More Efficient Estimation for Logistic Regression with Optimal Subsamples

Nils Kornacker

Diese Seminarausarbeitung dient dazu den Zuhörern eine Übersicht über die wichtigsten Lemmata und Theoreme aus [Wan19] sowie insbesondere alle dort nicht explizit genannten aber nützlichen Lemmata zu geben. Zusätzlich wurde eine Symbolübersicht erstellt.

Inhaltsverzeichnis

1 Vorbereitende Lemmata **2**

2 Lemmata und Theoreme aus [Wan19] **3**

3 Beweis von Lemma 4 **7**

4 Symbolübersicht **9**

Literatur **11**

1 Vorbereitende Lemmata

Im Folgenden seien $(X_n)_{n\in\mathbb{N}}$ und X $\mathbb{R}^k$-wertige Zufallsvariablen.

v.Definition 1. *($o_P(r_n), O_P(r_n)$, S.12 in [van98])*

$$\begin{aligned} X_n = o_P(1) \quad &\Leftrightarrow \quad X_n \to_P 0 \\ X_n = O_P(1) \quad &\Leftrightarrow \quad \forall \varepsilon > 0\ \exists M:\ \sup_n \mathbf{P}(\|X_n\| > M) < \varepsilon \\ X_n = o_P(r_n) \quad &\Leftrightarrow \quad X_n = r_n o_P(1) \\ X_n = O_P(r_n) \quad &\Leftrightarrow \quad X_n = r_n O_P(1) \end{aligned}$$

v.Definition 2. *(S. 7 in [Wan19])*
$X_n = O_{P|\mathcal{D}_N}(a_n) \Leftrightarrow$ *Für alle $\varepsilon > 0$ gibt es ein δ_ε, so dass für $n, N \to \infty$*

$$\mathbf{P}(\sup_n \mathbf{P}\left(\|X_n\| > a_n\delta_\varepsilon|\mathcal{D}_N\right) \le \varepsilon) \to 1.$$

v.Lemma 1. *(Continuous Mapping, 2.3 in [van98])*
Sei $g : \mathbb{R}^k \mapsto \mathbb{R}^m$ stetig in jedem Punkt von einer Menge C so, dass $\mathbf{P}(X \in C) = 1$. Es gilt:

$$X_n = X + o_P(1) \quad \Rightarrow \quad g(X_n) = g(X) + o_P(1).$$

v.Lemma 2. *(Prohorov, 2.4 in [van98])*
Wenn $X_n \rightsquigarrow X$, dann ist $X_n = O_P(1)$. $\rightsquigarrow$ meint Konvergenz in Verteilung.

v.Lemma 3. *(2.7 in [van98])*
Sei c eine Konstante: $X_n = c + o_P(1) \quad \Leftrightarrow \quad X_n \rightsquigarrow c.$

v.Lemma 4. *(Slutsky, 2.8 in [van98])*
Seien $X_n \rightsquigarrow X$ und $Y_n \rightsquigarrow c$, c eine Konstante, c, X_n, Y_n, X konform Matrixwertig (dies schließt Vektor- und Skalarwertig ein).

$$\begin{aligned} &X_n + Y_n \rightsquigarrow X + c \\ &Y_n X_n \rightsquigarrow cX \\ &Y_n^{-1} X_n \rightsquigarrow c^{-1}X, \ \textit{gegeben } c \textit{ invertierbar.} \end{aligned}$$

v.Lemma 5. *(Lemma 2.1 in [Mus88])*
Sei $C \subset \mathbb{R}^k$ eine konvexe Borelmenge und g eine messbare konvexe Funktion von C nach $(-\infty, \infty]$ so, dass $C' = \{x \in C : g(x) < \infty\} \neq \emptyset$. Für $i = 1, \ldots, k$ sei X_i eine reellwertige messbare Funktion auf Ω ($(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum) mit $\mathbf{E}[|X_i|] < \infty$. Ist $X(\omega) \in C$ für alle $\omega \in \Omega$, wobei $X = (X_1, \ldots, X_k)$, so gilt

$$g(\mathbf{E}[X|\mathcal{F}]) \le \mathbf{E}[g(X)|\mathcal{F}] \textit{ f.s. für alle } \mathcal{F} = \sigma(\mathcal{F}) \subset \mathcal{A}.$$

v.Lemma 6. *(Theorem 1.3.6 in [Ser02])*
Seien $X_n = X + o_P(1)$, $\|X_n\| \le \|Y\|$ f.s. $\forall n$ und $\mathbf{E}[\|Y\|^r] < \infty$. Dann ist $X_n \to_{\mathcal{L}^r} X$.

