More Efficient Estimation for Logistic Regression with Optimal Subsamples

Seminarvortrag Nils Kornacker

22.06.2021

Inhaltsverzeichnis

∘ Erweiterte ∘ Einführung

Theorem 1
Lemma 1-4, Basic Corollary, Theorem 1 Beweis, Lemma 1 Beweis, Proposition 1, Bemerkungen

Poisson Subsampling

Theorem 2
Lemma 5-6, Theorem 2 Beweis, Proposition 2 & Beweis, Bemerkungen

Misspecifications

Theorem 3
(Beweis)

Theorem 4
(Beweis), Bemerkungen

Theorem 5
Bemerkungen

Theorem 6
Bemerkungen

Zusammenfassung

Literatur

Einführung

Basic:

Computing Power ist beschränkt aber die Menge der $\{$Daten$\}$ groß. $\overset{\texttt{Pr1}}{\Rightarrow}$
Wähle (i.g.S.) optimale Stichprobe (Subsample) - mittels optimaler Subsampling Probabilities. $\overset{\texttt{Pr2}}{\Rightarrow}$
Finde mittels dieser (i.g.S.) optimaleren Schätzer.

`Pr1`: Wang et al. (2018) unter *A-optimality Criterion*
`Pr2`: Kornacker.

Einführung$_{\pounds}$

Das Modell ist benannt durch

$$\text{LRM:} \quad \mathbf{P}(y = 1|x) = p(x, \beta) = \frac{e^{x^T\beta}}{1 + e^{x^T\beta}},$$

wobei $y \in \{0, 1\}, \ x \in \mathbb{R}^d, \ \beta \in \mathrm{K} \ \subset \mathbb{R}^d$ kompakt.

Schätze β mittels $\hat{\beta}_{\text{MLE}} = \arg\max_{\beta} \sum_{i=1}^{N} y_i x_i^T \beta - \log(1 + e^{x_i^T\beta})$ $\quad \Big|\underset{\text{Daten}}{\mathcal{D}_N}$

Problem: großer Rechenaufwand $O(\zeta N d^2)$.

ζ: Anzahl der Iterationen in der Optimierungsprozedur.

$_{\text{ung}}$Wang et al. (2018): $\hat{\beta}_w^{\pi} = \arg\max_{\beta} \sum_{i=1}^{n} \frac{1}{\pi_i^*}(y_i^* x_i^{*T} \beta - \log(1 + e^{x_i^{*T}\beta}))$

$\{\}^*$ - Subsample

π_i Subsampling Probabilities

Einführung$_\pounds$

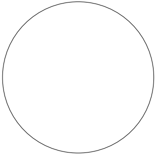
Subsample durch Ziehen mit Wiederholung gemäß $\{\pi_i\}$. Verschiedene Wahlen von π_i:

$$\pi_i^{\mathrm{OS}}(\hat{\beta}_{\mathrm{MLE}}) = \frac{|y_i - p(x_i, \hat{\beta}_{\mathrm{MLE}})|h(x_i)}{\sum_{j=1}^N |y_j - p(x_j, \hat{\beta}_{\mathrm{MLE}})|h(x_j)}$$

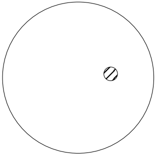
Wang et al. (2018): Spezielle Wahlen: $h(x) = \|M_N^{-1}x\|$ (π_i^{Aopt}), $h(x) = \|x\|$ (π_i^{Lopt}) minimieren *trace* der (transformierten) Kovarianzmatrix von $\hat{\beta}_w^\pi$. π_i^{Aopt} minimiert den asymptotischen mittleren quadr. Fehler von $\hat{\beta}_w^\pi$.

Problem: $\hat{\beta}_{\mathrm{MLE}}$ nicht bestimmt.
Jedoch π_i abhängig von $\underline{\hat{\beta}_{\mathrm{MLE}}/\beta}$

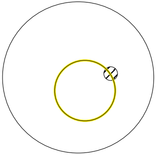
<u>Lösung</u>: Wir verwenden einen pilot estimator von einem pilot subsample.



$$\mathcal{D}_N = ((x_1, y_1), \dots, (x_N, y_N))$$



Pilot Subsample $((x_1^{*0}, y_1^{*0}), \dots, (x_{n_0}^{*0}, y_{n_0}^{*0}))$
$\Rightarrow$ Pilot Estimator $\hat{\beta}_0$



2.nd. Stage Subsample $((x_1^{*}, y_1^{*}), \dots, (x_{n}^{*}, y_{n}^{*}))$
$\Rightarrow$ Estimator $\hat{\beta}$

Einführung

Im Allgemeinen sind ungewichtete Schätzer im logistischen Regressionsmodell, wenn man ein Subsample wie hier wählt, verzerrt. Dies und der Fakt, dass $\{\pi_i^{\mathrm{OS}}(\hat{\beta}_0)\}_{i=1}^N$ von $\{y_i\}_{i=1}^N$ abhängen unterstützt den *weighted Estimator* in seiner präliminären Wahl. Eine bessere (dieses besser soll bald begründet werden vergleiche Proposition 1 und vorhergehende Bemerkungen) Wahl ist die Folgende.

More Efficient Estimator

Es sei $((x_1^*, y_1^*), ..., (x_n^*, y_n^*))$ ein Subsample erhalten durch Ziehen mit Zurücklegen, gemäß $\pi_i^{\mathrm{OS}}(\hat{\beta}_0)_{i=1}^N$. $\hat{\beta}_0$ sei ein pilot estimator für β_t mit $\hat{\beta}_0 = \beta_t + o_P(1)$ - vergleiche meinen Vortrag im letztjährigen Seminar vom 28.07.2020[1].
Wir definieren

$$\tilde{\beta}_{\mathrm{uw}} := \arg\max_{\beta} \sum_{i=1}^{n} (y_i^* {x_i^*}^T \beta - \log(1 + e^{{x_i^*}^T \beta})),$$

$$\hat{\beta}_{\mathrm{uw}} := \tilde{\beta}_{\mathrm{uw}} + \hat{\beta}_0.$$

$\hat{\beta}_{\mathrm{uw}}$ ist ein unverzerrter Schätzer (asymptotisch) für β_t und der, respektive ein More Efficient Estimator. Siehe folgendes Theorem 1.

