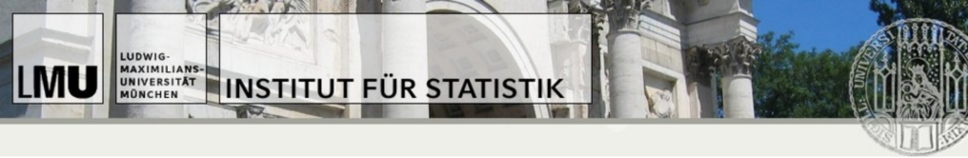


Kausale Inferenz

Kapitel 6: Inverse Probability Weighting

Sommersemester 2017
Version: 24. Juli 2017

Michael Schomaker
University of Cape Town, CIDER, South Africa
Ludwig-Maximilians-Universität München, Institut für Statistik

Horvitz-Thompson-Schätzer

- Sei π_i die Auswahlwahrscheinlichkeit für das i-te Individuum in die Stichprobe aufgenommen zu werden
- Sei N die Anzahl der Beobachtungen in der Grundgesamtheit, n die Größe der Stichprobe
- Bei einer einfachen Zufallsauswahl ist $\pi_i = n/N$
- Sei $y_1, \ldots, y_n$ eine Stichprobe bezüglich der Variable Y

Horvitz-Thompson-Schätzer [1]

Der Horvitz-Thompson Schätzer für das arithmetische Mittel in der Grundgesamtheit ist

$$\hat{\bar{Y}}_{\text{HT}} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

Beispiel

Beispiel:

4 Individuen mit $Y_1 = 10,\ Y_2 = 20,\ Y_3 = 80,\ Y_4 = 90$

Stichprobe vom Umfang $n = 2$

Y_4 habe höhere Auswahlwahrscheinlichkeit:
$\pi_{1,2,3} < \pi_4$ (Urne mit 5 Kugeln: 1,2,3,4,4)

Gezogene Individuen		$\bar{y}$	Wahrscheinlichkeit
1	2	15	$1/10 = 1/5 \cdot 1/4 + 1/5 \cdot 1/4$
1	3	45	$1/10$
1	4	50	$7/30 = 1/5 \cdot 2/4 + 2/5 \cdot 1/3$
2	3	50	$1/10$
2	4	55	$7/30$
3	4	85	$7/30$

Beispiel [Fortsetzung]

Beispiel:

Verteilung des arithmetischen Mittels:

$\bar{y}$	15	45	50	55	85
$P(\bar{y})$	1/10	1/10	1/3	7/30	7/30

Auswahlwahrscheinlichkeiten:

i	1	2	3	4
π_i	13/30	13/30	13/30	21/30

Kommen Individuen 1 und 2 in die Stichprobe:

$$
\begin{aligned}
\bar{y} &= 15 \\
\hat{\bar{Y}}_{\text{HT}} &= \frac{1}{N}\sum_{i=1}^{n}\frac{y_i}{\pi_i} = \frac{1}{4}\left(\frac{10}{13/30} + \frac{20}{13/30}\right) = 17.31
\end{aligned}
$$

Beispiel [Fortsetzung]

Beispiel: Verteilung des Horvitz-Thompson-Schätzers:

$\hat{\hat{Y}}_{HT}$	17.31	37.91	43.68	51.92	57.69	78.30
$P(\hat{\hat{Y}}_{HT})$	1/10	7/30	7/30	1/10	1/10	7/30

Das heißt:

$$E(\hat{\hat{Y}}_{HT}) = \left(17.31 \cdot \frac{1}{10} + \ldots + 78.30 \cdot \frac{7}{30}\right) = 50$$

$$E(\bar{y}) = \left(15 \cdot \frac{1}{10} + \ldots + 85 \cdot \frac{7}{30}\right) = 55.33$$

Erwartungstreue

Sei I_i der Indikator, der angibt, ob das i-te Individuum in die Stichprobe gezogen wird.

$$\hat{\bar{Y}}_{\text{HT}} = \frac{1}{N}\sum_{i=1}^{n}\frac{y_i}{\pi_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i I_i}{\pi_i}$$

I ist eine Bernoulli-Variable mit $E(I_i) = P(I_i = 1) = \pi_i$. Damit können wir schreiben:

$$E(\hat{\bar{Y}}_{\text{HT}}) = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i E(I_i)}{\pi_i} = \bar{Y}$$

Inverse Probability of Treatment Weighting

- Idee: Kreiere Pseudo-Population, bei der Confounding aufgehoben wird

$$
\begin{tikzcd}
L \arrow[r, dashed, "\times" marking] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

- Damit wird aus Assoziation Kausalität

- Verwende hierzu die Gewichte

$$w \;=\; 1/\pi = 1/p = \frac{1}{f(A|L)}$$

bzw. für diskretes A

$$w \;=\; 1/\pi = 1/p = \frac{1}{P(A=a|L)}$$

Der IPTW-Schätzer

IPTW-Schätzer (auf Basis des Horvitz-Thomson-Schätzers)

Unter den Annahmen der Positivität, bedingten Austauschbarkeit und Konsistenz gilt

$$E(Y^a) \quad = \quad E\left(\frac{Y \cdot I(A=a)}{p}\right)$$

Ein Schätzer für den Average Treatment Effect ist damit

$$\hat{\psi}_{\text{ATE}} \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\frac{y_i \cdot I(A_i=1)}{P(A_i=1|L_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{y_i \cdot I(A_i=0)}{P(A_i=0|L_i)}$$

Schätzung

1. Schätze den Mechanismus der Behandlungszuweisung: $P(A = 1|L)$

   (z.B. logistische Regression)

