Explikative und deduktive Komponenten der "Herleitung" statistischer Vorhersagen aus psychologischen Hypothesen

Albrecht Iseler, Freie Universität Berlin

1. Zwei Verständnisse von Herleitung

"Herleitung" in einem Schritt	
Psychologische Hypothese (PH)	
Unter Bedingung *a* ist ein interessierendes Merkmal tendenziell schwächer ausgeprägt als unter Bedingung *b*.	
"$\Downarrow$"	**"Herleitung"**
Statistische Hypothese (SH)	
In der Popultaion (bzw. bei Randomisierung einer Stichprobe) ist unter Bedingung a ... *SH 1* ... der Durchschnitt (bzw. der Erwartungswert) der AV kleiner ... *SH 2* ... der Median der AV kleiner ... *SH 3* ... die Ausprägung der AV tendenziell kleiner i.S. des U-Test-Kriterium (s.u.) ... *SH 4* ... die kumulative Wahrscheinlichkeit jedes möglichen Werts der AV mindestens so groß und bei mindestens einem Wert streng größer ... *SH 5* ... die mittlere Lösungswahrscheinlichkeit aller Items kleiner ... *SH 6* ... die Lösungswahrscheinlichkeit bei jedem einzelnen Item höchstens so groß und bei mindestens einem Item streng kleiner ... *SH 7* ... die Lösungswahrscheinlichkeit jedes einzelnen Items kleiner ... ... als unter Bedingung b.	

Meehl: 'Loose derivation'.

Herleitung in zwei Schritten

Vorstatistische psychologische Hypothese (VSPH)

Unter Bedingung *a* ist ein interessierendes Merkmal *tendenziell schwächer ausgeprägt* als unter Bedingung *b*.

$\Downarrow$ **Explikation**

Statistisch explizierte psychologische Hypothese (SEPH)

Bei jeder ein Individuum charakterisierenden Wahrscheinlichkeitsverteilung ...	In einer Aggregat-Verteilung ...

ist unter Bedingung a ...
SEPH 1 ... der Erwartungswert (Populationsmittel bzw. true score) der AV kleiner ...
SEPH 2 ... der Median der AV kleiner ...
SEPH 3 ... die Ausprägung der AV tendenziell kleiner i.S. des U-Test-Kriterium (s.u.) ...
SEPH 4 ... die kumulative Wahrscheinlichkeit jedes möglichen Werts der AV mindestens so groß und bei mindestens einem Wert streng größer ...
SEPH 5 ... die mittlere Lösungswahrscheinlichkeit aller Items kleiner ...
SEPH 6 ... die Lösungswahrscheinlichkeit bei jedem einzelnen Item höchstens so groß und bei mindestens einem Item streng kleiner ...
SEPH 7 ... die Lösungswahrscheinlichkeit jedes einzelnen Items kleiner ...
... als unter Bedingung b.

$\Downarrow$ **Deduktion**

Prüfbare statistische Aggregathypothese (SAH)

Bei jeder Individuenauswahl ...	Bei repräsentativer Individuenauswahl ...

... und randomisierter Bedingungszuweisung ist unter Bedingung a ...
SAH 1 ... der Erwartungswert der AV kleiner ...
SAH 2 ... *(der Median der AV kleiner ...)**
SAH 3 ... *(die Ausprägung der AV tendenziell kleiner i.S. des U-Test-Kriterium (s.u.) ...)**
SAH 4 ... die kumulative Wahrscheinlichkeit jedes möglichen Werts der AV mindestens so groß und bei mindestens einem Wert streng größer ...
SAH 5 ... die mittlere Lösungswahrscheinlichkeit aller Items kleiner ...
SAH 6 ... die Lösungswahrscheinlichkeit bei jedem einzelnen Item höchstens so groß und bei mindestens einem Item streng kleiner ...
SAH 7 ... die Lösungswahrscheinlichkeit jedes einzelnen Items kleiner ...
... als unter Bedingung b.

