Matematisk statistik
Stockholms universitet

Matematiska modeller inom sakförsäkring

Björn Johansson

Kompendium
Februari 1997 – Nytryck 2014

Postadress:
Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm
Sverige

Internet:
http://www.math.su.se

Matematisk statistik
Stockholms universitet
Kompendium 1997
http://www.math.su.se

Matematiska modeller inom sakförsäkring

Björn Johansson *

Februari 1997 – Nytryck 2014

Förord

Detta kompendium handlar om grundläggande modeller inom sakförsäkring, samt beräkningsmetoder utifrån dessa modeller. Kompendiet kom till eftersom jag saknade en enhetlig framställning av detta område för undervisningen i sakförsäkringsmatematik.

Vidare saknade jag också ordentliga övningsmaterial för studenterna att pröva metoderna på. I anslutning till varje kapitel finns därför "övningscase" bestående av tvättäkta problemställningar och omfattande datamaterial från svenska försäkringsbolag. Jag vill tacka Skandia, Skandia International, Länsförsäkringsbolagens AB, samt AMF och Folksam, som vänligen ställt dessa material till förfogande.

Sammanställandet av casen skedde inom ramen för ett projekt kallat "Casemetodik inom undervisningen i försäkringsmatematik". Projektet finansierades av Grundutbildningsrådet, som härmed också tackas.

För varje övningscase ges först en beskrivning av vad det hela handlar om, samt vilka frågor man vill ha svar på. Därefter ger jag ett förslag till hur man kan börja nysta i problematiken. Den oförvägne läsaren kan dock söka sig egna vägar och för de mer komplexa casen finns många alternativa analysmöjligheter.

Kompendiet innehåller också övningar av mer teoretisk karaktär. Dessa har dock också ett praktiskt värde, eftersom man ofta är tvungen att utvidga eller modifiera existerande metoder, vilket inte är möjligt utan kunskaper om hur metoderna är konstruerade. I slutet av kompendiet finns lösningar till samtliga övningar.

Grundläggande kunskaper i matematisk statistik förutsätts. I de två första kapitlen går jag igenom vad man behöver kunna. För arbetet med övningscasen krävs också att man behärskar lämplig programvara (t ex MatLab, Mathematica, Excel, SAS,. . .).

Slutligen vill jag tacka alla de kursdeltagare, många av dem från försäkringsbranschen, som lämnat värdefulla synpunkter på tidigare versioner av kompendiet.

Kräftriket 12 februari 1997

Björn Johansson

Nytryck 2008 och 2014 med ny typografisk form och rättade tryckfel.

* E-post: bjorn.johansson@lansforsakringar.se. Postadress: Försäkringsekonomi, Länsförsäkringar AB, 106 50 Stockholm.

Innehåll

1 Grundläggande Sannolikhetsteori

Det här är en sammanställning av sannolikhetsteoretiska begrepp och resultat som används i kompendiet. Läsaren förväntas ha sett det mesta tidigare, varför framställningen är något kortfattad.

1.1 Sannolikhetsfördelningar

Låt X vara en stokastisk variabel med fördelningsfunktionen F, dvs $F(x) = P(X \leq x)$. Om det finns en funktion f, så att

$$F(y) = \int_{-\infty}^{y} f(x)\, dx$$

säges X ha en *absolutkontinuerlig* fördelning och f kallas *täthetsfunktionen* till X.

Exempel 1.1 En fördelning som vi kommer att stöta på upprepade gånger är gammafördelningen, som har tätheten

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\beta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

där Γ betecknar gammafunktionen:

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1}\, e^{-x}\, dx$$

Fördelningen har två parametrar, $\alpha > 0$ och $\beta > 0$.

Fördelningsfunktionen för en gammafördelning,

$$F(y) = \int_0^y \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\beta x}\, dx$$

finns som standardfunktion i moderna programpaket. □

Om en stokastisk variabel X med sannolikhet 1 antar något av ett ändligt eller uppräkneligt oändligt antal värden $x_0, x_1, x_2, \ldots$, säges X ha en *diskret* fördelning. Motsvarigheten till täthetsfunktionen ovan är den så kallade *sannolikhetsfunktionen*

$$p_k := P(X = x_k), \qquad k = 0, 1, 2, \ldots$$

Fördelningsfunktionen kan skrivas som

$$F(x) = P(X \le x) = \sum_{k:\, x_k \le x} p_k$$

Ett vanligt specialfall är en *heltalsfördelning*: $x_k = k$, $k \ge 0$.

Exempel 1.2 Den förmodligen mest kända diskreta fördelningen är *binomialfördelningen*, som är fördelningen för antalet lyckade av n oberoende försök, där vart och ett har en viss sannolikhet q att lyckas. Binomialfördelningens sannolikhetsfunktion är

$$p_k = \binom{n}{k} q^k \,(1-q)^{n-k}; \qquad k = 0, 1, \ldots, n$$

Fördelningsfunktionen F för binomialfördelningen beräknas enkelt rekursivt med hjälp av en dator: För $k \ge 1$ gäller att

$$\begin{cases} p_k = \frac{q}{1-q}\,\frac{n-k+1}{k}\, p_{k-1} \\ F(k) = F(k-1) + p_k \end{cases}$$

Som startvärden för rekursionerna har man $p_0 = (1-q)^n$, $F(0) = p_0$. □

Inom sakförsäkringsmatematiken råkar man inte sällan ut för fördelningar av *blandad* typ, dvs fördelningar som består av en absolutkontinuerlig del och en diskret del. Fördelningsfunktionen kan då skrivas som

$$F(x) = \alpha\, F_1(x) + (1-\alpha)\, F_2(x)$$

där F_1 är fördelningsfunktionen för en absolutkontinuerlig fördelning, F_2 är fördelningsfunktionen för en diskret fördelning och $0 \le \alpha \le 1$.

Exempel 1.3 Inom så kallad *excess of loss*-återförsäkring förekommer följande konstruktion: Om ett skadebelopp uppgår till X kronor, betalar återförsäkringsbolaget beloppet

$$Y := \begin{cases} 0, & \text{om } X < a \\ X - a, & \text{om } a \le X < b \\ b - a, & \text{om } X \ge b \end{cases}$$

där a och b är några gränser sådana att $a < b$.

Vi låter G beteckna fördelningsfunktionen för X och vill beräkna fördelningsfunktionen F för Y. Låt oss anta att G är absolutkontinuerlig. För $x < 0$ gäller att $F(x) = 0$ och för $0 \leq x < b - a$ får vi att

$$\begin{aligned} F(x) &= P(Y \leq x) = P(X - a \leq x) \\ &= P(X \leq a + x) = G(a + x) \end{aligned}$$

För $x > b - a$ gäller att $F(x) = 1$.

