Stockholms universitet, Matematiska institutionen

Tentamen: MT7028 Prissättning inom sakförsäkring, 2023-02-24, 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.

ML-ekvationerna för modellanpassning vid GLM-modellering ges av

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

$I\{H\}$ tar värdet 1 om händelsen H är sann, och värdet 0 annars.

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 $(2,1)$ i termer av lösningar till GLM-modellens specifika ML-ekvationer med storheter angivna utifrån den givna informationen. (10 poäng)

Uppgift 2

Antag att antalet skador för en försäkringprodukt under ett försäkringsår är Poissonfördelat med väntevärde $\lambda(x) = \exp(\beta_1 + \beta_2 1\{x = 1\})$, där $x \in \{0,1\}$. Data är på formen $\{(x_i, z_i) : i = 1, \ldots, n\}$, där z_i är antalet skador för kontrakt i. Härled ekvationssystemet i variablerna β_1 och β_2 som ger ML-skattningen för β_1 och β_2. (10 poäng)

Uppgift 3

Antag endast två premieargument: "fordonsålder"(x) och "ägarålder"(y) som bägge ses som numeriska variabler. Antag en modell för förväntad skadefrekvens på formen $\lambda(x,y) = \sum_{j=1}^{J} c_j I\{(x,y) \in R_j\}$. Antag en GLM-analys med log-länk och kodning i kategoriska variabler av x och y enligt intervall

$$A_{x,1} = (0, 1], \quad A_{x,2} = (1, 5], \quad A_{x,3} = (5, \infty),$$
$$A_{y,1} = (0, 20], \quad A_{y,2} = (20, 30], \quad A_{y,3} = (30, 70], \quad A_{y,4} = (70, \infty),$$

med bascell $x \in (1, 5]$ och $y \in (30, 70]$, resulterat i regressionskoefficienter $\widehat{\beta}_i$ (intercept) och $\widehat{\beta}_{z,k}$ där $z \in \{x, y\}$ och k är heltalsvärd. Uttryck $\lambda(x, y)$ givet denna information: bestäm J och c_j, R_j för $j = 1, \ldots, J$. (10 poäng)

Uppgift 4

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 poäng)

(b) Antag att $w_{j,t}$ är approximativt av samma storlek för alla j, t. Antag att $n_1 = n_2 = 10$ och illustrera kvalitativt i en figur med datapunkterna $Y_{1,1}, \ldots, Y_{1,10}$ (med symbolen $\circ$) och $Y_{2,1}, \ldots, Y_{2,10}$ (med symbolen $\times$) de två situationerna "stort σ och litet τ" och "litet σ och stort τ" (markera datapunkter på tallinje, en tallinje för vardera situationen). (5 poäng)

Uppgift 5

Antag endast två premieargument: "fordonsålder"(x) och "ägarålder"(y) som bägge ses som numeriska variabler. Antag en modell för förväntad skadefrekvens på formen $\lambda(x, y) = \sum_{j=1}^{J} c_j I\{(x, y) \in R_j\}$. Antag att ett regressionsträd anpassat till skadefrekvensdata givit

$$\lambda(x, y) = 0.004 I\{x \geq 13, y \geq 31\} + 0.001 I\{0 < x < 13, y \geq 31\}$$
$$+ 0.03 I\{0 < y < 31\}.$$

(a) Rita regressionsträdet. (5 poäng)

(b) Betrakta en GLM-modell med log-länk och kodning i kategorivariabler enligt intervall

$$A_{x,1} = (0, 13), \quad A_{x,2} = [13, \infty), \quad A_{y,1} = (0, 31), \quad A_{y,2} = [31, \infty).$$

Bascell väljs som $x \geq 13$ och $y \geq 31$. Givet att $\lambda(x, y)$ är sann, ange möjliga värden för regressionskoefficienter för GLM-modellen. (5 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 (2,1) är

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

Uppgift 2

Låt $p_z = \exp(-\lambda(x))\lambda(x)^z/z!$. Loglikelihood-funktionen är

$$\begin{aligned}\sum_{i=1}^{n} \log p_{z_i} &= \sum_{i=1}^{n}\Big(-\lambda(x_i) + z_i \log \lambda(x_i) - z_i!\Big)\\ &= \sum_{i=1}^{n}\Big(-e^{\beta_1+\beta_2 I\{x_i=1\}} + z_i(\beta_1 + \beta_2 I\{x_i = 1\}) - z_i!\Big)\end{aligned}$$

vars partialderivata map β_1 är

$$\sum_{i=1}^{n} z_i - \sum_{i=1}^{n} e^{\beta_1+\beta_2 I\{x_i=1\}}$$

och vars partialderivata map β_2 är

$$\sum_{i=1}^{n} z_i I\{x_i = 1\} - \sum_{i=1}^{n} e^{\beta_1+\beta_2} I\{x_i = 1\}.$$

sätter vi dessa uttryck lika med noll fås ekvationssystemet för ML-skattning av β_1, β_2.