2. Schätze damit p_i:

   - Für diejenigen, die tatsächlich behandelt wurden ($A_i = 1$): verwende einfach $\hat{P}(A = 1|L = l_i)$
   - Für diejenigen, die nicht behandelt wurden ($A_i = 0$): verwende $1 - \hat{P}(A = 1|L = l_i)$

3. Jedes Individuum erhält das Gewicht $\hat{w}_i = 1/\hat{p}_i$

4. Schätze dann z.B.

$$E(Y^1) = \frac{1}{n}\sum_{i=1}^{n}[y_i \times I(A_i = 1) \times w_i]$$

   (oder marginales strukturelles Modell, siehe unten)

Beispiel (aus Kapitel 5)

i	L	A	Y	p_i	w_i
1	0	0	0	?	?
2	0	0	1	?	?
3	0	0	0	?	?
4	0	1	0	?	?
5	0	1	1	?	?
6	0	1	1	?	?
7	0	0	0	?	?
8	1	0	0	?	?
9	1	1	0	?	?
10	1	1	1	?	?
11	1	0	1	?	?
12	1	0	1	?	?

Beispiel

$$
\begin{aligned}
p_{11} = P(A=1|L=1) &= \frac{2}{5}\\
p_{01} = P(A=0|L=1) &= \frac{3}{5}\\
p_{10} = P(A=1|L=0) &= \frac{3}{7}\\
p_{00} = P(A=0|L=0) &= \frac{4}{7}
\end{aligned}
$$

und damit

$$
\begin{aligned}
w_{11} = 1/P(A=1|L=1) &= \frac{5}{2}\\
w_{01} = 1/P(A=0|L=1) &= \frac{5}{3}\\
w_{10} = 1/P(A=1|L=0) &= \frac{7}{3}\\
w_{00} = 1/P(A=0|L=0) &= \frac{7}{4}
\end{aligned}
$$

Beispiel

i	L	A	Y	p_i	w_i
1	0	0	0	$\frac{4}{7}$	$\frac{7}{4}$
2	0	0	1	$\frac{4}{7}$	$\frac{7}{4}$
3	0	0	0	$\frac{4}{7}$	$\frac{7}{4}$
4	0	1	0	$\frac{3}{7}$	$\frac{7}{3}$
5	0	1	1	$\frac{3}{7}$	$\frac{7}{3}$
6	0	1	1	$\frac{3}{7}$	$\frac{7}{3}$
7	0	0	0	$\frac{4}{7}$	$\frac{7}{4}$
8	1	0	0	$\frac{3}{5}$	$\frac{5}{3}$
9	1	1	0	$\frac{2}{5}$	$\frac{5}{2}$
10	1	1	1	$\frac{2}{5}$	$\frac{5}{2}$
11	1	0	1	$\frac{3}{5}$	$\frac{5}{3}$
12	1	0	1	$\frac{3}{5}$	$\frac{5}{3}$

Beispiel

$$
\begin{aligned}
E(Y^1) &= \frac{1}{n}\sum_{i=1}^{n} \frac{y_i \cdot I(A_i = 1)}{P(A_i = 1|L_i)} \\
&= \frac{1}{12}\left(\frac{7}{3}\cdot 0 + \frac{7}{3}\cdot 1 + \frac{7}{3}\cdot 1 + \frac{5}{2}\cdot 0 + \frac{5}{2}\cdot 1\right) \\
&= \frac{1}{12}\left(\frac{28}{6} + \frac{15}{6}\right) = 0.5972
\end{aligned}
$$

$$
\begin{aligned}
E(Y^0) &= \frac{1}{n}\sum_{i=1}^{n} \frac{y_i \cdot I(A_i = 0)}{P(A_i = 0|L_i)} \\
&= \frac{1}{12}\left(\frac{7}{4}\cdot 0 + \frac{7}{4}\cdot 1 + \frac{7}{4}\cdot 0 + \frac{7}{4}\cdot 0 + \frac{5}{3}\cdot 0 + \frac{5}{3}\cdot 1 + \frac{5}{3}\cdot 1\right) \\
&= \frac{1}{12}\left(\frac{7}{4} + \frac{10}{3}\right) = 0.4236
\end{aligned}
$$

$$
\hat{\psi}_{\text{ATE}} = 0.5972 - 0.4236 = 0.1736
$$

$\rightarrow$ korrektes Ergebnis!

Gleiches Ergebnis bei Verwendung logistischer Regression

```
m.w           <- glm(A~L,family=binomial)
weight        <- 1/predict(m.w, type="response")
weight[A==0]  <- 1/(1-predict(m.w, type="response")[A==0])

mean(Y*(A==1)*weight)
mean(Y*(A==0)*weight)
```

```
> weight
   1    2    3    4    5    6    7    8    9   10   11   12
1.75 1.75 1.75 2.33 2.33 2.33 1.75 1.67 2.50 2.50 1.67 1.67
```

```
> mean(Y*(A==1)*weight)
[1] 0.5966667
> mean(Y*(A==0)*weight)
[1] 0.4241667
```

Der IPTW-Schätzer schätzt $E(Y^a)$

Für den diskreten Fall:

$$
\begin{aligned}
E_Y\left(\frac{Y\cdot I(A=a)}{p}\right) &\overset{\text{Kons.}}{=} E_{Y^a}\left(\frac{Y^a\cdot I(A=a)}{p}\right)\\
&\overset{\text{BE}}{=} E_L\left\{E_{Y^a|L}\left(\frac{Y^a\cdot I(A=a)}{P(A=a|L)}\right)|L\right\}\\
&\overset{\text{Aust.}}{=} E_L\left\{\underbrace{E_{Y^a|L}\left(\frac{I(A=a)}{P(A=a|L)}|L\right)}_{=1}\cdot E_{Y^a|L}(Y^a|L)\right\}\\
&= E_L\left\{E_{Y^a|L}(Y^a|L)\right\}\\
&\overset{\text{BE}}{=} E(Y^a)
\end{aligned}
$$

Stabilisierte Gewichte

- Die Ergebnisse eines marginalen strukturellen Modells ändern sich nicht wenn statt der Gewichte

$$w = \frac{1}{P(A = a|L)}$$

die Gewichte

$$w = \frac{c}{P(A = a|L)}, \qquad 0 < c \leq 1$$

verwendet werden [2].

