Logistic Regression for Data w/ Rare Events

Ideen und Beweise basierend auf Paper von Wang, 2020 (arXiv:2006.00683)

Statistical Learning for Imbalanced Datasets Seminar

Jean Chammas

08.06.2021

- Haben einen Datensatz $\mathcal{D}_n = \{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$ mit n Datenpunkten, davon n_1 cases und n_0 controls.
- Es ist $n_1 << n_0$.
- Wir wollen eine Logistische Regression durchführen. Gängige Ansätze in der Praxis: Over- und Undersampling. Auswirkung der Ansätze auf den Schätzer im unbalancierten Fall ist aber noch unbekannt.
- $\Rightarrow$ Das wird nun untersucht! (Müssen aber noch die mathematische Modellierung präzisieren).

Modellierung Unbalancierter Daten

Annahmen:

- $n_1/n_0 \to 0$ bzw. $n_1/n \to 0$ für $n \to \infty$.
- Es soll trotzdem $n_1 \to \infty$ für $n \to \infty$ gelten.
- Es folgt also

$$\mathbb{P}(y = 1) \to 0 \quad \text{und } n\mathbb{P}(y = 1) \to \infty \tag{1}$$

Was heißt das für die Regressionsgerade?

- Es seien α_t, β_t die wahren geraden Parameter $(\alpha_t + \beta_t^T x)$. Es gelte $\beta_t = \text{const.}, \forall n \in N$ und $\alpha_t \to -\infty$ mit einer passenden Rate, so dass gilt:

$$\frac{n_1}{n} = \mathbb{P}(y = 1)\{1 + o_p(1)\} = \mathbb{E}\left[\frac{\exp(\alpha_t + \beta_t x)}{1 + \exp(\alpha_t + \beta_t x)}\right]\{1 + o_p(1)\} \quad (2)$$

Modellierung Standardfall

- Maximum-Likelihood-Ansatz: wir maximieren die Log-Likelihood Funktion

$$l(\theta) = \sum_{i=1}^{n} (y_i z_i^T \theta - \log(1 + \exp(z_i^T \theta))) \tag{3}$$

- Wir bezeichnen mit $\hat{\theta}$ den maximierer von (3)

Modellierung Subsampling

- Wollen alle cases im Datensatz beibehalten und einige controls weglassen
- Definiere die Auswahlfunktion für den i-ten Datenpunkt:

$$\delta_i = y_i + (1 - y_i) I_{u_i < \pi_0}$$

wobei $u_i \sim \mathbb{U}[0, 1]$ i.i.d. sind und $\pi_0 \in [0, 1]$ gewählt werden kann.

Modellierung Subsampling

- Maximum-Likelihood-Ansatz: wir maximieren die Log-Likelihood Funktion
$$l(\theta) = \sum_{i=1}^{n} \frac{\delta_i}{\pi_i}(y_i z_i^T \theta - \log(1 + \exp(z_i^T \theta))) \tag{4}$$
wobei $\pi_i = \mathbb{E}\left[\delta_i | \mathcal{D}_n\right] = y_i + (1 - y_i)\pi_0 = \pi_0 + (1 - \pi_0)y_i$.
- Wir bezeichnen mit $\hat{\theta}^w_{under}$ den Maximierer von (4)

Modellierung Oversampling

- Wollen alle controls im Datensatz beibehalten und einige cases mehrfach zählen.
- Definiere die Auswahlfunktion für den i-ten Datenpunkt:

$$\tau_i = 1 + y_i v_i$$

wobei $v_i \sim \mathbb{POI}(\lambda_n)$ zum Parameter λ_n i.i.d. sind.

