Additive hazard Regression mit zensierte Datensätzen

Paul Schulz

19.01.2015

1 Ausgangsproblematik

Wir möchten von zensierte Datensätze ausgehen und in diesen die Überlebenswahrscheinlichkeiten und die Risikofaktoren für die verschiedenen Individuen schätzen. Unter zensierten Datensätzen verstehen wir solche, in denen es nicht nur möglich ist, dass Individuen durch einen Ausfall aus der Studie wegfallen, sondern auch aus anderen Gründen aus der Studie ausscheiden.

2 Regularisierte Schätzer für Ereigniszeiten

Sei T die Ausfallszeit, C die Abbruchszeit (Zeit bis zu der maximal beobachtet wird), $\Delta = I\left(T \leq C\right)$ der Fehlerindikator, $X = T \wedge C$ Ausscheidezeit und $Z\left(\cdot\right)$ ein Vektor von vorhersagbaren Kovarianzprozessen. Wir beobachten $\left(X_i, \Delta_i, Z_i\right)$ und nehmen an das die bedingte Hazardfunktion gegeben ist durch:

$$\lambda\left(t|Z\right) = \lambda_0(t) + \beta_0^T Z\left(t\right) \tag{1}$$

Die bedingte Hazardfunktion, gibt an mit welcher Wahrscheinlichkeit ein Individuum ausfällt gegeben Z. λ_0 ist der Basehazard, welcher beschreibt wie wahrscheinlich ein Ausfall generell ist und $\beta_0^T Z$ modelliert den Einfluss der Risikofaktoren auf die Ausfallswahrscheinlichkeit.

Sei $N_i(t) = I(X_i \leq t, \Delta_i = 1)$ der 'Ausfall beobachtet'-Zählprozess und $Y_i(t) = I(X_i \geq)$ der Risikoindikator. Dann lässt sich $N_i(t)$ nach der Doob-Zerlegung umschreiben zu :

$$N_i(t) = M_i(t) + \int_0^t Y_i(s) \left\{ \lambda_0(s) + \beta_0^T Z_i(s) \right\} ds \tag{2}$$

,wobei $M_i(t)$ein Martingal und $\int_0^t Y_i(s) \left(\lambda_0(s) + \beta_0^T Z_i(s) \right) ds$ vorhersagbar ist. Wir möchten β_0 mit Hilfe folgender Schätzfunktion annähern:

$$U(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \left\{ Z_i(t) - \bar{Z}_i(t) \right\} \left\{ dN_i(t) - Y_i(t) \beta_0^T Z_i(t) \, dt \right\}$$

,wobei $\beta \in \mathbb{R}$, $\bar{Z}(t) = \sum_{i=1}^n Y_i(t) Z_i(t) / \sum_{i=1}^n Y_i(t)$ und τ die maximale Beobachtungszeit ist. Dies können wir umschreiben zu $U(\beta) = b - V\beta$ mit

$$b = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ Z_i(t) - \bar{Z}(t) \right\} dN_i$$

'Mittlere Abweichung der Parameter von $-\bar{Z}$ zum Ausfallzeitpunkt',

$$V = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \left\{ Z_i(t) - \bar{Z}(t) \right\}^{\otimes 2} dt$$

'Empirische Covarianzmatrix aller Aktiven gemittelt über den Beobachtungszeitraum'.
Wir erhalten durch integrieren von $-U(\beta)$ nach β eine Fehlerfunktion

$$L(\beta) = \frac{1}{2} \beta^T V \beta - b^T \beta \tag{3}$$

$L(\beta)$ lässt sich als empirischer Vorhersagefehler interpretieren.
Den regularisierten Schätzer $\hat{\beta}$ definieren wir durch:

$$\hat{\beta} = arg\ min_{\beta \in (R)^p} \left\{ Q(\beta) \equiv L(\beta) + \sum_{i=1}^{p} p_\lambda(|\beta_j|) \right\} \tag{4}$$

Mögliche penalty-Funktionen:

- Lasso, L_1-penalty, $\rho(\Theta) = \Theta, \Theta > 0$.
- SICA, Familie von penalty-Funktionen welche glatt zwischen L_0- und L_1-penalties liegt,

$$\rho(\Theta) = \frac{(a+1)^{\Theta}}{a+\Theta}, \Theta \leq 0, a > 0 \text{ fest.} \tag{5}$$

3 Schwache Oracle Eigenschaft

3.1 Bedingungen und Notation

Wir definieren

$$D = E\left[\int_0^{\tau} \{Z(t) - E[Y(t)Z(t)]/E[Y(t)]\}^{\otimes 2} dt\right],$$

