \in]0, 1[$. Now let a set Π and a relation H be defined by Assertions (ii) and (iv) of Lemma 2.1, and consider an aggregation rule $\Phi:\Pi \to V$, where Φ_π is the map v* specified by Equation (2.52). Then the set system T of an SSA (V, Π, Φ, H, T) is specified by SSA-Axiom (iv). Although Lemma 2.1 cannot be applied directly to this SSA (due to its specific aggregation rule), an SSA-isomorphism allows an

---

[101] Note that the surjectivity of the map $f:\Pi \to \Pi'$, which is assumed in Corollary 3.10, is not required by Definition 3.11.

indirect application. For this purpose, we resume some other definitions of Section 2.2.6. With $Q :=]0, 1[$, let V' be the set of all continuous, strictly increasing maps $Q \to \mathbb{R}$, and let a bijective map $g:V \to V'$ be defined such that g(v) is the map $v':Q \to \mathbb{R}$, where v'(q') is the (unique) real number ξ with $v(\xi) = q'$.[102] Furthermore, let an SSA (V', Π', Φ', H', T') be given by the specifications of Lemma 2.1. In particular, recall (for the subsequent application of Isomorphism-Corollary 3.10) that T' is the system of all convex subsets of V'. Finally, let $f(g, \pi)$ for $\pi = \{(\lambda_i, v_i)\}_{i=1..n}$ be given by $f(g, \pi) := \{(\lambda_i, g(v_i))\}_{i=1..n}$. Then all premissas of the Isomorphism-Corollary can be easily verified, and the corollary leads to the conclusion that the set system T consists of those subsets S of V where the set g(S) is convex.

The use of the concept of an isomorphism raises the question whether the denotation SSA-homomorphism shouldn't be introduced for a similar situation without the requirement of injectivity for the map $g:V \to V'$. However, the underlying Mapping-Theorem 3.8 allows conclusions about the aggregation stability of a subset S of the vocabulary set V only if there is a subset S' of V' such that $S = g^{-1}(S')$. If the map g is injective, the set $S' := g(S)$ fulfills this requirement for every subset S of the vocabulary set V. But for a non-injective map g, the equivalences $S \in T \Leftrightarrow S' \in T'$ and $S \in T_e \Leftrightarrow S' \in T'_e$ do not hold for all subsets S of V and suitable subsets S' of V', and this property would be necessary to speak of a homomorphism.

Nevertheless, we can apply a well known approach, which is commonly used to transform a homomorphism $h:A \to B$ into an isomorphism. With an equivalence relation ~ on A, which holds for elements a' and a'' with h(a') = h(a''), the resulting set A/~ of equivalence classes in A can be mapped isomorphically onto h(A). As a first step in this approach, a relational structure on the set A, which is mapped into a similar structure on the set B, must be translated into a relational structure on the set A/~ of equivalence classes in A. This transition is described separately in the following lemma, since it can also be applied to equivalence relations on a vocabulary set which are not derived from a map into the vocabulary set of another SSA.

Lemma 3.12: Let an SSA (V, Π, Φ, H, T) and an equivalence relation ~ on the vocabulary set V of the SSA be given, and with $V^* := V/\sim$, let a map $g':V \to V^*$ be given such that g'(v) is the equivalence class of V. With $\Pi^* := \Pi$, let a map $Z^*:\Pi^* \to V^*$, a relation $H^* \subseteq \mathsf{P}V^* \times \Pi^*$, and set systems $T^* \subseteq \mathsf{P}V^*$ and $T^*_e \subseteq \mathsf{P}V^*$ be given by the equation

$$Z^*_{\pi^*} := g'(\Phi_{\pi^*}), \tag{3.45}$$

for every $\pi^* \in \Pi^*$, and by

$$H^* := \{(S^*, \pi^*) \in \mathsf{P}V^* \times \Pi^*: (g'^{-1}(S^*), \pi^*) \in H\}, \tag{3.46}$$

$$T^* := \{S^* \subseteq V^*: g'^{-1}(S^*) \in T\} \tag{3.47}$$

and

$$T^*_e := \{S^* \subseteq V^*: g'^{-1}(S^*) \in T_e\} \tag{3.48}$$

Then the ordered quintuple (V*, Π*, Z*, H*, T*) is an SSA, and for that SSA, the set system of Equation (1.1) is T^*_e.

The lemma is proved in Section 6.17. Its application to an equivalence relation based on a map between the vocabulary sets of two SSAs is decribed in the following corollary, which is proved in Section 6.18.

---

[102] See Section 6.34 for a more detailed analysis of the map $g:V \to V'$.

Corollary 3.13: Let $g:V \to V'$ be a map between the vocabulary sets of two SSAs (V, Π, Φ, H, T) and $(V', \Pi', \Phi', H', T')$, and let an equivalence relation ~ on V be given by the definition

$$v_1 \sim v_2 \;:\Leftrightarrow g(v_1) = g(v_2) \tag{3.49}$$

for all elements v_1 and v_2 of the set V. Furthermore, derive a map $g':V \to V^*$ and an SSA $(V^*, \Pi^*, Z^*, H^*, T^*)$ as described in Lemma 3.12, and let g'' be the unique map $g'':V^* \to V'$ with $g = g'' \circ g'$. Finally, assume that the map g is surjective and that there exists a surjective map $f:\Pi \to \Pi'$ such that Equation (3.36) as well as Implication (3.35) and its reversal hold for every $\pi \in \Pi$ and for all subsets S of V and S' of V' with $S = g^{-1}(S')$.

Then the map g'' is an SSA-isomorphism of the SSA $(V^*, \Pi^*, Z^*, H^*, T^*)$ onto the SSA $(V', \Pi', \Phi', H', T')$.

The denotation *Homomorphism-Corollary* for the above corollary has to be understood as a summary of the idea that a potentially non-injective map $g:V \to V'$ is treated in a similar way as a homomorphisms in the transformation into an isomorphism.

Certainly, a limitation for applications of the Homomorphisms-Corollary is the surjectivity-premissa for the maps $g:V \to V'$ and $f:\Pi \to \Pi'$; but sometimes this limitation can be overcome by suitable V'-Π'-restrictions according to Restriction-Lemma 3.4. If the map $f:\Pi \to \Pi'$ is non-surjective, we can first define $\Pi^{\sim} := f(\Pi)$ and replace the SSA $(V', \Pi', \Phi', H', T')$ by its $\Pi^{\sim}$-restriction in the way outlined immediately after Mapping-Theorem 3.8. So we may assume that the map f is surjective (possibly after a $\Pi^{\sim}$-restriction) in the treatment of a non-surjective map $g:V \to V'$. In this situation, we can define $V^{\sim} := g(V)$ and replace the SSA $(V', \Pi', \Phi', H', T')$ by its $V^{\sim}$-restriction.[103]

The treatment of maps between vocabulary sets should be rounded up by some comments referring to the map $f:\Pi \to \Pi'$ in Mapping-Theorem 3.8, Isomorphism-Corollary 3.10 and Homomorphism-Corollary 3.13 and to the map $f:G \times \Pi \to \Pi'$ in Mapping-Corollary 3.9. In Comment d to the definition of SSAs (Definition 1.2), it has been pointed out that the set Π of aggregates in an SSA is typically based upon the vocabulary set V by a higher order relationship, which can be verbalised as 'being composed of elements of V' in some way; but for purposes of generality, this property isn't formalised directly in the SSA-Axioms. The above applications of Mapping-Theorem 3.8 and Mapping-Corollary 3.9 to the CDF-SSA and to Interval-SSAs demonstrate this generality. An element of the set Π - i.e., a distribution π of a random variable U with values in the domain set D - is related to the vocabulary set V by a map $\phi:D \to V$ such that ϕ_U is the element of the vocabulary set characterising the randomly selected unit U. The relationship between the sets Π' and V' in the Interval-SSAs is much more direct: The elements of Π' are probability distributions of random variables with values in V'. Now the maps $f:\Pi \to \Pi'$ and $f:G \times \Pi \to \Pi'$ specify, how both kinds of 'being

[103] As a premissa of Lemma 3.4, the properties $\Phi'_{\pi'} \in V^{\sim}$ and $(V^{\sim}, \pi') \in H'$ must hold for every $\pi' \in \Pi'$; but under the assumption of a surjective map $f:\Pi \to \Pi'$, these properties follow from the premissas of Homomorphism-Corollary 3.13. To verify this claim, let an arbitrary element π' of Π' be given, and an element π of Π with $\pi' = f(\pi)$. Then $\Phi'_{\pi'} \in V^{\sim}$ follows from Equation (3.36). Furthermore, since the relation $(V, \pi) \in H$ is granted by SSA-Axiom (ii) and the equation $V = g^{-1}(V^{\sim})$ by the definition of the set $V^{\sim}$, we obtain $(V^{\sim}, \pi') \in H'$ from the reversal of Implication (3.35), the validity of this reversal being a premissa of the Homomorphism-Corollary.

composed of elements of the respective vocabulary set' are linked by the map $g:V\to V'$. So the maps $g:V\to V'$ and $f:\Pi\to\Pi'$ play a formally different role, which is expressed in Definition 3.11: Under certain conditions, the map g is an SSA-isomorphism, and the existence of a suitable map $f:\Pi\to\Pi'$ is one of these conditions.

It is typical that the random variable $g(\phi_U)$ played a central role in the specification of the maps f for our applications. But it is a price of the intended generality that results for such concatenations cannot be derived directly from the SSA-Axioms. So the properties needed for fruitful conclusions are introduced as premissas in Theorem 3.8 and in Corollaries 3.9, 3.10 and 3.13, and these premissas must be verified for applications. In Section 4.1, we will define a rather general class of SSAs, where the interpretation of the set Π as a set of probability measures is introduced as an additional axiom, and then it will be possible to verify the premissas once and for all for large classes of maps between vocabulary sets.

Observe also that the map f isn't necessarily the result of a concatenation. Consider e.g. the first demonstration of an SSA-isomorphism (immediately after Definition 3.11), where the vocabulary set V of the CDF-SSA was mapped into a set V' of ordered pairs of probability measures on B. In this case, the map $f:\Pi\to\Pi'$ was the identity map in Π. To generalise this approach, we can take a map $g:V\to V'$ of the vocabulary set V of an SSA (V, Π, Φ, H, T) into an arbitrary set V' and define an SSA $(V', \Pi', \Phi', H', T')$ by $\Pi' := \Pi$, $\Phi' := g \circ \Phi$, and

$$H' := \{(S', \pi) \in \mathsf{P}V'\times\Pi' : (g^{-1}(S'), \pi) \in H\}. \tag{3.50}$$

It is easily verified that the validity of SSA-Axioms (i), (ii) and (iii) follows for this relation H' from their (assumed) validity for the original relation H, whereas set systems T' and T'_e are determined by SSA-Axiom (iv) and Equation (1.1). With the identity-map $f:\Pi\to\Pi'$, everything is prepared to apply Assertion (iv) in Mapping-Theorem 3.8 to all subsets S of V and S' of V' with $S = g^{-1}(S')$.

The just outlined approach can be demonstrated by an alternative proof for the aggregation stability of the sets S'_w, S''_w and S^*_w in the CDF-SSA. With $V' := [-1, 1]$, let the map $g:V\to V'$ be the map g_w of Equatopn (3.28) for a given element w of the set W (i.e., $g(v) := v(a, w) - g(b, w)$ for every $v \in V$). With $\Pi' := \Pi$, the aggregation rule $\Phi':\Pi'\to V'$ is given by the equation

$$\Phi'_\pi := g(\Phi_\pi) = \mathsf{E}_{U\sim\pi}\, g(\phi_U) \tag{3.51}$$

for every $\pi \in \Pi'$, and the relation H' of Equation (3.50) consists of all ordered pairs (S', π) where S' is a subset of V' and π a probability distribution on A_D such that the properties $\pi(A) = 1$ and $g(\phi(A)) \subseteq S'$ hold for a suitable element A of the σ-algebra A_D. In other words, $(S', \pi) \in H'$ means that the random variable $g(\phi_U)$ is almost surely contained in S' for a random variable U with distribution π. Obviously, this implies $\Phi'_\pi \in S'$, if S' is a subinterval of V', which means that the set system T' contains all such intervals. So the identity map $f:\Pi'\to\Pi$ can be used to derive from Assertion (iv) of Mapping-Theorem 3.8 that the sets $S'_w = g^{-1}([0, 1])$, $S''_w = g^{-1}(]0, 1])$ and $S^*_w = g^{-1}([0, 0])$ are elements of T.

3.6 Intersections of Stable Sets

In Subsection 1.3 and Section 2, where convexity of a subset of a vocabulary set was sufficient for its stability under aggregation, a property of convex sets could be used as a powerful tool: If $\{S_i\}_{i\in I}$ is a family of convex subsets of a real vector space, then their intersection (i.e., the set $\bigcap_{i\in I} S_i$) is also convex. The present section will show that this approach can be generalised to intersections of sets, which are stable under an arbitrary kind of aggregation.

We start again with a demonstration of the underlying logical pattern for a subset of the vocabulary set V in the CDF-SSA of Example 3.3. In Subsection 3.5, we have already proved the aggregation stability of the sets S'_w representing the property $v(a, w) \geq v(b, w)$ for a given element w of the set W. Now consider the set S_O, which is the intersection of all these sets S'_w. According to SSA-Axiom (iv), the aggregation stability of this set is established, if we show that the property $(S_O, \pi) \in H$ implies $\Phi_\pi \in S_O$. For this purpose, observe that the inclusion $S_O \subseteq S'_w$ follows for every $w \in W$ from the definition of S_O. So the property $(S'_w, \pi) \in H$ is obtained from SSA-Axiom (iii) for every $\pi \in \Pi$ with $(S_O, \pi) \in H$. But since the sets S'_w are stable under aggregation, we have $\Phi_\pi \in S'_w$, and then Φ_π must also be contained in the intersection of the sets S'_w, i.e., in the set S_O.

The first claim of the following theorem (which is proved in Section 6.19) generalises the result. For equivalential aggregation stability (i.e., membership in the set system T_e), an additional requirement upon the relation H, which is expressed by Implication (3.52), allows a similar conclusion. Although its demonstration is again postponed until the treatment of equivalential aggregation stability in Subsection 3.8, it should be mentioned that the validity of the additional premissa will be established for a large class of SSAs by Lemma 4.3, but only for situations with a finite or countable index set I.

Theorem 3.14: For an SSA (V, Π, Φ, H, T) and an arbitrary non-empty set I, let $\{S_i\}_{i\in I}$ be a family of subsets of V with $S_i \in T$ for every $i \in I$, and define $S := \bigcap_{i\in I} S_i$.
Then the set S is also contained in T.
Furthermore, if $S_i \in T_e$ for every $i \in I$ and the implication
$$(\forall\, i \in I: (S_i, \pi) \in H) \Rightarrow (S, \pi) \in H \tag{3.52}$$
holds for every $\pi \in \Pi$, then S is contained in T_e.

The central role of intersections in the above theorem motivates the denotation *Intersection-Theorem*. Recalling that subsets of the vocabulary set are typically used to represent properties of interest, we can restate the central claim of the Intersection-Theorem: A property of interest is stable under a considered kind of aggregation, if it is equivalent with the joint occurence of (typically simpler) properties $\{P_i\}_{i\in I}$, which are stable under the same kind of aggregation.

For another application to the CDF-SSA of Example 3.3, consider the set S_L of those elements v of the vocabulary set V, where the equality $v(a, w) = v(b, w)$ holds for every $w \in W$. Since this set is the intersection of the sets S^*_w, whose aggregation stability has been proved in Subsection 3.5, the role of the index set I in Theorem 3.14 can be transferred to the set W, and $S_L \in T$ follows from the theorem.

For a slightly more complex application, we replace the assumption $C = \{a, b\}$ in Example 3.3 by a more general one: Let C be a set of experimental conditions, and let the vocabulary set V of a suitable SSA consist of all maps $v:C\times W\to\mathbb{R}$, where all partial maps $v(c, .\,)$ are CDFs with domain W. Furthermore, add a map $\phi:D\to V$ as well as the further components of an SSA as in Example 3.3.

Now assume that a quasi order $\preceq$ is defined on C, which is associated with the following hypothesis: If a and b are elements of C with $a \preceq b$, then the inequality $v(a, w) \geq v(b, w)$ holds for every $w \in W$. So let S_{OC} be the set of all elements of V with this property. For a more concrete interpretation of the quasi order $\preceq$, one can think of a multidimensional set C of treatment combinations. Then the relation $a \preceq b$ may be given for all ordered pairs (a, b) of elements of C, where treatment combination a isn't more 'favourable' (with respect to the dependent variable) in any treatment dimension.

To study the aggregation stability of the set S_{OC} , let the index set I of Intersection-Theorem 3.14 consist of all ordered pairs (a, b) of elements of C with $a \preceq b$. For convenience, we will also write ab instead of (a, b). For every such pair, let S_{ab} the set of all elements v of the set V with $v(a, w) \geq v(b, w)$ for every $w \in W$. Then it is easy to derive $S_{ab} \in T$ from Mapping-Theorem 3.8, the role of the SSA (V', Π', Φ', H', T') in that theorem being transferred to the original CDF-SSA of Example 3.3. But since S_{OC} is the intersection of all these sets S_{ab}, the property $S_{OC} \in T$ follows from Intersection-Theorem 3.14.

The aggregation stability of intersections of elements of the set system T has been introduced as a generalisation of a similar property of convex sets. We will now consider a closely related generalisation. If the vectors x_i in a convex linear combination $\sum_{i=1..n} \lambda_i x_i$ are elements of a non-convex set S, then the result isn't generally contained in S; but it isn't entirely unpredictable: It must be contained in the so-called convex hull of S, and this is the smallest convex set which includes S as a subset. A similar concept is prepared by the following corollary, which doesn't need a formal proof, since it follwos immediately from Intersection-Theorem 3.14.[104]

Corollary 3.15: For an SSA (V, Π, Φ, H, T), let S be a subset of V, and let the set system T_S be given by the definition

$$T_S := \{S' \in T: S \subseteq S'\}.$$

Then T_S is non-empty, and for every family $\{S_i\}_{i \in I}$ of elements of T_S with an arbitrary non-empty index set I, the intersection $\bigcap_{i \in I} S_i$ is contained in T_S. In particular, the intersection of all elements of T_S is the smallest element of T_S.

The main result of the corollary for a given set $S \subseteq V$ is the existence of a smallest subset of V, which is stable under aggregation and includes S. In analogy to the terminology for convex sets, this set will subsequently called the T-hull of S.[105]

Definition 3.16: In an SSA (V, Π, Φ, H, T), the T-hull of a subset S of V is the smallest subset of V which is contained in the set system T and includes S as a subset.

---

[104] Recall that the set V is always contained in the set system T. Hence it is contained in T_S, and T_S is non-empty.

[105] Note that the concept of a hull is also used in other set systems, which are closed under arbitrary intersection. For instance in a toplogy, the system of closed sets is has this property, and the closed hull of a set S is the intersection of those closed sets which includes S as a subset. See Birkhoff (###) for a general treatment of this issue.

The central properties of this T-hull are very similar to those of the convex hull in real vector spaces. Since they are immediate consequences of Definition 3.16, Corollary 3.15 and SSA-Axioms (iii) and (iv), they can be summarised without a formal proof in the subsequent corollary, which will be denoted as the *T-hull-Corollary*.

Corollary 3.17: For every subset S of the vocabulary set V in an SSA (V, Π, Φ, H, T), the T-hull of S has the following properties:

(i) If π is an element of Π with $(S, \pi) \in H$, then Φ_π is contained in the T-hull of S.

(ii) The T-hull of S is the intersection of all subsets of V, which are contained in T and include S as a subset.

(iii) The T-hull of S is the set S itself, if and only if S is contained in the set system T.

3.7 Conditional Stability under Aggregation

In the study of stability under convex linear combinations for strict stochastic order (Section 2.4), the set S_{st} turned out to be stable under the condition, that all elements of the vocabulary set V entering a convex linear combination with non-zero weight are contained in the set S_O. This conditional stability allowed to derive the stability under convex linear combinations of the set S_S, since this set is the intersection of S_O and S_{st}, and S_O is stable under convex linear combinations.

The present subsection will formalise the underlying concept of conditional stability under aggregation and the associated pattern of deduction.

According to SSA-Axiom (iv), a subset S of the vocabulary set V is stable under aggregation (i.e., contained in the set system T), if and only if the implication $(S, \pi) \in H \Rightarrow \Phi_\pi \in S$ holds for every $\pi \in \Pi$. On this background, aggregation stability is conditional upon a set S_0 (representing the condition), if the above implication is required only for elements π of the set Π with $(S_0, \pi) \in H$. Using a verbalisation, which has been introduced in Comments f and g to Definition 1.2, we can also say: The implication is required only for aggregates, which are 'sheerly composed' of elements fulfilling the condition.

Formally, this concept looks more general than membership in the set system T: Since the ordered pair (V, π) is contained in H for every $\pi \in \Pi$ (see SSA-Axiom (ii)), one could conceive membership in T as a special case of conditional aggregation stability with $S_0 = V$. But the following theorem opens another approach: Its Assertion (i) claims conditional aggregation stability in a the above introduced understanding, and the theorem states logical relationships between this assertion and other properties of a set S, which are explicated in Assertion (ii) and (iii) of the theorem. The relevance of these logical relationships will be demonstrated immediately.

Theorem 3.18: For an SSA (V, Π, Φ, H, T), let S_0 be a non-empty subset of V such that $(S_0, \pi) \in H$ for some element π of the set Π. Then there exists a unique SSA $(V, \Pi^\sim, \Phi^\sim, H^\sim, T^\sim)$ with

$$\Pi^\sim = \{\pi \in \Pi: (S_0, \pi) \in H\}, \tag{3.53}$$

$$Z\sim_\pi = \Phi_\pi \tag{3.54}$$

for every $\pi \in \Pi^\sim$, and

$$H^\sim = \{(S, \pi) \in \mathsf{P}V \times \Pi^\sim: (S, \pi) \in H\}. \tag{3.55}$$

In this situation, the implications (i)⇔(ii)⇒(i)⇒(iii) of the following properties hold for every subset S of V:

(i) The implication $(S_0, \pi) \in H \Rightarrow ((S, \pi) \in H \Rightarrow \Phi_\pi \in S)$ holds for every $\pi \in \Pi$.

(ii) $S \in T^{\sim}$.

(iii) The implication $(S \cap S_0, \pi) \in H \Rightarrow \Phi_\pi \in S$ holds for every $\pi \in \Pi$.

Furthermore, these properties are equivalent, if S is a subset of V such that the implication

$$((S, \pi) \in H \wedge (S_0, \pi) \in H) \Rightarrow (S \cap S_0, \pi) \in H \tag{3.56}$$

holds for every $\pi \in \Pi$.

The theorem, which will subsequently benoted as the *Conditioning-Theorem*, is proved in Section 6.20. To prepare an application to the CDF-SSA of Example 3.3, recall that we are mainly interested in the aggregation stability of the set S_S, which is the intersection of S_O and S_{st} So let π be an element of Π with $(S_S, \pi) \in H$.[106] Then the properties $(S_O, \pi) \in H$ and $(S_{st}, \pi) \in H$ are obtained from SSA-Axiom (iii). Furthermore, since the aggregation stability of the set S_O has already been established in Subsection 3.6, we get $\Phi_\pi \in S_O$ from $(S_O, \pi) \in H$. Given this situation, it suffices to verify the property $\Phi_\pi \in S_{st}$, and $\Phi_\pi \in S_S$ will follow immediately.

Now the implication (ii)⇒(iii) in Conditioning-Theorem 3.18 opens a way for this task: If we transfor the roles of the sets S_0 and S in the theorem to our sets S_O and S_{st} (in this order), then $\Phi_\pi \in S_{st}$ will follow, if we verify Assertion (ii) - i.e., $S_{st} \in T^{\sim}$. But then we can also use the implication (ii)⇒(i), since the properties $(S_O, \pi) \in H$ and $(S_{st}, \pi) \in H$ have already been established for the considered situation.

Before the property $S_{st} \in T^{\sim}$ will be proved in the next subsection, the general logical status of Conditioning-Theorem 3.18 should be recapitulated: Its Assertion (i) formalises the first intuitive understanding of conditional aggregation stability. The equivalence of Assertions (i) and (ii) has two noteworthy consequences. First, it is unnecessary to provide a separate concept of conditional aggregation stability in the axiomatisation of SSAs, since this property is equivalent with unconditional aggregation stability in a suitably restricted SSA. Furthermore, the equivalence (i)⇔(ii) can be used to prove conditional aggregation stability: The property $S \in T^{\sim}$ can be verified by the general tools for unconditional aggregation stability. Finally, the implications (i)⇒(iii) and (ii)⇒(iii) show, how conditional aggregation stability can be used to prove the aggregation stability of the sets $S \cap S_0$ in situations, where the property $\Phi_\pi \in S_0$ can be derived otherwise (e.g. from the property $S_0 \in T$ as in our example). Note also that the validity of Formula (3.56) - the premissa for an equivalence of Assertions (i), (iii) and (iii) - will be established for a large class of SSAs by Lemma 4.3.

[106] The existence of a suitable $\pi \in \Pi$ with $(S_S, \pi) \in H$ can be assumed without loss of generality: Otherwise the implication $(S_S, \pi) \in H \Rightarrow \Phi_\pi \in S_S$ is true for every $\pi \in \Pi$, and $S_S \in T$ follows immediately. But this situation is uninsterting.

3.8 Equivalential Stability under Aggregation

Although some results referring to equivalential aggregation tability (i.e., to the set system T_e in an SSA (V, Π, Φ, H T), which is defined by Equation (1.1)) have been stated in previous theorems and corollaries, their demonstration by applications has been postponed until the present subsection. So it may be helpful to recall the very first example, which has been used to demonstrate the equivalential stability, which is represented by that set system: In the process of mixing cocktails, the aggregate (i.e., the cocktail) is non-alcoholic, if and only if (!) all mixed ingredients have been non-alcoholic.[107] As a generalisation of this example, the set system T_e in an SSA (V, Π, Φ, H T) consists of those subsets S of the vocabulary set V, where the properties $(S, \pi) \in H$ and $\Phi_\pi \in S$ are equivalent for all elements π of the set Π:

It has already been mentioned in a comment to Definition 1.2 that the empty set and the entire vocabulary set V are always contained in the set system T_e. In some SSAs (e.g. in those Interval-SSAs of Lemma 3.7 where the vocabulary set V' is an open interval), these two sets are the only elements of T_e, and in other cases, this property is usually trivial. E.g. in the CDF-SSA of Example 3.3, it isn't at all surprising that every set S defined by

$$S := \{v \in V: v(c, w) = 0\} \tag{3.57}$$

with $c \in C$ and $w \in W$ is contained in T_e: Due to the CDF-interpretation of the partial maps v(c, .), the random variable $\phi_U(c, w)$ is almost surely contained in the interval [0, 1], and this implies that the expectation $E_{U\sim\pi}\, \phi_U(c, w)$ is equal to 0 if and only if $\phi_U(c, w)$ is almost surely equal to 0. In other words, for every such set S the properties $\Phi_\pi \in S$ and $(S, \pi) \in H$ are equivalent for every $\pi \in \Pi$,[108] which means $S \in T_e$. The same considerations apply, if the property v(c, w) = 0 in the above definition of a set S is replaced by v(c, w) = 1. Since the triviality and the minor relevance of the conclusion $S \in T_e$ for such sets is rather typical, equivalential stability is not represented by a separate component of an SSA in Definition 1.2, but only by a notational convention.

However, an exception is noteworthy: If an SSA (V, $\Pi^\sim$, $\Phi^\sim$, $H^\sim$, $T^\sim$) is derived from an SSA (V, Π, Φ H, T) according to Conditioning-Theorem 3.18, then equivalential aggregation stability in the derived SSA - i.e., membership in the set system $T^\sim_e$ - may be untrivial and useful for the analysis of conditional aggregation stability. For a demonstration of this claim, we resume the application of Conditioning-Theorem 3.18 to the CDF-SSA of Example 3.3, which has been presented in the preceeding subsection and ended in the result that the proof of the property $S_S \in T$ will be completed, if $S_{st} \in T^\sim$ is verified. In the present subsection, we will first show that the set S_L of all elements v of V with v(a, w) = v(b, w) is equivalentially stable in the derived SSA, and then it will be easy to derive $S_{st} \in T^\sim$.

The property $S_L \in T^\sim$ will be verified in two steps: We will first use Mapping-Corollary 3.9 to

---

[107] To be precise, a side condition of the statement has also to be recalled: It applies to mixtures, where alcohol is neither destroyed nor produced by chemical reactions.

[108] See the definition of the aggregation rule Φ and the relation H in Assertions (iv) and (v) of Example 3.3 for this reformulation, and note that the set $\phi^{-1}(S)$ (with S given by Equation (3.57)) is contained in A_D, since the map $u \to \phi_u(c, w)$ is measurable (see Footnote 80).

show that every set S^*_w (i.e., every set of all elements v of V with v(a, w) = v(b, w) for a given element w of the set W) is contained in $T^\sim_e$, and then the rest will be derived from Intersection-Theorem 3.14.

For the first step, let Π'' be the set of all probability measures π'' on the σ-algebra B of Borel sets in ℝ with π''([0, 1]) = 1, and let (V', Π'', Φ'', H'', T'') be the Π''-restriction of the Interval-SSA with V' := [−1, 1].[109] For the subsequent application of Mapping-Corollary 3.9, it suffices that the sets {0} and {1} are contained in T''_e. So let G be the set of the maps g_w:V→V' given by $g_w(v)$:= v(a, b) − v(b, w), and observe that the equality $S^*_w = g_w^{-1}(\{0\})$ follows immediately for every w ∈ W. Furthermore, let a map f:G×Π~→Π'' be given by the following specification: For every w ∈ W and every D-valued random variable U with distribution π ∈ Π~, the distribution of the random variable $g_w(\phi_U)$ is f(g_w, π).[110] Then everything is prepared to obtain $S^*_w \in T^\sim_e$ from {0} ∈ T''_e by an application of Mapping-Corollary 3.9.

To derive $S_L \in T^\sim_e$ from this result by Intersection-Theorem 3.14, recall that the set S_L could be redefined in Equation (3.19) as the intersection of all sets $S^*_{w^*}$ with w* ∈ W*, where W* is the set of all rational numbers contained in W. Hence the role of the set family $\{S_i\}_{i\in I}$ in Intersection-Theorem 3.14 can be transferred to the family $\{S^*_{w^*}\}_{w^*\in W^*}$, and the set S_L can take the role of the set S. Implication (3.52) is cleared by Lemma 6.2 in Subsection 6.15 of the Appendix. So the relation $S_L \in T^\sim_e$ follows from the Intersection-Theorem.

It is left to derive $S_{st} \in T^\sim$ from this result. So let π be an element of the set Π~ with (S_{st}, π) ∈ H~, and we will verify the property $\Phi^\sim_\pi \in S_{st}$. Now recall that the set Π~ contains only elements π with (S_O, π) ∈ H, and since S_O ∈ T has already been established, we obtain $\Phi_\pi \in S_O$ and $\Phi^\sim_\pi \in S_O$. Furthermore, it follows from the definition of the relation H (Example 3.3.(v)) that the assumption (S_{st}, π) ∈ H~ excludes (S_L, π) ∈ H~, since the sets S_{st} and S_L are disjoint.[111] But since

---

[109] See Lemma 3.7 for Inverval-SSAs, and Definition 3.5 for restrictions.

[110] To make sure that the specification of f(g_w, π) leads to a map f:G×Π~→Π'', it has to be verified that the distribution of the random variable $g_w(\phi_U)$ is an element of Π'' for every w ∈ W and every random variable U with distribution π ∈ Π~. So let w and π be given elements of W resp. Π~, and recall that Π~ contains only probability measures π on A_D with (S_O, π) ∈ H. According to the definition of the relation H for the CDF-SSA, there must exist an element A of A_D with π(A) = 1 and ϕ(A) ⊆ S_O. Then the definition of the set S_O leads to

$$0 \leq_{a.s.} \phi_U(b, w) \leq_{a.s.} \phi_U(a, w) \leq_{a.s.} 1$$

for every random variable U with distribution π (where a.s. stands for 'almost surely'), and this implies

$$0 \leq_{a.s.} g_w(\phi_U) \leq_{a.s.} 1.$$

So the distribution of the random variable $g_w(\phi_U)$ is an element of Π'', indeed.

[111] More generally, if S_1 and S_2 are disjoint subsets of V, and π is an arbitrary element of Π, then the joint presence of the relations (S_1, π) ∈ H and (S_2, π) ∈ H would imply the existence of elements A_1 and A_2 of the σ-algebra A_D with the properties ϕ(A_1) ⊆ S_1, ϕ(A_2) ⊆ S_2, and π(A_1) = π(A_2) = 1 (see Example 3.3.(v) for this conclusion). But for disjoint sets S_1 and S_2, sets A_1 and A_2 with the said properties would also be disjoint, and then the probabilites π(A_1) and π(A_2) couldn't both be 1.
(continued...)

the property $S_L \in T^{\sim}_e$ has already been established, we obtain $\Phi^{\sim}_\pi \notin S_L$ from $(S_L, \pi) \notin H^{\sim}$. In other words, there must be an element w of W with $\Phi^{\sim}_\pi(a, w) \neq \Phi^{\sim}(b, w)$. But since $\Phi^{\sim}_\pi(a, w) < \Phi^{\sim}(b, w)$ would be incompatible with the former result $\Phi^{\sim}_\pi \in S_O$, we have $\Phi^{\sim}_\pi(a, w) > \Phi^{\sim}(b, w)$, and this implies $\Phi^{\sim}_\pi \in S_{st}$.

Note that the redefinition of the set S_L by Equation (3.19) has been used in the above argumentation, and recall that the fruitfulness of such redefinitions has been claimed in Subsection 3.3 in a pleading for definitions of vocabulary sets, which implement semantic aspects of the intended interpretations of function values like v(c, w). But of course, the mere fact that we needed these additional properties for our proof doesn't exclude that the set S_L is also equivalentially stable under aggregation under weaker assumptions. So a contrary example is even more informative than the reference to the usage of the redefinition of the set S_L in the above argumentation; but the analysis of this example (which covers the rest of the present subsection) can be skipped without loss of continuity.

This contrary example is identical with the CDF-SSA of Example 3.3 with the exception of a seemingly minute alteration in the definition of the vocabulary set: Equation (3.6) is assumed only for c = a, and for c = b it is replaced by

$$v(b, w) = \lim_{n\to\infty} v(b, w - (w-w')/n) \tag{3.58}$$

for every $w \in W$. In other words, only the partial map v(a, .) is right-continuous like a CDF in the understanding of Definition 2.14; but the partial map v(b, .) is assumed to be left-continuous.[112] Whereas this alteration is a material deviation from the assumptions of Example 3.3, a further assumption is only a specification, which is introduced to facilitate the subsequent analysis: It is assumed that the set W is the interval [0, 1]. All other assumptions of Example 3.3 including the set definitions in Equations (3.10) through (3.16) are transferred to the new vocabulary set.[113]

It has to be admitted that it would be hard to find a plausible semantic interpretation for a vocabulary set with these properties; but this very fact emphasises the point which has to be made by the example: It may be fruitful to implement semantic aspects in the definition of mathematical objects making up the vocabulary set.

To verify that the set S_L is no longer contained in $T^{\sim}_e$ after the alteration, it suffices to construct a situation with a selection distribution $\pi \in \Pi^{\sim}$ such that $(S_L, \pi) \notin H^{\sim}$ and $\Phi^{\sim}_\pi \in S_L$. For this purpose, let a family $\{v_\lambda\}_{\lambda\in]0, 1[}$ of elements of V be defined such that

$$v_\lambda(c, w) = \tfrac{1}{2} w \tag{3.59}$$

[111] (...continued)
So the relations $(S_1, \pi) \in H$ and $(S_2, \pi) \in H$ exclude each other.

[112] If the partial maps v(a, .) and v(b, .) are both left-continuous like the CDFs of Bauer (###, see our Footnote ###), then minor changes in the proof of Lemma ### show that the redefinitions of the sets S_O, S_L and S_{st} by Equations ### are again equivalent.

[113] To verify that the map Φ_π:C×W→ℝ defined by Equation (3.9) is contained in the new vocabulary set, the proof in Footnote 81 has to be adapted suitably.

for $w < \lambda$, and

$$v_\lambda(c, w) = ½ (w + 1) \tag{3.60}$$

for $w > \lambda$. For $w = \lambda$, Equation (3.60) applies for $c = a$, and Equation (3.59) for $c = b$.[114] Furthermore, assume that there is a measurable map $t{:}D \to \mathbb{R}$ with $t(D) =]0, 1[$ such that $\phi_u = v_{t(u)}$ for every $u \in D$. Finally, let π be a probability measure on A_D (i.e., an element of Π) with the following property: If U is a random variable with distribution π, then the random variabl t(U) has a rectangular distribution over the interval $]0, 1[$.[115] In this situation, $\pi \in \Pi^{\sim}$ follows from $(S_O, \pi) \in H$, and the definition of the relation H in Assertion (v) for Example 3.3 leads to $(S_L, \pi) \notin H^{\sim}$;[116] but the aggregation rule (Equation (3.9)) yields the equation

$$\Phi^{\sim}_\pi(a, w) = \Phi^{\sim}_\pi(b, w) = w \tag{3.61}$$

for every $w \in W$,[117] and this implies $\Phi^{\sim}_\pi \in S_L$. So the properties $(S_L, \pi) \in H^{\sim}$ and $\Phi^{\sim}_\pi \in S_L$ are not equivalent for the given π, and this means that S_L is not contained in $T^{\sim}_e$.

---

[114] An application of Equation (3.59) with $c = b$ and $w = 1$ would lead to a violation of the property $v(b, 1) = 1$, which must hold for every $v \in V$ (see Inequality (3.5)). But this case woul occur only for $\lambda = 1$, whereas the index set for λ is the interval $]0, 1[$.

[115] In other words, it is assumed that the equation $\pi(t^{-1}(]0, \lambda])) = \lambda$ holds for every $\lambda \in]0, 1[$. To spare debates about the existence of a probability measure with these properties, we can add (only for the modified SSA) the assumption that the σ-algebra A_D is generated by all sets $t^{-1}(]0, \lambda[)$ with $\lambda \in]0, 1[$. Certainly, the map $t{:}D \to \mathbb{R}$ is measurable with respect to this σ-algebra (see Bauer, ###, for this conclusion). To make sure that Assertion (ii) of Example 3.3 holds for this σ-algebra, it has to be verified that A_D contains all subsets A of D, which can be written in the form of Equation (3.7) with $(c, w) \in C \times W$ and $\lambda \in [0, 1]$. For instance, the assumptions imply

$$\{u \in D: \phi_u(c, w) \le \lambda\} = \{u \in D: t(u) > w\} = t^{-1}(]w, 1[)$$

for $c = a$ and $w \le 2\lambda < w + 1$, and this set is contained in A_D, if the map $t{:}D \to]0, 1[$ is measurable. For sets given by Equation (3.7) with other configurations of c, w and λ, the containment in A_D can be checked in a similar way.

[116] Let A be subset A of D with $A \in \mathsf{A}_D$ and $\pi(A) = 1$, which implies that A is non-empty. Then the assumed situation implies the existence of a real number $\lambda \in]0, 1[$ such that $v_\lambda \in \phi(A)$. But since every such v_λ is not contained in S_L, we cannot have $\phi(A) \subseteq S_L$. In summary, there is no subset A of D with $A \in \mathsf{A}_D$, $\pi(A) = 1$ and $\phi(A) \subseteq S_L$. In this situation, $(S_L, \pi) \notin H$ follows from the definition of the relation H.

[117] For $t(U) \le w$ (i. e., with probability w), we have $\phi_U(a, w) = ½ (w + 1)$, and $\phi_U(a, w) = ½ w$ for $t(U) > w$ (i.e., whith probability $1 - w$). This leads to

$$\Phi_\pi(a, w) = ½(w (w + 1) + (1 - w) w) = w.$$

For $\phi_U(b, w)$, the respective conditions are $t(U) < w$ and $t(U) \ge w$; but the probabilities of both events are identical.

As a side result, it can be noted that the set S_S representing strict stochastic order is no longer stable under aggregation after the alteration of the vocabulary set: Although the ordered pair (S_S, π) with the above π is contained in the relation H,[118] we obtain $\Phi_\pi \notin S_S$. According to SSA-Axiom (iv), a single π with these properties suffices to exclud $S_S \in T$.

For a deeper understanding of the example, it may be interesting to note an implication of the alteration in the vocabulary set. The above defined elements v_λ of this set demonstrate that the equality v(a, w) = v(b, w) may be violated in an isolated element w of W, whereas it holds for all other elements. But for the vocabulary set in the original CDF-SSA of Example 3.3, the inequality v(a, w) > v(b, w) implied the existence of a real number $\delta > 0$ such that $v(a, w+\alpha) > v(b, w+\alpha)$ for every $\alpha \in]0, \delta[$, and a similar implication holds for v(a, w) < v(b, w).[119] As a consequence, every ordered pair (S^*_w, π) with the above considered π and $w \in W$ is contained in the relation $H^\sim$;[120] but although S_L is the intersection of these sets, the ordered pair (S_L, π) is not contained in $H^\sim$. So we have an example for a problem, which has been mentioned in Comment h to Definition 1.2: If $\{S_i\}_{i\in I}$ is a family of subsets of a vocabulary set and π an element of Π such that $(S_i, \pi) \in H$ for every $i \in I$, then the ordered pair $(\bigcap_{i\in I} S_i, \pi)$ may nevertheless be not contained in the relation H, although the verbal interpretation of the relation H as 'sheerness' may suggest an axiom forbidding such situations. Further, less artificial examples will be presented in Section ###. Since an axiom of this kind has not been introduced, it is necessary to include Implication (3.52) as a premissa for the treatment of equivalential stability in Intersection-Theorem 3.14. For the vocabulary set in the original CDF-SSA of Example ###, an implication of this kind is stated by Lemma 6.3 in Section 6.15 of the Appendix.

3.9 Analytical and Empirical Aggregation Stability

In the preceeding subsections, we have derived the aggregation stability of subsets of the vocabulary set V in the CDF-SSO (Example 3.3) from analytical properties of the subsets. In the present subsection, an example will show that membership in the set system T can also result from limitations of the elements v of the vocabulary set V which can be realised by the selection of a suitable element u of the set D with $\phi_u = v$. The example will first by used for a postponed demonstration referring to Mapping-Theorem 3.8. A further analysis of the example will lead to a distinction between analytical and empirical agregation stability, and some consequences for the following sections will be outlined.

Example 3.19: For the CDF-SSA (V, Π, Φ, H, T) of Example 3.3 and the Interval-SSA with $V' := [-1, 1]$ (Lemma 3.7), let w be an element of W with w < w", define a map $g_w: V \to V'$ by $g_w(v) := v(a, w) - v(b, w)$, and assume that the inequality $g_w(\phi_u) \leq 0.6$ holds for every element

---

[118] Since all members of the family $\{v_\lambda\}_{\lambda\in]0, 1[}$ are contained in S_S, we have $A \in \mathsf{A}_D$, $\pi(A) = 1$ and $\phi(A) \subseteq S_S$ for A := D.

[119] See the proof of Lemma 6.2 in Section 6.15.

[120] For a given $w \in W$, the set $A := t^{-1}(]0, 1[\setminus\{w\})$ has the properties $A \in \mathsf{A}_D$, $\pi(A) = 1$ and $\phi(A) \subseteq S^*_w$ required by the definition of the relation H.

u of the domain set D; i.e.,

$$g_w(\phi(D)) \subseteq [-1, 0.6]. \tag{3.62}$$

As in previous applications of Mapping-Theorem 3.8, let a map $f_w:\Pi\to\Pi'$ be given such that $f_w(\pi)$ is the distribution of the random variable $g_w(\phi_U)$ in situations, where U is a D-valued random variable with distribution π. Then the maps g_w and f_w can take the role of the maps $g:V\to V'$ and $f:\Pi\to\Pi'$ in the Mapping-Theorem, and it is easily verified that the premissas of the theorem apply to all sets $S \subseteq V$ and $S' \subseteq V'$ with $S = g^{-1}(S')$, but with an exception referring to the additional premissa of Assertions (iii) and (iv): The map f_w is non-surjective. If π' is an element of Π' with $\pi'(]0.6, 1]) > 0$, then there can be no element π of Π with $\pi' = f_w(\pi)$, since Formula (3.62) implies that the random variable $g_w(\phi_U)$ cannot be contained in the interval]0.6, 1] with a non-zero probability.

To ascertain the relevance of this property, let S' be the union of the intervals [0, 0.3] and [0.7, 1], and define $S := g_w^{-1}(S')$. Since S' is not an interval, it follows immediately from Lemma 3.7 that the set S' is not contained in T'. Indeed, since our set S' is a Borel set in $\mathbb{R}$, the definition of the relation H' in Lemma 3.7 implies for this set that the property $(S', \pi') \in H'$ is equivalent with the equation $\pi'(S') = 1$, and there are many such elements π' of Π' with $\Phi'_{\pi'} \notin S'$. But both properties can hold jointly only for elements π' of Π' with $\pi'([0.7, 1]) > 0$, whereas Formula (3.62) implies that there can be no $\pi \in \Pi$ such that the random variable $g_w(\phi_U)$ is contained in the interval [0.7, 1] with a non-zero probability. If the property $(S, \pi) \in H$ holds for the distribution π of a D-valued random variable U, then the random variable $g_w(\phi_U)$ is almost surely contained in the interval [0, 0.3],[121] and this leads to $g_w(\Phi_\pi) \in [0, 0.3]$ and hence to $\Phi_\pi \in S$. In summary, we have $S \in T$ and $S' \notin T'$.