Uppgift 3

$J = 12$, R_j är på formen $B_{k,l} = A_{x,k} \times A_{y,l}$, $k = 1, \dots, 3$, $l = 1, \dots, 4$.

$$\lambda(x, y) = \sum_{k=1}^{3} \sum_{l=1}^{4} e^{\beta_i + \beta_{x,k} + \beta_{y,l}} I\{(x, y) \in B_{k,l}\},$$

där $\beta_{x,2} = \beta_{y,3} = 0$. Vi kan ordna enligt

$$\begin{aligned}
&(c_1, R_1) = (e^{\beta_i + \beta_{x,1} + \beta_{y,1}}, B_{1,1}), \quad (c_2, R_2) = (e^{\beta_i + \beta_{x,1} + \beta_{y,2}}, B_{1,2}),\\
&(c_3, R_3) = (e^{\beta_i + \beta_{x,1} + \beta_{y,3}}, B_{1,3}), \quad (c_4, R_4) = (e^{\beta_i + \beta_{x,1} + \beta_{y,4}}, B_{1,4}),\\
&(c_5, R_5) = (e^{\beta_i + \beta_{x,2} + \beta_{y,1}}, B_{1,1}), \quad (c_6, R_6) = (e^{\beta_i + \beta_{x,2} + \beta_{y,2}}, B_{1,2}),\\
&(c_7, R_7) = (e^{\beta_i + \beta_{x,2} + \beta_{y,3}}, B_{1,3}), \quad (c_8, R_8) = (e^{\beta_i + \beta_{x,2} + \beta_{y,4}}, B_{1,4}),\\
&(c_9, R_9) = (e^{\beta_i + \beta_{x,3} + \beta_{y,1}}, B_{1,1}), \quad (c_{10}, R_{10}) = (e^{\beta_i + \beta_{x,3} + \beta_{y,2}}, B_{1,2}),\\
&(c_{11}, R_{11}) = (e^{\beta_i + \beta_{x,3} + \beta_{y,3}}, B_{1,3}), \quad (c_{12}, R_{12}) = (e^{\beta_i + \beta_{x,3} + \beta_{y,4}}, B_{1,4}).
\end{aligned}$$

Uppgift 4

(a) Kredibilitetsskattaren är den linjära funktion $h(Y) = \sum_{k=1}^{J} \sum_{t=1}^{n_k} a_{k,t} Y_{k,t}$ som minimerar $\mathrm{E}[(h(Y) - V_j)^2]$.

(b) τ kan ses som ett mått på förväntat avstånd mellan stickprovsmedelvärden $\overline{Y}_{j,\cdot}$, $j = 1, \dots, J$. σ kan ses som ett mått på förväntad spridning inom stickproven $\{Y_{j,1}, \dots, Y_{j,n_j}\}$, $j = 1, \dots, J$.

Uppgift 5

Regressionsträdet:

$$\begin{aligned}
\lambda_R(x, y) = {} & 0.004 I\{(x, y) \in A_{x,2} \times A_{y,2}\} + 0.001 I\{(x, y) \in A_{x,1} \times A_{y,2}\}\\
& + 0.03 I\{(x, y) \in (A_{x,1} \cup A_{x,2}) \times A_{y,1}\}
\end{aligned}$$

GLM-modellen:

$$\begin{aligned}
\lambda_G(x, y) = {} & e^{\beta_i} I\{(x, y) \in A_{x,2} \times A_{y,2}\} + e^{\beta_i + \beta_{x,1}} I\{(x, y) \in A_{x,1} \times A_{y,2}\}\\
& + e^{\beta_i + \beta_{y,1}} I\{(x, y) \in A_{x,2} \times A_{y,1}\} + e^{\beta_i + \beta_{x,1} + \beta_{y,1}} I\{(x, y) \in A_{x,1} \times A_{y,1}\}
\end{aligned}$$

Vi noterar att $e^{\beta_i} = 0.004$ och $e^{\beta_i + \beta_{x,1}} = 0.001$, och att $\beta_{x,1} < 0$. För konsistens måste följande gälla:

$$e^{\beta_i + \beta_{x,1} + \beta_{y,1}} \leq 0.03 \leq e^{\beta_i + \beta_{y,1}}.$$

Den första olikheten kan skrivas $e^{\beta_{y,1}} \leq 30$. Den andra olikheten kan skrivas $e^{\beta_{y,1}} \geq 0.03/0.004 = 30/4$. Alltså: Ja, det existerar regressionskoefficienter så att en multiplikativ modell kan vara konsistent med regressionsträdet.