Stockholms Universitet, Matematiska Institutionen

Tentamen: Prissättning inom sakförsäkring, MT7028

30 januari 2019 9–14

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

ML-ekvationerna på allmän form för skattning av regressionskoefficienter för GLM med länkfunktion g baserat på EDM med variansfunktion v ges av

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

Uppgift 1

Försäkringsföretaget JanuariFörsäkring är bara öppet för tecknande av 1-åriga försäkringar under januari varje år och tecknade försäkringar kan inte sägas upp under året. Kunder som önskar teckna försäkring måste meddela detta redan i december och premien sätts först i januari. Företaget har en tariff bestående av tre tariffceller $k = 1, 2, 3$. Som grund för tariffen har JanuariFörsäkrings aktuarie en stokastisk modell för antal skador och skadebelopp för ett givet år. För cell k är N_k antalet skador och $S_k = X_{k,1} + \cdots + X_{k,N_k}$ totala skadebeloppet. Skadebeloppen $X_{k,1}, X_{k,2}, \dots$ antas oberoende och likafördelade och oberoende av N_k. Skadebelopp och antal skadehändelser för olika celler antas oberoende. Under januari 2019 har 20, 50,100 försäkringar tecknats för de respektive tariffcellerna. Det antas att

$$\mathrm{E}[N_k] = \mu_k^{(F)} d_k, \ \mathrm{Var}(N_k) = \phi^{(F)} \mu_k^{(F)} d_k$$

och

$$\mathrm{E}[X_{k,1}] = \mu_k^{(S)}, \ \mathrm{Var}(X_{k,1}) = \phi^{(S)} (\mu_k^{(S)})^2,$$

där d_k betecknar duration (försäkringsår). Alla parametrar antas kända då de skattats baserat på historisk data bestående av 200, 300, 400 1-åriga försäkringar. Premien för en försäkring hörande till tariffcell k är vald som väntevärde plus en standardavvikelse för riskpremien total skadekostnad delat med total duration för cell k. Beräkna premien för en försäkringstagare som just tecknat en försäkring och hör till tariffcell 2. (10 poäng)

Uppgift 2

Aktuarien på försäkringsföretaget JanuariFörsäkring tycker att premien bör sättas enligt varje försäkringstagares bidrag till den totala risken, här svarande mot väntevärde plus en standardavvikelse av det totala skadebeloppet (för alla tre kategorier av försäkringstagare tillsammans). Därför sätts premien enligt p_k/d_k där

$$p_k = \frac{\partial f}{\partial x_k}(x)\Big|_{x_1=x_2=x_3=1},$$

där

$$f(x) = \mathrm{E}[S(x)] + \mathrm{Var}(S(x))^{1/2}, \quad S(x) = x_1S_1 + x_2S_2 + x_3S_3.$$

Beräkna premien för en försäkringstagare som just tecknat en försäkring och hör till tariffcell 2 i termer av $\mathrm{E}[S_k]$ och $\mathrm{Var}(S_k)$ och, vid behov, storheter givna i Uppgift 1, för $k = 1, 2, 3$. (10 poäng)

Uppgift 3

Antag att vi är intresserade av prissättning proportionell mot förväntad skadefrekvens genom GLM analys med länkfunktion $x \mapsto x^3$ då antal skadehändelser olika år antas vara oberoende och Poissonfördelade. Låt $n_{i,j}$ och $d_{i,j}$ beteckna antalet historiska skadehändelser och motsvarande duration för premieargument nr 1 i klass i och premieargument nr 2 i klass j. Antag att vi betraktar två premieargument med 3 respektive 3 klasser. Bestäm storheterna explicit som ingår i det allmänna uttrycket för ML-ekvationerna då $(3, 2)$ är bascell. Kan vi uttrycka premien multiplikativt i termer av relationstal? (10 poäng)