[1] https://www.stochastik.uni-freiburg.de/lehre/ss-2020/seminar-mathematical-foundations-of-statistical-learning-ss-2020

Annahme 1-3

Annahme 1: $\mathbf{E}[\phi(\beta_t)h(x)xx^T]$ ist endlich und positiv-definit.
Annahme 2: $\mathbf{E}[\|x\|^2 h^2(x)] < \infty,\ \mathbf{E}[\|x\|^2 h(x)] < \infty.$
Annahme 3: Wenn $n \to \infty$, $n\mathbf{E}[h(x)I(\|x\|^2 > n)] \to 0.$

$$\phi(\beta_t) = p(x, \beta_t)(1 - p(x, \beta_t))$$

Die Annahmen 2 und 3 sind insbesondere dann erfüllt, wenn
$h(x) = 1$ und $\mathbf{E}[\|x\|^2] < \infty$ oder $h(x) = \|x\|$ und $\mathbf{E}[\|x\|^4] < \infty$.

Wir werden zusätzlich die Annahme(n) $h(x) \overset{\text{f.s.}}{>} 0$ integrierbar treffen.
(Theorem 1,3,5, Lemma 3,4)

Theorem 1

Unter Annahmen 1-3, bedingt auf $\mathcal{D}_N$, wenn $\hat{\beta}_0$ konsistent, $n_0, n, N \to \infty$

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \Sigma_{\beta_t}).$$

Wenn $n/N \to 0$, dann

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \beta_t) \to_d \mathbb{N}(0, \Sigma_{\beta_t}).$$

$\hat{\beta}_{\text{wMLE}} = \arg\max_{\beta} \sum_{i=1}^{N} |y_i - p(x_i, \hat{\beta}_0)| h(x_i)[y_i x_i^T(\beta - \hat{\beta}_0) - \log(1 + e^{x_i^T(\beta - \hat{\beta}_0)})]$

ein gewichteter MLE, $\Sigma_\beta = \left[\frac{\mathbf{E}[\phi(\beta)h(x)xx^T]}{4\Phi(\beta)}\right]^{-1}$ und $\Phi(\beta) = \mathbf{E}[\phi(\beta)h(x)]$.

Wenn $\hat{\beta}_0$ *uniform* aus einer Substichprobe der Größe n_0 bestimmt wurde mit $n_0/\sqrt{N} = o(1)$ oder $\hat{\beta}_0$ unabhängig von $\mathcal{D}_N$, dann

$$\sqrt{N}(\hat{\beta}_{\text{wMLE}} - \beta_t) \to_d \mathbb{N}(0, \Sigma_{\text{wMLE}}),$$

wobei

$\Sigma_{\text{wMLE}} = [\mathbf{E}[\phi(\beta_t)h(x)xx^T]]^{-1}\mathbf{E}[\phi(\beta_t)h^2(x)xx^T][\mathbf{E}[\phi(\beta_t)h(x)xx^T]]^{-1}.$

Beweis siehe folgende Folien.

Beweis von Theorem 1

Lemma 1

Seien $v_1, \ldots, v_N$ i.i.d. Vektoren mit selber Verteilung wie v, sowie g_{1N} eine beschränkte und g_2 eine von N unabhängige Funktion. Wenn $g_{1N}(v) = o_P(1)$ und $\mathbf{E}[|g_2(v)|] < \infty$, dann

$$\frac{1}{N}\sum_{i=1}^{N} g_{1N}(v_i)g_2(v_i) = o_P(1).$$

Beweis im Vortrag.

Beweis von Theorem 1

Lemma 2

Sei $\eta_i = |\psi_i(\hat{\beta}_0)|\psi_i(\beta_t - \hat{\beta}_0)h(x_i)x_i$, wobei $\psi_i(\beta) = y_i - p(x_i, \beta)$. Unter Annahmen 1,2 und bedingt auf den konsistenten Schätzer $\hat{\beta}_0$, wenn $n_0/\sqrt{N} \to 0$, dann

$$\sqrt{N}(\hat{\beta}_{\text{wMLE}} - \beta_t) = \frac{\Sigma_{\beta_t}}{2\mathbf{E}[\phi(\beta_t)h(x)]} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta_i + o_P(1),$$

wobei obiges gegen eine Normalverteilung mit Erwartungswertvektor 0 und Kovarianzmatrix $\mathbf{E}[\phi(\beta_t)h(x)xx^T]^{-1}\mathbf{E}[\phi(\beta_t)h^2(x)xx^T]\mathbf{E}[\phi(\beta_t)h(x)xx^T]^{-1}$ konvergiert, für n_0 und N gegen unendlich.

Beweis siehe Wang (2019).

Beweis von Theorem $1_{£}$

Basic Corollary (siehe [Hjort11])

Sei $A_n(s) = \frac{1}{2}s^T Vs + U_n^T s + C_n + R_n(s)$ eine Folge von konvexen Funktionen, wobei V symmetrisch und positiv definit sei, $U_n = O_P(1)$, C_n beliebig und $R_n(s) = o_P(1)$ für alle s. Dann gilt für α_n, das arg min von A_n, dass $\alpha_n = \beta_n + o_P(1)$, wobei $\beta_n = -V^{-1}U_n$ das arg min von $\frac{1}{2}s^T Vs + U_n^T s + C_n$. Wenn zusätzlich $U_n \to_d U$, dann $\alpha_n \to_d -V^{-1}U$.

Ist $A_n(s) = \frac{1}{2}s^T V_n s + U_n^T s + C_n + R_n(s)$ konvex, V_n positiv semidefinit, symmetrisch mit $V_n = V + o_P(1)$, V positiv definit, hält das Ergebnis mit V_n ebenso $(R_n(s)' = R_n(s) + \frac{1}{2}s^T(V - V_n)s)$.

Beweis von Theorem 1

Lemma 3

Sei $\dot{\lambda}^*_{\mathrm{uw}}(\beta_t) = \sum_{i=1}^n (y_i^* - p_i^*(\beta_t - \hat{\beta}_0))x_i^*$. Unter Annahmen 1,2, bedingt auf $\mathcal{D}_N$ und den konsistenten Schätzer $\hat{\beta}_0$, wenn n_0, n und N gegen unendlich* streben

$$\frac{\dot{\lambda}^*_{\mathrm{uw}}(\beta_t)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\Psi_N(\hat{\beta}_0)} \to_d \mathbb{N}(0, \Sigma_{\beta_t}^{-1}),$$

wobei $\Psi_N(\beta) = N^{-1}\sum_{i=1}^N |y_i - p(x_i, \beta)| h(x_i)$.

Beweis siehe Wang (2019).

Lemma 4

Unter Annahmen 1-3, wenn n_0, n und N gegen unendlich streben und $s_n = o_P(1)$

$$\frac{1}{n}\sum_{i=1}^n \phi_i^*(\beta_t - \hat{\beta}_0 + s_n)\|x_i^*\|^2 - \sum_{i=1}^N \pi_i(\hat{\beta}_0)\phi_i(\beta_t - \hat{\beta}_0)\|x_i\|^2 = o_P(1).$$

Beweis siehe Wang (2019).