Följande allmänna procedur kan användas för att dela upp fördelningen i en diskret och en absolutkontinuerlig del: Fördelningen för Y har två punktmassor, i punkten $x_0 = 0$ och i punkten $x_1 = b - a$. Vi finner att

$$\begin{aligned} q_0 &:= P(Y = x_0) = P(X \leq a) = G(a) \\ q_1 &:= P(Y = x_1) = P(X > b) = 1 - G(b) \end{aligned}$$

Vi definierar nu α så att

$$1 - \alpha = \sum_k q_k$$

vilket ger $\alpha = G(b) - G(a)$. Därefter definierar vi F_2 via

$$F_2(x) := \frac{1}{1 - \alpha} \sum_{k:\, x_k \leq x} q_k$$

vilket ger

$$F_2(x) = \begin{cases} 0, & x < 0 \\ G(a)/(G(a) + 1 - G(b)), & 0 \leq x < b - a \\ 1, & x \geq b - a \end{cases}$$

Slutligen definierar vi F_1 så att

$$F(x) = \alpha\, F_1(x) + (1 - \alpha)\, F_2(x)$$

dvs

$$F_1(x) = \frac{1}{\alpha}\big(F(x) - (1 - \alpha)\, F_2(x)\big)$$

Vi får att

$$F_1(x) = \begin{cases} 0, & x < 0 \\ (G(a + x) - G(a))/(G(b) - G(a)), & 0 \leq x < b - a \\ 1, & x \geq b - a \end{cases}$$

Uppenbarligen gäller att F_1 är absolutkontinuerlig och att F_2 är diskret. □

Om X har en absolutkontinuerlig fördelning med täthetsfunktion f, definieras väntevärdet för X av

$$E[X] := \int_{-\infty}^{\infty} x\, f(x)\, dx$$

givetvis förutsatt att integralen existerar.

Om h är en funktion på $(-\infty, \infty)$, gäller att den stokastiska variabeln $h(X)$ har väntevärdet

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx \tag{1.1}$$

Exempel 1.4 Om X är gammafördelad, finner man att

$$\begin{aligned}
E[X^k] &= \int_{-\infty}^{\infty} x^k\, f(x)\, dx = \int_0^{\infty} x^k\, \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\beta x}\, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+k-1}\, e^{-\beta x}\, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} \left(\frac{y}{\beta}\right)^{\alpha+k-1} e^{-y}\, \frac{1}{\beta}\, dy \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)\, \beta^{\alpha+k}} \int_0^{\infty} y^{\alpha+k-1}\, e^{-y}\, dy = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\, \beta^k}
\end{aligned}$$

Det sista uttrycket kan förenklas genom att använda att $\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$, $\Gamma(\alpha+2) = (\alpha+1)\alpha\,\Gamma(\alpha)$, osv. □

Om X i stället har en diskret fördelning, med $p_k = P(X = x_k)$, definieras väntevärdet för X analogt som

$$E[X] = \sum_k x_k\, p_k$$

Man har också att

$$E[h(X)] = \sum_k h(x_k)\, p_k$$

Exempel 1.5 Om X är binomialfördelad, får man att

$$\begin{aligned}
E[X] &= \sum_{k=0}^{n} k\, p_k = \sum_{k=1}^{n} k \binom{n}{k} q^k\, (1-q)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!\,(n-k)!}\, q^k\, (1-q)^{n-k} \\
&= nq \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,(n-1-(k-1))!} q^{k-1}\, (1-q)^{n-1-(k-1)} \\
&= nq \sum_{j=0}^{n-1} \binom{n-1}{j} q^j\, (1-q)^{n-1-j} \\
&= nq\, (q+1-q)^{n-1} = nq
\end{aligned}$$

□

Antag slutligen att X har en blandad fördelning, dvs fördelningsfunktionen kan skrivas som

$$F(x) = \alpha\, F_1(x) + (1-\alpha)\, F_2(x)$$

där F_1 är absolutkontinuerlig, F_2 är diskret och $0 \le \alpha \le 1$. Om f är täthetsfunktionen till F_1 och p sannolikhetsfunktionen till F_2, definieras väntevärdet av $h(X)$ som

$$E[h(X)] := \alpha \int_{-\infty}^{\infty} h(x)\, f(x)\, dx + (1-\alpha) \sum_k h(x_k) p_k \tag{1.2}$$

Exempel 1.6 Låt X och Y vara som i Exempel 1.3. Låt g beteckna täthetsfunktionen till X. Enligt vad vi fann i Exempel 1.3, gäller att

$$F_1(x) = \int_{-\infty}^{x} f(x)\, dx$$

där

$$f(x) = \begin{cases} 0, & x \le 0 \\ g(a+x)/(G(b)-G(a)) & 0 < x < b-a \\ 0, & x \ge b-a \end{cases}$$

Enligt definitionen (1.2) ovan har vi alltså att

$$\begin{aligned}
E[Y] &= \alpha \int_0^{b-a} x\, g(a+x)/(G(b)-G(a))\, dx \\
&+ (1-\alpha) \left(0 \cdot \frac{G(a)}{1-\alpha} + (b-a)\, \frac{1-G(b)}{1-\alpha} \right) \\
&= \int_0^{b-a} x\, g(a+x)\, dx + (b-a)(1-G(b))
\end{aligned}$$

Vi noterar att $Y = h(X)$, där

$$h(x) = \begin{cases} 0, & x < a \\ x - a, & a \leq x < b \\ b - a, & x \geq b \end{cases}$$

Vi bör alltså få samma sak om vi beräknar $E[Y]$ via (1.1):

$$\begin{aligned} E[Y] &= E[h(X)] = \int_{-\infty}^{\infty} h(x)\, g(x)\, dx \\ &= \int_{-\infty}^{a} h(x)\, g(x)\, dx + \int_{a}^{b} h(x)\, g(x)\, dx + \int_{b}^{\infty} h(x)\, g(x)\, dx \\ &= 0 + \int_{a}^{b} (x-a)\, g(x)\, dx + \int_{b}^{\infty} (b-a)\, g(x)\, dx \\ &= \int_{0}^{b-a} x\, g(x+a)\, dx + (b-a)\,(1 - G(b)) \end{aligned}$$

Detta illustrerar att definitionen (1.1) är vettig. □

Ett viktigt hjälpmedel vid studiet av fördelningar inom sakförsäkringsmatematiken är den så kallade *väntevärdesfunktionen*. Den definieras för en godtycklig stokastisk variabel av

$$L(a) := E\big[\min(X, a)\big] \tag{1.3}$$

där

$$\min(x, a) := \begin{cases} x, & x \leq a \\ a, & x > a \end{cases}$$

Väntevärdet för variabeln Y i föregående exempel kan skrivas som

$$E[Y] = E[h(X)] = L(b) - L(a)$$

Det är behändigt att införa en beteckning för uttrycket i (1.2):

$$\int h(x)\, dF(x) := \alpha \int_{-\infty}^{\infty} h(x)\, f(x)\, dx + (1-\alpha) \sum_k h(x_k)\, p_k$$

Som specialfall, för $\alpha = 1$ och $\alpha = 0$, får vi väntevärdet $E[h(X)]$ då X har en absolutkontinuerlig respektive en diskret fördelning.

1.2 Moment och genererande funktioner

Låt X ha en godtycklig fördelning, dvs absolutkontinuerlig, diskret eller blandad.