Uppgift 4

Ett försäkringsföretag har utifrån en GLM-analys skapat en tariff baserat på förväntad riskpremie. Tariffen baseras på två multiplikativa Tweedie(p)-modeller med $p = 1$ respektive $p = 2$ för skadefrekvens och medelskada. För bägge nyckeltalen görs samma indelning i tre premieargument med, respektive, 2, 3 och 7 klasser. Skadedata har gjort troligt att premieargument (och klasser) endast behövs för nyckeltalet skadefrekvens, dvs att tariffen kan förenklas. För att undersöka om tariffen kan förenklas beräknas teststorheten

$$\frac{1}{\widehat{\phi}_X}\big(D(y,\widehat{\mu}) - D(y,\widehat{\mu}^*)\big),$$

där $D(y,\widehat{\mu}^*)$ betecknar deviansen mellan den mättade modellen och den ursprungliga modellen, och $D(y,\widehat{\mu})$ betecknar deviansen mellan den mättade modellen och den föreslagna enklare modellen. Vad är y i teststorheten? Vilken stokastisk modell svarar D i teststorheten mot? Vad ska teststorheten jämföras mot och hur för att avgöra om försäkringsföretaget bör gå vidare med den föreslagna enklare modellen? Ett felaktigt beslut pga ren otur accepteras om sannolikheten för detta är högst av storleksordningen 5%. (10 poäng)

Uppgift 5

Antag att vi är intresserade av prissättning proportionell mot förväntad skadefrekvens genom GLM analys med loglänkfunktion då antal skadehändelser antas Poissonfördelade, med vanliga oberoendeantaganden för händelser över tid och för olika försäkringstagare. Antag att vi betraktar två premieargument med 2 respektive 2 klasser. Antag vidare historisk data i form av en lång lista baserad på kundnummer enligt tabellen. Bevisa eller ge ett motexempel till påståendet att ML-skattningarna av relationstalen inte påverkas om historisk data aggregeras till tariffceller. (10 poäng)

Class	Age	Duration	N. of claims	Claim cost
1	1	1.6	0	0
2	1	0.8	0	0
2	1	2.1	1	8352
1	2	3.1	0	0
1	2	2.9	2	12171
2	2	2.2	0	0
2	1	1.6	0	0
⋮	⋮	⋮	⋮	⋮

Uppgift 1

Premien:

$$\begin{aligned}
&\mathrm{E}[S_k/d_k] + \mathrm{Var}(S_k/d_k)^{1/2} \\
&\quad = \frac{1}{d_k}\Big(\mathrm{E}[S_k] + \mathrm{Var}(S_k)^{1/2}\Big) \\
&\quad = \frac{1}{d_k}\Big(\mathrm{E}[N_k]\,\mathrm{E}[X_{k,1}] + \Big(\mathrm{Var}(N_k)\,\mathrm{E}[X_{k,1}]^2 + \mathrm{E}[N_k]\,\mathrm{Var}(X_{k,1})\Big)^{1/2}\Big) \\
&\quad = \frac{1}{d_k}\Big(d_k\mu_k^{(F)}\mu_k^{(S)} + \Big(\phi^{(F)}\mu_k^{(F)}(\mu_k^{(S)})^2 + \phi^{(S)}\mu_k^{(F)}(\mu_k^{(S)})^2\Big)^{1/2}\Big) \\
&\quad = \mu_k^{(F)}\mu_k^{(S)} + \frac{1}{d_k^{1/2}}\Big((\phi^{(F)} + \phi^{(S)})\mu_k^{(F)}(\mu_k^{(S)})^2\Big)^{1/2},
\end{aligned}$$

där $d_1 = 20$, $d_2 = 50$, $d_3 = 100$. Den sökta premien svarar mot $k = 2$.

Uppgift 2

Vi ska beräkna p_k/d_k där

$$p_k = \frac{\partial f}{\partial x_k}(x)\Big|_{x=1}, \quad f(x) = \mathrm{E}[S(x)] + \mathrm{Var}(S(x))^{1/2}, \quad S(x) = \sum_{k=1}^{3} x_k S_k.$$

Notera att

$$\begin{aligned}
\frac{\partial f}{\partial x_k}(x) &= \mathrm{E}[S_k] + \frac{1}{2\,\mathrm{Var}(S(x))^{1/2}}\frac{\partial}{\partial x_k}\mathrm{Var}(S(x)) \\
&= \mathrm{E}[S_k] + \frac{1}{2\,\mathrm{Var}(S(x))^{1/2}}\frac{\partial}{\partial x_k}\sum_{l=1}^{n}\sum_{m=1}^{n} x_l x_m \,\mathrm{Cov}(S_l, S_m) \\
&= \mathrm{E}[S_k] + \frac{1}{\mathrm{Var}(S(x))^{1/2}}\sum_{l=1}^{n} x_l \,\mathrm{Cov}(S_k, S_l) \\
&= \mathrm{E}[S_k] + \frac{\mathrm{Cov}(S_k, S(x))}{\mathrm{Var}(S(x))^{1/2}}.
\end{aligned}$$