*und $n_0/\sqrt{N} \to 0$ oder $\hat{\beta}_0$ unabhängig von $\mathcal{D}_N$

Beweis von Theorem 1

$\hat{\beta}_{\text{uw}}$ maximiert $\lambda^*_{\text{uw}}(\beta) = \sum_{i=1}^n ((\beta - \hat{\beta}_0)^T x_i^* y_i^* - \log(1 + e^{(\beta-\hat{\beta}_0)^T x_i^*}))$
$\Rightarrow \sqrt{n}(\hat{\beta}_{\text{uw}} - \beta_t)$ maximiert $\gamma(s) = \lambda^*_{\text{uw}}(\beta_t + s/\sqrt{n}) - \lambda^*_{\text{uw}}(\beta_t)$.
TE: $\gamma(s) = \frac{1}{\sqrt{n}} s^T \dot{\lambda}^*_{\text{uw}}(\beta_t) - \frac{1}{2n} \sum_{i=1}^n \phi_i^*(\beta_t - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{n}})(s^T x_i^*)^2$. $\acute{s}$ zwischen 0 und s.

Lemma 4 $\Rightarrow \frac{1}{n} \sum_{i=1}^n \phi_i^*(\beta_t - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{n}}) x_i^* x_i^{*T} - \sum_{i=1}^N \pi_i(\hat{\beta}_0)\phi_i(\beta_t - \hat{\beta}_0) x_i x_i^T = o_P(1)$.
Lemma 1 und das swGGZ $\Rightarrow$
$\sum_{i=1}^N \pi_i(\hat{\beta}_0)\phi_i(\beta_t - \hat{\beta}_0) x_i x_i^T = \frac{\frac{1}{N}\sum_{i=1}^N |\psi_i(\hat{\beta}_0)| h(x_i)\phi_i(\beta_t - \hat{\beta}_0) x_i x_i^T}{\Psi_N(\hat{\beta}_0)} = \Sigma_{\beta_t}^{-1} + o_P(1)$.

$\Rightarrow \; n^{-1} \sum_{i=1}^n \phi_i^*(\beta_t - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{n}}) x_i^* x_i^{*T} = \Sigma_{\beta_t}^{-1} + o_P(1)$

Lemma 2, Lemma 3 $\Rightarrow \; \dot{\lambda}^*_{\text{uw}}(\beta_t)/\sqrt{n} = O_P(1)$. Basic Corollary $\Rightarrow$ Der Maximierer $\sqrt{n}(\hat{\beta}_{\text{uw}} - \beta_t)$ von $\gamma(s)$ erfüllt gegeben $\mathcal{D}_N$ und $\hat{\beta}_0$

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \beta_t) = \Sigma_{\beta_t} \frac{1}{\sqrt{n}} \dot{\lambda}^*_{\text{uw}}(\beta_t) + o_P(1)$$

Beweis von Theorem 1

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \beta_t) = \Sigma_{\beta_t} \frac{1}{\sqrt{n}} \dot{\lambda}^*_{\text{uw}}(\beta_t) + o_P(1)$$

Damit ist

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \hat{\beta}_{\text{wMLE}}) = \Sigma_{\beta_t} \left(\frac{1}{\sqrt{n}} \dot{\lambda}^*_{\text{uw}}(\beta_t) - \Sigma^{-1}_{\beta_t} \sqrt{n}(\hat{\beta}_{\text{wMLE}} - \beta_t) \right) + o_P(1).$$

Lemma 2 $\Rightarrow$

$$\Sigma^{-1}_{\beta_t} \sqrt{n}(\hat{\beta}_{\text{wMLE}} - \beta_t) = \frac{\sqrt{n} \sum_{i=1}^{N} \eta_i}{2N\mathbf{E}[\phi(\beta_t) h(x)]} + o_P(1) = \frac{\sqrt{n} \sum_{i=1}^{N} \eta_i}{N \Psi_N(\hat{\beta}_0)} + o_P(1)$$

und aus Slutsky + Lemma 3 folgt die Aussage. $\square$

Bemerkungen

Gezeigt: $\hat{\beta}_{\text{uw}} - \hat{\beta}_{\text{wMLE}} = O_{P|\mathcal{D}_N}(n^{-1/2}) \Rightarrow O_P(n^{-1/2})$
$\hat{\beta}_{\text{wMLE}} - \beta_t = O_P(N^{-1/2}) \Rightarrow \hat{\beta}_{\text{uw}} - \beta_t = O_P(n^{-1/2} + N^{-1/2}) = O_P(n^{-1/2})$ ($\Rightarrow \hat{\beta}_{\text{uw}}$ unbedingt $\sqrt{n}$-konsistent für β_t)

Der gewichtete Schätzer $\hat{\beta}_w$ hat asymptotische Kovarianzmatrix $\frac{1}{n} V_N^{\text{OS}} = \frac{1}{n} M_N^{-1} V_{Nc}^{\text{OS}} M_N^{-1}$ mit $M_N = \frac{1}{N}\sum_{i=1}^N \phi_i(\hat{\beta}_{\text{MLE}}) x_i x_i^T$, $V_{Nc}^{\text{OS}} = \left(\frac{1}{N}\sum_{i=1}^N |y_i - p(x_i, \hat{\beta}_{\text{MLE}})| h(x_i)\right)\left(\frac{1}{N}\sum_{i=1}^N \frac{|y_i - p(x_i, \hat{\beta}_{\text{MLE}})| x_i x_i^T}{h(x_i)}\right)$.
Unter der zusätzlichen Voraussetzung $\mathbf{E}[\|x\|^2/h(x)] < \infty$ folgt mit Lemma 1 und dem swGGZ $V_N^{\text{OS}} = V^{\text{OS}} + o_P(1)$, $V^{\text{OS}} = M^{-1} V_c^{\text{OS}} M^{-1}$, mit $M = \mathbf{E}[\phi(\beta_t) x x^T]$ und $V_c^{\text{OS}} = 4\Phi(\beta_t)\mathbf{E}\left[\frac{\phi(\beta_t) x x^T}{h(x)}\right]$. Insbesondere ist gegeben $\mathcal{D}_N$ die Verteilung von $\hat{\beta}_w$ um $\hat{\beta}_{\text{MLE}}$ zentriert und es kann unter Annahme 1 und 2 gezeigt werden

$$\sqrt{N}(\hat{\beta}_{\text{MLE}} - \beta_t) \to_d \mathbb{N}(0, M^{-1}).$$

Bemerkungen

Damit sind für die Effizienzbestimmung von $\hat{\beta}_w$ als Schätzer für β_t beide Matrizen $n^{-1}V^{\mathrm{OS}}$ und $N^{-1}M^{-1}$ von Bedeutung (im Falle von $\hat{\beta}_{\mathrm{uw}}$ genauso $n^{-1}\Sigma_{\beta_t}$ und $N^{-1}\Sigma_{\mathrm{wMLE}}$). Da im Allgemeinen aber $n \ll N$ ist die Qualitätsbemessung hinsichtlich Effizienz vor allem von $n^{-1}V^{\mathrm{OS}}$ (und $n^{-1}\Sigma_{\beta_t}$) abhängig. Damit ist für den Vergleich von $\hat{\beta}_{\mathrm{uw}}$ und $\hat{\beta}_w$ die Relation Σ_{β_t} zu V^{OS} wesentlich. Dazu folgende Proposition