Before we draw conclusions from the example, note that the underlying assumption, which is expressed by Formula (3.62), can very well apply to concrete experimental situations. Recalling that v(a, w) and v(b, w) are probabilities of the event $Y \leq w$ under conditions a and b (where the random variable Y is the dependent variable of an experiment), we can reformulate the assumption in the following way: The probability of the event $Y \leq w$ may be greater under condition a than under condition b; but then the difference of this probability under the two conditions (i.e., the difference $v(a, w) - v(b, w)$) isn't greater than 0.6 for any element u of the domain set D.

A first generalisation of the example refers to the premissa of a surjective map $f:\Pi\to\Pi'$ in Assertion (iii) of Mapping-Theorem 3.8, whose claim can be rewritten by the implication $S' \notin T' \Rightarrow S \notin T$. Now $S' \notin T'$ means that there exists an element π' of Π', which prevents the membership of S' in T' by the simultaneous occurence of the properties $(S', \pi') \in H'$ and $\Phi'_{\pi'} \notin S'$. If there exists an element π of Π with $f(\pi) = \pi'$, then this π is suitable to prevent $S \in T$, since the

[121] For a given element π of Π with $(S, \pi) \in H$, let A be a subset of D with $A \in \mathsf{A}_D$, $\pi(A) = 1$, and $\phi(A) \subseteq S$ (see Assertion (v) of Example 3.3). The last property implies $g_w(\phi(A)) \subseteq g_w(S)$, and since A is a subset of D, we have $A = A \cap D$. Combining these results with Formula (3.62), we obtain

$$g_w(\phi(A)) = g_w(\phi(A \cap D)) = g_w(\phi(A)) \cap g_w(\phi(D)) \subseteq g_w(S) \cap [-1, 0.3] = [0, 0.3].$$

properties $(S, \pi) \in H$ and $\Phi_\pi \notin S$ follow under the other assumptions underlying the assertion.[122] But if no π with these properties exists, then S is contained in T, even if S' is not contained in T'. Obviously, such situations can be excluded, if the map $f:\Pi\rightarrow\Pi'$ is surjective; but weaker assumptions, which can replace this premissa, are specified in a note after the proof of the Mapping-Theorem in Section 6.16.

Beyond demonstrating the role of the surjectivity-premissa, Example 3.19 points out a fundamental difference between the CDF-SSA and the Interval-SSAs. An explication of this difference is facilitated, if we introduce a suitable σ-algebra A_V in V, which is the coursest σ-algebra, where all projection maps $pr_{cw}:V\rightarrow\mathbb{R}$ are measurable.[123] Then the map $\phi:D\rightarrow V$ is measurable,[124] and ϕ_U is a V-valued random variable, whose distribution depends upon the distribution π of the random variable U: For every subset A of V, which is contained in A_V, the prohability of the event $\phi_U \in A$ is $\pi(\phi^{-1}(A))$. But it depends on the (typically unknown) set $\phi(D)$ in a given empirical situation, which distributions of ϕ_U can be realised by suitable selection distributions. E.g., in the situation described by Formula (3.62), there is no distribution π of the random variable U such that the random variable ϕ_U is contained in the set $g_w^{-1}(]0.6, 1])$ with a non-zero probability. For the Interval-SSAs, the matter is different: It follows from the definition of these SSAs, which distributions of V'-valued random variables are realised by suitable elements of the set Π'.

It is an immediate consequence of this property of Interval-SSAs that the set systems T' and T'_e are logically determined by the other components of the SSA.[125] For the CDF-SSA, where we do not know the distributions of the random variable ϕ_U which can be realised by a suitable selection distribution π, it is almost remarkable that some statements about all empirically realisable distributions of ϕ_U can be made. For instance, one of these statements is: If the distribution is such that ϕ_U is almost surely contained in S'_w, then Φ_π is also contained in S'_w. So the property $S'_w \in T$ can be derived analytically, and the same holds for all sets whose aggregation stability has been proved in the preceeding subsections. But for some subsets of the vocabulary set V (e.g. the set S of Example 3.19), their membership in the set system T depends on the set $\phi(D)$ of a concrete empirical situation. In other words, the set system T can be partioned into two subsystems: For some subsets of V, their aggregation stability (i.e., their membership in T) is analytical, and for others it is empirical.[126]

---

[122] See the proof of Mapping-Theorem 3.8 in Section 6.16.

[123] See Equation (3.39) for the definition of the projection maps.

[124] The definition of A_D is equivalent with the requirement that A_D is generated by the system of all sets $\{v \in V: v(c, w) \leq \lambda\}$ with $(c, w) \in C\times W$ and $\lambda \in [0, 1]$. So the measurability of the map $\phi:D\rightarrow V$ is granted by Assertion (ii) of Example 3.3.

[125] Formally, the description of the set system T' was introduced as a premissa and not as a consequence in Lemma 3.7. But according to SSA-Axiom (iv), the set system T' is completely determined by the other components of the SSA.

[126] Note that the concept of 'emprical' membership in T has to be understood as a summary of the (continued...)

Undoubtedly, this situation is somewhat unsatisfactory: In studies of stability under aggregation, we are typically interested only in analytical stability. It could be considered to define the general concept of SSAs in a way excluding such consequences. But in concrete research situations, we have to cope with the fact that we do not know, which distributions of a random variables like ϕ_U or $g_w(\phi_U)$ can be realised empirically. In a discussion of Example 4.12 in Section 4.3, we will point out problems in the derivation of testable predictions from hypotheses referring to all elements of a domain set like the set D in the CDF-SSA of Example 3.3, and these problems could be hard to detect under a definition of aggregation stability confined to analytical stability. Therefore, the CDF-SSA of Example 3.3 has been introduced deliberately to demonstrate not only the problem, but also a way to its solution, which is based on a modification of the SSA.

Whereas Example 3.3 was based on a random variable U representing a randomly selected element of a set D of persons or persons in situations such that ϕ_U is an element of the vocabulary set characterising the selected unit U, we can also contract both ideas and start with a random variable U' with values in the vocabulary set, which may be identified with the former ϕ_U. But whereas the set of possible distributions of ϕ_U was empirically confined by the fact that ϕ_U is always contained in the set $\phi(D)$ with probability 1, the modified SSA has a set of aggregates consisting of all probability measures on a suitable σ-algebra in the vocabulary set. To keep the two SSAs apart, the components of the modified SSA are denoted as V', Π', Φ', H' and T' in the following specification of the modified SSA.

Example 3.20: Let the vocabulary set V' of an SSA be identical with the vocabulary set V in the CDF-SSA of Example 3.3. Furthermore, let $\mathsf{A}_{V'}$ be the coursest σ-algebra in V', where all projection maps pr_{cw}:V'→$\mathbb{R}$ are measurable, and let Π' the set of all probability measures on $\mathsf{A}_{V'}$. For an aggregation rule Φ':Π'→V', let $\Phi'_{\pi'}$ be given by the equation

$$\Phi'_{\pi'}(c, w) := \mathsf{E}_{U' \sim \pi'}\, U'(c, w) \tag{3.63}$$

for every π' ∈ Π' and every (c, w) ∈ C×W. Finally, let a relation H' ⊆ PV'×Π' be given such that an ordered pair (S', π') with S' ⊆ V' and π' ∈ Π' is contained in H' iff there exists a subset A of S' with A ∈ $\mathsf{A}_{V'}$ and π'(A) = 1. Then SSA-Axioms (i), (ii) and (iii) are easily verified for the relation H', and the set systems T' and T'_e in an SSA (V', Π', Φ', H', T') are determined by SSA-Axiom (iv) and Equation (1.1).

Under the aspect of analytical and empirical aggregation stability, the SSA of Example 3.20 is much more similar to the Interval-SSAs than to the CDF-SSA of Example 3.3: Membership in the set systems T and T_e is entirely determined analytically by the definitions of the other components of

---

126 (...continued)
idea that this membership depends upon the set $\phi(D)$ in a given empirical situation. Certainly, this membership is not empirical in the sense that it can be determined uniquely by empirical operations. However, some sampling strategies, which will be described in Section 5.3, could be used to test the hypothesis that the set $\phi^{-1}(g_w^{-1}[0.7, 1]))$ is empty, and under this assumption, the set S of Example 3.19 is contained in T. So the denotations 'empirical membership in T' or 'empirical aggregation stability' are not entirely nonsensical. The philosophical distinction of analytical and synthetical assertions could be used to speak of analytical and synthetical aggregation stability; but to spare philosophical overload, the denotation 'empirical aggregation stability' will be maintained.

the SSA without any reference to empirical facts. So the SSA of Example 3.20 will subsequently called the *analytical CDF-SSA*, and the denotation *empirical CDF-SSA* will be used for Example 3.3. The terminology can be generalised: An SSA is analytical iff its components are defined without reference to unknown empirical facts, and otherwise it is empirical.

Whereas Example 3.19 has demonstrated difficulties with the surjectivity-premissa in Assertion (iii) of Mapping-Theorem 3.8 in the study of the empirical CDF-SSA, this premissa is much more uncomplicated for the analytical CDF-SSA. E.g., if the roles of the SSAs (V, Π, Φ, H, T) and (V', Π', Φ', H', T') in the Mapping-Theorem are transferred to the analytical CDF-SSA and to the Interval-SSA with V' := [−1, 1], we can apply the well known map g_w:V→V' given by $g_w(v) := v(a, w) - v(b, w)$ and define a map f_w:Π→Π' by $f(\pi)(A) := \pi(g_w^{-1}(A))$ for every $\pi \in \Pi$ and every Borel set A in $\mathbb{R}$. Then it is easily verified that the map f_w is surjective for w < w'' (where w'' is the upper boundary of W), and then the implication S ∈ T ⇒ S' ∈ T', which results from Assertion (iii) of Mapping-Theorem 3.8, can be rewritten as S' ∉ T' ⇒ S ∉ T to show that the set S of Example 3.19 is not stable under aggregation in the analytical SSA, since it is identical with $g_w^{-1}(S')$ and S' is not contained in T'. More generally, the surjectivity premissa underlying Assertions (iii) and (iv) of Mapping-Theorem 3.8 will typically be available much more easily in analytical SSAs than in empirical ones.

It may also contribute to an understanding of the relationship between the two CDF-SSAs, if we apply the results of Section 3.5 for maps between the vocabulary sets of SSAs to the identity map g:V→V', a suitable map f:Π→Π' being given by[127]

$$f(\pi)(A') := \pi(\phi^{-1}(A')) \tag{3.64}$$

for every $\pi \in \Pi$ and every $A' \in \mathsf{A}_{V'}$. In other words, the element ϕ_U of the vocabulary set characterising a unit U selected with distribution π is considered as a random variable on $\mathsf{A}_{V'}$, and f(π) is the distribution of this random variable. Now this map will be non-surjective if the set ϕ(D) is a proper subset of the vocabulary set V; but this limitation can be overcome: With $\Pi^{\sim} := f(\Pi)$, let $(V', \Pi^{\sim}, \Phi^{\sim}, H^{\sim}, T^{\sim})$ be the $\Pi^{\sim}$-restriction of the analytical CDF-SSA, and transfer the role of the SSA (V', Π', Φ', H', T') in Mapping-Theorem 3.8 to that $\Pi^{\sim}$-restriction.[128] Some problems with premissas of the Mapping-Theorem can be avoided if we add the assumptions that the set ϕ(D) is contained in $\mathsf{A}_{V'}$ and that A_D is the coarsest σ-algebra in D where the map ϕ is A_D-$\mathsf{A}_{V'}$-measurable, the second one of these assumptions being equivalent with the equation

$$\mathsf{A}_D = \{\phi^{-1}(A') : A' \in \mathsf{A}_{V'}\}. \tag{3.65}$$

Under these assumptions, it can be derived from Isomorphism-Corollary 3.10 that the identity map

[127] See Footnote 124 for the A_D-$\mathsf{A}_{V'}$-measurability of the map ϕ:D→V, which is assumed for Equation (3.64).

[128] A similar transition to a $\Pi^{\sim}$-restriction with $\Pi^{\sim} := f(\Pi)$ may also be necessary in the analysis of analytical SSAs to fulfill the surjectivity premissa of Theorem 3.8.(iii) and (iv). Note, however, that the set f(Π) is determined analytically in analytical SSAs, whereas it depends upon the empirically determined set ϕ(D) in the empirical CDF-SSA.

$g{:}V \to V'$ is an SSA-isomorphism of the emprical CDF-SSA upon the $\Pi^{\sim}$-restriction of the analytical CDF-SSA[129], which implies $T = T^{\sim}$ and $T_e = T^{\sim}_e$. With the notation T' and T'_e for the respective set systems in the analytical CDF-SSA, the inclusions $T' \subseteq T^{\sim}$ and $T'_e \subseteq T^{\sim}_e$ follow from Restriction-Lemma 3.4.(ii). Combining these results, we obtain $T' \subseteq T$ and $T'_e \subseteq T_e$. So analytical implicative (resp. equivalential) aggregation stability can be identified with membership in T' (resp. in T'_e), and emiprical implicative (resp. equivalential) aggregation stability with membership in $T \backslash T'$ resp. $T_e \backslash T'_e$.

The matter becomes more complicated if the above premissas referring to the σ-algebras A_D and $A_{V'}$ are violated. Since the consequences of such violations are hard to follow up for the two CDF-SSAs, their discussion is postponed until Section 4.3, where the linkage between analytical and empirical aggregation stability by applications of the Mapping-Theorem and of SSA-restrictions will be resumed in comments to Corollary 4.11 and to Example 4.12.

3.10

3.10 Concluding Remarks

The results of the preceeding subsections have been announced to be tools for the study of stability under aggregation. However, it may be questioned whether they help much. Do they really reduce the trouble of analysing the class of mathematical objects forming a vocabulary set, or isn't this trouble only shifted to the verification of premissas?

The empirical CDF-SSA of Example 3.3 has deliberately been chosen for applications to demonstrate that it may be necessary, indeed, to use analytical properties of a vocabulary set in the verification of premissas. E.g., a seemingly minute alteration of the vocabulary set by Equation (3.58) turned out to invalidate a premissa of Intersection-Theorem 3.14. More generally, careful mathematical analysis isn't rendered unnecessary by the tools presented in preceeding subsections; but these tools can serve as heuristic aid in the search for useful properties. E.g., Mapping-Theorem 3.8 can be seen as an invitation to look whether a set S of interest can be conceived as an inverse image $g^{-1}(S')$ of a set S' in a better known SSA under a suitable map g, the criteria of suitability being stated in the premissas of the theorem. Similarly, Intersection-Theorem 3.14 suggests to ask whether a set can be fruitfully represented as an intersection of sets, whose aggregation stability is analysed more easily, and Conditioning-Theorem 3.18 is a prototypical example of capitalising on $\Pi^{\sim}$-restrictions of SSAs.

To some extent, the necessity to verify analytically the premissas of theorems and corallaries presented in the preceeding subsections is also due to the fact that these results were entirely based upon the SSA-Axioms of Definition 1.2. In Section 4, these axioms will be supplemented by other weak assumptions, which apply to many aggregation processes in psychology. Under these additional axioms, some premissas of the preceeding results will be verified once and for all.

---

[129] It may be a useful exercise for the reader to verify the equations and equivalences needed for an applications of the Isomorphism-Corollary. But the claimed isomorphism property of the identity map in the common vocabulary set of the two CDF-SSAs can be derived more directly from Corollary 4.11.(ii), whose proof in Section 6.25 generalises the application of the Isomorphism-Corollary.

4 Stability under Stochastic Aggregation

In the general concept of aggregation stability explicated in Definition 1.2, it has been left almost entirely unspecified, in which way an aggregate is composed of elements, and how these elements and the whole aggregate are related to a vocabulary set, whose subsets represent properties of interest. In the present section, we will confine the analysis by additional weak assumptions. The direction of this confinement can be demonstrated by some aggregates, which are excluded by the added assumptions. A sentence is an aggregate of words, and a computer graphic is an aggregate of pixels; but in both cases, the aggregate isn't sufficiently specified by the quantitative proportions of the occurence of the elements.

Under a suitably general understanding of the concept of a probability distribution, we can say that the set Π of aggregates in all hitherto analysed SSAs was a set of probability distributions of random variables, whose values were either contained in the vocabulary set or mapped into this set by a map underlying the relation H.[130] This property will be introduced as a formal axiom of *stochastic SSAs* in Subsection 4.1. Some properties of the vocabulary set V and the aggregation rule Φ, which have also been present in some examples, will not be constitutive for stochastic SSAs; but they will be introduced by additional definitions, since they will enable useful conclusions. In particular, it has been frequently assumed that the vocabulary set V is a set of maps $v{:}Q\to\mathbb{R}$ with a suitable set Q, i.e., a subset of the function space $\mathbb{R}^Q$. In most such cases, the aggregation rule Φ used the expectations of real valued random variables, which were based on the projection maps $pr_q{:}V\to\mathbb{R}$ given by $pr_q(v) := v(q)$. Such SSAs will be called projection-based.

In Subsections 4.3 and 4.6, some basic properties will be derived from the definitions of Subsection 4.1. In particular, applications of Mapping-Theorem 3.8 and the corollaries derived from that theorem will be supported by some properties of maps from the vocabulary set of a stochastic SSA into the set of real numbers, which are presented in Subsections 4.3, 4.6 and 4.5. The association between the convexity of a set and its stability under aggregation, which has become questionable by Example 3.1, will be reconsidered in Subsection 4.7. For readers who are familiar with the theory of locally convex Hausdorff spaces, an approach to stochastic SSAs in the framework of this theory will be outlined in Subsection 4.8.

4.1 Stochastic Structures of Aggregation Stability

In this subsection, we will introduce a rather general class of SSAs covering many processes of aggregation, which are relevant in psychology. Their common property consists in the interpretation of the set Π of aggregates as a set of probability measures, and a corresponding definition of the

---

[130] In the SSAs for stability under averaging or convex linear combinations (Lemmas 1.4 and 2.1), the set Π has been formally defined as a set of finite sequences. But in both cases, it is easy to define an equivalence relation ~ on Π such that $\pi' \sim \pi''$ holds for elements π' and π'' of Π iff the resulting distribution of elements of the vocabulary set V is identical. Then the resulting set $\Pi/\!\sim$ of equivalence classes would be a set of probability distributions, and the equivalence class of an element of Π determines all relevant properties: For $\pi' \sim \pi''$, we have $\Phi_{\pi'} = \Phi_{\pi''}$, and the relations $(S, \pi') \in H$ and $(S, \pi'') \in H$ are equivalent for all subsets S of the vocabulary set V.

relation H.

For a first approach to the subsequent definition of stochastic SSAs, consider the emprical CDF-SSA (Example 3.3). The main part of the definition generalises the formal roles of the measurable space (D, A_D) and the map ϕ:D→V in that SSA. After some comments on the meaning of the assumed properties in this context, other possible interpretations of the measurable space an the map ϕ underlying a stochastic SSA will be pointed out.

Definition 4.1: An SSA (V, Π, Φ, H, T) is ϕ-*stochastic* iff there exists a measurable space (Ω_0, A_0) with the following properties (v), (vi) and (vii):

(v) Π is a set of probability measures on A_0.

(vi) ϕ is a map $\phi:\Omega_0 \to V$.

(vii) An ordered pair (S, π) with $S \subseteq V$ and $\pi \in \Pi$ is an element of H iff there is an element A of the σ-algebra A_0 such that

$$\phi(A) \subseteq S \tag{4.1}$$

and

$$\pi(A) = 1. \tag{4.2}$$

In this situation, we will also say that the SSA (V, Π, Φ, H, T) is based on the measurable space (Ω_0, A_0) and on the map $\phi:\Omega_0 \to V$, and the ordered triple (Ω_0, A_0, ϕ) will be called a stochastic base of the SSA.

An SSA is a *stochastic SSA* if there exists a measurable space (Ω_0, A_0) and a map ϕ such that the SSA is ϕ-stochastic.

An SSA (V, Π, Φ, H, T) is *identity-based* iff it is ϕ-stochastic for the identity map ϕ:V→V.

The numbering of the assumed properties begins with (v), since they supplement the SSA-Axioms (i), (ii), (iii) and (iv) introduced in Definition 1.2. Therefore we will also speak of *SSA-Axioms* (v), (vi) and (vii).

The new SSA-Axioms are generalisations of similar assumptions in the empirical CDF-SSA. (See Example 3.3.(ii), (iii) and (v), and note a difference: In the empirical CDF-SSA, the set Π is the set of *all* probability measures on A_D; but SSA-Axiom (v) requires only that Π is a set of probability measures on A_0 and leaves it open that some probability measures on A_0 are not contained in the set Π.) Again it is easily verified that the relation H specified by SSA-Axiom (vii) complies with SSA-Axioms (i), (ii) and (iii) in Definition 1.2.[131]

One concept explicated in Definition 4.1 is left to be demonstrated: The identity-based stochastic SSA. Indeed, the empirical CDF-SSA doesn't belong to this class, since the map ϕ:D→V of that SSA is not an identity map. For the analytical CDF-SSA, a set Ω_0, a σ-algebra A_0 and a map $\phi:\Omega_0 \to V'$ have not been specified in its description (Example 3.20); but the definition of identity-based stochastic SSAs requires only their existence, which is easily verified: With Ω_0 := V' and $\mathsf{A}_0 := \mathsf{A}_{V'}$,

[131] Formally, it could be criticised that the SSA-Axioms are not independendent, since SSA-Axiom (vii) implies SSA-Axioms (i), (ii) and (iii). But we should put up with this property, which may occur in specialisation networks of axiom systems.

the role of the map $\phi:\Omega_0\to V'$ can be taken by the identity map in V', and then the definitions of the set Π' and the relation H' in Example 3.20 comply with SSA-Axioms (v) and (vii). The same approach can also be used to show that the Interval-SSAs of Lemma 3.7 are identity-based stochastic SSAs.

For notational convenience, it will be useful to agree upon the following rule: In an identity-based stochastic SSA, the σ-algebra in the vocabulary set will always be denoted by A subscipted with the denotation of the vocabulary set. Indeed, this rule has already been followed in the definitions of Interval-SSAs and of the analytical CDF-SSA, where the σ-algebra in the vocabulary set V' is $\mathsf{A}_{V'}$.

At first glance, it may look like a roundabout way to reconceive the analytical CDF-SSA and the Interval-SSAs in this way. Indeed, if all stochastic SSAs of interest would be identity-based, then a more direct description (as in Example 3.20 and in Lemma 3.7) without a set Ω_0, a σ-algebra A_0 and a map $\phi:\Omega_0\to V$ would be preferable. But since SSAs like the empirical CDF-SSA will be used for applications, it is advantageous to have a common formalisation such that results derived from the definition apply to both types of stochastic SSAs. Furthermore, some useful results for maps between the vocabulary sets of stochastic SSAs couldn't be obtained, if the property of being identity-based would be constitutive for stochastic SSAs.[132]

Another interpretation of the set Ω_0 and the σ-algebra A_0 is also worth mentioning: They can be identical with the first two components of the most basic probability space governing random events, whose typical denotation is (Ω, A, P). For such situations, SSA-Axiom (v) implies that Π is a set of probability measures on A, and the aggregation rule Φ can be any assignment of elements of a set V to such probability measures. Furthermore, if a map $\phi:\Omega\to V$ is measurable with respect to the σ-algebra A and a suitable σ-algebra in the vocabulary set of the considered SSA, then we can also speak of a random variable ϕ with values in V, and $\Phi(P)$ may be an element of the set V characterising the distribution of the random variable ϕ, e.g., the expectation of ϕ[133]. Then a statement claiming the aggregation stability of a subset S of the vocabulary set V has the following format: For every probability measure $P \in \Pi$ where the random variable ϕ is almost surely contained in S, the expectation of ϕ is also contained in S.

The formal simplicity of the situation resulting from an identification of the set Ω_0 and the σ-algebra A_0 with the first two component of a most basic probability space may raise the question why this identification isn't introduced generally in the definition of stochastic SSAs. But it should be recalled from Section 3.9 that some useful properties for the CDF-SSAs can be lost if the σ-algebra A_D is finer than the coarsest one complying with Example 3.3.(ii). More generally, some results to be presented in later subsections will work only under premissas which would be violated by a too fine σ-algebra A_0. So it is advantageous in many situations to work with a concatenation $\phi \circ U$ of a map $U:\Omega\to\Omega_0$ and a map $\phi:\Omega_0\to V$.

It may also be questioned why Definition 4.1 doesn't introduce a σ-algebra in the vocabulary set

---

[132] For examples, see Corollaries 4.9 and 4.27, and compare their results with the more complicated ones reported in Corollary 4.10 and Theorem 4.25 for situations, where the vocabulary set of any stochastic SSA is mapped into the vocabulary st of an identity-based SSA.

[133] Of course, an interpretation of $\Phi(P)$ as expectation of the random variable ϕ is possible only under a suitable definition of this expectation. See e.g. Subsection 4.8 for a generalised conecpt of expectations of random values with values in a real vectur space.

V to enable a conceptualisation of ϕ as a random variable with values in V. But for some considerations in subsequent sections, it will be advantageous to have some freedom in the choice of a σ-algebra in V, and then it would be impedimental to have introduced it in advance. Note, however, that Definition 4.1 provides a special class of stochastic SSAs with a σ-algebra in the vocabulary set V: The identity-based SSAs. The hitherto introduced examples of such SSAs - the Interval-SSAs (Lemma 3.7) and the analytical CDF-SSA (Example 3.20) - demonstrate that identity-based SSAs are typically handled more easily than the general case of potentially non-identity-based stochstic SSAs. For this reason, the following subsections will present some results generalising a treatment of non-identity-based stochastic SSAs, which has been demonstrated repeatedly in Section 3 for the empirical CDF-SSA (Example 3.3): Maps of its vocabulary set V into the vocabulary set V' of an identity-based stochastic SSA like Interval-SSAs (Lemma 3.7) or the analytical CDF-SSA (Example 3.20).

It has already been mentioned that SSA-Axiom (vii) generalises the definition of the relation H in the empirical CDF-SSA (Example 3.3.(v)). As a consequence, the relation H in stochastic SSAs can always be interpreted in the way demonstrated for the empirical CDF-SSA imediately after Equation (3.9): An ordered pair (S, π) with $S \subseteq V$ and $\pi \in \Pi$ is contained in H, if and only if the element ϕ_U of the vocabulary set V is almost surely contained in S for an Ω_0-valued random variable U with distribution π. So the general format of propositions claiming stability under stochastic aggregation for a set S can be stated as follows: Fore every Ω_0-valued random variable U with distribution $\pi \in \Pi$ where ϕ_U is almost surely contained in S, the respective element Φ_π of the vocabulary set is also contained in S. Under a sufficiently general understanding of the concept of probability distributions, where relative frequencies in sample distributions are understood as probabilities, the above propositional format covers many kinds of aggregation in psychology.

For an identity-based stochastic SSA, the interpretation of the relation H and the format of propositions claiming stability under stochastic aggregation becomes simpler: Then the condition that the element ϕ_U of the vocabulary set V is almost surely contained in S can be reduced to the requirement that a random variable with distribution π is almost surely contained in S.

For subsets S of the vocabulary set, where the set $\phi^{-1}(S)$ is contained in the α-algebra A_0, the relation H has another simple interpretation, which is stated in the following lemma.[134]

Lemma 4.2: In a stochastic SSA (V, Π, Φ, H, T) with a basis $(\Omega_0, \mathsf{A}_0, \phi)$, let S be a subset of V such that $\phi^{-1}(S) \in \mathsf{A}_0$. Then the properties $(S, \pi) \in H$ and $\pi(\phi^{-1}(S)) = 1$ are equivalent for every $\pi \in \Pi$.

Another rationale of the definition of the relation H in SSA-Axiom (vii) can be derived from SSA-Axiom (iii): If A is an element of A_0 with $\pi(A) = 1$, then the relation $(\phi(A), \pi) \in H$ is intuitively meaningful, and then $(S, \pi) \in H$ follows from SSA-Axiom (iii) for every subset S of the vocabulary set V with $\phi(A) \subseteq S$. So the relation H specified by SSA-Axiom (vii) is the smallest subset of $\mathsf{P}V \times \Pi$ conforming with SSA-Axiom (iii) and with the property $(\phi(A), \pi) \in H$ for

[134] Proof: If $(S, \pi) \in H$, then SSA-Axiom (vii) grants the existence of an element A of A_0 with $\pi(A) = 1$ and $\phi(A) \subseteq S$, and $\pi(\phi^{-1}(S)) = 1$ follows immediately. Converseley, if $\pi(\phi^{-1}(S)) = 1$, then the set $\phi^{-1}(S)$ can take the role of the set A in SSA-Axiom (vii).

$\pi(A) = 1$.

We will now come back upon some properties of the relation H in SSAs, which have been mentioned in Comment h to the general definition of SSAs (Definition 1.2) as plausible, but not axiomatically constitutive assumptions. First, the said comment referred to situations with a family $\{S_i\}_{i\in I}$ of subsets of the vocabulary set V and an element π of Π such that $(S_i, \pi) \in H$ for every $i \in I$. It was questioned whether the property $(\bigcap_{i\in I} S_i, \pi) \in H$ shouldn't be stipulated for such situations by an additional SSA-Axiom. However, an axiom of this kind would be incompatible with SSA-Axiom (vii) for many stochastic SSAs. Take e.g. the Interval-SSA $(V', \Pi', \Phi', H', T')$ of Lemma 3.7 with $V' := [0, 1]$, let π be the rectangular distribution over the interval [0, 1], and define $I := [0.5, 1]$ and $S_i := [0, 1]\backslash\{i\}$ for every $i \in I$. Then it is easily verified that $(S_i, \pi) \in H'$ holds for every $i \in I$. However, the intersection of the sets S_i is the interval [0, 0.5[, and a random variable with distribution π isn't alsmost surely contained in this interval. So the ordered pair $(\bigcap_{i\in I} S_i, \pi)$ is not contained in H'.

But the following lemma shows that the matter is different for a finite or countable family $\{S_i\}_{i\in I}$ of subsets of V - i.e., a family with a finite or countable index set I.

Lemma 4.3: In a stochstic SSA (V, Π, Φ, H, T), let $\{S_i\}_{i\in I}$ a family of subsets of the vocabulary set V with a finite or countable index set I. Furthermore, define $S := \bigcap_{i\in I} S_i$, and let π be an element of the set Π such that $(S_i, \pi) \in H$ for every $i \in I$.
Then $(S, \pi) \in H$.

The lemma (which is proved in Section 6.21) can particularly be used to establish the validity of premissas in Intersection-Theorem 3.14 (Implication (3.52)[135]) and in Conditioning-Theorem 3.18 (Implication (3.56)).

An immediate consequence of Lemma 4.3 is the following corollary. It validates for stochastic SSAs another conjecture, which has been mentioned as plausible, but not axiomatically constitutive for SSAs in Comment h to Definition 1.2: If S' and S'' are disjoint subsets of the vocabulary set of an SSA, then an aggregate cannot simultaneously be sheerly composed of elements of S' and sheerly composed of elements of S''.

Corollary 4.4: In a stochastic SSA (V, Π, Φ, H, T), let S' and S'' be disjoint subsets of the vocabulary set V, and π an element of the set Π. Then the properties $(S', \pi) \in H$ and $(S'', \pi) \in H$ are mutually exclusive.

Indeed, since the intersection of disjoint sets S' and S'' is the empty set, the simultaneous presence of the properties $(S', \pi) \in H$ and $(S'', \pi) \in H$ would lead to $(\varnothing, \pi) \in H$ by Lemma 4.3,[136] which is impossible by SSA-Axiom (i).

Observe also that Definition 4.1 doesn't explicitly require the existence of a unique map $\phi:\Omega_0 \to V$

[135] Note, however, that Lemma 4.3 verifies Implication (3.52) only in situations with a finite or countable index set I, whereas the Intersection-Theorem refers to arbitrary index sets.

[136] For this conclusion, let the family $\{S_i\}_{i\in I}$ in Lemma 4.3 consist of the sets S' and S''.

and a unique measurable space (Ω_0, A_0) with the properties specified by the SSA-axioms, but only the existence of a (i.e., at least one stochastic) map and a measurable space. However, the measurable space (Ω_0, A_0) is implicitly determined by SSA-Axiom (v). Since Π must be a set of probability measures on A_0, and since every probability measure is a probability measure on one and only one σ-algebra, the second component of the stochastic base is completely determined by the elements of the set Π, and then the definition of σ-algebras implies that Ω_0 is the largest set contained in A_0. So the following lemma doesn't need a formal proof.

Lemma 4.5: If (Ω_0, A_0) and (Ω'_0, A'_0) are measurable spaces underlying the same SSA, then $\Omega'_0 = \Omega_0$, and $\mathsf{A}'_0 = \mathsf{A}_0$.

But for the the map $\phi:\Omega_0 \to V$, which makes an SSA ϕ-stochastic, the matter is different. The following lemma, which is proved in Section 6.22, states a condition, which allows to replace the map ϕ in a stochastic base (Ω_0, A_0, ϕ) by another map ϕ'.

Lemma 4.6: For a ϕ-stochastic SSA (V, Π, Φ, H, T), let V' be a subset of V such that (V', π) $\in$ H for every $\pi \in \Pi$. Furthermore, let $\phi':\Omega_0 \to V$ be a map with $\phi'(\omega) = \phi(\omega)$ for every $\omega \in \phi^{-1}(V')$. Then the ordered triple (Ω_0, A_0, ϕ') is another stochastic base of the considered SSA.

In particular, the property stated by this lemma is relevant for restrictions of the basic sets V and Π of SSAs, which have turned out to be useful tools in the analysis of stability under aggregation. If V' and Π' are subsets of the sets V and Π in a stochastic SSA (V, Π, Φ, H, T) with $\Phi_\pi \in V'$ and (V', π) $\in$ H for every $\pi \in \Pi'$, then the V'-Π'-restriction of the considered SSA is well defined (Definition 3.5). But it may be questioned whether the SSA resulting from this restriction is a stochastic SSA, i.e., whether there exists a stochastic base of the SSA. There are no problems, if a stochastic base (Ω_0, A_0, ϕ) of the original SSA is given such that $\phi(\Omega_0)$ is a subset of V': Then ϕ can be reconceived as a map $\phi:\Omega_0 \to V'$, and it is easily verified that the ordered triple (Ω_0, A_0, ϕ) with this reinterpretation of the map ϕ is a stochastic base of the V'-Π'- restriction. But if the set $\phi(\Omega_0)$ is not a subset of V', then a stochastic SSA with vocabulary set V' cannot be based on the map ϕ.

Recall, however, that a V'-Π'-restriction is defined only in situations with (V', π) $\in$ S and $\Phi_\pi \in V'$ for every $\pi \in \Pi'$. So let $\phi':\Omega_0 \to V'$ be a map with $\phi'(\omega) = \phi(\omega)$ for $\omega \in \phi^{-1}(V')$, and arbitrary elements of V' as function values $\phi'(\omega)$ for $\omega \notin \phi^{-1}(V')$. Then Lemma 4.6 implies that the ordered triple (Ω_0, A_0, ϕ') is a stochastic base of the Π'-restriction of the considered SSA.[137] The same holds for the V'-restriction of the Π'-restriction, which is identical with the V'-Π'-restriction of the original SSA (see Corollary 3.6). So the following corollary doesn't need a formal proof:

Corollary 4.7: For a stochastich SSA (V, Π, Φ, H, T), let V' and Π' be subsets of V resp. Π such that $\Phi_\pi \in V'$ and (V', π) $\in$ H for every $\pi \in \Pi'$. Then the V'-Π'-restriction of the considered SSA is a stochastic SSA.

Of course, the transition to a new map $\phi':\Omega_0 \to V$ requires careful consideration of consequences.

---

[137] Exercise!

For instance, in the emprical CDF-SSA (Example 3.3), let V' be identical with the set S_O of all elements v of V with $v(a, w) \geq v(b, w)$ for every $w \in W$, and let Π' be the set of all elements π of Π with $(V', \pi) \in H$ and $\Phi_\pi \in V'$. Now suppose that there are elements u of the set D with $\phi_u \notin V'$. Then the original triple (D, A_D, ϕ) is not a stochastic V'-Π'-restriction of the SSA. If we replace the original map ϕ:D→V by another map ϕ':D→V in the above outlined way, then ϕ'_u is no longer the element of the vocabulary set characterising unit u, unless $\phi_u \in V'$. However, a more coutious interpretation can be derived from the definition of Π': For every D-valued random variable U with distribution $\pi \in \Pi'$, ϕ'_U is π-almost-surely identical with the element of the vocabulary set characterising the selected unit U; i.e., $\phi'_U =_{\pi\text{-a.s.}} \phi_U$. Under a condition, which will be discussed immediately, this implies the equation

$$\mathsf{E}_{U\sim\pi}\, \phi'_U(c, w) = \mathsf{E}_{U\sim\pi}\, \phi_U(c, w) = \Phi_\pi(c, w) = \Phi'_\pi(c, w) \tag{4.3}$$

for every $(c, w) \in (C{\times}W)$ and every $\pi \in \Pi'$, the second equality being the aggregation rule of the original SSA (Example 3.3.(iv)), and the last Equality following from the definition of a V'-Π'-restriction (Definition 3.5).

Of course, the map $pr_{cw} \circ \phi'$ must be A_D-B-measurable for the first equality in Equation (4.3), and this requirement reduces the arbitrariness in the definition of the map ϕ'. It can be shown that the following specification for every $u \in D$ and every $w \in W$ is sufficient:[138]

- If $\phi_u(a, w) \leq \phi_u(b, w)$, then $\phi'_u(a, w) := \phi_u(a, w)$ and $\phi'_u(b, w) := \phi_u(b, w)$.
- If $\phi_u(a, w) > \phi_u(b, w)$, then $\phi'_u(a, w) := \phi_u(b, w)$ and $\phi'_u(b, w) := \phi_u(a, w)$.

The A_0-B-measurability of the maps $pr_{cw} \circ \phi'$, which depends upon the choice of a suitable map $\phi':\Omega_0 \to V'$, is an instance of certain properties, which will subsequently be called base-dependent. (A formal explication of this concept will follow in Definition 4.8.) Such properties raise questions of the range of identifiers like ϕ: How long can we assume that the identifier ϕ refers to the same map? In formalised languages (like predicate calculus or computer programming languages), there are precise rules giving an answer to such questions. The well known ambiguities of current language can be diminished by the following convention: A stochastic SSA can be introduced as $(V, \Pi, \Phi, H, T)_\phi$, which has two consequences: First, it is assumed that the SSA is ϕ-stochastic. Furthermore, one and the same map ϕ is declared to be referenced (explicitly or implicitly) in the further analysis of the SSA.[139]

---

[138] Exercise! Note: It is almost trivial to show that the inequality $\phi'_u(a, w) \leq \phi'_u(b, w)$ for every $u \in D$ and every $w \in W$ results from the definition of ϕ'. But admittedly, it requires more endeavour to show that the maps $pr_{cw} \circ \phi'$ are A_0-B-measurable and that every ϕ'_u is an element of the set V, i.e., a map $v:C{\times}W \to \mathbb{R}$ with the properties specified by Inequality (3.5) and Equation (3.6).

[139] It could also be considered to introduce stochastic SSAs as ordered sextuples, whose first five elements form an SSA, whereas the sixth element would be our map ϕ. But it is good practice to maintain the number and the order of components of n-tuples denoting a class of relational structures. Formally correct, but awkward would be a definition as an ordered pair (S, ϕ), where S is an SSA, and ϕ a map with properties specified by suitable axioms.

Definition 4.8: In a stochastic SSA (V, Π, Φ, H, T), which is based on a measurable space (Ω_0, $\mathcal{A}_0$), a property is base-dependent iff there are maps $\phi:\Omega_0 \to V$ and $\phi':\Omega_0 \to V$ such that the SSA is ϕ-stochastic and ϕ'-stochastic and that the property holds for only one of the two maps.
A stochastic SSA $(V, \Pi, \Phi, H, T)_\phi$ is a ϕ-stochastic SSA, where one and only one map ϕ is referenced (explicitly or implicitly[140]) in the role of the equally named map in Definition 4.1 and all derived concepts and results.

A map ϕ, which has been introduced in this way, will subsequently be called the *prominent* map underlying the considered SSA.

Section 4.2
4.2 Deriving Statistical Predictions via Stochastic Structures of Aggregation Stability

###
Alte Textbausteine zu $\phi(D) \subseteq S$:

Now assume that we want to derive a testable prediction from a research hypothesis referring to a subset S of the common vocabulary set of the two CDF-SSAs, the claim of the hypothesis being $\phi(D) \subseteq S$; i.e.: For very element u of the domain set D, the respective element ϕ_u of the vocabulary set V is contained in S.[141] Of course, single case studies are the method of choice if the claim $\phi_u \in S$ can be tested for a single unit u. But if tests of this type are impossible, it could be considered to come back upon the approach outlined in Example 2.17, which consists in a transformation of a hypothesis referring to individuals into a hypothesis about RSO-processes. Indeed, a membership of the set S in the set system T would allow the following consideration: If the hypothesis $\phi(D) \subseteq S$ is true, then every ordered pair (S, π) with $\pi \in \Pi$ will be contained in the relation H, since the set D can take the role of the set A in SSA-Axiom (vii). So the element Φ_π of the vocabulary set characterising an RSO-process with selection distribution π will be contained in S for every $\pi \in \Pi$, if S is contained in T.

A generalisation of these results refers to stochastic SSAs (V, Π, Φ, H, T), where the measurable space (Ω_0, $\mathcal{A}_0$) in Definition 4.1 is (D, $\mathcal{A}_D$) with a set D being a domain of empirical units as in the empirical CDF-SSAs. In such stochastic SSAs, a hypothesis of the type $\phi(D) \subseteq S$ with $S \subseteq V$ implies that ϕ_U is almost surely contained in S for every random variable U on $\mathcal{A}_D$, and this means $(S, \pi) \in H$ for every $\pi \in \Pi$.

---

[140] An example of an indirect reference to the map ϕ in a stochastic SSA $(V, \Pi, \Phi, H, T)_\phi$ will be discussed after Definition 4.13.

[141] For an example, reconsider the research hypothesis of Example 2.17, which claimed that the true score in a dependent variable Y is less under condition a than under condition b for every unit u of the domain set D. In the empirical CDF-SSA, this hypothesis can be written as $\phi(D) \subseteq S$, where S is the set of those elements v of the vocabulary set V, where the expectation of a random variable with a CDF given by the partial map v(a, .) is smaller than the one derived from v(b, .).

Auch: Ableitung von Hypothesen aus Π'-restriction (vorbereitet in Section 3.4).
aus V'-restriction bei Hilfs-Annahmen. Beispiel: V=Menge aller W-Maße auf B. V' := Menge aller W-Maße auf B mit endlichem Erwartungswert. Hilfsannahme: $\phi(D) \subseteq V'$, $\Phi(\Pi) \subseteq V'$. => T = T', wenn (V', Π', Φ', H', T') die V'-restriction von (V, Π, Φ, H, T) ist.

Auch: raffiniertere Hypothesen, z.B. Kognitionspsychologie: Erdfelder?? s. Dresdner Kongreß!

'Derive empirical predictions' etc..
Auch: single case studies are the method of choice.....
Stichprobenmodell??? oder schon unter conv lincomb:
A_i=Menge der units, bei denen Person i beteiligt ist. Ist $S \in T$ und $\phi(A_i) \subseteq S$, dann $\Phi_\pi \in S$ für alle π mit $\pi(A_i) = 1$.
Ist aber nur Spezialfall von $\phi(D) \subseteq S$: A_i übernimmt die Rolle von D.
Auch: Begriff 'empirical map ϕ'; wird verwendet in Kommentar zu Corollary 4.9.
###

4.3

4.3 Maps between the Vocabulary Sets of Stochastic Structures of Aggregation Stability

In this subsection, we will point out some consequences of the specification of the relation H by SSA-Axiom (vii), which provide interfaces to Mapping-Theorem 3.8 and the corollaries derived from that theorem. In particular, these interfaces can be very useful to derive properties of a stochastic SSA under study from those of a better known one. Some premissas of the Mapping-Theorem and its corollaries refer to implications between the relations H and H' of SSAs (V, Π, Φ, H, T) and (V', Π', Φ', H', T'), whose vocabulary sets are the domains of maps $g:V \to V'$. A general statement of results, which can be derived from SSA-Axiom (vii), requires a rather complicated set theoretical structure. So these results are stated by Lemma 6.4 in Section 6.24, and applications to situations with additional assumptions will subsequently be presented as corollaries.