Modellierung Oversampling

- Maximum-Likelihood-Ansatz: wir maximieren die Log-Likelihood Funktion
$$l(\theta) = \sum_{i=1}^{n} \frac{\tau_i}{w_i}(y_i z_i^T \theta - \log(1 + \exp(z_i^T \theta))) \tag{5}$$
wobei $w_i = \mathbb{E}[\tau_i | \mathcal{D}_n, y_i = 1] = 1 + \lambda_n$.
- Wir bezeichnen mit $\hat{\theta}^w_{over}$ den Maximierer von (5)

Theorem 1

Es sei $\mathbb{E}\left[\exp(t\,||x||)\right] < \infty$ für alle $t > 0$, $\mathbb{E}\left[\exp(\beta_t^T x) z z^T\right]$ sei positiv definit. Es gelten weiter die Annahmen aus (1) und (2). Dann gilt falls $n \to \infty$:

$$\sqrt{n_1}(\hat{\theta} - \theta_t) \to_{\mathcal{D}} \mathcal{N}\left(0, V_f\right) \tag{6}$$

mit

$$V_f = \mathbb{E}\left[\exp(\beta_t^T x)\right] M_f^{-1}$$
$$M_f = \mathbb{E}\left[\exp(\beta_t^T x) z z^T\right]$$

Theorem 2 (Undersampling)

Es sei $\mathbb{E}\left[\exp(t\,||x||)\right] < \infty$ für alle $t > 0$, $\mathbb{E}\left[\exp(\beta_t^T x) z z^T\right]$ sei positiv definit. Es gelte $c_n = \exp(\alpha_t)/\pi_0 \to c \in [0, \infty)$. Es gelten weiter die Annahmen aus (1) und (2). Dann gilt falls $n \to \infty$:

$$\sqrt{n_1}(\hat{\theta}^w_{under} - \theta_t) \to_{\mathcal{D}} \mathcal{N}\left(0, V^w_{under}\right) \tag{7}$$

mit

$$V^w_{under} = \mathbb{E}\left[\exp(\beta_t^T x)\right] M_f^{-1} M^w_{under} M_f^{-1}$$
$$M^w_{under} = \mathbb{E}\left[\exp(\beta_t^T x)(1 + c\exp(\beta_t^T x)) z z^T\right]$$

Theorem 3 (Oversampling)

Es sei $\mathbb{E}\left[\exp(t\,||x||)\right] < \infty$ für alle $t > 0$, $\mathbb{E}\left[\exp(\beta_t^T x) z z^T\right]$ sei positiv definit. Es gelte $\lambda_n \to \lambda \geq 0$. Es gelten weiter die Annahmen aus (1) und (2). Dann gilt falls $n \to \infty$:

$$\sqrt{n_1}(\hat{\theta}^w_{over} - \theta_t) \to_{\mathcal{D}} \mathcal{N}\left(0, V^w_{over}\right) \tag{8}$$

mit

$$V^w_{over} = \frac{(1+\lambda)^2 + \lambda}{(1+\lambda)^2} \mathbb{E}\left[\exp(\beta_t^T x)\right] M_f^{-1}$$

Vergleich der Kovarianzen

- Es gilt falls $c > 0$, $V_{under}^{w} \geq V_f$ in der Loewner Ordnung.
- Es gilt falls $\lambda \neq \{0, \infty\}$, $V_{over}^{w} \geq V_f$ in der Loewner Ordnung.
- Die Effizienz wir in beiden Fällen schlechter

Hilfsaussagen

ZGWS von Lindeberg (van der Vaart, 1998 S. 20)

Für jedes n seien $Y_{n,1}, \ldots, Y_{n,k_n}$ unabhängige Zufallsvektoren mit endlicher Varianz so dass für jedes $\epsilon > 0$ gilt

$$\sum_{i=1}^{k_n} \mathbb{E}\left[||Y_{n,i}||^2 \, I_{||Y_{n,i}||>\epsilon} \right] \to 0 \quad \text{und}$$

$$\sum_{i=1}^{k_n} \mathrm{Cov}(Y_{n,i}) \to \Sigma$$

Dann gilt

$$\sum_{i=1}^{k_n} (Y_{n,i} - \mathbb{E}[Y_{n,i}]) \to_{\mathcal{D}} \mathcal{N}(0, \Sigma)$$

Hilfsaussagen

Basic Corollary (Hjort and Pollard, 2011 S. 2)

