methevalreport
3 (1) 2001

Christof Nachtigall & Rolf Steyer

Unkonfundiertheit: Äquivalente Bedingungen und Perspektiven für die Testung

metheval
psychologische methodenlehre
und evaluationsforschung jena

Impressum

methevalreport
erscheint seit 1999
in unregelmäßigen Abständen
als „graue“ Schriftenreihe des Lehrstuhls für
Psychologische Methodenlehre und Evaluationsforschung
am Institut für Psychologie der Friedrich-Schiller-Universität Jena

Herausgeber:
Prof. Dr. Rolf Steyer
Sekr.: +49 (3641) 945 230
Durchwahl: +49 (3641) 945 231
Fax: +49 (3641) 945 232

rolf.steyer@uni-jena.de

Redaktion:
Dipl.Psych. Friedrich Funke
sff@uni-jena.de

Typographie:
cand.psych. Silke Zachariae
zachariae@web.de

Standort:
Thüringer Universitäts- und Landesbibliothek
Lesesaal Zweigstelle Psychologie

Internet
http://www.uni-jena.de/svw/metheval/report/

Bestellungen:
Methodenlehre und Evaluationsforschung
Institut für Psychologie
Steiger 3 Haus 1
D-07743 Jena
Deutschland

Unkonfundiertheit:

Äquivalente Bedingungen und

Perspektiven für die Testung

Christof Nachtigall & Rolf Steyer

Zusammenfassung

Die Unkonfundiertheit einer Treatmentregression (Steyer et al., 1996, 2000 b) stellt eine empirisch überprüfbare Bedingung dar, welche die kausale Interpretation von Effekten auch außerhalb randomisierter Experimente ermöglicht. In diesem Beitrag werden für den Fall von diskreten Treatment- und Störvariablen eine Reihe von Bedingungen vorgestellt, welche zur Unkonfundiertheit äquivalent sind. Diese Bedingungen bilden die Basis für die Weiterentwicklung von Konfundierungstests.

1 Der Satz über Unkonfundiertheit

Im Folgenden bezeichne X eine Treatmentvariable, $W = f(U)$ eine potenzielle Störvariable und U die Personenprojektion in dem in Steyer et al. (2000 a) beschriebenen Zufallsexperiment. Vereinfacht gesprochen beschreibt dieses Zufallsexperiment die "Ziehung" von Units (U) (im psychologischen Kontext in der Regel Personen), deren mögliche Zuweisung zu verschiedenen Treatments (X) und die Response (Y) auf diese Treatments. Dabei ist Y numerisch, X und U diskret und $P(X=x, U=u) > 0$ für alle x, u. Der folgende Satz nennt eine Reihe von Bedingungen, die zur Unkonfundiertheit einer Treatmentregression (vgl. Steyer et al., 1996; Nachtigall et al., 2000) für diskrete Regressoren äquivalent sind bzw. zur Definition von Unkonfundiertheit herangezogen werden können (vgl. Steyer et al., 2000 b).

Satz: Sei $E(Y \mid X)$ eine Treatmentregression, W eine potenzielle Störvariable mit Werten w. Mit $P_{W=w}$ wird die durch $W=w$ bedingte Verteilung und mit $E_{W=w}(Y \mid X=x)$ wird der unter $W=w$ bedingte Erwartungswert $E(Y \mid X=x,\ W=w)$ bezeichnet. Folgende Bedingungen sind äquivalent:

i) $\forall x, \forall W,$ *" w gilt:* $P_{W=w}(X=x \mid U=u)=P_{W=w}(X=x)\ \forall u$

oder

$E_{W=w}(Y \mid X=x) = E_{W=w}(Y \mid U=u, X=x)\ \forall u$

ii) $\forall x$ *gilt:* $P(X=x \mid U=u)=P(X=x)\ \forall u$

oder

$E(Y \mid X=x) = E(Y \mid U=u, X=x)\ \forall u$

iii) $\forall x, \forall W$ *gilt:* $P(X=x \mid W=w)=P(X=x)\ \forall w$

oder

$E(Y \mid X=x) = E(Y \mid W=w, X=x)\ \forall w$

iv) $\forall x, \forall W, \forall w$ *gilt:* $P(X=x \mid W=w)=P(X=x)$

oder

$E(Y \mid X=x) = E(Y \mid W=w, X=x)$

v) $\forall x, \forall W$ *gilt:* $E(Y \mid X=x) = \sum_{w} E(Y \mid X=x, W=w)P(W=w)$