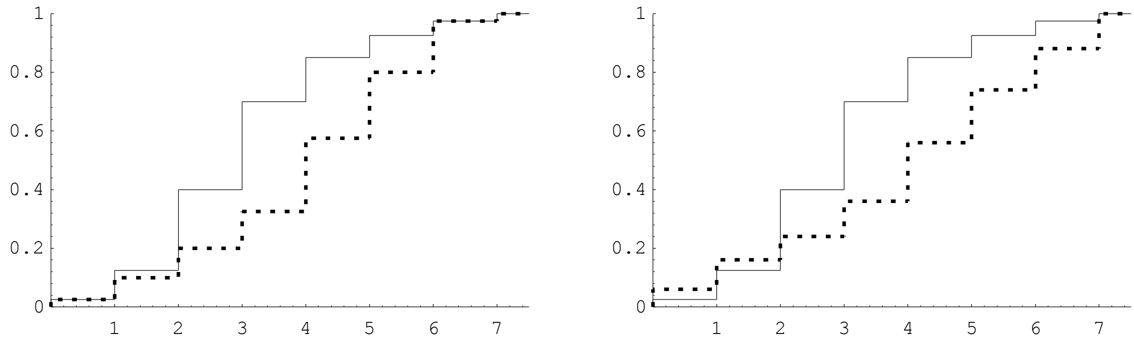
** SAH 2 und SAH 3 folgen deduktiv nur aus Aggregat-bezogener SEPH und bei repräsentativer Auswahl. "Median-Paradox" bei Individuen-bezogener SEPH 2!*

2. Deduktive Reduktion der Beliebigkeit von Explikationen

2.1 Das extensionale Kriterium trennscharfer hypothetischer Fälle

2.1.1: Individuen- oder Aggregat-bezogen?
Frage an den Vertreter der Hypothese: Wenn es Einzelpersonen oder Untergruppen ("Minderheiten") gibt, bei denen das Merkmal unter Bedingung *a* tendenziell ausgeprägter ist als unter Bedingung *b* und wenn diese statistisch "in der Mehrheit untergehen", wäre dann Ihre Hypothese voll und ganz als erfüllt zu betrachten?

2.1.2: Verteilungsbezogene Kriterien, z.B. Erwartungswert-Unterschied (*SEPH 1)* oder "strenge stochastische Ordnung" *(SEPH 4)*?
Gespräch mit dem Vertreter der Hypothese: Sehen Sie sich die folgenden Kurven an. (Interpretation je nach Antwort zu 2.1.1 als kumulative Populationsverteilung oder als kumulative intraindividuelle Wahrscheinlichkeitsverteilung.)



Zwei Verteilungspaare ohne und mit Ogiven-Überkreuzung.
In beiden Fällen ist
$\mu_a = 3$ (durchgezogene Ogive)
$\mu_b = 4$ (gepunktete Ogive)

Wenn nun in einer Population (bzw. bei einer Person) die in der rechten Abbildung dargestellte Überkreuzung vorläge, wäre dann Ihre Hypothese voll und ganz als erfüllt zu betrachten? Oder erst bei Mittelwerts-Unterschied ohne solche Überkreuzung?

2.2 Axiomatisierung der Relation "tendenziell schwächer ausgeprägt"

W sei eine endliche Menge von Merkmalsausprägungen in einem empirischen Relativ (z.B. mögliche "Lösungsmuster") oder einem numerischen (mögliche Meßwerte, auch mehrdimensional)
Π sei die Menge der Wahrscheinlichkeitsmaße auf $\wp W$
$\preceq$ sei eine zweistellige Relation auf Π. Für zwei Wahrscheinlichkeitsmaße P_1 und P_2 auf $\wp W$ bedeutet $P_1 \preceq P_2$: Bei P_1 ist die Tendenz zur Ausprägung des interessierenden Merkmals höchstens so groß wie bei P_2.

Die auf 'unit' (Person oder Aggregat*) *u* bezogene SEPH lautet

$$P_{ua} \prec P_{ub}, \text{ d.h.: } P_{ua} \preceq P_{ub} \wedge \neg(P_{ub} \preceq P_{ua}).$$

*Bakan (1955): An aggregate is not 'the general', but rather 'a particular'.

Erwägenswerte Postulate ("Axiome") für die Relation $\preceq$.

1. Die Relation $\preceq$ ist eine Quasiordnung (also reflexiv und transitiv).

Verletzung der Transitivität beim U-Test-Kriterium!

