Stockholms universitet, Matematiska institutionen

Tentamen: MT7028 Prissättning inom sakförsäkring, 2025-01-08, 8–13

Examinator: Filip Lindskog, lindskog@math.su.se
Återlämning: Meddelas via kursforum.
Tillåtna hjälpmedel: Inga.
Allmänt: Resonemang skall vara tydliga och lätta att följa.

Uppgift 1

För att modellera skadefrekvens vill vi bestämma parametervektorn β som minimerar

$$\sum_{i=1}^{n} d(n_i, w_i\mu(x_i;\beta)), \quad d(y,m) = y\log\frac{y}{m} - y + m,$$

där $\{(n_i, w_i, x_i) : i = 1, \dots, n\}$ är realisationer av oberoende likafördelade tripletter där n_i är antalet skador för kontrakt nr i, w_i är kontraktsduration och x_i är en kovariatvektor. Antag att kovariatvektorn $x = (x^{(1)}, x^{(2)})$ är 2-dimensionell och att funktionen $\mu(x;\beta)$ ges av

$$\mu(x;\beta) = \exp\left(\beta_1 + \beta_2 \mathbb{1}_{\{x^{(1)}<a\}} + \beta_3 \mathbb{1}_{\{x^{(2)}<b\}}\right)$$

för några numeriska värden a, b. $\mathbb{1}_{\{x^{(1)}<a\}}$ antar värdet 1 om $x^{(1)} < a$, annars värdet 0. Bestäm ett ekvationsystem som de optimala parametervärdena $\beta_1, \beta_2, \beta_3$ löser. (10 p)

Uppgift 2

Antag premieargumenten fordonsålder, med nivåer *under 3 år* respektive *över 3 år*, och ägarålder, med nivåer *under 30 år* respektive *över 30 år*. Antag en modell för (förväntad betingad) skadefrekvens på formen $\mu_{i,j} = \gamma_0\gamma_{1,i}\gamma_{2,j}$, $i = 1, \dots, m_1$ och $j = 1, \dots, m_2$. Ange m_1, m_2 och uttryck $\mu_{i,j}$ på formen i Uppgift 1 samt ange designmatrisen inklusive hur du väljer att ordna cellerna (i, j). (10 p)

Uppgift 3

För ett enskilt försäkringskontrakt är N är antalet skador, W är kontraktsdurationen och X är en kovariatvektor som beskriver försäkringstagaren och det försäkrade objektet. Antag att N givet W och X är Poissonfördelad enligt:

$$P(N = n \mid W = w, X = x) = e^{-w\mu(x)} \frac{(w\mu(x))^n}{n!}, \quad n = 0, 1, \dots$$

Antag att skadebeloppen $C_1, \dots, C_N$ för de N skadorna för kontraktet är oberoende och likafördelade, och oberoende av N, W, X. Låt $\pi(x)$ beteckna 1-årspremien för ett försäkringskontrakt med kovariatvektor x som betalas för den tid (upp till ett år) som kontraktet är i bruk (kan avslutas i förtid). Bestäm $\pi(x)$ om det väljs så att förväntad premieintäkt för ett försäkringskontrakt med kovariatvektor x är lika med förväntad skadekostnad för samma kontrakt. (10 p)

Uppgift 4

Antag att vi vill modellera (förväntad betingad) skadefrekvens som funktion av ägarålder med ett polynom $p(x;\theta) = \theta_0 + \theta_1 x + \theta_2 x^2$. Vi har endast data

$$\{(y_1, w_1, x_1), (y_2, w_2, x_2), (y_3, w_3, x_3), (y_4, w_4, x_4)\}.$$

För varje kombination av $k \in \{1, \dots, 4\}$ och $\lambda \in \{\lambda_1, \dots, \lambda_{10}\}$, $0 < \lambda_1 < \cdots < \lambda_{10}$, beräknar vi den minimerande parametervektorn

$$\widehat{\theta}_{\lambda,k} = \operatorname{argmin}_\theta \left(\sum_{i \neq k} w_i (y_i - p(x_i;\theta))^2 + \lambda \int_{x_{\min}}^{x_{\max}} (p''(x;\theta))^2 dx \right).$$

Vi avser välja parametervektorn θ med hjälp av korsvalidering så att $p(x;\theta)$ passar data väl samtidigt som vi undviker överanpassning. Beskriv hur vi ska använda vektorerna $\widehat{\theta}_{\lambda,k}$ för att välja ett optimalt θ? (10 p)

Uppgift 5

Antag att vi har tre observationer

$$(y_1, x_1) = (5, 60), \quad (y_2, x_2) = (12, 25), \quad (y_3, x_3) = (7, 40),$$

av oberoende likafördelade par (Y_i, X_i), där Y_i är skadekostnad och X_i är ägarålder. Anpassa ett regressionsträd $x \mapsto T(x)$ med två löv till datamaterialet. Beräkna generaliseringsfelet

$$E[(Y - T(X))^2],$$

där det anpassade regressionsträdet betraktas som känt, under antagande att X är likformigt fördelad på $(20, 80)$ och Y givet $X = x$ är normalfördelad med väntevärde

$$\mu(x) = 10\mathbb{1}_{\{x \leq 30\}} + 4\mathbb{1}_{\{x > 30\}}$$

och varians 20. (10 p)