Alltså fås

$$p_k = \mathrm{E}[S_k] + \frac{\mathrm{Cov}(S_k, S)}{\mathrm{Var}(S)^{1/2}}.$$

Eftersom S_1, S_2, S_3 antas oberoende fås

$$p_k = \mathrm{E}[S_k] + \frac{\mathrm{Var}(S_k)}{\mathrm{Var}(S)^{1/2}}.$$

Uppgift 3

Vi har modell på formen $g(\mu_{ij}) = \gamma_0 + \gamma_{1i} + \gamma_{2j}$, $i, j = 1, \ldots, 3$. Eftersom $(3, 2)$ är bascell sätts $\gamma_{13} = \gamma_{22} = 0$. Totalt har vi 5 regressionsparametrar $\beta_1 = \gamma_0$, $\beta_2 = \gamma_{11}$, $\beta_3 = \gamma_{12}$, $\beta_4 = \gamma_{21}$, $\beta_5 = \gamma_{23}$ vilket ger, på listform, $g(\mu_i) = \sum_{j=1}^{5} x_{ij}\beta_j$ med designmatris $[x_{ij}]$ av dimension 9×5 enligt

$$\left[\begin{array}{c|ccccc}
(1,1) & 1 & 1 & 0 & 1 & 0 \\
(1,2) & 1 & 1 & 0 & 0 & 0 \\
(1,3) & 1 & 1 & 0 & 0 & 1 \\
(2,1) & 1 & 0 & 1 & 1 & 0 \\
(2,2) & 1 & 0 & 1 & 0 & 0 \\
(2,3) & 1 & 0 & 1 & 0 & 1 \\
(3,1) & 1 & 0 & 0 & 1 & 0 \\
(3,2) & 1 & 0 & 0 & 0 & 0 \\
(3,3) & 1 & 0 & 0 & 0 & 1
\end{array}\right]$$

Den "vanliga" log-länkfunktionen $\widetilde{g} = \log$ ger $\mu_i = \exp\{\sum_{j=1}^{5} x_{ij}\beta_j\}$ vilket är ett uttryck för μ_i som en produkt av relationstal. Med $g(x) = x^3$ fås $\mu_i = (\sum_{j=1}^{5} x_{ij}\beta_j)^{1/3}$ vilket inte svarar mot en multiplikativ modell.

Uppgift 4

Vi ska testa modellen för medelskada, en Tweediemodell med $p = 2$ ("Gamma"). Modeller $H_s \subset H_r$, $s < r$, dvs H_s är ett specialfall av den allmännare modellen H_r. H_s har s fria β-parametrar. Hypotesen är "Data kommer från modellen H_s". Om hypotesen sann gäller att teststorheten LRT, en likelihood-kvot, är approximativt $\chi^2(r - s)$-fördelad. Om LRT överstiger $F_{\chi^2(r-s)}(0.95)$ är det antingen ett mycket osannolikt datamaterial eller, mer troligt, så är hypotesen falsk och förkastas på signifikansnivån 5%. Alltså accepteras sannolikhet 5% för att förkasta en sann hypotes.

Totalt har vi för den stora modellen H_r, $r = 1 + (2 - 1) + (3 - 1) + (7 - 1) = 10$ regressionskoefficienter. H_s svarar mot att vi plockar bort samtliga premieargument vilket ger H_s med $s = 1$, och därmed $r - s = 9$.

Uppgift 5

Skriv $\widetilde{w}_k$, $\widetilde{x}_k$ och $\widetilde{y}_k = \widetilde{x}_k/\widetilde{w}_k$ för storheterna som svarar mot data som inte är aggregerad enligt tariffceller. Låt $\mathcal{I}$ vara en indexmängd som svarar mot radnummer för en given kovariatkombination, tex premieargument 1 i klass 1 och premieargument 2 i klass 1, dvs paret $(1, 1)$. Notera att designmatrisen av dimension $m \times 3$ för icke-aggregerad data innehåller raderna i designmatrisen av dimension 4×3 för den aggregerade datan, men med flera identiska rader. De 4 raderna svarar mot cellerna $(1, 1), (1, 2), (2, 1), (2, 2)$.