v.Theorem 1. *(Lindeberg-Feller Theorem 2.27 in [van98])*
Seien für $n \in \mathbb{N}$ $X_{n,1}, \ldots, X_{n,k_n}$ unabhängige (zufällige) Vektoren mit endlichen Varianzen, sodass

$$\sum_{i=1}^{k_n} \mathbf{E}[\|X_{n,i}\|^2 I(\|X_{n,i}\| > \varepsilon)] \to 0 \qquad \forall \varepsilon > 0,$$

$$\sum_{i=1}^{k_n} \mathbf{Cov}[X_{n,i}] \to \Sigma.$$

Dann gilt $\sum_{i=1}^{k_n}(X_{n,i} - \mathbf{E}[X_{n,i}]) \rightsquigarrow \mathbb{N}(0, \Sigma)$.

2 Lemmata und Theoreme aus [Wan19]

Annahmen. $\mathbf{E}[\phi(\beta_t)h(x)xx^T]$ *ist endlich und positiv-definit.* ***(1)***

$\mathbf{E}[\|x\|^2 h^2(x)] < \infty,\ \mathbf{E}[\|x\|^2 h(x)] < \infty.$ ***(2)***

$n\mathbf{E}[h(x)I(\|x\|^2 > n)] \overset{n\to\infty}{\longrightarrow} 0.$ ***(3)***

Zusätzlich: $h(x) \overset{f.s.}{>} 0$ *integrierbar (Theorem 1,3,5, Lemma 3,4).*

Lemma 1. *(Lemma 28 in [Wan19])*
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*

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

Lemma 2. *(Lemma 29 in [Wan19])*
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*

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

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.*

Bemerkung. *Lemma 3 & 5 und Theorem 1 & 2 benötigen vermutlich die Voraussetzung* $n_0/\sqrt{N} \to 0$ *oder* $\hat\beta_0$ *unabhängig von* $\mathcal{D}_N$ *ebenso.*

Basic Corollary. *(Basic Corollary in [HP11])*
Sei $A_n(s) = \frac{1}{2}s^TVs + U_n^Ts + C_n + R_n(s)$ *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* $\frac{1}{2}s^TVs + U_n^Ts + C_n$*. Wenn zusätzlich* $U_n \rightsquigarrow U$*, dann* $\alpha_n \rightsquigarrow -V^{-1}U$*.*

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

Lemma 3. *(Lemma 30 in [Wan19])*
Sei

$$\dot\lambda^*_{uw}(\beta_t) = \sum_{i=1}^{n}(y_i^* - p_i^*(\beta_t - \hat\beta_0))x_i^*.$$

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*

$$\frac{\dot\lambda^*_{uw}(\beta_t)}{\sqrt{n}} - \frac{\sqrt{n}\sum_{i=1}^{N}\eta_i}{N\Psi_N(\hat\beta_0)} \rightsquigarrow \mathbb{N}(0, \Sigma_{\beta_t}^{-1}),$$

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

Lemma 4. *(Lemma 31 in [Wan19])*
Unter Annahmen 1-3, wenn n_0, n und N gegen unendlich streben und $s_n = o_P(1)$

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

Ebenso gilt

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

Für den Beweis siehe hier in diesem Blatt S.7.