- Stabilisierte Gewichte:

$$w_s = \frac{P(A = a)}{P(A = a|L)}$$

Wozu stabiliserte Gewichte?

- Einzelne nichtstabilisierte Gewichte können sehr hoch sein, stabilisierte Gewichte reduzieren die Spannweite der Gewichte
- Wichtig um Schätzgrenzen einzuhalten, z.B. dass $\hat{P}(Y = 1) \leq 1$
- Reduzierung der Breite der Konfidenzintervalle

In *R*

```
> m.w          <- glm(A~L,family=binomial)
> weight       <- 1/predict(m.w, type="response")
> weight[A==0] <- 1/(1-predict(m.w, type="response")[A==0])
>
> m.w2 <- glm(A~1,family=binomial)
> weight2                <- 1/predict(m.w2, type="response")
> weight2[A==0]    <- 1/(1-predict(m.w2, type="response")[A==0])
>
> sweight <- weight/weight2
>
> summary(weight)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.667   1.729   1.750   2.000   2.333   2.500
> summary(sweight)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.9722  0.9722  0.9965  1.0000  1.0210  1.0420
```

Beispiel [Fortsetzung]

$$
\begin{aligned}
E(Y^1) &\overset{Def.}{=} \frac{1}{n}\sum_{i=1}^{n} y_i \cdot I(A_i = 1)\cdot w_i \\
&\overset{Erw.}{=} \frac{1}{n}\sum_{i=1}^{n} y_i \cdot I(A_i = 1)\cdot \frac{1}{P(A_i = 1|L)}\cdot\frac{P(A_i = 1)}{P(A_i = 1)} \\
&= \frac{1}{n}\sum_{i=1}^{n} y_i \cdot I(A_i = 1)\cdot \underbrace{\frac{P(A_i = 1)}{P(A_i = 1|L)}}_{w_{s,i}}\cdot\frac{1}{P(A_i = 1)} \\
&= \frac{1}{12}\left(\frac{0\cdot\frac{5/12}{3/7}}{5/12}+\frac{1\cdot\frac{5/12}{3/7}}{5/12}+\frac{1\cdot\frac{5/12}{3/7}}{5/12}+\frac{0\cdot\frac{5/12}{2/5}}{5/12}+\frac{1\cdot\frac{5/12}{2/5}}{5/12}\right) \\
&= \frac{1}{12}\left(\frac{7}{3}+\frac{7}{3}+\frac{5}{2}\right) \\
&= 0.5972
\end{aligned}
$$

Aber:

$$
\frac{1}{n}\sum_{i=1}^{n} y_i \cdot I(A_i = 1) w_{s,i} = 0.5972\cdot\frac{5}{12} = 0.249
$$

Beispiel [Fortsetzung]

Das heisst:

$$\psi_{ATE} \quad = \quad \hat{\psi}_{1,w_s} \cdot \frac{1}{P(A=1)} - \hat{\psi}_{0,w_s} \cdot \frac{1}{P(A=0)}$$

Besser:

- verwende Gewichte in marginalem strukturellen Modell
- für den ATE: lineares gewichtetes Modell

Marginales strukturelles Modell

Gegeben sei ein lineares strukturelles Modell

$$E(Y^a) = \beta_0 + \beta_1 A$$

Dann ist

$$\begin{aligned} E(Y^a) &= \beta_0 && \text{für} \quad A = 0 \\ E(Y^a) &= \beta_0 + \beta_1 && \text{für} \quad A = 1 \end{aligned}$$

Das heisst:

$$\psi_{\text{ATE}} = \beta_1$$

In der durch Gewichtung entstandenen Pseudopopulation ist Assoziation = Kausalität und das gewichtete lineare Modell

$$E(Y|A) = \beta_0 + \beta_1 A$$

liefert den ATE über β_1.

Marginales strukturelles Modell

...liefert korrektes Ergebnis ($\hat{\psi} = 0.1736$) für normale und stabilisierte Gewichte

```
> lm(Y~A, weights=weight)

Call:
lm(formula = Y ~ A, weights = weight)

Coefficients:
(Intercept)            A
     0.4236       0.1736

> lm(Y~A, weights=sweight)

Call:
lm(formula = Y ~ A, weights = sweight)

Coefficients:
(Intercept)            A
     0.4236       0.1736
```

Warum nicht den [causal] odds ratio berechnen?