Notera att $\widetilde{\mu}_k = \widetilde{\mu}_{k'}$ och $\widetilde{x}_{k,j} = \widetilde{x}_{k',j}$ för $k, k' \in \mathcal{I}$. Alltså kan ML-ekvationerna $\partial l/\partial\beta_j = 0$ skrivas

$$\begin{aligned}
0 &= \frac{1}{\phi}\sum_{k=1}^{m}\widetilde{w}_k\frac{\widetilde{y}_k - \widetilde{\mu}_k}{\widetilde{\mu}_k^{p-1}}\widetilde{x}_{k,j} \\
&= \frac{1}{\phi}\sum_{k=1,k\notin\mathcal{I}}^{m}\widetilde{w}_k\frac{\widetilde{y}_k - \widetilde{\mu}_k}{\widetilde{\mu}_k^{p-1}}\widetilde{x}_{k,j} + \frac{1}{\phi}\sum_{k\in\mathcal{I}}\widetilde{w}_k\frac{\widetilde{y}_k - \widetilde{\mu}_k}{\widetilde{\mu}_k^{p-1}}\widetilde{x}_{k,j}
\end{aligned}$$

Skriv $\widetilde{\mu}_{\mathcal{I}}$ och $\widetilde{x}_{\mathcal{I},j}$ för $\widetilde{\mu}_k$ och $\widetilde{x}_{k,j}$ för k sådant att $k \in \mathcal{I}$:

$$\begin{aligned}
\frac{1}{\phi}\sum_{k\in\mathcal{I}}\widetilde{w}_k\frac{\widetilde{y}_k - \widetilde{\mu}_k}{\widetilde{\mu}_k^{p-1}}\widetilde{x}_{k,j} &= \frac{\widetilde{x}_{\mathcal{I},j}}{\phi\widetilde{\mu}_{\mathcal{I}}^{p-1}}\sum_{k\in\mathcal{I}}(\widetilde{x}_k - \widetilde{w}_k\widetilde{\mu}_{\mathcal{I}}) \\
&= \frac{\widetilde{x}_{\mathcal{I},j}}{\phi\widetilde{\mu}_{\mathcal{I}}^{p-1}}\Big(\sum_{k\in\mathcal{I}}\widetilde{x}_k - \widetilde{\mu}_{\mathcal{I}}\sum_{k\in\mathcal{I}}\widetilde{w}_k\Big) \\
&= \frac{1}{\phi}w_{\mathcal{I}}\frac{y_{\mathcal{I}} - \widetilde{\mu}_{\mathcal{I}}}{\widetilde{\mu}_{\mathcal{I}}^{p-1}}\widetilde{x}_{\mathcal{I},j},
\end{aligned}$$

där

$$w_{\mathcal{I}} = \sum_{k\in\mathcal{I}}\widetilde{w}_k, \quad y_{\mathcal{I}} = \frac{1}{w_{\mathcal{I}}}\sum_{k\in\mathcal{I}}\widetilde{x}_k.$$

På samma sätt för alla 4 indexmängder $\mathcal{I}_k$, $k = 1, \ldots, 4$ svarande mot, respektive, $(1,1), (1,2), (2,1), (2,2)$. Alltså:

$$\begin{aligned}
\frac{1}{\phi}\sum_{k=1}^{m}\widetilde{w}_k\frac{\widetilde{y}_k - \widetilde{\mu}_k}{\widetilde{\mu}_k^{p-1}}\widetilde{x}_{k,j} &= \frac{1}{\phi}\sum_{k=1}^{4}w_{\mathcal{I}_k}\frac{y_{\mathcal{I}_k} - \widetilde{\mu}_{\mathcal{I}_k}}{\widetilde{\mu}_{\mathcal{I}_k}^{p-1}}\widetilde{x}_{\mathcal{I}_k,j} \\
&= \frac{1}{\phi}\sum_{k=1}^{4}w_k\frac{y_k - \mu_k}{\mu_k^{p-1}}x_{k,j}
\end{aligned}$$

vilket visar att ML-ekvationerna för β-parametrarna sammanfaller i de två fallen med aggregerad respektive icke-aggregerad data.