Theorem 1. *(Theorem 1 in [Wan19])*
Unter Annahmen 1-3, bedingt auf $\mathcal{D}_N$, wenn $\hat{\beta}_0$ konsistent, $n_0, n, N \to \infty$

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

Wenn $n/N \to 0$, dann

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

$\hat{\beta}_{wMLE} = \arg\max_{\beta} \sum_{i=1}^{N} |y_i - p(x_i, \hat{\beta}_0)| h(x_i)[y_i x_i^T(\beta - \hat{\beta}_0) - \log(1 + e^{x_i^T(\beta - \hat{\beta}_0)})]$ *ist 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 Stichprobe der Größe n_0 gesamplet wurde mit $n_0/\sqrt{N} = o(1)$ oder $\hat{\beta}_0$ unabhängig von $\mathcal{D}_N$, dann gilt

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

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

Proposition 1. *(Proposition 4 in [Wan19])*
Wenn M, V_c^{OS} und Σ_{β_t} endliche und positiv definite Matrizen sind, dann gilt

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

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

Bemerkung. $A \leq B \Rightarrow C^T AC \leq C^T BC$

Lemma 5. *(Lemma 32 in [Wan19])*
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 und bedingt auf $\mathcal{D}_N$ sowie die konsistenten Schätzer $\hat{\beta}_0$ und $\hat{\Psi}_0$ gilt, wenn $n/N \to 0$, dass

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

Konvergiert n/N gegen $\rho \in (0,1)$, dann

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

Lemma 6. *(Lemma 33 in[Wan19])*
Unter Annahmen 1 und 2, wenn $n_0, n, N \to \infty$, und $s_n = o_P(1)$

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

Theorem 2. *(Theorem 6 in [Wan19])*
Es gelten Annahmen 1 und 2, und $\hat{\beta}_0$ sei konsistenter Schätzer für β_t. Dann ist bedingt auf $\mathcal{D}_N$ für n_0, n und $N \to \infty$ mit $n/N \to 0$

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

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

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

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

Proposition 2. *(Proposition 8 in [Wan19])*
Ist $\rho > 0$ und Σ_{β_t} eine endliche und positiv definite Matrix, dann gilt

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

Die folgenden Theoreme 3 und 4 unterstellen jeweils, dass das logistische Regressionsmodell missspezifiziert ist und die wahre, den Daten unterliegende Verteilung ist:

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

$p_t(x)$ unbekannt.

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

Theorem 3. *(Theorem 24 in [Wan19])*
Sei das pilot Sample unabhängig von $\mathcal{D}_N$, $\sqrt{n_0}(\hat{\beta}_0 - \beta_l) \rightsquigarrow \mathbb{N}(0, \Sigma_0)$ 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 $\mathcal{D}_N$

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

wobei

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

$\hat{\beta}_{wMLE}$ erfüllt

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

wobei

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

Bemerkung. *Wir benötigen in Theorem 3 vermutlich zusätzlich κ_a positiv definit.*

Theorem 4. *(Theorem 26 in [Wan19])*
Angenommen das pilot Sample ist unabhängig von $\mathcal{D}_N$ und die pilot Estimators erfüllen $\sqrt{n_0}(\hat{\beta}_0 - \beta_l) \rightsquigarrow \mathbb{N}(0, \Sigma_0)$ 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 $n_0, n, N \to \infty$

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

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

Theorem 5. *(Theorem 18 in [Wan19])*
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 $n, N \to \infty$

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

Ist $n/N \to 0$, dann

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

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

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

erfüllt.

Theorem 6. *(Theorem 21 in [Wan19])*
Sei das logistische Regressionsmodell korrekt spezifiziert und die pilot Estimators unabhängig von $\mathcal{D}_N$ und 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

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

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

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

3 Beweis von Lemma 4

Wir beginnen mit der folgenden Zerlegung,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\phi_i^*(\beta_t-\hat\beta_0+s_n)x_i^*x_i^{*T}\\
=&\frac{1}{n}\sum_{i=1}^{n}\phi_i^*(\beta_t-\hat\beta_0+s_n)x_i^*x_i^{*T}I(\|x_i^*\|^2\le n)+\frac{1}{n}\sum_{i=1}^{n}\phi_i^*(\beta_t-\hat\beta_0+s_n)x_i^*x_i^{*T}I(\|x_i^*\|^2> n)\\
&\equiv \Delta_1+\Delta_2
\end{aligned}$$