The first one may be exemplified by an alternative approach to the aggregation stability of the sets S'_w, S''_w and S^*_w (with $w \in W$) in the empirical CDF-SSA of Example 3.3, whose definition by Equations (3.10), (3.11) and (3.12) could be rewritten as $S'_w := g^{-1}([0, 1])$, $S''_w := g^{-1}(]0, 1])$, and $S^*_w := g^{-1}(\{0\})$ with a map $g:V \to [-1, 1]$ given by $g(v) := v(a, w) - v(b, w)$. In the demonstration of Mapping-Theorem 3.8 in Section 3.5, the interval $[-1, 1]$ was considered as the vocabulary set V' of an Interval-SSA (V', Π', Φ', H', T') in the understanding of Lemma 3.7, i.e., of an identity-based stochastic SSA. Indeed, maps into the vocabulary set of identity-based stochastic SSAs will turn out to be a useful tool for the analysis of non-identity-based stochastic SSAs, and this approach will be generalised in Corollary 4.10; but beforehand, recall that the application of Interval-SSAs looked like a roundabout way in Section 3.5. It would have been more direct to characterise every unit u of the domain set D by a real number ϕ'_u defined as $\phi'_u := g(\phi_u)$ (or more explicitly as the difference $\phi'_u := \phi_u(a, w) - \phi_u(b, w)$). This approach can be formalised as another, very simple stochastic SSA, which is hased on the same measurable space (D, A_D) and a map $\phi':D \to V'$ with $V' := [-1, 1]$. The following corollary (which is proved in Section 6.23) generalises this situation.

Corollary 4.9: For stochastic SSAs (V, Π, Φ, H, T) and $(V', \Pi', \Phi', H', T')$ with stochastic bases $(\Omega_0, \mathsf{A}_0, \phi)$ and $(\Omega_0, \mathsf{A}_0, \phi')$, asumme that the sets Π and Π' are identical, and let $g:V\to V'$ be a map such that $\phi' = g \circ \phi$.
Then the equivalence

$$(g^{-1}(S'), \pi) \in H \;\Leftrightarrow\; (S', \pi) \in H' \tag{4.4}$$

holds for every subset S' of the set V' and every $\pi \in \Pi$.
Now assume that the equality $\Phi'_\pi = g(\Phi_\pi)$ holds for every $\pi \in \Pi$.
Then the equivalences $g^{-1}(S') \in T \Leftrightarrow S' \in T'$ and $g^{-1}(S') \in T_e \Leftrightarrow S' \in T'_e$ hold for every subset S' of the set V'. In particular, the map g is an SSA-isomorphism of the SSA (V, Π, Φ, H, T) onto the SSA $(V', \Pi', \Phi', H', T')$, if it is bijective.

An aggregation rule $\Phi':\Pi'\to V'$, which is consistent with the assumptions of the corollary and with the aggregation rule Φ of the empirical CDF-SSA (i.e., with Equation (3.9)) is $\Phi'_\pi := \mathsf{E}_{U\sim\pi}\, \phi'_U$. Under this aggregation rule, it is easily verified that every subset S' of V', which is an interval of real numbers, is contained in T', since $(S', \pi) \in H'$ implies that ϕ'_U is almost surely contained in S' for a random variable U with distribution π. So Corollary 4.9 can be used to obtain $g^{-1}(S') \in T$.

The same application can also be used to demonstrate a limitation of the corollary. Whereas the property of being a subinterval of V' is necessary and sufficient for the membership in T' in an Interval-SSA, it is sufficient, but not necessary in the above application of Corollary 4.9.[142] More generally, the map $\phi':\Omega_0\to V'$ of the corollary inherits all uncertainties about the map $\phi:\Omega_0\to V$ in situations, where ϕ is an empirical map in the understanding of Section 4.2. So the corollary is mainly designed for situations, where the property $S' \in T'$ is derived easily from the mathematical structure of an aggregation rule Φ' like $\Phi'_\pi = \mathsf{E}_{U\sim\pi}\, \phi'_U$ in the above applications. A class of more complex stochastic SSAs, which can take the role of the SSA $(V', \Pi', \Phi', H', T')$ in the corollary, will be introduced in Section 4.7.

The outlined limitations of Corollary 4.9 motivate the consideration of maps $g:V\to V'$ into the vocabulary set of a stochastic SSA $(V', \Pi', \Phi', H', T')$, where the entire set system T' can be derived analytically from the definition of the SSA as in the Interval SSAs. Since many such SSAs are identity based, the following corollary considers maps $g:V\to V'$, where V' is the vocabulary set of an identity-based SSA. For a first approach to the corollary, recall the applications of Mapping-Theorem 3.8 in Section 3.5, where the SSA $(V', \Pi', \Phi', H', T')$ was the Interval-SSA (Lemma 3.7) with $V' = [-1, 1]$ and with the same map $g:V\to V'$ as above.

Corollary 4.10: Let (V, Π, Φ, H, T) and $(V', \Pi', \Phi', H', T')$ be stochastic SSAs, the first one being based on a measurable space (Ω_0, A_0) and a map $\phi:\Omega_0\to V$, and the second one being identity-based with a σ-algebra $\mathsf{A}_{V'}$ in V'. Furthermore, let a map $g:V\to V'$ be given such that the map $g \circ \phi$ is A_0-$\mathsf{A}_{V'}$-measurable.
Then the implication (i)$\Rightarrow$(ii) holds for the following assertions:

---

[142] For instance, in the situation assumed in Example 3.19, the set S' given by the definition $S' := [0, 0.3] \cup [0.7, 1]$ is not contained in the set system T' of an Interval-SSA; but it is contained in the set system T' of an SSA $(V', \Pi', \Phi', H', T')$ constructed by Corollary 4.9. (Apply the equivalence $g^{-1}(S') \in T \;\Leftrightarrow\; S' \in T'$, which is claimed by the corollary, and recall that the property $g^{-1}(S') \in T$ has been established in Example 3.19.)

(i) The σ-algebras A_0 and A'_0 have the properties

$$\mathsf{A}_0 = \{\phi^{-1}(g^{-1}(A')) \colon A' \in \mathsf{A}_{V'}\}, \tag{4.5}$$

and

$$g(\phi(\Omega_0)) \in \mathsf{A}_{V'}. \tag{4.6}$$

(ii) For every $A \in \mathsf{A}_0$, the set $g(\phi(A))$ is an element of $\mathsf{A}_{V'}$.

Furthermore, if Equation (4.6) is given and the equation $A = \phi^{-1}(g^{-1}(g(\phi(A))))$ holds for every $A \in \mathsf{A}_0$, then Assertions (i) and (ii) are equivalent.

Now let a map $f{:}\Pi \to \Pi'$ be given such that the equation

$$f(\pi)(A') = \pi(\phi^{-1}(g^{-1}(A'))) \tag{4.7}$$

holds for every $\pi \in \Pi$ and every $A' \in \mathsf{A}_{V'}$.

Then the following properties follow:

(iii) The implication

$$(S', f(\pi)) \in H' \;\Rightarrow\; (g^{-1}(S'), \pi) \in H \tag{4.8}$$

holds for every subset S' of the set V' and every $\pi \in \Pi$.

(iv) The reversal of Implication (4.8) holds for every $\pi \in \Pi$ and every subset S' of V' which is an element of $\mathsf{A}_{V'}$.

(v) If either Assertion (i) or Assertion (ii) holds for the σ-algebras A_0 and $\mathsf{A}_{V'}$, then the reversal of Implication (4.8) holds for every $\pi \in \Pi$ and every subset S' of the set V'.

Although Assertions (i) and (ii) are mainly used as premissas in Assertion (v), they are introduced before the definition of the map $f{:}\Pi \to \Pi'$ (Equation (4.7)) to make clear that their logical relations (the implication (i)⇒(ii) and the conditional equivalence (i)⇔(ii)) do not need the map f, but only the measurablity of the map $g \circ \phi$. The relevance of these premissas of Assertion (v) will be demonstrated by Example 4.12 after Corollary 4.11. - The interpretation of the map $f{:}\Pi \to \Pi'$ in Corollary 4.10 is the same as in the applications of the Mapping-Theorem and its corollaries to the empirical CDF-SSA in Subsection 3.5: If U is a random variable with values in Ω_0 and distribution π, then Equation (4.7) specifies the distribution $f(\pi)$ of a random variable $g(\phi_U)$ with values in V'. Note that the assumed existence of a map $f{:}\Pi \to \Pi'$ with this property implies a premissa about the set Π' in the second SSA: It must contain as its elements all probality measures $f(\pi)$ on $\mathsf{A}_{V'}$ resulting from an application of Equation (4.7) to an element π of the set Π.

As mentioned beforehand, Corollary 4.10 is designed to establish premissas of Mapping-Theorem 3.8 and its corollaries; but a limitation should be noted. To derive $S \in T$ from $S' \in T'$ by Assertion (i) of that theorem, we need the reversal of Implication (4.8), and this reversal is claimed only under additional premissas in Corollary 4.10.(iv) and (v). Indeed, in the applications of Mapping-Theorem 3.8 to the empirical CDF-SSA and to Interval-SSAs in Section 3.5, the set S' was always a subinterval of V' (with V' = [−1, 1]) and hence contained in the σ-algebra $\mathsf{A}_{V'}$ of the respective Interval-SSA. So the reversal of Implication (4.8) can be derived from Corollary 4.10.(iv).

More generally, if we want to use Mapping-Theorem 3.8 do derive the aggregation stability (or its absence) for subsets of V from a better known SSA (V', Π', Φ', H', T'), then the claim of Corollary 4.10.(iv) suffices for subsets of V, which can be conceived as inverse images $g^{-1}(S')$ of elements S' of $\mathsf{A}_{V'}$. The extension to all subsets S' of V' in Assertion (v) is mainly used, if Implication (4.8) and its reversal are needed to make a bijective map $g{:}V \to V'$ an SSA-isomorphism. Since linitations resulting from these premissas are easier to follow up in the special situation of the following corollary, their discussion is postponed.

The additional assumption of Corollary 4.11 has already been demonstrated in Section 3.9 for the empirical and the analytical CDF-SSA (Examples 3.3 and 3.20): The vocabulary sets V and V' of the two considered SSAs were identical, and the map $g:V \to V'$ was the identity map in V. An application of Corollary 4.10 to such situations leads to the following result, which is proved in Section 6.25:

Corollary 4.11: In the situation of Corollary 4.10, assume that the vocabulary sets V and V' of the two considered SSAs are identical, and that the map $g:V \to V'$ is the identity map in V.
Then the specification of a map $f:\Pi \to \Pi'$ by Equation (4.7) can be rewritten as
$$f(\pi)(A') = \pi(\phi^{-1}(A')) \tag{4.9}$$
for every $A' \in \mathsf{A}_{V'}$.
Furthermore, define $\Pi^{\sim} := f(\Pi)$, and assume that the equation
$$\Phi'(f(\pi)) = \Phi(\pi) \tag{4.10}$$
holds for every $\pi \in \Pi$
Then the subsequent properties follow:
(i) The implication $S \in T' \Rightarrow S \in T$ holds for every subset S of the common vocabulary set which is an element of the σ-algebra $\mathsf{A}_{V'}$.
(ii) If the σ-algebra $\mathsf{A}_{V'}$ contains all sets $\phi(A)$ with $A \in \mathsf{A}_0$, then the map $g:V \to V'$ is an SSA-isomorphism of the SSA (V, Π, Φ, H, T) onto the $\Pi^{\sim}$-restriction of the SSA (V', Π', Φ', H', T').
(iii) If the map f is surjective, then $T \subseteq T'$.

Corollary 4.11 is designed to support the transition from a non-identity-based stochastic SSA to an identity-based one with the same vocabulary set. The premissas expressed by Equations (4.9) and (4.10) can again be demonstrated by an application to the two CDF-SSAs (Examples 3.3 and 3.20), the measurable space (Ω_0, A_0) of Corollary 4.11 being denoted as (D, A_D) in the empirical CDF-SSA. For every distribution π of a D-valued random variable U, Equation (4.9) tells that f(π) is the distribution of the random variable ϕ_U, and Equation (4.10) assumes that the elements of the common vocabulary set assigned to both distributions by the respective aggregation rules are identical. Indeed, this property follows from the definition of the two aggregation rules by Equations (3.9) and (3.63); but Equation (4.10) would also hold if both aggregation rules would be based e.g. on medians instead of expectations.[143] Nevertheless, Equation (4.10) implies a limitation for applications of the corollary: For elements π_1 and π_2 of the set Π, where Equation (4.9) leads to identical elements $f(\pi_1)$ and $f(\pi_2)$ of Π', the function values $\Phi(\pi_1)$ and $\Phi(\pi_2)$ of the aggregation rule in the first SSA must be identical.

Corollary 4.11 should also be taken as an opportunity to resume a problem, which has been postponed in the general discussion of empirical and analytical aggregation stability (Section 3.9). The problem is about the additional premissa referring to the σ-algebras A_0 and $\mathsf{A}_{V'}$. In Corollary 4.11.(ii), it was assumed that $\mathsf{A}_{V'}$ contains all sets $\phi(A)$ with $A \in \mathsf{A}_0$. According to Corollary 4.10, this premissa (which was Assertion (ii) in that corollary) is generally weaker than another one stated by Corollary 4.10.(i). In the simplified situation of Corollary 4.11, this premissa requires that the set

[143] Note, however, that a transition to aggregation rules based on medians would require further assumptions granting the existence of well defined medians such that the maps $\Phi_\pi:C \times W \to \mathbb{R}$ and $\Phi'_{\pi'}:C \times W \to \mathbb{R}$ resulting from these medians are elements of the common vocabulary set. For the aggregation rules based on expectations, this property has been verified in Footnote 81.

$\phi(\Omega_0)$ in contained in the σ-algebra $A_{V'}$ and that A_0 is the coarsest σ-algebra in Ω_0 where the map $\phi:\Omega_0\to V'$ is A_0-$A_{V'}$-measurable. Under this assumption, it was shown in Section 3.9 that the set systems T and T_e of the empirical CDF-SSA include the set systems T' and T'_e of the analytical CDF-SSA, and in this situation, a partition of the set systems T and T_e resulted in a clear distinction of analytical aggregation stability (represented by T' and T'_e) and empirical aggregation stability (T\T' resp. $T_e \backslash T'_e$).

Violations of the premissa underlying this approach have some implications, whose discussion may seem trifling as long as only the conceptualisation of analytical and empirical aggregation stability is considered;[144] but due to some critical consequences for applications, we should take an (admittedly laborious) closer look at such violations.[145]

Since it may be hard to follow up a demonstration of the problems by the two CDF-SSAs, we start with a pair of very simple stochastic SSAs specified by the following example.

Example 4.12: For an application of Corollary 4.11, let (Ω_0, A_0, ϕ) be a basis of a stochastic SSA (V, Π, Φ, H, T) with the following specifications:

(i) The set Ω_0 consists of elements ω_1, ω_2 and ω_3, and the σ-algebra A_0 is the power set of Ω_0.

(ii) The vocabulary set V consists of elements v_1, v_2 and v_3, and the map $\phi:\Omega_0\to V$ is given by $\phi(\omega_i) := v_i$ for i = 1..3.

(iii) The set Π consists of probability measures π_1, π_2 and π_3 on A_0 with

$$\pi_1(\{\omega_1\}) = \pi_1(\{\omega_2\}) = \pi_2(\{\omega_2\}) = \pi_2(\{\omega_3\}) = 0.5,$$
$$\pi_3(\{\omega_3\}) = 1,$$

and

$$\pi_1(\{\omega_3\}) = \pi_2(\{\omega_1\}) = \pi_3(\{\omega_1\}) = \pi_3(\{\omega_2\}) = 0.$$

(iv) The aggregation rule Φ is given by $\Phi(\pi_1) = \Phi(\pi_2) = v_1$, and $\Phi(\pi_3) = v_3$.

As in Corollary 4.11, let a second, identity-based stochastic SSA (V', Π', Φ', H', T') with V = V' be given with the following property:

(v) The σ-algebra $A_{V'}$ consists of the sets ∅, $\{v_1, v_2\}$, $\{v_3\}$ and V', and the set Π' contains only the probability measures $\{f(\pi_i)\}_{i=1..3}$ given by Equation (4.9). Finally, the aggregation rule Φ' is obtained from Equation (4.10).

Then the relations H and H' are completely determined by SSA-Axiom (vii), and the set system T, T_e, T' and T'_e by SSA-Axiom (iv) and Equation (1.1).

In this example, consider the subset S := $\{v_2, v_3\}$ of the vocabulary set V. The membership of S

---

[144] For the mere conceptualisation, a loss of the inclusion $T' \subseteq T$ could be answered by partitioning the power set of V into four subsystems without giving them names: The set system $T \cap T'$ consists of all subsets of V, which are elements of T and T'. Subsets of V which are contained only in one of the set systems are elements of T\T' or T'\T, and the remaining subsets of V, which are contained neither in T nor in T', form the set system $PV \setminus (T \cup T')$. Of course, similar considerations can be applied to the sets systems T_e and T'_e.

[145] Readers not interested in details can skip without loss of continuity to the summarising last two paragraphs of Subsection 4.3.

in the set systems T and T_e is prevented by the combination of the properties (S, π_2) $\in$ H and $\Phi(\pi_2) \notin$ S, which are immediate consequences of Assumptions (iii) and (iv) and of SSA-Axiom (vii). But in the second SSA, we have (S, f(π_2)) $\notin$ H', since the set V' is the only elements A' of $\mathsf{A}_{V'}$ with f(π_2)(A') = 1, and V' isn't a subset of S. But if the ordered pair (S, f(π_2)) is not contained in H', then the property $\Phi'(f(\pi_2)) \notin$ S doesn't prevent the membership of S in the set systems T' and T'_e. Indeed, since f(π_3) is the only element π' of Π' with (S, π') $\in$ H', and $\Phi'(f(\pi_3)) \in$ S follows from Assumption (iv) and Equation (4.10), we have S $\in$ T' by SSA-Axiom (iv), and S $\in$ T'_e can also be derived from Equation (1.1).

A somewhat paradoxical issue of this situation may become clearer, if we reformulate it in the language of almost sure properties without any reference to the framework of stochastic SSAs. So assume first that (Ω_0, A_0, π_2) is the basic probability space governing random events, let U be a random variable on A_0 with distribution π_2, and define a random variable U' on $\mathsf{A}_{V'}$ by U' := ϕ. Then the distribution of U' is the probabiliy measure f(π_2) specified by Equation (4.10), and (V', $\mathsf{A}_{V'}$, f(π_2)) is another probability space. In this situation, the random variable U' is almost surely contained in the set S as long as U' is considered as a random variable in the probability space (Ω_0, A_0, π_2): The set A := {ω_2, ω_3) is an element of A_0 with π_2(A) = 1, and U'(ω) is contained in S for every $\omega \in$ A. But if we consider the probability space (V', $\mathsf{A}_{V'}$, f(π_2)) in itself, then a random variable U' with distribution f(π_2) isn't almost surely contained in S. (Recall that the set V' is the only element A' of $\mathsf{A}_{V'}$ with f(π_2)(A') = 1, and V' contains the element v_1, which is not contained in S.) Indeed, this situation is only an example of a more general one: If (Ω, A, P) and (Ω', A', P') are probability spaces and X:$\Omega \rightarrow \Omega'$ is a measurable map such that P' is the distribution of the random variable X, then two aspects of the almost sure containment of X in a subset S' of Ω' are commonly distinguishied by a reference to the two probability measures P and P': X is P-almost-surely contained in S' iff there exists an element A of the σ-algebra A_0 such that P(A) = 1 and X(A) $\subseteq$ S'. But X is P'-almost-surely contained in S' iff there is an element A' of the σ-algebra A' such that P'(A') = 1 and A' $\subseteq$ S'. With the notations X $\in_{\text{P-a.s}}$ S' and X $\in_{\text{P'-a.s.}}$ S' for both situations, the implication X $\in_{\text{P'-a.s.}}$ S' $\Rightarrow$ X $\in_{\text{P.-a.s}}$ holds always, and the reversed implication holds for S' $\in \mathsf{A}'$.

In Example 4.12, the consequence for aggregation stability can also be formulated without reference to the framework of stochastic SSAs. For this approach, we may consider $\Phi'(\pi')$ with $\pi' \in \Pi'$ as an element of the set V' characterising the distribution π' of the random variable U'. As long as we consider only probability spaces (V', $\mathsf{A}_{V'}$, π') with $\pi' \in \Pi'$, the set S of Example 4.12 is stable under aggregation, since the following implication holds for every $\pi' \in \Pi'$: If a random variable U' on $\mathsf{A}_{V'}$ with distribution π' is π'-almost-surely contained in S, then $\Phi'(\pi')$ is also contained in S. But if U' is considered as a random variable derived from the probability space (Ω_0, A_0, π) with $\pi \in \Pi$, the aggregation stability of the set S would require that the following implication holds for every $\pi \in \Pi$: If π is an element of Π such that the random variable U' is π-almost-surely contained in S, then the element $\Phi'(f(\pi))$ of the vocabulary set V characterising the distribution of U' is also contained in S. But this implication doesn't hold for $\pi = \pi_2$.

A discussion of such subtleties wouldn't be worth its trouble unless it had consequnces for applications. To demonstrate such consequences, we resume the application of Corollary 4.11 to the two CDF-SSAs. In Section 4.2, we have demonstrated the derivation od statistical predictions from hypotheses of the type ϕ(D) $\subseteq$ S, where S is the subset of the vocabulary set representing a property, which is claimed for all elements of a set D. In a situation of this kind, it would be helpful, if the membership of S in the set system T could be derived from the property S $\in$ T', which may be verified for the analytical CDF-SSA. Indeed, this conclusion is obtained from Corollary 4.11.(ii), if

S is an element of the σ-algebra $\mathsf{A}_{V'}$. But otherwise, it would be necessary to know whether there isn't an element π^* of the set Π such that the ordered pair (S, π^*) has the critical properties of the ordered pair (S, π_2) in Example 4.12; i.e.:

- (S, π^*) is contained in the relation H, but $(S, f(\pi^*))$ is not contained in H'.
- The identical function value $\Phi(\pi^*)$ and $\Phi'(f(\pi^*))$ of the two aggregation rules is not contained in S.

If a selection distribution π^* with these properties can exist, then the prediction $\Phi_\pi \in S$ would fail to be valid for $\pi = \pi^*$, even if the set S is contained in T'.

Now the implication $S \in T' \Rightarrow S \in T$ as well as $S \in T'_e \Rightarrow S \in T_e$ can also be verified under the assumption that the σ-algebra $\mathsf{A}_{V'}$ contains all sets $\phi(A)$ with $A \in \mathsf{A}_D$. To verify this premissa, the implication (i)⇒(ii) in Corollary 4.10 may be helpful: If the hypothesis under study implies that the set $\phi(D)$ is contained in $\mathsf{A}_{V'}$ and if we assume that the selection of units is governed by a probability measure on the coarsest σ-algebra in D with an A_D-$\mathsf{A}_{V'}$-measurable map ϕ, then everything is prepared to derive the above implications from a combination of the research hypothesis with Corollary 4.11.(ii) and Restriction-Lemma 3.4.(i).[146] However, research hypotheses allowing the conclusion that $\phi(D)$ is contained in $\mathsf{A}_{V'}$ will be rare exceptions. In general, it will be hard to derive $S \in T$ from $S \in T'$, unless the set S is contained in $\mathsf{A}_{V'}$.

More generally, a hypothesis of the type $\phi(D) \subseteq S$ with $S \subseteq V$ implies that ϕ_U is almost surely contained in S for every random variable U on A_D, and this means $(S, \pi) \in H$ for every $\pi \in \Pi$. If a stochastic SSA of this kind is mapped into an identity based one in the way assumed in Corollary 4.11, then the hypothesis $\phi(D) \subseteq S$ implies that a random variable $U' := \phi(U)$ is almost surely contained S, as long as this almost sure containment refers to a probability space (D, A_D, π) with $\pi \in \Pi$, but not necessary for the derived probability space $(V', \mathsf{A}_{V'}, f(\pi))$, which would mean $(S, f(\pi)) \in H'$. As a consequence, a testable prediction $\Phi_\pi \in S$ can be derived from the hypothesis $\phi(D) \subseteq S$, if the set S is contained in the set system T, whereas $S \in T'$ isn't generally sufficient for this derivation.

Nevertheless, Corollary 4.11 can be useful to derive aggregation stability in the stochastic SSA (V, Π, Φ, H, T) from the (typically more easily analysed) SSA $(V', \Pi', \Phi', H', T'))$. It has to be considered rather an exception than the rule that the SSA-isomorphism claimed by Corollary

---

[146] For an explicit proof of the implications $S \in T' \Rightarrow S \in T$ and $S \in T'_e \Rightarrow S \in T_e$, we have to examine the matching of the set systems in Restriction-Lemma 3.4 with the set systems T, T_e, T' and T'_e of the two CDF-SSAs and the respective set systems in the $\Pi^\sim$-restriction of the analytical CDF-SSA, which will be denoted as $T^\sim$ resp. $T^\sim_e$. Since the $\Pi^\sim$-restriction is applied to the analytical CDF-SSA, the set systems T and T_e of the Restriction-Lemma are the set systems T' and T'_e of the analytical CDF-SSA, and by Equations (3.26) and (3.27), they are identical with the set systems T* and T^*_e of the Restriction-Lemma, since the vocabulary set is unchanged under the $\Pi^\sim$-restriction. The set systems T' and T'_e of the Restriction-Lemma are identical with the set systems $T^\sim$ resp. $T^\sim_e$ in the $\Pi^\sim$-restriction of the analytical CDF-SSA. Finally, $T^\sim$ and $T^\sim_e$ are identical with the set systems T and T_e of the empirical CDF-SSA, if the identity map $g:V \rightarrow V'$ is an SSA-isomorphism of the empirical CDF-SSA onto the $\Pi^\sim$-restriction of the analytical CDF-SSA. But this property follows from Corollary 4.11.(ii), if the σ-algebra $\mathsf{A}_{V'}$ contains all sets $\phi(A)$ with $A \in \mathsf{A}_D$. - In summary, the inclusions $T^* \subseteq T'$ and $T^*_e \subseteq T'_e$ in Restriction-Lemma 3.4.(i) have to be written as $T' \subseteq T$ and $T'_e \subseteq T_e$ in the notation of the two CDF-SSAs.

4.11.(ii) can be used to derive empirically testable hypotheses. But results to be presented in later subsections imply that the set S in many research hypotheses of the type $\phi(D) \subseteq S$ is contained in the σ-algebra $\mathsf{A}_{V'}$ of a suitable identity-based stochastic SSA (V', Π', Φ', H', T') like the analytical CDF-SSA.[147] In such situations, Corollary 4.11.(i) allows to derive $S \in T$ from $S \in T'$. If the set S is not contained in $\mathsf{A}_{V'}$, we can capitalise upon a certain flexibility in the choice of a σ-algebra in the vocabulary set. For instance, the σ-algebra $\mathsf{A}_{V'}$ in the analytical CDF-SSA can be replaced by a finer one containing a set S of interest or an entire class of sets.[148] Although the transition to a finer σ-algebra in the vocabulary set of the SSA (V', Π', Φ', H', T') in Corollary 4.11 will usually lead to changes in all compenents of the SSA with the exception of the vocabulary set V', it will be easier in many cases to analyse aggregation stability in this modification of the SSA (V', Π', Φ', H', T') than in the SSA (V, Π, Φ, H, T), which is based on a typically unknown map $\phi:\Omega_0 \rightarrow V$.[149]

---

[147] For the analytical CDF-SSA, the membership in $\mathsf{A}_{V'}$ can be derived for the sets defined by Equations (3.10) through (3.16) from the assumed measurability of the projection maps $pr_q:V' \rightarrow \mathbb{R}$ and from Equations (3.18) through (3.20)

[148] A limitation for the transition to a finer σ-algebra $\mathsf{A}_{V'}$ containing a set S' of interest is given by the assumed A_0-$\mathsf{A}_{V'}$-measurability of the map ϕ: The σ-algebra A_0 must contain the set $\phi^{-1}(S')$, and this means that every probability measure on A_0 must specify a probability of obtaining an element ω of Ω_0, where $\phi(\omega)$ has the property of interest, which is represented by membership in the set S'. Indeed, this is a rather weak assumption, and in the derivation of predictions from the hypothesis $\phi(\Omega_0) \subseteq S'$, it is granted by the property $\phi^{-1}(S') = \Omega_0$, which is an immediate consequence of the hypothesis. (Ω_0 is contained in every σ-algebra in Ω_0!).

[149] Note, however, that Corollary 4.11 can also be used to study changes as well as invariances under the transition to a finer σ-algebra in the vocabulary set. A premissa of Corollary 4.10, which requires that the map $g \circ \phi$ in that corollary must be measurable, is fulfilled iff the SSA with the finer σ-algebra is the SSA (V, Π, Φ, H, T) of Corollary 4.11. So let A_V be a σ-algebra in the identical vocabulary set, which is strictly finer than $\mathsf{A}_{V'}$, and assume that the premissas of Corollary 4.11 expressed by Equations (4.9) and (4.10) hold. Then a formal change in the set Π is obvious, since probability measures on A_V are not probability measures on $\mathsf{A}_{V'}$; but Equation (4.9) requires for every probability measure π on A_V that f(π) is its restriction to $\mathsf{A}_{V'}$. For the subsequent considerations, assume that Π is the set of all probability measures on A_V, whose restriction to $\mathsf{A}_{V'}$ is contained in Π'. Furthermore, assume that the map $f:\Pi \rightarrow \Pi'$ is surjective, which means that every probability measure π' on $\mathsf{A}_{V'}$, which is contained in Π', has an extension to a probability measure on A_V. Then the main change in the aggregation rule Φ has to be seen in the fact that it is a map $\Pi \rightarrow V$ and not $\Pi' \rightarrow V'$; but according to Equation (4.10), the result Φ(π) of aggregation rule Φ must be identical with the result Φ'(f(π)), which is assigned by aggregation rule Φ' to the restriction of π to $\mathsf{A}_{V'}$. For the relations H and H', Implication (4.8) can be rewritten as $(S, f(\pi)) \in H' \Rightarrow (S, \pi) \in H$ for every subset S of the common vocabulary set, and according to Corollary 4.10.(iv), the reversal of this implication holds, if $S \in \mathsf{A}_{V'}$. But for a subset S of the vocabulary set not contained in $\mathsf{A}_{V'}$, the combination of $(S, \pi) \notin H$ and $(S, f(\pi)) \in H'$ may occur, if there exists an element A of $\mathsf{A}_V \backslash \mathsf{A}_{V'}$ with $A \subseteq S$ and $\pi(A) = 1$, whereas all elements A' of $\mathsf{A}_{V'}$ with $f(\pi)(A') = 1$ are not subsets of S. In a

(continued...)

4.4 Expectational Maps and Projectional Structures of Aggregation Stability

The definition of stochastic SSAs (i.e., Definition 4.1) contains some axioms referring to the components Π and H of a stochastic SSA (V, Π, Φ, H, T), and the set system T is completely determined by SSA-Axiom (iv). But up to this point, there are no specifications of the vocabulary set V and the aggregation rule Φ in stochastic SSAs. The present subsection doesn't add additional axioms, but two useful concepts referring to the aggregation rule and the vocabulary set: Expectational maps, and projectional SSAs.

The first one of these concepts has been prepared by applications of Mapping-Theorem 3.8 and its corollaries to the empirical CDF-SSA (Example 3.3). In these applications, we have used maps $g:V\to\mathbb{R}$, where the equation

$$g(\Phi_\pi) = \mathsf{E}_{U\sim\pi}\, g(\phi_U) \tag{4.11}$$

holds for every element π of the set Π. Examples of such maps are the projection maps $pr_{cw}:V\to\mathbb{R}$ and their linear combinations.[150] Such maps will be called expectational by the following Definition 4.13. In later sections, we will be confronted with situations, where Equation (4.11) doesn't hold for all elements of the set Π, but only for elements of a subset Π^* of Π. In particular, this limitation applies to situations, where the expectation $\mathsf{E}_{U\sim\pi}\, g(\phi_U)$ is infinite or doesn't exist for some elements π of the set Π. To support applications of subsequently derived results to such situations, Definition 4.13.(i) introduces the property of being expectational in Π^*.

Observe also that the expectation $\mathsf{E}_{U\sim\pi}\, g(\phi_U)$ is undefined in situations, where the map $g \circ \phi$ isn't A_0-B-measurable. But again, a weaker property is sufficient for some results to be reported in the following subsections. Therefore, the concept of maps, which are almost expectational in Π^*, is introduced in Definition 4.13.(ii) to support the formulation of weak premissas in later applications. The term χ_S, which is used is the definition of such maps, refers to a *characteristic function*, whose general definition is given in Section ?.?. For the present situation, the following specifications are sufficient: For every subset S of the set V, the map $\chi_S:V\to\mathbb{R}$ is given by $\chi_S(v) := 1$ for $v \in S$ and $\chi_S(v) := 0$ for $v \in V\backslash S$. (To some readers, the function may be known as the indicator function $1_A:\Omega\to\{0, 1\}$, where Ω is the most basic set of a probability space.) Furthermore, the term $\chi_S \cdot g$ refers to a map resulting from pointwise multiplication of the maps χ_S and g. In other words, $\chi_S \cdot g$ can be conceived as a map $g^*:V\to\mathbb{R}$ with $g^*(v) := \chi_S(v) \cdot g(v)$ for very $v \in V$.
(i) (ii) (iii) (ii.a) (ii.b) (ii.c) (ii.d)
Definition 4.13: For a stochastic SSA (V, Π, Φ, H, $T)_\phi$, which is based on a measurable space

---

[149] (...continued)
situation of this kind, π prevents membership of the set S in the set system T, if $\Phi(\pi) \notin S$, whereas S may be contained in T'. (An example can be obtained from Example 4.12 under the additional assumption $\omega_i = v_i$ for i = 1..3.). Finally, the implication $S \in T \Rightarrow S \in T'$ can be obtained from Corollary 4.11.(iii).

[150] See Equation (3.39) for the definition of the projection maps $pr_{cw}:V\to\mathbb{R}$ in the empirical CDF-SSA.

(Ω_0, A_0), let Π^* be a subset of Π. Then expectational and almost expectational maps $V \to \mathbb{R}$ are defined as follows:

(i) A map $g:V \to \mathbb{R}$ is *expectational in* Π^* iff the map $g \circ \phi$ is A_0-B-measurable and Equation (4.11) holds for every $\pi \in \Pi^*$.

(ii) A map $g:V \to \mathbb{R}$ is *almost expectational in* Π^* iff a subset S of the set V with the following properties exists for every element π of the set Π:

(ii.a) $(S, \pi) \in H$.

(ii.b) $\phi^{-1}(S) \in A_0$.

(ii.c) The map $(\chi_S \cdot g) \circ \phi$ is A_0-B-measurable.

(ii.d) $g(\Phi_\pi) = E_{U \sim \pi} (\chi_S(\phi_U) \cdot g(\phi_U))$. (4.12)

(iii) A map $g:V \to \mathbb{R}$ is called '(almost) expectational in the SSA (V, Π, Φ, H, T)' or just '(almost) expectational' iff it is (almost) expectational in Π.

It goes almost without saying that the additional specification 'almost' in Part (iii) of the definition has to be treated consistently: A map $g:V \to \mathbb{R}$ is expectational, iff it is expectational in Π, and it is almost expectational iff it is almost expectational in Π.

Note that the property of being (almost) expectational (in Π^*) is defined only for maps $V \to \mathbb{R}$. An extension of the definition to maps $X \to \mathbb{R}$, where X is a set with $V \subseteq X$, will be given in Definition 4.28.

At first glance, it may seem redundant that the A_0-B-measurability of the map $g \circ \phi$ is explicitly required in Definition 4.13.(i). Without this measurability, Equation (4.11) cannot hold, since the expectation on its right hand side is undefined. But beyond reminding of this implicit premissa, there is a formal reason to mention the measurability requirement: If Π^* is the empty set, then Equation (4.11) holds for every element of Π^*, even if the map $g \circ \phi$ isn't A_0-B-measurable. So the explcit measurability requirement avoids the unwanted consequence that every map $g:V \to \mathbb{R}$ would be expectational in the empty subset of Π.[151]

For subsequent references to this measurability premissa, note that the explicit references to the σ-algebras A_0 and B is redundant. Since Π is a set of probability measures on one and only one σ-algebra, the role of this σ-algebra for map $g \circ \phi$ is implicitly given. Furthermore, it is common practice to consider the Borel sets in $\mathbb{R}$ as the default for for a σ-algebra in $\mathbb{R}$.

It has already been demonstrated that this measurability is a base-dependent property. (See Definition 4.8 and the example before that definition.) The expectational properties explicated by Definition 4.13 are base-dependent for still another reason. There may be maps $\phi:\Omega_0 \to V$ and $\phi':\Omega_0 \to V$ such that the considered SSA is both ϕ-stochastic and ϕ'-stochastic, and that the maps $g \circ \phi$ and $g \circ \phi'$ are both measurable. But this doesn't exclude that the expectations $E_{U \sim \pi}\, g(\phi_U)$ and $E_{U \sim \pi}\, g(\phi'_U)$ are different for some π, and then only one of them can be identical with Φ_π.
##Wirklich? Sonst Darstellung von ϕ-expectational auf measurability beschränken!
So it could be considered to introduce concepts of being (almost) ϕ-expectational (in Π^*) to remove

[151] Since the property of being expectational in the empty subset of Π doesn't have any practical implications, unwanted consequences for $\Pi^* = \varnothing$ could also be avoided by the assumption that Π^* is a non-empty subset of Π. But this assumption would have to be checked in all application, and a short reminder of the measurability premissa is useful in itself and not only for the discussed formal reasons.

all ambiguities. But although the base-dependence of expectational properties is a challenge to logical precision, it should also be considered that stochastic SSAs are typically introduced for one specific map $\phi:\Omega_0\to V$, which is the 'prominent' one, whereas other maps ϕ' with similar mathematical properties are a nuisance rather than a fruitful enrichment. For such reasons, Definition 4.13 defines expectational properties only for stochastic SSAs $(V, \Pi, \Phi, H, T)_\phi$, the meaning of the subscript being explicated by Definition 4.8. As a consequence, expectational properties are declared only for stochastic SSA, where a prominent map has been introduced. Then the map ϕ in Definition 4.13 is implicitly referenced in any statement of expectational properties, and it has to be identified with the prominent map of the considered SSA.

With respect to the definition of almost expectational maps, note that Definition 4.13.(ii) doesn't require the existence of a unique set S, which has Properties (ii.a) through (ii.d) for every $\pi \in \Pi^*$. It suffices, if a set S with these properties exists for every $\pi \in \Pi^*$, and this requirement may be fulfilled by different sets S for different elements of the set Π^*.[152]

For convenient reference to the properties of a set S in Definition 4.13.(ii), a subset of the set V will be said to have Property 4.13.(ii.d) for a map $V\to\mathbb{R}$ and an element of the set Π, if the set, the map and the element of Π can take the roles of S, g and π in Definition 4.13.(ii.d). The specification of a map $V\to\mathbb{R}$ or an element of Π may be omitted in situations where a map $g:V\to\mathbb{R}$ resp. an element π of the set Π is specified in the local context. In a similar way, we will speak of Properties 4.13.(ii.a), 4.13.(ii.b) and 4.13.(ii.c), with the exception that a specification of a map $V\to\mathbb{R}$ is not required for Properties 4.13.(ii.a) and 4.13.(ii.b), and an element of Π is unnecessary for Properties 4.13.(ii.b) and 4.13.(ii.c).

The following lemma, which is proved in Section 6.27, states some interrelationships between expectational and almost expectational maps.

Lemma 4.14: For a stochastic SSA (V, Π, Φ, H, T), let $g:V\to\mathbb{R}$ be a map, and Π^* a subset of Π. Then the following assertions hold:

(i) The map g is expectational in Π^* iff there exists a stochastic base $(\Omega_0, \mathsf{A}_0, \phi)$ such that the map $g \circ \phi$ is A_0-B-measurable and that the properties of a set S in Definition 4.13.(ii) hold for $S = V$ and every $\pi \in \Pi^*$.

(ii) If the map g is expectational in Π^*, then it is almost expectational in Π^*.

(iii) If the map g is almost expectational in Π^*, then it is expectational in Π^* iff there exists a stochastic base $(\Omega_0, \mathsf{A}_0, \phi')$ of the SSA such that the map $g \circ \phi'$ is A_0-B-measurable.

(iv) If the map g is almost expectational in Π^* and S is a subset of V, where the map $(\chi_S \cdot g) \circ \phi$ is A_0-B-measurable and the properties $\phi^{-1}(S) \in \mathsf{A}_0$, $(S, \pi) \in H$ and $\Phi_\pi \in S$ hold for every $\pi \in \Pi^*$, then the map $\chi_S \cdot g$ is expectational in Π^*.

(v) If the map g is expectational (resp. almost expectational) in Π^*, then it is expectational (resp. almost expectational) in every subset of Π^*.

(vi) If $\{\Pi_i\}_{i\in I}$ is a family of subsets of Π such that g is expectational (resp. almost expectational) in every set Π_i of the family, and $\Pi^* = \bigcup_{i\in I} \Pi_i$, then g is expectational (resp. almost expectational in Π^*.

[152] In other words, the property of almost expectational maps definition in Definition 4.13.(ii) has the logical format $\exists\ (\Omega_0, \mathsf{A}_0, \phi)\text{: } \forall\ \pi \in \Pi^*\text{: } \exists\ S$: (ii.a) .. (ii.d), and this requirement is weaker than $\exists\ (\Omega_0, \mathsf{A}_0, \phi)\text{: } \exists\ S \subseteq V\text{: } \forall\ \pi \in \Pi^*$: (ii.a) .. (ii.d).

Assertion (iv) of the above lemma is mainly designed for situations, where only almost expectational maps are available, and where a set S with the assumed properties exists. Then all subsequently reported results for expectational maps can be applied to the map $\chi_S \cdot g$.

It has been mentioned that the differentiation of concepts in Definition 4.13 is introduced to cover certain maps $g:V\to\mathbb{R}$, which do not have all properties of expectational maps. Since concrete instances of such maps will turn up in the context of results to be presented in Subsection 4.5, a systematic account of reasons motivating the conceptual framework is postponed until the end of that subsection, where Corollary 4.22 is applied to a concrete experiment.Up to this point, an approach to the thread of considerations may be supported, if the reader thinks of maps, which are expectational in an entire SSA like the projection maps and their linerar combinations in the emprirical CDF-SSA.[153]

Expectational and almost expectational maps are the basis of another concept, which has also been prepared by the emprical CDF-SSA (Example 3.3). With the definition $Q := C\times W$, the vocabulary set V of that SSA can be conceived as a set of maps $v:Q\to\mathbb{R}$. In other words, the set V is a subset of the function space $\mathbb{R}^Q$. Furthermore, the projection maps $pr_{cw}:V\to\mathbb{R}$ are expectational, and they can now be reconceived as projection maps $pr_q:V\to\mathbb{R}$, where q is an element of the set Q, i.e., an ordered pair (c, w). The concept of projectional SSAs, which is explicated in the following definition, is a generalisation of this situation. In accordance with the differentiation of expectational maps, the definition provides weaker concepts for situations, where the projection maps $pr_q:V\to\mathbb{R}$ are expectational or almost expectational in a subset Π^* of the set Π.

Definition 4.15: An SSA (V, Π, Φ, H, T) is *(almost) projectional* (in Π^*, with $\Pi^* \subseteq \Pi$) iff it is a stochastic SSA with the following property: There is a non-empty set Q such that $V \subseteq \mathbb{R}^Q$, and all projection maps $pr_q:V\to\mathbb{R}$ are (almost) expectational (in Π^*).

Again, the additional specifications in parentheses have to be treated consistently: A stochastic SSA with $V \subseteq \mathbb{R}^Q$ is projectional iff the projection maps are expectational, and it is almost projectional iff the projection maps are almost expectational. Similarly, the specification of a subset Π^* of Π has to be considered as being twice present or twice absent. With these variations, Definition 4.15 explicates four concepts.

The simplest and most specific one of these concepts - the projectional SSA - formalises an idea of aggregation, which distinguishes the approach of Estes (1956) from his predecessors Sidman (1952) and Bakan (1954). Whereas these authors considered only aggregation by pointwise averaging of individual curves, Estes analysed expectations of random variables representing curve points and parameter values. However, these expectations have a different formal function. The aggregation is assumed to be based on expectations of random variables representing curve points,

[153] See Footnote 80 for the A_0-B-measurability of the maps $pr_{cw} \circ \phi$, and note that the specification of the aggregation rule Φ by Equation (3.9) can be rewritten as $pr_{cw}(\Phi_\pi) = E_{U\sim\pi}\, pr_{cw}(\phi_U)$. A similar property for linear combinations of projection maps is proved in Footnote 100.

and then it is questioned under what conditions this aggregation results in a curve, which is a member of the same parametric family as the aggregated curves. If this property is given, then it is a second question whether the parameters of the resulting curve are the expectations of the parameters of the aggregated curves. In other words, it is assumed that the projection maps $pr_q:V\to\mathbb{R}$ are expectational, and then it is questioned whether other maps $V\to\mathbb{R}$ representing parameters are also expectational. Example 4.18 in Section 4.5 will not only give an answer this question; it will also demonstrate a contribution of expectational maps to the solution of the more fundamental problem whether the curve resulting from expectational projection maps is a member of a parametric family of function.