Det j:te *momentet* (kring origo) för X definieras av

$$\alpha_j := E[X^j] = \int x^j \, dF(x) \qquad j \geq 1$$

Det j:te *centralmomentet* för X definieras av

$$\mu_j := E\big[(X - E[X])^j\big] \qquad j \geq 2$$

Vi definierar också $\mu_1 := \alpha_1$. Det andra centralmomentet kring origo kallas *variansen* för X. Storheten

$$\gamma := \frac{\mu_3}{\mu_2^{3/2}}$$

brukar kallas *skevheten* för X.

Via enkla räkningar kan centralmomenten uttryckas i termer av momenten kring origo och tvärtom. Exempelvis har man att

$$\begin{aligned} \mu_2 &= \alpha_2 - \alpha_1^2 \\ \mu_3 &= \alpha_3 - 3\,\alpha_2\,\alpha_1 + 2\,\alpha_1^3 \end{aligned}$$

Den *momentgenererande funktionen* för X (eller F) definieras av

$$M(t) := E\big[e^{tX}\big] = \int e^{tx} \, dF(x)$$

Man har alltid $M(0) = 1$, men det behöver inte gälla att $M(t) < \infty$ för alla $t \neq 0$.

Exempel 1.7 Om X är gammafördelad, gäller att

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \, dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-t)x} \, dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\beta - t}\right)^{\alpha-1} e^{-y} \frac{1}{\beta - t} \, dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)\,(\beta - t)^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} \, dy = \left(\frac{\beta}{\beta - t}\right)^\alpha \end{aligned}$$

Integralerna ovan är ändliga, dvs beräkningarna är giltiga, endast då $t < \beta$. För $t \geq \beta$ har vi $M(t) = \infty$. $\square$

Exempel 1.8 Om X är binomialfördelad, får man att

$$\begin{aligned} M(t) &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} q^k \, (1-q)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (q\, e^t)^k \, (1-q)^{n-k} \\ &= (q\, e^t + 1 - q)^n = (1 + q(e^t - 1))^n \end{aligned}$$

För binomialfördelningen gäller att $M(t) < \infty$ för alla värden på t. $\square$

Man kan visa att den momentgenererande funktionen entydigt bestämmer fördelningen, förutsatt att $M(t) < \infty$ för alla t i en omgivning av origo.

Den momentgenererande funktionen har fått sitt namn av att momenten kring origo kan erhållas från dess derivator. Förutsatt att $M(t) < \infty$ i en omgivning av origo, har man att

$$\begin{aligned} M^{(k)}(t) &:= \frac{d^k}{dt^k} M(t) = \frac{d^k}{dt^k} \int e^{tx} \, dF(x) \\ &= \int \frac{d^k}{dt^k} e^{tx} \, dF(x) = \int x^k \, e^{tx} \, dF(x) \end{aligned}$$

vilket ger

$$\lim_{t \to 0} M^{(k)}(t) = \int x^k \, dF(x) = E[X^k]$$

Om man har n stycken oberoende stokastiska variabler $X_1, \ldots, X_n$ med momentgenererande funktioner $M_1, \ldots, M_n$, gäller att den stokastiska variabeln $Y := X_1 + \cdots + X_n$ har den momentgenererande funktionen

$$\begin{aligned} M(t) &= E\big[e^{tY}\big] = E\big[e^{t(X_1 + \cdots + X_n)}\big] \\ &= E\Big[\prod_{k=1}^{n} e^{tX_k}\Big] = \prod_{k=1}^{n} E\big[e^{tX_k}\big] = \prod_{k=1}^{n} M_k(t) \end{aligned} \tag{1.4}$$

Den *kumulantgenererande funktionen* Ψ till X definieras av

$$\Psi(t) := \log M(t)$$

och den j:te *kumulanten* för X är

$$\kappa_j := \lim_{t\to 0} \frac{d^j}{dt^j}\,\Psi(t)$$

Man kan visa att

$$\kappa_j = \mu_j \qquad j = 1, 2, 3, \tag{1.5}$$

men sambandet gäller inte för $j \geq 4$. I många situationer är det enklaste sättet att räkna ut de tre första centralmomenten att räkna ut kumulanterna.

Vi noterar att likheten (1.4) kan skrivas som

$$\Psi(t) = \sum_{k=1}^{n} \Psi_k(t)$$

om Ψ och Ψ_k betecknar de kumulantgenererande funktionerna för Y respektive X_k.

Exempel 1.9 Enligt vad vi fann i Exempel 1.7, ges den kumulantgenererande funktionen för gammafördelningen av

$$\Psi(t) = \log M(t) = \alpha \log \beta - \alpha \log(\beta - t)$$

Från detta får vi enkelt att

$$\begin{aligned} \mu_1 &= \kappa_1 = \Psi'(0) = \frac{\alpha}{\beta} \\ \mu_2 &= \kappa_2 = \Psi''(0) = \frac{\alpha}{\beta^2} \\ \mu_3 &= \kappa_3 = \Psi'''(0) = \frac{2\alpha}{\beta^3} \end{aligned}$$

Om $X_1, \ldots, X_n$ är oberoende och X_k är gammafördelad med parametrar α_k och β, så blir den momentgenererande funktionen för $X = X_1 + \cdots + X_n$

$$M(t) = \prod_{k=1}^{n} \left(\frac{\beta}{\beta - t}\right)^{\alpha_k} = \left(\frac{\beta}{\beta - t}\right)^{\alpha_1 + \cdots + \alpha_n}$$

dvs X är gammafördelad med parametrar $\alpha_1 + \cdots + \alpha_n$ och β. $\square$

För heltalsvärda stokastiska variabler är det behändigt att införa de s k faktorialmomenten:

$$\begin{aligned} \gamma_1 &:= E[N] \\ \gamma_j &:= E\big[N(N-1)\cdots(N-j+1)\big] = \sum_{k=j}^{\infty} k(k-1)\cdots(k-j+1)\,p_k \quad j \geq 2. \end{aligned}$$

Det är lätt att räkna fram sambanden mellan faktorialmomenten och momenten kring origo. Man har till exempel att

$$\begin{aligned}\gamma_2 &= \alpha_2 - \alpha_1 \\ \gamma_3 &= \alpha_3 - 3\alpha_2 + 2\alpha_1\end{aligned}$$