Δ_2 ist $o_P(1) \Leftrightarrow \|\Delta_2\|$ ist $o_P(1)$. Es gilt jedoch $\|\Delta_2\| \le \frac{1}{n}\sum_{i=1}^{n}\phi_i^*(\beta_t-\hat\beta_0+s_n)\|x_i^*\|^2 I(\|x_i^*\|^2 > n) = o_P(1)$ (siehe Seite 36 in [Wan19]) und damit $\|\Delta_2\| = o_P(1)$. Bemerke insbesondere $\|xx^T\| = \sqrt{\sum_{i,j=1}^{d}|x_ix_j|^2} = \sum_{i=1}^{d}x_i^2 = \|x\|^2$.
Weiter gilt

$$\begin{aligned}
&\left\|\mathbf{E}[\Delta_1|\mathcal{D}_N,\hat\beta_0]-\sum_{i=1}^{N}\pi_i(\hat\beta_0)\phi_i(\beta_t-\hat\beta_0)x_ix_i^T\right\|\\
&=\left\|\sum_{i=1}^{N}\pi_i(\hat\beta_0)\phi_i(\beta_t-\hat\beta_0+s_n)x_ix_i^T I(\|x_i\|^2\le n)-\sum_{i=1}^{N}\pi_i(\hat\beta_0)\phi_i(\beta_t-\hat\beta_0)x_ix_i^T\right\|\\
&=\left\|\sum_{i=1}^{N}\pi_i(\hat\beta_0)x_ix_i^T(\phi_i(\beta_t-\hat\beta_0+s_n)-\phi_i(\beta_t-\hat\beta_0))-\sum_{i=1}^{N}\pi_i(\hat\beta_0)\phi_i(\beta_t-\hat\beta_0+s_n)x_ix_i^T I(\|x_i\|^2> n)\right\|\\
&\le\sum_{i=1}^{N}\pi_i(\hat\beta_0)\|x_i\|^2 o_P(1)+\sum_{i=1}^{N}\pi_i(\hat\beta_0)\phi_i(\beta_t-\hat\beta_0+s_n)\|x_i\|^2 I(\|x_i\|^2>n)\\
&\overset{\star}{\le}\frac{\frac{1}{N}\sum_{i=1}^{N}|\psi_i(\hat\beta_0)|h(x_i)\|x_i\|^2 o_P(1)}{\Psi_N(\hat\beta_0)}+o_P(1)\\
&\le\frac{\frac{1}{N}\sum_{i=1}^{N}h(x_i)\|x_i\|^2 o_P(1)}{\Psi_N(\hat\beta_0)}+o_P(1)\\
&=o_P(1),
\end{aligned}$$

wobei für den letzten Schritt Lemma 1 mit $\phi_i(\beta_t-\hat\beta_0+s_n)-\phi_i(\beta_t-\hat\beta_0) = o_P(1)$ und betragsmäßig durch 1 beschränkt, sowie das schwache Gesetz der großen Zahlen genutzt wurde. $\star$ siehe S.36 in [Wan19].

Damit müssen wir nur noch zeigen, dass $\Delta_1-\mathbf{E}[\Delta_1|\mathcal{D}_N,\hat\beta_0] = o_P(1) \;\Leftrightarrow\; \|\Delta_1-\mathbf{E}[\Delta_1|\mathcal{D}_N,\hat\beta_0]\| = o_P(1)$. Hierzu folgender Trick:

$$\begin{aligned}
\|\Delta_1-\mathbf{E}[\Delta_1|\mathcal{D}_N,\hat\beta_0]\| &= \sqrt{\sum_{k,j=1}^{d}|\Delta_{1kj}-\mathbf{E}[\Delta_{1kj}|\mathcal{D}_N,\hat\beta_0]|^2} \le \sqrt{\left(\sum_{k,j=1}^{d}|\Delta_{1kj}-\mathbf{E}[\Delta_{1kj}|\mathcal{D}_N,\hat\beta_0]|\right)^2}\\
&=\sum_{k,j=1}^{d}|\Delta_{1kj}-\mathbf{E}[\Delta_{1kj}|\mathcal{D}_N,\hat\beta_0]|
\end{aligned}$$