Mit den Werten von oben erhalten wir:

$$\psi_{\mathrm{OR}} = \frac{P(Y^1 = 1)/P(Y^1 = 0)}{P(Y^0 = 1)/P(Y^0 = 0)}$$

$$= \frac{0.59722/(1 - 0.59722)}{0.42361/(1 - 0.42361)} = 2.02$$

Ein gewichtetes logistisches Modell in *R* liefert:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.308      0.765   -0.40     0.69
A              0.702      1.190    0.59     0.56
```

Korrekt, da $exp(0.702) = 2.02$!

Noch etwas Intuition zu MSMs

- Wir haben ein lineares Modell

$$\mathcal{M}(A,\beta) : E(Y|A) \quad = \quad \beta_0 + \beta_1 A$$

- Mögliche Schätzung über kleinste Quadrate:

$$\hat{\beta} = \text{argmin}_\beta \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 a_i)^2$$

- z.B. Schätzung für β_1 über partielle Ableitung und Nullsetzen:

$$\sum_{i=1}^{n} \frac{\partial}{\partial \beta_1} (y_i - \beta_0 - \beta_1 a_i)^2 = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 a_i) a_i \,.$$

also

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 a_i) a_i = 0$$

Noch etwas Intuition zu MSMs

- Etwas allgemeiner:

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\mathcal{M}(a_i,\beta)(y_i - \mathcal{M}(a_i,\beta))$$

- Der IPTW-Schätzer ist die Lösung der Schätzgleichung

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{g(A)}{h(A|L)}\frac{\partial}{\partial\beta}\mathcal{M}(a_i,\beta)(y_i - \mathcal{M}(a_i,\beta))$$

...und deswegen können wir gewichtete Regression verwenden...

Zwischenbilanz

- Gewichtungsprinzip kann auch in kausaler Inferenz verwendet werden

- Wir erzeugen eine Pseudopopulation der es kein Confounding gibt und damit Assoziation Kausalität ist

- Hierzu müssen wir ein Modell für $P(A|L)$ schätzen (und ggf. auch $P(A)$ berechnen)

- Ein gewichtetes lineares Modell liefert oft ψ_{ATE} und ein gewichtetes logistisches Modell ψ_{OR}

- Am einfachsten: gewichtete Regression $\overset{\text{oft}}{=}$ marginales strukturelles Modell

Zusammenfassung der Gewichte

- Einzelne Gewichte sollten nicht zu hoch sein
- Das arithmetische Mittel der stabilisierten Gewichte sollte etwa 1 sein

Beispiel: Krebsdaten

	Min	1st Qu.	Median	Mean	3rd Qu.	Max.
w	1	1.3	1.6	2.0	2.1	137.0
w_s	0.5	0.6	0.8	1.0	1.1	50.2

$\rightarrow$ arithmetisches Mittel von 1 (gut)

$\rightarrow$ aber hohe Gewichte (siehe Positivität)

Was, wenn das arithmetische Mittel nicht 1 ist?

- Progressive Trunkierung wird oft verwendet
- Vorgehen:
  1. Bestimme das α und $1-\alpha$-Quantil (q_α; $q_{1-\alpha}$) der Gewichte
  2. Gewichte, die kleiner als q_α sind werden trunkiert und durch das entsprechende Quantil ersetzt
  3. Gewichte, die größer als $q_{1-\alpha}$ sind werden trunkiert und durch das entsprechende Quantil ersetzt
- $\rightarrow$ Reduzierung der Variabilität auf Kosten von Bias!

Beispiel [3]

Variation der Effektschätzung bezüglich der Trunkierung

Truncation percentiles	Estimated weights for the outcome Death or LTF: Mean (SD)	Estimated weights for the outcome Death or LTF: Minimum/ maximum	Estimates of effect of club on Death or LTF: Estimate	Estimates of effect of club on Death or LTF: Standard Error
0,100	1.06 (0.54)	0.01/11.12	0.43	0.16
1,99	1.09 (0.42)	0.05/3.02	0.44	0.14
5,95	1.01 (0.07)	0.92/1.22	0.44	0.13
10,90	1.01 (0.03)	0.97/1.07	0.45	0.13
25,75	1.00 (0.01)	0.99/1.01	0.46	0.13

Konsequenz

- Konsistenz des IPTW-Schätzers hängt von der Konsistenz der Schätzung von $P(A|L)$ ab
- Trunkierung führt dazu, dass der Schätzer nicht mehr konsistent ist
- Trunkierung kann gut sein, muss es aber nicht
- Der perfekte "Bias-Variance-Tradeoff" ist schwer zu finden
- Im Zweifel besser eine alternative Methode verwenden (z.B. Standardisierung)

Simulation von Petersen et al. [4]

	Bias	Var	MSE
G-COMP			
None	0.007	0.009	0.009
[0.025,0.975]	0.007	0.009	0.009
[0.05,0.95]	0.007	0.009	0.009
[0.1,0.9]	0.007	0.009	0.009
IPTW			
None	0.544	0.693	0.989
[0.025,0.975]	1.080	0.090	1.257
[0.05,0.95]	1.437	0.059	2.123
[0.1,0.9]	1.935	0.043	3.787

Beispiel

Beispiel: Krebsdaten

Einige Personen haben sehr hohe Gewichte (> 30)

```
> Odat1[sweight>30,]
       ID    w1               w2        w3       w4 a y
1152 1152   old         2_middle     other stage 4 0 1
1586 1586   old 3_lower middle     other stage 4 0 1
1716 1716 young         2_middle     other stage 4 0 1
4314 4314 young 3_lower middle alzheimer stage 4 0 1
```

Dies scheinen Leute zu sein, die nicht behandelt wurdem ($a = 0$), trotz fortgeschrittener Erkrankung (stage 4 & Begleiterkrankungen)

Gibt es Positivitätsprobleme?

Antwort: ja