Proposition 1

Wenn M, V_c^{OS} und Σ_{β_t} endliche und positiv definite Matrizen sind, dann gilt

$$\Sigma_{\beta_t} \leq V^{\mathrm{OS}}$$

in der Loewner-Ordnung [A und B positiv semidefinit: $A \geq B \Leftrightarrow A - B$ ist positiv semidefinit]. Für $h(x) = 1$ gilt Gleichheit.

Beweis siehe Wang (2019). Dieser verwendet das starke Gesetz der großen Zahlen und eine Eigenschaft von Projektionsmatrizen.

Bemerkungen

Damit folgt, dass $\hat{\beta}_{\text{uw}}$ für $n/N \to 0$ ein effizienterer Schätzer (MEE) für β_t als $\hat{\beta}_w$ ist. Stärker kann man dies für $n/N \to \rho$ mit $0 < \rho < 1$ hinreichend klein zeigen.

(Bedingt auf $\mathcal{D}_N$)

Theorem 1

$$\sqrt{N}(\hat{\beta}_{\text{uw}} - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \rho^{-1}\Sigma_{\beta_t})$$
$$\sqrt{N}(\hat{\beta}_{\text{wMLE}} - \beta_t) \to_d \mathbb{N}(0, \Sigma_{\text{wMLE}}).$$

$$\sqrt{N}(\hat{\beta}_w - \hat{\beta}_{\text{MLE}}) \to_d \mathbb{N}(0, \rho^{-1}V^{\text{OS}})$$
$$\sqrt{N}(\hat{\beta}_{\text{MLE}} - \beta_t) \to_d \mathbb{N}(0, M^{-1})$$

($\Sigma_{\text{wMLE}} = M^{-1}$ wenn $\Sigma_{\beta_t} = V^{\text{OS}}$)

Ist demnach ρ hinreichend klein, so sind die dominierenden Matrizen

$$\rho^{-1}\Sigma_{\beta_t} \le \rho^{-1}V^{\text{OS}}.$$

Poisson subsampling

$\hat{\beta}_{\text{uw}}$; Ziehen mit Zurücklegen.
$\hat{\beta}_p$; Zufällige Stichprobengröße (aber kein Zurücklegen).
Vorteil: *Gewichtsanpassung* in der Maximierungsfunktion, weniger Arbeitspeicher, 'one pass through the data'

Sei $\hat{\Psi}_0$ pilot estimator von $\mathbf{E}[|\psi(\beta_t)|h(x)]$

$$= \lim_{N\to\infty} N^{-1} \underbrace{\sum_{i=1}^{N} |y_i - p(x_i, \beta_t)|h(x_i)}_{\text{Nenner von } \pi_i^{\text{OS}}}$$

Prinzip: Berechne $\pi_i^p = |y_i - p(x_i, \hat{\beta}_0)|h(x_i)/(N\hat{\Psi}_0)$, generiere $u_i \sim U(0,1)$ und inkludiere (x_i, y_i, π_i^p) in das Subsample, wenn $u_i \le n\pi_i^p$. Nun berechne

$$\tilde{\beta}_p = \arg\max_{\beta} \sum_{i=1}^{n^*} (\boxed{n\pi_i^{p*} \vee 1})(\beta^T x_i^* y_i^* + \log(1 + e^{\beta^T x_i^*}))$$

und setze $\hat{\beta}_p = \tilde{\beta}_p + \hat{\beta}_0$.

Theorem 2

Es gelten Annahmen 1 und 2 und $\hat{\beta}_0$ sei konsistenter Schätzer. Dann ist bedingt auf $\mathcal{D}_N$ für n_0, n und $N \to \infty$ mit $n/N \to 0^*$

$$\sqrt{n}(\hat{\beta}_p - \beta_t) \to_d \mathbb{N}(0, \Sigma_{\beta_t}).$$

Ist $n/N \to \rho \in (0, 1)$, so gilt

$$\sqrt{n}(\hat{\beta}_p - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \Sigma_{\beta_t} \Lambda_\rho \Sigma_{\beta_t})$$

mit $\Lambda_\rho = \mathbf{E}[|\psi(\beta_t)|h(x)(\Psi(\beta_t) - \rho|\psi(\beta_t)|h(x))_+ xx^T]/(4\Psi^2(\beta_t))$.

Beweis siehe folgende Folien.

*und $n_0/\sqrt{N} \to 0$ oder $\hat{\beta}_0$ unabhängig von $\mathcal{D}_N$

Beweis von Theorem 2

Lemma 5

Sei $\dot{\lambda}_p(\beta_t) = \sum_{i=1}^N \delta_i^{\hat{\beta}_0}(n\pi_i^p(\hat{\beta}_0) \vee 1)(y_i - p(x_i, \beta_t - \hat{\beta}_0))x_i$. Unter Annahmen 1 und 2, bedingt auf $\mathcal{D}_N$ sowie die konsistenten Schätzer $\hat{\beta}_0$ und $\hat{\Psi}_0$ gilt, wenn $n/N \to 0$, dass*

$$\frac{\dot{\lambda}_p(\beta_t)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0} \to_d \mathbb{N}(0, \Sigma_{\beta_t}^{-1}),$$

wobei hier $\delta_i^{\hat{\beta}_0} := I(u_i \leq n\pi_i^p(\hat{\beta}_0))$. Konvergiert n/N gegen $\rho \in (0,1)$, dann

$$\frac{\dot{\lambda}_p(\beta_t)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0} \to_d \mathbb{N}(0, \Lambda_\rho).$$

Beweis siehe Wang (2019).

*und $n_0/\sqrt{N} \to 0$ oder $\hat{\beta}_0$ unabhängig von $\mathcal{D}_N$

Lemma 6

Unter Annahmen 1 und 2, wenn $n_0, n, N \to \infty$, und $s_n = o_P(1)$

$$\frac{1}{n}\sum_{i=1}^{N}\delta_i^{\hat{\beta}_0}(n\pi_i^p(\hat{\beta}_0) \vee 1)\phi_i(\beta_t - \hat{\beta}_0 + s_n)\|x_i\|^2 - \sum_{i=1}^{N}\pi_i^p(\hat{\beta}_0)\phi_i(\beta_t - \hat{\beta}_0)\|x_i\|^2 = o_P(1).$$

Beweis siehe Wang (2019).

