Stockholms Universitet, Matematiska Institutionen

Tentamen: Prissättning inom sakförsäkring, MT7028

22 augusti 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

Givet w_1, w_2, låt X_1, X_2 vara Poissonfördelade med parametrar $w_1\mu_1, w_2\mu_2$. Visa att sannolikhetsfunktionen för $Y_i = X_i/w_i$ kan skrivas på formen av en EDM

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

för y_i så att $w_i y_i$ är ett icke-negativt heltal. Identifiera storheterna θ_i, ϕ och funktionerna b, c. (5 poäng)

Härled ett uttryck för väntevärdet för en EDM med sannolikhetsfunktion/täthetsfunktion på formen ovan. (5 poäng)

Uppgift 2

Tabell 1 visar historisk skadedata enligt en uppdelning i tariffceller. Bestäm så explicit som möjligt ML-ekvationerna för skattning av förväntad riskpremie (pure premium) i samband med en GLM-analys baserad på en multiplikativ Tweedie(p) modell.

Uppgift 3

Betrakta en GLM-analys av skadefrekvens baserad på 3 premieargument med 2, 2 respektive 20 klasser med data på formen y_{ijk} (observerad skadefrekvens) och w_{ijk} (duration) för $i = 1, 2$, $j = 1, 2$, $k = 1, \ldots, 20$. Antag en multiplikativ modell för förväntad skadefrekvens där relationstalen som svarar mot premieargumentet med flest klasser (bilmodell) antas kända medan de övriga ska skattas. Antag en lämplig Tweedie-modell för skadefrekvens och ange ML-ekvationerna så explicit som möjligt. (10 poäng)

Uppgift 4

En GLM-analys har gjorts för att skapa en multiplikativ Tweedie(p) modell för förväntad medelskada. Efter analysen studeras Pearson residualerna

$$r_{Pi} = \frac{y_i - \widehat{\mu}_i}{\sqrt{v(\widehat{\mu}_i)/w_i}},$$

där y_i betecknar observerad medelskada för tariffcell i, $\widehat{\mu}_i$ motsvarande skattad förväntad medelskada, v variansfunktionen för Tweedie(p) modellen och w_i duration. Figur 1 visar talparen $(\widehat{\mu}_i, r_{Pi})$. Figuren antyder att GLM-analysen bör göras om med ett ändrat val av p. Förklara hur och hur detta antyds av figuren. (10 poäng)

Uppgift 5

Ett försäkringsföretag har utifrån en GLM-analys skapat en tariff baserat på förväntad riskpremie. Tariffen baseras på en multiplikativ Tweedie(p)-modell för riskpremie med 3 premieargument med, respektive, 2, 8 och 4 klasser. Skadedata har gjort troligt att det tredje premieargumentet inte påverkar förväntad riskpremie, 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? 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)

Klass	Ålder	Zon	Duration	Antal skador	Skadekostnad
1	1	1	62.9	17	310 352
1	1	2	112.9	7	95 424
1	2	1	352.1	52	428 064
1	2	2	840.1	69	511 842
2	1	1	191.6	43	333 422
2	1	2	237.3	34	235 722
2	2	1	844.8	94	444 432
2	2	2	1296.0	99	420 948

Table 1: Skadedata, duration i år, skadekostnad i kronor

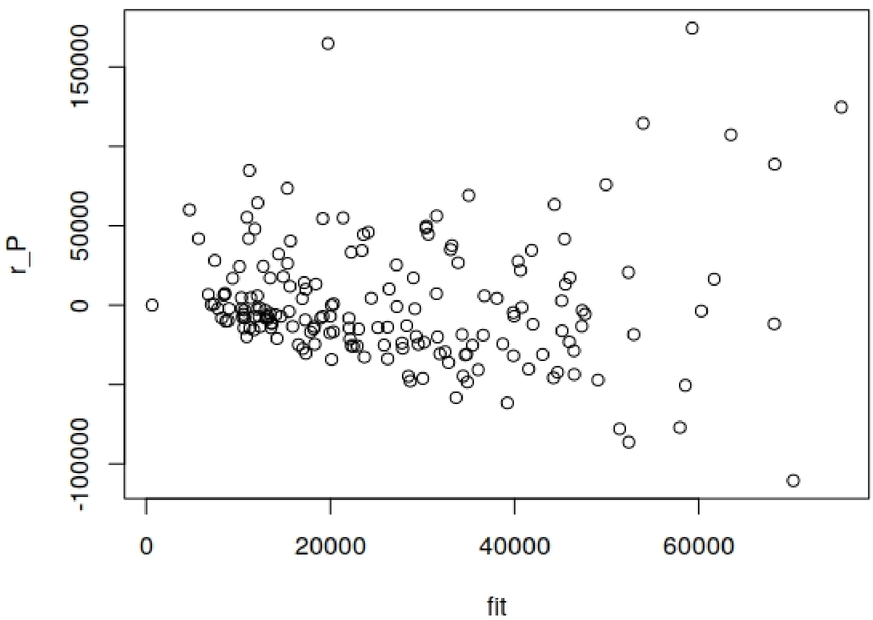

Figure 1: Pearson residualer $(\widehat{\mu}_i, r_{Pi})$

Uppgift 1

Svar finns i kursboken.