More generally, the concepts of expectional maps and projectional SSAs are not directly related to aggregation stability, but helpful tools in its analysis. Some results supporting this tool function will be presented in Subsections 4.6 and 4.7. To enable demonstrations of results of the next subsection by applications, we take allowance to anticipate a proposition, which is systematically a corollary to Lemma 4.23. Therefore, it is proved together with the lemma in Section 6.31. For an exercise, the reader is invited to derive the corollary directly from the definition of almost expectational maps by an application of Mapping-Theorem 3.8, the role of the SSA (V', Π', Φ', H', T') in the theorem being taken by the Interval-SSA with $V' = \mathbb{R}$. (See Lemma 3.7 for this SSA).[154]

Corollary 4.16: Let a stochastic SSA (V, Π, Φ, H, T), a map $g:V\to\mathbb{R}$, a subset Π* of Π, and an interval S' of real numbers be given with the following properties: The map g is almost expectational in Π*, and the set Π* contains all elements π of Π with $(g^{-1}(S'), \pi) \in H$.
Then the set $g^{-1}(S')$ is contained in the set system T.

For an application of the corollary to the empirical CDF-SSA (Example 3.3), let S' be a subtinterval of the interval [−1, 1], and consider the set $pr_{cw}{}^{-1}(S')$ (with $(c, w) \in C\times W$), which consists of all elements v of the vocabulary set V with $v(c, w) \in S'$. Since the projection map pr_{cw} is expectational, it is almost expectational by Lemma 4.14.(ii). So Corollary 4.16yields the aggregation stability (i.e., the membership in the set system T) of the considered set.

Corollary 4.16 can also be taken as an opportunity to point out a motivation for introducing expectational maps not by an additional SSA-Axiom in Definition 4.1, but by the separate Definition 4.13. Although an aggregation by medians instead of expectations of random variables like $g(\phi_U)$ would be compatible with the basical Axioms (v) and (vii) of stochastic SSAs,[155] there are good arguments to prefer the aggregation by expectations (see ###). But an obligatory axiom requiring an aggregation rule Φ, which is entirely based on expectational maps, would make necessary a set of expectational maps $V\to\mathbb{R}$, which is sufficiently large to identify a unique element Φ_π of the vocabulary set V for every element π of the set Π. For instance, projectional SSAs like the empirical

---

[154] Since the concept of weakly expectational maps is rather abstract, the derivation of Corollary 4.16 is easier, if the map $g:V\to\mathbb{R}$ is assumed to be expectational in a set Π* with the properties specified in the corollary.

[155] See Footnote 143 for some limitations, which have to be observed in an aggregation by medians instead of expectations.

CDF-SSAs (Example 3.3) have this property, since the equation $pr_q(\Phi_\pi) = E_{U\sim\pi}\, pr_q(\phi_U)$ for every $q \in Q$ specifies a unique element Φ_π for every $\pi \in \Pi$. But requiring this property for all stochastic SSAs would be an unnecessary loss of generality. If there is just one expectational or almost expectational map $g:V\to\mathbb{R}$ such that a set of interest can be specified as $g^{-1}(S')$ with a suitable interval of real numbers S', then its aggregation stability can be derived from Corollary 4.16, and we do not need to know whether there are other (almost) expectational maps $V\to\mathbb{R}$ determining unique elements Φ_π of the vocabulary set V.

It may be objected that the above application of Corollary 4.16 doesn't yield more than a rather trivial result, which could very well be derived without any theory of expectational maps. Indeed, this objection holds for all situations, where the expectational property of a map of interest is part of premissas formulated in an aggregation rule. Bute more interesting results can be obtained, if we derive such properties for other maps. For instance, in the empirical CDF-SSA, maps $g_a:V\to\mathbb{R}$, $g_b:V\to\mathbb{R}$ and $g:V\to\mathbb{R}$ may be defined as follows: For every $v \in V$, $g_a(v)$ and $g_b(v)$ are the expectations of random variables with CDF v(a, .) resp. v(b, .), and $g(v) := g_b(v) - g_a(v)$. Now consider the hypothesis that the true score (i.e., the expectation) of a dependent variable is smaller under condition a than under condition b for every element u of the domain set D. This hypothesis can be expressed as $\phi(D) \subseteq g^{-1}(S')$ with $S' :=]0, w''-w']$.[156] Although an expectational property of the maps g_a, g_b and g isn't assumed explicitly by the aggregation rule Φ of the considered CDF-SSA, it can be derived from a corollary to be presented in the following subsection. More generally, the contribution of expectational maps to the study of stability under aggregation becomes untrivial, if two approaches are combined: The derivation of expectational properties of maps $g:V\to\mathbb{R}$ from simpler assumptions in an aggregation rule, and conclusions about aggregation stability, which are based on results like Corollary 4.16. These approaches are the objectives of the following two subsections.

4.5 Extending a Set of Expectational Maps

The objective of the present subsection has already been motivated at the end of preceeding Subsection 4.6: Statements of stability under aggregation, which are based on expectational maps $g:V\to\mathbb{R}$, are often trivial, if the expectational property of the considered map is introduced as a premissa by an aggregation rule. But even if the property of being expectational or almost expectational is given only for simple maps like the projection maps in projectional SSAs, basic methods of integration theory can be used to derive the same property for other maps. Combined with other facts like those stated in Corollary 4.16, the derived expectational property of more complex maps may yield more interesting results.

A simple extension of a given set of expectational maps has already been used fruitfully in a demonstration of Mapping-Corollary 3.9. A side result of this demonstration can be translated into the framework of expectational maps by the following statement: Since the projection maps $pr_{cw}:V\to\mathbb{R}$ in the empirical CDF-SSA of Example 3.3 are expectational, the result of every linear

[156] Since random variables with CDFs v(a, .) or v(b, .) are confined to the interval [w', w''] in the empirical CDF-SSA, the same holds for their expectations $g_a(v)$ and $g_b(v)$. So g(v) cannot be greater than $w'' - w'$.

combination of these projection maps is also an expectational map. The following lemma (whose proof in given in Section 6.28) generalises this result and adds a claim for constant maps $V \to \mathbb{R}$.

Lemma 4.17: For a stochstastic SSA (V, Π, Φ, H, T), let Π* be a subset of Π.
Then all constant maps $V \to \mathbb{R}$ are expectational in Π. Furthermore, the set of those maps $V \to \mathbb{R}$ which are expectational in Π* is a linear subspace of the function space $\mathbb{R}^V$. In particular, if a map $g: V \to \mathbb{R}$ is the result of a linear combination of maps $V \to \mathbb{R}$, which are expectational in Π*, then g is also expectational in Π*. Similarly, every linear combination of maps $V \to \mathbb{R}$, which are almost expectational in Π*, results in a map, which is almost expectational in Π*.

For another application of Lemma 4.17, we resume the approach of Estes (1956) to the aggregation stability of parametric families of functions, whose reformulation in Equation (2.10) and Estes-Theorem 2.4 (Section 2.2.3) was confined to aggregation by convex linear combinations. Among others, Lemma 4.17 enables an amendment of a certain weakness of Estes' argumentation. It has already been mentioned in Section 4.4 that Estes analysed expectations of random variables representing curve points and parameter values. In full accordance with the standards of Mathematical Psychology in his time, he used sophisticated tools of differential calculus (in particular Taylor srieses) for this analysis, but without formalising the expectations in a probability space. Instead, he worked under an implicit premissa, which can be explicated as follows: If maps belonging to a parametric family of functions are aggregated by pointwise expectations, then the parameter values underlying the aggregated maps can be considered as random variables, which have an expectation.[157] Under this tacid assumption, Estes proved that these expectations are the parameters of the map resulting from the aggregation, if the parametric family is based on a representation function of his Class A. Nowadays, such problems are frequently solved (or rather avoided) by explicit premissas referring to the measurability of maps and to the existence of finite expectations. But the following approach to the situation shows that the non-parametric representation of prametric families presented in Lemma 2.9 can be combined with Lemma 4.17 to fill the gap without additional premissas beyond the basic model of aggregating curves by pointwise expectations, which is formalised by the concept of projectional SSAs.

Example 4.18: Let Q be a non-empty set, and (V, Π, Φ, H, T) a projectional SSA with $V \subseteq \mathbb{R}^Q$.
Furthermore, let m be a natural number, and $\psi: \mathbb{R}^m \to \mathbb{R}^Q$ an affine and injective parametrisation map of a parametric family of functions $Q \to \mathbb{R}$. Finally, let a subset S* of the set V be given by the definition $S^* := V \cap \psi(\mathbb{R}^m)$. Then Lemma 2.9.(iv) and (v) enable a redescription of the parameters θ and of the set S* by families $\{h_k\}_{k=1..m}$ and $\{g_q\}_{q \in Q}$ of maps $h_k: V \to \mathbb{R}$ resp. $g_q: V \to \mathbb{R}$, whose specifications by Equations (2.36) and (2.37) can be rewritten in terms of projection maps $pr_q: V \to \mathbb{R}$ and $pr_{q(j)}: V \to \mathbb{R}$ as

$$h_k = \xi_k + \sum_{j=1..m} z_{kj} \cdot pr_{q(j)} \tag{4.13}$$

for k = 1..m, and

$$g_q = \tau_q + pr_q + \sum_{j=1..m} \zeta_{qj} \cdot pr_{q(j)} \tag{4.14}$$

157 Stated more formally, Estes (1956) introduced terms for the expectations of parameter values without discussing whether the parameter values underlying the 'individual' maps are at all real valued random variables, and if so, whether they have an expectation

for every $q \in Q$. In these equations, the term q_j refers to the members of a family $\{q_j\}_{j=1..m}$ of elements of Q, whereas ξ_k, z_{kj}, τ_q and ζ_{qj} stand for constant real numbers.[158] So the maps h_k and g_q are expectational by Lemma 4.17, since they are linear combinations of constant maps (ξ_k resp. τ_q) and of projection maps. (See Lemma 4.17 for the expectational property of constant maps, and Definition 4.15 for the projection maps in projectional SSAs.)

Several conclusions can be drawn from the example. First (and most important under the perspective of extending a set of expectational maps), it is demonstrated that the almost trivial expectational properties of constant maps and linear combinations of expectational maps, which are stated by Lemma 4.17, may become fruitful, if they are combined with the results of an analysis of a field of application. In particular, the claim of Lemma 2.9.(v) can again be restated by the equality $S^* = \bigcap_{q \in Q} S_q$, where S_q is the set of all elements v of the set V with $g_q = 0$. If this definition is rewritten as $S_q := g_q^{-1}(\{0\})$ and the set $\{0\}$ is conceived as an interval of real numbers, then the expectational property of the maps g_q can be used to obtain $S_q \in T$ from Corollary 4.16, and Intersection-Theorem 3.14 yields $S^* \in T$.

For a demonstration of the result of Example 4.18 for the maps $h_k{:}V \to \mathbb{R}$, consider an RSO-process such that the considered SSA is based on a measurable space (D, A_D) and a map $\phi{:}D \to V$, where D is a set of 'units'. Since the maps h_k are expectational, the equality

$$h_k(\Phi_\pi) = \mathsf{E}_{U \sim \pi}\, h_k(\phi_U) \tag{4.15}$$

holds for k = 1..m and for every $\pi \in \Pi$. For an interpretation of this equality, recall that Lemma 2.9.(iv) gives a special interpretation of the function values $h_k(v)$ for an element v of the set S*: They are the k^{th} components of a parameter θ underlying th map v. So let π be an element of the set Π with $(S^*, \pi) \in H$, and let the unit selected in an RSO-process be represented by a D-valued random variable U with distribution π. Then the premissa $(S^*, \pi) \in H$ assumes that the map $\phi_U{:}Q \to \mathbb{R}$ characterising the randomly selected unit U is π-almost-surely an element of the set S* - i.e., a member of the considered family of functions -, and the same holds for the map $\Phi_\pi{:}Q \to \mathbb{R}$ characterising the entire RSO-process, since $\Phi_\pi \in S^*$ follows (via SSA-Axiom (iv)) from the former result $S^* \in T$ and the present premissa $(S^*, \pi) \in H$. Combining these properties with the above recalled interpretation of $h_k(v)$ for $v \in S^*$, we can say: Since ϕ_U is almost surely contained in S* and Φ_π is contained in S*, $h_k(\phi_U)$ is almost surely the k^{th} component of a parameter underlying the map ϕ_U, and $h_k(\Phi_\pi)$ is the k^{th} component of a parameter for the map Φ_π. On the background of this interpretation, Equation (4.15) reformulates the basic result of Estes (1956) for his Class A: If the functions ('curves') entering a process of aggregation by pointwise expectations are (almost surely) elements of the considered family, then the result of the aggregation is a member of the family, and the components of the underlying parameters are the expectations of random variables representing the components of the aggregated curves. Furthermore, it is a side result that the above named tacid assumptions of Estes are verified: The expectational property of the maps $h_k{:}V \to \mathbb{R}$, which has been shown in Example 4.18, implies the measurability of the map $h_k \circ \phi$. So $h_k(\phi_U)$ is a real valued

[158] The more explicit specification of the real numbers ξ_k, z_{kj}, τ_q and ζ_{qj} in the proof of Lemma 2.9 (Section 6.10) shows that these numbers are constants in the sense of being independen of the argument v in terms like $h_k(v)$ and $g_q(v)$.

random variable, and it has a finite expectation, since the term $h_k(\Phi_\pi)$ in Equation (4.15) is a finite real number.[159]

Whereas Lemma 4.17 (which underlies Example 4.18) reformulates (in the framework of expectational maps) some basic facts referring to measurability and to expectations based on linear combinations, the following theorem uses another result of integration theory, which is commonly called Levi's theorem of monotone convergence.

Theorem 4.19: For a stochastic SSA (V, Π, Φ, H, T), which is based on a measurable space (Ω_0, A_0) and a map $\phi:\Omega_0 \rightarrow V$, let $\{g_n\}_{n=1..\infty}$ be an infinite sequence of maps $g_n:V \rightarrow \mathbb{R}$ such that the maps $g_n \circ \phi$ are A_0-B-measurable and that the inequality

$$0 \leq g_n(v) \leq g_{n+1}(v) \tag{4.16}$$

holds for n = 1..∞ and every $v \in V$. Furthermore, let S* be the set of all elements v of the set V with $\lim_{n\rightarrow\infty} g_n(v) < +\infty$.

Then the set $\phi^{-1}(S^*)$ is an element of the σ-algebra A_0.

Now let $g:V \rightarrow \mathbb{R}$ be a map such that

$$g(v) = \lim_{n\rightarrow\infty} g_n(v) \tag{4.17}$$

for $v \in S^*$, and $g(v) = 0$ for $v \in V \setminus S^*$.

Then the map $g \circ \phi$ is A_0-B-measurable.

Finally, let $\Pi^\sim$ be a subset of Π with $(S^*, \pi) \in H$ for every $\pi \in \Pi^\sim$, and Π* a subset of $\Pi^\sim$ with $\Phi_\pi \in S^*$ for every $\pi \in \Pi^*$, and assume that the maps g_n are expectational in $\Pi^\sim$.

Then the equivalence

$$\Phi_\pi \in S^* \;\Leftrightarrow\; \mathsf{E}_{U\sim\pi}\, g(\phi_U) < +\infty \tag{4.18}$$

holds for every $\pi \in \Pi^\sim$, and the map g is expectational in Π*.

Due to the central role of the convergence premissa (Equation (4.17)), Theorem 4.19 will subsequently be called the Convergence-Theorem. The proof of the theorem in Section 6.29 is followed by some notes referring to weaker premissas, and by a lemma with weaker premissas (Lemma 6.6), which is based upon Lebesgues' theorem of dominated convergence.

Another kind of convergence - the pointwise convergence of sums of maps - can be subsumed easily under the Convergence-Theorem. Let $\{g_i\}_{i=1..\infty}$ be a sequence of maps $g_i:V \rightarrow [0, +\infty[$, and S* the subset of those elements of V, where the sum $\sum_{i=1..\infty} g_n(v)$ is finite. Furthermore, let $g:V \rightarrow \mathbb{R}$ be a map with $g(v) = \sum_{i=1..\infty} g_n(v)$ for $v \in S^*$, and $g(v) = 0$ for $v \in V \setminus S^*$. Then the role of the maps g_n

[159] It may be objected that the maps $h_k:V \rightarrow \mathbb{R}$ in the approach to parametric families via Lemma 2.9 depend heavily upon the arbitrary choice of the family $\{q_j\}_{j=1..m}$ of elements of Q. Indeed, this objection can be based on the proof the lemma in Section 6.10, but an exception should be recalled: For all elements v of the set S*, the function values $h_k(v)$ are the components θ(k) of the unique parameter θ underlying the map $v:Q \rightarrow \mathbb{R}$. (See Lemma 2.9.(iv) for this conclusion.) So it doesn't matter, which elements $\{q_j\}_{j=1..n}$ are used to reconstruct the underlying parameter θ. Now the conclusion that the parameter underlying Φ_π is the result of a pointwise aggregation of parameters is confined to elements of Π with $(S^*, \pi) \in H$. For a random variable U with a distribution π of this property, it follows from SSA-Axiom (vii) that ϕ_U is almost surely contained in S*, and then the function values $h_k(\phi_U)$ are almost surely independent of the choice of the vectors $\{q_j\}_{j=1..m}$.

in Convergence-Theorem 4.19 can be taken by maps $g'_n:V\to\mathbb{R}$ defined by $g'_n(v) := \sum_{i=1..n} g_i(v)$ for $n = 1..\infty$ and every $v \in V$. In a similar way, Lemma 6.6 can be applied to pointwise converging sums of maps $g_i:V\to\mathbb{R}$, which may have negative function values.

Beyond demonstrating an application of the Convergence-Theorem, the following example prepares a further result, which will be statet in Corollary 4.21.

Example 4.20: Let (W, A_W) be a measurable space, and (V, Π, Φ, H, T) a projectional SSA, where V is the set of all probability measures on A_W. (In other words, the σ-algebra A_W takes the role of the set Q in Definition 4.15.) Furthermore, let $X:W\to[0, +\infty[$ be an A_W-B-measurable map, S* the subset of V consisting of all probability measures v on A_W with a finite expectation $\mathsf{E}_{w\sim v}\, X(w)$, and Π^* the set of all elements π of the set Π with the properties $\Phi_\pi \in S^*$ and $(S^*, \pi) \in H$. Finally, define a map $g:V\to\mathbb{R}$ by $g(v) := \mathsf{E}_{w\sim v}\, X(w)$ for $v \in S^*$ and $g(v) := 0$ for $v \in V\setminus S^*$. In this situation, Convergence-Theorem 4.19 can be used to show that the map g is expectational in Π^*.

To verify this claim, observe that the assumed measurability of the map X implies the existence of a sequence $\{X_n\}_{n=1..\infty}$ of maps $X_n:W\to\mathbb{R}$ with $X(w) = \lim_{n\to\infty} X_n(w)$ for every $w \in W$, and with the following properties for $n = 1..\infty$:[160]

(i) $0 \le X_n(w) \le X_{n+1}(w) \le X(w)$ for every $w \in W$.

(ii) There is a natural number m(n), a sequence $\{\xi_{nj}\}_{j=1..m(n)}$ of real numbers ξ_{nj}, and a sequence $\{A_{nj}\}_{j=1..m(n)}$ of mutually disjoint elements A_{nj} of A_W with $\bigcup_{j=1..m(n)} A_{nj} = W$, and $X_n(w) = \xi_{nj}$ for every $w \in A_{nj}$.

For a given sequence of maps $X_n:W\to\mathbb{R}$ with these properties, define a sequence $\{g_n\}_{n=1..\infty}$ of maps $g_n:V\to\mathbb{R}$ by the equation

$$g_n(v) := \mathsf{E}_{w\sim v}\, X_n(w) = \sum_{j=1..m(n)} \xi_{nj}\, v(A_{nj}) \tag{4.19}$$

for $n = 1..\infty$ and every $v \in V$. The second equality shows that every map $g_n:V\to\mathbb{R}$ is a linear combination (with coefficients ξ_{nj}) of the expectational projection maps $pr_{A(nj)}:V\to\mathbb{R}$; so the maps g_n are expectational (Lemma 4.17).

With these specifications, it can be left to the reader[161] to show that the definitions of the sets S* and Π^* and of the map $g:V\to\mathbb{R}$ comply with the premissas of the Convergence-Theorem and to derive that the map g is expectational in Π^*, indeed.

In particular, Example 4.20 can be used to demonstrate the role of the set S* in the Convergence-Theorem and the motivation of providing a concept of maps, which are expectational only in a subset Π^* of the set Π in a stochastic SSA. We will first consider a situation, where this precaution is unnecessary. Assume that the map X is bounded, which means the existence of a (finite) real number ξ such that $|X(w)| \le \xi$ for every $w \in W$. In this situation, a finite expectation $\mathsf{E}_{w\sim v}\, X(w)$

[160] See e.g. Bauer (1992, p. 70, Proposition 11.6), and note the underlying definition of a set E* on his p. 66.

[161] Exercise!

exists for every $v \in V$, and this implies $S^* = V$ and $\Pi^* = \Pi$.[162] As a consequence, the equation $g(v) = \mathsf{E}_{w\sim v} X(w)$ holds for every element v of the set V (i.e., for every probability measure v on the σ-algebra A_W), and the map $g:V\to\mathbb{R}$ with this specification is expectational in Π.

Conversely, if the map X is unbounded, then there are elements v of the set V without a finite expectation $\mathsf{E}_{w\sim v} X(w)$,[163] and then the set S* is a proper subset of V. Some formal problems resulting from this situation form the main motivation to consider maps, which may be expectational or almost expectational only in a proper subset Π* of the set Π. In fact, this variant has been introduced in Definition 4.13.(i) and (ii) to enable a solution of the said problems by a combination of results like Convergence-Theorem 4.19 with elementary facts of our everyday background knowledge. The approach will be demonstrated by a concrete application after Corollary 4.22.

The following corollary is a generalisation of Example 4.20 in two points: In the example, it is assumed that the vocabulary set V is the set of all probability measures on A_W; but the corollary assumes only that every element of the vocabulary set specifies a probability measure by a map $\gamma:V\to V'$, where V' is a set of probability measures on A_W. (For instance, the elements v of the vocabulary set in the empirical CDF-SSA of Example 3.3 are not probability measures; but for every such v, the partial map v(a, .) specifies a probability measure on A_W, since it is a CDF with domain W.) Furthermore, Example 4.20 admits only finite, non-negative real numbers as function values of the map X; but the corollary can be applied to every A_0-B^*-measurable map $X:W\to\mathbb{R}^*$, where $\mathbb{R}^*$ is the set of 'extended real numbers' including $-\infty$ and $+\infty$, and B^* the σ-algebra of Borel sets in $\mathbb{R}^*$.[164]

Corollary 4.21: Let (V, Π, Φ, H, T) be a stochastic SSA, which is based on a measurable space (Ω_0, A_0) and a map $\phi:\Omega_0\to V$. Furthermore, let S_0 be a subset of V, Π_0 a subset of Π, (W, A_W) a measurable space, S a system of subsets of W with $\mathsf{S} \subseteq \mathsf{A}_W$, V' a set of maps $v':\mathsf{A}_W\to\mathbb{R}$, which

---

[162] According to its definition in Example 4.20, the set S* is the set of all elements v of the set V with a finite expectation $\mathsf{E}_{w\sim v} X(w)$. So it is identical with V, if this expectation is finite for every $v \in V$. Then SSA-Axiom (ii) implies $(S^*, \pi) \in H$ for every $\pi \in \Pi$, and $\Phi_\pi \in S^*$ must also hold for every $\pi \in \Pi$, since the function values of the aggregation rule $\Phi:\Pi\to V$ are elements of the vocabulary set V. But then the definition of the set Π* in Example 4.20 implies $\Pi^* = \Pi$. (Note a difference between the specifications of the set Π* in Convergence-Theorem 4.19 and Example 4.20: Whereas the theorem requires only that Π* is a subset of Π with $(S^*, \pi) \in H$ and $\Phi_\pi \in S^*$ for every $\pi \in \Pi^*$, the set Π* in Example 4.20 is the set of all elements of Π with these properties.)

[163] More explicitly, the assumption of an unbounded map X implies the existence of a sequence $\{\xi_n\}_{n=1..\infty}$ of elements of the set X(W) with $0 < 2\,(\xi_{n+1} - \xi_n) \le \xi_{n+2} - \xi_{n+1}$ for $n = 1..\infty$. For a sequence with these properties, define a sequence $\{y_n\}_{n=1..\infty}$ of maps $y_n:\mathsf{A}_W\to[0, 1]$ by $y_n(A) := 2^{-n}$ for $\xi_n \in A$, and $y_n(A) := 0$ for $\xi_n \notin A$. It can be left to the reader to verify that the map $v:\mathsf{A}_W\to\mathbb{R}$ given by $v(A) := \sum_{n=1..\infty} y_n(A)$ is a probability measure on A_W (i.e., an element of the set V), and that the expectation $\mathsf{E}_{w\sim v} X(w)$ is infinite.

[164] Formally: $\mathbb{R}^* := \mathbb{R} \cup \{-\infty, +\infty\}$, and B^* is the system of all subsets A of $\mathbb{R}^*$ with $(A \cap \mathbb{R}) \in \mathsf{B}$.

are probability measures on A_W, and γ:V→V' a map, and assume the following properties of these objects:

(i) For every A ∈ S, the map $pr_A \circ \gamma$ is expectational in Π_0.

(ii) A_W is the σ-algebra in W generated by the set system S.

(iii) The intersection of two elements of S results in an element of S.

Then the following property follows:

(iv) For every A ∈ A_W, the map $pr_A \circ \gamma$ is expectational in Π_0.

Now let X:W→$\mathbb{R}$* be an A_W-B*-measurable map, and let sets S*, $\Pi^{\sim}$ and Π* as well as a map g:V→$\mathbb{R}$ be given with the following properties:

(v) S* is the set of all elements v of the set V with a finite expectation $\mathsf{E}_{w\sim\gamma(v)}$ X(w).

(vi) $\Pi^{\sim}$ is a subset of Π_0 with (S*, π) ∈ H for every π ∈ $\Pi^{\sim}$.

(vii) Π* is a subset of $\Pi^{\sim}$ with Φ_π ∈ S* for every π ∈ Π*.

(viii) g(v) = $\mathsf{E}_{w\sim\gamma(v)}$ X(w) for v ∈ S*, and g(v) = 0 for v ∈ V\S*.

Then the following properties follow:

(ix) For every π ∈ $\Pi^{\sim}$, a finite expectation $\mathsf{E}_{U\sim\pi}$ g(ϕ_U) exists iff Φ_π ∈ S*.

(x) The map g is expectational in Π*.

The corollary is mainly designed for the study of RSO-processes by a class of stochastic SSAs, which is prototypically exemplified by the empirical CDF-SSA (Example 3.3), and this motivates the denotation as RSO-Corollary. In the proof of the corollary in Section 6.30, its Premissas (ii) and (iii) are replaced by weaker assumptions.

The - admittedly abstract - set theoretical structure, which is assumed by the premissas of the RSO-Corollary, may become easier to follow up by an application to the empricial CDF-SSA. So let A_W be the σ-algebra consisting of all Borel sets in $\mathbb{R}$, which are subsets of the set W. Although the vocabulary set V is not a set of probability measures on A_W, its elements specify probabilities for some elements of the σ-algebra (the intervals [w', w] with w ∈ W), and implicitly, these specifications are sufficient to derive probability measures on A_W under the conditions a and b. Confining the following considerations to condition a, we can also say that there is a map γ:V→V', where V' is a set of probability measures on A_W, and γ(v) is the probability measure on A_W specified by the partial map v(a, .). (Recall that this partial map is assumed to be a CDF with domain W.)

The intervals [w', w] with w ∈ W form a system of subsets of W, and the denotation S for this set system is used in the corollary. In the CDF-SSA, the probability assigned to an interval [w', w] with w ∈ W is specified by the expectational projection map pr_{aw}:V→$\mathbb{R}$. The assignment of probabilities by expectational maps is also assumed in Premissa (i) of the RSO-Corollary. However, since the corollary doesn't assume that the considered SSA is projectional, the assignment of probabilities to subsets of the set W is formalised in a more general way. Nevertheless, this formalisation can again be demonstrated by the example of the empirical CDF-SSA. If v' is an element of the set V' (i.e., a probability measure on A_W), then v'(A) is the probability assigned to an element A of the σ-algebra A_W. Conceiving the probability measure v' on A_W as a map v':A_W→$\mathbb{R}$, we can also write pr_A(v') instead of v'(A), and this notation can be used for a formalisation of the specification of probabilities by element of the vocabulary set V. If v is an element of the vocabulary set V, then γ(v) is the probability measure on A_W specified by the CDF v(a, .), and the probability assigned to an element A of the σ-algebra A_W is $pr_A(\gamma(v))$. In other words, the assignment of probabilities to A is performed by the map $pr_A \circ \gamma$. In the empirical CDF-SSA, this map is identical with the projection map pr_{aw}:V→ℝ for A = [w', w]. The maps $pr_A \circ \gamma$ are used in the RSO-Corollary to formulate both a premissa and a result. Premissa (i) assumes that the map $pr_A \circ \gamma$ is expectational (in a subset Π_0 of the set Π) for all elements of the set system S (e.g., for all intervals [w', w] with w ∈ W in the empirical CDF-SSA). Under this premissa and two other ones, which will be commented immediately, Assertion (iv) claims the same property for the map $pr_A \circ \gamma$ of all elements A of the σ-algebra A_W. Since the projection maps pr_{aw} (which are reconceived as maps $pr_A \circ \gamma$ in the corollary) are expectional (in Π) in the CDF-SSA, the set Π_0 of the corollary is identical with Π in this application.

Two properties of the set system S, whose validity is obvious in the CDF-SSA, are introduced in the RSO-Corollary by its Premissas (ii) and (iii). The assumption that A_W is the σ-algebra generated by the set system S doesn't need a comment. Furthermore, if [w', w_1] and [w', w_2] are intervals with $w' \le w_1 \le w_2 \le w''$ (i.e., elements of the set system S), then then the intersection of the two intervals is also an element of S. This property of 'being closed under all intersections of two of its elements' is assumed by premissa (iii).

Since the premissas of the RSO-Corollary have been established for the CDF-SSA, its results are summarised in the following corollary both for a special class of maps X:W→ℝ and for general maps X:W→ℝ*.

Corollary 4.22: For the vocabulary set V of the empirical CDF-SSA (Example 3.3), let maps g_{cn}:V→ℝ with c ∈ C and n ∈ ℕ be defined by $g_{cn}(v) = E_{Y\sim v(c,\,.\,)}Y^n$.[165]
Then every such map and every linear combination of such maps is expectational in the empirical CDF-SSA and also in the analytical CDF-SSA (Example 3.20).
More generally, let X:W→ℝ* be an A_0-B*-measurable measurable map, Π^* a subset of Π, and subsets S^*_a and S^*_b of the set V as well as maps g_a:V→ℝ and g_b:V→ℝ be given with the following properties for c = a and c = b:
(i) S^*_c is the set of all elements v of the set V where the expectation $E_{Y\sim v(c,\,.\,)}X(Y)$ is finite, and $g_c(v)$ is this expectation for $v \in S^*_c$; furthermore, $g_c(v) := 0$ for $v \in V\backslash S^*_c$.
(ii) For every $\pi \in \Pi^*$, the expectation $E_{U\sim\pi}\, g_c(\phi_U)$ is finite, and $(S^*_c, \pi) \in H$.
Then every such map g_c and every linear combination of these maps is expectational in Π^*.

A combined application of this corollary and of Corollary 4.16 has already been outlined at the end of Subsection 4.4. Reconsider the hypothesis that the true score in the dependent variable Y is lower under condition a than under condition b for every unit belonging to the domain set D. This hypothesis can be formalised as $\phi(D) \subseteq g^{-1}(S')$, where S' is the interval]0, w''−w'], and a map g:V→ℝ is defined (in the notation of Corollary 4.22) as $g := g_{b1} - g_{a1}$. Now Corollary 4.22 tells that the just defined map g:V→ℝ is expectational, since it is a linear combination (with coefficients −1 and 1) of the maps g_{a1} and g_{b1}. So the aggregation stability of the set $g^{-1}(S')$ can be obtained from Corollary 4.16, and this result can be used to derive testable statistical predictions in the way outlined in Subsection 4.2.

---

[165] More explicitly, $g_{cn}(v)$ is the expectation of Y^n for a real valued random variable Y, whose CDF is the partial map v(c, .). See Definition 3.2 for the term $E_{Y\sim v(c,\,.\,)}Y^n$.

Another example, which is based on the more general claim of Corollary 4.22, can be used to motivate three (admittedly abstract) issues in the definition of expectational and almost expectational maps (Definition 4.13). First, only maps $g{:}V \to \mathbb{R}$ can be expectational or almost expectational, but not maps $g{:}V \to \mathbb{R}^*$, where the function value $g(v)$ is infinite for some $v \in V$. Furthermore, the definition contains a concept of maps, which are almost expectational, but not expectational. Finally, there are maps, which are expectational or almost expectational only in a subset Π^* of the set Π in a stochastic SSA (V, Π, Φ, H, T). For the example, assume that the empirical CDF-SSA is applied to a situation, where a subject has to work on two tasks t' and t'', and that the experimental conditions a and b (i.e., the elements of the set C) are informations about the tasks. Being interested in the effect of such informations upon the time dispositions of subjects, the experimenter assesses a dependent variable Y, which is the percentage of total time spent for task t'. So the set W in the CDF-SSA is the interval [0, 100], i.e., $w' = 0$ and $w'' = 100$. However, the hypothesis underlying the experiment doesn't refer directly to this dependent variable Y, but to the derived variable $(100 - Y) / Y$ (i.e., to the reatio of times spent for the two tasks, and the hypothesis claims that the true score (i.e., the individual expectation) of this ratio is greater under condition b than under condition a. So define a map $X{:}W \to \mathbb{R}^*$ by $X(w) := (100 - w) / w$ for every $w \in W$, including $X(0) := +\infty$. Furthermore, define subset S^*_c of V and maps $g_c{:}V \to \mathbb{R}$ as in the corollary. The definition $g_c(v) := 0$ for $v \in V \backslash S^*_c$ will be discussed later, but for the moment, this issue becomes irrelevant by the following assumption, which is rather safe, if task t' is presented first: For every $u \in D$, there is a non-zero element w_u of the set W with $\phi_u(c, w_u) = 0$. In other words, even if the time for task t' is used only for the decision to switch to task t'', its percentage (relative to the total time spent for both tasks) is almost surely greater than w_u. Then the following property follows immediately: If Y is a random variable with CDF $v(c, .)$, then $0 \leq X(Y) \leq_{a.s.} X(w_u)$, and this implies

$$0 \leq \mathsf{E}_{Y \sim v(c,\,.)}\, X(Y) \leq X(w_u). \tag{4.20}$$

So the random variable $X(Y)$ has a finite expectation, and $g_c(\phi_u)$ is the true score of interest for unit u under condition c. Furthermore, since the assumption implies $\phi(D) \subseteq S^*$, we obtain $(S^*_c, \pi) \in H$ for every $\pi \in \Pi$ from the definition of the relation H in Example 3.3.(v).

The introduction of a subsets Π^* in Corollary 4.22 is motivated by a problem, which can also be demonstrated by the outlined experiment. If only the existence of individual elements w_u of the set W with the above property is assumed, then it cannot be excluded that the set Π contains an element π, where the expectation $\mathsf{E}_{U \sim \pi}\, g_c(\phi_U)$ is infinite, and then this expectation cannot be identical with $g_c(\Phi_\pi)$ as in an expectational map. (Note that the definition of the map g_c implies that $g_c(\Phi_\pi)$ must be a finite real number.) But the final claim of Corollary 4.22 can be combined with Definition 4.13.(i) to obtain the equality $g_c(\Phi_\pi) = \mathsf{E}_{U \sim \pi}\, g_c(\phi_U)$ for every element π of the set Π^*. Now a minimal strengthening of the above assumption suffices to obtain the assumed properties of the set Π^* for the entire set Π: If we assume the existence of a non-zero element w^* of the set W with $\phi_u(c, w^*) = 0$ for every $u \in D$, then the element w_u in Inequality (4.20) can be replaced by w^*, and the inequality $0 \leq \mathsf{E}_{U \sim \pi}\, g_c(\phi_U) \leq X(w^*)$ follows for every $\pi \in \Pi$.[166] But this implies that the entire set Π has all properties required for the set Π^* of Corollary 4.22. So the maps g_c are expectational in Π. Moreover, if we define a map $g{:}V \to \mathbb{R}$ by $g(v) := g_b(v) - g_a(v)$, then this map (being a linear

[166] Exercise!

combination of the maps g_a and g_b) is also expectational.

This result can be used to derive a testable prediction. Define a subset S of the set V by $S := g^{-1}(]0, +\infty[)$, and observe that the hypothesis under study can be written as $\phi(D) \subseteq S$.[167] Furthermore, since Corollary 4.16 yields $S \in T$, the hypotheses implies for an RSO-process with arbitrary selection distribution $\pi \in \Pi$ that the expectation of the derived dependent variable X(Y) is smaller under condition a than under condition b.

The following generalisation of the example describes a contribution of expectational and almost expectational maps to an interface between the mathematical analysis of stochastic SSAs and its application. In situations, where the mathematical analysis yields an expectational property of a map $g:V\to\mathbb{R}$ only in a subset Π^* of the set Π, background informations from the intended field of application may sometimes allow assumptions, which imply the identity of the sets Π^* and Π, and then results like Corollary 4.16 can be applied to derive the aggregation stability of a set of interest.[168] To some readers, paying any attention to problems with non-existing or infinite expectations may look like wasted effort. But others may find it satisfactory to account for assumptions, whose introduction seems frequently motivated only by the removal of seemingly irrelevant formal difficulties. The above application of Corollary 4.22 demonstrates that paying careful attention to such problems instead of ignoring them may lead to results describing a class of situations, where the problems can be excluded, and to the conclusion that a concrete application belongs to this class.

Since references to background informations are frequently misunderstood as requirements upon the availability of well established scientific results, it should be noted that elementary everyday knowledge may very well be sufficient and that a suitable planning of experimental situations may contribute to their validity. So the above considered experiment could be realised in the following way. The tasks t' and t'' are presented on a computer monitor; but at every point of time, only one of the tasks is being displayed. The subject can switch between the two tasks by a keystroke, the first keystroke after an instruction always leading to the presentation of task t'. The dependent variable Y is the precentage of total time where task t' was on the screen. More precisely, total time is the time used for the completion of both tasks; but if the tasks are not completed after two hours, the experiment is stopped, and then Y is the percentage of the two hours where task t' was on the screen. Since technical limitations imply a minimal time difference between the first two key strokes (say 1 millisecond), task t' is displayed for at least this time. Since 1 millisecond is 1/72000 percent of two hours, the property $v(c, w^*) = 0$ holds for $w^* := 1 / 72000$.

Another way of using maps $g:V\to\mathbb{R}$, which are expectational or almost expectational only in a subset Π^* of the set Π is based on a variant of Corollary 4.16. Let (V, Π, Φ, H, T) be a stochastic SSA, Π^* a subset of Π, and $g:V\to\mathbb{R}$ a map, which is almost expectational in Π^*. Furthermore, let

---

[167] More explicitly, the hypothesis is equivalent with $\phi(D) \subseteq S$ under the premissa $\phi(D) \subseteq (S^*_a \cap S^*_b)$, the validity of this premissa following from the assumption referring to the element w*. Without the premissa, an element u of the set D with $\mathsf{E}_{Y\sim h(u)(a,\,.\,)}\, X(Y) = +\infty$ and $0 < \mathsf{E}_{Y\sim h(u)(b,\,.\,)}\, X(Y) < +\infty$ would violate the hypothesis, although $\phi_u \in S$.

[168] Observe also that the identity of the sets Π^* and Π isn't necessary for an application of Corollary 4.16. It suffices to verify (possibly under assumptions derived from background informations) that the set Π^* contains all elements π of the set Π with $(g^{-1}(S'), \pi) \in H$.

(V', Π', Φ', H', T') be the Π*-restriction of the considered SSA. Then the set system T' contains all sets $g^{-1}(S')$, where S' is an interval of real numbers.[169] So a prediction for an RSO-process may be derived by an approach to Π'-restrictions, which has been outlined at the end of Section 3.4.

To motivate the definition $g_c(v) := 0$ for $v \in V\backslash S^*_c$ in Corollary 4.22.(i), observe first that the maps g_c and g are claimed to be expectational only in the set Π*. Since $(S^*_c, \pi) \in H$ is required for every $\pi \in \Pi^*$, we have almost surely $\phi_U \in S^*_c$ for a random variable U with distribution $\pi \in \Pi^*$. So the definition of a function value $g_c(v)$ for $v \in V\backslash S^*_c$ doesn't affect the existence or the value of the expectation $\mathsf{E}_{U\sim\pi}\, g(\phi_U)$, which must be identical with $g(\Phi_\pi)$ for an expectational map. Insofar (!), the zero function value $g_c(v)$ for $v \in V\backslash S^*_c$ is a sort of dummy value without any consequence for the expectations of interest, and it could be replaced by any other rule, but with one exception: The expectation $\mathsf{E}_{U\sim\pi}\, g(\phi_U)$ is undefined if the map $g_c \circ \phi$ isn't A_0-B-measurable, and this may happen, if arbitrary function values g(v) are assigned for $v \in V\backslash S^*_c$, whereas the measurability requirement is fulfilled under the definition $g_c(v) = 0$ for $v \in V\backslash S^*$.[170]

The same measurability problem motivates the concept of almost expectational maps. In some situations, one may be interested in maps $g'_c{:}V\to\mathbb{R}$, which are identical with the maps g_c of Corollary 4.22 for $v \in S^*_c$. If function values $g_c(v)$ for $v \in V\backslash S^*_c$ are chosen such that the maps $g'_c \circ \phi$ are A_0-B-measurable, then the maps g'_c are expectational. But if there are doubts about this measurbility, then the set S^*_c of the corollary can take the role of the set S in the definition of almost expectational maps (Definition 4.13.(ii)). So the maps g'_c are almost expectational, and all results requiring only almost expectational maps (like Corollary 4.16) become available.

Finally, the above application of Corollary 4.22 demonstrates why maps $g{:}V\to\mathbb{R}^*$ with infinite function values g(v) are excluded from being expectational. If one is only interested in the maps $g_c{:}V\to\mathbb{R}$, it could be considered to admit infinite function values $g_c(v)$ in situations with an infinite expectation $\mathsf{E}_{Y\sim\pi}\, X(Y)$. But then the definition $g(v) := g_b(v) - g_a(v)$ couldn't be applied to elements v of the set V with $g_a(v) = g_b(v) = +\infty$. More generally, the expectational properties of linear combinations of expectational maps, which have been stated in Lemma 4.17, are immediately relevant for applications, and they are also used in the proof of RSO-Corollary 4.21. But linear combinations of maps with infinite function values may be undefined, and the loss of a central result[171] would be a too high price for the admission of expectational maps with infinite function values. In particular, maps with infinite function values are rarely relevant in psychology, and the above application of Corollary 4.22 demonstrates that infinite function values can frequently be replaced without loss of fruitful consequences by finite dummy values.

---

[169] Proof: It is easily derived from Definitions 3.5 and 4.13.(ii) that the property of being almost expectational in Π* is maintained in the transition to the SSA (V', Π', Φ', H', T'). So an application of Corollary 4.16 to this SSA yields $g^{-1}(S') \in T'$ for every interval S' of real numbers.

[170] If the assignment $g_c(v) := 0$ for $v \in V\backslash S^*$ in Corollary 4.22 wouldn't lead to A_0-B-measurable maps $g_c \circ \phi$, then the maps g_c couldn't be expectational. (See Definition 4.13.(i) for this argument.)

[171] Of course, the result for linear combinations of expectational maps could be stated under the premissa that the linear combination leads to a well defined map $V\to\mathbb{R}^*$. In the rare cases, where infinite function values cannot be replaced by finite dummy values, it is left to the user to check whether the premissas and the proofs of subsequently reported results can be adapted suitably.

4.6 Applications of Expectational Maps

Indeed, many subsets of a vocabulary set V, which may be used in the explication of psychological hypotheses, can be defined in this way by a single expectational or almost expectational map.

In Subsection 4.3, we have derived some properties of the relation H in stochastic SSAs from the specification by SSA-Axiom (vii). In particular, some results provided an interface to Mapping-Theorem 3.8 and its corollaries: Some premissas referring to the relation H (in particular Implications (3.35) and (3.37) and their reversals) could be verified for rather general classes of applications. Similarly, the present subsection will show how the concept of expectational maps $g:V\rightarrow\mathbb{R}$ in Definition 4.13 can be used for an interface to other premissas of Mapping-Theorem 3.8 and its corollaries, which require the validity of Equation (3.36) resp. (3.38) for certain elements π of the set Π. The first equation refers to the aggregation rules Φ and Φ' of two SSAs (V, Π, Φ, H, T) and $(V', \Pi', \Phi', H', T')$, which are interrelated by maps $g:V\rightarrow V'$ and $f:\Pi\rightarrow\Pi'$, and the premissa requires that function values $\Phi'(f(\pi))$ of the second aggregation rule Φ' can be obtained by an application of the map g to the function value $\Phi(\pi)$ of the first aggregation rule. The following lemma states this property for the simple case, where the SSA $(V', \Pi', \Phi', H', T')$ in the Mapping-Theorem is an Interval-SSA (Lemma 3.7). Since the concept of maps, which are expectational only in a proper subset Π^* of the set Π hasn't yet been motivated, a first approach to the lemma may be supported by the additional assumption $\Pi^* = \Pi$.