När man studerar heltalsvärda stokastiska variabler, arbetar man ofta med den s k *sannolikhetsgenererande funktionen*

$$Q(s) := E\big[s^N\big] = \sum_{k=0}^{\infty} s^k\, p_k$$

i stället för den momentgenererande funktionen. Sambandet mellan dem är enkelt: $M(t) = Q(e^t)$. Den sannolikhetsgenererande funktionen kan användas till att räkna ut faktorialmomenten:

$$\gamma_j = \lim_{s\to 1} \frac{d^j}{ds^j}\, Q(s)$$

Ibland händer det att man har den sannolikhetsgenererande funktionen given och vill beräkna sannolikheterna p_k; $k = 0, 1, \ldots$ Det kan man göra via sambandet

$$p_k = \frac{1}{k!} \lim_{s\to 0} \frac{d^k}{ds^k}\, Q(s)$$

Om $N_1, \ldots, N_m$ är oberoende heltalsvärda variabler med sannolikhetsgenererande funktioner $Q_1, \ldots, Q_m$, har $N_1 + \cdots + N_m$ den sannolikhetsgenererande funktionen

$$Q(s) = \prod_{k=1}^{m} Q_k(s)$$

Exempel 1.10 Faktorialmomenten för binomialfördelningen är lätta att räkna ut direkt:

$$\begin{aligned}\gamma_j &= \sum_{k=0}^{n} k(k-1)\cdots(k-j+1)\, p_k \\ &= \sum_{k=j}^{n} k(k-1)\cdots(k-j+1)\, \frac{n!}{k!\,(n-k)!}\, q^k\, (1-q)^{n-k} \\ &= n(n-1)\cdots(n-j+1)\, q^j \sum_{k=j}^{n} \frac{(n-j)!}{(k-j)!\,(n-j-(k-j))!}\, q^{k-j}\, (1-q)^{n-j-(k-j)} \\ &= n(n-1)\cdots(n-j+1)\, q^j \sum_{i=0}^{n-j} \binom{n-j}{i} q^i\, (1-q)^{n-j-i} \\ &= n(n-1)\cdots(n-j+1)\, q^j\, (q+1-q)^{n-j} = n(n-1)\cdots(n-j+1)\, q^j\end{aligned}$$

Den sannolikhetsgenererande funktionen för binomialfördelningen är

$$Q(s) = M(\log s) = (1 + q(s-1))^n$$

Om $N_1, \ldots, N_m$ är oberoende och N_k är binomialfördelad med parametrar n_k och p_k, blir den sannolikhetsgenererande funktionen för $N_1 + \cdots + N_m$

$$Q(s) = \prod_{k=1}^{m} (1 + q_k(s-1))^{n_k}$$

Om $q_1 = q_2 = \cdots = q_m = q$, får vi

$$Q(s) = (1 + q(s-1))^{n_1 + \cdots + n_m}$$

dvs $N_1 + \cdots + N_m$ är binomialfördelad med parametrar $n_1 + \cdots + n_m$ och q. Om däremot q_k:na är olika, så är $N_1 + \cdots + N_m$ inte binomialfördelad. □

1.3 Betingade väntevärden

Det betingade väntevärdet för en stokastisk variabel X, givet att en annan stokastisk variabel Y antagit värdet y, skriver vi $E\big[X \mid Y = y\big]$. Vi erinrar om följande viktiga likhet, som används många gånger i kompendiet:

$$E\big[h(X)\big] = \int E\big[h(X) \mid Y = y\big] dF_Y(y) \tag{1.6}$$

där F_Y betecknar fördelningsfunktionen för Y.

Exempel 1.11 Låt $Y, X_1, X_2, \ldots$ vara oberoende stokastiska variabler, antag att Y är Poissonfördelad med parameter λ och att alla X_k är gammafördelade med parametrar α och β. Sätt $X := X_1 + \cdots + X_Y$, dvs X är summan av ett Poissonfördelat antal gammavariabler. Låt oss beräkna den momentgenererande funktionen för X:

$$\begin{aligned}
M(t) &= E\big[e^{tX}\big] = \sum_k E\big[e^{t(X_1+\cdots+X_k)} \mid Y = k\big] P(Y = k) \\
&= \sum_k E\big[e^{t(X_1+\cdots+X_k)}\big] P(Y = k) \\
&= \sum_k \left(\left(\frac{\beta}{\beta - t}\right)^{\alpha}\right)^{k} \frac{\lambda^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} \sum_k \left(\lambda \left(\frac{\beta}{\beta - t}\right)^{\alpha}\right)^{k} \frac{1}{k!} \\
&= e^{-\lambda} \exp\left\{\lambda \left(\frac{\beta}{\beta - t}\right)^{\alpha}\right\} = \exp\left\{\lambda\left(\left(\frac{\beta}{\beta - t}\right)^{\alpha} - 1\right)\right\}
\end{aligned}$$

Utifrån detta kan t ex kumulanterna enkelt beräknas. □

Vid omfattande beräkningar med betingade väntevärden är det behändigt att använda sig av ett antal räkneregler, som kan skrivas på ett kompakt sätt.

Låt $Y, X_1, \dots, X_n$ vara stokastiska variabler. Vi antar för enkelhets skull att alla har tätheter, men detta är inte nödvändigt. Vi använder följande beteckningar:

$f_Y(\cdot) =$ tätheten för Y,

$f_n(\cdot, \dots, \cdot) =$ simultana tätheten för $(X_1, \dots, X_n)$,

$f_{Y,n}(\cdot, \cdot, \dots, \cdot) =$ simultana tätheten för $(Y, X_1, \dots, X_n)$,

$f_Y(\cdot | X_1 = x_1, \dots, X_n = x_n) =$ tätheten för Y givet $X_1 = x_1, \dots, X_n = x_n$, dvs

$$f_Y(y|X_1 = x_1, \dots, X_n = x_n) = \frac{f_{Y,n}(y, x_1, \dots, x_n)}{f_n(x_1, \dots, x_n)}$$

Det betingade väntevärdet av Y givet $X_1 = x_1, \dots, X_n = x_n$ är

$$\begin{aligned} \varphi_n(x_1, \dots, x_n) &:= E\big[Y \bigm| X_1 = x_1, \dots, X_n = x_n\big] \\ &= \int y \, f_Y(y|X_1 = x_1, \dots, X_n = x_n) \, dy. \end{aligned}$$

Vi inför nu *beteckningen*

$$E\big[Y \bigm| X_1, \dots, X_n\big] := \varphi_n(X_1, \dots, X_n)$$

Vi ska visa ett antal användbara räkneregler.

1. $E[Y] = E\big[E\big[Y \bigm| X_1, \dots, X_n\big]\big]$

Detta följer av att

$$\begin{aligned} E[Y] &= \int \cdots \int E\big[Y \bigm| X_1 = x_1, \dots, X_n = x_n\big] f_n(x_1, \dots, x_n) \, dx_1 \cdots dx_n \\ &= \int \cdots \int \varphi_n(x_1, \dots, x_n) \, f_n(x_1, \dots, x_n) \, dx_1 \cdots dx_n \\ &= E\big[\varphi_n(X_1, \dots, X_n)\big] = E\big[E\big[Y \bigm| X_1, \dots, X_n\big]\big] \end{aligned}$$

2. Om $m < n$, gäller att

$$E\big[Y \mid X_1, \dots, X_m\big] = E\big[E\big[Y \mid X_1, \dots, X_n\big] \mid X_1, \dots, X_m\big]$$

För att inse detta, notera först att den betingade tätheten för $(X_{m+1}, \dots, X_n)$ givet $X_1 = x_1, \dots, X_m = x_m$ är $f_n(x_1, \dots, x_n)/f_m(x_1, \dots, x_m)$, så

$$\begin{aligned}
\Psi(x_1, \dots, x_m) &:= E\big[\varphi_n(X_1, \dots, X_n) \mid X_1 = x_1, \dots, X_m = x_m\big] \\
&= E\big[\varphi_n(x_1, \dots, x_m, X_{m+1}, \dots, X_n) \mid X_1 = x_1, \dots, X_m = x_m\big] \\
&= \int \cdots \int \varphi_n(x_1, \dots, x_m, x_{m+1}, \dots, x_n) \, \frac{f_n(x_1, \dots, x_n)}{f_m(x_1, \dots, x_m)} \, dx_{m+1} \cdots dx_n = (*)
\end{aligned}$$