'Erwartete Covarianzmatrix' und 'Populationsgegenstück zu V'

$$A: \text{ 'Aktive Menge', } s = |A|,$$

$$\Omega_L: \text{ '}sup_{t\in[0,\tau]} |Z_j(t)| \text{ ist durch } L \leq 0 \text{ gleichmäßig beschränkt.'}$$

$$d = \frac{1}{2} min_{j\in A} |\beta_{0j}|,$$

'Parameter um die minimale Größe der Einträge zu kontrollieren (vgl. β_{min})'

$$\varphi = \left\|D_{AA}^{-1}\right\|_{\infty},$$

$$\mu = \Lambda_{min}(D_{AA}) - \lambda\kappa_0,$$

wobei $\Lambda_{min}(\cdot)$ den kleinsten Eigenwert bezeichnet und κ_0 die Krümmung der penalty-Funktion kontrolliert. Im weiteren Verlauf möchten wir $\mu > 0$ haben. Ein größeres μ erlaubt uns eine einfachere Schätzung.

Bedingung 1 Die Funktion $\rho_\lambda(\Theta)$ ist monoton wachsend und konkav in $\Theta \in$

$[0, \infty)$ und hat die stetige Ableitung $\rho_\lambda^{'}(\Theta)$ auf $(0, \infty)$. Außerdem ist $\rho_\lambda^{'}(\Theta)$ monoton wachsend in λ und $\rho_\lambda^{'}(0+) \equiv \rho^{'}(0) > 0$ ist unabhängig λ

Bedingung 2 (1)$\int_0^\tau \lambda_0(t)\, dt < \infty$.(2) $P\{Y(\tau) > 0\}$.
(3)Es gibt Konstanten $D, K, r > 0$, so dass

$$P\left(sup_{t\in[0,\tau]} |Z_i(t)| > x\right) \leq D \exp(-Kx^r),$$

für alle $x > 0$ und $i = 1, \ldots, p$.
(4) Die Pfade von $Z_j, j = 1, \ldots, p$, sind von gleichmäßig beschränkter Variation.

Bedingung 3 Es gibt Konstanten $\alpha \in (0, 1], \gamma \in \left[, \frac{1}{2}\right]$ und $c > 0$, so dass

$$\left\|D_{A^C A} D_{AA}^{-1}\right\|_\infty \leq \left((1-\alpha)\frac{\rho^{'}(0+)}{\rho_\lambda^{'}(d)}\right) \wedge (cn^\gamma)$$

'A undA_C sind nicht zu stark korreliert' und 'D hat nicht zu kleine Einträge'

3.2 Vorbereitende Lemmata

Lemma 1 *(Massart, 2000, Theorem 9)* Sei $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit Werten in $\mathbb{R}^N$. Für gewisse reellen Zahlen $a_{i,t}$ und $b_{i,t}$, so dass $a_{i,t} \leq X_{i,t} \leq b_{i,t}$, für alle $i \leq n$ und alle $t \leq N$. Sei

$$Z = sup_{1\leq t\leq N} \sum_{i=1}^{n} X_{i,t}$$

und definiere $L^2 = sup_{1\leq t\leq N} \sum_n^{i=1} (b_{i,t} - a_{i,t})^2$. Dann gilt für alle positiven x,

$$P[Z \geq E[Z] + x] \leq \exp\left(-\frac{x^2}{2L^2}\right).$$

Lemma 2 *(Konzentration von empirischen Matrizen)* Unter Bedingung $1-3$, $\mu > 0$ und $\varphi \vee \mu^{-1} = O(\sqrt{n}/s)$, gibt es Konstanten $D, K > 0$, so dass

$$P\left(\left\|V_{AA}^{-1}\right\|_\infty \geq 2\left\|D_{AA}^{-1}\right\|_\infty \mid \Omega_L\right) \leq s^2 D \exp\left(-K\frac{n}{L^4}\left(\frac{1}{\varphi^2 s^2} \wedge 1\right)\right), \qquad (6)$$

$$
\begin{aligned}
P\left(\left\|V_{A^CA}V_{AA}^{-1}\right\|_\infty \geq \left((1-\frac{\alpha}{2}\right)\frac{\rho^{'}(0+)}{\rho_\lambda^{'}(d)} \wedge (2cn^\gamma)\left\{\geq \left\|D_{A^CA}D_{AA}^{-1}\right\|_\infty\right\} \mid \Omega_L\right) \\
\leq (p-s)sD\exp\left(-K\frac{n}{L^4}\left(\frac{\left(\rho_\lambda^{'}(d)^{-1}\wedge n^\gamma\right)^2}{\varphi^2 s^2}\wedge 1\right)\right) \\
+s^2\exp\left(-K\frac{n}{L^4}\left(\frac{1}{\varphi^2 s^2}\wedge 1\right)\right)
\end{aligned}
\tag{7}
$$