Lemma 4.23: Let (V, Π, Φ, H, T) be a stochastic SSA, which is based on a measurable space (Ω_0, A_0) and a map $\phi:\Omega_0\rightarrow V$. Furthermore, let Π^* be a subset of Π, and $g:V\rightarrow\mathbb{R}$ a map, which is expectational in Π^*. Finally, let $(V', \Pi', \Phi', H', T')$ be the Interval-SSA with $V' = \mathbb{R}$, and let a map $f:\Pi^*\rightarrow\Pi'$ be given such that

$$f(\pi)(A) = \pi(\phi^{-1}(g^{-1}(A))) \tag{4.21}$$

for every $g \in G$, every $\pi \in \Pi^*$ and every Borel set A in $\mathbb{R}$.
Then the equation $\Phi'(f(\pi)) = g(\Phi(\pi))$ holds for every $\pi \in \Pi^*$.

See Section 6.31 for a proof of the lemma. The interpretation of the map $f:\Pi\rightarrow\Pi'$ specified by Equation (4.21) is the same as in former applications of Mapping-Theorem 3.8 to the emprical CDF-SSA and to Interval-SSAs: If U is an Ω_0-valued random variable with distribution π, then $f(\pi)$ is the distribution of the real valued random variable $g(\phi_U)$.

Combining Lemma 4.23 with Corollary 4.10.(iv) and Mapping-Theorem 3.8.(i), we obtain the membership in the set system T for every set $g^{-1}(S')$, where S' is an interval of real numbers and $g:V\rightarrow\mathbb{R}$ is a map, which is expectational in a set Π^* containing all elements π of Π^* with $(g^{-1}(S'), \pi) \in H$. Corollary 4.16, which has been anticipated in Section 4.4 for demonstrations, states this property under the weaker premissa that the map g is almost expectational (and not necessarily expectational) in Π^*. The proof of the corollary is combined with the proof of Lemma 4.23 in Section 6.31.

Expectational and almost expectational maps can also be used to study the aggregation stability of sets, which are defined in a more complex way than the hitherto considered sets $g^{-1}(S')$ with an expectational or almost expectational map $g:V\to\mathbb{R}$ and an interval S' of real numbers. In particular, the function values of expectational maps can be transformed by a convex function[172], and a map $f:V\to\mathbb{R}$, which is a sum of such concatenations, can be used to study subsets $f^{-1}(]-\infty, \xi[)$ and $f^{-1}(]-\infty, \xi])$ of the vocabulary set, where ξ can be any real number. The following lemma, whose proof in Section 6.32 is based on the well known Jensen-inequality, states the aggregation stability of such sets.

Lemma 4.24: For a stochastic SSA (V, Π, Φ, H, T), let S be a subset of V, and Π^* a subset of Π containing all elements of Π with $(S, \pi) \in H$. Furthermore, let $\{f_i\}_{i=1..n}$ be a finite sequence of maps $f_i:V\to\mathbb{R}$ with the following properties for i = 1..n: There is a non-empty interval S_i of real numbers, a map $g_i:V\to\mathbb{R}$ with $g_i(V) \subseteq S_i$, which is almost expectational in Π^*, and a convex map $h_i:S_i\to\mathbb{R}$ such that $f_i = h_i \circ g_i$. Finally define a map $f:V\to\mathbb{R}$ by $f := \sum_{i=1..n} f_i$, and let ξ be a real number such that S is one of the sets $f^{-1}(]-\infty, \xi[)$ and $f^{-1}(]-\infty, \xi])$.
Then S is contained in T.

For a demonstration, we resume (from Section 2.5) the pooling of intraindividual and interindividual variance and reconsider the set S_ξ of those elements v of the vocabulary set V in the empirical CDF-SSA where the variance of a random variable with CDF v(a, .) is greater than some positive constant ξ. In the notation of Corollary 4.22, this variance is $g_{a2}(v) - g_{a1}(v)^2$. So the set S_ξ can be defined by the equation

$$S_\xi := \{v \in V: g_{a2}(v) - g_{a1}(v)^2 > \xi\}, \tag{4.22}$$

and this definition can be brought into the format of Lemma 4.24 by the following substitutions: With n := 2, define $S_1 := S_2 := \mathbb{R}$, $g_1 := g_{a1}$, $g_2 := g_{a2}$, $h_1(\xi') := \xi'^2$, and $h_2(\xi') := -\xi'$ for every $\xi' \in \mathbb{R}$. Then the maps h_1 and h_2 are convex, and with maps $f_i:V\to\mathbb{R}$ and $f:V\to\mathbb{R}$ as in Lemma 4.24, the set S_ξ can be redefined as $f^{-1}(]-\infty, -\xi[)$. So the aggregation stability of S_ξ follows from the lemma. Conversely, if we are interested in the set S of all elements of the vocabulary set with a smaller variance under condition a than under condition b, we can add $S_3 := S_4 := \mathbb{R}$, $g_3 := g_{b1}$, $g_4 := g_{b2}$, and work with maps $h_1(\xi') := -\xi'^2$, $h_2(\xi') := \xi$, $h_3(\xi') := \xi'$ and $h_4(\xi) := \xi$, then the equation $S = f^{-1}(]-\infty, 0[)$ is easily verified; but since the map h_1 is non-convex, an application of Lemma 4.24 fails.

There is still another useful approach to situations where several expectational or almost expectational maps have to be considered simultaneously: The involved maps can be conceived as a family $\{g_q\}_{q\in Q}$ of almost expectational maps $g_q:V\to\mathbb{R}$, and for every element v of the vocabulary set V, the function values $g_q(v)$ can be collected in a map $Q\to\mathbb{R}$. Since these maps are elements of the function space $\mathbb{R}^Q$, we can formalise the underlying idea by a map $g:V\to\mathbb{R}^Q$ with the following interpretation: For every element v of V, the element g(v) of the function space $\mathbb{R}^Q$ is the map

[172] If S is a non-empty interval of real numbers, then a map $h:S\to\mathbb{R}$ is convex iff the inequality
$h(\lambda x + (1-\lambda) y) \le \lambda h(x) + (1-\lambda) h(y)$
holds for all elements x and y of S and every $\lambda \in [0, 1]$.

$v':Q\to\mathbb{R}$ with $v'(q) = g_q(v)$ for every $q \in Q$. The following theorem summarises this definition in Equation (4.23). Now the map g can be used for an application of Mapping-Theorem 3.8, where the SSA $(V', \Pi', \Phi', H', T')$ of that theorem is an identity-based projectional SSA. The main results of this approach are stated in Assertions (v) and (vi) of the subsequent theorem.

Theorem 4.25: For a stochastic SSA (V, Π, Φ, H, T), which is based on a measurable space (Ω_0, A_0) and a map $\phi:\Omega_0\to V$, let $\{g_q\}_{q\in Q}$ be a family of maps $g_q:V\to\mathbb{R}$. Furthermore, let V' be the set of all maps $v':Q\to\mathbb{R}$ where an element v of V exists such that $v'(q) = g_q(v)$ for every $q \in Q$, and define a map $g:V\to V'$ by

$$g(v)(q) := g_q(v) \tag{4.23}$$

for every $v \in V$ and $q \in Q$.

Then the map g is surjective, and the following properties (i) and (ii) are equivalent:

(i) For every $q \in Q$, the map $g_q \circ \phi$ is A_0-B-measurable.

(ii) There exists a σ-algebra $\mathsf{A}_{V'}$ in V', where the map $g \circ \phi$ is A_0-$\mathsf{A}_{V'}$-measurable, and where all projection maps $pr_q:V'\to\mathbb{R}$ are $\mathsf{A}_{V'}$-B-measurable.

Now let $\mathsf{A}_{V'}$ be a σ-algebra with the properties specified in (ii), and Π^* a non-empty subset of Π such that the maps g_q are expectational in Π^*. Furthermore, let Π' be the set of all probability measures π' on $\mathsf{A}_{V'}$, where the expectation $\mathsf{E}_{U'\sim\pi'}\, U'(q)$ is finite, define a map $f:\Pi^*\to\Pi'$ by

$$f(\pi)(A') := \pi(\phi^{-1}(g^{-1}(A'))) \tag{4.24}$$

for every $\pi \in \Pi^*$ and $A' \in \mathsf{A}_{V'}$, and let a subset $\Pi^\sim$ of Π' be given by the definition $\Pi^\sim := f(\Pi^*)$. Then there exist unique identity-based projectional SSAs $(V', \Pi', \Phi', H', T')$ and $(V', \Pi^\sim, \Phi^\sim, H^\sim, T^\sim)$, and the following assertions hold for these SSAs:

(iii) The SSA $(V', \Pi^\sim, \Phi^\sim, H^\sim, T^\sim)$ is the $\Pi^\sim$-restriction of the SSA $(V', \Pi', \Phi', H', T')$.

(iv) The equation

$$\Phi'_{f(\pi)} = \Phi^\sim_{f(\pi)} = g(\Phi_\pi) \tag{4.25}$$

holds for every $\pi \in \Pi^*$.

(v) The equivalences $g^{-1}(S') \in T \Leftrightarrow S' \in T^\sim$ and $g^{-1}(S') \in T_e \Leftrightarrow S' \in T^\sim_e$ hold for every subset S' of V', which is an element of the σ-algebra $\mathsf{A}_{V'}$, and where the set Π^* contains all elements π of Π with $(S, \pi) \in H$.

(vi) Now assume that the map g is injective, that the sets Π^* and Π are identical, and that one of the following properties (vi.a) and (vi.b) holds for the σ-algebras A_0 and $\mathsf{A}_{V'}$:

(vi.a): The set $g(\phi(\Omega_0))$ is contained in $\mathsf{A}_{V'}$, and A_0 is the coarsest σ-algebra in Ω_0 where the map $g \circ \phi$ is measurable.

(vi.b): $\mathsf{A}_{V'}$ contains all sets $g(\phi(A))$ with $A \in \mathsf{A}_0$.

Then the map $g:V\to V'$ is an SSA-isomorphism of the SSA (V, Π, Φ, H, T) onto the SSA $(V', \Pi^\sim, \Phi^\sim, H^\sim, T^\sim)$.

Before applications of Theorem 4.25 are demonstrated, it should be noted that the equivalence (i)⇔(ii) is stated before any assumptions about expectational properties of the maps $g_q:V\to\mathbb{R}$. Due to this equivalence, the theorem works only under the assumption that the maps g_q are expectational in Π^*, whereas almost expectational maps are sufficient in Corollary 4.16 and Lemma 4.24. Since this is a rather technical issue, it is discussed (together with weakened premissas for Assertions (v) and (vi)) in a note after the proof of the theorem in Section 6.33.

Assertions (v) and (vi) of Theorem 4.25 have a noteworthy consequence for applications. Although identity-based projectional SSAs form a combination of two special subclasses of stochastic SSAs, results for such SSAs can also be used for the study of aggregation stability in other stochastic SSAs, if the properties of interest can be expressed in terms of almost expectational maps: Patterns of function values for these maps can be conceived as maps $Q \to \mathbb{R}$, and a property of interest can be represented by a set S' of all patterns indicating the presence of the property. Then the set $g^{-1}(S')$ of Assertion (v) is the set of all elements of the original vocabulary set V with the property under study.[173] Under the additional premissas of Assertion (vi), the entire set systems T and T_e can be derived (via Equations (3.43) and (3.44)) from $T^{\sim}$ and $T^{\sim}_e$. So the study of aggregation stability in stochastic SSAs, which are not projectional or not identity-based, may be supported by the interface to identity-based projectional SSAs supplied by Theorem 4.25.

For a first demonstration of this approach, we can generalise a result referring to Vincentising, whose presentation in Section 2.2.6 was limited to aggregation by convex linear combinations of quantiles. Under this limitation, an isomorphism onto an SSA with a vocabulary set V' consisting of all continuous, strictly increasing maps $]0, 1[\to \mathbb{R}$ has already been presented in Section 3.5, and this isomorphism will now be generalised. So let the vocabulary set V of an identity-based stochastic SSA (V, Π, Φ, H, T) consist of all non-decreasing maps $v:\mathbb{R} \to [0, 1]$, where a unique real number ξ with $v(\xi) = q$ exists for every $q \in]0, 1[$. With $Q :=]0, 1[$, let a family $\{g_q\}_{q \in Q}$ of maps $g_q:V \to \mathbb{R}$ be defined such that $g_q(v)$ is the said unique number ξ with $v(\xi) = q$. Furthermore, let A_V be the coarsest σ-algebra in V, where all maps g_q are measurable, and let Π be the set of those probability measures π on A_V, where all expectations $\mathsf{E}_{U \sim \pi}\, g_q(U)$ with $q \in Q$ are finite. Then the basic approach of Vincentising is formalised by an aggregation rule $\Phi:\Pi \to V$ requiring that the maps g_q are expectational, which means that the equation

$$\Phi_\pi(\mathsf{E}_{U \sim \pi}\, g_q(U)) = q \tag{4.26}$$

holds for every $\pi \in \Pi$ and every $q \in Q$. The existence and the uniqueness of an element Φ_π of V with this property is proved in Section 6.34. Finally, the components H and T of an identity-based SSA (V, Π, Φ, H, T) are given by SSA-Axioms (vii) and (iv).

Now the definitions in Theorem 4.25 lead to the set $V' \subseteq \mathbb{R}^Q$ and the map $g:V \to V'$ which have already been introduced in Section 2.2.6. In particular, recall that the map g is bijective.[174] With the coarsest σ-algebra $\mathsf{A}_{V'}$ in V' where all projection maps $pr_q:V' \to \mathbb{R}$ are measurable, everything is prepared to derive from Theorem 4.25.(vi) that g is an SSA-Isomorphism.

Another application of Theorem 4.25 demonstrates how it can be used for the study of stochastic SSAs, which are projectional, but not identity-based. The SSAs (V, Π, Φ, H, T) and (V', Π', Φ', H', T') of the theorem can be the empirical and the analytical CDF-SSA; but the σ-algebra $\mathsf{A}_{V'}$ of the analytical CDF-SSA can also be replaced by a finer one to fulfill the premissa $S' \in \mathsf{A}_{V'}$ in Assertion (v).[175] Although the resulting situation has already been shown to be covered

---

[173] As demonstrated for a similar situation at the end of Subsection 4.3, the premissa $S' \in \mathsf{A}_{V'}$ can be covered by the choice of a suitably fine σ-algebra $\mathsf{A}_{V'}$.

[174] See Section 6.34 for a more detailed analysis.

[175] However, since the σ-algebra $\mathsf{A}_{V'}$ in the analytical CDF-SSA is the coarsest one where all (continued...)

by Corollary 4.11, a contribution of Theorem 4.25 can be seen in Equation (4.25), which had to be introduced as a premissa (in the form of Equation (4.10)) for Corollary 4.11. Observe also that Theorem 4.25 doesn't completely specify the σ-algebra $\mathsf{A}_{V'}$, but only a lower bound of its fineness.[176] As in former considerations referring to Corollary 4.11, this flexibility can be used to fulfill the premissas of Theorem 4.25.(v) and (vi) referring to the σ-algebra $\mathsf{A}_{V'}$.

The injectivity of the map $g{:}V \to V'$ in Theorem 4.25 is of interest not only as a premissa of Assertion (vi), but also for the aggregation rule Φ in the first SSA: It is specified by Equation (4.25), if the map g is injective. Whereas this injectivity is trivial for situations where it is an identity map (as in the above application to the CDF-SSAs), a necessary and sufficient condition is stated by the following corollary:[177]

Corollary 4.26: In the situation of Theorem 4.25, the map $g{:}V \to V'$ is injective iff the family of maps $\{g_q\}_{q \in Q}$ is V-separating, i.e., if and only if the equivalence

$$(\forall q \in Q{:}\ g_q(v_1) = g_q(v_2)) \;\Leftrightarrow v_1 = v_2 \tag{4.27}$$

holds for all elements v_1 and v_2 of V.

In other words, the map g is injective iff for different elements v_1 and v_2 of V there is at least one element q of the index set Q such that $g_q(v_1) \neq g_q(v_2)$. It is easily verified that this property is present in all projectional SSAs, and typically the same holds for all stochastic SSAs where the aggregation rule is fully specified by the almost expectational maps $\{g_q\}_{q \in Q}$.[178]

---

[175] (...continued)
projection maps $pr_q{:}V' \to \mathbb{R}$ are measurable, the transition to an even coarser σ-algebra would violate the premissa in Theorem 4.25 requiring the measurability of these projection maps.

[176] As long as the σ-algebra A_0 is considered fixed, the assumed A_0-$\mathsf{A}_{V'}$-measurability of the map $g \circ \phi$ implies an upper bound for the fineness of $\mathsf{A}_{V'}$, too. But in many applications, there will be some flexibility in the choice of the σ-algebra A_0.

[177] Proof: If Equivalence (4.27) holds for all elements v_1 and v_2 of V and v_1 and v_2 are different elements of V, then the inequality $g_q(v_1) \neq g_q(v_2)$ must hold for some $q \in Q$, and then $g(v_1) \neq g(v_2)$ follows from the definition of the map $g{:}V \to V'$. But if $v_1 \neq v_2$ implies $g(v_1) \neq g(v_2)$ for all elements v_1 and v_2 of V, then the map g is injective. Conversely, if g is injective (i.e., if $v_1 \neq v_2$ implies $g(v_1) \neq g(v_2)$), then the forward implication in Equivalence (4.27) follows again from the definition of the map g, and the backward implication is trivial.

[178] If there would be an element π of the set Π and two different elements v_1 and v_2 of V such that the equation

$$g_q(v_1) = g_q(v_2) = \mathsf{E}_{U \sim \pi}\, g_q(\phi_U)$$

holds for every $q \in Q$, then the maps g_q wouldn't specify a unique element Φ_π of the vocabulary set. However, violations of Equivalence (4.27) for elements v_1 and v_2 of the set $V \setminus \Phi(\Pi)$ do not exclude that Φ_π is uniquely determined by the maps $\{g_q\}_{q \in Q}$ for every $\pi \in \Pi$. In other words, the (continued...)

If the map g is non-injective, it may be worthwile to try the approach of Corollary 3.13: An equivalence relation ~ on the vocabulary set V given by $v_1 \sim v_2 :\Leftrightarrow g(v_1) = g(v_2)$ for all elements v_1 and v_2 of V can be used for the transition to a new vocabulary set $V^* := V/\sim$. Then Equation (4.25) and results of Subsection 4.3 may supply the premissas of Corollary 3.13.

For applications of Theorem 4.25, the properties claimed by its Assertions (v) and (vi) have the greatest immediate relavance. But such applications are confined by the assumption $S' \in \mathsf{A}_{V'}$ resp. by the additional premissa that either (vi.a) or (vi.b) must hold for the σ-algebras A_0 and $\mathsf{A}_{V'}$. If these limitations impede the study of aggregation stability for a set $g^{-1}(S')$, the following corollary yields the premissas for an application of Corollary 4.9:

Corollary 4.27: For a stochastic SSA (V, Π, Φ, H, T), which is based on a measurable space (Ω_0, A_0) and a map $\phi:\Omega_0 \to V$, let Π^* be a subset of Π, and $\{g_q\}_{q \in Q}$ a family of maps $g_q:V \to \mathbb{R}$, which are expectational in Π^*. Furthermore, define a set V' and a map $g:V \to V'$ as in Theorem 4.25, and a map $\phi':\Omega_0 \to V'$ by $\phi' := g \circ \phi$.
Then there exists a unique projectional SSA (V', Π', Φ', H', T') with Π' = Π, which is based on the measuruble space (Ω_0, A_0) and the map ϕ'. In this SSA, the equation

$$\Phi'_\pi = g(\Phi_\pi)$$

holds for every $\pi \in \Pi^*$.

See ### for a proof of the corollary and for weakened premissas.

4.7 Linear Stochastic Structures of Aggregation Stability

The objective of the present subsection has been prepared by two approaches to SSAs, whose vocabulary set is a set of maps $v:Q \to \mathbb{R}$ (i.e., a subset of the function space $\mathbb{R}^Q$), where Q may be any non-empty set. In the review of examples from the pertinent literature in Section 2, it has turned out helpful to conceive the set of all maps $Q \to \mathbb{R}$ as a real vector space $\mathbb{R}^Q$. In these operations, the elements of the vocabulary set are treated as wholes. Another approach to such vocabulary sets is based upon the simultaneous considerations of many expectational maps $V \to \mathbb{R}$. For instance, projectional SSAs are based on the assumption that all projection maps are expectational (Definition 4.15). Furthermore, results of Section 4.5 could be used to derive expectational properties of other maps (e.g. the maps $h_k:V \to \mathbb{R}$ and $g_q:V \to \mathbb{R}$ in Example 4.18 or the maps $g_{cn}:V \to \mathbb{R}$ and $g_c:V \to \mathbb{R}$ in Corollary 4.22), and these maps turned out even more useful than the projection maps in the study of complex problems of aggregation stability.

The vector space approach and the analysis by expectational maps are closely related. Since the vector space structure of $\mathbb{R}^Q$ is based upon the operations of pointwise addition and pointwise multiplication by scalars, we can also say that convex linear combinations of elements of the vocabulary set are based upon expectations of projection maps. Conversely, the analysis of parametric families of functions in Example 4.18 was based upon expectational maps $h_k:V \to \mathbb{R}$ and

---

178 (...continued) aggregation rule is fully determined by these expectational maps iff Equivalence (4.27) holds whenever v_1 or v_2 is identical with Φ_π for some $\pi \in \Pi$.

g_q:V→ℝ, whose relevance for the considered function family was obtained by means of the vector space approach underlying Lemma 2.9.

The present subsection adds another concept integrating the two approches. Before the concept is introduced formally, recall that many properties of real vector spaces can be expressed in terms of linear maps, and many expectational maps are restrictions of linear maps to the vocabulary set of a stochastic SSA, this vocabulary set being a subset of a real vector space. For instance, the projection maps pr_q:V→ℝ in a projectional SSA (V, Π, Φ, H, T) with V ⊆ $\mathbb{R}^Q$ and q ∈ Q can be conceived as restrictions to V of a projection map pr_q:$\mathbb{R}^Q$→ℝ, which is again based on the defining equality pr_q(v) := v(q). The only difference between the two projection maps lies in the domain, which is the set V for the projection map pr_q:V→ℝ, whereas the entire function space $\mathbb{R}^Q$ is the domain of the projection map pr_q:$\mathbb{R}^Q$→ℝ. We can also say that the projection map pr_q:V→ℝ is the restriction to V of the projection map pr_q:V→ℝ.[179]

Now it is easily verified that the projection maps pr_q:$\mathbb{R}^Q$→ℝ are linear: For elements x and y of $\mathbb{R}^Q$ (i.e., for maps x:Q→ℝ and y:Q→ℝ) and for every real number ξ, the equalities $pr_q(x + y) = pr_q(x) + pr_q(y)$ and $pr_q(\lambda \cdot x) = \lambda \cdot pr_q(x)$ follow immediately from the pointwise addition and multiplication by scalars.[180] So the expectational projection maps pr_q:V→ℝ are restrictions to V of the linear projection map pr_q:V→ℝ. References to such situations become more convenient by an extension of the concept of expectational maps, whose explication by Definition 4.13 was confined to maps V→ℝ.

Definition 4.28: Let (V, Π, Φ, H, T) be a stochastic SSA, Π* a subset of Π, and X a set such that V is a subset of X. Furthermore, let g:V→ℝ and g':X→ℝ be maps such that the map g is the restriction to V of the map g'. Then the map g' is (almost) expectational (in Π*) iff the map g is (almost) expectational (in Π*).

Transferring the role of the set X in the definition to the set $\mathbb{R}^Q$ in a projectional SSA with V ⊆ $\mathbb{R}^Q$, we can also say that the projection maps pr_q:$\mathbb{R}^Q$→ℝ are expectational, since their restrictions to V (i.e., the projection maps pr_q:V→ℝ) are expectational.

Another property of the projection maps pr_q:$\mathbb{R}^Q$→ℝ in a projectional SSA (V, Π, Φ, H, T) with V ⊆ $\mathbb{R}^Q$ is also basic for the following conceptualisation. If we conceive therm as members of a family $(pr_q\}_{q \in Q}$ of maps pr $_q$:$\mathbb{R}^Q$→ℝ, then this family is V-separating. (See Section ?.? for this concept.) More explicitly, if v' and v" are non-identical elements of the set V, then there is an

[179] See Section ?.? for the definition of the restriction of maps, including the note referring to the practice of using the same symbol for a map and its restriction.

[180] More explicitly, if z = x + y, then the third equality in

$pr_q(x + y) = pr_q(z) = z(q) = x(q) + y(q) = pr_q(x) + pr_q(y)$

follows from the pointwise addition of maps, whereas the second equality and the last one are based on the definition of projection maps. Similarly, the pointwise multiplication of maps by scalars leads to

$pr_q(\lambda \cdot x) = pr_q(w) = w(q) = \lambda \cdot x(q) = \lambda \cdot pr_q(x)$

for w = λ · x.

element q of Q with $pr_q(v') \neq pr_q(v'')$. In other words, the projection maps $pr_q:\mathbb{R}^Q \to \mathbb{R}$ form a V-separating family $(pr_q\}_{q \in Q}$ of maps.

###In the present draft-version, the remainder of Section 4.7 has to be considered as an outline consisting of definitions and results, which will have to be commented.

Definition 4.29: A map $g:V\to\mathbb{R}$ is linearly extendable iff V is a subset of a real vector space E and the map g is the restriction to V of a linear map $E\to\mathbb{R}$.

Definition 4.30: If E is a real vector space, then a stochastic SSA (V, Π, Φ, H, T) is linear in E iff it has the following properties:

(i) The vocabulary set V is a subset of the vector space E.

(ii) There exists a V-separating set of linearly extendable and expectational maps $V\to\mathbb{R}$.

Furthermore, a stochastic SSA is linear iff there is a real vector space E such that the SSA is linear in E.

Theorem 4.31: Every projectional SSA is linear, and every linear stochastic SSA is SSA-isomorphic with a projectional SSA.

Lemma 4.32: Let E be a finite dimensional real vector space, and (V, Π, Φ, H, T) a stochastic SSA, which is linear in E. Then all linear maps $g:E\to\mathbb{R}$ are expectational in Π.

Lemma 4.28: In a linear stochastic SSA (V, Π, Φ, H, T), which is based on a measurable space (Ω_0, A_0) and a map $\phi:\Omega_0\to V$, let π be an element of Π, and ω an element of Ω_0 such that $\pi = \varepsilon_\omega$. Then $\Phi_\pi = \phi_\omega$.

Check:

Lemma 4.33: Let E be a real vector space, and (V, Π, Φ, H, T) a stochastic SSA, which is linear in E. Furthermore, let F be a linear subspace of E, V' a subset of $V \cap F$, and Π' a subset of Π such that $\Phi_\pi \in V'$ and $(V', \pi) \in H$ for every $\pi \in \Pi'$. Then the V'-Π'-restriction of (V, Π, Φ, H, T) is linear in F.

Conceptual: Convex linear combinations of elements of Π: $\sum_{i=1..n} \lambda_i \pi_i$ is a map $\pi:\mathsf{A}_0\to\mathbb{R}$ with $\pi(A) = \sum_{i=1..n} \lambda_i \pi_i(A)$ for every $\pi \in \mathsf{A}_0$.
Fact: π is a probability measure on A_0.
Concept: Convex set Π.

Lemma 4.34: In a linear stochastic SSA (V, Π, Φ, H, T), let $\{\pi_i\}_{i=1..n}$ be a finite sequence of elements of Π, and π and element of Π such that $\pi = \sum_{i=1..n} \lambda_i \pi_i$ with real numbers λ_i.
Then

$$\Phi_\pi = \sum_{i=1..n} \lambda_i \Phi_{\pi(i)}.$$

Theorem 4.35: Let (V, Π, Φ, H, T) be a linear stochastic SSA with the following properties:

(i) The set Π is convex.

(ii) Φ(Π) = V. (I.e., for every $v \in V$, the equality $\Phi_\pi = v$ holds for some $\pi \in \Pi$.)

Then the set system T contains only convex subsets of V.

Examples for (ii):

- $\phi(\Omega_0) = V$, and Π contains all Dirac measures e_ω with $\omega \in A_0$.

In particular:

- Identity-based, and Π contains all Dirac measures e_v with $v \in V$.

###The following paragraph, which was originally placed in another section, has to be adapted.

As a last application of expectional maps, it will now be demonstrated how the aggregation by convex linear combination, which was the objective of Section 2, can be subsumed under the theory of identity-based stochastic SSAs with expectational maps. We resume the SSA (V, Π, Φ, H, T) described in Lemma 2.1, whose vocabulary set V is a subset of a real vector space, Π being the set of all finite sequences $\pi = \{\lambda_i, v_i\}_{i=1..n(\pi)}$ with $\lambda_i \in \mathbb{R}$ and $v_i \in V$, and the aggregation rule being given by $\Phi_\pi := \sum_{i=1..n(\pi)} \lambda_i v_i$ for very such π. With the notation E for the real vector space including V, let $\{g_q\}_{q\in Q}$ be a V-separating[181] family of maps $g_q:V\to\mathbb{R}$, which can be extended to linear maps $g'_q:E\to\mathbb{R}$. Now consider a second, identity-based SSA (V', Π', Φ', H', T') with V' = V and a σ-algebra $A_{V'}$ in V', where all maps $g_q:V\to\mathbb{R}$ are $A_{V'}$-B-measurable. With a set Π' consisting of all probability measures on $A_{V'}$ with a representation as a convex linear combination of Dirac-measures,[182] there is a unique aggregation rule Φ':Π'→V', where all maps $g_q:V\to\mathbb{R}$ are expectational, and then the relation H' and the set systems T' and T'_e are determined by SSA-Axioms (vii) and (iv) and Equation (1.1). Then it is easily verfied that the identity map in V is an SSA-isomorphism of (V, Π, Φ, H, T) onto (V', Π', Φ', H', T').

Lemma 4.36: Let E be a finite dimensional real vector space, and (V, Π, Φ, H, T) a stochastic SSA,

[181] See Section ?.? for the definition of a separating family of maps.

[182] In the assumed situation, the Dirac-measure ε_v in an element v of the set V is the unique probability measure on $A_{V'}$ with $\varepsilon_v(A) = 1$ for every $A \in A_{V'}$. Then a probablity measure π' on $A_{V'}$ has a representation as a convex linear combination of Dirac-measures on $A_{V'}$ iff there are non-negative real numbers $\{\lambda_i\}_{i=1..n}$ and elements $\{v_i\}_{i=1..n}$ of V' such that $\pi' = \sum_{i=1..n} \lambda_i \varepsilon_{v(i)}$, i.e., $\pi'(A) = \sum_{i=1..n} \lambda_i \varepsilon_{v(i)}(A)$ for every $A \in A_{V'}$. In this situation, the requirement that all maps $g_q:V'\to\mathbb{R}$ are expectational and have an extension to a linear map $E\to\mathbb{R}$ leads to $g_q(\Phi'_{\pi'}) = \sum_{i=1..n} \lambda_i v_i$ for every $q \in Q$, and this property specifies a unique element $\Phi'_{\pi'}$ of V', since the family $\{g_q\}_{q\in Q}$ is assumed to be V-separating.

which is linear in E.
Then all linear linear maps $E \to \mathbb{R}$ are expectational.

Theorem 4.37: Let E be a finite dimensional real vector space, and (V, Π, Φ, H, T) a stochastic SSA, which is linear in E.
Then the set system T contains all convex subsets of V.

Corollary 4.38: Let E be a real vector space, and (V, Π, Φ, H, T) a stochastic SSA, which is linear in E. Furthermore, let F be a finite dimensional linear subspace of E, and $\{g_q\}_{q \in Q}$ a family of linear maps $g_q{:}E \to \mathbb{R}$ with the following properties:
(i) $V \cap F$ generates F; i.e., every element of F is a linear combination of elements of $V \cap F$.
(ii) For every $q \in Q$, the restriction to V of the map g_q is expectational.
(iii) F is the set of all elements x of the vector space E with $g_q(x) = 0$ for every $q \in Q$.
Then the set system T contains all convex subsets of $V \cap F$.

Probably simpler and more general:

Corollary 4.39: Let E be a real vector space, and (V, Π, Φ, H, T) a stochastic SSA, which is linear in E. Furthermore, let F be a finite-dimensional linear subspace of E, and V' a subset of $V \cap F$, which is contained in T.
Then T contains all convex subsets of V'

Application:

Example 4.40:
Estes: ψ(S) with convex S.; $V' = V \cap \psi(\mathbb{R}^m)$. (i.e., V' = S* of Example 4.18)

To be continued

Send email to iseler@zedat.fu-berlin.de if you want to be informed about later versions, and communicate the 'date and time of present version' on page 1.

Further provided sections:

4.8 Expectations of Random Variables in Locally Convex Hausdorff Spaces

5 Applications

5.1 Models for the Random Selection and Observation of Units

5.2 Singular and Universal Aggregate Hypotheses

5.3 Hypothesis Testing in Deliberately Biased Samples

5.4 Explication and Deduction in the Derivation of Statistical Hypotheses

5.5 Hypotheses for Quasi-Experimentation

6 Appendix: Proofs and Other Mathematical Details

6.1 Preliminaries

6.2 Characterisation of Aggregates by Subsets of a Vocabulary Set

In Comment e to Definition 1.2, it has been pointed out that the formalisation of an aggregation rule by a map $\Phi:\Pi\rightarrow V$ excludes situations, where aggregates are not charectarised by unique elements of the vocabulary set V, but by subsets of V. We will now demonstrate by an example an approach, which has been only outlined in that comment: The transition to a new set Π' consisting of ordered pairs like (π, v_1), (π, v_2) etc..

Let Ω be a finite set of potential outcomes of a random process, and V a set of ordered pairs (P, y) representing games of hazard based on these outcomes such that P is a probability measure on $\mathsf{P}\Omega$ (the power set of Ω), and y a payoff-function $\Omega\rightarrow\mathbb{R}$ with zero expectation under P. (I.e., the games are objectively 'fair'.) Furthermore, let Π be a system of nonempty subsets of Ω, and let the map $\Phi:\Pi\rightarrow\mathsf{P}V$ be such that Φ_π is the set of all games where the probability of obtaining an outcome belonging to the chosen subset π is 1. Now assume that a participant of the game is allowed to choose an element π of Π with the effect that his opponent can select an element of Φ_π (i.e., a game), which will be played. In this situation, we do not have a map $\Phi:\Pi\rightarrow V$, but a suitable reformalisation of this situation could be based on a redefined set Π'. To fulfill the assumptions of Definition 1.2 about the aggregation rule Φ and the relation H, we would also have to introduce a map $\Phi':\Pi'\rightarrow V$ and a relation $H' \subseteq \mathsf{P}V \times \Pi'$, e.g. by the following definitions:

$\Pi' := \{(\pi, v): \pi \in \Pi \wedge v \in \Phi_\pi\}$,

$\Phi'(\pi, v) := v$,

and

$H' := \{(S, (\pi, v)): S \subseteq V \wedge (S, \pi) \in H\}$.

To prepare a question of aggregation stability, we have to give an interpretation to the original relation H. So let the subsets S_1 and S_2 of V consist of those games, which are 'attractive' for player 1 resp. 2. (Of course, the above assumption of objective fairness doesn't exclude variations in the attractiveness due to preferences for more or less risky games.) Then we may define a relation H as follows: An ordered pair (S, π) is an element of H, iff all elements of Φ_π, which are attractive for player 2, are elements of S (i.e., iff $\Phi_\pi \cap S_2 \subseteq S$). It is left to the reader to find out (for the revised formalisation with Π', Φ' and H')

a) a condition which is necessary to fulfill the SSA-Axioms,

b) for each one of the sets S_1 and S_2, a sufficient conditions for its stability under aggregation.

Note that the reformalisation would be unnecessary, if the definition of an SSA would assume a map $\Phi:\Pi\rightarrow\mathsf{P}V$ such that situations with a set Φ_π containing exactly one element would be a special case. As a consequence, the expression $\Phi_\pi \in S$ in SSA-Axiom (iv) would have to be changed into $\Phi_\pi \subseteq S$. However, Definition 1.2 covers most cases of interest in psychology, and it can be left to the reader to verify, that the set system T defined by the revised SSA-Axiom (iv) would be kept unchanged by a transition to Π', Φ' and H' with the above definitions. Hence, it wouldn't materially pay off to sacrifice the greater simplicity of Definition 1.2, where the aggregation rule is a map $\Phi:\Pi\rightarrow V$.

6.3 *Proof of Lemma 1.3*

(i): If a set S is contained in T_e, then the equivalence $(S, \pi) \in H \Leftrightarrow \Phi(\pi) \in S$ must hold for every $\pi \in \Pi$, and then the implication $(S, \pi) \in H \Rightarrow \Phi(\pi) \in S$ must also hold for every $\pi \in \Pi$.
(ii): For the empty set, we obtain $(\varnothing, \pi) \notin H$ from SSA-Axiom (i), and the property $\Phi_\pi \notin \varnothing$ is obvious. Hence $\varnothing \in T_e$. Similarly, $(V, \pi) \in H$ follows from SSA-Axiom (ii), and $\Phi_\pi \in V$ from the assumption that Φ is a map $\Phi:\Pi \to V$, leading to $V \in T_e$. Furthermore, Assertion (i) allows to derive $\varnothing \in T$ and $V \in T$ from $\varnothing \in T_e$ and $V \in T_e$.
(iii): It is obvious that the implication $(S, \pi) \in H \Rightarrow \Phi(\pi) \in S$ holds for every $\pi \in \Pi$ under the premissa $\Phi(\Pi) \subseteq S$.
(iv): If the implication $(S, \pi) \in H \Rightarrow \Phi_\pi \in S$ holds for very $\pi \in \Pi$ (i.e., if $S \in T$), then this implication can be sharpened to an equivalence, if and only if the reversed implcation holds as well.
(v): Under the assumption $H \subseteq H'$, let S be a subset of V with $S \in T'$. It suffices to prove that the property $\Phi_\pi \in S$ holds for every $\pi \in \Pi$ with $(S, \pi) \in H$: Then $S \in T$ is granted by SSA-Axiom (iv). But under the premissa $H \subseteq H'$, the relation $(S, \pi) \in H$ implies $(S, \pi) \in H'$, and then $\Phi_\pi \in S$ follows from the assumption $S \in T'$ by another reference to SSA-Axiom (iv).□

6.4 *A Generalisation of Convexity*

The generalisation of convexity to be presented in this subsection is based on the following idea: A subset S of a real vector space is convex iff all linear combinations of elements of S with coefficients contained in the interval [0, 1] and summing up to 1 are elements of S. But in some kinds of aggregation, only a subset of the interval [0, 1] is available for coefficients in linear combinations. E.g., the averaging of elements of a vector space can be conceived as a linear combination, where the coefficients are rational numbers contained in the interval [0, 1] and summing up to 1. The subsequent lemma specifies a class of subsets of the interval [0, 1] such that a central property of convex sets is conserved: If Λ is a set with the property specified in the premissa of the lemma, then Assertions (i) and (ii) are equivalent for every subset S of a real vector space.

Lemma 6.1: Let Λ be a subset of the interval [0, 1] with the following property: For every finite sequence $\{\lambda_i\}_{i=1..n}$ of elements Λ with $0 < \sum_{i=1..n} \lambda_i \leq 1$, the numbers $1 - \lambda_i$ and $\lambda_i / \sum_{j=1..n} \lambda_j$ with i = 1..n are also elements of Λ. Then the following properties are equivalent for every subset S of a real vector space:

(i) For all elements x_1 and x_2 of S and every element λ of Λ, the vector $\lambda\, x_1 + (1-\lambda)\, x_2$ is an element of S.

(ii) For every natural number n, every sequence $\{x_i\}_{i=1..n}$ of elements of S and every sequence $\{\lambda_i\}_{i=1..n}$ of elements of Λ with $\sum_{i=1..n} \lambda_i = 1$, the vector $\sum_{i=1..n} \lambda_i\, x_i$ is an element of S.

<u>Poof of Lemma 6.1</u>

It suffices to verify the implications (i)⇒(ii)⇒(i) for a given subset Λ of the interval [0, 1] with the properties specified in the premissa of the lemma, and an arbitrary subset S of a real vector space.

(i)⇒(ii): We will prove (ii) by induction over n under the assumed validity of (i). For $n = 1$, the validdity of (ii) is trivial, since the assumption $\sum_{i=1..n} \lambda_i = 1$ implies $\lambda_1 = 1$. Now let k be an arbitrary natural number, and assume that (ii) holds for $n = k$. For $n = k + 1$, let sequences $\{x_i\}_{i=1..n}$ and $\{\lambda_i\}_{i=1..n}$ be given with the properties assumed in (ii). For $\lambda_n = 1$, the claim of (ii) is again trivial, since $\lambda_i = 0$ follows for $i = 1..n-1$ from the assumptions $\lambda_i \in \Lambda$ (which implies $\lambda_i \geq 0$) and $\sum_{i=1..n} \lambda_i = 1$. For $\lambda_n < 1$, define a sequence $\{\gamma_i\}_{i=1..n-1}$ of real numbers by

$$\gamma_i := \lambda_i / \textstyle\sum_{j=1..n-1} \lambda_j, \tag{6.1}$$

and observe that the numbers γ_i are elements of the set Λ with $\sum_{i=1..n-1} \gamma_i = 1$. Furthermore, define a vector x' by

$$x' := \textstyle\sum_{i=1..n-1} \gamma_i \, x_i. \tag{6.2}$$

Then the assumed validity of (ii) for $n = k$ yields $x' \in S$. Furthermore, the definition of x' implies

$$\textstyle\sum_{i=1..n} \lambda_i \, x_i = \lambda_n \, x_n + (1-\lambda_n) \cdot x'. \tag{6.3}$$

But an application of (i) to the right hand side of this equation shows that the result of the linear combination on the left hand side is an element of S.
(ii)⇒(i): For a given element λ of the set Λ and elements x_1 and x_2 of S, define $\lambda_1 := \lambda$ and $\lambda_2 := 1 - \lambda$. Then the definition of the set Λ yields $\lambda_2 \in \Lambda$, and a reference to the assumed validity of (ii) completes the proof.□

The properties required for a set Λ by the first premissa of Lemma 6.1 are easily verified for situations where Λ is the entire interval [0, 1] or the set of all rational numbers contained in that interval. So stability under convex linear combinations and stability under averaging are special cases of the kind of stability treated by the lemma.

Qv 6.5
6.5 Proof of Lemma 1.4

In the situation assumed in Lemma 1.4, let S be an arbitrary subset of V. It suffices to verify the implications (i)⇒(ii)⇒(iii)⇒(i).

(i)⇒(ii): Assume that S is an element of T, let v_1 and v_2 be elements of S, and λ a rational number such that $0 \leq \lambda \leq 1$. Furthermore, let j be an integer and k a natural number such that $\lambda = j/k$. (Certainly, such numbers exist, if λ is a rational number.) Then the assumption $0 \leq \lambda \leq 1$ can be rewritten as $0 \leq j \leq k$. Now consider the element $\pi' = \{v'_i\}_{i=1..n(\pi')}$ of Π given by $n(\pi') := k$, $v'_i := v_1$ for $1 \leq i \leq j$, and $v'_i := v_2$ for $j < i \leq k$. These definitions imply $\Phi_{\pi'} = \lambda \cdot v_1 + (1-\lambda) \cdot v_2$; hence it suffices to verify $\Phi_{\pi'} \in S$. But $(S, \pi') \in H$ follows from the definition of H in Lemma 1.4 and the assumption that v_1 and v_2 are elements of S; hence $\Phi_{\pi'} \in S$ is granted by the assumption $S \in T$ and SSA-Axiom (iv).

(ii)⇒(iii): This implication follows immediately from Lemma 6.1, the set of all rational numbers in the interval [0, 1] taking the role of the set Λ in that lemma.