Bemerkung: In den Bedingungen i) bis iv) kann die Gleichheit von $P(X=x \mid W=w)=\mathrm{P}(X=x)$ auch äquivalent durch $P(W=w \mid X=x)=P(W=w)$ beschrieben werden, da die stochastische Unabhängigkeit von Ereignissen eines Zufallsexperimentes eine symmetrische Relation ist. Im Beweis von iv)⇒iii) wird davon Gebrauch gemacht.

Beweis[1]: i)⇒ ii): Mit $W\equiv 1$ folgt ii)

ii)⇒ i): Beweis von Theorem 5 aus Steyer et al. (2000 b).

[1] Für Teile dieses Beweises siehe Steyer et al., (1996, 2000 b).

ii)⇒iii): Gegeben seien beliebige, aber feste x, W und w. Gilt $P(X=x \mid U=u)=P(X=x)\ \forall u$, dann folgt

$$P(X=x \mid W=w)P(W=w)=P(X=x \cap W=w)$$

$$= \sum_{u\in W^{-1}(w)} P(X=x \cap U=u)P(U=u) / P(U=u)$$

$$= \sum_{u\in W^{-1}(w)} P(X=x|U=u)P(U=u) = P(X=x)\sum_{u\in W^{-1}(w)} P(U=u) = P(X=x)P(W=w).$$

Wegen $P(W=w)>0$ gilt $P(X=x \mid W=w)=P(X=x)$. Gilt $E(Y \mid U=u,\ X=x) = E(Y \mid X=x)\ \forall u$, so folgt auch $E(Y \mid W=w,\ X=x) = E(Y \mid X=x)\ \forall w$, da $W=f(U)$.

iii)⇒ii): U ist eine spezielle potenzielle Störvariable W.

iii)⇒iv): trivial

iv)⇒iii): Indirekter Beweis (analog zum zweiten Teil des Beweises von Theorem 2 aus Steyer et al., 2000 b). Angenommen, es existiert ein x und ein W mit Werten w_1, w_2, für die

$P(W=w_1 \mid X=x) \neq P(W=w_1)$ und $E(Y \mid X=x,\ W=w_2) \neq E(Y \mid X=x)$ ist.

Gilt $w_1 = w_2$ so folgt sofort Widerspruch zu iv). Sei also $w_1 \neq w_2$. Nach Voraussetzung gilt

$P(W=w_2 \mid X=x)=P(W=w_2)$ und $E(Y \mid X=x,\ W=w_1)=E(Y \mid X=x)$.

Betrachte $I_{\{w1,\ w2\}}$ die Indikatorvariable der Menge $\{w_1,w_2\}$. Es ist

$$P(I_{\{w1,\ w2\}}=1 \mid X=x) = P(W=w_1 \mid X=x)+P(W=w_2 \mid X=x) \neq P(W=w_1)+P(W=w_2) = P(I_{\{w1,\ w2\}}=1)$$

Für $I_{\{w1,\ w2\}}$ trifft also die erste der „Oder"-Bedingungen aus iv) nicht zu. Wir zeigen, dass auch

$$E(Y \mid X=x,\ I_{\{w1,\ w2\}}=1) \neq E(Y \mid X=x)$$

und führen damit den Widerspruch herbei. $E(Y \mid X=x,\ I_{\{w1,\ w2\}}=1)$ kann aufgrund der Rechenregeln für bedingte Erwartungswerte $E(Y \mid f(Z))=E(E(Y \mid Z) \mid f(Z))$ mit $Z=(X,\ W)$ und $f(Z) = (X,\ I_{\{w1,\ w2\}})$ (vgl. Steyer & Eid, 2000, Box G1 v.) umgeformt werden zu