Men

$$\varphi_n(x_1, \dots, x_n) = \int y \, \frac{f_{Y,n}(y, x_1, \dots, x_n)}{f_n(x_1, \dots, x_n)} \, dy$$

så $(*)$ blir

$$\begin{aligned}
&\int \cdots \int \int y \, \frac{f_{Y,n}(y, x_1, \dots, x_n)}{f_m(x_1, \dots, x_m)} \, dy \, dx_{m+1} \cdots dx_n \\
&= \int y \, \frac{1}{f_m(x_1, \dots, x_m)} \int \cdots \int f_{Y,n}(y, x_1, \dots, x_m, x_{m+1}, \dots, x_n) \, dx_{m+1} \cdots dx_n \, dy \\
&= \int y \, \frac{1}{f_m(x_1, \dots, x_m)} \, f_{Y,m}(y, x_1, \dots, x_m) \, dy \\
&= \int y \, f_Y(y | X_1 = x_1, \dots, X_m = x_m) \, dy = \varphi_m(x_1, \dots, x_m)
\end{aligned}$$

Vi får alltså att $\Psi(X_1, \dots, X_m) = \varphi_m(X_1, \dots, X_m)$, dvs regel 2 gäller.

3. $E\big[Y \, g(X_1, \dots, X_n) \mid X_1, \dots, X_n\big] = E\big[Y \mid X_1, \dots, X_n\big] g(X_1, \dots, X_n)$

Detta är mer eller mindre självklart:

$$\begin{aligned}
\Psi(x_1, \dots, x_n) &:= E\big[Y \, g(X_1, \dots, X_n) \mid X_1 = x_1, \dots, X_n = x_n\big] \\
&= E\big[Y \, g(x_1, \dots, x_n) \mid X_1 = x_1, \dots, X_n = x_n\big] \\
&= E\big[Y \mid X_1 = x_1, \dots, X_n = x_n\big] g(x_1, \dots, x_n) \\
&= \varphi_n(x_1, \dots, x_n) \, g(x_1, \dots, x_n),
\end{aligned}$$

så att

$$\Psi(X_1, \dots, X_n) = \varphi_n(X_1, \dots, X_n) \, g(X_1, \dots, X_n)$$

dvs regel 3 gäller.

4. Om Y är betingat oberoende av $(X_1, \ldots, X_m)$ givet $(X_{m+1}, \ldots, X_n)$, gäller att

$$E\big[Y \mid X_1, \ldots, X_n\big] = E\big[Y \mid X_{m+1}, \ldots, X_n\big], \qquad m < n$$

Detta följer direkt av definitionen av oberoende.

5. Om vi inför den betingade variansen för Y givet $X_1, \ldots, X_n$, dvs

$$V\big[Y \mid X_1, \ldots, X_n\big] := E\big[Y^2 \mid X_1, \ldots, X_n\big] - \big(E\big[Y \mid X_1, \ldots, X_n\big]\big)^2$$

gäller att

$$V[Y] = E\big[V\big[Y \mid X_1, \ldots, X_n\big]\big] + V\big[E\big[Y \mid X_1, \ldots, X_n\big]\big]$$

Detta följer av regel 1:

$$\begin{aligned}
V[Y] &= E[Y^2] - (E[Y])^2 \\
&= E\big[E\big[Y^2 \mid X_1, \ldots, X_n\big]\big] - \big(E[Y]\big)^2 \\
&= E\big[V\big[Y \mid X_1, \ldots, X_n\big] + \big(E\big[Y \mid X_1, \ldots, X_n\big]\big)^2\big] - (E[Y])^2 \\
&= E\big[V\big[Y \mid X_1, \ldots, X_n\big]\big] + E\big[\big(E\big[Y \mid X_1, \ldots, X_n\big]\big)^2\big] \\
&- \big(E\big[E\big[Y \mid X_1, \ldots, X_n\big]\big]\big)^2 \\
&= E\big[V\big[Y \mid X_1, \ldots, X_n\big]\big] + V\big[E\big[Y \mid X_1, \ldots, X_n\big]\big]
\end{aligned}$$

2 Parameterskattning

Vi ska här gå igenom hur man för en given fördelning finner den bästa anpassningen till ett datamaterial. Det finns inget entydigt svar på vad man menar med bästa anpassning, men vi kommer att ta upp två vanliga metoder att angripa problemet: anpassning till den empiriska fördelningen och ML-skattning, där ML står för *maximum likelihood*.

2.1 Anpassning till den empiriska fördelningen

Det antagande vi utgår från är att vi observerar utfallen $x_1, \dots, x_n$ av n oberoende stokastiska variabler $X_1, \dots, X_n$, alla med samma fördelningsfunktion F. Ett naturligt sätt att skatta $F(x)$, sannolikheten att ett utfall blir mindre än eller lika med x, är andelen observationer som är mindre än eller lika med x. Vi kan skriva detta som

$$\hat{F}(x) = \sum_{k:\, x_k \le x} \frac{1}{n}$$

Funktionen $\hat{F}$ kallas för den empiriska fördelningsfunktionen. Givet att alla observationer är olika, är den empiriska fördelningen en diskret fördelning med sannolikhetsmassan $p_k = 1/n$ i punkten x_k; $k = 1, \dots, n$. Momenten för den empiriska fördelningen kommer vi senare att använda för att skatta parametrar enligt den s k momentmetoden. Väntevärdet är

$$\bar{x} := \sum_{k=1}^{n} x_k \, p_k = \frac{1}{n} \sum_{k=1}^{n} x_k$$

och variansen

$$s^2 := \sum_{k=1}^{n} x_k^2 \, p_k - \bar{x}^2 = \frac{1}{n} \sum_{k=1}^{n} x_k^2 - \bar{x}^2$$

Man övertygar sig lätt om att uttrycken i högerleden för väntevärdet och variansen blir desamma även om flera observationer är lika, vilket typiskt händer för en diskret fördelning. Den empiriska väntevärdesfunktionen, dvs väntevärdesfunktionen för den empiriska fördelningen (se 1.3), blir

$$\hat{L}(x) = \frac{1}{n} \sum_{k=1}^{n} \min(x, x_k)$$

Man kan tänka sig att använda den empiriska fördelningen som modell när man utför beräkningar, med det finns ett par skäl att hellre anpassa en parametrisk fördelning till data (förutsatt att anpassningen är god). För det första är den empiriska fördelningen svårhanterlig om man ska göra omfattande beräkningar. För det andra har den empiriska fördelningen sannolikhetsmassan noll i intervall där det inte råkat hamna några observationer. Exempelvis skattas sannolikheten för att få en observation som är större än $\max(x_1, \ldots, x_n)$ till noll. Detta kan vara olämpligt, särskilt då man sysslar med långsvansade fördelningar.