Benötigte Grundlage: Ordnungsrelation $\preceq^*$ auf W.
Ist U die Zahl der "hypothesenkonformen Vergleiche", dann ist $E(U) = N_1 \cdot N_2 \cdot V_{12}$
mit $V_{12} := \sum_{w \in W} P_2(w) \cdot (0.5 \cdot P_1(w) + \sum_{w' \prec^* w} P_1(w'))$

Kriterium: $P_1 \preceq P_2 \;:\Leftrightarrow\; V_{12} \geq 0.5$.
Beispiel (mit $w_i \prec^* w_{i+1}$ für i = 1..5):

	w_1	w_2	w_3	w_4	w_5	w_6
P_a	0.20	0.00	0.00	0.80	0.00	0.00
P_b	0.00	0.50	0.00	0.00	0.50	0.00
P_c	0.00	0.00	0.80	0.00	0.00	0.20

Vergleich nach U-Test-Kriterium:

V_{ab} =	0.60	$P_a \prec P_b$
V_{bc} =	0.60	$P_b \prec P_c$
V_{ac} =	0.36	$P_c \prec P_a$

2. Die Relation $\preceq$ ist "konvex kürzbar".

d.h.: Für alle Wahrscheinlichkeitsmaße P_1, P_2 und P_3 auf $\wp W$ und jedes $\alpha \in]0, 1[$ gilt
$\alpha \cdot P_1 + (1-\alpha) \cdot P_3 \;\preceq\; \alpha \cdot P_2 + (1-\alpha) \cdot P_3 \;\Leftrightarrow\; P_1 \;\preceq\; P_2$.

Beispiel (mit α = 0.7)

Zu vergleichende WVen	Lösungsmuster			
	− −	+ −	− +	+ +
P_1	0.20	0.30	0.40	0.10
P_2	0.20	0.40	0.30	0.10
P_3	0.10	0.40	0.30	0.20
$0.7 \cdot P_1 + 0.3 \cdot P_3$	0.17	0.33	0.37	0.13
$0.7 \cdot P_2 + 0.3 \cdot P_3$	0.17	0.40	0.30	0.13

Klar: $P_1 \prec P_3$ und $P_2 \prec P_3$. Aber P_1 und P_2???
Darauf gibt konvexe Kürzbarkeit keine Antwort! Verlangt wird aber, daß zwischen den beiden letzten Zeilen dieselbe Relation besteht wie zwischen P_1 und P_2.

3. Die Relation $\preceq$ ist konnex.

Falls konvex kürzbare schwache Ordnung vorliegt, ergeben sich Differenz-Strukturen in Π und W.

Zu vergleichende WVen	Lösungsmuster			
	$--$	$+-$	$-+$	$++$
$P_1 = \varepsilon_{--}$	1.00	0.00	0.00	0.00
$P_2 = \varepsilon_{+-}$	0.00	1.00	0.00	0.00
$P_3 = \varepsilon_{++}$	0.00	0.00	0.00	1.00
$0.1 \cdot P_1 + 0.9 \cdot P_3$	0.10	0.00	0.00	0.90
$0.2 \cdot P_1 + 0.8 \cdot P_3$	0.20	0.00	0.00	0.80
$0.3 \cdot P_1 + 0.7 \cdot P_3$	0.30	0.00	0.00	0.70
$0.4 \cdot P_1 + 0.6 \cdot P_3$	0.40	0.00	0.00	0.60
$0.5 \cdot P_1 + 0.5 \cdot P_3$	0.50	0.00	0.00	0.50
$0.6 \cdot P_1 + 0.4 \cdot P_3$	0.60	0.00	0.00	0.40
$0.7 \cdot P_1 + 0.3 \cdot P_3$	0.70	0.00	0.00	0.30
$0.8 \cdot P_1 + 0.2 \cdot P_3$	0.80	0.00	0.00	0.20
$0.9 \cdot P_1 + 0.1 \cdot P_3$	0.90	0.00	0.00	0.10

Klar: $P_1 \prec P_2 \prec P_3$

Für konvex kürzbare schwache Ordnung ergibt sich (vgl. Iseler, 1998): Es gibt ein $\delta \in [0, 1]$ mit

$\alpha \cdot P_1 + (1-\alpha) \cdot P_3 \prec P_2$ für $0 \leq \alpha < \delta$

$\alpha \cdot P_1 + (1-\alpha) \cdot P_3 \succ P_2$ für $\delta < \alpha \leq \delta$

Unklar: Was ist bei $\alpha = \delta$? (Dazu ist ein zusätzliches 'archimedisches' Axiom erforderlich, s.u.)

Grundlage einer Differenzstruktur in Π: $d(P_1, P_2) / d(P_1, P_3) := 1 - \delta$
"Differenzquotienten-Struktur" in W:
$d^*(--, +-) / d^*(--, ++) := d(\varepsilon_{--}, \varepsilon_{+-}) / d(\varepsilon_{--}, \varepsilon_{++}) = 1 - \delta$.