Es sei $A_n(s)$ convex und darstellbar der Form $s^T VS + U_n^T s + C_n + r_n(s)$ wobei V symmetrisch und positiv definit ist, U_n stochastisch beschränkt, C_n beliebig, und $r_n(s) \in o_p(1)$. Dann gilt für den Minimierer α_n von A_n: α_n und der Minimierer $\beta_n = -V^{-1}U_n$ von $s^T VS + U_n^T s + C_n$ unterscheiden sich um $o_p(1)$. Gilt zusätzlich $U_n \rightarrow_{\mathcal{D}} U$ so folgt $\alpha_n \rightarrow_D -V^{-1}U$.

Hilfsaussagen

Theorem von Slutsky (Rohde, WS2020/2021 S.58)

Gilt für die Folgen von ZVAs $A_n \to_p A$ und $B_n \to_{\mathcal{D}} B$ so folgt für das Produkt: $A_n B_n \to_{\mathcal{D}} AB$.

Hilfsaussagen

- Definiere $a_n := \sqrt{n \exp(\alpha_t)}$
- Die Annahme $\mathbb{E}\left[\exp(t\,||x||)\right] < \infty$ für $t > 0$ liefert $\mathbb{E}\left[\exp(t_1\,||x||)\,||z||^{t_2}\right] < \infty$ mit $t_1, t_2 > 0$. Wählt man für $t_1, t_2 > 0$ $t > t_1$ und $k > t_2$ folgt die Aussage aus

$$\exp(t\,||x||) \geq \exp(-t)\exp(t\,||z||)$$

$$= \exp(-t)\exp(t_1\,||z||)\exp((t-t_1)\,||z||)$$

$$\geq \frac{\exp(-t)(t-t_1)^k}{k!}\,||z||^k \exp(t_1\,||x||)$$

$$\geq \frac{\exp(-t)(t-t_1)^k}{k!}\,||z||^{t_2} \exp(t_1\,||x||)$$

mit Wahrscheinlichkeit 1.

Hilfsaussagen

Es gilt $n_1 = n \exp(\alpha_t)\mathbb{E}\left[\exp(\beta_t^T x)\right]\{1 + o_p(1)\}$. Dazu betrachte man

$$
\begin{aligned}
\mathbb{P}(y = 1|x)(1 + \exp(\alpha_t + \beta_t^T x)) &= \exp(\alpha_t + \beta_t^T x) \\
\Rightarrow \mathbb{E}\left[\mathbb{P}(y = 1|x)(1 + \exp(\alpha_t + \beta_t^T x))\right] &= \mathbb{E}\left[\exp(\alpha_t + \beta_t^T x)\right] \\
\Rightarrow \underbrace{\mathbb{E}\left[\mathbb{P}(y = 1|x)\right]}_{=\mathbb{P}(Y=1)=\mathbb{E}[Y]}\{1 + o_p(1)\} &= \exp(\alpha_t)\mathbb{E}\left[\exp(\beta_t^T x)\right]
\end{aligned}
$$

Hilfsaussagen

Gleichzeitig gilt $\frac{n_1}{n} = \frac{\sum_{i=1}^n y_n}{n} \rightarrow_p \mathbb{E}[Y]$. Einsetzen liefert

$$\frac{n_1}{n}\{1 + o_p(1)\}\{1 + o_p(1)\} = \exp(\alpha_t)\mathbb{E}\left[\exp(\beta_t^T x)\right]$$
$$\Rightarrow n_1 = n \exp(\alpha_t)\mathbb{E}\left[\exp(\beta_t^T x)\right]\{1 + o_p(1)\}$$

Theorem 1

Beweis Idee:

- Schreibe Loglikelihood-Funktion um in die Form $\frac{1}{2}s^T Vs + B_n^T s + C_n + r_n(s)$, V symmetrisch und $r_n(s) \in o_{\mathbb{P}}(1)$
- ZGW Linderberg liefert $B_n \to_{\mathcal{D}} \mathcal{N}(0, W_1)$
- Zeigen direkt $V \to_p W_2$
- Basic corollary und Slutsky liefern die Aussage