```
, ,  = other,  = stage 4

  1_upper middle 2_middle 3_lower middle 4_working 5_non-working
0              0        2              1         0             0
1              3       25             88        34            22
```

z.B.: 1(!) Person repräsentiert die komplette Schicht “a=0, $W_2 = 4$, $W_3 =$ other, $W_4 = 4$” in den Daten und bekommt ein hohes Gewicht

z.B: Keiner wurde nicht behandelt in der Schicht “$W_2 = 4$, $W_3 =$ other, $W_4 = 5$”

Mehr zu Positivität

- Große Gewichte können auf Positivitätsprobleme zurückzuführen sein
- Gibt es keine großen Gewichte, heißt das nicht automatisch, dass es keine Positivitätsprobleme gibt
- Der IPTW-Schätzer ist sehr sensibel gegenüber Positivitätsverletzung
- Das ist ein Argument dafür andere Schätzer zu verwenden
- Wird ein marginales strukturelles Modell über gewichtete Regression geschätzt, so kann das Modell in “leere Regionen” extrapolieren...
- ...dies funktioniert nur, wenn das marginale Modell korrekt spezifiziert ist (andere Methoden können generell besser extrapolieren)

Was tun?

- Trunkierung; eher als Sensitivitätsanalyse um zu sehen, ob die Effektschätzung stabil ist, anstatt als Lösung
- Analyse von Gewichten:
  - um zu sehen, ob über verbesserte Modellspezifizierung (für $P(A|L)$) sich das aritmetische Mittel verbessert
  - und hohe Gewichte reduziert werden können
- Veränderung der Fragestellung, z.B. Restriktion auf eine Untermenge der Daten
- Verwendung eines besseren Schätzers(!)

Konfidenzintervalle

1. Verwende robusten Varianzschätzer über GEE's (z.B. `geeglm() in` *R*)
   - Gewichtung macht aus einer Beobachtung mehrere Beobachtungen, deswegen ist normaler OLS-Schätzer nicht passend
   - fast immer verwendet
   - Problem: Dieser Standardfehler ist generell konservativ, Konfidenzintervalle sind zu groß [5]
   - Gewichte sind nicht fest, sondern werden geschätzt
2. Bootstrapping
   - aufwändiger, aber häufig besser
   - Modellselektion (oder allgemeiner: machine learning) während des Bootstraps nicht optimal

Beispiel

Beispiel: Krebsdaten