Det finns flera tänkbara sätt att anpassa en parametrisk fördelning till den empiriska fördelningen. En klassisk metod, momentmetoden, är att bestämma parametrarna så att de första momenten överensstämmer. Man får samma skattningar vare sig man anpassar centralmomenten eller någon av de andra typerna av moment som vi tog upp i föregående avsnitt.

Exempel 2.1 Låt oss anpassa en gammafördelning med hjälp av momentmetoden. Gammafördelningen har två parametrar, vilket innebär att vi kan få de två första momenten i gammafördelningen att stämma överens med motsvarande moment för den empiriska fördelningen. Om de två första centralmomenten ska stämma överens, måste det gälla att

$$\begin{cases} \alpha/\beta = \bar{x} \\ \alpha/\beta^2 = s^2 \end{cases}$$

Lösningen till ekvationssystemet ger *momentmetodens* skattningar av α och β: $\hat{\alpha} = \bar{x}^2/s^2$, $\hat{\beta} = \bar{x}/s^2$. □

Exempel 2.2 Antag nu att observationerna är binomialfördelade med parametrar r och q (vi använder r i stället för som ovan n, eftersom vi för tillfället använder n för att beteckna antalet observationer). Det vanligaste är att r är känd och man vill skatta q. Man kan då bara få väntevärdena att överensstämma, vilket ger skattningen $\hat{q} = \bar{x}/r$. □

En annan naturlig ansats är att försöka få fördelningsfunktionen för den parametriska fördelningen att likna den empiriska så mycket som möjligt. Ett sätt att göra detta är att låta en dator rita upp fördelningsfunktionerna för den parametriska respektive empiriska fördelningen tillsammans och variera parametervärdena tills anpassningen inte går att förbättra nämnvärt. Man kan också tänka sig att mäta avståndet mellan fördelningsfunktionerna på lämpligt sätt och låta datorn finna de parametervärden som minimerar avståndet. Ett vanligt mått på avståndet mellen F och $\hat{F}$ är

$$\sum_{i=1}^{k} w_i \big(F(y_i) - \hat{F}(y_i)\big)^2$$

där $\{y_1, \ldots, y_k\}$ är några väl valda punkter och $\{w_1, \ldots, w_k\}$ är vikter. Vikten w_i kan väljas förhållandevis stor om man anser det viktigt att anpassningen är särskilt god i punkten y_i.

I försäkringssammanhang kan det ofta vara mer relevant att anpassa fördelningens väntevärdesfunktion till den empiriska väntevärdesfunktionen. Tillvägagångssättet är detsamma: Anpassa kurvan grafiskt med hjälp av en dator eller formulera ett minimeringsproblem, på samma sätt som för fördelningsfunktionerna.

2.2 ML-metoden

Vi antar fortfarande att vi observerat utfallen $x_1, \ldots, x_n$ av n oberoende variabler $X_1, \ldots, X_n$ med samma fördelning. Vi antar först att denna fördelning har en täthet f. Den simultana täthetsfunktionen för $(X_1, \ldots, X_n)$ ges då av

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{k=1}^{n} f(x_k)$$

Idén bakom ML-metoden är att välja de värden på parametrarna i fördelningen som gör observationerna så “troliga” som möjligt. En täthetsfunktion antar höga värden i områden där utfallen har stor sannolikhet att hamna, så “troligheten” hos utfallet $(x_1, \ldots, x_n)$ mäts av

$$L = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

De värden på parametrarna som ger det högsta värdet på L, den så kallade *likelihood-funktionen*, kallas ML-skattningarna. Hur det kan gå till att räkna ut ML-skattningarna framgår av följande exempel.

Exempel 2.3 Antag att variablerna $X_1, \ldots, X_n$ är gammafördelade, dvs har täthetsfunktionen

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\beta x}, \qquad x > 0$$

Likelihoodfunktionen blir då (vi skriver $L(\alpha, \beta)$ för att betona att vi betraktar L som en funktion av parametervärdena)

$$\begin{aligned} L(\alpha,\beta) &= \prod_{k=1}^{n} f(x_k) = \prod_{k=1}^{n} \left\{ \frac{\beta^\alpha}{\Gamma(\alpha)}\, x_k^{\alpha-1}\, e^{-\beta x_k} \right\} \\ &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \left(\prod_{k=1}^{n} x_k\right)^{\alpha-1} \exp\left\{-\beta \sum_{k=1}^{n} x_k\right\} \end{aligned}$$

Vi söker paret (α, β) som maximerar $L(\alpha, \beta)$. Det är ekvivalent, och för de flesta vanliga fördelningar enklare, att söka maximum till logaritmen av likelihoodfunktionen:

$$\begin{aligned}\ell(\alpha, \beta) &:= \log L(\alpha, \beta) \\ &= \alpha n \log \beta - n \log \Gamma(\alpha) + (\alpha - 1) \sum_{k=1}^{n} \log x_k - \beta \sum_{k=1}^{n} x_k\end{aligned}$$

Det vanligaste sättet att söka maximum är att finna den punkt (förutsatt att det finns en och endast en) där de partiella derivatorna är noll, dvs lösa ekvationssystemet

$$\begin{cases} \frac{\partial}{\partial \alpha} \ell(\alpha, \beta) = 0 \\ \frac{\partial}{\partial \beta} \ell(\alpha, \beta) = 0 \end{cases}$$

Dessa ekvationer kallas *ML-ekvationerna*. Vi finner att

$$\frac{\partial}{\partial \alpha} \ell(\alpha, \beta) = n \log \beta - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{k=1}^{n} \log x_k$$

och

$$\frac{\partial}{\partial \beta} \ell(\alpha, \beta) = \frac{\alpha n}{\beta} - \sum_{k=1}^{n} x_k$$

Vi låter $\bar{x}$ och $\tilde{x}$ beteckna de aritmetiska respektive geometriska medelvärdena av observationerna:

$$\bar{x} := \frac{1}{n} \sum_{k=1}^{n} x_k \qquad \tilde{x} := \left(\prod_{k=1}^{n} x_k \right)^{1/n}$$

Vi inför också beteckningen

$$\Psi(\alpha) := \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

Funktionen Ψ kallas digammafunktionen och finns som standardfunktion i t ex *Mathematica*. ML-ekvationerna kan nu skrivas

$$\begin{cases} \log \beta - \Psi(\alpha) + \log \tilde{x} = 0 \\ \alpha - \beta \bar{x} = 0 \end{cases}$$

Om man löser ut β i den andra ekvationen och stoppar in i den första, får man följande ekvation för α:

$$\log \alpha - \log \bar{x} - \Psi(\alpha) + \log \tilde{x} = 0$$

Denna ekvation måste lösas numeriskt (vi återkommer till numerisk ekvationslösning nedan), vilket ger ML-skattningen $\hat{\alpha}$ av α. Denna stoppas in i den andra ekvationen ovan, vilket ger ML-skattningen av β: $\hat{\beta} = \hat{\alpha}/\bar{x}$. □