Theorem 1

Beweis Idee:

- **Schreibe Loglikelihood-Funktion um in die Form** $\frac{1}{2}s^T V s + B_n^T s + C_n + r_n(s)$**,** V **symmetrisch und** $r_n(s) \in o_{\mathbb{P}}(1)$
- ZGW Linderberg liefert $B_n \to_{\mathcal{D}} \mathcal{N}(0, W_1)$
- Zeigen direkt $V \to_p W_2$
- Basic corollary und Slutsky liefern die Aussage

- $\hat{\theta}$ maximiert

$$l(\theta) = \sum_{i=1}^{n} \Big((\alpha + \beta x_i^T) y_i - log(1 + \exp(\alpha + \beta x_i^T)) \Big) \tag{9}$$

- Somit maximiert $u_n = a_n(\hat{\theta} - \theta_t)$

$$\gamma(u) = l(\theta_t + a_n^{-1} u) - l(\theta_t) \tag{10}$$

- Wir bilden die Taylorentwicklung von γ bei $u = 0$ bis zur zweiten Ordnung:

$$\gamma(u) = a_n^{-1} u^T \dot{l}(\theta_t) + 0.5 a_n^{-2} \sum_{i=1}^{n} \phi_i(\theta_t + a_n \grave{u})(z_I^T u)^2 \qquad (11)$$

wobei $\phi_i(\theta) = p_i(\theta)(1 - p_i(\theta))$, $\dot{l}$ der Gradient der Likelihood Funktion ist und $\grave{u} \in [0, u]$.

- Es folgt mit der Kettenregel :

$$\dot{l}(\theta) = \sum_{i=1}^{n} z_i y_i - z_i \underbrace{\frac{\exp(\alpha + \beta x_i^T)}{1 + \exp(\alpha + \beta x_i^T)}}_{= p_i(\theta)} \tag{12}$$

- Mit der Produktregel folgt $\dot{p}_i(\theta) = p_i(\theta)(1 - p_i(\theta))z_i$ und somit :

$$\frac{d^2}{d\theta^2} l(\theta) = - \sum_{i=1}^{n} \phi_i(\theta_t + a_n^{-1} u) z_i z_i^T \tag{13}$$

- Ziel ist es, zu zeigen dass

$$a_n^{-1}\dot{l}(\theta) \to_{\mathcal{D}} \mathcal{N}(0, M_f) \tag{14}$$

mit $M_f = \mathbb{E}\left[\exp(\beta_t^T x) z z^T\right]$.

- Desweiteren wollen wir

$$-a_n^{-2}\sum_{i=1}^{n}\phi_i(\theta_t + a_n^{-1}\grave{u}) z_i z_i^T \to_p M_f \tag{15}$$

für alle $\grave{u} \in [0, u]$.

Theorem 1

Beweis Idee:

- Schreibe Loglikelihood-Funktion um in die Form $\frac{1}{2}s^T Vs + B_n^T s + C_n + r_n(s)$, V symmetrisch und $r_n(s) \in o_{\mathbb{P}}(1)$
- **ZGW Linderberg liefert** $B_n \to_{\mathcal{D}} \mathcal{N}(0, W_1)$
- Zeigen direkt $V \to_p W_2$
- Basic corollary und Slutsky liefern die Aussage

Wir zeigen zuerst (14) und überprüfen dazu die Voraussetzungen.