$$\begin{aligned} E(Y \mid X=x, I_{\{w1,\ w2\}}=1) \quad &= \quad E(Y \mid X=x,\ W=w_1)P(X=x,\ W=w \mid X=x,\ I_{\{w1,\ w2\}}=1) \\ &\quad + E(Y \mid X=x,\ W=w_2)P(X=x,\ W=w_2 \mid X=x,\ I_{\{w1,\ w2\}}=1) \\ &= \quad E(Y \mid X=x)(1 - P(W=w_2 \mid X=x,\ I_{\{w1,\ w2\}}=1)) \\ &\quad + E(Y \mid X=x,\ W=w_2)P(W=w_2 \mid X=x,\ I_{\{w1,\ w2\}}=1). \end{aligned}$$

Das zweite Gleichheitszeichen folgt durch Einsetzen der Definition für bedingte Wahrscheinlichkeit und elementare Rechnung. Es folgt, dass

$$E(Y \mid X=x) - E(Y \mid X=x,\ I_{\{w1,\ w2\}}=1)$$

$$= (E(Y \mid X=x) - E(Y \mid X=x,\ W=w_2))P(W=w_2 \mid X=x,\ I_{\{w1,\ w2\}}=1) \neq 0,$$

was ein Widerspruch zur Voraussetzung ist.

iv)⇒ v): Es werden diejenigen Werte w zu Teilmengen der Wertemenge von W zusammengefasst, die jeweils einer der vorausgesetzten Bedingungen genügen.

$$A := \{w: P(X=x \mid W=w) = P(X=x)\},\ B := \{w: E(Y \mid X=x) = E(Y \mid W=w, X=x) \wedge w \notin A\}$$

Auf diese Weise erhalten wir disjunkte Mengen, die nach Voraussetzung den gesamten Wertebereich von W umfassen, es gilt also $P(I_A=1)+P(I_B=1)=1$.[2] Zu zeigen ist

$$E(Y \mid X=x) = \sum_w E(Y|W=w, X=x) \cdot P(W=w).$$

Es gilt immer

$$E(Y \mid X=x) = \sum_w E(Y|W=w, X=x) \cdot P(W=w|X=x).$$

Wir betrachten

$$\sum_w E(Y|W=w, X=x) \cdot P(W=w)$$

$$= \sum_{w\in A} E(Y|W=w, X=x) \cdot P(W=w) \quad + \sum_{w\in B} \mathrm{E}(\mathrm{Y}|\mathrm{W}=\mathrm{w},\ \mathrm{X}=\mathrm{x}) \cdot P(W=w)$$

$$= \sum_{w\in A} E(Y|W=w, X=x) \cdot P(W=w|X=x) + \sum_{w\in B} \mathrm{E}(\mathrm{Y}|\mathrm{X}=\mathrm{x}) \cdot P(W=w). \qquad (1)$$

Weiter ist

$$\sum_{w\in B} P(W=w) = P(I_B=1).$$

Nun ist $I_A=1$ stochastisch unabhängig von $X=x$, denn

$$P(I_A=1|X=x) = P(\bigcup_{w\in A}\{W=w\}|X=x) = \sum_{w\in A} P(W=w|X=x) = \sum_{w\in A} P(W=w) = P(I_A=1).$$

Da $B = A^c$, so ist $I_B=1$ stochastisch unabhängig von $X=x$ und es gilt

$$P(I_B=1) = P(I_B=1 \mid X=x) = \sum_{w\in B} P(W=w|X=x).$$

Damit kann (1) umgeformt werden zu

$$\sum_{w\in A} E(Y|W=w, X=x) \cdot P(W=w|X=x) \ + \sum_{w\in B} \mathrm{E}(\mathrm{Y}|\mathrm{X}=\mathrm{x}) \cdot P(W=w|X=x)$$

$$= \sum_w E(Y|W=w, X=x) \cdot P(W=w|X=x)$$

$$= E(Y \mid X=x).$$

[2] Da wir uns auf diskrete Zufallsvariablen beschränken, sind I_A und I_B messbar.