$\hat{\beta}_p$ maximiert
$\lambda_p(\beta) = \sum_{i=1}^N \delta_i^{\hat{\beta}_0}(n\pi_i^p(\hat{\beta}_0) \vee 1)((\beta - \hat{\beta}_0)^T x_i y_i - \log(1 + e^{(\beta-\hat{\beta}_0)^T x_i}))$
$\Rightarrow$ $\sqrt{n}(\hat{\beta}_p - \beta_t)$ maximiert $\gamma_p(s) = \lambda_p(\beta_t + s/\sqrt{n}) - \lambda_p(\beta_t)$.
TE: $\gamma_p(s) = \frac{1}{\sqrt{n}} s^T \dot{\lambda}_p(\beta_t) - \frac{1}{2n}\sum_{i=1}^N \delta_i^{\hat{\beta}_0}(n\pi_i^p(\hat{\beta}_0) \vee 1)\phi_i(\beta_t - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{n}})(s^T x_i)^2$,
$\acute{s}$ zwischen 0 und s.

Lemma 5 & 6 $\Rightarrow$

$$\frac{1}{n}\sum_{i=1}^{N}\delta_i^{\hat\beta_0}(n\pi_i^p(\hat\beta_0)\vee 1)\phi_i(\beta_t-\hat\beta_0+\frac{\acute{s}}{\sqrt{n}})x_ix_i^T=\Sigma_{\beta_t}^{-1}+o_P(1)$$

bedingt auf $\mathcal{D}_N$ und $\hat\beta_0$.
Lemma 5 $\Rightarrow$ $\dot\lambda_p(\beta_t)/\sqrt{n}$ konvergiert gegen $\mathbb{N}(0,\Sigma_{\beta_t}^{-1})$, resp. ist $O_P(1)$ (wichtig). Basic Corollary $\Rightarrow$

$$\sqrt{n}(\hat\beta_p-\beta_t)=\Sigma_{\beta_t}\frac{1}{\sqrt{n}}\dot\lambda_p(\beta_t)+o_P(1)$$

gegeben $\mathcal{D}_N,\hat\beta_0$ und $\hat\Psi_0$. Slutsky und Lemma 5 liefern im ersten Fall die Aussage. Die zweite Konvergenz sehen wir wie folgt ein:

$$\begin{aligned}\sqrt{n}(\hat\beta_p-\hat\beta_{\text{wMLE}})&=\sqrt{n}(\hat\beta_p-\beta_t)+\sqrt{n}(\beta_t-\hat\beta_{\text{wMLE}})\\&\overset{\text{Lemma 2}}{=}\Sigma_{\beta_t}\left(\frac{\dot\lambda_p(\beta_t)}{\sqrt{n}}-\frac{\sqrt{n}\sum_{i=1}^N\eta_i}{N\hat\Psi_0}+\frac{\sqrt{n}\sum_{i=1}^N\eta_i}{N\hat\Psi_0}\right)+o_P(1)\\&\quad-\Sigma_{\beta_t}\frac{\sqrt{n}\sum_{i=1}^N\eta_i}{N\Psi(\beta_t)}+\frac{\sqrt{n}}{\sqrt{N}}o_P(1)\end{aligned}$$

Beweis von Theorem 2

$$= \Sigma_{\beta_t}\left(\frac{\dot{\lambda}_p(\beta_t)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0} + \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0}\right) - \Sigma_{\beta_t}\frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N(\hat{\Psi}_0 + o_P(1))} + o_P(1)$$

$$\stackrel{\text{Lemma 2}}{=} \Sigma_{\beta_t}\left(\frac{\dot{\lambda}_p(\beta_t)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0} + \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0}\right) - \Sigma_{\beta_t}\frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0} + o_P(1)$$

$$= \Sigma_{\beta_t}\left(\frac{\dot{\lambda}_p(\beta_t)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0}\right) + o_P(1) \stackrel{\text{Lemma 5}}{\to_d} \mathbb{N}(0, \Sigma_{\beta_t}\Lambda_\rho\Sigma_{\beta_t})$$

da $\frac{\sum_{i=1}^N \eta_i}{\sqrt{N}} = O_P(1)$. $\square$

Bemerkungen

Analog zu Theorem 1 gilt also $\hat{\beta}_p - \hat{\beta}_{\text{wMLE}} = O_{P|\mathcal{D}_N}(n^{-1/2})$ ($\Rightarrow$ $\hat{\beta}_p - \hat{\beta}_{\text{wMLE}} = O_P(n^{-1/2})$) und $\hat{\beta}_p - \beta_t = O_P(n^{-1/2})$.

Auch sind nach Theorem 1 und 2 die Schätzer $\hat{\beta}_{\text{uw}}$ und $\hat{\beta}_p$ für $n/N \to 0$ asymptotisch identisch verteilt (intuitiv).
Ist nun aber $n/N \to \rho \neq 0$ so unterscheiden sich (der Erwartungswert ist in beiden Fällen gleich) die asymptotischen Kovarianzmatrizen.

Proposition 2

Ist $\rho > 0$ und Σ_{β_t} eine endliche und positiv definite Matrix, dann gilt

$$\Sigma_{\beta_t} \Lambda_\rho \Sigma_{\beta_t} < \Sigma_{\beta_t}.$$

(wobei hierfür vermutlich die Voraussetzungen nicht ausreichen)

Beweis im Vortrag.

Bemerkungen

Damit ist $\hat{\beta}_p$ effizienter als $\hat{\beta}_{\mathrm{uw}}$ und somit Poisson Subsampling effizienter als Ziehen mit Zurücklegen.

Misspecifications I

Im Folgenden nehmen wir an, dass die pilot Estimators $\hat{\beta}_0, \hat{\Psi}_0$ unabhängig von den Daten $\mathcal{D}_N$ sind. Dies ist begründet durch die Tatsache, dass jene oft mittels differenter Datenquellen berechnet werden.

Frage: Was passiert wenn unser angenommenes (logistisches) Modell inkorrekt ist und eigentlich eine (uns unbekannte) Funktion p_t die bedingte Wahrscheinlichkeit spezifiziert:

$$\mathbf{P}(y = 1|x) = p_t(x)?$$

Der (pilot) Estimator $\hat{\beta}$ schätzt nun ...?

Misspecifications I

Problem: Wir brauchen eine Neuerung des Konsistenzbegriffes.