und es reicht zu zeigen, dass $\Delta_{1kj}-\mathbf{E}[\Delta_{1kj}|\mathcal{D}_N,\hat\beta_0] = o_P(1)\ \forall\ k,j$. Dafür zeigen wir, dass die bedingte Varianz $\mathbf{V}[\Delta_{1kj}|\mathcal{D}_N,\hat\beta_0] = \mathbf{E}[(\Delta_{1kj}-\mathbf{E}[\Delta_{1kj}|\mathcal{D}_N,\hat\beta_0])^2|\mathcal{D}_N,\hat\beta_0]$ stochastisch gegen 0 konvergiert.

$$\begin{aligned}
\mathbf{V}[\Delta_{1kj}|\mathcal{D}_N, \hat{\beta}_0] &= \mathbf{V}\left[\frac{1}{n}\sum_{i=1}^{n} \phi_i^*(\beta_t - \hat{\beta}_0 + s_n)(x_i^*)_k (x_i^*)_j I(\|x_i^*\|^2 \leq n)|\mathcal{D}_N, \hat{\beta}_0\right] \\
&\overset{iid}{=} \frac{1}{n}\mathbf{V}\left[\phi^*(\beta_t - \hat{\beta}_0 + s_n)(x^*)_k (x^*)_j I(\|x^*\|^2 \leq n)|\mathcal{D}_N, \hat{\beta}_0\right] \\
&\leq \frac{1}{n}\mathbf{E}\left[\phi^*(\beta_t - \hat{\beta}_0 + s_n)^2((x^*)_k (x^*)_j)^2 I(\|x^*\|^2 \leq n)|\mathcal{D}_N, \hat{\beta}_0\right] \\
&\leq \frac{1}{16n}\mathbf{E}\left[\sum_{k,j=1}^{d}((x^*)_k (x^*)_j)^2 I(\|x^*\|^2 \leq n)|\mathcal{D}_N, \hat{\beta}_0\right] \\
&= \frac{1}{16n}\mathbf{E}\left[\|x^*\|^4 I(\|x^*\|^2 \leq n)|\mathcal{D}_N, \hat{\beta}_0\right] \\
&\vdots \qquad \text{(siehe S.36, [Wan19])} \\
&\leq \frac{1}{16n}\left(1 + \sum_{i=1}^{n-1} 3i\mathbf{P}(\|x^*\|^2 > i|\mathcal{D}_N, \hat{\beta}_0)\right) = o_P(1),
\end{aligned}$$

wobei die letzte Gleichheit aus [Wan19], S.37 folgt. Bemerke, dass eigentlich richtig ist $\frac{1}{n}\sum_{i=1}^{n} i\mathbf{P}(\|x^*\|^2 > i|\mathcal{D}_N, \hat{\beta}_0) = \frac{1}{n}\sum_{i=1}^{n} i\sum_{j=1}^{N}\pi_j(\hat{\beta}_0)I(\|x_j\|^2 > i)$.

4 Symbolübersicht

$\beta_t =$ wahrer Parametervektor (das " = " ist als 'ist gleich' zu lesen, im Folgenden nicht mehr angemerkt)

$\hat{\beta}_{\text{MLE}} = \arg\max_{\beta} \sum_{i=1}^{N} \{y_i x_i^T \beta - \log(1 + e^{\beta^T x_i})\}$

$\hat{\beta}_w^{\pi} = \arg\max_{\beta} \sum_{i=1}^{n} \frac{y_i^* \beta^T x_i^* - \log(1 + e^{\beta^T x_i^*})}{\pi_i^*}$

$\hat{\beta}_w = \arg\max_{\beta} \sum_{i=1}^{n} \frac{y_i^* \beta^T x_i^* - \log(1 + e^{\beta^T x_i^*})}{\pi_i^{\text{OS}}(\hat{\beta}_0)^*}$