Uppgift 2

$\beta_1 = \log \gamma_0$, $\beta_2 = \log \gamma_{11}$, $\beta_3 = \log \gamma_{21}$, $\beta_4 = \log \gamma_{31}$ och (tex)

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

$r = 4$, $n = 8$, $v(x) = x^p$, $g(x) = \log(x)$, $g'(x) = x^{-1}$, $g^{-1}(x) = \exp(x)$. ML-ekvationerna ges av

$$\sum_{i=1}^{n} w_i \frac{y_i - \mu_i}{\mu_i^{p-1}} x_{ik} = 0, \quad k = 1, \ldots, r, \quad \mu_i = \exp\Big(\sum_{j=1}^{r} \beta_j x_{ij}\Big).$$

Låt w_i and c_i beteckna värdena i 4:e och 6:e kolumnen i Tabell 1. ML-ekvationerna ges av

$$\begin{aligned}
&\frac{c_1 - w_1\mu_1}{\mu_1^{p-1}} + \cdots + \frac{c_8 - w_8\mu_8}{\mu_8^{p-1}} = 0, \\
&\frac{c_1 - w_1\mu_1}{\mu_1^{p-1}} + \frac{c_2 - w_2\mu_2}{\mu_2^{p-1}} + \frac{c_3 - w_3\mu_3}{\mu_3^{p-1}} + \frac{c_4 - w_4\mu_4}{\mu_4^{p-1}} = 0, \\
&\frac{c_1 - w_1\mu_1}{\mu_1^{p-1}} + \frac{c_2 - w_2\mu_2}{\mu_2^{p-1}} + \frac{c_5 - w_5\mu_5}{\mu_5^{p-1}} + \frac{c_6 - w_6\mu_6}{\mu_6^{p-1}} = 0, \\
&\frac{c_1 - w_1\mu_1}{\mu_1^{p-1}} + \frac{c_3 - w_3\mu_3}{\mu_3^{p-1}} + \frac{c_5 - w_5\mu_5}{\mu_5^{p-1}} + \frac{c_7 - w_7\mu_7}{\mu_7^{p-1}} = 0.
\end{aligned}$$

Uppgift 3

Skriv på listform enligt (tex)

$$
\begin{array}{ccc}
1 & 1 & 1 \\
1 & 1 & 2 \\
\vdots & & \\
1 & 1 & 20 \\
1 & 2 & 1 \\
\vdots & & \\
1 & 2 & 20 \\
2 & 1 & 1 \\
\vdots & & \\
2 & 1 & 20 \\
2 & 2 & 1 \\
\vdots & & \\
2 & 2 & 20
\end{array}
$$

dvs totalt 80 rader. Modellen är $\mathrm{E}[Y_{ijk}] = \gamma_0 \gamma_1^i \gamma_2^j \gamma_3^k$ där γ_3^k är känd. På listform blir modellen

$$\mathrm{E}[Y_i] = u_i \mu_i = u_i \exp\Big(\sum_{j=1}^{r} \beta_j x_{ij}\Big), \quad i = 1, \dots, 80, \quad r = 1+(2-1)+(2-1) = 3,$$

där

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

och $u_i \in \{\gamma_3^1, \dots, \gamma_3^{20}\}$ enligt position i listan. ML-ekvationerna ges av

$$\sum_{i=1}^{80} w_i \frac{y_i - u_i \mu_i}{(u_i \mu_i)^{p-1}} x_{ik} = 0, \quad k = 1, \dots, 3, \quad \mu_i = \exp\Big(\sum_{j=1}^{3} \beta_j x_{ij}\Big)$$

där $p = 1$ är lämpligt val. Om vi inför nya variabler

$$\widetilde{y}_i = y_i/u_i, \quad \widetilde{w}_i = w_i u_i^{2-p}$$

kan ML-ekvationerna skrivas

$$\sum_{i=1}^{80} \widetilde{w}_i \frac{\widetilde{y}_i - \mu_i}{\mu_i^{p-1}} x_{ik} = 0, \quad k = 1, \ldots, 3, \quad \mu_i = \exp\Big(\sum_{j=1}^{3} \beta_j x_{ij}\Big)$$

Uppgift 4

Genom att byta p i $v(\mu) = \mu^p$ till ett större värde kommer punktmolnet ändra form från en strut (samlade punkter till vänster, spridda punkter till höger) till ett punktmoln där värdena varierar liknande oberoende av position på x-axeln. Så bör Pearson residualerna se ut och därför bör GLM-analysen göras om med ett större värde på p.

Uppgift 5

Vi ska testa modellen för riskpremie, en Tweediemodell. 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 $F^{-1}_{\chi^2(r-s)}(0.95)$ 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, $r = 1 + (2-1) + (8-1) + (4-1) = 12$ regressionskoefficienter. H_s svarar mot att vi plockar bort tredje premieargument vilket ger H_s med $s = 9$, och därmed $r - s = 3$. Vi ska alltså se hur teststorheten förhåller sig till kvantilen $F^{-1}_{\chi^2(3)}(0.95)$.