Vi antar nu att variablerna $X_1, \ldots, X_1$ antar något av värdena $0, 1, 2, \ldots$ och sätter

$$p_k = P(X_i = k); \qquad k = 0, 1, 2, \ldots$$

När man har heltalsvariabler råkar man ofta ut för att ett stort antal observationer är lika. Vi låter n_k beteckna antalet observationer som är lika med k och noterar att $n_0 + n_1 + n_2 + \cdots = n$. Sannolikheten för det observerade utfallet $(x_1, \ldots, x_n)$ är

$$L = P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{j=1}^{n} p_{x_j} = \prod_k p_k^{n_k}$$

Vi söker de parametervärden som gör observationerna så sannolika som möjligt, dvs maximerar L, eller ekvivalent logaritmen av L:

$$\ell := \log L = \sum_k n_k \log p_k$$

Exempel 2.4 Antag att de observerade variablerna är binomialfördelade med parametrar r och q, dvs att

$$p_k = \binom{r}{k} q^k \, (1-q)^{r-k}; \qquad k = 0, 1, 2, \ldots, r$$

och antag att parametern r är känd. Logaritmen av likelihoodfunktionen är

$$\begin{aligned} \ell(\lambda) &= \sum_{k=0}^{m} n_k \left[\log \binom{r}{k} + k \log q + (r-k) \log(1-q) \right] \\ &= \sum_{k=0}^{m} n_k \log \binom{r}{k} + n\bar{x} \log q + n(r - \bar{x}) \log(1-q). \end{aligned}$$

Om man deriverar denna med avseende på q och sätter derivatan till noll, får man

$$\frac{n\bar{x}}{q} - \frac{n(r-\bar{x})}{1-q} = 0$$

Löser man ut q, får man $q = \bar{x}/r$, dvs ML-skattningen är identisk med momentmetodens skattning. $\square$

Som vi såg i exemplet med gammafördelningen, kan man råka ut för att man måste lösa en ekvation numeriskt (eller t o m ett system av ekvationer, men det tar vi inte upp här). Färdiga rutiner för detta finns i många moderna programpaket, men låt oss ändå titta på de vanligaste metoderna.

Antag att vi ska lösa ekvationen $f(x) = 0$. *Newton-Raphsons metod* går till så här: Man väljer ett startvärde x_0 och räknar ut tangenten till kurvan f i denna punkt, dvs den räta linjen

$$g(x) = f(x_0) + (x - x_0)\, f'(x_0)$$

Om vi startat tillräckligt nära roten bör tangenten skära x-axeln i en punkt som ligger närmare roten än x_0 (rita en figur!). Vi kallar denna punkt x_1:

$$f(x_0) + (x_1 - x_0)\, f'(x_0) = 0$$

dvs

$$x_1 = x_0 - f(x_0)/f'(x_0)$$

Så upprepar vi proceduren utgående från punkten x_1 och hamnar (förhoppningsvis) ännu närmare roten. Detta leder till en iterativ procedur:

$$x_n = x_{n-1} - f(x_{n-1})/f'(x_{n-1})$$

Metoden hittar för det mesta snabbt fram till roten, dvs iterationerna konvergerar, förutsatt att man startat någotsånär i närheten.

En alternativ metod, som är särskilt användbar när f' är tidskrävande att beräkna, är *sekant-metoden*. Denna utgår från två startvärden x_0 och x_1. Den räta linje som går genom punkterna $(x_0, f(x_0))$ och $(x_1, f(x_1))$ ges av

$$g(x) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}\, x + \frac{f(x_0)\, x_1 - f(x_1)\, x_0}{x_1 - x_0}$$

Denna linjen skär x-axeln i en punkt x_2, som ges av

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0}\, x_2 + \frac{f(x_0)\, x_1 - f(x_1)\, x_0}{x_1 - x_0} = 0$$

dvs

$$x_2 = \frac{f(x_1)\, x_0 - f(x_0)\, x_1}{f(x_1) - f(x_0)}$$

Man hoppas att x_2 ligger närmare roten än x_0 och x_1 (rita en figur!), upprepar proceduren utgående från x_1 och x_2, osv. Man får alltså iterationsformeln

$$x_n = \frac{f(x_{n-1})\, x_{n-2} - f(x_{n-2})\, x_{n-1}}{f(x_{n-1}) - f(x_{n-2})}$$

3 Modeller för skadebeloppen

Inom sakförsäkring förekommer ofta att man vill anpassa en sannolikhetsfördelning till data över utbetalda ersättningsbelopp för en viss typ av försäkring. En vanlig egenskap hos ett sådant datamaterial är att ett relativt litet antal stora skador svarar för en avsevärd andel av det sammanlagda skadebeloppet. Som exempel, betrakta nedanstående datamaterial från Länsförsäkringsbolagen (för mer information, se avsnitt 5.9). Data utgörs av utbetalda ersättningsbelopp för skador på lantbruksegendomar i Malmöhus län, orsakade av en storm 1990.

272	338	436	545	632	686	787	902	953
1090	1090	1090	1090	1090	1090	1166	1199	1253
1362	1473	1525	1634	1634	1743	1765	1810	2173
2179	2179	2179	2179	2289	2315	2322	2416	2615
2615	2664	2684	2724	2760	2857	3016	3132	3160
3190	3193	3210	3269	3269	3269	3269	3269	3322
3596	3633	3705	3782	3784	3814	3879	4183	4341
4358	4358	4358	4358	4358	4358	4364	4369	4462
4467	4510	4607	4634	4853	4893	4903	4903	4903
4911	4931	4937	4960	4974	5018	5060	5230	5257
5276	5448	5448	5448	5448	5448	5498	5729	5993
5993	6014	6033	6168	6226	6228	6429	6538	6538
6668	6805	6815	6842	6967	6973	7082	7162	7366
7679	7763	8017	8025	8028	8037	8045	8172	8286
8337	8362	8390	8624	8717	9008	9106	9697	9697
9806	9934	9965	10327	10395	10504	10825	10875	10896
10896	11314	11986	12687	12688	12853	13075	13075	13095
13144	13511	13668	13789	14165	14406	14428	14529	14758
15017	15037	15113	15419	15612	16420	17325	17434	18086
18523	18548	18659	18800	19099	20159	20760	21827	21972
22009	23564	23827	24196	25061	25269	25435	25785	26220
27240	28001	31384	37046	37424	39812	43880	44038	48383
54480	55141	56119	56768	67501	73831	80192	101276	198118
325326								

Tabell 3.1: *Ersättningsbelopp för stormskador.*

Beloppen är med hjälp av lämpliga index omräknade i 1994 års penningvärde. En del skadebelopp förekommer flera gånger, vilket oftast beror på att det icke inflationskorrigerade beloppet är ett “jämnt” tal. Exempelvis var beloppet 4358, som förekommer sex gånger, lika med 4000 före korrigering för inflation. Datamaterialet är typiskt “långsvansat”: De tjugo största av de över 200 skadorna svarar för ungefär hälften av det sammanlagda skadebeloppet.

Vi ska studera de två mest använda så kallade långsvansade fördelningarna inom sakförsäkringsmatematiken, Paretofördelningen och lognormalfördelningen. För att illustrera hur man genom att införa fler parametrar kan uppnå bättre anpassning, tar vi också upp en generalisering av Paretofördelningen, den så kallade Burrfördelningen.