$\hat{\beta}_0 =$ (konsistenter) pilot Estimator von β_t, Achtung! bei Theorem 3-6 siehe β_l und β_0

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

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

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

$\tilde{\beta}_p = \arg\max_{\beta} \sum_{i=1}^{n^*} (n\pi_i^{p*}(\hat{\beta}_0) \vee 1)(\beta^T x_i^* y_i^* + \log(1 + e^{\beta^T x_i^*}))$

$\hat{\beta}_p = \tilde{\beta}_p + \hat{\beta}_0$

$\beta_l =$ Minimierer von $\mathbf{E}[-p_t(x)h(x)x^T\beta + h(x)\log(1 + e^{\beta^T x})]$, $\hat{\beta} - \beta_l = o_P(1)$ (Theorem 3 und 4)

$\beta_0 :\ \beta_0 = \hat{\beta}_0 + o_P(1)$ (Theorem 5 und 6)

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

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

$\eta_i = |\psi_i(\hat{\beta}_0)| \psi_i(\beta_t - \hat{\beta}_0) h(x_i) x_i$

$\eta_i = |\psi_i(\hat{\beta}_0)| \psi_i(\beta_l - \hat{\beta}_0) h(x_i) x_i$ (Theorem 3 und 4)

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

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

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

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

$\dot{\lambda}_{uw}^*(\beta_t) = \sum_{i=1}^{n} (y_i^* - p_i^*(\beta_t - \hat{\beta}_0)) x_i^*$

$\lambda_p(\beta) = \sum_{i=1}^{N} \delta_i^{\hat{\beta}_0} (n\pi_i^p(\hat{\beta}_0) \vee 1)((\beta - \hat{\beta}_0)^T x_i y_i - \log(1 + e^{(\beta - \hat{\beta}_0)^T x_i}))$

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

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

$\Lambda_u = \dfrac{\mathbf{E}[|\psi(\beta_t)|(\rho|\psi(\beta_t)|h(x) \vee \Psi(\beta_t))h(x)xx^T]}{4\Psi^2(\beta_t)}$

$M_N = \dfrac{1}{N}\sum_{i=1}^{N} \phi_i(\hat{\beta}_{\mathrm{MLE}})x_i x_i^T$

$M = \mathbf{E}[\phi(\beta_t)xx^T]$

n_0: Größe des Pilot-Subsamples

n: Größe des Second-Stage-Subsamples (Ziehen mit Zurücklegen)

n^*: Größe des Second-Stage-Subsamples (Poisson Subsampling) (zufällig)

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

$p_t(x) = \mathbf{P}(y = 1|x)$ (Theorem 3 und 4)

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

$\pi_i^p(\hat{\beta}_0) = \dfrac{|y_i - p(x_i, \hat{\beta}_0)|h(x_i)}{N\hat{\Psi}_0}$ oder auch einfach nur π_i^p

$\rho = \lim_{n,N\to\infty} n/N$ (an manchen Stellen)

$\rho_0 =$ Grenzwert von n_0/N (an manchen Stellen)

$\varsigma_a = \mathbf{E}[(1 - p(x, \beta_t))p(x, \beta_0)p(x, \beta_t - \beta_0)h(x)xx^T]$

$\varsigma_b = \mathbf{E}[\phi(\beta_0)\phi(\beta_t - \beta_0)h^2(x)xx^T]$

$\varsigma_c = \mathbf{E}\left[|\psi(\beta_0)|(1 - \rho\Psi_0^{-1}|\psi(\beta_0)|h(x))_+ \psi^2(\beta_t - \beta_0)h(x)xx^T\right]$

$\Sigma_\beta = \left[\dfrac{\mathbf{E}[\phi(\beta)h(x)xx^T]}{4\Phi(\beta)}\right]^{-1}$

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

$\Sigma_0 :$ asymptotische Kovarianzmatrix von $\sqrt{n_0}(\hat{\beta}_0 - \beta_l)$

