Stockholms Universitet, Matematiska Institutionen

Tentamen: Prissättning inom sakförsäkring, MT7028 2021-01-07 13–19

Examinator: Filip Lindskog, lindskog@math.su.se
Återlämning: Meddelas via kursforum.
Tillåtna hjälpmedel: Alla hjälpmedel som inte innebär kommunikation med någon annan.
Allmänt: Resonemang skall vara klara och tydliga att följa.

Uppgift 1

Betrakta data (w_i, y_i), $i = 1, \dots, n$, från en exponentiell spridningsmodell (EDM). Visa steg för steg från definitionen av EDM att deviansen kan skrivas

$$D(\mathbf{y}, \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} w_i d(y_i, \widehat{\mu}_i)$$

och uttryck d i storheterna som definierar EDM. (10 poäng)

Uppgift 2

Tabell 1 visar historisk skadedata enligt en uppdelning i tariffceller. Antag att skadefrekvens modelleras med en multiplikativ modell baserad på en överspridd Poissonmodell (ODP) och att skadebeloppen inte beror på premieargument och klasser. Skatta väntevärde och standardavvikelse för skadekostnaden för en ny ettårig försäkring som hör till tariffcell $(2, 2, 2)$. Beskriv antaganden och analys. Skattningarna behöver inte beräknas numeriskt men du ska ange hur de beräknas från Tabell 1. (10 poäng)

Uppgift 3

Betrakta en GLM-analys av historisk skadefrekvensdata baserad på 4 premieargument med 2, 2, 3 respektive 2 klasser med data på formen y_{ijkl} (observerad skadefrekvens) och w_{ijkl} (duration) för $i = 1, 2$, $j = 1, 2$, $k = 1, 2, 3$, $l = 1, 2$. Antag en multiplikativ modell för förväntad skadefrekvens där relationstalen som svarar mot premieargumentet med index i antas kända: $i = 1$ betyder bonuskund och svarar mot relationstalet 0.8, $i = 2$ betyder ej bonuskund och svarar mot relationstalet 1.0. Antag en lämplig exponentiell spridningsmodell (EDM) och ange explicit det ekvationssystem vars lösning ger skattningar av relationstalen. Förklara de ingående storheterna och hur relationstalen bestäms. (10 poäng)

Uppgift 4

Betrakta historisk riskpremie y_{jt}, $j = 1, \ldots, 10$, $t = 1, \ldots, 10$, (historisk skadekostnad dividerat med historisk duration) för företagsförsäkring baserat på 10 branscher och 10 observationsår för varje bransch: alltså sammanlagt 100 observationer av historisk riskpremie. Antag att den historiska durationen för varje bransch och varje obserationsår är 100. Figur 1 visar både histogram för historisk riskpremie för olika branscher, med olika färg för olika branscher, samt histogram för historisk riskpremie för alla branscher tillsammans, dvs utan att visa vilka observationer som svarar mot vilken bransch. Antag Bühlmann-Straub-modellen. Skatta från figurerna parametrarna μ, τ, σ. Beräkna kredibilitetsskattningen av riskpremien för ett företag vars historiska branschdata visas i rött i figur 1. (10 poäng)

Uppgift 5

Ett aktuarie vill anpassa en kurva med målet att beskriva hur förväntad skadefrekvens beror av försäkringstagarens ålder. Som utgångspunkt finns historisk data i form av tripletter

i	1	2	3	4
(a_i, w_i, x_i)	$(22, 1000, 50)$	$(25, 500, 27)$	$(40, 4000, 100)$	$(67, 1500, 60)$

där (a, w, x) betyder att historisk data innehåller x rapporterade skadehändelser för försäkringstagare som är a år gamla baserat på w försäkringsår för denna ålder. Kurvanpassningen ska ges som lösningen $\widehat{f}_\lambda$ till

$$\operatorname*{argmin}_{f \in \mathcal{F}} \sum_{i=1}^{4} w_i (y_i - f(a_i))^2 + \lambda \int_{20}^{70} (f''(t))^2 dt,$$

där funktionsrummet $\mathcal{F}$ har basfunktionerna $a \mapsto 1$, $a \mapsto a$, $a \mapsto a^2$, $a \mapsto a^3$. Bortse från effekter av andra kovariater än försäkringstagarens ålder.