```
msm_cancer <- geeglm(y~a, weights=sweight, data=Odat, id=c(seq(1:5000)))
```

Schätzer	S.E.
Naive	0.0125
Robust	0.0339
Bootstrapping	0.0310

Fehlende Daten

- Hängt die Fehlwahrscheinlichkeit von beobachtbaren (bzw. unbeobachtbaren) Variablen ab, so ist in der Regel[1] die Schätzung auf Basis der vollständigen Fälle verzerrt

- Idee: interveniere auf die fehlenden Daten und schätze

$$Y^{A=a,C=0}$$

  wobei $C_i = 0$ für einen vollständigen Fall steht

- *C* kann allgemein für fehlende Werte stehen, für fehlende Werte von Y, oder für zensierte Daten bei Longitudinaldaten

- wir betrachten den allgemeinen Fall bei Querschnittsdaten

[1] Beispiel für eine Ausnahme: Regressionsparameter, wenn Fehlwahrscheinlichkeit nicht von der Zielgröße abhängt

Fehlende Daten bei IPTW

- Erzeuge eine Pseudopopulation, mit derselben Größe wie die ursprüngliche Population ohne fehlende Daten, und in der es kein Confounding gibt

- Die Wahrscheinlicheit für eine vollständige Beobachtung ist

$$P(C = 0|A, L, Y)$$

  und kann über logistische Regression oder Stratifizierung geschätzt werden

- Multipliziere diese inverse (stabilisierte) Wahrscheinlichkeit an das (stabilisierte) Gewicht bezüglich der Behandlungszuweisung

Der IPTW-Schätzer bei fehlenden Daten

IPCTW-Schätzer

Unter den Annahmen der Positivität, bedingten Austauschbarkeit und Konsistenz gilt

$$E(Y^{a,C=0}) = E\left(\frac{Y \cdot I(A=a, C=0)}{P(A=a|L) \times P(C=0|A,L,Y)}\right)$$

Ein Schätzer für den Average Treatment Effect ist damit

$$\hat{\psi}_{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i \cdot I(A_i=1, C_i=0)}{P(A_i=1|L_i) \cdot P(C_i=0|A_i,L_i,Y_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{y_i \cdot I(A_i=0, C_i=0)}{P(A_i=0|L_i) \cdot P(C_i=0|A_i,L_i,Y_i)}$$

Der IPTW-Schätzer bei fehlenden Daten

IPCTW-Schätzer (II)

Unter Verwendung von

$$w_i^{aC} = \Bigg\{ \underbrace{P(A_i = a|L_i)}_{w_i^a} \cdot \underbrace{P(C_i = 0|A_i, L_i, Y_i)}_{w_i^C} \Bigg\}^{-1}$$

lässt sich $E(Y^{a,C=0})$ über

$$\hat{\psi}_{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n} w_i^{1C} \cdot y_i \cdot I(A_i = 1, C_i = 0) - \frac{1}{n}\sum_{i=1}^{n} w_i^{0C} \cdot y_i \cdot I(A_i = 0, C_i = 0)$$

schätzen.

Beispiel

i	L	A	Y	C	p_i^a	w_i^a	p_i^C	w_i^C	w_i^{aC}
1	0	0	0	1	$\frac{4}{7}$	$\frac{7}{4}$	?	?	?
2	0	0	1	0	$\frac{4}{7}$	$\frac{7}{4}$	?	?	?
3	0	0	0	0	$\frac{4}{7}$	$\frac{7}{4}$	?	?	?
4	0	1	0	0	$\frac{3}{7}$	$\frac{7}{3}$	?	?	?
5	0	1	1	0	$\frac{3}{7}$	$\frac{7}{3}$	?	?	?
6	0	1	1	0	$\frac{3}{7}$	$\frac{7}{3}$	?	?	?
7	0	0	0	0	$\frac{4}{7}$	$\frac{7}{4}$	?	?	?
8	1	0	0	0	$\frac{3}{5}$	$\frac{5}{3}$	?	?	?
9	1	1	0	0	$\frac{2}{5}$	$\frac{5}{2}$	?	?	?
10	1	1	1	0	$\frac{2}{5}$	$\frac{5}{2}$	?	?	?
11	1	0	1	0	$\frac{3}{5}$	$\frac{5}{3}$	?	?	?
12	1	0	1	1	$\frac{3}{5}$	$\frac{5}{3}$	?	?	?

Beispiel

$$
\begin{aligned}
P(C=1|L=1,A=1,Y=1) &= 0\\
P(C=1|L=1,A=0,Y=1) &= 1/2\\
P(C=1|L=0,A=1,Y=1) &= 0\\
P(C=1|L=0,A=0,Y=1) &= 0
\end{aligned}
$$

$$
\begin{aligned}
P(C=1|L=1,A=1,Y=0) &= 0\\
P(C=1|L=1,A=0,Y=0) &= 0\\
P(C=1|L=0,A=1,Y=0) &= 0\\
P(C=1|L=0,A=0,Y=0) &= 1/3
\end{aligned}
$$

$$
\begin{aligned}
P(C=0|L=1,A=1,Y=1) &= 1\\
P(C=0|L=1,A=0,Y=1) &= 1/2\\
P(C=0|L=0,A=1,Y=1) &= 1\\
P(C=0|L=0,A=0,Y=1) &= 1
\end{aligned}
$$

$$
\begin{aligned}
P(C=0|L=1,A=1,Y=0) &= 1\\
P(C=0|L=1,A=0,Y=0) &= 1\\
P(C=0|L=0,A=1,Y=0) &= 1\\
P(C=0|L=0,A=0,Y=0) &= 2/3
\end{aligned}
$$

Beispiel

i	L	A	Y	C	p_i^a	w_i^a	p_i^C	w_i^C	w_i^{aC}
1	0	0	0	1	$\frac{4}{7}$	$\frac{7}{4}$	0	0	0
2	0	0	1	0	$\frac{4}{7}$	$\frac{7}{4}$	1	1	$\frac{7}{4}$
3	0	0	0	0	$\frac{4}{7}$	$\frac{7}{4}$	$\frac{2}{3}$	$\frac{3}{2}$	$\frac{21}{8}$
4	0	1	0	0	$\frac{3}{7}$	$\frac{7}{3}$	1	1	$\frac{7}{3}$
5	0	1	1	0	$\frac{3}{7}$	$\frac{7}{3}$	1	1	$\frac{7}{3}$
6	0	1	1	0	$\frac{3}{7}$	$\frac{7}{3}$	1	1	$\frac{7}{3}$
7	0	0	0	0	$\frac{4}{7}$	$\frac{7}{4}$	$\frac{2}{3}$	$\frac{3}{2}$	$\frac{21}{8}$
8	1	0	0	0	$\frac{3}{5}$	$\frac{5}{3}$	1	1	$\frac{5}{3}$
9	1	1	0	0	$\frac{2}{5}$	$\frac{5}{2}$	1	1	$\frac{5}{2}$
10	1	1	1	0	$\frac{2}{5}$	$\frac{5}{2}$	1	1	$\frac{5}{2}$
11	1	0	1	0	$\frac{3}{5}$	$\frac{5}{3}$	$\frac{1}{2}$	1	$\frac{10}{3}$
12	1	0	1	1	$\frac{3}{5}$	$\frac{5}{3}$	0	0	0

Beispiel

$$E(Y^{1,C=0}) = \frac{1}{n}\sum_{i=1}^{n} w_i^{1C} \cdot y_i \cdot I(A_i = 1, C_i = 0)$$
$$= \frac{1}{12}\left(0 + \frac{7}{3} + \frac{7}{3} + 0 + \frac{5}{2}\right) = 0.5972$$

$$E(Y^{0,C=0}) = \frac{1}{n}\sum_{i=1}^{n} w_i^{0C} \cdot y_i \cdot I(A_i = 1, C_i = 0)$$
$$= \frac{1}{12}\left(0 + 0 + \frac{7}{4} + 0 + 0 + 0 + \frac{10}{3}\right) = 0.424$$

D.h. $\hat{\psi}_{\text{ATE}} \approx 0.173 \rightarrow$ korrekt

In *R*

```
m.w            <- glm(A~L,family=binomial)
weight         <- 1/predict(m.w, type="response")
weight[A==0] <- 1/(1-predict(m.w, type="response")[A==0])

m.w2 <- glm(A~1,family=binomial)
weight2                <- 1/predict(m.w2, type="response")
weight2[A==0]    <- 1/(1-predict(m.w2, type="response")[A==0])

m.c <- glm(C~A*L*Y, family="binomial")
weight3                <- 0
weight3[C==0]          <- 1/(1-predict(m.c, type="response")[C==0])
weight3[is.na(weight3)] <- 0