$V_N^{\mathrm{OS}} = M_N^{-1} V_{\mathrm{Nc}}^{\mathrm{OS}} M_N^{-1}$

$V_{\mathrm{Nc}}^{\mathrm{OS}} = \left(\dfrac{1}{N}\sum_{i=1}^{N} |y_i - p(x_i, \hat{\beta}_{\mathrm{MLE}})|h(x_i)\right)\left(\dfrac{1}{N}\sum_{i=1}^{N} \dfrac{|y_i - p(x_i, \hat{\beta}_{\mathrm{MLE}})|x_i x_i^T}{h(x_i)}\right)$

$V^{\mathrm{OS}} = M^{-1} V_c^{\mathrm{OS}} M^{-1}$

$V_c^{\mathrm{OS}} = 4\Phi(\beta_t)\mathbf{E}\left[\dfrac{\phi(\beta_t)xx^T}{h(x)}\right]$

$\phi(\beta) = p(x, \beta)(1 - p(x, \beta)),\ \phi_i(\beta) = p(x_i, \beta)(1 - p(x_i, \beta))$

$\Phi(\beta) = \mathbf{E}[\phi(\beta)h(x)]$

$\psi(\beta) = y - p(x,\beta),\ \psi_i(\beta) = y_i - p(x_i,\beta)$

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

$\Psi(\beta) = \mathbf{E}[|y - p(x,\beta)| h(x)]$

$\hat{\Psi}_0 :$ pilot Estimator von $\Psi(\beta_t)$

$\Psi_0 :\ \Psi_0 = \hat{\Psi}_0 + o_P(1)$ (Theorem 5 und 6)

$\omega = \mathbf{E}[(p_t(x) - 2p_t(x)p(x,\beta_l) + p(x,\beta_l))h(x)]$

Relative Efficiency $= \dfrac{MSE(\check{\beta}_w)}{MSE(\check{\beta}_{new})}$

$MSE(\check{\beta}) = \dfrac{1}{S}\sum_{s=1}^{S} \|\check{\beta}^{(s)} - \beta_t\|^2$

Literatur

[CH10] Guang Cheng und Jianhua Z. Huang. “Bootstrap consistency for general semiparametric M-estimation”. English. In: *The Annals of Statistics* 38.5 (2010), S. 2884–2915. ISSN: 0090-5364; 2168-8966/e.

[FH14] William Fithian und Trevor Hastie. “Local case-control sampling: efficient subsampling in imbalanced data sets”. English. In: *The Annals of Statistics* 42.5 (2014), S. 1693–1724. ISSN: 0090-5364; 2168-8966/e.

[HP11] Nils Lid Hjort und David Pollard. “Asymptotics for minimisers of convex processes”. In: (19. Juli 2011). arXiv: `1107.3806v1 [math.ST]`.

[Mus88] Dieter Mussmann. “Sufficiency and Jensen’s inequality for conditional expectations”. English. In: *Annals of the Institute of Statistical Mathematics* 40.4 (1988), S. 715–726. ISSN: 0020-3157; 1572-9052/e.

[Ser02] Robert J. Serfling. *Approximation theorems of mathematical statistics*. English. New York, NY: Wiley, 2002, S. 371. ISBN: 0-471-21927-4/pbk.

[van98] A. W. van der Vaart. *Asymptotic statistics*. Bd. 3. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998, S. xvi+443. ISBN: 0-521-49603-9; 0-521-78450-6. DOI: `10.1017/CBO9780511802256`.

[Wan19] 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`.

[WZM18] HaiYing Wang, Rong Zhu und Ping Ma. “Optimal subsampling for large sample logistic regression”. English. In: *Journal of the American Statistical Association* 113.522 (2018), S. 829–844. ISSN: 0162-1459; 1537-274X/e.

[XL08] Shifeng Xiong und Guoying Li. “Some results on the convergence of conditional distributions”. English. In: *Statistics & Probability Letters* 78.18 (2008), S. 3249–3253. ISSN: 0167-7152.

Die Quellen [CH10], [FH14], [WZM18], [XL08] als zusätzliche Vortragsliteratur.