Das oft verwendete Kriterium ist, dass der Grenzwert den erwarteten Verlust bezüglich einer Verlustfunktion minimiert. Wir notieren den Grenzwert von $\hat{\beta}$ als β_l und definieren ihn als Minimierer von

$$\mathbf{E}\left[-p_t(x)h(x)x^T\beta + h(x)\log(1 + e^{\beta^T x})\right].$$

Dann erfüllt β_l

$$\mathbf{E}[(p_t(x) - p(x, \beta_l))h(x)x] = 0 \qquad (\star)$$

mit $p(x, \beta) = e^{x^T\beta}(1 + e^{x^T\beta})^{-1}$.

Theorem 3

Sei das pilot Subsample unabhängig von $\mathcal{D}_N$, $\sqrt{n_0}(\hat{\beta}_0 - \beta_l) \to_d \mathbb{N}(0, \Sigma_0)$ und Annahmen 1-3 gelten. Wenn $n_0/N \to \rho_0,\ n/N \to \rho$ mit $\rho_0, \rho \in (0,1)$, dann ist für $n_0, n, N \to \infty$ bedingt auf $\mathcal{D}_N$

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \omega\kappa_a^{-1}),$$

wobei $\kappa_a = \frac{1}{4}\mathbf{E}[(p_t(x) - 2p_t(x)p(x,\beta_l) + p(x,\beta_l))h(x)xx^T],\ \omega = \mathbf{E}[(p_t(x) - 2p_t(x)p(x,\beta_l) + p(x,\beta_l))h(x)]$.

$\hat{\beta}_{\text{wMLE}}$ erfüllt

$$\sqrt{N}(\hat{\beta}_{\text{wMLE}} - \beta_l) \to_d \mathbb{N}(0, \kappa_a^{-1}(\kappa_b + \rho_0^{-1}\kappa_c\Sigma_0\kappa_c)\kappa_a^{-1}),$$

wobei $\kappa_b = \frac{1}{4}\mathbf{E}[(p_t(x) - 2p_t(x)p(x,\beta_l) + p^2(x,\beta_l))h^2(x)xx^T],\ \kappa_c = \frac{1}{4}\mathbf{E}[(1 - 2p(x,\beta_l))(p_t(x) - p(x,\beta_l))h(x)xx^T]$.[2]

Beweis siehe folgende Folien.

[2] Vermutlich benötigen wir κ_a pos. def.!

Beweis von Theorem 3

Bereits eingesehen (mit β_t): $\sqrt{n}(\hat{\beta}_{\mathrm{uw}} - \beta_l)$ maximiert $\gamma(s) = \frac{1}{\sqrt{n}} s^T \dot{\lambda}^*_{\mathrm{uw}}(\beta_l) - \frac{1}{2n} \sum_{i=1}^n \phi_i^*(\beta_l - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{n}})(s^T x_i^*)^2$, $\acute{s}$ zwischen 0 und s. Wir bemerken $\dot{\lambda}^*_{\mathrm{uw}}(\beta_l) = \sum_{i=1}^n (y_i^* - p(x_i^*, \beta_l - \hat{\beta}_0))x_i^*$ und „abuse the notation“: $\eta_i = |\psi_i(\hat{\beta}_0)|\psi_i(\beta_l - \hat{\beta}_0)h(x_i)x_i$. Man kann sehr leicht zeigen, dass

$$\mathbf{E}[(y^* - p(x^*, \beta_l - \hat{\beta}_0))x^* | \mathcal{D}_N, \hat{\beta}_0] = \frac{\sum_{i=1}^N \eta_i}{N\Psi_N(\hat{\beta}_0)} \quad \text{sowie}$$

$$\mathbf{V}[(y^* - p(x^*, \beta_l - \hat{\beta}_0))x^* | \mathcal{D}_N, \hat{\beta}_0] = \frac{\kappa_a}{\omega} + o_P(1).$$

Anwendung des bedingten Lindeberg-Feller Theorems liefert bedingt auf $\mathcal{D}_N$ und $\hat{\beta}_0$, dass

$$\frac{\dot{\lambda}^*_{\mathrm{uw}}(\beta_l)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\Psi_N(\hat{\beta}_0)} \xrightarrow[d]{n_0, n, N \to \infty} \mathbb{N}\left(0, \frac{\kappa_a}{\omega}\right).$$

Eine Rechnung zeigt

$$\sum_{i=1}^N \eta_i = \frac{1}{2}\sum_{i=1}^N (2y_i - 1)^2 (y_i - p_i(x_i, \beta_l))h(x_i)x_i + \kappa_c N(\hat{\beta}_0 - \beta_l) + o_P(N)(\hat{\beta}_0 - \beta_l).$$

Der ZGWS, Slutsky, die Unabhängigkeit von $\hat{\beta}_0$ und $\mathcal{D}_N$ sowie $\star$ liefern

$$\frac{\sum_{i=1}^N \eta_i}{\sqrt{N}} \to_d \mathbb{N}\left(0, \kappa_b + \frac{\kappa_c \Sigma_0 \kappa_c}{\rho_0}\right).$$

Ähnlich wie in Lemma 4 ist $\frac{1}{n}\sum_{i=1}^n \phi^*(\beta_l - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{n}})x_i^* x_i^{*T} = \frac{\kappa_a}{\omega} + o_P(1)$. κ_a ist positiv definit, $\dot{\lambda}^*_{\mathrm{uw}}(\beta_l)/\sqrt{n} = O_P(1)$ B.C.

$$\sqrt{n}(\hat{\beta}_{\mathrm{uw}} - \beta_l) = \omega\kappa_a^{-1}\dot{\lambda}^*_{\mathrm{uw}}(\beta_l)/\sqrt{n} + o_P(1)$$

gegeben $\mathcal{D}_N$ und $\hat{\beta}_0$.Wie im Beweis von Lemma 2 folgt $\sqrt{N}(\hat{\beta}_{\mathrm{wMLE}} - \beta_l)$ maximiert $\frac{1}{\sqrt{N}}s^T\sum_{i=1}^N \eta_i - \frac{1}{2N}\sum_{i=1}^N |\psi_i(\hat{\beta}_0))|h(x_i)\phi_i(\beta_l - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{N}})(s^T x_i)^2$,
$\acute{s}$ zwischen 0 und s. Lemma 1 $\Rightarrow$

$$\frac{1}{N}\sum_{i=1}^N |\psi_i(\hat{\beta}_0))|h(x_i)\phi_i(\beta_l - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{N}})x_i x_i^T = \kappa_a + o_P(1).$$

B.C. $\Rightarrow \sqrt{N}(\hat{\beta}_{\mathrm{wMLE}} - \beta_l) = \kappa_a^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \eta_i + o_P(1)$. Teil 2 ist gezeigt.