weight4 <- weight*weight3

# ATE
mean(Y*(A==1 & C==0)*weight4)-
mean(Y*(A==0 & C==0)*weight4)

# alternativ als MSM
lm(Y~A, weights=weight4, data=as.data.frame(cbind(Y,A)))
```

Augmented IPTW

AIPTW-Schätzer [6]

Unter den Annahmen der Positivität, bedingten Austauschbarkeit und Konsistenz ist

$$\hat{\psi}_{\text{ATE, AIPTW}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{I(A_i=1)}{P(A_i=1|L_i)} - \frac{I(A_i=0)}{P(A_i=0|L_i)} \times \left(Y_i - \hat{Y}_i\right)\right\} + \hat{\psi}_{\text{ATE, g-formula}}$$

ein doppelt robuster Schätzer für den Average Treatment Effect ψ_{ATE}.

Wozu?

- Der AIPTW-Schätzer ist doppelt robust (siehe z.B. [7]):
  1. Sind sowohl das Modell für die Zielgröße[2] und die Behandlunszuweisung[3] korrekt spezifiziert ist der Schätzer konsistent und asymptotisch effizient
  2. Ist eines der beiden Modelle mis-spezifiziert, ist der Schätzer immer noch konsistent, jedoch nicht mehr asymptotisch effizient
  3. sind beide Modelle mis-spezifiziert, so ist der Schätzer nicht mehr konsistent
- Der Schätzer ist dennoch weiterhin volatil bei Positivitätsverletzungen

---

[2] verwendet für $\hat{Y}_i$ und $\hat{\psi}_{\text{ATE}}$, g-formula

[3] also: $P(A|L)$

- Wenn das Regressionsmodell für Y korrekt spezifiziert ist, liefert die g-formula korrekte Ergebnisse

- In diesem Fall “reduziert sich der AIPTW-Schätzer zum g-formula Schätzer, der in einer endlichen Stichprobe nur von diesem zufällig auf Basis der geschätzten Residuen abweicht”

- Wenn das Regressionsmodell für Y mis-spezifiziert ist, wird der Schätzer bezüglich ψ_{ATE} korrigiert, über passende Gewichtung der Residuen

Beispiel

i	L	A	Y	$\frac{I(A_i=1)}{P(A_i=1\vert L_i)} - \frac{I(A_i=0)}{P(A_i=0\vert L_i)}$	$\hat{Y}$
1	0	0	0	?	?
2	0	0	1	?	?
3	0	0	0	?	?
4	0	1	0	?	?
5	0	1	1	?	?
6	0	1	1	?	?
7	0	0	0	?	?
8	1	0	0	?	?
9	1	1	0	?	?
10	1	1	1	?	?
11	1	0	1	?	?
12	1	0	1	?	?

Beispiel

Schätze $P(Y = 1|A, L)$ über logistische Regression:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.099      1.155  -0.951    0.341
A              1.792      1.683   1.064    0.287
L              1.792      1.683   1.064    0.287
A:L           -2.485      2.517  -0.987    0.323
```

z.B: Schätze $P(Y = 1|A = 1, L = 1)$ über:

$$\frac{1}{1 + exp(-[-1.099 + 1.792 \cdot 1 + 1.792 \cdot 1 - 2.485 \cdot 1 \cdot 1])} = 0.5$$

Beispiel

i	L	A	Y	$\frac{I(A_i=1)}{P(A_i=1\|L_i)} - \frac{I(A_i=0)}{P(A_i=0\|L_i)}$	$\hat{Y}$
1	0	0	0	?	0.25
2	0	0	1	?	0.25
3	0	0	0	?	0.25
4	0	1	0	?	0.67
5	0	1	1	?	0.67
6	0	1	1	?	0.67
7	0	0	0	?	0.25
8	1	0	0	?	0.67
9	1	1	0	?	0.50
10	1	1	1	?	0.50
11	1	0	1	?	0.67
12	1	0	1	?	0.67

Beispiel

Mit

$$
\begin{aligned}
p_{11} = P(A=1|L=1) &= \frac{2}{5} \\
p_{01} = P(A=0|L=1) &= \frac{3}{5} \\
p_{10} = P(A=1|L=0) &= \frac{3}{7} \\
p_{00} = P(A=0|L=0) &= \frac{4}{7}
\end{aligned}
$$

Beispiel

i	L	A	Y	$\frac{I(A_i=1)}{P(A_i=1\|L_i)} - \frac{I(A_i=0)}{P(A_i=0\|L_i)}$	$\hat{Y}$
1	0	0	0	$-\frac{7}{4}$	0.25
2	0	0	1	$-\frac{7}{4}$	0.25
3	0	0	0	$-\frac{7}{4}$	0.25
4	0	1	0	$\frac{7}{3}$	0.67
5	0	1	1	$\frac{7}{3}$	0.67
6	0	1	1	$\frac{7}{3}$	0.67
7	0	0	0	$-\frac{7}{4}$	0.25
8	1	0	0	$-\frac{5}{3}$	0.67
9	1	1	0	$\frac{5}{2}$	0.50
10	1	1	1	$\frac{5}{2}$	0.50
11	1	0	1	$-\frac{5}{3}$	0.67
12	1	0	1	$-\frac{5}{3}$	0.67

Beispiel

$$
\begin{aligned}
\hat{\psi}_{\text{ATE, AIPTW}} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{I(A_i=1)}{P(A_i=1|L_i)} - \frac{I(A_i=0)}{P(A_i=0|L_i)}\right. \\
&\quad \left. \times \left(Y_i - \hat{Y}_i\right)\right\} + \underbrace{\hat{\psi}_{\text{ATE, g-formula}}}_{\text{Kapitel 5: } 0.17359} \\
&= 0.17359 + \underbrace{\frac{1}{12}\left\{-\frac{7}{4}\cdot(0-0.25) + \ldots - \frac{5}{3}(1-0.67)\right\}}_{=0} \\
&= 0.17359
\end{aligned}
$$