(a) Uttryck $\widehat{f}_\lambda$ som en linjärkombination av basfunktionerna och ange vektorn med vikter för denna linjärkombination som ett explicit uttryck i lämpliga matriser vars element är numeriska värden med undantag för parametern λ. (5 poäng)

(b) Bestäm funktionsvärdena $\widehat{f}_1(50)$ och $\widehat{f}_{10^4}(50)$. Kommentera huruvida $\lambda = 1$ eller $\lambda = 10^4$ är lämpligast val av parameter. (5 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

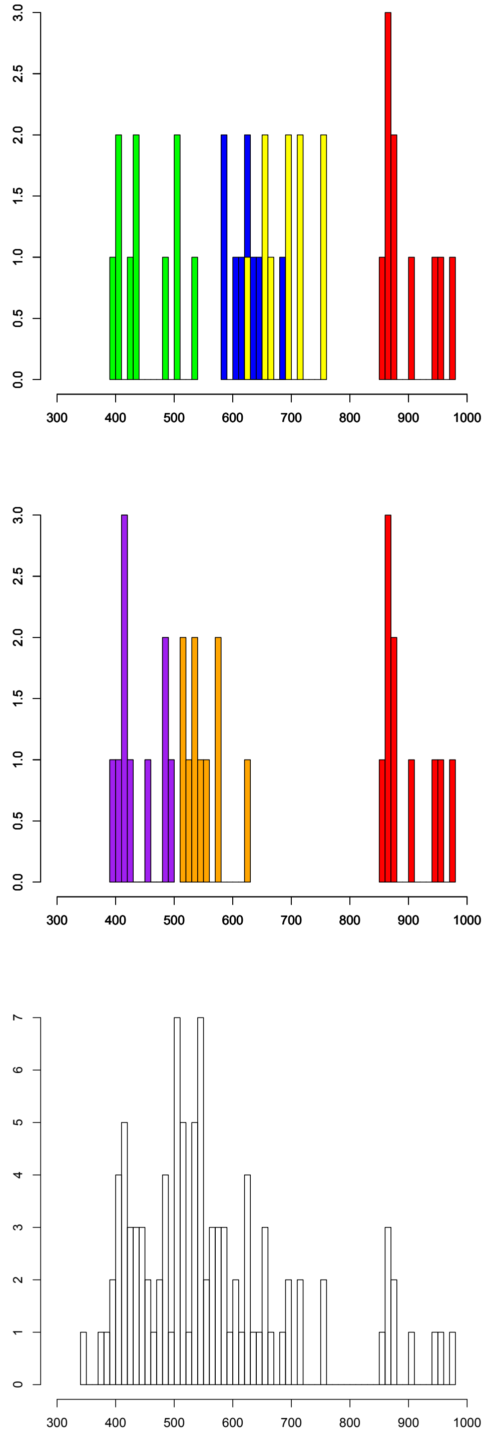

Figure 1: Övre figuren: histogram baserat på $\{y_{jt} : t = 1, \ldots, 10\}$ för 4 olika index j. Mellersta figuren: histogram baserat på $\{y_{jt} : t = 1, \ldots, 10\}$ för 3 olika index j. Undre figuren: histogram baserat på $\{y_{jt} : j = 1, \ldots, 10, t = 1, \ldots, 10\}$.

Uppgift 1

Täthetsfunktionen (sannolikhetsfunktionen) för EDM ges av

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

vilket (baserat på stickprov bestående av n oberoende och likafördelade variabler) ger log-likelihood funktionen (för ML-skattning av $\boldsymbol{\theta}$)

$$l(\boldsymbol{\theta};\phi,\mathbf{y}) = \frac{1}{\phi}\sum_i w_i(y_i\theta_i - b(\theta_i)) + \sum_i c(y_i,\phi,w_i)$$

Eftersom $\mu_i = b'(\theta_i)$ och $\theta_i = h(\mu_i)$ där $h := (b')^{-1}$ är en (monoton) 1-1 avbildning kan log-likelihood funktionen omparametriseras till en funktion $\widetilde{l}(\boldsymbol{\mu}) := l(\mathbf{h}(\boldsymbol{\mu});\phi,\mathbf{y})$ där $\mathbf{h}(\boldsymbol{\mu}) := (h(\mu_1),\ldots,h(\mu_n))$. Vi skriver $l(\boldsymbol{\mu})$ istället för $\widetilde{l}(\boldsymbol{\mu})$. Notera att

$$l(\boldsymbol{\mu}) = \frac{1}{\phi}\sum_i w_i(y_i h(\mu_i) - b(h(\mu_i))) + \sum_i c(y_i,\phi,w_i)$$