4. Die Relation $\preceq$ ist "vor-archimedisch".

d.h.: Für alle Wahrscheinlichkeitsmaße P_1, P_2 und P_3 auf $\wp W$ ist die Menge

$\{\alpha \in [0, 1]: P_1 \preceq \alpha P_2 + (1-\alpha) P_3\}$

topologisch abgeschlossen.

Konsequenz für obiges Beispiel (s. Iseler, 1998):
Ist die Relation $\preceq$ eine konvex kürzbare, vorarchimedische schwache Ordnung, dann ist $0 < \delta < 1$, und

$\alpha \cdot P_1 + (1-\alpha) \cdot P_3 \sim P_2$ für $\alpha = \delta$

3. Deduktive Herleitung von Vorhersagen mit "Kennziffern"

Zu *SAH 1, 2, 3 und 7* gibt es unmittelbar auf diese Hypothesen bezogene Prüfverfahren.

Zu *SAH 4*
- Zwei einseitige Kolmogoroff-Smirnov-Tests
- U-Test-Kriterium (als *SAH* hier möglich, auch wenn als *SEPH* fragwürdig).
- Äquivalent mit *SAH 4* ist die Vorhersage: Für jede streng monotone numerische Repräsentation $\phi: W \to \mathbb{R}$ ist $E(\phi(X))$ unter Bedingung a kleiner.
- Bei Individuen-bezogener *SEPH 4* gilt dies sogar, wenn individuell verschiedene Repräsentationen $\phi_i: W \to \mathbb{R}$ verwendet werden, solange deren Wahl stochastisch unabhängig vom Treatment erfolgt.

Aus *SAH 5* folgt Erwartungswert-Unterschied bei streng positiver Gleichgewichtung aller Items.

Äquivalent mit *SAH 6* ist die Vorhersage: *SAH5* gilt bei jeder Item-Gewichtung, bei der jedes Item ein (streng) positives Gewicht erhält. Bei Individuen-bezogener *SEPH 6* gilt dies sogar, wenn individuell verschiedene Item-Gewichtungen verwendet werden, solange deren Wahl statistisch unabhängig vom Treatment erfolgt.

Aus *SAH 7* folgt *SAH 6*.

Schlußbemerkung: Strenge, Wohlwollen und Fairness; kontextuelle Bedeutungshaltigkeit

Strenge, Wohlwollen und Fairness der Hypothesenprüfung sind nicht absolut, sondern relativ zur *SEPH* zu sehen.
- Bei Individuen-bezogener *SEPH* ist "Rasterfahndung nach Gesetzesbrechern" angemessen streng, bei Aggregat-bezogener *SEPH* dagegen ein Verstoß gegen die wohlwollende Prüfung.
- Die Erwartungswert-bezogene *SEPH 1* ist nur bei Differenzstruktur bedeutungshaltig ('meaningful') und ist dann auch mit Meßwerten auf Intervallskalen-Niveau zu prüfen. Bei *SEPH 4* ist es dagegen angemessen streng, Mittelwertsunterschiede aufgrund derjenigen Ordinalskala zu prüfen, bei der (z.B. aufgrund empirischer Vorinformationen) noch am ehesten Falsifikationen zu erwarten sind. Bei Individuen-bezogener *SEPH 4* sogar mit individueller Skalierung
- Zur Überprüfung von *SEPH 6* ist es angemessen streng, diejenige Item-Gewichtung zu wählen, bei der noch am ehesten Falsifikationen zu erwarten sind, bei Individuen-bezogener *SEPH 6* sogar mit individueller Item-Gewichtung.

Luce & Krumhansl (1988): 'It is not the statistic per se that is or is not meaningful, but rather the proposition in which it appears' (p. 19).
Weitergeführt: 'It is not the proposition containig a statistic per se that is or is not meaningful, but rather the deductive context in which it appears.'

Literatur

Bakan, D. (1955). The general and the aggregate: A methodological distinction. *Perceptual and Motor Skills*, *5*, 211-212.

Iseler, A. (1998). Über richtungsbasierte Relationen in reellen Vektorräumen. In K. Ch. Klauer & H. Westmeyer (Hrsg.), *Psychologische Methoden und soziale Prozesse - Festschrift für Hubert Feger* (S. 80-121). Lengerich: Pabst. s.a.
userpage.fu-berlin.de/~iseler/papers/dirbas_mat.htm