(iii)⇒(i): According to SSA-Axiom (iv), we have to verify that $(S, \pi) \in H$ implies $\Phi_\pi \in S$ for every

$\pi \in \Pi$. So let an arbitrary element $\pi = \{v_i\}_{i=1..n(\pi)}$ of Π be given such that $(S, \pi) \in H$. Then the definition of the relation H in Lemma 1.4 implies $v_i \in S$ for $i = 1..n(\pi)$, and with $\lambda_i := n(\pi)^{-1}$ for $i = 1..n(\pi)$ we can write $\Phi_\pi = \sum_{i=1..n(\pi)} \lambda_i \cdot v_i$. But then $\Phi_\pi \in S$ follows, if (iii) holds.□

6.6 *Proof of Lemma 2.1*

Certainly all convex subsets of V must be elements of T, since every convex set is closed under convex linear combinations of its elements. Conversely, if S is a non-convex subset of V, there are elements v_1 and v_2 of S and a scalar $\lambda \in [0, 1]$ such that $\lambda\, v_1 + (1-\lambda)\, v_2 \notin S$. In other words, $\Phi_\pi \notin S$ for $\pi := \{(\lambda_i, v_i)\}_{i=1..2}$ with $\lambda_1 := \lambda$ and $\lambda_2 := 1-\lambda$. But since (S, π) is an element of H, SSA-Axiom (iv) leads to $S \notin T$.□

6.7 *Proof of Estes-Theorem 2.4*

Troughout this proof, let a map $\psi_0:\mathbb{R}^m \to \mathbb{R}^Q$ be defined by

$$\psi_0(\theta) := \psi(\theta) - \psi(\eta_0). \tag{6.4}$$

The equivalence of Assertions (i) through (vi) will be established by proofs of the implications (ii)⇒(i)⇒(ii)⇒(iv)⇒(v)⇒(vi)⇒(iii)⇒(ii).

The further claims of the theorem will be side results.

(ii)⇒(i): If $\{\gamma_k\}_{k=0..m}$ is a family of maps $\gamma_k Q \to \mathbb{R}$ such that Equation (2.8) holds, then the first order partial derivative of $f(q, \theta)$ with respect to $\theta(k)$ is $\gamma_k(q)$. So this derivative is independent of every component $\theta(k')$ of θ, and Equation (2.11) follows immediately.

(i)⇒(ii): Assertion (i) implies that the first order partial derivatives of $f(q, \theta)$ for components of θ may depend on q, but must be independent of all components of θ. So let maps $\gamma_k:Q \to \mathbb{R}$ for $k = 1..m$ be given such that $\gamma_k(q)$ is the first order partial derivative of $f(q, \theta)$ with respect to $\theta(k)$. Furthermore, let a map $\gamma':Q\times\mathbb{R}^m \to \mathbb{R}$ be defined by

$$\gamma'(q, \theta) := f(q, \theta) - \sum_{k=1..m} \theta(k) \cdot \gamma_k(q). \tag{6.5}$$

Now the definition of the maps γ_k implies that all first order partial derivatives of $\gamma'(q, \theta)$ with respect to components of θ must be zero. In other words, $\gamma'(q, \theta)$ may depend on q, but must be independent of θ. So we can take an arbitrary element θ^* of $\mathbb{R}^m$, define $\gamma_0(q) := \gamma'(\theta^*, q)$, and we obtain Equation (2.8).

(ii)⇒(iv): For given maps $\gamma_k:Q \to \mathbb{R}$ underlying Assertion (ii), Equation (2.16) rewrites Equation (2.8) in terms of (pointwise) addition and scalar multiplication of whole maps. Inserting a given linear combination $\sum_{i=1..n} \lambda_i\, \theta_i$ for θ in Equation (2.16), we obtain Equation (2.12) by a rearrangement of sums. Side result: An application of Equation (2.16) to the vector η_k leads to Equations (2.14) and (2.15).

(iv)⇒(v): The assumption $\sum_{i=1..n} \lambda_i = 1$ can be rewritten as $1 - \sum_{i=1..n} \lambda_i = 0$. But the addition of $0\, \psi(\eta_0)$ in Equation (2.12) can be neglected.

(v)⇒(vi): For given elements θ_1 and θ_2 of $\mathbb{R}^m$ and $\lambda \in [0, 1]$, apply Equation (2.10) with $n := 2$,

$\lambda_1 := \lambda$, and $\lambda_2 := 1-\lambda$.
(vi)⇒(iii): It suffices to prove the linearity of the map ψ_0 defined by Equation (6.4), i.e., the properties

$$\psi_0(\xi\ \theta) = \xi\ \psi_0(\theta) \tag{6.6}$$

and

$$\psi_0(\theta_1 + \theta_2) = \psi_0(\theta_1) + \psi_0(\theta_2). \tag{6.7}$$

Equation (6.6) will be verified separately for $\xi = 0$, $\xi < 0$, and $\xi > 0$.
For $\xi = 0$, Equation (6.6) reduces to

$$\psi_0(\eta_0) = \gamma_z, \tag{6.8}$$

since $0\ \theta = \eta_0$ and $0\ \psi_0(\theta) = \gamma_z$ are granted by the definitions of η_0 and γ_z. But the validity of Equation (6.8) follows immediately, if we replace the vector θ in Equation (6.4) by η_0. (The definition of γ_z implies $\psi(\eta_0) - \psi(\eta_0) = \gamma_z$.
For $\xi < 0$, define $\lambda := (1-\xi)^{-1}$, $\theta_1 := \xi\ \theta$ and $\theta_2 := \theta$, leading to

$$\lambda\ \theta_1 + (1-\lambda)\ \theta_2 = (1\text{-}\xi)^{-1} \cdot (\xi\ \theta - \xi\ \theta) = \eta_0. \tag{6.9}$$

Combining this result with Equation (6.8), we obtain the first line of

$$\begin{aligned}\gamma_z\ &= \psi_0(\lambda\ \theta_1 + (1-\lambda)\ \theta_2)\\ &= \psi(\lambda\ \theta_1 + (1-\lambda)\ \theta_2) - \psi(\eta_0)\\ &= \lambda\ \psi(\theta_1) + (1-\lambda)\ \psi(\theta_2) - \psi(\eta_0)\\ &= (1-\xi)^{-1}\ (\psi_0(\xi\ \theta) - \xi\ \psi_0(\theta)),\end{aligned} \tag{6.10}$$

the second equality being based on Equation (6.4), and the third one on Equation (2.13), whose validity may be assumed in the proof of the implication (vi)⇒(iii). The last line is obtained, if we replace λ, θ_1 and θ_2 according to their definitions, apply Equation (6.4), and rearrange terms. Now the factor $(1\text{-}\xi)^{-1}$ is certainly greater than 0 for $\xi < 0$; so the last line of Equation (6.10) can be equal to γ_z (the zero-element of $\mathbb{R}^Q$) only if Equation (6.6) holds.
For $\xi^* > 0$, the last two equalities in

$$\psi_0(\xi^*\ \theta) = \psi_0((\text{-}1)\ (-\xi^*)\ \theta) = (-1)\ \psi_0((-\xi^*)\ \theta) = (-1)\ (-\xi^*)\ \psi_0(\theta) \tag{6.11}$$

result from an application of Equation (6.6), where ξ is replaced by the negative numbers -1 resp. $-\xi^*$.
Having verified Equation (6.6), we can use it for the last equality in the following proof of Equation (6.7):

$$\begin{aligned}\psi_0(\theta_1 + \theta_2) &= \psi_0(0.5 \cdot 2\,\theta_1 + 0.5 \cdot 2\,\theta_2)\\ &= 0.5\,\psi_0(2\,\theta_1) + 0.5\,\psi_0(2\,\theta_2)\\ &= 0.5 \cdot 2\,\psi_0(\theta_1) + 0.5 \cdot 2\,\psi_0(\theta_2).\end{aligned} \tag{6.12}$$

(The second equality is another application of Equation (2.13) with $\lambda := 0.5$.)
(iii)⇒(ii): Under the assumption that the map ψ is affine, it suffices to show the existence of maps $\{\gamma_k\}_{k=0..m}$ such that Equation (2.16) holds for every $\theta \in \mathbb{R}^m$. Then Equation (2.8) will result from a pointwise application of the partial maps f(. , θ), which are equal to $\psi(\theta)$ by the assumed validity of Equation (2.5).

We claim that maps $\{\gamma_k\}_{k=0..m}$ fulfilling Equations (2.14) and (2.15) have these properties.[183] To verify this claim, observe that the definition of the vectors η_k allows to treat each element θ of $\mathbb{R}^m$ as a linear combination of these vectors, the components of θ serving as coefficients:

$$\theta = \textstyle\sum_{k=1..m} \theta(k)\,\eta_k. \tag{6.13}$$

So we get

$$\begin{aligned}\psi_0(\theta) &= \psi_0(\textstyle\sum_{k=1..m} \theta(k)\,\eta_k)\\ &= \textstyle\sum_{k=1..m} \theta(k)\,\psi_0(\eta_k)\\ &= \textstyle\sum_{k=1..m} \theta(k)\,\gamma k,\end{aligned} \tag{6.14}$$

the second equality being granted by the linearity of the map ψ_0 (which follows from Equation (6.4) and the assumed affinity of ψ), and the last one by Equations (2.15) and (6.4). This result can be combined with Equations (2.14) and (6.4) to obtain Equation (2.16).

It is left to prove the claims referring to the linearity and the injectivity of the map ψ. Certainly, ψ is linear iff $\psi(\eta_0) = \gamma_z$, and this equality is equivalent with $\gamma_0 = \gamma_z$. (See Equation (2.14) for this conclusion.) Furthermore, ψ is injective iff ψ_0 has this property. Now Equation (6.14) alllows to consider $\psi_0(\theta)$ as a result of a linear combination of the vectors $\{\gamma_k\}_{k=1..m}$, the components of θ serving as coefficients. But these coefficients are unique iff the vectors $\{\gamma_k)_{k=1..m}$ are linearly

[183] A potential reproof of circularity could argue that Equations (2.14) and (2.15) may not be used in a proof of the implication (iii)⇒(ii), since they have been derived under the assumption validity of Assertion (ii). So we should make clear the logical structure of both references to these equations.
- In the proof of the implication (ii)⇒(iv), we obtained the side results, that - given the validity of Assertion (ii) and suitable maps γ_k - Equations (2.14) and (2.15) must hold.
- For a proof of the implication (iii)⇒(ii) we will show that - given the validity of Assertion (iii) - maps γ_k fulfilling Equations (2.14) and (2.15) are suitable for Assertion (ii).

In other words, we don't refer to the equations as results, which have already been proved, but introduce 'new' maps γ_k fulfilling the equations and show that they are suitable.

independent vectors in $\mathbb{R}^Q$.□

6.8 Proof of Lemma 2.5

It suffices to show that Equation (2.22) holds for every $\theta \in \Theta$, since Equation (2.24) is tantamount with this claim, and Equation (2.25) is an immediate consequence. So let θ be an arbitrary element of Θ. Then $\psi(\theta)$ and $\psi^{\sim}(t(\theta))$ are the maps $v{:}Q\rightarrow Y$ resp $v^{\sim}{:}Q\rightarrow Y$, where the equations $v(q) = f(q, \theta)$ and $v^{\sim}(q) = f^{\sim}(q, t(\theta))$ hold for every $q \in Q$. Equation (2.23) yields $v(q) = v^{\sim}(q)$ for every $q \in Q$, and this is the claim of Equation (2.22).□

6.9 Proof of Lemma 2.6

Before we prove the claims of the lemma, we should verify that the underlying situation is well defined. Certainly, a sequence $\{k_j\}_{j=1..n}$ with the required property exists, since every subset of a real vector space includes a basis of the linear subspace spanned by the subset (see ###). The assumption $f(q, \theta) \neq \gamma_0(q)$ for some (q, θ) implies that $\gamma_k(q) \neq 0$ must hold for some $k < 0$ and $q \in Q$; hence there is a suitable $\gamma_{k(1)}$. Furthermore, every element of a vector space has a unique representation as a linear combination of elements of a given basis. Hence the coefficients λ_{jk} as well as the maps $f^{\sim}$ and t are uniquely defined for a given sequence $\{k_j\}_{j=1..n}$ by Equations (2.26), (2.27) and (2.28). It is also obvious that the map t is linear.

For the concluding claims, note that the linear independence of the maps $\{\gamma^{\sim}_j\}_{j=1..n}$ follows immediately from their property as a basis of F. Furthermore, Equation (2.23) can be derived step by step by the equation

$$\begin{aligned} f(q, \theta) &= \gamma_0(q) + \textstyle\sum_{k=1..m} \theta(k) \cdot \sum_{j=1..n} \lambda_{jk} \cdot \gamma^{\sim}_j(q) \\ &= \gamma_0(q) + \textstyle\sum_{j=1..n} (\sum_{k=1..m} \lambda_{jk} \cdot \theta(k)) \cdot \gamma^{\sim}_j(q) \\ &= f^{\sim}(q, t(\theta)) \end{aligned} \tag{6.15}$$

(with arbitrary $(q, \theta) \in Q\times\mathbb{R}^m$), which uses Equations (2.8) and (2.27) for the first line, and a rearrangement of sums for the second one. The last equality is granted by the definitions of the maps $f^{\sim}$ and t in Equations (2.26) and (2.28).□

6.10 Proof of Lemma 2.9

Throughout this proof, $\sum_j$ and $\sum_k$ stand for $\sum_{j=1..m}$ and $\sum_{k=1..m}$.

The proof will be given in three steps. A first step will show the existence of a family $\{q_j\}_{j=1..m}$ of elements of the set Q such that an m×m-matrix B with elements $b_{jk} := \gamma_k(q_j)$ is non-singular. A second step will give definitions of the matrix Z and the other families used for the final claims, and a third step will verify these claims.

A basic property for the construction of a suitable family $\{q_j\}_{j=1..m}$ of elements of Q is the linear independence of the maps $\{\gamma_k\}_{k=1..m}$, which follows from the concluding claim in Estes-Theorem 2.4. (Recall that the injectivity of the prametrisation map ψ is a premissa of the lemma. Furthermore, Assertion (ii) of the Estes-Theorem holds, since the validity of Equation (2.9) is assumed, and Equation (2.8) is equivalent with Equation (2.9).) This linear independence will now be used to prove the following claim for n = 1..m by induction over n:

(vi) There is a a sequence $\{q_j\}_{j=1..n}$ of elements of Q such that the n×n-matrix B with elements $b_{jk} := \gamma_k(q_j)$ (with j = 1..n and k = 1..n) is non-singular.

For n = 1, the claim is fulfilled by any element q_1 of Q with $\gamma_1(q_1) \neq 0$. (Certainly, an element q_1 with this property must exist: Otherwise g_1 would be the zero-element of $\mathbb{R}^Q$, and then the maps γ_k:Q→$\mathbb{R}$ wouldn't be lineraly independent vectors $\mathbb{R}^Q$.) Now let i be a natural number with i < m such that (vi) holds for n = i, and let a suitable sequence $\{q_j\}_{j=1..i}$ and the corresponding matrix B specified in (vi) be given. Furthermore, let $\{\lambda_k\}_{k=1..i}$ be the unique real numbers with the property $\gamma_{i+1}(q_j) = \sum_{k=1..i} \lambda_k \cdot \gamma_k(q_j)$ for j = 1..i. (Indeed, these real numbers exist and are unique, since the matrix B is non-singular.) Then the linear independence of the maps γ_k implies the existens of an element q_{i+1} of Q with $\gamma_{i+1}(q_{i+1}) \neq \sum_{k=1..i} \lambda_k \cdot \gamma_k(q_{i+1})$. (Otherwise, the equality $\gamma_{i+1} = \sum_{k=1..i} \lambda_k \gamma_k$ would follow, and the maps γ_k wouldn't be linearly independent vectors in $\mathbb{R}^Q$.) But if an element q_{i+1} with this property exists, then (vi) holds for n = i+1.

For the rest of the proof, let a family $\{q_j\}_{j=1..m}$ of elements of Q and a non-singular m×m-matrix B with elements $b_{jk} := \gamma_k(q_j)$ be given. Furthermore, let an m×m-matrix Z with elements z_{kj} be the inverse of the matrix B (i.e., $Z := B^{-1}$), and observe that the non-singularity of Z follows immediately from this definition. Suitable definitions of the families $\{\xi_k\}_{k=1..m}$, $\{\tau_q\}_{q\in Q}$ and $\{\zeta_{qj}\}_{q\in Q,\, j=1..m}$ of real numbers are

$$\xi_k := -\sum_j z_{kj} \cdot \gamma_0(q_j), \tag{6.16}$$

$$\tau_q := -\gamma_0(q) + \sum_j \sum_k z_{kj} \cdot \gamma_0(q_j) \cdot \gamma_k(q), \tag{6.17}$$

and

$$\zeta_{qj} := -\sum_k z_{kj} \cdot \gamma_k(q). \tag{6.18}$$

Finally, Equations (2.36) and (2.37) can be taken as definitions of a family $\{h_k\}_{k=1..m}$ of maps h_k:V→$\mathbb{R}$ and a family $\{g_q\}_{q\in Q}$ of maps g_q:V→$\mathbb{R}$. Then Assertions (i) and (ii) follow immediately, and it is left to prove Assertions (iii), (iv) and (v).

(iii): Replacing the real numbers τ_q and ζ_{qj} in Equation (2.36) by the right hand sides of Equations (6.17) and (6.18) lead to

$$g_q(v) = -\gamma_0(q) + \sum_j \sum_k z_{kj} \cdot \gamma_0(q_j) \cdot \gamma k(q) + v(q) + \sum_j (-\sum_k z_{kj} \cdot \gamma_k(q)) \cdot v(q_j). \tag{6.19}$$

Similarly, the equation

$$v(q) - (\gamma_0(q) + \sum_k h_k(v) \cdot \gamma_k(q)) = v(q) - (\gamma_0(q) + \sum_k (-\sum_j z_{kj} \cdot \gamma_0(q_j) + \sum_j z_{kj} \cdot v(q_j)) \cdot \gamma_k(q)) \tag{6.20}$$

results, if the term $h_k(v)$ on its left hand side is substituted by the right hand side of Equation (2.36) and Equation (6.16) is applied for the number ξ_k. Now it is easily verified that the right hand sides

of Equations (6.19) and (6.20) are identical up to a rearrangement of sums. So we may equate their left hand sides to obtain Equation (2.38).
(iv): For given elements v and θ of V resp. $\mathbb{R}^m$ with $v = \psi(\theta)$, define another element x of $\mathbb{R}^m$ by the equation

$$x(j) := v(q_j) - \gamma_0(q_j) \tag{6.21}$$

for its j^{th} component (j = 1..m). Then Equation (2.9) (whose validity is a premissa of the lemma) can be combined with the definition of the matrix B to obtain

$$x(j) = \sum_k b_{jk} \cdot \theta(k) \tag{6.22}$$

for j = 1..m. Conceiving x and θ as column vectors, we can rewrite Equation (6.22) in matrix notation as $x = B \cdot \theta$. A premultiplation of both sides of this equality leads to $Z \cdot x = Z \cdot B \cdot x$, and the definition of Z as the inverse of B yields $Z \cdot x = \theta$, i.e.,

$$\theta(k) = \sum_j z_{kj} \cdot x(j) = \sum_j z_{kj} \cdot (v(q_j) - \gamma_0(q_j)) \tag{6.23}$$

for k = 1..m. Now recall that Equation (2.36) has been used as a definition of the maps $h_k{:}V \to \mathbb{R}$. Replacing the real number ξ_k in that equation by the right hand side of Equation (6.16) and rearranging sums, we get

$$h_k(v) = \sum_j z_{kj} \cdot (v(q_j) - \gamma_0(q_j)) \tag{6.24}$$

for k = 1..m, and then the equality $\theta(k) = h_k(v)$ is obtained by a combination of Equations (6.23) and (6.24).
(v): It suffices to establish for a given element v of the set V a chain of equivalences between the following properties:
(v.a): $v \in S^*$
(v.b): There is an element θ of $\mathbb{R}^m$ such that $v = \psi(\theta)$.
(v.c): There is an element θ of $\mathbb{R}^m$ such that

$$v(q) = \gamma_0(q) + \sum_k \theta(k) \cdot \gamma_k(q) \tag{6.25}$$

for every $q \in Q$.
(v.d): The equation

$$v(q) = \gamma_0(q) + \sum_k h_k(v) \cdot \gamma_k(q) \tag{6.26}$$

holds for every $q \in Q$.
(v.e): The equality $g_q(v) = 0$ holds for every $q \in Q$.
So let v be an arbitrary element of V. Then the equivalence (v.a)⇔(v.b) follows immediately from the definition of the set S* by Equation (2.35). Furthermore, since the validity of Equation (2.9) for every $\theta \in \mathbb{R}^m$ is a premissa of the lemma, the equivalence (v.b)⇔(v.c) can be based on the argument that the equality $v = \psi(\theta)$ is equivalent with the validity of Equation (6.25) for every $q \in Q$. To prove the equivalence (v.c)⇔(v.d), assume first that (v.c) holds, and recall that the equivalence (v.b)⇔(v.c) and Assertion (iv) have already been proved. Combining both results, we obtain the equality $\theta(k) = h_k(v)$ for k = 1..m, and Equation (6.26) follows for every $q \in Q$ from the assumed validity of Equation (6.25). Conversely, if Equation (6.26) holds for every $q \in Q$, then an element θ

of $\mathbb{R}^Q$ fulfilling Equation (6.25) is obtained by the definition $\theta(k) := h_k(v)$ for k = 1..m. Finally, the equivalence (v.d)⇔(v.e) is an immediate consequence of Equation (2.38), which has already been proved.□

Note that the injectivity of the parametrisation map ψ is not only sufficient for the claims of the lemma, but also necessary, if the set S* is non-empty. (For a non-injective map ψ, there would be an infinite set of suitable parameters for every $v \in S^*$. But Assertion (iv) cannot hold for different parameters θ underlying the same element v of the set V. It would be possible to restate the assertion as follows: If v is an element of V and if there are elements θ of $\mathbb{R}^m$ such that $v = \psi(\theta)$, then there is one such θ with $\theta(k) = h_k(v)$ for k = 1..m. But since parametric families with non-injective parametrisation maps are highly untypical and can alway be reparametrised such that an injective parametrisation map results, further discussions of this issue are not worth to be undertaken in the present context.

Observe also, that Lemma 2.9 can be reversed: Let Q be a non-empty set, V a set of maps $v:Q\to\mathbb{R}$, and $\{q_j\}_{j=1..m}$ a family of elements of Q. Furthermore, let$\{\tau_q\}_{q\in Q}$ and $\{\zeta_{qj}\}_{q\in Q,\, j=1..m}$ be families of real numbers fulfilling certain conditions (which will be commented immeditately), and let a family $\{g_q\}_{q\in Q}$ a maps $g_q:V\to\mathbb{R}$ be given by Equation (2.37). Finally, let S* be the set of all elements v of V with $g_q(v) = 0$. Then there exists a family $\{\gamma_k\}_{k=0..m}$ of maps $\gamma_k:Q\to\mathbb{R}$ such that the map $\psi:\mathbb{R}^m\to\mathbb{R}^Q$ given by the definition $\psi(\theta) := \gamma_0 + \sum_{k=1..m} \theta_k \cdot \gamma_k$ is injective and that the equality $S^* = V \cap \psi(\mathbb{R}^m)$ holds for this map.

For the side conditions, note that the definitions of the matrices B and Z can be combined with Equation (6.18) to obtain $\zeta_{q(j)j} = -1$, and $\zeta_{q(j)i} = 0$ for $j \neq i$. Furthermore, the construction principle implies $g_{q(j)}(v) = 0$ for j = 1..m and every $v \in V$. So these properties must hold, if the numbers τ_q and ζ_{qj} are required to have the same relationship to the elments q_j of the set Q as in the above proof. Certain violations of these properties could be compensated by a transition to another family $\{q'_j\}_{j=1..m}$ of elements of Q or to other families $\{\tau'_q\}_{q\in Q}$ and $\{\zeta'_{qj}\}_{q\in Q,\, j=1..m}$ of real numbers, which may take the respective roles in the situation, which is assumed in the proof. But we shouldn't go further into these details, since the construction of a suitable parametrisation map for given sets V and S* is of subordinate relevance for a theory of aggregation stability. Furthermore, some principles of constructing parametrisation maps for a given set of members of a family will be presented in Section 6.13.

Qv 6.11

6.11 Proof of Lemma 2.11

Although Equation (2.48) has been introduced for a special situation, it can be generalised, since the validity of Equation (2.23) and the membership of $f^{\sim}$ in Class A are assumed in the lemma. For the last line, the existence of an element θ^* fulfilling Equation (2.49) is granted by the assumed convexity of the set $t(\Theta)$.□

6.12 Proof of Lemma 2.12

In the situation assumed by the lemma, let $\{y_k\}_{k=1..m}$ be an arbitrary finite sequence of linear independent elements of $\psi(\Theta)$. (Certainly, a sequence with this property exists for m = 1, since $\psi(\Theta)$ is assumed to contain a non-zero element.) It suffices to verify that $m \le n + 1$ must hold for every such sequence: Then we can take the maximal number m, where a sequence with this property exists, and use any sequence of that length to represent every element of $\psi(\Theta)$ as a linear combination of the elements of the sequence. (If there would be an element of $\psi(\Theta)$ without a representation of this kind, it would be linearly independent of the elements of the sequence, and m wouldn't be maximal.) Furthermore, the representation of a vector as a linear combination of other linearly independent vectors is unique, if it exists.

In the proof of the inequality $m \le n + 1$ for the assumed situation, we may assume $m > 1$. (Otherwise, nothing is left to be proved.) We will construct a surjective map $h{:}E \to \mathbb{R}^{m-1}$. Then $m - 1 \le n$ will follow immediately, since a real vector space cannot be mapped surjectively upon a vector space of higher dimension.

For the construction of the map h, let an arbitrary bijective map $g{:}]0, 1[\to \mathbb{R}$ be given (e.g. $g(\xi) = \log(\xi/(1-\xi))$), and for every $x \in E$, let h(x) be defined by the following rules:

- If $x \in \Theta$ and $\psi(x)$ can be represented as a convex linear combination $\psi(x) = \sum_{k=1..m} \lambda_k y_k$ with non-zero coefficients λ_k, then h(x) is the vector z in $\mathbb{R}^{m-1}$ with components[184]

$$z(1) := g(\lambda_1), \tag{6.27}$$

and

$$z(k) := g(\lambda_k / (1 - \textstyle\sum_{j=1..k-1} \lambda_j)) \tag{6.28}$$

for $k = 2..m-1$.

- If $x \notin \Theta$ or real numbers $\{\lambda_k\}_{k=1..m}$ with the above properties don't exist, then h(x) is the zero-element of $\mathbb{R}^{m-1}$.

To verify that the map $h{:}E \to \mathbb{R}^{m-1}$ defined by these rules is surjective, let z be an arbitrary element of $\mathbb{R}^{m-1}$, and we will prove the existence of an element x of Θ such that $z = h(x)$. So let real numbers $\{\lambda_k\}_{k=1..m}$ be given recursively by

$$\lambda_1 := g^{-1}(z(1)), \tag{6.29}$$

$$\lambda_k := g^{-1}(z(k)) \cdot (1 - \textstyle\sum_{j=1..k-1} \lambda_j), \tag{6.30}$$

for $k = 2..m-1$, and

$$\lambda_m := 1 - \textstyle\sum_{j=1..m-1} \lambda_j. \tag{6.31}$$

---

[184] Note that suitable coefficients λ_k are unique, if they exist, since the vectors y_k are assumed to be linearly independent.

Now we claim that the definition of a vector y by

$$y := \sum_{k=1..m} \lambda_k \, y_k \tag{6.32}$$

represents y as a convex linear combantion of the vectors y_k with non-zero coefficients. If this claim can be verified, then y is an element of the set $\psi(\Theta)$, since this set is assumed to be convex and to contain the vectors y_k. In other words, $y = \psi(x)$ holds for some element x of Θ, and $z = h(x)$ follows from a straightforward application of Equations (6.27) and (6.28).

To verify that Equation (6.32) represents y as a convex linear combination of the vectors y_k with non-zero coefficients, observe first that the inequality

$$0 < g^{-1}(z(k)) < 1 \tag{6.33}$$

must hold for $k = 1..m-1$, since g is a bijective map $]0, 1[\rightarrow\mathbb{R}$. So it suffices to show that the inequality

$$0 < \sum_{j=1..k} \lambda_j < 1 \tag{6.34}$$

holds for $k = 1..m-1$: Then the properties $\lambda_k > 0$ for $k = 1..m$ and $\sum_{k=1..m} = 1$ will follow from Equations (6.30) and (6.31) and Inequality (6.33).

Inequality (6.34) can be easily verified by induction over k. For $k = 1$, it is tantamount with Inequality (6.34) (see Equation (6.29) for this conclusion). For $1 < k \leq m-1$, Equation (6.30) is used for the first line in the equation

$$\begin{aligned} \sum_{j=1..k} \lambda_j &= \sum_{j=1..k-1} \lambda_k + g^{-1}(z(k)) \cdot (1 - \sum_{j=1..k-1} \lambda_j) \\ &= 1 - (1 - g^{-1}(z(k))) \cdot (1 - \sum_{j=1..k-1} \lambda_j), \end{aligned} \tag{6.35}$$

where the second equality can be verified by resolving parantheses in the second line. But then the inductive assumption $0 < \sum_{j=1..k} \lambda_j < 1$ and Inequality (6.33) lead to Inequality (6.34).□

6.13 Proof of Corollary 2.13

(i)⇒(ii): If S contains only the zero-element of $\mathbb{R}^Q$, the implication is trivial. Otherwise, let a finite sequence $\{\gamma_k\}_{k=1..m}$ of elements of S be given, which form a basis of the linear subspace of $\mathbb{R}^Q$ generated by S.[185] Furthermore, let a map $f:Q\times\mathbb{R}^m\rightarrow\mathbb{R}$ be given by Equation (2.30), derive a map

[185] Readers, who are familiar with the theory of vector spaces, will know that a sequence with this property exists. For others, the existence can be proved by the following recursive construction:

- Start with $m = 1$, and let γ_1 be an arbitrary non-zero element of S.
- Recursion cycle: If S contains elements, which are linear independent of the vectors $\{\gamma_k\}_{k=1..m}$, take any one of them and add it to the sequence; otherwise stop.

(continued...)

$\psi:\mathbb{R}^m \to \mathbb{R}^Q$ from f by Equation (2.5), and define $\Theta := \psi^{-1}(S)$. Then the set S is a parametric family of maps $Q \to \mathbb{R}$ with parameter space Θ, whose representation function is the restriction of f to $Q\times\Theta$. Finally, the stability of S under convex linear combinations follows from Lemma 2.1, if S is convex.
(ii)⇒(i): If the set S is stable under convex linear combinations, it must be convex by Lemma 2.1. Now let a natural number m, a subset Θ of $\mathbb{R}^m$ and a function $f:Q\times\Theta\mathbb{R}$ be given such that S consists of all members of a parametric family of maps $Q \to \mathbb{R}$ with parameter space Θ and representation function f. Furthermore, derive the parametrisation map $\psi:\Theta \to \mathbb{R}^Q$ of the family by Equation (2.5), which implies $S = \psi(\Theta)$. Finally, (referring to Lemma 2.12) let a sequence $\{\gamma_k\}_{k=1..n}$ of linearly independent elements of S be given such that they form a basis of the linear subspace of $\mathbb{R}^Q$, which is spanned by S. Certainly, this linear subspace is finite dimensional, since it has a finite basis.□

6.14 Proof of Lemma 2.15

(i): Let v be a given element of V_Q, P a probability measure on the Borel sets in $\mathbb{R}$ fulfilling Assertions (i) and (ii) of Definition 2.14 for the given v, and A a Borel in $\mathbb{R}$ set with $A \subseteq Q$ and $P(A) = 1$. Since a probability measure on the Borel sets in $\mathbb{R}$ is uniquely specified by the probabilities of all intervals $]-\infty, \xi]$ with $\xi \in \mathbb{R}$, it suffices to show that Equation (2.55) must hold for every $\xi \in \mathbb{R}$. Now the assumption $P(A) = 1$ can be combined with Equation (2.54) to obtain $v(q) = P(A \cap\,]-\infty, q]$ for every $q \in Q$, and $P(]-\infty, \xi]) = \sup P(A \cap\,]-\infty, \xi])$ for every $\xi \in \mathbb{R}$. But the second equation is equivalent with Equation (2.55) in the assumed situation.
(ii): Let v_1 and v_2 be elements of V_Q, and λ a real number in the interval $]0, 1[$. Furthermore, let P_1 and P_2 be the unique probability measures on the Borel sets in $\mathbb{R}$ with $P_i(]-\infty\ q]) = v_i(q)$ for every $q \in Q$, and let another probability measure P on these Borel sets be given as $P := \lambda\, P_1 + (1-\lambda)\, P_2$ (which means $P(A) = \lambda\, P_1(A) + (1-\lambda)\, P_2(A)$ for every Borel set A in $\mathbb{R}$). Finally, let A_1 and A_2 be Borel sets in $\mathbb{R}$, which are subsets of Q such that $P_1(A_1) = P_2(A_2) =1$, and let A be the union of A_1 and A_2, which is again a Borel set in $\mathbb{R}$ and a subset of Q. Then $P_1(A) = P_2(A) = 1$ is granted by the inclusions $A_1 \subseteq A$ and $A_2 \subseteq A$, and the definition of P leads to $P(A) = 1$. So the map $v:Q\to\mathbb{R}$ given by Equation (2.54) is an element of V_Q, and $v = \lambda\, v_1 + (1-\lambda)\, v_2$ follows from the above definitions. But if every such convex linear combination of elements v_1 and v_2 of V results in an element of V_Q, then V is convex.
(iii): Let m be an arbitrary natural number, and let the map $g:Q\to V_Q$ be given such that g(q) is the (unique) element of V_Q, where the equation $P(\{q\}) = 1$ holds for the probability measure P fulfilling

[185] (...continued)
If S is included in a linear subspace of $\mathbb{R}^Q$ with finite dimension n, then the stopping condition will be reached after not more than n cycles. (A vector space with finite dimension n cannot contain more than n linearly independent elements.) So take the sequence $\{\gamma_k\}_{k=1..m}$ after the stopping. Then every element of S can be represented as a linear combination of the vector γ_k. (If S would contain an element without a representation of this kind, it would be linearly independent of the γ_k, and the stopping condition wouldn't have applied.) So the linear subspace of $\mathbb{R}^Q$, which is spanned by the vectors γ_k, includes S, and since all vectors γ_k are linearly independent elements of S, these vectors form a basis of the smallest linear subspace of $\mathbb{R}^Q$ including S.

Assertions (i) and (ii) of Definition 2.14. Now asumme first that m isn't greater than the number of elements of Q, and let $\{q_k\}_{k=1..m}$ be a sequence of pairwise different elements of Q. Then the maps $\{v_k\}_{k=1..m}$ defined by $v_k := g(q_k)$ are linearly independent. Conversely, if the number of elements of Q is n with n < m, then V_Q is a subset of $\mathbb{R}^Q$, and this function space has dimension n. So it cannot contain more that n linearly independent elements.
(iv): Let real numbers $\{\lambda_k)_{k=1..m}$ and elements $\{v_k\}_{k=0..m}$ of V_Q be given such that $v_0 = \sum_{k=1..m} \lambda_k v_k$. Furthermore, let $\{P_k\}_{k=0..m}$ be the respective probability measures on the Borel sets in $\mathbb{R}$ fulfilling Assertions (i) and (ii) of Definition 2.14, and $\{A_k\}_{k=0..m}$ Borel sets in $\mathbb{R}$ with $A_k \subseteq Q$ and $P_k(A_k) = 1$. With the definition $A := \bigcup_{k=0..m} A_k$, we obtain another Borel set in $\mathbb{R}$ with $A \subseteq Q$ and $P_k(A) = 1$ for k = 0..m, and $P_0(A) = \sum_{k=1..m} \lambda_k P_k(A)$. Taken together, the results imply $\sum_{k=1..m} \lambda_k = 1$.□

6.15 Auxiliary Lemmas for the Analysis of Strict Stochastic Order

The subsequent lemma will establish the equivalence of the definitions of the sets S_O, S_L and S_{st} in Equations (3.13), (3.14) and (3.15) and in Equations (3.18), (3.19) and (3.20). For notational convenience, only the sets resulting from the original definitions are referenced as S_O, SL resp. S_{st}, and the denotations $S^{\sim}_O$, $S^{\sim}_L$ and $S^{\sim}_{st}$ are used for the redifined sets. The lemma states the respective identities.

Lemma 6.2: In the CDF-SSA of Example 3.3, let W* be the set of all rational numbers contained in W, and define sets $S^{\sim}_O$, $S^{\sim}_L$ and $S^{\sim}_{st}$ by

$$S^{\sim}_O := \bigcap_{w^* e W^*} S'_{w^*}, \tag{6.36}$$

$$S^{\sim}_L := \bigcap_{w^* e W^*} S^*_{w^*} \tag{6.37}$$

and

$$S^{\sim}_{st} := \bigcup_{w^* \in W^*} S''_{w^*}. \tag{6.38}$$

Then the identities $S^{\sim}_O = S_O$, $S^{\sim}_L = S_L$, and $S^{\sim}_{st} = S_{st}$ hold for these sets and those defined by Equations (3.13), (3.14) and (3.15).

<u>Proof of Lemma 6.2</u>:

Since the inclusions $S_O \subseteq S^{\sim}_O$, $S_L \subseteq S^{\sim}_L$ and $S^{\sim}_{st} \subseteq S_{st}$ are obvious consequences of $W^* \subseteq W$, it suffices to verify the reversed inclusions.

We start with $S_{st} \subseteq S^{\sim}_{st}$, since a side result of the following proof can also be used for the other inclusions. So let v be an element of S_{st}, and w an element of W such that $v \in S''_w$, which means $v(a, w) > v(b, w)$. This implies $w < w''$, since $v(a, w'') = v(b, w'') = 1$ follows from Inequality (3.5). Furthermore, combining the assumption $v(a, w) > v(b, w)$ with Equation (3.6), we can conclude that there must be a real number δ with $0 < \delta \le w'' - w$ such that the inequality $v(a, w) > v(b, w+\alpha)$ holds for every $\alpha \in]0, \delta[$. So let δ be a real number with these properties, and let α be a number in the interval]0, δ[such that the number $w + \alpha$ is rational and hence contained in W*. (Like every non-empty open interval of real numbers, the interval]w, w + δ[contains rational numbers!) With the definition $w^* := w + \alpha$, we have $w^* \in W^*$ and $v(a, w^*) \ge v(a, w) > v(b, w^*)$, the first inequality following from Inequality (3.5). In summary, we have $v(a, w^*) > v(b, w^*)$ for a suitable element w* of the set W*, and the properties $v \in S''_{w^*}$ and $v \in S^{\sim}_{st}$ follow from this result.

For the inclusion $S^{\sim}_O \subseteq S_O$, let v be an element of V with $v \notin S_O$, and we will show that v is not contained in $S^{\sim}_O$. The assumption $v \notin S_O$ is equivalent with the existence of an element w of W with v(a, w) < v(b, w). So the roles of v(a, w) and v(b, w) in the preceeding proof can be interchanged to show the existence of an element w* of W* with v(a, w*) < v(b, w*). So $v \notin S'_{w*}$ follows, and this implies $v \notin S\sim_O$.

Finally, for the inclusion $S^{\sim}_L \subseteq S_L$, let v be an element of V with $v \notin S_L$, which implies the existence of an element w of W with v(a, w) ≠ v(b, w). For v(a, w) > v(b, w), the existence of an element w* of W* with v(a, w*) > v(b, w*) can be proved in the same way as for the inclusion $S_{st} \subseteq S^{\sim}_{st}$, and for v(a, w) < v(b, w), the roles of v(a, w) and v(b, w) in that proof can again be interchanged to obtain an element w* of W* with v(a, w*) < v(a, w*). In both cases, the existence of an element w* of W* with $v \notin S'_{w*}$ follows from $v \notin S_L$, and this means $v \notin S^{\sim}_L$.□

The following lemma is used in Section 3.8 to verify Implication (3.52) for an application of Intersection-Theorem 3.14. The set definitions in Example 3.3 are valid for the lemma, and W* is again the set of all rational numbers contained in W.

Lemma 6.3: In the SSA $(V, \Pi^{\sim}, \Phi^{\sim}, H^{\sim}, \Phi^{\sim})$ of Section 3.8, let an element π of the set $\Pi^{\sim}$ be given such that $(S^*_{w*}, \pi) \in H^{\sim}$ for every $w^* \in W^*$. Then $(S_L, \pi) \in H^{\sim}$.

Since Lemma 6.3 is an immediate application of Lemma 4.3, a proof is postponed until Subsection 6.21.

6.16 Proof of Mapping-Theorem 3.8

(i): To prove the implication $S' \in T' \Rightarrow S \in T$, assume that the set S' is contained in T', and let π be an arbitrary element of Π such that $(S, \pi) \in H$. According to SSA-Axiom (iv), it suffices to verify $\Phi(\pi) \in S$ for this situation. Combining the above assumptions with Implication (3.35) and SSA-Axiom (iv), we obtain $\Phi'(f(\pi)) \in S'$, and then $\Phi_\pi \in S$ follows from Equations (3.36) and (3.34).
(ii): Under the assumption $S' \in T'_e$, we obtain $S' \in T'$ from Lemma 1.3.(i), and $S \in T$ follows as above. According to Lemma 1.3.(iv), it is left to show that the implication $\Phi_\pi \in S \Rightarrow (S, \pi) \in H$ holds for every $\pi \in \Pi$. So let π an element of Π with $\Phi_\pi \in S$. Then $\Phi'(f(\pi)) \in S'$ is granted by Equations (3.34) and (3.36), and $(S', f(\pi)) \in H'$ by the assumption $S' \in T'_e$. Finally, $(S, \pi) \in H$ is obtained from the reversal of Implication (3.35).
(iii): Assume $S \in T$, and let π' an arbitrary element of Π' with $(S', \pi') \in H'$. To derive $\Phi'_{\pi'} \in S'$ from these assumptions, refer to the assumed surjectivity of the map f, and let an element π of Π be given such that π' = f(π). Then $(S, \pi) \in H$ follows from the reversal of Implication (3.35), and the assumption $S \in T$ yields $\Phi_\pi \in S$ by SSA-Axiom (iv). But then $\Phi'_{\pi'} \in S'$ can be obtained from Equations (3.34) and (3.36).
(iv): Since the implication $S \in T \Rightarrow S' \in T'$ as well as $S' \in T' \Rightarrow S \in T$ and $S' \in T'_e \Rightarrow S \in T_e$ have already been established under weaker premissas, it suffices to show under the assumption $S \in T_e$ that the implication $\Phi'_{\pi'} \in S' \Rightarrow (S', \pi') \in H'$ holds for every $\pi' \in \Pi'$: Then $S' \in T'_e$ will follow from Lemma 1.3.(i) and (iv) . So let π' be an element of Π' with $\Phi'_{\pi'} \in S'$, and π an element of Π such that π' = f(π). Then $\Phi_\pi \in S$ follows from Equations (3.36) and (3.34), and this implies $(S, \pi) \in H$ under the assumption $S \in T_e$. Finally, $(S', \pi') \in H'$ is obtained from Implication (3.35).□

Note that the premissa of a surjective map $f:\Pi\rightarrow\Pi'$ in Assertions (iii) and (iv) of Theorem 3.8 can be weakened, since the existence of an element π with $\pi' = f(\pi)$ has been assumed in the proof only for some elements π' of Π'. A necessary and sufficient condition for the implication $S \in T \Rightarrow S' \in T'$ in Assertion (iii) can be stated as follows: If there is an element π' of Π' with $(S', \pi') \in H'$ and $\Phi'_\pi \notin S'$, then there must be an element π of Π such that the equation $\pi' = f(\pi)$ holds for some π' with these properties. Similarly, the following assumption is necessary and sufficient for the implication $S \in T_e \Rightarrow S' \in T'_e$ in Assertion (iv): If there is an element π' of Π' such that one and only one of the properties $(S', \pi') \in H'$ and $\Phi'_{\pi'} \in S'$ is present, then the equation $\pi' = f(\pi)$ must hold for some such π' and some element π of Π. Obviously, both conditions follow, if the map $f:\Pi\rightarrow\Pi'$ is surjective, but weaker assumptions implying one of the conditions are also sufficient for the respective equivalence. (Of course, the two conditions are required in addition to the premissas of Assertions (iii) and (iv) referring to Implication (3.35) and its reversal.)

6.17 Proof of Lemma 3.12

The validity of SSA-Axioms (i), (ii) and (iii) for the relation H^* follows immediately from its definition by Equation (3.46), if they hold for the unlerlying relation H. Then SSA-Axiom (iv) and Equation (1.1) specify unique set systems $T^\sim \subseteq PV^*$ and $T^\sim_e \subseteq PV^*$ such that the ordered quintuple $(V^*, \Pi^*, Z^*, H^*, T^\sim)$ is an SSA. With the definition of a map $f:\Pi\rightarrow\Pi^*$ as the identity map in Π, everything is prepared to derive from Mapping-Theorem 3.8.(iv) that the properties $S^* \in T^\sim$ and $S^* \in T^\sim_e$ are equivalent with $g'^{-1}(S^*) \in T$ resp. with $g'^{-1}(S^*) \in T_e$ for every subset S^* of V^*. But then the set systems T^* and T^*_e defined by Equations (3.47) and (3.48) are identical with $T^\sim$ resp. with $T^\sim_e$.

6.18 Proof of Homomorphism-Corollary 3.13

Before main conclusion of the corollary is proved, it should be established that the equivalence relation ~ and the map $g'':V^*\rightarrow V'$ are well defined. Obviously, the relation ~ defined by Equivalence (3.49) is reflexive, symmetric and transitive, i.e., an equivalence relation. Furthermore, the property $g = g'' \circ g'$ holds for a map $g'':V^*\rightarrow V'$ iff $g(v^*)$ with $v^* \in V^*$ is the (identical) function value $g(v)$ of all elements of the equivalence class v^*.