Enligt definition gäller $D^*(\mathbf{y},\widehat{\boldsymbol{\mu}}) = 2(l(\mathbf{y}) - l(\widehat{\boldsymbol{\mu}}))$ och $D(\mathbf{y},\widehat{\boldsymbol{\mu}}) = \phi D^*(\mathbf{y},\widehat{\boldsymbol{\mu}})$. Därför gäller

$$\begin{aligned}
D(\mathbf{y},\widehat{\boldsymbol{\mu}}) &= \phi 2(l(\mathbf{y}) - l(\widehat{\boldsymbol{\mu}}))\\
&= \phi\frac{2}{\phi}\sum_i w_i\big(y_i h(y_i) - b(h(y_i)) - y_i h(\widehat{\mu}_i) + b(h(\widehat{\mu}_i))\big)\\
&= \sum_i w_i d(y_i,\widehat{\mu}_i)
\end{aligned}$$

Uppgift 2

Med bascell $(2,2,2)$ pga störst duration väljs $\beta_1 = \log\gamma_0$, $\beta_2 = \log\gamma_{11}$, $\beta_3 = \log\gamma_{21}$, $\beta_4 = \log\gamma_{31}$ och designmatrisen blir

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

För skadefrekvens är Tweedie(1) lämplig vilket ger ML-ekvationerna

$$\sum_{i=1}^{8} d_i(y_i - \mu_i)x_{ik} = 0, \quad k = 1,\ldots,4, \quad \mu_i = \exp\Big(\sum_{j=1}^{4}\beta_j x_{ij}\Big),$$

där d_i fås från rad i i kolumnen "Duration" och $y_i = n_i/d_i$ där n_i fås från rad i i kolumnen "Antal skador" enligt Tabell 1. Skattad förväntad skadefrekvens för cellen $(2,2,2)$ är $\widehat{\mu}_8^{(F)} = \exp(\widehat{\beta}_1)$, där $\widehat{\beta}_1$ kommer från lösningen till ML-ekvationerna. Spridningsparametern skattas som Pearson residual:

$$\widehat{\phi}^{(F)} = \frac{1}{8-4} \sum_{i=1}^{8} d_i \frac{(y_i - \widehat{\mu}_i)^2}{\widehat{\mu}_i}$$

Variansen för antal skador för ett ny ettårig försäkring som tillhör cellen $(2,2,2)$ skattas som $\widehat{\mu}_8^{(F)} \widehat{\phi}^{(F)}$.

För medelskada är Tweedie(2) lämplig. Eftersom skadebeloppen inte beror på kovariater blir väntevärdesskattningen

$$\widehat{\mu}^{(S)} = \frac{s_1 + \cdots + s_8}{n_1 + \cdots + n_8}$$

där n_i och s_i fås från rad i i kolumnerna "Antal skador" respektive "Skadekostnad". Spridningsparametern skattas som Pearson residual:

$$\widehat{\phi}^{(S)} = \frac{1}{8-1} \sum_{i=1}^{8} n_i \frac{\left(y_i - \widehat{\mu}^{(S)}\right)^2}{\left(\widehat{\mu}^{(S)}\right)^2}$$

där $y_i = s_i/n_i$. Variansen för ett skadebelopp skattas som $\left(\widehat{\mu}^{(S)}\right)^2 \widehat{\phi}^{(S)}$.

Väntevärde och varians för skadekostnad för en ny ettårig försäkring som tillhör cellen $(2,2,2)$ skattas som $\widehat{\mu}_8^{(F)} \widehat{\mu}^{(S)}$ respektive

$$\widehat{\mu}_8^{(F)} \widehat{\phi}^{(F)} \left(\widehat{\mu}^{(S)}\right)^2 + \widehat{\mu}_8^{(F)} \left(\widehat{\mu}^{(S)}\right)^2 \widehat{\phi}^{(S)} = \widehat{\mu}_8^{(F)} \left(\widehat{\mu}^{(S)}\right)^2 \left(\widehat{\phi}^{(F)} + \widehat{\phi}^{(S)}\right)$$

Uppgift 3

Skriv på listform enligt (tex)