Beweis von Theorem 3

Es folgt sofort

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \hat{\beta}_{\text{wMLE}}) = \omega\kappa_a^{-1}\left(\frac{\dot{\lambda}^*_{\text{uw}}(\beta_l)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{\omega N}\right) + o_P(1)$$

und damit (und Slutsky) auch Teil 1. □

Ist $p_t(x) = p(x, \beta_t)$ (keine Misspecification) stimmt Theorem 3 mit Theorem 1 überein.

Da $\sqrt{N}(\hat{\beta}_{\text{wMLE}} - \beta_l) \to_d \mathbb{N}(0, \kappa_a^{-1}(\kappa_b + \rho_0^{-1}\kappa_c\Sigma_0\kappa_c)\kappa_a^{-1})$ und $n_0 \ll N$ kann die Kovarianzmatrix groß sein.

Theorem 4

Angenommen das pilot subsample ist unabhängig von $\mathcal{D}_N$ und die pilot estimators erfüllen $\sqrt{n_0}(\hat{\beta}_0 - \beta_l) \to_d \mathbb{N}(0, \Sigma_0)$ und $\hat{\Psi}_0 = \omega + o_P(1)$. Unter Annahmen 1 und 2, wenn $n_0/N \to \rho_0$ und $n/N \to \rho$ mit $\rho_0, \rho \in (0,1)$ ist bedingt auf $\mathcal{D}_N$ für $n_0, n, N \to \infty$

$$\sqrt{n}(\hat{\beta}_p - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \underline{\omega} \kappa_a^{-1} \kappa_d \kappa_a^{-1}),$$

wobei $\kappa_d = \frac{1}{4}\mathbf{E}\left[|\psi(\beta_l)|(1 - \rho\omega^{-1}|\psi(\beta_l)|h(x))_+ h(x)xx^T\right]$.

Beweis siehe folgende Folien.

Bemerke insbesondere, dass hier keine Konvergenzrate für $\hat{\Psi}_0$ gefordert ist.

Beweis von Theorem 4

Wie im Beweis zu Theorem 2: $\sqrt{n}(\hat{\beta}_p - \beta_l)$ maximiert
$\frac{1}{\sqrt{n}} s^T \dot{\lambda}_p(\beta_l) - \frac{1}{2n} \sum_{i=1}^N \delta_i^{\hat{\beta}_0}(n\pi_i^p(\hat{\beta}_0) \vee 1)\phi_i(\beta_l - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{n}})(s^T x_i)^2$,
$\acute{s}$ zwischen 0 und s und

$$\dot{\lambda}_p(\beta_l) = \sum_{i=1}^N \delta_i^{\hat{\beta}_0}(n\pi_i^p(\hat{\beta}_0) \vee 1)(y_i - p(x_i, \beta_l - \hat{\beta}_0))x_i.$$

Gegeben $\mathcal{D}_N, \hat{\beta}_0$ und $\hat{\Psi}_0$ sind die Summanden unabhängig und wie im Beweis von Lemma 5 (siehe Wang) ist die bedingte Lindeberg-Feller Bedingung erfüllt. Wir berechnen den (bed.) Erwartungswert und die Kovarianzmatrix

$$\frac{1}{\sqrt{n}} \mathbf{E}[\dot{\lambda}_p(\beta_l) | \mathcal{D}_N, \hat{\beta}_0, \hat{\Psi}_0] = \frac{\sqrt{n}}{\sqrt{N}} \frac{\sum_{i=1}^N \eta_i}{\hat{\Psi}_0 \sqrt{N}} \overset{\text{s.o.}}{=} O_P(\sqrt{n/N}).$$

$$\frac{1}{n} \mathbf{V}[\dot{\lambda}_p(\beta_l) | \mathcal{D}_N, \hat{\beta}_0, \hat{\Psi}_0] \overset{*}{=} \frac{1}{n} \sum_{i=1}^N \mathbf{E}[\delta_i^{\hat{\beta}_0}](n\pi_i^p(\hat{\beta}_0) \vee 1)^2 (y_i - p(x_i, \beta_l - \hat{\beta}_0))^2 x_i x_i^T$$
$$- \frac{n}{N} \frac{\frac{1}{N} \sum_{i=1}^N \psi_i^2(\hat{\beta}_0) \psi_i^2(\beta_l - \hat{\beta}_0) h^2(x_i) x_i x_i^T}{\hat{\Psi}_0^2}$$

$^*\delta_i^{\hat{\beta}_0} = I(u_i \leq n\pi_i^p(\hat{\beta}_0))$

$$= \frac{1}{n}\sum_{i=1}^{N}\left(n\pi_i^p(\hat{\beta}_0)\wedge 1\right)\left(n\pi_i^p(\hat{\beta}_0)\vee 1\right)^2\left(y_i - p(x_i,\beta_l-\hat{\beta}_0)\right)^2 x_i x_i^T$$
$$- \frac{n}{N}\frac{\frac{1}{N}\sum_{i=1}^{N}\psi_i^2(\hat{\beta}_0)\psi_i^2(\beta_l-\hat{\beta}_0)h^2(x_i)x_i x_i^T}{\hat{\Psi}_0^2}$$
$$= \frac{\frac{1}{N}\sum_{i=1}^{N}|\psi_i(\hat{\beta}_0)|\left(n\pi_i^p(\hat{\beta}_0)\vee 1\right)\psi_i^2(\beta_l-\hat{\beta}_0)h(x_i)x_i x_i^T}{\hat{\Psi}_0} \qquad (=\Delta)$$
$$- \frac{n}{N}\frac{\frac{1}{N}\sum_{i=1}^{N}\psi_i^2(\hat{\beta}_0)\psi_i^2(\beta_l-\hat{\beta}_0)h^2(x_i)x_i x_i^T}{\hat{\Psi}_0^2} \qquad \left(= -\rho\frac{\kappa_b}{\omega^2} + o_P(1)\right).$$

Es gilt nach dem Beweis von Lemma 5 (siehe Wang)

$$\Delta = \frac{\mathbf{E}\left[|\psi(\beta_l)|\left((\rho|\psi(\beta_l)|h(x))\vee\omega\right)h(x)xx^T\right]}{4\omega^2} + o_P(1)$$

und damit

Beweis von Theorem 4

$$\frac{1}{n}\mathbf{V}[\dot{\lambda}_p(\beta_l)|\mathcal{D}_N, \hat{\beta}_0, \hat{\Psi}_0] = \frac{\mathbf{E}[|\psi(\beta_l)|((\rho|\psi(\beta_l)|h(x)) \vee \omega)h(x)xx^T]}{4\omega^2} - \frac{4\rho\frac{1}{4}\mathbf{E}[(p_t(x) - 2p_t(x)p(x,\beta_l) + p^2(x,\beta_l))h^2(x)xx^T]}{4\omega^2} + o_P(1)$$