Uppgift 1

Med $\mu_i := \mu(x_i;\beta)$, $m_i = w_i\mu_i$, $d_i = d(n_i, m_i)$ fås

$$\frac{\partial}{\partial\beta_j}\sum_{i=1}^n d(n_i, w_i\mu(x_i;\beta)) = \sum_{i=1}^n \frac{\partial d_i}{\partial m_i}\frac{\partial m_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial\beta_j} = \sum_{i=1}^n \Big(1 - \frac{n_i}{w_i\mu_i}\Big) w_i \frac{\partial\mu_i}{\partial\beta_j}$$

där $\partial\mu_i/\partial\beta_1 = \mu_i$, $\partial\mu_i/\partial\beta_2 = \mu_i \mathbb{1}_{\{x_i^{(1)}<a\}}$, $\partial\mu_i/\partial\beta_3 = \mu_i \mathbb{1}_{\{x_i^{(2)}<b\}}$. Sätter vi dessa uttryck lika med noll fås

$$\sum_{i=1}^n (n_i - w_i\mu_i) = 0, \quad \sum_{i=1}^n (n_i - w_i\mu_i)\mathbb{1}_{\{x_i^{(1)}<a\}} = 0, \quad \sum_{i=1}^n (n_i - w_i\mu_i)\mathbb{1}_{\{x_i^{(2)}<b\}} = 0.$$

Uppgift 2

$m_1 = m_2 = 2$. Ordna enligt $(i, j) = (1, 1), (1, 2), (2, 1), (2, 2)$ med $(2, 2)$ som bascell, där $i = 1$ svarar mot $x^{(1)} < 3$, $i = 2$ svarar mot $x^{(1)} > 3$, $j = 1$ svarar mot $x^{(2)} < 30$, $j = 2$ svarar mot $x^{(2)} > 30$. $\beta_1 = \log(\gamma_0)$, $\beta_2 = \log(\gamma_{1,1})$, $\beta_3 = \log(\gamma_{2,1})$, och designmatrisen är (för data aggregerade till tariffceller)

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Uppgift 3

Enligt uppgiften gäller $E[\pi(X)W \mid X] = E[Z \mid X]$ där $Z = \sum_{k=1}^N C_k$ är skadekostnad. P.g.a. antaget oberoende gäller

$$E[Z \mid X] = E[E[Z \mid N, X] \mid X] = E[NE[C] \mid X] = E[C]E[N \mid X].$$

Därför gäller att

$$\begin{aligned}
\pi(x) = \frac{E[Z \mid X = x]}{E[W \mid X = x]} &= E[C]\frac{E[N \mid X = x]}{E[W \mid X = x]} \\
&= E[C]\frac{E[E[N \mid W, X = x] \mid X = x]}{E[W \mid X = x]} \\
&= E[C]\frac{E[W\mu(x) \mid X = x]}{E[W \mid X = x]} \\
&= E[C]\mu(x)
\end{aligned}$$

Uppgift 4

Prediktionsförmågan för modellen som svarar mot parametervärdet λ kvantifieras av

$$c(\lambda) = \sum_{k=1}^{4} w_i(y_k - p(x_k; \widehat{\theta}_{\lambda,k}))^2, \quad \lambda \in \{\lambda_1, \ldots, \lambda_{10}\}.$$

Låt λ^* vara det värde som minimerar $c(\lambda)$. Vi skattar parametervektorn θ som

$$\widehat{\theta} = \text{argmin}_\theta \left(\sum_{i=1}^{4} w_i(y_i - p(x_i; \theta))^2 + \lambda^* \int_{x_{\min}}^{x_{\max}} (p''(x; \theta))^2 dx \right).$$

Uppgift 5

Vi ska välja ett värde x_0 så att $x \leq x_0$ definierar ett löv och $x > x_0$ definierar det andra lövet. Genom att välja $x_0 \in [25, 40)$ hamnar $(12, 25)$ i ena lövet och $(7, 40), (5, 60)$ tillsammans i andra lövet. Ett sådant val ger minimal kvadratsumma (svarar mot val av kvadratisk förlustfunktion)

$$\left(12 - \frac{12}{1}\right)^2 + \left(7 - \frac{7+5}{2}\right)^2 + \left(5 - \frac{7+5}{2}\right)^2$$

och vi får träd-prediktorn (för $x_0 = 25$)

$$T(x) = 12\mathbb{1}_{\{x \leq 25\}} + \frac{7+5}{2}\mathbb{1}_{\{x > 25\}} = 12\mathbb{1}_{\{x \leq 25\}} + 6\mathbb{1}_{\{x > 25\}}$$

Generaliseringsfelet blir då, där $Z \sim N(0, 20)$,

$$\begin{aligned}
E[(Y - T(X))^2] &= \frac{5}{60}E[(10 - 12 + Z)^2] + \frac{5}{60}E[(10 - 6 + Z)^2] + \frac{50}{60}E[(4 - 6 + Z)^2] \\
&= \frac{1}{12}\Big((4 + 20) + (16 + 20) + 10(4 + 20)\Big) \\
&= \frac{300}{12} = \frac{50}{2} = 25
\end{aligned}$$