Stockholms universitet, Matematiska institutionen

Tentamen: MT7028 Prissättning inom sakförsäkring, 2023-01-04, 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.

ML-ekvationerna för modellanpassning vid GLM-modellering ges av

$$\sum_{i=1}^{n} w_i \frac{y_i - \mu_i}{g'(\mu_i)v(\mu_i)} x_{ij} = 0, \quad j = 1, \dots, r, \quad g(\mu_i) = \sum_{j=1}^{r} x_{ij}\beta_j$$

Uppgift 1

Antag en multiplikativ modell för förväntad riskpremie där skadefrekvens och medelskada modelleras separat, bägge med multiplikativa modeller. Antag två premieargument, vardera med två nivåer med bascell $(1,1)$ för både skadefrekvens och medelskada. Antag data på formen $(d_i, n_i, z_i, t_i)_{i=1}^m$ där d, n, z, t betecknar duration, antal skador, skadekostnad och kovariatvärden. D.v.s. t_i är $(1,1)$, $(1,2)$, $(2,1)$ eller $(2,2)$.

Ange ett uttryck för skattad förväntad riskpremie för ett kontrakt i tariffcellen $(2,2)$ i termer av lösningar till GLM-modellens specifika ML-ekvationer med storheter angivna utifrån den givna informationen. (10 poäng)

Uppgift 2

Antag en multiplikativ modell för förväntad skadefrekvens och en multiplikativ modell för förväntad riskpremie där endast skadefrekvens beror på kovariater. Antag premieargumenten “stad”(nivåer “Göteborg” och “Stockholm”) och “ålder” (nivåer “ung” och “gammal”). Antag att skadebeloppens fördelning inte beror på premieargumenten. Antag att försäkringsbolaget vill prissätta med GLM. Antag data på formen $(d_i, n_i, z_i, t_i)_{i=1}^m$ där d, n, z, t betecknar duration, antal skador, skadekostnad och kovariatvärden. Antag att (“Göteborg”, “ung”) är bascell och ange skattad förväntad riskpremie för en ung Göteborgare uttryckt i given data. (10 poäng)

Uppgift 3

Betrakta data $\mathcal{D} = \{(w_{j1}, Y_{j1}), \dots, (w_{jn_j}, Y_{jn_j}); j = 1, \dots, J\}$ för ett stort antal J bilmodeller, där Y_{j1} 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

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

Förklara varför detta är en rimlig storhet och redogör för de olika ingående variablerna. (10 poäng)

Uppgift 4

En aktuarie vill anpassa en slät funktion till skadefrekvens som funktion av ägarålder (0 till 100 år) och får tipset att lösa optimeringsproblemet

$$\min_{f \in \mathcal{F}} \left(\sum_{i=1}^{n} w_i d(y_i, f(x_i)) + \lambda \int_0^{100} \left| \frac{d^j f}{dx^j}(x) \right|^k dx \right),$$

där $\mathcal{F}$ betecknar en funktionsklass. Data ges av $(w_i, y_i, x_i)_{i=1}^n$.

(a) Aktuarien väljer $d(y,m) = (y-m)^2$, $j = 1$, $k = 4$, $\mathcal{F}$ som polynom av grad högst 7, men ett alldeles för stort värde för λ. Vilken funktion (approximativt) löser då optimeringsproblemet? (4 poäng)

(b) Efter litteraturstudier väljs istället $\mathcal{F}$ som kubiska splines och lämpligare val görs för j och k: vilka? Hur bör λ väljas? (3 poäng)

(c) Aktuarien provar två varianter vid val av knutpunkter, först $\{0, 25, 60, 100\}$ och sedan $\{0, 4, 8, 12, \ldots, 96, 100\}$ (alla multipler av 4 upp till maxålder). Ingen av varianterna medger perfekt passning till data. På vilket sätt skiljer sig de optimala värdena för λ för de två valen av knutpunkter? (3 poäng)

Uppgift 5

Täthetsfunktionen för en EDM kan skrivas

$$f_Y(y_i; \theta_i, \phi) = \exp\left(\frac{y_i \theta_i - b(\theta_i)}{\phi / w_i} + c(y_i, \phi, w_i) \right)$$

Antag att vi har aggregerad data $(w_i, z_i, x_i)_{i=1}^n$ för n tariffceller där z_i ses som en observation av $Z_i \sim N(w_i \mu(x_i; \beta_1, \ldots, \beta_r), w_i \sigma^2)$ och där $Z_1, \ldots, Z_n$ är oberoende. Låt $Y_i = Z_i / w_i$. Deviansen ges av

$$D(y, \widehat{\mu}) = 2\phi\Big(l(y_1, \ldots, y_n) - l(\widehat{\mu}_1, \ldots, \widehat{\mu}_n) \Big), \quad \widehat{\mu}_i = \mu(x_i; \widehat{\beta}_1, \ldots, \widehat{\beta}_n)$$

där log-likelihood-funktionen för $(Y_1, \ldots, Y_n)$ här uttrycks som funktion av cellmedelvärdena. Visa att fördelningen för Y_i är en viss EDM och att $D(y, \widehat{\mu}) = \sum_{i=1}^n w_i d(y_i, \widehat{\mu}_i)$ för en viss funktion d. (10 poäng)

Uppgift 1

Vi väljer en Tweedie(1)-modell för skadefrekvens och en Tweedie(2)-modell för medelskada.

$$\gamma_{ij} = \gamma_{F,ij}\gamma_{S,ij}, \quad i = 1,2, \quad j = 1,2,$$
$$\gamma_{F,ij} = \gamma_{F,0}\gamma_{F,1i}\gamma_{F,2j}, \quad \gamma_{S,ij} = \gamma_{S,0}\gamma_{S,1i}\gamma_{S,2j}$$