- Es gilt $\mathbb{E}[(y_i - p_i)z_i] = \mathbb{E}[\mathbb{E}[(y_i - p_i)z_i|z_i]] = 0$. Linearität liefert $\mathbb{E}[a_n^{-1}\dot{l}(\theta_t)] = 0$.
- Für die Varianz gilt

$$
\begin{aligned}
V(a_n^{-1}\dot{l}(\theta_t)) &= a_n^{-2}\sum_{i=1}^{n} V((y_i - p_i)z_i) \\
&= a_n^{-2}\sum_{i=1}^{n} \mathbb{E}[(y_i - p_i)^2 z_i z_i^T] \\
&= a_n^{-2}\sum_{i=1}^{n} \mathbb{E}[\mathbb{E}[(y_i - p_i)^2|z_i] z_i z_i^T]
\end{aligned}
$$

$$= a_n^{-2} \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}[(1-p_i)^2 I_{y_i=1} | z_i] z_i z_i^T] + \mathbb{E}[\mathbb{E}[p_i^2 I_{y_i=0} | z_i] z_i z_i^T]$$

$$= a_n^{-2} \sum_{i=1}^{n} \mathbb{E}[(1-p_i)^2 \underbrace{\mathbb{E}[I_{y_i=1} | z_i]}_{=p_i} z_i z_i^T] + \mathbb{E}[p_i^2 \underbrace{\mathbb{E}[I_{y_i=0} | z_i]}_{=(1-p_i)} z_i z_i^T]$$

$$= a_n^{-2} n \mathbb{E}[p_i(1-p_i) z_i z_i^T]$$

$$= a_n^{-2} n \exp(\alpha_t) \mathbb{E}\left[\frac{\exp(\beta_t^T x)}{(1+\exp(\alpha_t + \beta_t x_i^T))^2} z z^T\right]$$

$$= \mathbb{E}\left[\frac{\exp(\beta_t^T x)}{(1+\exp(\alpha_t + \beta_t x_i^T))^2} z z^T\right]$$

- Wir bemerken dass

$$\frac{\exp(\beta_t^T x)}{(1+\exp(\alpha_t+\beta_t x^T))^2} zz^T \rightarrow_{f.s.} \exp(\beta_t^T x) zz^T$$

- Weiter gilt $\frac{\exp(\beta_t^T x)}{(1+\exp(\alpha_t+\beta_t x^T))^2} \|z\|^2 \leq \exp(\beta_t^T x) \|z\|^2$ und $\mathbb{E}\left[\exp(\beta_t^T x) \|z\|^2\right] < \infty$
- Dominierte Konvergenz liefert also $V(a_n^{-1} \dot{l}(\theta_t)) = \mathbb{E}\left[\frac{\exp(\beta_t^T x)}{(1+\exp(\alpha_t+\beta_t x^T))^2} zz^T\right] \rightarrow \mathbb{E}\left[\exp(\beta_t^T x) zz^T\right]$

Wir zeigen nun die Lindeberg-Bedingung. Sei $\epsilon > 0$:

$$\sum_{i=1}^{n} \mathbb{E}\left[||(y_i - p_i)\, z_i||^2 \, I_{||(y_i - p_i) z_i|| \geq \epsilon a_n} \right]$$

$$= n\mathbb{E}\left[||(y - p)\, z||^2 \, I_{||(y-p) z|| \geq \epsilon a_n} \right]$$

$$= n \Bigg(\mathbb{E}\left[\mathbb{E}\left[(1-p)^2 I_{y=1} | z \right] ||z||^2 \, I_{||(1-p) z|| \geq \epsilon a_n} \right]$$

$$+ \mathrm{E}\left[\mathbb{E}\left[p^2 I_{y=0} | z \right] ||z||^2 \, I_{||(-p) z|| \geq \epsilon a_n} \right]$$

$$
\begin{aligned}
&= n\left(\mathbb{E}\left[p(1-p)^2\,||z||^2\, I_{||(1-p)z||\geq \epsilon a_n}\right] + \mathbb{E}\left[p^2(1-p)\,||z||^2\, I_{||pz||\geq \epsilon a_n}\right]\right) \\
&\leq n\left(\mathbb{E}\left[p\,||z||^2\, I_{||(1-p)z||\geq \epsilon a_n}\right] + \mathbb{E}\left[p^2\,||z||^2\, I_{||pz||\geq \epsilon a_n}\right]\right) \\
&\leq a_n^2\mathbb{E}\left[\exp\left(||\beta_t||\,||x||\right)||z||^2\, I_{||z||\geq \epsilon a_n}\right] \\
&\quad + a_n^2\mathbb{E}\left[\exp\left(||\beta_t||\,||x||\right)||z||^2\, I_{||z||\geq \epsilon a_n}\right] \\
&= o(a_n^2)
\end{aligned}
$$