1	1	1	1
1	1	1	2
1	1	2	1
1	1	2	2
1	1	3	1
1	1	3	2
1	2	1	1
1	2	1	2
1	2	2	1
1	2	2	2
1	2	3	1
1	2	3	2
2	1	1	1
2	1	1	2
2	1	2	1
2	1	2	2
2	1	3	1
2	1	3	2
2	2	1	1
2	2	1	2
2	2	2	1
2	2	2	2
2	2	3	1
2	2	3	2

dvs totalt $2 \cdot 2 \cdot 3 \cdot 2 = 24$ rader. Modellen är $\mathrm{E}[Y_{ijkl}] = \gamma_0 \gamma_{1i} \gamma_{2j} \gamma_{3k} \gamma_{4l}$ där $\gamma_{11} = 0.8$ och $\gamma_{12} = 1.0$ är givna. På listform blir modellen

$$\mathrm{E}[Y_i] = u_i \mu_i = u_i \exp\left(\sum_{j=1}^{r} \beta_j x_{ij} \right), \quad i = 1, \ldots, 36, \quad r = 1 + 1 + 2 + 1 = 5.$$

Med $(1,1,1)$ som bascell för den multiplikativa modellen för μ_i fås

$\beta_1 = \log \gamma_0,\ \beta_1 = \log \gamma_{12},\ \beta_2 = \log \gamma_{22},\ \beta_3 = \log \gamma_{32},\ \beta_4 = \log \gamma_{33},\ \beta_5 = \log \gamma_{42}$

vilket ger designmatrisen

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

Notera att $u_i = 0.8$ för $i = 1, \dots, 12$ och $u_i = 1.0$ för $i = 13, \dots, 24$. ML-ekvationerna ges av (Tweedie(1))

$$\sum_{i=1}^{24} \widetilde{w}_i(\widetilde{y}_i - \mu_i)x_{ik} = 0, \quad k = 1, \dots, 5, \quad \mu_i = \exp\Big(\sum_{j=1}^{5} \beta_j x_{ij}\Big)$$

där $\widetilde{y}_i = y_i/u_i$ och $\widetilde{w}_i = w_i u_i$ där y_i och w_i betecknar observerade skadefrekvenser och durationer enligt ursprunglig data.

Uppgift 4

För vardera av histogramen med olika färger gäller att den betingade fördelningen för Y_{jt} givet V_j ser symmetrisk och lättsvansad ut: $N(V_j, \sigma^2/w_{jt})$, där $w_{jt} = 100$, är en rimlig gissning. För normalfördelningen gäller att sannolikheten att ett slumpat värde hamnar inom 2 standardavvikelser från väntevärdet är ca 0.95. Sannolikheten att 10 oberoende slumpade värden alla hamnar inom 2 standardavvikelser från väntevärdet är därför ca $0.95^{10} \approx$

0.63. En rimlig skattning är därför att $4\sigma/10$ är ungefär bredden på ett histogram av en given färg. Då bredden är ungefär 150 löses ekvationen $4\sigma/10 = 150$ vilket ger $\widehat{\sigma} = 375$. Vi noterar att fördelningen för Y_{jt} har väntevärde μ och varians $\tau^2 + \sigma^2/100$. Med samma argument som ovan löses ekvationen $4^2(\tau^2 + \sigma^2/100) = (980 - 330)^2$ vilket ger $\widehat{\tau} = 158.1139$. Från figur gissar vi $\widehat{\mu} = 550$ och $\overline{y}_{j\cdot} \approx 900$ för observationerna som svarar mot färgen röd. Kredibilitetsvikten skattas till

$$\frac{1000}{1000 + 375^2/158^2} \approx 0.994$$

vilket ger kredibilitetsskattningen

$$\frac{1000}{1000 + 375^2/158^2} \cdot 900 + \left(1 - \frac{1000}{1000 + 375^2/158^2}\right) \cdot 550 \approx 900$$

Notera att det går att göra mer noggranna skattningar.