v)$\Rightarrow$ iv) Es seien x, W und w gegeben. Wir betrachten die potenzielle Störvariable $I_{W=w}$, also die Indikatorvariable für $W=w$. Nach Voraussetzung gilt

$$E(Y \mid X=x) = E(Y \mid I_{W=w}=1, X=x)P(I_{W=w}=1) + E(Y \mid I_{W=w)}=0, X=x)P(I_{W=w}=0) \quad (2)$$

und es gilt immer

$$E(Y \mid X=x) = E(Y \mid I_{W=w}=1, X=x)P(I_{W=1}=1 \mid X=x) + E(Y \mid I_{W=w)}=0, X=x)P(I_{W=w}=0 \mid X=x). \quad (3)$$

Zur Abkürzung und Vereinfachung schreiben wir: $p_1 := P(I_{W=w}=1)$, $p_2 := P(I_{W=w}=1 \mid X=x)$, $a := E(Y \mid I_{W=w}=1, X=x)$, $b := E(Y \mid I_{W=w}=0, X=x)$, $c := E(Y \mid X=x)$. Dann lauten die Gleichungen (2) und (3)

$c = ap_1 + b(1-p_1)$

$c = ap_2 + b(1-p_2)$.

Subtrahieren wir die zweite Gleichung von der ersten, so erhalten wir

$0 = a(p_1-p_2) + b(1-p_1-1+p_2)$

$\Leftrightarrow b(p_1-p_2) = a(p_1-p_2)$.

Daraus folgt, dass entweder $p_1-p_2=0$ oder $a=b$ sein muss, und damit $P(W=w \mid X=x) = P(W=w)$ oder $E(Y \mid X=x) = E(Y \mid I_{W=w}=1, X=x) = E(Y \mid W=w, X=x)$.

Damit ist der Beweis vollständig. ÿ

Eine dieser äquivalenten Bedingungen kann nun zur Definition des Begriffes der Unkonfundiertheit verwendet werden (vgl. Steyer et al., 1996, 2000 b).

Bemerkung 1: Aus Bedingung i) folgt, dass die Unkonfundiertheit einer Treatmentregression auch die Unkonfundiertheit in Subpopulationen $W=w$ impliziert.

Bemerkung 2 Die Unkonfundiertheit einer Treatmentregression ist empirisch falsifizierbar. Bedingung iii), iv), und v) des Unkonfundiertheitssatzes enthalten ausschließlich empirisch schätzbare Größen.

Eine Treatmentregression ist z. B. genau dann konfundiert, wenn es eine Treatmentbedingung x und einen Wert einer potenziellen Störvariable $W=w$ gibt, so dass

$$P(X=x \mid W=w) \neq P(X=x) \text{ und } E(Y \mid X=x) \neq E(Y \mid W=w, X=x)$$

erfüllt sind. Dies ergibt sich durch logische Verneinung von Bedingung iii). Auf diese Weise können auf der Basis logischer Verneinung von Bedingung iii), iv), oder v) statistische Tests für die Unkonfundiertheit abgeleitet werden.

Bemerkung 3: Für einzelne potenzielle Störvariable $W = f(U)$ sind die Bedingungen des Unkonfundiertheitssatzes nicht äquivalent.

Der Beweis von Bemerkung 3 ergibt sich aus folgendem fiktivem Beispiel: Betrachten wir ein dichotomes X und eine Population bestehend aus drei Units u_1, u_2 und u_3. Die folgende Tabelle zeigt die Wahrscheinlichkeit für das "Ziehen" der Units u_i, die bedingten Wahrscheinlichkeiten sowie die bedingten Erwartungswerte von Y.

u	$P(U=u)$	$P(X=x_1 \mid U=u)$	$P(U=u \mid X=x_1)$	$E(Y \mid U=u, X=x_1)$	$E(Y \mid U=u, X=x_2)$
u_1	1/3	3/8	1/4	1	4
u_2	1/3	3/4	1/2	0	2
u_3	1/3	3/8	1/4	-1	0

Tabelle 2: *Fiktive Werte einer Treatmentregression sowie der Zuweisungswahrscheinlichkeiten*.