'V weicht nicht stark von D ab'

$$
P\left(\Lambda_{min}\left(V\right) \leq \lambda\kappa_0 \middle| \, \Omega_L\right) \leq s^2\exp\left(-K\frac{n}{L^4}\left(\frac{\mu^2}{s^2}\wedge 1\right)\right) \tag{8}
$$

'Der Effekt von V wird nicht durch den Effekt von ρ überlagert'

Lemma 3 *(Charakterisierung des regularisierten Schätzers).* Unter Bedingung 1 ist $\hat{\beta} \in \mathbb{R}^\rho$ ein strikter lokaler Minimierer von 4, falls gilt:

$$
U_{\hat{A}}\left(\hat{\beta}\right) - \lambda\rho_\lambda^{'}\left(\left|\hat{\beta}_{\hat{A}}\right|\right) \circ sgn\left(\hat{\beta}_{\hat{A}}\right) = 0, \tag{9}
$$

$$
\left\|U_{\hat{A}^C}\left(\hat{\beta}\right)\right\|_\infty < \lambda\rho\left(0+\right), \tag{10}
$$

$$
\Lambda_{min}\left(V_{\hat{A}\hat{A}}\right) > \lambda_\kappa\left(\rho_\lambda; \hat{\beta}_{\hat{A}}\right) \tag{11}
$$

'Bedingung vom Typ Ableitung verschwindet'.

3.3 Schwache Oracel Eigenschaft

Satz 1 *(Schwache Oracel Eigenschaft) Zusätzlich zu Bedingung $1-3$ gelte:*

$$
\frac{\left(\rho_\lambda^{'}(d)^{-1}\wedge n^\gamma\right)^2}{\varphi^2 s^2\left(\log p\right)^{r_1}} \to \infty, \quad \frac{\left(\varphi^{-1}\wedge \mu\right)}{s^2\left(\log s\right)^{r_1}} \to \infty, \tag{12}
$$

$$
\frac{n\lambda^2}{\left(\log p\right)^{r_1}} \to \infty, \quad \frac{n^{1-2\gamma}\lambda^2}{\left(\log s\right)^{r_1}} \to \infty, \tag{13}
$$

$$
d \geq c_1\varphi\lambda\rho^{'}\left(0+\right), \tag{14}
$$

mit $\mu > 0, r_1 = (r+4)/r$, *und* $c_1 = 2 + 1/(4c)$. *Dann gilt für Konstanten* $D, K > 0$, *mit Wahrscheinlichkeit mindestens*

$$1 - D\exp\left(-Kn^{(1/r_1)}\left(\frac{(\varphi^{-1}\wedge\mu)}{s^2}\wedge 1\right)^{(1/r_1)}\right) \tag{15}$$

$$-D\exp\left(-Kn^{(1/r_1)}\left(\frac{(\lambda^2)}{n^2\gamma}\wedge 1\right)^{(1/r_1)}\right) \to 1 \tag{16}$$

existiert ein regulärer Schätzer $\hat{\beta}$, *welcher folgende Bedingungen erfüllt:*

- *(Sparsamkeit)* $\hat{\beta}_{A^C} = 0$
- *(Abweichung in der* L_∞*-Norm)* $\left\|\hat{\beta} - \beta_{0A}\right\|_\infty \leq c_1\varphi\lambda\rho'(0+)$

4 Methode des Koordinatenweisen Abstiegs

Wie schon beim Lasso-Schätzer benutzen wir um tatsächlich ein $\hat{\beta}$ zu berechnen die Methode des Koordinatenweisen Abstiegs. Hierbei ist zu beachten, dass unsere penalty-Funktion nicht konvex ist. Damit die Optimierung trotzdem funktioniert, müssen wir die Konvexität von L ausnutzen.

5 Verweise

Eine empirische Betrachtung für verschiedene penalty-Funktionen ist in „High-Dimensional Sparse Additive Hazards Regression“ von Wei Lin und Jinchi LV in den Tabellen 1-3 zu finden. Figur 1-3 zeigen die Anzahl der korrekt ausgewählten Parameter mit steigender Modellgröße.
Wei Lin stellt auf seiner Hompage http://162.105.204.96/teachers/linw/software.html ein vorgeschriebenes R Programm in form seines Ordners sahr zur Verfügung.