Given this map $g'':V^*\rightarrow V'$, the assumptions of the corollary about the maps $g:V\rightarrow V'$ and $f:\Pi\rightarrow\Pi'$ can be combined with the definitions in Lemma 3.12 and with Isomorphism-Corollary 3.10 to obtain the properties of an SSA-isomorphism.□

6.19 Proof of Intersection-Theorem 3.14

For every $\pi \in \Pi$ with $(S, \pi) \in H$, the property $(S_i, \pi) \in H$ follows for every $i \in I$ from SSA-Axiom (iii), since $S_i \subseteq S$ is granted by the definition $S := \bigcap_{i\in I} S_i$. Then SSA-Axiom (iv) leads to $\Phi_\pi \in S_i$ for every $i \in I$, and the conclusion $\Phi_\pi \in S$ summarises these containments. But if (S, π) implies $\Phi_\pi \in S$ for every $\pi \in \Pi$, then $S \in T$ is obtained by another reference to SSA-Axiom.

Under the additional assumption that the sets S_i are contained in the set system T_e, the implication $(S, \pi) \in H \Rightarrow \Phi_\pi \in S$ follows as above for every $\pi \in \Pi$. Conversely, if $\Phi_\pi \in S$, then $\Phi_\pi \in S_i$ for every $i \in I$, and $(S_i, \pi) \in H$ is yielded by the definition of the set system T_e. But then $(S, \pi) \in H$ follows, if Implication (3.52) holds for every $\pi \in \Pi$.□

6.20 Proof of Conditioning-Theorem 3.18

Since the SSA $(V, \Pi^\sim, \Phi^\sim, H^\sim, T^\sim)$ is the $\Pi^\sim$-restriction of the SSA (V, Π, Φ, H, T), the premissas of Restriction-Lemma 3.4 (with suitable adaptations of notation) are part of the assumed situation.

So let S be an arbitrary subset of V, and we will prove the implications (i)⇔(ii)⇒(i)⇒(iii), and (iii)⇒(i) under the additional assumption that Implication (3.56) holds for every $\pi \in \Pi$.

(i)⇒(ii): Let π be an arbitrary element of $\Pi^\sim$ such that $(S, \pi) \in H^\sim$. Then $(S_0, \pi) \in H$ follows from Equation (3.53), and $(S, \pi) \in H$ from Equation (3.21). Under the assumed validity of Assertion (i), we obtain $\Phi_\pi \in S$, and the definition of $\Phi^\sim$ as a restriction of Φ yields $\Phi^\sim{}_\pi \in S$. But if $(S, \pi) \in H^\sim$ implies $\Phi^\sim{}_\pi \in S$ for every $\pi \in \Pi^\sim$, then $S \in T^\sim$ follows from SSA-Axiom (iv).

(ii)⇒(i): Assume that S is contained in $T^\sim$, and let π be an arbitrary element of Π such that $(S_0, \pi) \in H$ and $(S, \pi) \in H$. (Certainly, the implication claimed by Assertion (i) cannot be false for an element π of Π without these properties.) Then $\pi \in \Pi^\sim$ follows from Equation (3.53), and $(S, \pi) \in H^\sim$ from Equation (3.21). So SSA-Axiom (iv) yields $\Phi^\sim{}_\pi \in S$, and $\Phi_\pi \in S$ can be obtained from the property of $\Phi^\sim$ as a restriction of Φ.

(i)⇒(iii): For every element π of Π with $(S \cap S_0, \pi) \in H$, the properties $(S, \pi) \in H$ and $(S_0, \pi) \in H$ follow from SSA-Axiom (iii), since the set $S \cap S_0$ is a subset of S and of S_0. So $\Phi_\pi \in S$ follows, if Assertion (i) holds.

(iii)⇒(i): Under the additional assumption that Implication (3.56) holds for every $\pi \in \Pi$, let π an element of Π with $(S_0, \pi) \in H$ and $(S, \pi) \in H$. Then $(S \cap S_0, \pi) \in H$ follows from Implication (3.56), and $\Phi_\pi \in S$ from the assumed validity of Assertion (iii).□

6.21 Proof and Application of Lemma 4.3

In the assumed situation, let a set Ω_0, a σ-algebra A_0 in Ω_0, and a map $\phi:\Omega_0 \to V$ be given such that SSA-Axioms (v) and (vii) hold. According to SSA-Axiom (vii), we have to verify the existence of a subset A of Ω_0 with the properties $A \in \mathsf{A}_0$, $\phi(A) \subseteq \bigcap_{i\in I} S_i$, and $\pi(A) = 1$. So let a family $\{A_i\}_{i\in I}$ of subsets of Ω_0 begiven such that $A_i \in \mathsf{A}_0$, $\pi(A_i) = 1$, and $\phi(A_i) \subseteq S_i$ for every $i \in I$, the existence of such sets being granted by the assumption $(S_i, \pi) \in H$ and SSA-Axiom (vii). We will show that the set $A := \bigcap_{i\in I} A_i$ has the required properties. Immediately, we obtain $\phi(A) \subseteq \bigcap_{i\in I} S_i$ from $\phi(A_i) \subseteq S_i$, and $A \in \mathsf{A}_0$ follows from $A_i \in \mathsf{A}_0$ and the assumption that the set I is finite or countable. For the probability $\pi(A)$, note that Ω_0 is the union of two disjoint sets, namely

$$\Omega_0 = A \cup (\textstyle\bigcup_{i\in I} (\Omega_0 \backslash A_i)). \tag{6.39}$$

(It follows from the definition of A that the union on the right-hand side of Equation (6.39) is identical with $\Omega_0 \backslash A$.) Furthermore, the last equality in

$$\pi(\bigcup_{i\in I} (\Omega_0\backslash A_i)) \leq \sum_{i\in I} \pi(\Omega_0\backslash A_i) = 0 \tag{6.40}$$

follows from $\pi(A_i) = 1$, and since a probability cannot be negative, Inequality (6.40) can be sharpened to an equation. Combining this result with Equation (6.39), we obtain

$$1 = \pi(\Omega_0) = \pi(A) + 0, \tag{6.41}$$

and the set A has the required properties.□

As an immediate application of Lemma 4.3, we can now bring up the proof of Lemma 6.3, which has been postponed in Subsection 6.15.

Transferring the roles of the set Ω_0, the σ-algebra A_0 and the map $\phi:\Omega_0\to V$ in Definition 4.1 to the set D, the σ-algebra A_D and the map $\phi:D\to V$ of Example 3.3, we can show that the SSA $(V, \Pi^{\sim}, \Phi^{\sim}, H^{\sim}, \Phi^{\sim})$ in Lemma 6.3 is a stochastic SSA. So the countable set W* can take the role of the index set I in Lemma 4.3, and Equation (3.19) allows to derive $(S_L, \pi) \in H^{\sim}$ from the premissas of Lemma 6.3.□

6.22 Proof of Lemma 4.6

It suffices to prove the following claim: For every subset S of V and every $\pi \in \Pi$, an element A of A_0 with $\phi'(A) \subseteq S$ and $\pi(A) = 1$ exists iff $(S, \pi) \in H$. So let a subset S of V and an element π be given. Furthermore, let A' be an element of A_0 with $\pi(A') = 1$ and $\phi(A') \subseteq V'$, the existence of a suitable A' being granted by the premissa $(V', \pi) \in H$.

Under the assumption $(S, \pi) \in H$, let A* be an element of A_0 with $\pi(A^*) = 1$ and $\phi(A^*) \subseteq S$, and we will show that the properties $\pi(A) = 1$ and $\phi'(A) \subseteq S$ hold for an element A of A_0 given by the definition $A := A' \cap A^*$. Almost trivially, $\pi(A) = 1$ follows from the assumptions $\pi(A') = 1$ and $\pi(A^*) = 1$ (see Section ?.?). Furthermore, the assumption $\phi(A') \subseteq V'$, which is equivalent with $A' \subseteq \phi^{-1}(V')$ (see Equivalence (?)), implies $A \subseteq \phi^{-1}(V')$. So the premissa $\phi'(\omega) = \phi(\omega)$ for $\omega \in \phi^{-1}(V')$ yields the initial equality in the formula

$$\phi'(A) = \phi(A) \subseteq \phi(A') \subseteq S, \tag{6.42}$$

the first inclusion in the formula being based on the inclusion $A \subseteq A'$ (an immediate consequence of the definition of A, see also Implication (?)), and the second inclusion having been assumed for A'.

Conversely, if A is an element of A_0 with $\phi'(A) \subseteq S$ and $\pi(A) = 1$, define $A^* := A' \cap A$, and $\pi(A^*) = 1$ follows again (Section ?.?), and the formula

$$\phi(A^*) = \phi'(A^*) \subseteq \phi'(A) \subseteq S \tag{6.43}$$

is obtained like Formula (6.42). So A* can take the role of the set A in SSA-Axiom (vii) to obtain $(S, \pi) \in H$.□

6.23 Proof of Corollary 4.9

To verify Equivalence (4.4), let S' be a subset of V', and π an element of the set Π. If $(g^{-1}(S'), \pi) \in H$, let A be an element of $\mathcal{A}_0$ with $\pi(A) = 1$ and $\phi(A) \subseteq g^{-1}(S')$, the existence of a suitable A being obtained from SSA-Axiom (vii). Then the first inclusion in the formula

$$\phi'(A) = g(\phi(A)) \subseteq g(g^{-1}(S')) \subseteq S' \tag{6.44}$$

is an application of Implication (?), the remaining parts of the formula being obtained from the premissa $\phi' = g \circ \phi$ and from Formula (?). Finally, Formula (6.44) leads to $(S', \pi) \in H'$ by another application of SSA-Axiom (vii). Conversely, if $(S', \pi) \in H'$, then similar applications of SSA-Axiom (vii) and of the premissa $\phi' = g \circ \phi$ yield the formula

$$\phi(A) \subseteq g^{-1}(g(\phi(A))) = g^{-1}(\phi'(A)) \subseteq g^{-1}(S') \tag{6.45}$$

for A with $\pi(A) = 1$ and $\phi'(A) \subseteq S'$, and this result implies $(g^{-1}(S'), \pi) \in H$.

Finally, since Equivalence (4.4) has been established, the identity map $f:\Pi\to\Pi'$ can be used in Mapping-Theorem 3.8.(iv) and Isomorphism-Corollary 3.10 to prove the remaining claims of the corollary.□

6.24 Generalisation and Proof of Corollary 4.10

Corollary 4.10 is an immediate application of the subsequent lemma to a situation, where the second SSA of the lemma is identity-based.

Lemma 6.4: Let (V, Π, Φ, H, T) and $(V', \Pi', \Phi', H', T')$ be stochastic SSAs, which are based on measurable spaces $(\Omega_0, \mathcal{A}_0)$ and $(\Omega'_0, \mathcal{A}'_0)$ and on maps $\phi:\Omega_0\to V$ and $\phi':\Omega'_0\to V'$. Furthermore, let a map $g:V\to V'$ be given, define a map $h:\mathcal{A}'_0\to P\Omega_0$ by

$$h(A') := \phi^{-1}(g^{-1}(\phi'(A'))), \tag{6.46}$$

for every $A' \in \mathcal{A}'_0$, and assume that the σ-algebra $\mathcal{A}_0$ contains all sets h(A') with $A' \in \mathcal{A}'_0$. Then the implication (i) ⇒ (ii) holds for the following assertions:

(i) The σ-algebras $\mathcal{A}_0$ and $\mathcal{A}'_0$ have the properties

$$\mathcal{A}_0 = \{h(A'): A' \in \mathcal{A}'_0\}, \tag{6.47}$$

and

$$\phi'^{-1}(g(\phi(\Omega_0))) \in \mathcal{A}'_0. \tag{6.48}$$

(ii) For every $A \in \mathcal{A}_0$, the equation $\phi'(A') = g(\phi(A))$ holds for some element A' of $\mathcal{A}'_0$.

Furthermore, if Equation (6.48) is given and the equation $A = \phi^{-1}(g^{-1}(g(\phi(A))))$ holds for every $A \in \mathcal{A}_0$, then Assertions (i) and (ii) are equivalent.

Now let a map $f:\Pi\to\Pi'$ be given such that the equation

$$f(\pi)(A') := \pi(h(A')) \tag{6.49}$$

holds for every $\pi \in \Pi$ and every $A' \in \mathcal{A}'_0$.

Then the following properties follow:

(iii) The implication

$$(S', f(\pi)) \in H' \;\Rightarrow\; (g^{-1}(S'), \pi) \in H \tag{6.50}$$

holds for every subset S' of the set V' and every $\pi \in \Pi$.

(iv) The reversal of Implication (6.50) holds for every $\pi \in \Pi$ and every subset S' of the set V' with the properties $\phi'(\phi'^{-1}(S')) = S'$ and $\phi'^{-1}(S') \in \mathsf{A}'_0$.

(v) The reversal of Implication (6.50) holds for every $\pi \in \Pi$ and every subset S' of the set V', if one of the properties (i) and (ii) holds for the σ-algebras A_0 and A'_0.

Proof of Lemma 6.4

(i)⇒(ii): To verify for a given element A of A_0 the existence of an element A' of A'_0 with the property $\phi'(A') = g(\phi(A))$, let elements A" and A* of A'_0 be given such that $A'' = \phi'^{-1}(g(\phi(\Omega_0)))$ and $A = h(A^*)$, the existence of suitable sets A" and A* being granted by the assumed validity of Assertion (i). With the definition $A' := A^* \cap A''$, we obtain another element of A'_0. Then Formulas (?), (?) and (?) can be used to derive the properties $\phi'(A') \subseteq g(\phi(\Omega_0))$ and $A = h(A')$, which are used for the outer equalities in the following equation:

$$\phi(A') = \phi(A') \cap g(\phi(\Omega_0)) = g(\phi(\phi^{-1}(g^{-1}(\phi'(A'))))) = g(\phi(h(A'))) = g(\phi(A)) \qquad (6.51)$$

See Formula (?) and Equation (6.46) for the second equality and the third one.

(i)⇔(ii): Since the forward implication has already been established and the validity of Equation (6.48) is assumed for the equivalence, it suffices to derive Equation (6.47) from Assertion (ii) under the additional premissas of the lemma for the equivalence. Now the property $h(A') \in \mathsf{A}_0$ for every $A' \in \mathsf{A}'_0$ is part of the general premissas. Hence Equation (6.47) will be established, if we prove for every $A \in \mathsf{A}_0$ the existence of an element A' of A'_0 with $A = h(A')$. So let A be an arbitrary element of A_0, and let an element A' of A'_0 be given such that $\phi'(A') = g(\phi(A))$, the existence of a suitable A' following from Assertion (ii). Recalling that the equivalence (i)⇔(ii) is claimed only under the additional premissa that the first equality in the following equation holds for every $A \in \mathsf{A}_0$, we obtain:

$$A = \phi^{-1}(g^{-1}(g(\phi(A)))) = \phi^{-1}(g^{-1}(\phi'(A')) = h(A') \qquad (6.52)$$

But then the property $A = h(A')$ is established, since the second equality follows immediately from the assumption $\phi'(A') = g(\phi(A))$, and the last one from Equation (6.46).

(iii): Let a subset S' of V' and an element π of Π be given. To verify Implication (6.50), assume that the ordered pair $(S', f(\pi))$ is contained in the relation H', and - referring to SSA-Axiom (vii) - let an element A' of A'_0 be given such that $f(\pi)(A') =1$ and $\phi'(A') \subseteq S'$. Then $\pi(h(A')) = 1$ follows from Equation (6.49), and the assumption $\phi'(A') \subseteq S'$ leads to $\phi(h(A')) \subseteq g^{-1}(S')$. Combining the results, we obtain $(g^{-1}(S'), \pi) \in H$ from SSA-Axiom (vii), the role of the set A in that axiom being taken by the set h(A').

Since Implication (6.50) has been established, only its reversal has to be proved for Assertions (iv) and (v).

(iv): Under the assumption $(g^{-1}(S'), \pi) \in H$, define $A' := \phi'^{-1}(S')$, and observe that the properties $A' \in \mathsf{A}'_0$ and $\phi'(A') = S'$ follow from the additional premissas $\phi'(\phi'^{-1}(S')) = S'$ and $\phi'^{-1}(S') \in \mathsf{A}'_0$. Now Equation (6.49) and Lemma 4.2 can be combined to obtain the equation

$$f(\pi)(A') = \pi(h(A')) = \pi(\phi^{-1}(g^{-1}(S'))) = 1. \qquad (6.53)$$

So the set A' has all properties to derive $(S', f(\pi)) \in H'$ from SSA-Axiom (vii).
(v): For $(g^{-1}(S'), \pi) \in H$, let A be an element of A_0 with $\pi(A) = 1$ and $\phi(A) \subseteq g^{-1}(S')$. Since the implication (i)⇒(ii) has already been proved, it suffices to derive under the assumed validity of Assertion (ii) the existence of an element A' of A'_0 with the properties $f(\pi)(A') = 1$ and $\phi'(A') \subseteq S'$: Then the reversal of Implication (6.50) will follow from rom SSA-Axiom (vii). We claim that every element A' of A'_0 with $\phi'(A') = g(\phi(A))$ has the desired properties, the existence of an element of A'_0 with this property following from Assertion (ii). First, $A \subseteq h(A')$ follows from Formula (?) and Equation (6.46), and the assumption $\pi(A) = 1$ yields $\pi(h(A')) = 1$, which implies $f(\pi)(A') = 1$ by Equation (6.49). Combining assumptions about the sets A and A', we obtain

$$\phi(A') = g(\phi(A)) \subseteq g(g^{-1}(S')) \subseteq S'. \tag{6.54}$$

The opening equality has been assumed, and the first inclusion follows from Formula (?) under the assumption $\phi(A) \subseteq g^{-1}(S')$, whereas the last inclusion is based on Formula (?).□

The rather complicated set theoretical structure of the Lemma 6.4 is a price for its generality: It applies to situations, where neither one of the two SSAs must be identity-based. But in many applications of Mapping-Theorem 3.8 and its corollaries to stochastic SSAs, the SSA $(V', \Pi', \Phi', H', T')$ is identity-based, and then we have the situation of Corollary 4.10, which doesn't need a separate proof, since it is an almost verbatim translation of Lemma 6.4 to situations where the second SSA is identity-based.

Note that the assumptions $\phi'(\phi'^{-1}(S')) = S'$ and $\phi'^{-1}(S') \in A'_0$ in Assertion (iv) of Lemma 6.4 can be replaced by the following weaker premissa: For every $\pi \in \Pi$ and every $A \in A_0$ with the properties $\pi(A) = 1$ and $\phi(A) \subseteq g^{-1}(S')$, the σ-algebra A'_0 contains an element with $\phi'(A') \subseteq S'$ and $A \subseteq h(A')$. Under this premissa, assume $(g^{-1}(S'), \pi) \in H$, let A be an element of A_0 with $\pi(A) = 1$ and $\phi(A) \subseteq g^{-1}(S')$, and A' an element of A'_0 with the properties described in the alternative assumption. Then $f(\pi)(A') = 1$ follows from Equation (6.49).

If this premissa holds for every subset S' of V', then this property can also replace the additional premissa of Assertion (v), which requires that A'_0 contains sets A' with the property $\phi'(A') = g(\phi(A))$ for every $A \in A_0$.

If even this weakened premissa doesn't hold, the reversal of Implication (6.50) may be regained in some situations by changes in the σ-algebras A_0 and A'_0. A_0 may be replaced by a coarser σ-algebra in Ω_0, where all elements of A_0 violating the premissa are dropped. Similarly, a finer σ-algebra in Ω'_0 instead of a given A'_0 can be obtained, if suitable sets A' are added to A'_0 and the σ-algebra generated by the resulting set system is used as a new σ-algebra in Ω'_0. However, the requirement that A_0 must contain all sets h(A') with $A' \in A'_0$ has to be observed as a limitation for such transitions. Furthermore, it has to be noted that all components of the SSAs with the exception of the vocabulary sets may be changed by alterations in the σ-algebras A_0 and A'_0.[186]

[186] See e.g. Footnote 149 for consequences of a transition to a finer σ-algebra in identity-based stochastic SSAs.

6.25 Proof of Corollary 4.11

(i): Let S be an element of the σ-algebra $\mathsf{A}_{V'}$. Then the implication $S \in T' \Rightarrow S \in T$ can be obtained from Mapping-Theorem 3.8.(i), since the premissas of that theorem are granted by Corollary 4.10.(iv) and Equation (4.10). (The identity of the sets S and $g^{-1}(S)$ is trivial, if the map $g{:}V \to V'$ is the identity map.)
(ii): The premissa of Isomorphism-Corollary 3.10 referring to the aggregation rules Φ and Φ' is reintroduced by Equation (4.10), whereas the necessary properties of the relations H and H' follow from Corollary 4.10.(iii) and (v).
(iii): Assume that the map $f{:}\Pi \to \Pi'$ is surjective, and let S be an arbitrary element of the set system T. To show that S is contained in T', let π' an element of Π' with $(S, \pi') \in H'$, and π an element of Π with $\pi' = f(\pi)$ (the existence of a suitable π being granted by the assumed surjectivity of the map f). Then $(S, \pi) \in H$ follows from Corollary 4.10.(iii), and this implies $\Phi(\pi) \in S$ under the assumption $S \in T$. But then $\Phi'_{\pi'} \in S$ is obtained from Equation (4.10).□

6.26 Tools for the Analysis of Expectational Maps

Whereas results with immediate relevance for applications of expectational maps are presented in Section 4, the following lemma is a sort of toolbox for the analysis of expectational maps such that the availability of the tools saves laborious discussions of mathematical subtleties in the following subsections. For convenience, an explicit mentioning of Definition 4.13.(ii) is omitted in references to its Properties (ii.a), (ii.b), (ii.c) and (ii.d).

Lemma 6.5: For a stochastic SSA (V, Π, Φ, H, T) with a stochastic base $(\Omega_0, \mathsf{A}_0, \phi)$, the following properties hold for every subset S of V and every subset Π* of Π:

(i) The set V has Properties (ii.a) and (ii.b) for every $\pi \in \Pi$.
(ii) The equality $\chi_V \cdot g = g$ holds for every map $g{:}V \to \mathbb{R}$.
(iii) If π is an element Π such that S has Properties (ii.a) and (ii.b), then $\pi(\phi^{-1}(S)) = 1$, and $\chi_S(\phi_U) =_{\pi\text{-a.s.}} 1$ for every random variable U with distribution π.
(iv) The map $\chi_S \circ \phi$ is A_0-B-measurable iff $\phi^{-1}(S) \in \mathsf{A}_0$.
(v) If S has Properties (ii.b) and (ii.c) for a given map $g{:}V \to \mathbb{R}$ and A is a Borel set in $\mathbb{R}$, then the set $S \cap g^{-1}(A)$ has Property (ii.b).
(vi) If S' is a subset of V with $S' \subseteq S$ and $\phi^{-1}(S') \in \mathsf{A}_0$, and $g{:}V \to \mathbb{R}$ is a map such that the map $(\chi_S \cdot g) \circ \phi$ is measurable, then the map $(\chi_{S'} \cdot g) \circ \phi$ is also measurable, and the equation
$$\mathsf{E}_{U\sim\pi}\,(\chi_{S'}(\phi_U) \cdot g(\phi_U)) = \mathsf{E}_{U\sim\pi}\,(\chi_S(\phi_U) \cdot g(\phi_U)) \qquad (6.55)$$
holds for every $\pi \in \Pi$ with $(S', \pi) \in H$, where one of the two expectations exists.
(vii) If $g{:}V \to \mathbb{R}$ is a map, π and element of Π, and S' is a subset of V with $S \subseteq S'$ such that S and S' have Properties (ii.a), (ii.b) and (ii.c) and S has Property (ii.d), then S' has Property (ii.d).
(viii) The set V has Properties (ii.a), (ii.b), (ii.c) and (ii.d) for every $\pi \in \Pi^*$ and every map $g{:}V \to \mathbb{R}$, which is almost expectational in Π* and where the map $g \circ \phi$ is A_0-B-measurable.
(ix) If π is an element of Π, $g{:}V \to \mathbb{R}$ a map, and S has Properties (ii.c) and (ii.d), then every subset of S with Properties (ii.a) and (ii.b) has Properties (ii.c) and (ii.d).
(x) If S has Properties (ii.a), (ii.b), (ii.c) and (ii.d) for a map $g{:}V \to \mathbb{R}$ and an element π of the set Π, and S' is a subset of V with Properties (ii.a) and (ii.b), then the set $S \cap S'$ has Properties

(ii.a), (ii.b), (ii.c) and (ii.d).

(xi) If S has Properties (ii.a), (ii.b), (ii.c) and (ii.d) for a map $g:V \to \mathbb{R}$ and an element π of the set Π, and S' is a subset of V with Properties (ii.a), (ii.b), and (ii.c), then S' has Property (ii.d).

(xii) If G is a finite or countable set of maps $V \to \mathbb{R}$, which are expectational in Π^*, and π an element of Π^*, then there exists a subset S' of V with Properties (ii.a), (ii.b), (ii.c) and (ii.d) for every $g \in G$.

Proof of Lemma 6.5

In this proof, measurability always means A_0-B-measurability.
(i): Property (ii.a) of the set V is granted by SSA-Axiom (ii), and the equality $\phi^{-1}(V) = \Omega_0$ yields Property (ii.b).
(ii): This property is an immediate consequence of the equality $\chi_V(v) = 1$ for every $v \in V$.
(iii): See Lemma 4.2 for this claim.
(iv): For a given set A, which is a Borel set in $\mathbb{R}$, the definition of the characteristic function χ_S in ?.? implies:
- If $0 \in A$ and $1 \in A$, then $\phi^{-1}(\chi_S{}^{-1}(A)) = \Omega_0$.
- If $0 \in A$ and $1 \notin A$, then $\phi^{-1}(\chi_S{}^{-1}(A)) = \Omega_0 \backslash \phi^{-1}(S)$.
- If $0 \notin A$ and $1 \in A$, then $\phi^{-1}(\chi_S{}^{-1}(A)) = \phi^{-1}(S)$.
- If $0 \notin A$ and $1 \notin A$, then $\phi^{-1}(\chi_S{}^{-1}(A)) = \varnothing$.

Obviously, the map $\chi_S \circ \phi$ isn't measurable, if $\phi^{-1}(S) \notin \mathsf{A}_0$, since $\phi^{-1}(\chi_S{}^{-1}(A)) \notin \mathsf{A}_0$ for $A = \{1\}$. Conversely, if $\phi^{-1}(S) \in \mathsf{A}_0$, then the above properties lead to $\phi^{-1}(\chi_S{}^{-1}(A)) \in \mathsf{A}_0$ for every Borel set A in $\mathbb{R}$, and then the map $\chi_S \circ \phi$ is measurable.
(v): Consider the equation

$$\phi^{-1}(S \cap g^{-1}(A)) = \phi^{-1}(S \cap (\chi_S \cdot g)^{-1}(A)) = \phi^{-1}(S) \cap \phi^{-1}((\chi_S \cdot g)^{-1}(A)). \tag{6.56}$$

For the first equality, note that membership in the sets $g^{-1}(A)$ and $(\chi_S \cdot g)^{-1}(A)$ is equivalent for elements of S, since $\chi_S(v) = 1$ for $v \in S$. So the intersection with S removes all differences between the two sets. The second equality in Equation (6.56) is an application of Equation (?). Furthermore, Equation (6.56) implies that the set $\phi^{-1}(S \cap g^{-1}(A))$ is contained in A_0, since it is the intersection of two elements of this σ-algebra. (For this argument, refer to the assumed Properties (ii.b) and (ii.c) of the set S.)
(vi): In the assumed situation, the equation

$$(\chi_{S'} \cdot g) \circ \phi = (\chi_{S'} \cdot \chi_S \cdot g) \circ \phi = (\chi_{S'} \circ \phi) \cdot ((\chi_S \cdot g) \circ \phi) \tag{6.57}$$

follows from ?.? and ?, and Assertion (iv) yields the measurability of the map $\chi_S \circ \phi$. So Equation (6.57) tells that the map $(\chi_{S'} \cdot g) \circ \phi$ is measurable, since it is the (pointwise) product of two measurable maps. (See ### for this conclusion.) Now let π be an element of Π with $(S', \pi) \in H$. Then $\pi(\phi^{-1}(S')) = 1$ is obtained from Lemma 4.2, and this implies $\chi_{S'}(\phi_U) =_{\pi.\text{-a.s.}} 1$ for a random variable U with distribution π. Combining this result with Equation (6.57), we obtain Equation (6.55), if one of the two expectations in this equation exists.
(vii): In the assumed situation, Equation (6.57) holds with interchanged roles of S and S'. So

Equation (6.55) can be derived as before, and the existence of the expectation on the right hand side of the equation is granted by the assumed Property (ii.d) of S.
(viii): Properties (ii.a) and (ii.b) of the set V have already been established (Assertion (i)), and Property (ii.c) is obtained from Assertion (ii) and from the assumed measurability of the map $g \circ \phi$. So the set V can take the role of the set S' in Assertion (vii) to verify Property (ii.d) of the set V for every $\pi \in \Pi$.
(ix): This assertion is an immediate consequence of Assertion (vi), the existence of the expectation on the right hand side of Equation (6.57) being again granted by the assumed Property (ii.d) of S.
(x): In the assumed siutation, Properties (ii.a) and (ii.b) of the set $S \cap S'$ are obtained from Lemma 4.3 and from the equation $\phi^{-1}(S \cap S') = \phi^{-1}(S) \cap \phi^{-1}(S')$, which is an application of Equation (?). So Properties (ii.c) and (ii.d) of the set $S \cap S'$ are obtained from Assertion (ix).
(xi): Since Assertion (x) yields Properties (ii.a), (ii.b), (ii.c) and (ii.d) of the set $S \cap S'$, this set can take the role of the set S in Assertion (vii) to obtain Property (ii.d) of the set S'.
(xii): In the assumed situation, let π be an arbitrary element of Π^*, and $\{S_g\}_{g \in G}$ a family of subsets of G such that Properties (ii.a), (ii.b), (ii.c) and (ii.d) hold for every $g \in G$ and the respective set S_g. With the definition $S' := \bigcap_{g \in G} S_g$, Property (ii.a) of the set S' follows from Lemma 4.3 and the assumed Property (ii.a) of the sets S_g. Similarly, the equality $\phi^{-1}(S') = \bigcap_{g \in G} \phi^{-1}(S_g)$, which is an immediate consequence of the definiton of S', can be combined with Property (ii.b) of the maps S_g to verify Property (ii.b) of the set S'. (Recall that the set G is assumed to be finite or countable.) Finally, let g be an arbitrary element of the set G, and transfer the role of the sets S in Assertion (ix) to the set S_g of the present situation to obtain Properties (ii.c) and (ii.d) of the set S' for the given g.□

6.27 Proof of Lemma 4.14

(i): For a given map $g:V \to \mathbb{R}$, Properties 4.13.(ii.a) and (ii.b) follow for the set V and every $\pi \in \Pi^*$ from Lemma 6.5.(i). Furthermore, a reference to Lemma 6.5.(ii) shows that Property 4.13.(ii.c) of the set V is equivalent with the A_0-B-measurability of the map $g \circ \phi$, and that Property 4.13.(ii.d) of the set V for an element π of the set Π^* is equivalent with Equation (4.11). Taken together, these results confirm the claimed equivalence.
(ii): If the map g is expectational in Π^*, then the existence of a stochastic base $(\Omega_0, \mathsf{A}_0, \phi)$ and a set S fulfilling the requirements of Definition 4.13.(ii) follows immediately from Assertion (i): We can take S = V.
(iii): Let $g:V \to \mathbb{R}$ be a map, which is almost expectational in Π^*. If there is no stochastic base $(\Omega_0, \mathsf{A}_0, \phi')$ such that the map $g \circ \phi'$ is A_0-B-measurable, then Definition 4.13.(i) implies that the map g cannot be expectational. Conversely, if there map $g \circ \phi$ is A_0-B-measurable, the Assertion (i) can be combined with Lemma 6.5.(viii) to show that the map g is expectational.
(iv): For a given element π of Π^*, the existence of a subset of V with Properties 4.13.(ii.a), (ii.b), (ii.c) and (ii.d) is granted by the assumption that the map g is almost expectational in Π^*. So the present set S can take the role of the set S' in Lemma 6.5.(xi) to obtain its Property 4.13.(ii.d), which can be rewritten as

$$\chi_S(\Phi_\pi) \cdot g(\Phi_\pi) = \mathsf{E}_{U \sim \pi} (\chi_S(\phi_U) \cdot g(\phi_U)), \tag{6.58}$$

since the premissa $\Phi_\pi \in S$ is equivalent with $\chi_S(\Phi_\pi) = 1$.
(v): If the requirements of Definition 4.13.(i) resp. (ii) hold for every element of the set Π^*, then they hold for every element of any subset of Π^*.
(vi): If the requirements of Definition 4.13.(i) resp. (ii) hold for every element of every member of the family, then they hold for every element of the union.□

6.28 Proof of Lemma 4.17

Let (Ω_0, A_0) be a measurable space, and $\phi:\Omega_0 \to V$ a map such that the considered SSA is based on them.

For a constant map $g:V\to\mathbb{R}$, let ξ the unique real number with $g(v) = \xi$ for every $v \in V$. Then the A_0-B-measurability of the map $g \circ \phi$ and the equation $\mathsf{E}_{U\sim\pi}\, g(\phi_U) = \xi = g(\Phi_\pi)$ for every $\pi \in \Pi^*$ follow from basic properties of constant maps. So the map g is expectational.

Now let G be the set of all maps $g:V\to\mathbb{R}$, which are almost expectational in Π^*, let elements g_1 and g_2 of G and real numbers λ_1 und λ_2 be given. To show that the map $g := \lambda_1\, g_1 + \lambda_2\, g_2$ is contained in G, let π be an element of Π^*, and S a subset of V such that the properties required for a almost expectational map by Definition 4.13.(ii) hold for the given π and for the maps g_1 and g_2. (To verify the existence of a suitable set S, the set $\{g_1, g_2\}$ can take the role of the set G in Lemma 6.5.(xii)). Furthermore, define maps $g':V\to\mathbb{R}$, $g'_1:V\to\mathbb{R}$ and $g'_2:V\to\mathbb{R}$ by $g' := \chi_S \cdot g$, $g'_1 := \chi_S \cdot g_1$, and $g'_2 := \chi_S \cdot g_2$. Combined with the definition of the map g, these definitions imply $g' \circ \phi = \lambda_1\,(g'_1 \circ \phi) + \lambda_2\,(g'_2 \circ \phi)$. So the map $g' \circ \phi$ inherits the A_0-B-measurability of the maps $g'_1 \circ \phi$ and $g'_2 \circ \phi$, which is a part of the assumed properties of the set S. The same assumptions yield the second equality in the equation

$$\mathsf{E}_{U\sim\pi}\, g'(\phi_U) = \lambda_1\, \mathsf{E}_{U\sim\pi}\, g'_1(\phi_U) + \lambda_2\, \mathsf{E}_{U\sim\pi}\, g'_2(\phi_U) = \lambda_1\, g'_1(\Phi_\pi) + \lambda_2\, g'_2(\Phi_\pi) = g'(\Phi_\pi) \tag{6.59}$$

So the set S has the properties required by Definition 4.13.(ii) for the map g and the given element π of the set Π^*. But if a set S with these properties exists for every $\pi \in \Pi^*$, then the map g is expectational in Π^*.

An application of this result to a situation with $\lambda_2 = 0$ shows that the set G is closed under scalar multiplication, and the property of being closed under addition is obtained from the same result for $\lambda_1 = \lambda_2 = 1$. But if the set G is closed under scalar multiplication and addition, then it is a linear subspace of the function space $\mathbb{R}^V$, and this implies that all linear combination of elements of G result in an element of G.

The claim of the lemma for expectational maps is obtained in the same way under the additional assumption $S = V$, i.e., $g' := g$, $g'_1 := g_1$, and $g'_2 := g_2$.□

6.29 Proof of Convergence-Theorem 4.19

For this proof, define $\mathbb{R}^* := \mathbb{R} \cup \{-\infty, +\infty\}$, and let B^* be the σ-algebra of all subsets A of $\mathbb{R}^*$ with $A \cap \mathbb{R} \in \mathsf{B}$.

To prove the claim $\phi^{-1}(S^*) \in \mathsf{A}_0$, let a sequence $\{f_n\}_{n=1..\infty}$ of maps $f_n:\Omega_0\to\mathbb{R}$ and a map $f:\Omega_0\to\mathbb{R}^*$ be defined by $f_n(\omega) := g_n(\phi(\omega))$, and $f(\omega) := \sup_{n\in\mathbb{N}} f_n(\omega)$ for every $\omega \in \Omega_0$, including $f(\omega) := +\infty$

for $\phi(\omega) \in V \backslash S^*$. Since the maps $g_n \circ \phi$ are assumed to be A_0-B-measurable, the A_0-B^*-measurability of the maps f_n and f is obtained from elementary results of measure theory.[187] But then the property $\phi^{-1}(S^*) \in A_0$ follows immediately, since the definition of the set S* implies $\phi^{-1}(S^*) = f^{-1}(\mathbb{R})$.

For the A_0-B-measurability of the map $g \circ \phi$, define a map $f^*:\Omega_0 \to \mathbb{R}$ by $f^*(\omega) := \chi_{S^*}(\phi(\omega))$. Obviously, this map is A_0-B^*-measurable, and the equality $g(\phi(\omega)) = f(\omega) \cdot f^*(\omega)$ follows from the definitions of the involved maps. So the map $g \circ \phi$ is A_0-B-measurable.[188]

To verify the concluding claims of the theorem, let π be an arbitrary element of the set $\Pi^{\sim}$, and note that the equation

$$\pi(\phi^{-1}(S^*)) = 1 \tag{6.60}$$

is obtained from $(S^*, \pi) \in H$ (definition of $\Pi^{\sim}$) and from Lemma 4.2. To verify both the forward implication in Equivalence (4.18) and the claimed expectational property of the map g, assume furthermore $\Phi_\pi \in S^*$. Then the definitions of the set S* and of the map g yield the first equality in the following equation, where sup and E stand for $\sup_{n\in\mathbb{N}}$ resp. $E_{U\sim\pi}$:

$$g(\Phi_\pi) = \sup g_n(\Phi_\pi) = \sup E\, g_n(\phi_U) = \sup E\, f_n(U) = E\, f(U) = E_{U\sim\pi}\, g(\phi_U). \tag{6.61}$$

The second equality is obtained from the premissa that the maps g_n are expectational in $\Pi^{\sim}$, and the third equality from the above definition of maps $f_n:\Omega_0 \to \mathbb{R}$. The fourth equality is based on Levi's theorem of monotone convergence.[189] For the last equality in Equation (6.61), observe that the equality $f(\omega) = g(\phi(\omega))$ is granted for $\omega \in \phi^{-1}(S^*)$ by the definitions of the maps f and g, and that Equation (6.60) allows to rewrite this property as $f(U) =_{\pi\text{-a.s.}} g(\phi_U)$ for a random variable U with distribution π. In summary, Equation (6.61) holds for every $\pi \in \Pi^{\sim}$ with $\Phi_\pi \in S^*$. So the forward implication in Equivalence (4.18) is established for every $\pi \in \Pi$, and the map g is expectational in Π^*.

It is left to prove the backward implication in Equivalence (4.18). Now the right hand side of this equivalence implies that the expecation $E_{U\sim\pi}\, g(\phi_U)$ is finite, since $g(v) \geq 0$ for every $v \in V$. Furthermore, combining the definition of the map g with Inequality (4.16) and Equation (6.60), we obtain $g(\phi_U) \geq_{\pi\text{-a.s.}} g_n(\phi_U)$ for a random variable U with distribution π, and this implies

$$E_{U\sim\pi}\, g(\phi_U) \geq E_{U\sim\pi}\, g_n(\phi_U) = g_n(\Phi_\pi) \tag{6.62}$$

under the premissa that the maps g_n are expectational in $\Pi^{\sim}$. In other words, Inequality (6.62) tells that the sequence $\{g_n(\Phi_\pi)\}_{n=1..\infty}$ has a finite upper bound, if $E_{U\sim\pi}\, g(\phi_U)$ is finite. Furthermore, this sequence is non-decreasing due to Inequality (4.16). But then $\lim_{n\to\infty} g_n(\Phi_\pi) < +\infty$ follows

---

187 See e.g. Bauer (1992, p. 59, Proposition 9.5) for f.

188 See e.g. Bauer (1992, p. 59, Proposition 9.4) to derive the A_0-B^*-measurability of the map $g \circ \phi$. Then the map $g \circ \phi$ is also A_0-B-measurable, since $g(\phi(\omega))$ is finite for every $\omega \in \Omega_0$.

189 See e.g. Bauer (1992, p. 68, Proposition 11.4).

immediately, i.e., $\Phi_\pi \in S^*$.□

Like Levi's theorem of monotone convergence, Convergence-Theorem 4.19 formulates a basic fact, whose premissas can be adapted to other situations. For instance, violations of the assumption $0 \le g_n(v)$ in Inequality (4.16) can be overcome by a simple redefinition, if $g_n(v) \le g_{n+1}(v)$ for $n = 1..\infty$ and for every $v \in V$. With a sequence $\{g'_n\}_{n=1..\infty}$ of maps $g'_n:V\to\mathbb{R}$ and a map $g':V\to\mathbb{R}$ given by $g'_n(v) := g_n(v) - g_1(v)$ for $v \in V$, and $g'(v) := g(v) - g_1(v)$ for $v \in S^*$, Lemma 4.17 can be used to show that the maps g'_n are expectational in $\Pi^\sim$ if the maps g_n have this property, and that the map g is expectational in Π^* if this holds for g' and g_1.[190] Note, however, that this approach can be applied only in situations, where a map $g:V\to\mathbb{R}$ of interest can be conceived as the difference of maps g' and g_1, where g' can be treated by the Convergence-Theorem, whereas the expectational property of the map g_1 has to be established separately.

A more general approach is the representation of a map g of interest as a difference $g^+ - g^-$ of maps with non-negative function values, which is well known from integration theory.[191] If suitable sequences $\{g^+_n\}_{n=1..\infty}$ and $\{g^-_n\}_{n=1..\infty}$ of maps exist, then results derived from Convergence-Theorem 4.19 for the maps g^+ and g^- can be combined. In particular, since g is a linear combination of the maps g^+ and g^-, Lemma 4.17 can be used to show that g is expectational in $\Pi^{*+} \cap \Pi^{*-}$, if g^+ and g^- are expectational in Π^{*+} resp. Π^{*-}.[192]

Observe also that the function value $g(v) = 0$ for $v \in V\backslash S^*$ is an arbitrary dummy value for situations with $\lim_{n\to\infty} g_n(v) = +\infty$. In fact, every assignment of real numbers, which results in an A_0-B-measurable map $g \circ \phi$, can be made as well for $v \in V\backslash S^*$.[193]

If the maps $g_n:V\to\mathbb{R}$ are only almost expectational in $\Pi^\sim$ and the A_0-B-measurability of the maps $g_n \circ \phi$ is not assumed, then the claims of the Convergence-Theorem may fail. But in some situations, there may be a subset S of the set V fulfilling the premissas of Lemma 4.14.(iv) for all maps g_n, and then the theorem may be applied to the maps $\chi_S \cdot g_n$, whose property of being expectational in Π^* is obtained from Lemma 4.14.(iv). In particular, if sets S^* and Π^* are specified on the basis of the maps g_n as in the theorem, then the property $\phi^{-1}(S^*) \in \mathsf{A}_0$ and the A_0-B-measurability of the map $(\chi_{S^*} \cdot g_n) \circ \phi$ are sufficient to transfer the role of the set S in Lemma 4.14.(iv) to the set S^*. (Recall

---

190 It can be left to the reader to show that the validity of premissas referring to the sets S^*, $\Pi^\sim$ and Π^* of Convergence-Theorem 4.19 is not changed by the transition.

191 More explicitly, the maps g^+ and g^- are defined by
$g^+(v) := g(v)$ and $g^-(v) := 0$ for $g(v) \ge 0$,
and
$g^+(v) := 0$ and $g^-(v) := |g(v)|$ for $g(v) < 0$.

192 Exercise!

193 Recall from Definition 4.13.(i) that a map $g:V\to\mathbb{R}$ cannot be expectational in Π^*, if the map $g \circ \phi$ isn't A_0-B-measurable. Conversely, if $g':V\to\mathbb{R}$ is a map with $g'(v) = g(v)$ for $v \in S^*$, where the map $g' \circ \phi$ is A_0-B-measurable, then the equality $g'(\phi_U) =_{\pi\text{-a.s.}} g(\phi_U)$ holds for a random variable U with distribution $\pi \in \Pi^\sim$. (See Equation (6.60) in the proof of Convergence-Theorem 4.19 in Section 6.29 for this conclusion). But this almost sure equality implies $\mathsf{E}_{U\sim\pi}\, g'(\phi_U) = \mathsf{E}_{U\sim\pi}\, g(\phi_U)$.

that the specification of the set Π^* in the Convergence-Theorem implies $(S^*, \pi) \in H$ and $\Phi_\pi \in S^*$ for every $\pi \in \Pi^*$.)