För listform väljs $\beta_{F,1} = \log\gamma_{F,0}$, $\beta_{F,2} = \log\gamma_{F,12}$, $\beta_{F,3} = \log\gamma_{F,22}$, och på samma sätt för modellen för medelskada, med designmatris (x_{ij}), $i = 1,\dots,m$, $j = 1,2,3$, med rad $(1,0,0)$ om $t_i = (1,1)$, rad $(1,0,1)$ om $t_i = (1,2)$, rad $(1,1,0)$ om $t_i = (2,1)$, rad $(1,1,1)$ om $t_i = (2,2)$. Regressionskoefficienterna löser

$$\sum_{i=1}^{m} d_i(n_i/d_i - \mu_{F,i})x_{ij} = 0, \quad j = 1,2,3, \quad \mu_{F,i} = \exp\Big(\sum_{j=1}^{3} x_{ij}\beta_{F,j}\Big)$$

och

$$\sum_{i=1,n_i\neq 0}^{m} n_i\frac{z_i/n_i - \mu_{S,i}}{\mu_{S,i}}x_{ij} = 0, \quad j = 1,2,3, \quad \mu_{S,i} = \exp\Big(\sum_{j=1}^{3} x_{ij}\beta_{S,j}\Big)$$

Den sökta skattade förväntade riskpremien för cell (2,2) är

$$\exp\Big(\beta_{F,1} + \beta_{F,2} + \beta_{F,3}\Big)\exp\Big(\beta_{S,1} + \beta_{S,2} + \beta_{S,3}\Big)$$

Uppgift 2

Vi väljer en Tweedie(1)-modell för skadefrekvens.

$$\gamma_{ij} = \gamma_{F,ij}\gamma_{S,0} = \gamma_{F,0}\gamma_{F,1i}\gamma_{F,2j}\gamma_{S,0}, \quad i = 1,2, \quad j = 1,2.$$

För listform väljs $\beta_{F,1} = \log\gamma_{F,0}$, $\beta_{F,2} = \log\gamma_{F,12}$, $\beta_{F,3} = \log\gamma_{F,22}$, med designmatris (x_{ij}), $i = 1,\dots,m$, $j = 1,2,3$, med rad $(1,0,0)$ om $t_i = (G,u)$, rad $(1,0,1)$ om $t_i = (G,g)$, rad $(1,1,0)$ om $t_i = (S,u)$, rad $(1,1,1)$ om $t_i = (S,g)$. Regressionskoefficienterna löser

$$\sum_{i=1}^{m} d_i(n_i/d_i - \mu_{F,i})x_{ij} = 0, \quad j = 1,2,3, \quad \mu_{F,i} = \exp\Big(\sum_{j=1}^{3} x_{ij}\beta_{F,j}\Big)$$

och speciellt gäller (för $j = 1$) att $\sum_{i=1}^{m} n_i = e^{\beta_{F,0}}\sum_{i=1}^{m} d_i$. Vidare gäller

$$\gamma_{S,0} = \frac{\sum_{i=1}^{m} z_i}{\sum_{i=1}^{m} n_i}$$

Den skattade förväntade riskpremien för en ung Göteborgare blir

$$\exp(\beta_{F,0})\gamma_{S,0} = \frac{\sum_{i=1}^{m} n_i}{\sum_{i=1}^{m} d_i}\frac{\sum_{i=1}^{m} z_i}{\sum_{i=1}^{m} n_i} = \frac{\sum_{i=1}^{m} z_i}{\sum_{i=1}^{m} d_i}$$

Uppgift 3

Se boken, kapitel 4.

Uppgift 4

(a) Om vi låter $f_\lambda(x)$ beteckna funktionen som löser optimeringsproblemet givet λ, då fås $\lim_{\lambda\to\infty} f_\lambda(x) = f_\infty(x) = c$ där c är det tal som löser

$$\min_c \sum_{i=1}^{n} w_i(y_i - c)^2.$$

Derivera m.a.p. c, sätt lika med 0 och lös ut c. Detta ger

$$c = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$

(b) $j = 2$, $k = 2$, λ väljs genom korsvalidering

(c) Få knutpunkter ger mindre flexibilitet och därgenom mindre behov av att straffa potentiell överanpassning, dvs korsvalidering bör ge ett mindre λ för få knutpunkter jämfört med att stort antal knutpunkter.

Uppgift 5

Eftersom $Y_i \sim N(\mu(x_i;\beta_1,\ldots,\beta_r),\sigma^2/w_i)$ fås, med $\mu_i = \mu(x_i;\beta_1,\ldots,\beta_r)$,

$$\begin{aligned} f_{Y_i}(y_i;\mu_i,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2/w_i}} \exp\Big(-(y_i-\mu_i)^2/(2\sigma^2/w_i)\Big) \\ &= \exp\left(\frac{y_i\mu_i - \mu_i^2/2}{\sigma^2/w_i} - y_i^2/(2\sigma^2/w_i) - \log\left(\sqrt{2\pi\sigma^2/w_i}\right)\right) \end{aligned}$$

d.v.s. EDM med $\theta_i = \mu_i$, $b(\theta_i) = \mu_i^2/2$ och $\phi = \sigma^2$. Vidare fås

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{w_i}} \exp\Big(-(y_i-\mu_i)^2/(2\sigma^2/w_i)\Big)$$

och

$$l(\mu) = \sum_{i=1}^{n}\left(-\frac{w_i}{2\sigma^2}(y_i-\mu_i)^2 - \log\left(\sqrt{2\pi}\frac{\sigma}{\sqrt{w_i}}\right)\right)$$

vilket ger $D(y,\widehat{\mu}) = \sum_{i=1}^{n} w_i(y_i - \widehat{\mu}_i)^2$