Uppgift 5

(a)

$$\widehat{f}_\lambda(a) = \sum_{j=1}^{4} \widehat{\theta}_j B_j(a), \quad \widehat{\boldsymbol{\theta}} := \left(\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{B} + \lambda\boldsymbol{\Omega}\right)^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{y}, \quad B_j(a) = a^{j-1}.$$

Eftersom $\mathbf{W} = \mathrm{diag}(w_1, \ldots, w_4)$ och $y_j = x_j/w_j$ fås $\mathbf{Wy} = \mathbf{x}$. $\mathbf{B}$ är 4×4 matris med $B_{ij} = B_j(x_i)$:

$$\mathbf{B} = \begin{bmatrix} 1 & a_1 & a_1^2 & a_1^3 \\ 1 & a_2 & a_2^2 & a_2^3 \\ 1 & a_3 & a_3^2 & a_3^3 \\ 1 & a_4 & a_4^2 & a_4^3 \end{bmatrix} = \begin{bmatrix} 1 & 22 & 484 & 10648 \\ 1 & 25 & 625 & 15625 \\ 1 & 40 & 1600 & 64000 \\ 1 & 67 & 4489 & 300763 \end{bmatrix}$$

$\boldsymbol{\Omega}$ är symmetrisk 4×4 matris med $\Omega_{jk} = \int_{20}^{70} B_j''(t)B_k''(t)dt$:

$$\Omega_{jk} = 0 \text{ om } j \leq 2 \text{ eller } k \leq 2,$$
$$\Omega_{33} = 2 \cdot 2 \cdot \int_{20}^{70} dt, \quad \Omega_{34} = 2 \cdot 6 \cdot \int_{20}^{70} t dt, \quad \Omega_{44} = 6 \cdot 6 \cdot \int_{20}^{70} t^2 dt$$

vilket ger

$$\boldsymbol{\Omega} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 4(70-20) & 12(70^2-20^2)/2 \\ 0 & 0 & 12(70^2-20^2)/2 & 36(70^3-20^3)/3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 200 & 27000 \\ 0 & 0 & 27000 & 4.02 \cdot 10^6 \end{bmatrix}$$

För parametervärden $\lambda = 1$ och $\lambda = 10^4$ fås

$$\boldsymbol{\theta}(\lambda = 1) = \begin{bmatrix} -1.967507 \cdot 10^{-1} \\ 2.285515 \cdot 10^{-2} \\ -6.467150 \cdot 10^{-4} \\ 5.348262 \cdot 10^{-6} \end{bmatrix}, \quad \boldsymbol{\theta}(\lambda = 10^4) = \begin{bmatrix} 4.748017 \cdot 10^{-2} \\ 2.306996 \cdot 10^{-3} \\ -1.240804 \cdot 10^{-4} \\ 1.312718 \cdot 10^{-6} \end{bmatrix}$$

(b) Se figur 2. Uppenbart att $\lambda = 1$ ger en orimlig kurva medan $\lambda = 10^4$ ger en rimligare kurva. Notera att $\widehat{f}_1(50) = -0.002247957$ och att $\widehat{f}_{10^4}(50) = 0.01671876$.

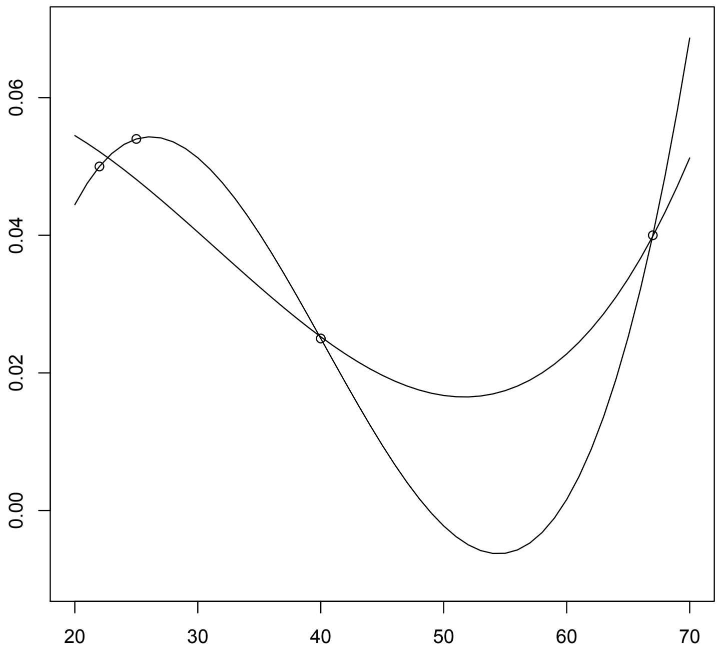

Figure 2: Kurvanpassning till skadefrekvenser för olika åldrar.