Theorem 1

Beweis Idee:

- Schreibe Loglikelihood-Funktion um in die Form $\frac{1}{2}s^T Vs + B_n^T s + C_n + r_n(s)$, V symmetrisch und $r_n(s) \in o_{\mathbb{P}}(1)$
- ZGW Linderberg liefert $B_n \to_{\mathcal{D}} \mathcal{N}(0, W_1)$
- **Zeigen direkt** $V \to_p W_2$
- Basic corollary und Slutsky liefern die Aussage

Es bleibt Gl.(15) zu zeigen:

$$a_n^{-2} \sum_{i=1}^{n} \phi_i(\theta_t + a_n^{-1} \grave{u}) z_i z_i^T \to_p M_f$$

für alle $\grave{u} \in [0, u]$.

- Dazu zeigen wir zuerst $\left| a_n^{-2} \sum_{i=1}^{n} \phi_i(\theta_t + a_n^{-1} \grave{u}) \, ||z_i||^2 - a_n^{-2} \sum_{i=1}^{n} \phi_i(\theta_t) \, ||z_i||^2 \right| \in o_p(1)$
- Nutzen anschließend $a_n^{-2} \sum_{i=1}^{n} \phi_i(\theta_t) z_i z_i^T \to_p \mathbb{E}\left[\exp(\beta_t x) z z^T\right]$ um insgesamt Gl.(15) zu erhalten.
- Wegen $A \to_p B$, $B \to_p C \Rightarrow A \to_p C$ gilt dann die Aussage

Es gilt

$$\left| a_n^{-2} \sum_{i=1}^{n} \phi_i(\theta_t + a_n^{-1} \grave{u}) \, ||z_i||^2 - a_n^{-2} \phi_i(\theta_t) \, ||z_i||^2 \right|$$

$$\leq a_n^{-2} \sum_{i=1}^{n} \left| \underbrace{\phi_i(\theta_t + a_n^{-1} \grave{u})}_{=\phi_i(\theta_t) + a_n^{-1} \grave{u} \dot{\phi}_i(\theta_t + a_n^{-1} \bar{u})} - \phi_i(\theta_t) \right| ||z_i||^2$$

mit $\bar{u} \in [0, \grave{u}]$.

Da $\dot{\phi}_i = \dot{p}_i(1 - p_i) + p_i(1 \dot{-} p_i) = z_i(p_i(1 - p_i)^2 - p_i^2(1 - p_i)) \leq z_i p_i$ gilt, folgt

$$\leq \left|\left| a_n^{-1} \grave{u} \right|\right| \, a_n^{-2} \sum_{i=1}^{n} p_i(\theta_t + a_n^{-1} \bar{u}) \, ||z_i||^3$$

$$= \frac{\left|\left| a_n^{-1} \grave{u} \right|\right|}{n} \sum_{i=1}^{n} \frac{\exp(\beta_t^T x_i + a_n^{-1} \bar{u}^T z_i)}{(1 + \exp(\theta_t^T z_i + a_n^{-1} \bar{u}^T z_i))^2} \, ||z_i||^3$$

$$\leq \frac{\left|\left|a_n^{-1}\grave{u}\right|\right|}{n} \sum_{i=1}^{n} \exp((||\beta_t||\,||\bar{u}||)(1 + ||x_i||))\,||z_i||^3$$

$$= o_p(1)$$

Es gilt zusätzlich

$$a_n^{-2} \sum_{i=1}^{n} \phi_i(\theta_t) z_i z_i^T = \frac{1}{n \exp(\alpha_t)} \sum_{i=1}^{n} \frac{\exp(\alpha_t + \beta_t^T x_i)}{1 + \exp(\alpha_t + \beta_t^T x_i)} z_i z_i^T$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(\beta_t^T x_i)}{1 + \exp(\alpha_t + \beta_t^T x_i)} z_i z_i^T$$