If only almost expectation maps are available and a set S with the assumed properties doesn't exist, another reconceptualisation may be helpful. For every element π of a set Π^*, the set $\{\pi\}$ can take the role of the set Π^* of the set Π^* in the Convergence-Theorem. Then a different set S and a different sequence of almost expectational maps g_n may be chosen for every such set $\{\pi\}$ in the above outlined application of Lemma 4.14.(iv). If this approch leads to the conclusion that the map g is almost expectational in every such set $\{\pi\}$, then Lemma 4.14.(vi) tells that the map is also almost expectational in their union, i.e., in the set Π^*.

Although such indirect uses of Convergence-Theorem 4.19 cover many applications, the following lemma may be more helpful in some situations, since its premissas are weaker.

Lemma 6.6: For a stochastic SSA (V, Π, Φ, H, T), which is based on an measurable space (Ω_0, A_0) and a map $\phi:\Omega_0 \to V$, let $g:V \to \mathbb{R}$ a map such that the map $g \circ \phi$ is A_0-B-measurable. Furthermore, let Π^* be a subset of Π with the following property: For every $\pi \in \Pi^*$, there is a subset S^* of the set V, a map $f:V \to \mathbb{R}$ and a sequence $\{g_n\}_{n=1..\infty}$ of maps $g_n:V \to \mathbb{R}$ such that the following assertions hold:

(i) $(S^*, \pi) \in H$.

(ii) The map $f \circ \phi$ is A_0-B-measurable, and the expectation $\mathsf{E}_{U\sim\pi}\, f(\phi_U)$ is finite.

(iii) The maps $g_n \circ \phi$ are A_0-B-measurable, and $\mathsf{E}_{U\sim\pi}\, g_n(\phi_U) = g_n(\Phi_\pi)$ for $n = 1..\infty$.

(iv) $|g_n(v)| \le f(v)$ for $n = 1..\infty$ and for every $v \in S^*$.

(v) $g(v) = \lim_{n\to\infty} g_n(v)$ for $v = \Phi_\pi$ and for every $v \in S^*$.

Then the map g is expectational in Π^*.

Before the lemma is proved, note that its premissas are weaker than those of Convergence-Theorem 4.19:

- In the Convergence-Theorem, the set S^* and the maps g_n must be identical for every $\pi \in \Pi^*$. Conversely, Lemma 6.6 leaves it open that different sets S^* and different maps g_n are used for every $\pi \in \Pi^*$ to fulfill Premissas (i) through (v).
- In the Convergence-Theorem, the set S^* is completely determined by the maps g_n.
- Inequality (4.16) is not required by Lemma 6.6.

At first glance, some premissas of Lemma 6.6 may seem stronger than those of the Convergence-Theorem: The measurability of the map $g \circ \phi$, the existence of a map $f:V \to \mathbb{R}$ fulfilling Premissas (ii) and (iv), and the convergence $g(v) = \lim_{n\to\infty} g_n(v)$ for $v = \Phi_\pi$ in Premissa (v) are explicitly assumed. But since both properties follow from the premissas of the Convergence-Theorem, their explicit introduction doesn't make the premissas of Lemma 6.6 materially stronger than those of the Convergence-Theorem. See the respective discussion in Section 6.29 after Example 4.20 for advantages of the stronger premissas of the Convergence-Theorem.

Proof of Lemma 6.6

Let an element π of the set Π^*, a subset S^* of the set V, a map $f:V \to \mathbb{R}$ and a sequence $\{g_n\}_{n=1..n}$ of maps $g_n:V \to \mathbb{R}$ be given with the properties specified by the premissas. Furthermore, let A be an element of A_0 with $\pi(A) = 1$ and $\phi(A) \subseteq S^*$, the existence of a suitable set A being granted by Premissa (i) and SSA-Axiom (vii). Then the first equality in the equation

$$g(\Phi_\pi) = \lim_{n\to\infty} g_n(\Phi_\pi) = \lim_{n\to\infty} \mathsf{E}_{U\sim\pi}\, g_n(\phi_U) = \mathsf{E}_{U\sim\pi}\, g(\phi_U). \tag{6.63}$$

follows immediately from Premissa (v), and the second equality from Premissa (iii). Furthermore, the properties $|g_n(\phi_U)| \leq_{\pi\text{-a.s.}} f(\phi_U)$ for n = 1..∞, and $g(\phi_U) =_{\pi.\text{-a.s.}} \lim_{n\to\infty} g_n(\phi_U)$ of a random variable U with distribution π follow from the assumptions $\pi(A) = 1$ and $\phi(A) \subseteq S^*$ and from Premissas (iv) and (v). Combining this result with Premissa (ii), we obtain the last equality in Equation (6.63) from an application of Lebesgues' theorem of dominatated convergence (###) to the maps $g \circ \phi$, $f \circ \phi$ and $g_n \circ \phi$. But if Equation (6.63) holds for every $\pi \in \Pi^*$, and if the map $g \circ \phi$ is A_0-B-measurable (general premissa), then the map g is expectational in Π^* (Definition 4.13.(i)).□

6.30 Proof of RSO-Corollary 4.21

The proof consists of two parts. First, the corollary will be proved under the following Premissas (ii') and (iii') instead of Premissas (ii) and (iii):

(ii') For every $A \in \mathsf{A}_W$, the map $pr_A \circ \gamma \circ \phi$ is A_0-B-measurable.

(iii') For every map $\mathsf{S}\to\mathbb{R}$, which has an extension to a probability measure on A_W, this extension is unique.

In a second part, we will show that Premissas (i), (ii) and (iii) of the corollary imply the validity of Premissas (ii') and (iii').

So assume first that Premissas (ii') and (iii') hold.

(iv): Let π be an element of Π_0, and define a map $v':\mathsf{A}_W\to\mathbb{R}$ by the equation

$$v'(A) := \mathsf{E}_{U\sim\pi}\, pr_A(\gamma(\phi_U)) \tag{6.64}$$

for every $A \in \mathsf{A}_W$. We will first show that this map is a probability measure on A_W, and then it will be easy to show that every map $pr_A \circ \gamma \circ \phi$ with $A \in \mathsf{A}_W$ is expectational in Π_0.

The property of being a probability distribution on A_W will be established for v', if we prove the following properties:

$v'(W) = 1$.

$0 \leq v'(A) \leq 1$ for every $A \in \mathsf{A}_W$.

$v'(\bigcup_{i=1..\infty} A_i) := \sum_{i=1..\infty} v'(A_i)$ for every sequence $\{A_i\}_{i=1..\infty}$ of mutually disjoint elements of A_W.

The first two properties follow immediately from Equation (6.64) and from the equality $pr_W(\gamma(\phi_\omega)) = 1$ resp. the inequality $0 \leq pr_A(\gamma(\phi_\omega)) \leq 1$, whose validity for every $\omega \in \Omega_0$ and every $A \in \mathsf{A}_W$ is granted by the premissa that the elements of the set V' are probability measures on A_W. For a given sequence $\{A_i\}_{i=1..\infty}$ of mutually disjoint elements of A_W, let an element A* of A_W be given by the definition $A^* := \bigcup_{i=1..\infty} A_i$. Furthermore, let a map $f:\Omega_0\to\mathbb{R}$ and a sequence $\{f_n\}_{n=1..\infty}$ of maps $f_n:\Omega_0\to\mathbb{R}$be defined by $f(\omega) := pr_{A^*}(\gamma(\phi_\omega))$ and $f_n(\omega) := \sum_{i=1..n} pr_{A(i)}(\gamma(\phi_\omega))$ for every $\omega \in \Omega_0$ and n = 1..∞. Then the outer equalities in the equation

$$v'(A^*) = \mathsf{E}_{U\sim\pi}\, f(U) = \sup_{n=1..\infty} \mathsf{E}_{U\sim\pi}\, f_n(U) = \sum_{i=1..\infty} v'(A_i) \tag{6.65}$$

follow from these definitions and from Equation (6.64), whereas the second equality is based on Levi's theorem of monotone convergence (see ###). For this conclusion, note that the properties $0 \leq f_n(\omega) \leq f_{n+1}(\omega) \leq f(\omega)$ for n = 1..∞, and $f(\omega) = \lim_{n\to\infty} f_n(\omega)$ are granted for $\omega \in \Omega_0$ by the definitions of the involved maps and by the premissa that the elements of the set V' are probability measures on A_W.

In summary, we have proved that v'(A) is a probability measure on A_W. Now let a map $v'':\mathsf{S}\to\mathbb{R}$ be the restriction of v' to the set system S, and note that Premissa (i) can be combined with Definition 4.13.(i) to obtain

$$pr_A(\gamma(\Phi_\pi)) = \mathsf{E}_{U\sim\pi}\, pr_A(\gamma(\phi_U)) = v''(A) \tag{6.66}$$

for every $A \in \mathsf{S}$. In other words, both $\gamma(\Phi_\pi)$ and v' are extensions of v'' to a probability measure on A_W. Since an extension of this kind is unique, if it exists (Premissa (iii')), we obtain $\gamma(\Phi_\pi) = v'$. So the first equality in Equation (6.66) holds for every $A \in \mathsf{A}_W$, and also for every $\pi \in \Pi_0$. But then the map $pr_A \circ \gamma$ is expectational in Π_0 for every $A \in \mathsf{A}_W$. (Recall that the A_0-B-measurability, which is also required for an expectational map by Definition 4.13.(i), is assumed for the maps $pr_A \circ \gamma \circ \phi$ by Premissa (ii').)

(ix): Let

$$X = X^+ - X^- \tag{6.67}$$

be the usual decomposition of the map $X:W\to\mathbb{R}^*$.[194] To prepare an application of Convergence-Theorem 4.19 to the present situation, let a sequence $\{X^+_n\}_{n=1..\infty}$ of maps $X^+_n:W\to\mathbb{R}$ be given with $X^+(w) = \lim_{n\to\infty} X^+_n(w)$ for every $w \in W$, and with the following properties for $n = 1..\infty$:[195]

$0 \le X^+_n(w) \le X^+_{n+1}(w) \le X^+(w)$ for every $w \in W$.

There is a natural number m(n), a sequence $\{\xi_{nj}\}_{j=1..m(n)}$ of real numbers ξ_{nj}, and a sequence $\{A_{nj}\}_{j=1..m(n)}$ of mutually disjoint elements A_{nj} of A_W with $\bigcup_{j=1..m(n)} A_{nj} = W$, and $X^+_n(w) = \xi_{nj}$ for every $w \in A_{nj}$.

For a given sequence of maps $X^+_n:W\to\mathbb{R}$ with these properties, define a sequence $\{g^+_n\}_{n=1..\infty}$ of maps $g^+_n:V\to\mathbb{R}$ by the equation

$$g^+_n(v) := \mathsf{E}_{w\sim\gamma(v)}\, X^+_n(w) = \textstyle\sum_{j=1..m(n)} \xi_{nj}\, \gamma_v(A_{nj}) \tag{6.68}$$

for $n = 1..\infty$ and every $v \in V$. The second equality shows that every map $g^+_n:V\to\mathbb{R}$ is a linear combination (with coefficients ξ_{nj}) of the maps $pr_{A(nj)} \circ \gamma$, whose property of being expectational in Π_0 has already been established. So the maps g^+_n are also expectational in Π_0 (Lemma 4.17).

Now we transfer the role of the maps $g_n:V\to\mathbb{R}$ in Convergence-Theorem 4.19 to the maps g^+_n, the validity of Inequality (4.16) being obvious for the maps g^+_n. With the notations S^{*+} and g^+ for the set S* and the map $g:V\to\mathbb{R}$ of the theorem, the first equality in the equation

[194] More explicitly, define maps $X^+:W\to[0, +\infty]$ and $X^-:W\to[0, +\infty]$ by

$X^+(w) := X(w)$ and $X^-(w) := 0$ for $X(w) \ge 0$,

as well as

$X^+(w) := 0$ and $X^-(w) := -X(w)$ for $X(w) < 0$.

[195] See e.g. Bauer (1992, p. 70, Proposition 11.6), and note the underlying definition of a set E* on his p. 66.

$$g^+(v) = \lim_{n\to\infty} \mathsf{E}_{w\sim\gamma(v)}\, X^+{}_n(w) = \mathsf{E}_{w\sim\gamma(v)}\, X^+(w) \tag{6.69}$$

for $v \in S^{*+}$ is obtained from Equation (6.68) and an adaptation of Equation (4.17) to the present situation, and the second equality follows from the definition of expectations by the Lebesgue integral.

Applying the same procedure to the map $X^-:W\to\mathbb{R}$, we obtain a sequence $\{g^-{}_n\}_{n=1..\infty}$ of maps $g^-{}_n:V\to\mathbb{R}$, a subset S^{*-} of the set V, and a map $g^-:V\to\mathbb{R}$ with

$$g^-(v) = \mathsf{E}_{w\sim\gamma(v)}\, X^-(w) \tag{6.70}$$

for $v \in S^{*-}$.

Another reference to the definition of expectations by the Lebesgue integral yields the following equivalence for every $v \in V$: A finite expectation $\mathsf{E}_{w\sim\gamma(v)}\, X(w)$ exists, if and only if the expectation $\mathsf{E}_{w\sim\gamma(v)}\, |X(w)|$ is finite; i.e., if and only if the expectations $\mathsf{E}_{w\sim\gamma(v)}\, X^+(w)$ and $\mathsf{E}_{w\sim\gamma(v)}\, X^-(w)$ are finite:

$$\mathsf{E}_{w\sim\gamma(v)}\, |X(w)| < +\infty \;\Leftrightarrow\; (\mathsf{E}_{w\sim\gamma(v)}\, X^+(w) < +\infty \;\wedge\; \mathsf{E}_{w\sim\gamma(v)}\, X^-(w) < +\infty). \tag{6.71}$$

In other words, $v \in S^*$ iff $v \in S^{*+}$ and $v \in S^{*-}$. (For this reformulation, see Premissa (v) of the RSO-Corollary, and combine the definition of the set S^* in Convergence-Theorem 4.19 with Equation (6.68).) In set theoretical notation, this result can be written as

$$S^* = S^{*+} \cap S^{*-}, \tag{6.72}$$

and then the equality

$$g(v) = g^+(v) - g^-(v) \tag{6.73}$$

for $v \in S^*$ is obtained from Premissa (viii) in combination with Equations (6.67), (6.69) and (6.70).

We are now sufficiently prepared to prove Assertion (ix) of the RSO-Corollary. Equation (6.72) can be combined with SSA-Axiom (iii) to obtain $(S^{*+}, \pi) \in H$ and (S^{*-}, π) for every element π of the set $\Pi^\sim$ specified by Premissa (vi) of the RSO-Corollary. Hence, this set $\Pi^\sim$ can take the role of the equally named set of Convergence-Theorem 4.19 in an application of that theorem to the maps g^+ and g^- of the present situation. So let π be an arbitrary element of the set $\Pi^\sim$, and consider the following chain of equivalences, where E stands for $\mathsf{E}_{U\sim\pi}$:

$$\Phi_\pi \in S^* \Leftrightarrow (\Phi_\pi \in S^{*+} \wedge \Phi_\pi \in S^{*-}) \Leftrightarrow (\mathsf{E}g^+(\phi_U) < +\infty \wedge \mathsf{E}g^+(\phi_U) < +\infty) \Leftrightarrow \mathsf{E}\, |g(\phi_U)| < +\infty. \tag{6.74}$$

The first equivalence in this chain is obtained from Equation (6.72), and the second one from an adaptation of Equivalence (4.18) to the sets S^{*+} and S^{*-} and to the maps g^+ and g^- of the present situation. To prepare a proof of the last equivalence in Formula (6.74), observe that an application of Convergence-Theorem 4.19 to the sets S^{*+} and S^{*-} yields $\phi^{-1}(S^{*+}) \in \mathsf{A}_0$ as well as $\phi^{-1}(S^{*-}) \in \mathsf{A}_0$. Then Equation (6.72) implies that the set $\phi^{-1}(S^*)$ is also contained in the σ-algebra A_0, and this result can be combined with Premissa (vi) and Lemma 4.2 to obtain $\pi(\phi^{-1}(S^*)) = 1$ for every $\pi \in \Pi^{\sim}$. Furthermore, the validity of Equation (6.73) for $v \in S^*$ can be combined with the property $g(v) = 0$ for $v \in V \backslash S^*$ (Premissa (viii)) to obtain the equation

$$g(v) = \chi_{S^*}(v) \cdot (g^+(v) - g^-(v)) \tag{6.75}$$

for every $v \in V$, and this implies

$$g(\phi_\omega) = \chi_{S^*}(\phi_\omega) \cdot (g^+(\phi_\omega) - g^-(\phi_\omega)) \tag{6.76}$$

for every $\omega \in \Omega_0$. Now the A_0-B-measurability of the maps $g^+ \circ \phi$, $g^- \circ \phi$ and $\chi_{S^*} \circ \phi$ is obtained from the expectational properties of the maps g^+ and g^- and from Lemma 6.5.(iv). So Equation (6.76) yields the A_0-B-measurability of the map $g \circ \phi$.[196] Furthermore, since the property $\pi(\phi^{-1}(S^*)) = 1$ implies the almost sure equality $\chi_{S^*}(\phi_U) =_{\pi\text{-a.s.}} 1$ for a random variable U with distribution π, Equation (6.76) yields

$$g(\phi_U) =_{\pi\text{-a.s.}} g^+(\phi_U) - g^-(\phi_U). \tag{6.77}$$

But then the last equivalence in Formula (6.74) follows immediately, and the formula as a whole establishes Assertion (ix) of the RSO-Corollary.
(x): The measurablity of the map $g \circ \phi$ has already been established. Furthermore, Premissa (vii) can be combined with Equation (6.72) to obtain $\Phi_\pi \in S^{*+}$ as well as $\Phi_\pi \in S^{*-}$ for every $\pi \in \Pi^*$. So the maps g^+ and g^- are expectational in Π^*. Then the same property of the map $g^+ - g^-$ is obtained from Lemma 4.17, and it follows for the map g from Equation (6.77).

After the RSO-Corollary has been proved under Premissas (ii') and (iii'), it is left to show that its Premissas (i), (ii) and (iii) imply the validity of Premissas (ii') and (iii'). So let D be the set system consisting of those elements A of the σ-algebra A_W where the map $pr_A \circ \gamma \circ \phi$ is A_0-B-measurable. We will first show that the set system D is a so-called Dynkin system in W, and then it will be easy to derive Premissas (ii') and (iii'). The property of being a Dynkin system in W is defined by the following properties:[197]

a: The set W is an element of D.
b: If A is an element of D, then the set W\A is also an element of D
c: For every sequence $\{A_i\}_{i=1..\infty}$ of mutually disjoint elements of D, the set $\bigcup_{i=1..\infty} A_i$ is an element of D.

To verify these properties, recall that γ is a map $V \to V'$, where V' is a set of probability measures on A_W. For notational convenience, we will write $\gamma_{h(\omega)}$ for the probability measure on A_W, which is assigned to an element ω of the set Ω_0 by the map $\gamma \circ \phi$, and $\gamma_{h(\omega)}(A)$ for the respective probability assigned to an element A of A_W. In other words, $\gamma_{h(\omega)}(A)$ is a result of the map $pr_A \circ \gamma \circ \phi$. With this notation, the property $W \in \mathsf{D}$ follows from the equlity $\gamma_{h(\omega)}(\omega) = 1$, which implies that the map $pr_W \circ \gamma \circ \phi$ is a constant map and hence A_0-B-measurable. Now let A be an

[196] See e.g. Proposition 9.4 in Bauer (1992, p.59) for this conclusion.

[197] See e.g. Definition 2.1 in Bauer (1992, p. 7) for the definition of Dynkin systems.

arbitrary element of D, and define $A^* := W \backslash A$. Then the equality $\gamma_{h(\omega)}(A^*) = \gamma_{h(\omega)}(W) - \gamma_{h(\omega)}(A)$ can be rewritten as $pr_{A^*} \circ \gamma \circ \phi = (pr_W \circ \gamma \circ \phi) - (pr_A \circ \gamma \circ \phi)$. So the map $pr_{A^*} \circ \gamma \circ \phi$ is the difference of two A_0-B-measurable maps, and then $W \backslash A \in \mathsf{D}$ follows immediately. For countable unions, let $\{A_i\}_{i=1..\infty}$ be a sequence of mutually disjoint elements of D, and define $A := \bigcup_{i=1..\infty} A_i$. Then the equality $\gamma_{h(\omega)}(A) = \sum_{i=1..\infty} \gamma_{h(\omega)}(A_i)$ can be rewritten as $pr_A \circ \gamma \circ \phi = \sum_{i=1..\infty} pr_{A(i)} \circ \gamma \circ \phi$, and the A_0-B-measurability of the map $pr_A \circ \gamma \circ \phi$ is obtained from the assumed membership of the sets A_i in the set system D.[198] So the set A is also contained in D.

In summary, it has been proved that the set system D is a Dynkin system in W. Furthermore, Premissa (i) of the RSO-Corollary can be written as $\mathsf{S} \subseteq \mathsf{D}$. Finally, Premissas (ii) and (iii) imply that the σ-algebra A_W is the smallest Dynkin system in W which includes the set sytem S.[199] So we obtain $\mathsf{A}_W \subseteq \mathsf{D}$, and the inclusion $\mathsf{D} \subseteq \mathsf{A}_W$ (an immediate consequence of the definition of D) leads to $\mathsf{D} = \mathsf{A}_W$. But this equality is equivalent with Premissa (ii'). (See the definition of the set system D.)

Premissa (iii') is equivalent with the following claim: If maps $v':\mathsf{A}_W \rightarrow \mathbb{R}$ and $v'':\mathsf{A}_W \rightarrow \mathbb{R}$ are probability measures on A_W with $v'(A) = v''(A)$ for every $A \in \mathsf{S}$, then the maps v' and v'' are identical. So let v' and v'' be maps with the the assumed properties, and let a system S^* of subsets of W be given by the definition $\mathsf{S}^* := \mathsf{S} \cup \{W\}$. Then Premissas (ii) and (iii) imply that the σ-algebra A_W is also generated by the set system S^* and that this set system is closed under intersection of two elements. (Let A' and A'' be elements of S^*. If one of them is the set W, then the other one is identical with the intersection $A' \cap A''$. Otherwise, A' and A'' are elements of S, and then the property $(A' \cap A'') \in \mathsf{S}^*$ is obtained from Premissa (iii).) Furthemore, the property of v' and v'' as probability measures on A_W implies the equality $v'(W) = v''(W) = 1$. So everything is prepared to derive the equality $v' = v''$from elementary results of measure theory.[200]□

It goes almost without saying that the proof of RSO-Corollary 4.21 has been given in two parts to support its application in situations, where Premissas (ii) and (iii) may be uncertain and where Premissas (ii') and (iii') may be verified otherwise.

6.31 Proof of Lemma 4.23 and Corollary 4.16

For a common framework for Lemma 4.23 and Corollary 4.16, let $g:V \rightarrow \mathbb{R}$ be a map, which is almost expectational in Π^*, π an element of Π^*, and S a subset of the set V with the properties required by Definition 4.13.(ii). Furthermore, let a probability measure π' on B be defined by the

---

[198] See e.g. Corollary 11.5 in Bauer (1992, p. 69) for this conclusion.

[199] See e.g. Proposition 2.4 in Bauer (1992, p. 8) for this conclusion.

[200] See e.g. Proposition 5.4 in Bauer (1992, p. 26), and note that the equality $W = \bigcup_{i=1..\infty} A_i$ and the property $v'(A_i) = v''(A_i) = 1 < +\infty$ for $i = 1..\infty$ hold for a sequence $\{A_i\}_{i=1..\infty}$ of elements of S^* with $A_i := W$ for $i = 1..\infty$.

equation $\pi'(A) := \pi(\phi^{-1}(S \cap (g^{-1}(A)))$ for every $A \in \mathsf{B}$.[201] Now consider the following equation, where Φ' is the aggregation rule of the Interval-SSA with $V' = \mathbb{R}$ (Lemma 3.7):

$$g(\Phi(\pi)) = \mathsf{E}_{U\sim\pi}\,(\chi_S(\phi_U) \cdot g(\phi_U)) = \mathsf{E}_{Y\sim\pi'}\,Y = \Phi'(\pi'). \tag{6.78}$$

The first equality in this equation is equivalent with the assumed Property 4.13.(ii.d) of the set S. Now let U be a random variable with distribution π. With the definition of a real valued random variable Y by $Y := \chi_S(\phi_U) \cdot g(\phi_U)$, the above specification of π' implies that π' is the distribution of Y. So the second equality in Equation (6.78) is also established, and the last equality follows from the aggregation rule Φ' in Lemma 3.7.

To apply this result to the situation of Lemma 4.23, observe that the map g of the lemma is assumed to be expectational in Π^*. So the set V can take the role of the set S in the derivation of Equation (6.78), and then the probability measure π' is identical with the probability measure $f(\pi)$ specified by Equation (4.21). So Equation (6.78) establishes the claim of the lemma.

Under the assumptions of Corollary 4.16, let S' be an interval of real numbers, and π an element of Π with $(g^{-1}(S'), \pi) \in H$. It suffices to verify $\Phi(\pi) \in g^{-1}(S')$ for this situation. Now observe that the premissas of the corollary imply $\pi \in \Pi^*$ and that the map g of the corollary is assumed to be almost expectational in Π^*. So let a subset S of V with the properties required by Definition 4.13.(ii) be given. Without loss of generality, we may assume $S \subseteq g^{-1}(S')$. (Otherwise, Lemma 6.5.(x) allows a transition to the set $S \cap g^{-1}(S')$.) Combining this assumption with the assumed Property 4.13.(ii.a) of the set S and with the above definition of a probability measure π', we obtain $\pi'(S') = 1$. So $\Phi'(\pi') \in S'$ follows from Lemma 3.7, and Equation (6.78) yields $g(\Phi(\pi)) \in S'$, i.e., $\Phi(\pi) \in g^{-1}(S')$.□

6.32 Proof of Lemma 4.24

Let a measurable space (Ω_0, A_0) and a map $\phi:\Omega_0 \to V$ be given such that the considered SSA is based on them. Furthermore, let π be an element of the set Π^*, and S' a subset of V fulfilling the requirements of Definition 4.13.(ii) for the given π and for the considered expectational maps $g_i:V \to \mathbb{R}$. (See Lemma 6.5.(xii) for the existence of a suitable set S'.) Finally, define maps $f':V \to \mathbb{R}$ and $g'_i:V \to \mathbb{R}$ for i = 1..n by $g'_i(v) := \chi_{S'}(v) \cdot g_i(v)$ and $f'(v) := \sum_{i=1..n} h_i(g'_v(v))$. We will first verify the inequality

$$f(\Phi_\pi) = \sum h_i(g_i(\Phi_\pi)) = \sum h_i(\mathsf{E}g'_i(\phi_U)) \le \sum \mathsf{E}h_i(g'_i(\phi_U)) = \mathsf{E}\sum h_i(g'_i(\phi_U)) = \mathsf{E}f'(\phi_U), \tag{6.79}$$

where $\sum$ stands for $\sum_{i=1..n}$, and E for $\mathsf{E}_{U\sim\pi}$. The outer equalities are obtained from the definitions of the maps f and f', and the second one from the definition of the maps g'_i and from the assumption that the maps g_i are almost expectational in Π^*. (Recall that π is an element of Π^*.) For the following inequality, note that Jensen's Inequality (see Bauer, 1991, p. 22, Proposition 3.9, and p. 25, Exercise 7) yields $h_i(\mathsf{E}g'_i(\phi_U)) \le \mathsf{E}h_i(g'_i(\phi_U)$. Finally, the last but one equality is an application of well known properties of expectations of sums.

For the case $S = f^{-1}(]-\infty, \xi[)$, let π be an arbitrary element of Π with $(S, \pi) \in H$, and A an

---

[201] See Equation (6.56) and Section ?.? for the claim that π' is a probability measure on B.

element of $\mathcal{A}_0$ with $\pi(A) = 1$ and $\phi(A) \subseteq S$ (see SSA-Axiom (vii)). Then $f(\phi_U) <_{\pi\text{-a.s.}} \xi$ follows for a random variable U with distribution π. Furthermore, the considered π must be an element of Π^*, since this set is assumed to contain all elements of Π with $(S, \pi) \in H$. Now the property $(S', \pi) \in H$ of the above introduced set S' implies $\chi_{S'}(\phi_U) =_{\pi\text{-a.s.}} 1$, and the definition of the map f' yields $f'(\phi_U) =_{\pi\text{-a.s.}} f(\phi_U)$. Combining the results, we obtain $f'(\phi_U) <_{\pi\text{-a.s.}} \xi$ and $\mathsf{E}_{U\sim\pi}\, f'(\phi_U) < \xi$. Finally, a combination of this inequality and Inequality (6.79) leads to $f(\Phi_\pi) < \xi$, i.e., $\Phi_\pi \in S$. But if Φ_π is an element of S for every π with $(S, \pi) \in H$, then S is contained in T by SSA-Axiom (iv).

For $S = f^{-1}(]-\infty, \xi])$, the proof of $S \in T$ is parallell: $(S, \pi) \in H$ leads to $f'(\phi_U) \leq_{\text{a.s.}} \xi$, and the resulting inequality $\mathsf{E}_{U\sim\pi}\, f'(\phi_U) \leq \xi$ can again be combined with Inequality (6.79) to obtain $f(\Phi_\pi) \leq \xi$ and $\Phi_\pi \in S$.□

6.33 Proof of Theorem 4.25

The surjectivity of the map g follows immediately from the definitions of the set V' and of the map by Equation (4.23).

(i)⇔(ii): Let $\mathcal{A}'$ be the coursest σ-algebra in V' where all projection maps $pr_q{:}V'\to\mathbb{R}$ are $\mathcal{A}_{V'}$-$\mathcal{B}$-measurable, and note that Equation (4.23) can be rewritten as $pr_q(g(v)) = g_q(v)$, i.e., $pr_q \circ g = g_q$. So Assertion (i) is equivalent with the $\mathcal{A}_0$-$\mathcal{B}$-measurability of all maps $pr_q \circ g \circ \phi$, and this property is equivalent with the $\mathcal{A}_0$-$\mathcal{A}'$-measurability of the map $g \circ \phi$. (See, e.g., Proposition 7.4 in Bauer, 1992, p. 42) for this conclusion.) With this result, the equivalence (i)⇔(ii) is easily verified. If Assertion (i) halds, then $\mathcal{A}'$ fulfills the requirements of Assertion (ii) upon a σ-algebra $\mathcal{A}_{V'}$. Conversely, if $\mathcal{A}_{V'}$ is a σ-algebra with these properties, then $\mathcal{A}' \subseteq \mathcal{A}_{V'}$ follows from the definition of $\mathcal{A}'$. So the $\mathcal{A}_0$-$\mathcal{A}_{V'}$-measurability of the map $g \circ \phi$ implies its $\mathcal{A}_0$-$\mathcal{A}'$-measurability, and then Assertion (i) is obtained from the above properties of $\mathcal{A}'$.

To prepare later considerations about weakened premissas, the rest of the proof is given under the assumption that the maps g_q are almost expectational in Π^*, whereas the theorem assumes that they are expectational in Π^*. (Lemma 4.14.(ii) allows this approach.)

Before further claims of the theorem are proved, it should be made certain that the map $f{:}\Pi^*\to\Pi'$ is well defined, which means that the probability measure $f(\pi)$ on $\mathcal{A}_{V'}$ given by Equation (4.24) is an element of the set Π'. So let π be an arbitrary element of the set Π^*, q an element of Q, and S a subset of V such that the properties required by Definition 4.13.(ii) hold for the map g_q and the given π. Then the finiteness of the expectation $\mathsf{E}_{U'\sim f(\pi)}\, U'(q)$ is obtained from the equation

$$\mathsf{E}_{U'\sim f(\pi)}\, pr_q(U') = \mathsf{E}_{U\sim\pi}\,(\chi_S(\phi_U)\cdot g_q(\phi_U)) = g_q(\Phi_\pi). \tag{6.80}$$

To verify this equation, observe first that the second equality is an application of Equation (4.12) to the present situation. (Recall that the properties of Definition 4.13.(ii) including Property 4.13.(ii.d) are assumed for the set S and the map g_q.) In particular, the second expectation in Equation (6.80) is finite, since $g_q(\Phi_\pi)$ is finite. For the first equality in this equation, let U be a random variable with distribution π. Then the definition of the map f by Equation (4.24) implies that $f(\pi)$ is the distribution of the random variable $g(\phi_U)$. So the first expectation in Equation (6.80) is the expectation of the random variable $pr_q(g(\phi_U))$. Furthermore, the assumptions about the set S imply $\chi_S(\phi_U) =_{\pi\text{-a.s.}} 1$ (see Lemma 6.5.(iii)). Combining this property with the previously verified equality $pr_q \circ g = g_q$, we obtain $\chi_S(\phi_U)\cdot g_q(\phi_U) =_{\pi\text{-a.s.}} pr_q(g(\phi_U))$. But then the first equality in Equation (6.80) is established, since the expectations of almost surely identical random variables are identical, if one of them has a finite expectation. But if the first expectation is finite for every $q \in Q$, then $f(\pi) \in \Pi'$ follows from the definition of the set Π'.

. The existence and uniqueness of an identity-based, projectional SSA $(V', \Pi', \Phi', H', T')$ is easily verified: For given components V' and Π', the SSA must be based on the measurable space $(V', \mathsf{A}_{V'})$, since Π' is a set of probability measures on $\mathsf{A}_{V'}$ (see SSA-Axiom (v)), and it must be based on the identity map in V' to be identity-based. Then the aggregation rule Φ' is determined by the requirement that all projection maps $pr_q:V' \to \mathbb{R}$ must be expectational in a projectional SSA, which leads to $\Phi'_{\pi'}(q) = \mathsf{E}_{U' \sim \pi'} U'(q)$ for every $\pi' \in \Pi'$ and $q \in Q$. Finally, the components H' and T' of the required SSA are determined by SSA-Axioms (vii) and (iv).

The existence and uniqueness of an identity-based, projectional SSA $(V', \Pi^{\sim}, \Phi^{\sim}, H^{\sim}, T^{\sim})$ is proved in the same way.

(iii): Since the SSA $(V', \Pi', \Phi', H', T')$ is identity-based and projectional, its $\Pi^{\sim}$-restriction has the same properties. But we have already seen that there is only one unique identity-based projectional SSA whose first two components are V' and $\Pi^{\sim}$.

(iv): Since the considered SSAs are identity-based and projectional, the first expectation in Equation (6.80) must be identical with $\Phi'_{f(\pi)}$ and with $\Phi^{\sim}_{f(\pi)}$.

(v): To derive this property from Mapping-Theorem 3.8.(iv), note that Equation (4.25) yields Equation (3.36). Furthermore, Implication (3.35) and its reversal can be derived from Corollary 4.10.(iii) and (iv) for $S' \in \mathsf{A}_{V'}$.

(vi): If the map $g:V \to V'$ is injective, then it is also bijective, since its surjectivity has already been established. Furthermore, the premissas (vi.a) and (vi.b) are reformulations of the properties specified by Corollary 4.10.(i) and (ii). So Implication (3.35) and its reversal follow from Corollary 4.10.(v). Finally, since Equation (3.36) is obtained as above, Equivalences (3.41) and (3.42) are obtained from Mapping-Theorem 3.8.(iv)□

Note that the premissa $S' \in \mathsf{A}_{V'}$ in Assertion (v) has been used only for an application of Corollary 4.10.(iv). Similarly, Premissas (vi.a) and (vi.b) are used only for an application of Corollary 4.10.(v). Hence, these premissas can be weakened in the way pointed out after the proof of Lemma 6.4 in Section 6.24.

The equivalance (i)⇔(ii) implies that a σ-algebra with the properties specified in Assertion (ii) doesn't exist, if only one of the maps $g_q:V \to \mathbb{R}$ is only almost expectational and not expectational. (Lemma 4.14.(iii) tells that the map $g_q \circ \phi$ wouldn't be expectational in this case.) If only almost expectational maps g_q are available, it could be considered to construct a almost projectional, identity based SSA with vocabulary set V'. Then the first expectation in Equation (6.80) would have to be replaced by $\mathsf{E}_{U' \sim f(\pi)} (\chi_{S^*}(U') \cdot pr_q(U'))$, where S^* is a subset of V' such that properties of the set S in Definition 4.13.(ii) hold for this set S^* with respect to the map $pr_q:V' \to \mathbb{R}$ and the probability measure $f(\pi)$. After this modification, the first equality in Equation (6.80) could be based upon the almost sure equality $\chi_S(\phi_U) \cdot g_q(\phi_U) =_{\pi\text{-a.s.}} \chi_{S^*}(g(\phi_U)) \cdot pr_q(g(\phi_U))$.

Another approach to a situation with almost expectational maps $g_q:V \to \mathbb{R}$ becomes possible under the assumption that there exists a family $\{S_q\}_{q \in Q}$ of subsets of V such that the maps $(\chi_{S(q)} \cdot g_q) \circ \phi$ are A_0-B-measurable and the properties $\phi^{-1}(S_q) \in \mathsf{A}_0$, $(S_q, \pi) \in H$ and $\Phi_\pi \in S_q$ hold for every $q \in Q$ and every $\pi \in \Pi$. Then Lemma 4.14.(iv) tells that the maps $\chi_{S(q)} \cdot g_q$ are expectational, and Theorem 4.25 can be applied to these maps.

6.34 Some Issues of Vincentising

In this subsection, some properties resulting from the definition of a vocabulary set V for aggregation by Vincentising are summarised for reference in various sections referring to this kind of aggregation. So let V be the set of those non-decreasing maps $v:\mathbb{R}\to[0, 1]$ where a p^{th} order quantile - i.e., a unique real number ξ with $v(\xi) = p$ - exists for every $p \in]0, 1[$. The subsequent lemma follows immediately from this definition:

Lemma 6.7: A map $v:\mathbb{R}\to\mathbb{R}$ is an element of the just described set V iff it has the following properties:

(i) $]0, 1[\subseteq v(\mathbb{R}) \subseteq [0, 1]$.

(ii) If ξ' and ξ'' are real numbers with $\xi' < \xi''$, then $v(\xi') \leq v(\xi'')$. Furthermore, if $0 < v(\xi') < 1$, then $v(\xi') < v(\xi'')$. In other words, the map v is strictly increasing with the exception of potential constancy at $v(\xi) = 0$ or $v(\xi) = 1$.

Now define $Q :=]0, 1[$, and let V' be the set of all continuous, strictly increasing maps $v':Q\to\mathbb{R}$. Furthermore, let a map $g:V\to V'$ be defined such that $g(v)$ is the (unique) map $v':Q\to\mathbb{R}$ where $v'(q)$ is the q^{th} order quantile of v. It follows immediately from Lemma 6.7. that this map v' is strictly increasing and continuous, i.e., an element of V'. Conversely, if v' is an element of V', then an element v of V with $g(v) = v'$ must have the following properties, if it exists:

- If $\xi = v'(q)$ for some $q \in Q$, then $v(\xi) = q$.
- If $\xi < v'(q)$ for every $q \in Q$, then $v(\xi) = 0$.
- If $\xi > v'(q)$ for every $q \in Q$, then $v(\xi) = 1$.

The first property follows from the definition of the map $g:V\to V'$. Given this property, the remaining ones follow from the fact that all elements of V are non-decreasing maps $\mathbb{R}\to[0, 1]$. But these three requirements specify a unique map $v:\mathbb{R}\to[0, 1]$. (For the first requirement, note that the equation $\xi = v'(q)$ can hold for at most one $q \in Q$, since all elements of V' are strictly increasing maps $Q\to\mathbb{R}$.) Furthermore, the properties specified by Lemma 6.7 follow immediately from the above specification and from the definition of the set V'. So v is an element of V. But if a unique element v of V with $g(v) = v'$ exists for every element v' of V', then the map g is bijective.

Now consider the stochastic SSA for Vincentising specified after Theorem 4.25 in Section 4.6. We will show that an aggregation rule $\Phi:\Pi\to V$ is well defined by the requirement that Equation (4.26) must hold for every $\pi \in \Pi$ and every $q \in Q$. In other words, we have to verify for every $\pi \in \Pi$ the existence and the uniqueness of an element Φ_π of the set V, where Equation (4.26) holds for every $q \in Q$. To derive this property, define a map $v'_\pi:Q\to\mathbb{R}$ for every $\pi \in \Pi$ by

$$v'_\pi(q) := \mathsf{E}_{U\sim\pi}\, g_q(U) \tag{6.81}$$

for every $q \in Q$. It suffices to prove that every such map v'_π is an element of V': Then the definition $\Phi_\pi := g^{-1}(v'_\pi)$ will specify an element Φ_π of V with the required property, and the uniqueness of this element will follow from the bijectivity of the map g.

So let π be an arbitrary element of Π and U a V-valued random variable with distribution π. To show that the map v'_π is strictly increasing, let q' and q'' be elements of Q with $q' < q''$. Then $g_{q'}(v) < g_{q''}(v)$ is obtained for every $v \in V$ from Lemma 6.7.(ii) and the definition of the maps g_q, and the inequality $g_{q'}(U) <_{a.s.} g_{q''}(U)$ follows immediately. But this implies $\mathsf{E}g_{q'}(U) < \mathsf{E}g_{q''}(U)$, and this is equivalent with $v'_\pi(q') < v'_\pi(q'')$.

For the proof that the map v'_π is continuous, let an element q of Q be given, and we will show that v'_π is continuous in q. For this purpose, it suffices to prove that the equation

$$\lim_{n\to\infty} v'_\pi(q(n)) = v'_\pi(q) \tag{6.82}$$

holds for every non-decreasing or non-increasing sequence $\{q(n)\}_{n=1..\infty}$ of elements of Q with $\lim_{n\to\infty} q(n) = q$. So assume first that $\{q(n)\}_{n=1..\infty}$ is a non-decreasing sequence with this property. Then the properties $g_{q(n)}(v) \le g_{q(n+1)}(v)$ and $g_q(v) = \sup_{n=1..\infty} g_{q(n)}(v)$ follow for every $v \in V$ from Lemma 6.7 and the definition of the maps g_q. Now let a sequence $\{f_n\}_{n=1..\infty}$ of maps $f_n{:}V\to\mathbb{R}$ be given by $f_n(v) := g_{q(n)}(v) - g_{q(1)}(v)$ for every $v \in V$. Then the above results can be rewritten as $0 \le f_n(v) \le f_{n+1}(v)$ and $\sup_{n=1..\infty} f_n(v) = g_q(v) - g_{q(1)}(v)$, and these properties are used in the following derivation of Equation (6.82):

$$\begin{aligned}
\lim_{n\to\infty} v'_\pi(q(n)) &= \sup_{n=1..\infty} v'_\pi(q(n)) \\
&= \sup_{n=1..\infty} \mathsf{E}g_{q(n)}(U) \\
&= \sup_{n=1..\infty} \mathsf{E}(g_{q(1)}(U) + f_n(U)) \\
&= \mathsf{E}g_{q(1)}(U) + \sup_{n=1..\infty} \mathsf{E}f_n(U) \\
&= \mathsf{E}g_{q(1)}(U) + \mathsf{E}\sup_{n=1..\infty} f_n(U) \\
&= \mathsf{E}g_{q(1)}(U) + \mathsf{E}(g_q(U) - g_{q(1)}(U)) \\
&= \mathsf{E}g_q(U) \\
&= v'(q).
\end{aligned} \tag{6.82'}$$

For a non-decreasing sequence $\{q(n)\}$, the first equality follows from the monotonicity of the map v'_π, which has already been established. The remaining equalities are based on the definitions of the maps v'_π and f_n and on basic properties of expectations, including an application of the monotone convergence theorem for the transition to the fifth line.

For a non-increasing sequence $\{q(n)\}_{n=1..\infty}$, the definition of a sequence $\{f_n\}_{n=1..\infty}$ of maps $f_n{:}V\to\mathbb{R}$ is changed to $f_n(v) := g_{q(1)}(v) - g_{q(n)}(v)$. Then a suitable adaptation of the arguments underlying Equation (6.82') leads to

$$\begin{aligned}
\lim_{n\to\infty} v'_\pi(q(n)) &= \inf_{n=1..\infty} v'_\pi(q(n)) \\
&= \inf_{n=1..\infty} \mathsf{E}g_{q(n)}(U) \\
&= \inf_{n=1..\infty} \mathsf{E}(g_{q(1)}(U) - f_n(U)) \\
&= \mathsf{E}g_{q(1)}(U) - \sup_{n=1..\infty} \mathsf{E}f_n(U) \\
&= \mathsf{E}g_{q(1)}(U) - \mathsf{E}\sup_{n=1..\infty} f_n(U) \\
&= \mathsf{E}g_{q(1)}(U) - \mathsf{E}(g_{q(1)}(U) - g_q(U)) \\
&= \mathsf{E}g_q(U) \\
&= v'(q).
\end{aligned} \tag{6.82''}$$

Content

To be continued

Send email to iseler@zedat.fu-berlin.de if you want to be informed about later versions, and communicate the 'date and time of present version' on page 1.