In *R*: korrektes Ergebnis, auch bei mis-spezifiziertem Modell

```
# 1) g-formula
m1 <- glm(Y~A*L, family=binomial)
# Setze A=1 und A=0
newdata1 <- as.data.frame(cbind(Y,A,L))
newdata1$A <- c(rep(1,12))
newdata0 <- as.data.frame(cbind(Y,A,L))
newdata0$A <- c(rep(0,12))
psi_gformula <- mean(predict(m1, type="response", newdata=newdata1))- # E(Y1)
mean(predict(m1, type="response", newdata=newdata0))  # E(Y0)

# 2) Aktualisierung mit AIPTW

(est_AIPTW <- mean(((A==1)/predict(m.w, type="response")-(A==0)/
                   (1-predict(m.w, type="response")))
                  *(Y-predict(m1, type="response")) + psi_gformula))

# korrekt auch bei Mis-spezifizierung

est_AIPTW2 <- mean(((A==1)/predict(m.w2, type="response")-(A==0)/
                  (1-predict(m.w2, type="response")))
                  *(Y-predict(m1, type="response"))) + psi_gformula
```

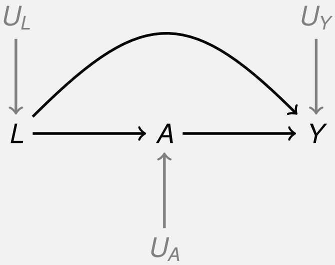
Ergebnis beide Male: $\hat{\psi}_{\text{AIPTW}} = 0.174$

Der graphische Ansatz

- gleiche Argumentation wie in Kapitel 5
- spezifiziere die Postinterventionsverteilung des DAGs für $A = a$ unter Verwendung der Markovbedingung
- Bestimme L mit Hilfe des Back-door Kriteriums
- Alternative Argumentation: “Pseudopopulation bei der es kein Confounding gibt”

Präinterventionsverteilung

Beispiel:

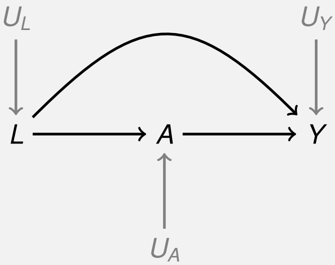


Über Markovbedingung (Kapitel 2):

$$P(A, L, Y) \quad = \quad P(Y|A, L) \times P(A|L) \times P(L)$$

Präinterventionsverteilung

Beispiel:



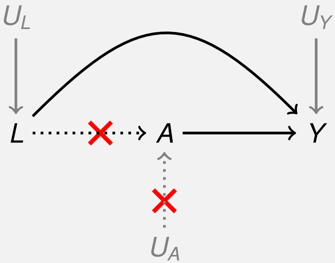
Strukturgleichungen:

$$
\begin{aligned}
L &= f_L(U_L) \\
A &= f_A(L, U_A) \\
Y &= f_Y(A, L, U_Y)
\end{aligned}
$$

Postinterventionsverteilung

Beispiel:

Streiche Pfeile, die in intervenierenden Knoten zeigen:



Strukturgleichungen:

$$
\begin{aligned}
L &= f_L(U_L) \\
A &= a \\
Y &= f_Y(A, L, U_Y)
\end{aligned}
$$

Postinterventionsverteilung

Verwendung der Markovbedingung sowie der Postinterventionsverteilung ergibt:

$$P(Y, L|do(A=a)) \quad = \quad P(Y|L, A=a) \times P(L)$$

und damit

$$\begin{aligned}
P(Y|do(A=a)) &= \sum_l P(Y|L=l, A=a) \times P(L=l) \\
&= \sum_l P(Y|L=l, A=a) \times P(L=l) \times \frac{P(A=a|L=l)}{P(A=a|L=l)} \\
&= \frac{P(Y=1|A=a) \times P(A=a)}{P(A=a|L)} \\
&= \frac{P(Y=1|A=a) \times P(A=a)}{P(A=a|L)} \\
&= \frac{1}{N} \sum_{i=1}^{N} \frac{y_i \cdot I(A_i=a)}{P(A_i=a|L_i)}
\end{aligned}$$

Literaturnachweise für Beispiele

[1] D. G. Horvitz and D. J. Thompson.
A generalization of sampling without replacement from a finite universe.
Journal of the American Statistical Association, 47(260):663–685, 1952.

[2] M. Hernan and J. Robins.
Causal inference.
Chapman & Hall/CRC, Boca Raton, 2017.

[3] M. A. Luque-Fernandez, G. Van Cutsem, E. Goemaere, K. Hilderbrand, M. Schomaker, N. Mantangana, S. Mathee, V. Dubula, N. Ford, M. A. Hernan, and A. Boulle.
Effectiveness of patient adherence groups as a model of care for stable patients on antiretroviral therapy in Khayelitsha, Cape Town, South Africa.
Plos One, 8(2):e56088, 2013.

[4] M. L. Petersen, K. E. Porter, S. Gruber, Y. Wang, and M. J. van der Laan.
Diagnosing and responding to violations in the positivity assumption.
Statistical Methods in Medical Research, 21(1):31–54, 2012.

[5] P. C. Austin.
Variance estimation when using inverse probability of treatment weighting (iptw) with survival analysis.
Statistics in Medicine, 35(30):5642–5655, 2016.

[6] J. M. Robins, A. Rotnitzky, and L. P. Zhao.
Estimation of regression-coefficients when some regressors are not always observed.
Journal of the American Statistical Association, 89(427):846–866, 1994.

[7] A. N. Glynn and K. M. Quinn.
An introduction to the augmented inverse propensity weighted estimator.
Political Analysis, 18(1):36–56, 2010.