Stockholms universitet, Matematiska institutionen

Tentamen: MT7028 Prissättning inom sakförsäkring, 2025-02-13, 14–19

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

$$\sum_{i=1}^{n} w_i(n_i/w_i - \mu(x_i;\beta))^2,$$

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, att $n = 4$ och att data (n_i, w_i, x_i), $i = 1, \dots, 4$, ges av

$$(0, 1, (28, 116)), \quad (1, 1, (23, 142)), \quad (0, 1, (58, 86)), \quad (1, 1, (49, 91)),$$

Funktionen $\mu(x;\beta)$ ges av

$$\mu(x;\beta) = \exp\Big(\beta_1 + \beta_2 \mathbb{1}_{\{x^{(1)}<30\}} + \beta_3 \mathbb{1}_{\{x^{(2)}<100\}}\Big)$$

Indikatorvariabeln $\mathbb{1}_{\{x^{(1)}<a\}}$ antar värdet 1 om $x^{(1)} < a$, annars värdet 0. Bestäm explicit 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*, ägarålder, med nivåer *under 30 år* respektive *över 30 år*, samt bostadsort, med nivåer *landsbygd* respektive *stad*. Antag en multiplikativ modell för (förväntad betingad) skadefrekvens. Formulera modellen matematiskt i termer av kategoriska variabler och beskriv hur dessa är relaterade till premieargument och nivåer. (10 p)

Uppgift 3

Betrakta data $\mathcal{D} = \{(w_{j,1}, Y_{j,1}), \ldots, (w_{j,n_j}, Y_{j,n_j}); j = 1, \ldots, J\}$ för ett stort antal J bilmodeller, där $Y_{j,1}$ betecknar det senaste årets riskpremie för modell j. Antag Bühlmann-Straub-modellen vilket innefattar slumpmässiga effekter i form av oobserverbara stokastiska variabler $V_1, \ldots, V_J$ där V_j kan tolkas som den teoretiskt sanna riskpremien om man hade vetat allt om bilmodell j. Idealiskt skulle man vilja bestämma $\mathrm{E}[V_j \mid \mathcal{D}]$ som substitut för det oobserverbara V_j, men istället används kredibilitetsskattaren

$$z_j \overline{Y}_{j,\cdot} + (1 - z_j)\mu, \quad z_j = \frac{w_{j,\cdot}}{w_{j,\cdot} + \sigma^2/\tau^2}.$$

(a) Ange det optimeringsproblem som kredibilitetsskattaren löser. (5 p)

(b) Förklara hur parametrarna σ och τ ska tolkas, eller hur de dyker upp i en matematiskt beskrivning av Bühlmann-Straub-modellen. (5 p)

Uppgift 4

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:

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

där $\mu(x) = \mu(x; \beta_1, \beta_2, \beta_3) > 0$ beror på parametrar $\beta_1, \beta_2, \beta_3$. Givet data i form av observationer (n_i, w_i, x_i), $i = 1, \ldots, n$, ange log-likelihoodfunktionen för skattning av $\beta_1, \beta_2, \beta_3$ och härled ett ekvationssystem som det optimala parametervärdena löser (uttryckt i $\mu(x; \beta_1, \beta_2, \beta_3)$ och lämpliga derivator). (10 p)

Uppgift 5

Antag att vi har observationer

$$(14, 22),\ (8, 44),\ (20, 29),\ (14, 69),\ (13, 46),\ (22, 28),\ (8, 51),\ (15, 88)$$

av par (Y_i, X_i), där Y_i är skadekostnad, X_i är ägarålder. Anpassa ett regressionsträd (enligt standardmetoden ”greedy binary split”) med två löv för prediktion av skadekostnad givet ägarålder baserat på de fyra första paren och skatta generaliseringsfelet baserat på de fyra sista paren. Använd kvadratisk förlustfunktion. (10 p)

Uppgift 1

Med $\mu_i := \mu(x_i;\beta)$ fås

$$\frac{\partial}{\partial \beta_j}\sum_{i=1}^{n} w_i(n_i/w_i - \mu(x_i;\beta))^2 = -2\sum_{i=1}^{n} w_i\Big(n_i/w_i - \mu_i\Big)\frac{\partial \mu_i}{\partial \beta_j}$$

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

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

Med data fås

$$\begin{aligned}
&-\mu_1^2 + (1-\mu_2)\mu_2 - \mu_3^2 + (1-\mu_4)\mu_4 = 0,\\
&-\mu_1^2 + (1-\mu_2)\mu_2 = 0,\\
&-\mu_3^2 + (1-\mu_4)\mu_4 = 0,
\end{aligned}$$

where

$$\mu_1 = e^{\beta_1+\beta_2}, \quad \mu_2 = e^{\beta_1+\beta_2}, \quad \mu_3 = e^{\beta_1+\beta_3}, \quad \mu_4 = e^{\beta_1+\beta_3}.$$

Uppgift 2

Ordna enligt

$$(i,j,k) = (1,1,1),(1,1,2),(1,2,1),(1,2,2),(2,1,1),(2,1,2),(2,2,1),(2,2,2)$$

med $(2,2,2)$ 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 $x^{(2)} > 30$, $k=1$ svarar mot landsbygd, $k=2$ svarar mot stad. $\beta_1 = \log(\gamma_0)$, $\beta_2 = \log(\gamma_{1,1})$, $\beta_3 = \log(\gamma_{2,1})$, $\beta_4 = \log(\gamma_{3,1})$, och designmatrisen är (för data aggregerade till tariffceller)

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

Uppgift 3

Se boken.

Uppgift 4

Log-likelihood-funktionen är, där $\mu_i = \mu(x_i; \beta_1, \beta_2, \beta_3)$,

$$\sum_{i=1}^{n} \big(- w_i\mu_i + n_i \log w_i + n_i \log \mu_i - \log n_i! \big)$$

som maximeras av $\beta_1, \beta_2, \beta_3$ som löser

$$\sum_{i=1}^{n} \left(\frac{n_i}{\mu_i} - w_i \right) \frac{\partial \mu_i}{\partial \beta_j} = 0, \quad j = 1, 2, 3$$

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 fås kvadratsummorna

$$\begin{aligned}
&x_0 \in [22, 29) : 0 + (20-14)^2 + (8-14)^2 + (14-14)^2 = 72 \\
&x_0 \in [29, 44) : (14-17)^2 + (20-17)^2 + (8-11)^2 + (14-11)^2 = 36 \\
&x_0 \in [44, 69) : (14-14)^2 + (20-14)^2 + (8-14)^2 = 72 \\
&x_0 = 69 : (14-14)^2 + (20-14)^2 + (8-14)^2 + (14-14)^2 = 72
\end{aligned}$$

vilket innebär att x_0 väljs $x_0 \in [29, 44)$ vilket ger två löv med vardera två observationer och prediktorn $T(x) = 17\mathbb{1}_{x \leq x_0} + 11\mathbb{1}_{x > x_0}$. Generaliseringsfelet skattas

$$\begin{aligned}
E[(Y - T(X))^2] &\approx \frac{1}{4}\Big((13-11)^2 + (22-17)^2 + (8-11)^2 + (15-11)^2\Big) \\
&= \frac{4 + 25 + 9 + 16}{4} = \frac{54}{4}
\end{aligned}$$