Schwaches Gesetz liefert:

$$= \mathbb{E}\left[\frac{\exp(\beta_t^T x_i)}{1 + \exp(\alpha_t + \beta_t^T x_i)} z_i z_i^T\right] + o_p(1)$$

Majorisierte Konvergenz liefert

$$= \underbrace{\mathbb{E}\left[\exp(\beta_t^T x_i) z_i z_i^T\right]}_{= M_f} + o_p(1)$$

Theorem 1

Beweis Idee:

- Schreibe Loglikelihood-Funktion um in die Form $\frac{1}{2}s^T V s + B_n^T s + C_n + r_n(s)$, V symmetrisch und $r_n(s) \in o_{\mathbb{P}}(1)$
- ZGW Linderberg liefert $B_n \to_{\mathcal{D}} \mathcal{N}(0, W_1)$
- Zeigen direkt $V \to_p W_2$
- **Basic corollary und Slutsky liefern die Aussage**

Als letztes kombinieren wir die obigen Resultate. Gl. (14) und (15) liefern folgende Darstellung von γ:

$$\gamma(u) = 0.5u^T M_f u + a_n^{-1} \dot{l}(\theta) u + o_p(1)$$

- Da $a_n^{-1} \dot{l}(\theta) \to_{\mathcal{D}} \mathcal{N}(0, M_f)$, liefert das "Basic Corollary"

$$a_n(\hat{\theta} - \theta_t) = \underbrace{M_f^{-1} a_n^{-1} \dot{l}(\theta)}_{\sim \mathcal{N}(0, M_f^{-1})} + o_{\mathcal{D}}(1)$$

- Wegen der Vorüberlegung gilt aber $a_n = \sqrt{n \exp(\alpha_t)} = \sqrt{\frac{n_1}{\mathbb{E}\left[\exp(\beta_t^T x_i)\right](1 + o_p(1))}}$ und es folgt somit

$$\sqrt{n_1}(\hat{\theta} - \theta_t) = \underbrace{\sqrt{\mathbb{E}\left[\exp(\beta_t^T x_i)\right]} M_f^{-1} a_n^{-1} \dot{l}(\theta)}_{\sim \mathcal{N}(0, \mathbb{E}\left[\exp(\beta_t^T x_i)\right] M_f^{-1})}(1 + o_p(1)) \qquad (16)$$

Theorem 2 und 3

Beweis Idee ist identisch zu Theorem 1!

- Schreibe Loglikelihood-Funktion um in die Form $\frac{1}{2}s^T Vs + B_n^T s + C_n + r_n(s)$, V symmetrisch und $r_n(s) \in o_{\mathbb{P}}(1)$
- ZGW Linderberg liefert $B_n \rightarrow_{\mathcal{D}} \mathcal{N}(0, W_1)$
- Zeigen direkt $V \rightarrow_p W_2$
- Basic corollary und Slutsky liefern die Aussage

Lassen den Beweis weg.

Diskussion

- Wir haben gesehen, dass die Grenzverteilungen mit einer Geschwindigkeit von $\frac{1}{\sqrt{n_1}}$ erreicht werden.
- Dies impliziert eine Inhomogene "Informationskonzetration" in den Daten (seltene Cases enthalten mehr "Information" als Controls)
- Subsampling verändert dementsprechend den Informationsinhalt des Datensatzes kaum und liefert eine gute Effizienz
- Oversampling verfälscht signifikant den Informationsinhalt des Datensatzens und liefert eine schlechtere Effizienz

Quellen

Hjort, Nils Lid and David Pollard (July 2011). "Asymptotics for minimisers of convex processes". In: arXiv: 1107.3806 [math.ST].

Rohde, Angelike (WS2020/2021). *Wahrscheinlichkeitstheorie Skript*.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*.

Wang, HaiYing (June 2020). "Logistic Regression for Massive Data with Rare Events". In: arXiv: 2006.00683 [stat.ML].