Alle Units haben die "Ziehungswahrscheinlichkeit" 1/3. Mit dem Satz der totalen Wahrscheinlichkeit läßt sich berechnen, dass $P(X = x_1) = P(X = x_2) = 1/2$. Als potenzielle Störvariable W wird nun die Unitvariable U selbst betrachtet. Es zeigt sich: U und X sind stochastisch abhängig, da $P(X = x_1) \neq P(X = x_1 \mid U = u)$ für alle u, Unit u_2 erhält z. B. mit erhöhter Wahrscheinlichkeit Treatment x_1. Auch ist die Identität $E(Y \mid X, U) = E(Y \mid X)$ verletzt, mithin liegt eine Falsifizierung der Unkonfundiertheit aufgrund Bedingung iii) und iv) des Satzes vor. Betrachten wir Bedingung v). Es ist

$$E(Y \mid X = x_1) = \sum_u E(Y \mid U = u,\ X = x_1) \cdot P(U = u \mid X = x_1)$$

$$= 1 \cdot 1/4 + 0 \cdot 1/2 - 1 \cdot 1/4 = 0.$$

Andererseits liefert

$$\sum_u E(Y \mid U = u,\ X = x_1) \cdot P(U = u) = 1 \cdot 1/3 + 0 \cdot 1/3 - 1 \cdot 1/3 = 0.$$

Bedingung v) wird also durch *diese* potenzielle Störvariable nicht verletzt[3].

[3] Aufrund des Unkonfundiertheitssatzes muss es eine andere Störvariable W geben, für die Bedingung v) nicht erfüllt ist. Die Variable

$$W = I_{u_1} = \begin{cases} 1 & \text{falls } U = u_1 \\ 0 & \text{falls } U \neq u_1 \end{cases}$$

leistet das Gewünschte, wie sich durch nachrechnen zeigt.

2 Diskussion

Der Satz über Unkonfundiertheit einer Treatmentregression liefert verschiedene äquivalente Bedingungen für Unkonfundiertheit im Falle diskreter Treatment- und Störvariablen. Einige davon sind empirisch falsifizierbar. Sie bilden die Basis für statistische Tests der Unkonfundiertheit. Allerdings sind diese Bedingungen bei einer konkret vorliegenden potenziellen Störvariable *W* nicht "gleich stark". So kann bei der Testung von Unkonfundiertheit mit einem gegebenen *W* beispielsweise Bedingung iv) bereits falsifiziert sein, Bedingung v) jedoch nicht. Daraus ergeben sich Fragen hinsichtlich Auswahl und Anwendung der auf den verschiedenen Bedingungen iii), iv) und v) beruhenden Konfundierungstests, die es in weiteren Arbeiten zu klären gilt.

3 Literatur

Nachtigall, C., Suhl, U. & Steyer, R.. (2000): Einführung in die Konfundierungsanalyse. *metheval report 2* (*1*). http://www.uni-jena.de/svw/metheval/report/.

Steyer, R. Gabler, S. & Rucai, A.A. (1996). Individual Causal Effects, Average Causal Effects, and Unconfoundedness in Regression Models. In F. Faulbaum and W. Bandilla (Eds.), *SoftStat '95. Advances in Statistical Software 5* (pp. 203-210). Stuttgart: Lucius & Lucius.

Steyer, R. & Eid, M. (2000): *Messen und Testen* (2. Aufl.). Berlin: Springer.

Steyer, R., Gabler, S., von Davier, A.A., Nachtigall, C. & Buhl, T. (2000a). Causal Regression Models I: Individual and Average Causal Effects. *Meth. Psych. Research-Online*, (*5*), 2, 39-71.

Steyer, R., Gabler, S., von Davier, A.A., & Nachtigall, C. (2000b). Causal Regression Models II: Unconfoundedness and Causal Unbiasedness. *Meth. Psych. Research-Online*, (*5*), 3, 55-86.