Bedingt auf $\mathcal{D}_N, \hat{\beta}_0$ und $\hat{\Psi}_0$ folgt $\boxed{\frac{\dot{\lambda}_p(\beta_l)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^N \eta_i}{N\hat{\Psi}_0} \to_d \mathbb{N}\left(0, \frac{\kappa_d}{\omega}\right)}$. Da zudem gilt $\frac{1}{n}\sum_{i=1}^N \delta_i^{\hat{\beta}_0}(n\pi_i^p(\hat{\beta}_0) \vee 1)\phi_i(\beta_l - \hat{\beta}_0 + \frac{\acute{s}}{\sqrt{n}})x_i x_i^T = \frac{\kappa_a}{\omega} + o_P(1)$ (siehe Wang (2019)) folgt aus dem B.C.

$$\sqrt{n}(\hat{\beta}_p - \beta_l) = \omega\kappa_a^{-1}\frac{\dot{\lambda}_p(\beta_l)}{\sqrt{n}} + o_P(1).$$

Zusammen mit $\sqrt{N}(\hat{\beta}_{\mathrm{wMLE}} - \beta_l) = \kappa_a^{-1}\frac{\sum_{i=1}^N \eta_i}{\sqrt{N}} + o_P(1)$ aus dem Bew. von Theorem 3 folgt das Theorem. $\square$

Bemerkungen

Die in Theorem 3 und Theorem 4 gezeigten Konvergenzen sind

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \omega\kappa_a^{-1})$$

und

$$\sqrt{n}(\hat{\beta}_p - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \omega\kappa_a^{-1}\kappa_d\kappa_a^{-1}).$$

Es gilt (wie in Proposition 2)

$$\kappa_a^{-1}\kappa_d\kappa_a^{-1} < \kappa_a^{-1}$$

und damit wie in Theorem 1 und 2, dass der Schätzer basierend auf Poisson-Sampling effizienter ist als der erstgenannte, respektive die bedingte Kovarianzmatrix ist kleiner.

Misspecifications II

Die pilot Estimators konvergieren gegen Parameter, die nicht die wahren sind.

Theorem 5

Sei das logistische Regressionsmodell korrekt spezifiziert und der pilot Estimator $\hat{\beta}_0$ (der unabhängig von $\mathcal{D}_N$ ist) inkonsistent, d.h. $\hat{\beta}_0 = \beta_0 + o_P(1), \; \beta_t \neq \beta_0$. Dann folgt unter Annahme 1-3 bedingt auf $\mathcal{D}_N$ für $n, N \to \infty$

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \Psi(\beta_0)\varsigma_a^{-1}).$$

Ist $n/N \to 0$, dann

$$\sqrt{n}(\hat{\beta}_{\text{uw}} - \beta_t) \to_d \mathbb{N}(0, \Psi(\beta_0)\varsigma_a^{-1}),$$

wobei hier $\hat{\beta}_{\text{wMLE}}$

$$\sqrt{N}(\hat{\beta}_{\text{wMLE}} - \beta_t) \to_d \mathbb{N}(0, \varsigma_a^{-1}\varsigma_b\varsigma_a^{-1})$$

erfüllt.

Bemerkungen

Ist der pilot Estimator 'sehr falsch', so, dass $\beta_t^T x x^T \beta_0 < 0$ (äquivalent ist $[p(x, \beta_t) > 0.5 > p(x, \beta_0)$ oder $p(x, \beta_t) < 0.5 < p(x, \beta_0)]$), dann gilt

$$\Psi(\beta_0)\varsigma_a^{-1} > \Sigma_{\beta_t}$$

und der Schätzer ist effizienter im konsistenten Fall.

Theorem 6

Sei das logistische Regressionsmodell korrekt spezifiziert und die pilot Estimators inkonsistent, d.h. $\hat{\beta}_0 = \beta_0 + o_P(1)$, $\beta_t \neq \beta_0$ und $\hat{\Psi}_0 = \Psi_0 + o_P(1)$, $\Psi(\beta_t) \neq \Psi_0$. Es folgt unter Annahme 1 und 2, bedingt auf $\mathcal{D}_N$, dass für $n, N \to \infty$, mit $n/N \to 0$ gilt

$$\sqrt{n}(\hat{\beta}_p - \beta_t) \to_d \mathbb{N}(0, \Psi_0 \varsigma_a^{-1}).$$

Ist $n/N \to \rho$, dann

$$\sqrt{n}(\hat{\beta}_p - \hat{\beta}_{\text{wMLE}}) \to_d \mathbb{N}(0, \Psi_0 \varsigma_a^{-1} \varsigma_c \varsigma_a^{-1}).$$

Bemerkungen

Es gilt

$$\Psi_0 \varsigma_a^{-1} \varsigma_c \varsigma_a^{-1} < \Psi_0 \varsigma_a^{-1}$$

und damit, falls $\Psi_0 = \Psi(\beta_0)$, dass der Schätzer basierend auf Poisson-Subsampling ($\hat{\beta}_p$) effizienter ist als $\hat{\beta}_{\text{uw}}$ (z.m.Z.).

(Sehr) interessant ist, die Schätzer sind (in beiden Fällen (!)) trotzdem konsistent.

Zusammenfassung

Basierend auf den optimal subsampling probabilities haben wir für das logistische Regressionsmodell einen ungewichteten und zugleich mit Hilfe eines pilot Estimators auch unverzerrten Schätzer ($\hat{\beta}_{\text{uw}}$) entwickelt dessen Kovarianzmatrix asymptotisch kleiner ist als die des gewichteten Schätzers aus Wang (2018).
Zusätzlich konnten wir zeigen, dass der auf Poisson-Subsampling basierende Schätzer $\hat{\beta}_{\boldsymbol{p}}$, der ebenso mithilfe von pilot Estimators konstruiert wurde konsistent, sowie asymptotisch normalverteilt ist. Nebst diesen Eigenschaften tut sich dieser durch eine tatsächlich in allen Fällen mindestens ebenso hohe Effizienz hervor.

Dieses Charakteristikum und auch die asymptotische Normalität bleiben (tatsächlich) unter Model- und Pilot-Misspecification (unter weiteren Bedingungen) erhalten.

Literatur

HaiYing Wang. „More Efficient Estimation for Logistic Regression with Optimal Subsamples“. In: *J. Mach. Learn. Res.* 20 (2019), 132:1-132:59. URL: `http://jmlr.org/papers/v20/18-596.html`

… und viele weitere Quellen in der Ausarbeitung

Vielen Dank für eure Aufmerksamkeit!

Fragen?