Efter en genomgång av varje fördelnings grundläggande egenskaper, går vi igenom hur man kan anpassa fördelningen till data av den typ som återfinns i Tabell 3. Vårt grundantagande kommer att vara att ersättningsbeloppen kan betraktas som utfall av oberoende och likafördelade stokastiska variabler.

Det är i allmänhet svårt att finna "fysikaliska"motiveringar för att använda en viss fördelning, utan man får empiriskt söka sig fram till modeller som fungerar bra. Med detta menas att de skall vara tillräckligt flexibla för att kunna anpassas väl till data och samtidigt vara bekväma att arbeta med. Dessa två egenskaper är motstridiga: ökad flexibilitet erhålls i allmänhet genom att göra modellen mer komplicerad och öka antalet parametrar, vilket leder till att beräkningar blir krångligare och parametrarna svårare att skatta.

Det finns åtskilliga andra fördelningar än de vi tar upp som använts för att modellera skadebelopp (ytterligare ett par exempel finns i övningarna till det här kapitlet). För en översikt, se van der Laan & Hop (1989). Hogg & Klugman (1984) är en hel bok som handlar om att anpassa skadefördelningar.

3.1 Paretofördelningen

Paretofördelningens täthet är

$$f(x) = \frac{\gamma\,\alpha^\gamma}{(\alpha+x)^{\gamma+1}}, \qquad x > 0$$

Fördelningsfunktionen är ovanligt enkel:

$$\begin{aligned} F(x) &= \int_0^x f(y)\,dy = \int_0^x \frac{\gamma\,\alpha^\gamma}{(\alpha+y)^{\gamma+1}}\,dy \\ &= \left[-\frac{\alpha^\gamma}{(\alpha+y)^\gamma}\right]_0^x = 1 - \left(\frac{\alpha}{\alpha+x}\right)^\gamma \end{aligned}$$

Väntevärdesfunktionen är också enkel, i den meningen att den kan uttryckas med hjälp av elementära funktioner. Man har, för $\gamma \neq 1$, att

$$\begin{aligned}
\int_0^y x\,f(x)\,dx &= \int_0^y x\,\frac{\gamma\,\alpha^\gamma}{(\alpha+x)^{\gamma+1}}\,dx \\
&= \left[-\frac{x\,\alpha^\gamma}{(\alpha+x)^\gamma}\right]_0^y + \int_0^y \frac{\alpha^\gamma}{(\alpha+x)^\gamma}\,dx \\
&= -\frac{y\,\alpha^\gamma}{(\alpha+y)^\gamma} + \left[-\frac{\alpha^\gamma}{(\gamma-1)(\alpha+x)^{\gamma-1}}\right]_0^y \\
&= -\frac{y\,\alpha^\gamma}{(\alpha+y)^\gamma} + \frac{\alpha^\gamma}{\gamma-1}\left(\frac{1}{\alpha^{\gamma-1}} - \frac{1}{(\alpha+y)^{\gamma-1}}\right) \\
&= -\frac{y\,\alpha^\gamma}{(\alpha+y)^\gamma} + \frac{\alpha}{\gamma-1}\left(1 - \left(\frac{\alpha}{\alpha+y}\right)^{\gamma-1}\right)
\end{aligned}$$

Väntevärdesfunktionen blir alltså

$$L(y) = \int_0^y x\,f(x)\,dx + y(1-F(y)) = \frac{\alpha}{\gamma-1}\left(1 - \left(\frac{\alpha}{\alpha+y}\right)^{\gamma-1}\right)$$

Om $\gamma = 1$, så får vi istället att

$$\begin{aligned}
\int_0^y x\,f(x)\,dx &= \int_0^y x\,\frac{\alpha}{(\alpha+x)^2}\,dx \\
&= \int_0^y \left(\frac{\alpha}{\alpha+x} - \frac{\alpha^2}{(\alpha+x)^2}\right)dx \\
&= \left[\alpha\log(\alpha+x) + \frac{\alpha^2}{\alpha+x}\right]_0^y \\
&= -\alpha\log\left(\frac{\alpha}{\alpha+y}\right) - \frac{\alpha y}{\alpha+y}
\end{aligned}$$

så att vi får väntevärdesfunktionen

$$L(y) = \int_0^y x\,f(x)\,dx + y(1-F(y)) = -\alpha\log\left(\frac{\alpha}{\alpha+y}\right)$$

Den momentgenererande funktionen existerar inte för $t > 0$ och är inte till någon hjälp vid studiet av fördelningen. Momenten går att räkna ut direkt, men som vi ska se, existerar de inte för alla parametervärden.

Väntevärdet kan erhållas genom att låta $y \to \infty$ i väntevärdesfunktionen. Förutsatt att $\gamma > 1$, fås att

$$\kappa_1 = \alpha_1 = \frac{\alpha}{\gamma-1}$$

Andramomentet kring origo kan räknas ut med hjälp av upprepad partiell integration. Här måste vi förutsätta att $\gamma > 2$:

$$\begin{aligned}
\alpha_2 &= \int_0^\infty x^2 f(x)\,dx = \int_0^\infty x^2\,\frac{\gamma\,\alpha^\gamma}{(\alpha+x)^{\gamma+1}}\,dx \\
&= \left[-x^2\,\frac{\alpha^\gamma}{(\alpha+x)^\gamma}\right]_0^\infty + \int_0^\infty 2x\,\frac{\alpha^\gamma}{(\alpha+x)^\gamma}\,dx \\
&= 0 + \left[-2x\,\frac{\alpha^\gamma}{(\gamma-1)(\alpha+x)^{\gamma-1}}\right]_0^\infty + \int_0^\infty 2\,\frac{\alpha^\gamma}{(\gamma-1)(\alpha+x)^{\gamma-1}}\,dx \\
&= 0 + \left[-\frac{2\,\alpha^\gamma}{(\gamma-1)(\gamma-2)(\alpha+x)^{\gamma-2}}\right]_0^\infty = \frac{2\,\alpha^2}{(\gamma-1)(\gamma-2)}
\end{aligned}$$

Vi lämnar åt läsaren att på liknande sätt verifiera att, om $\gamma > 3$,

$$\alpha_3 = \frac{6\,\alpha^3}{(\gamma-1)(\gamma-2)(\gamma-3)} \tag{3.1}$$

Från α_1, α_2 och α_3 kan de tre första centralmomenten räknas ut. Paretofördelningen tillhör de mer långsvansade fördelningarna, vilket återspeglas i det faktum att momenten inte alltid existerar.

Vi ska nu kortfattat härleda hur man räknar ut momentmetodens och ML-metodens skattningar av Paretofördelningens parametrar. Därefter ska vi räkna ut skattningarna baserade på data från Tabell 3 och jämföra de sålunda anpassade fördelningarna med den empiriska fördelningen.

De två första momenten kring origo för Paretofördelningen är, som vi nyss såg,

$$\alpha_1 = \frac{\alpha}{\gamma-1} \qquad \alpha_2 = \frac{2\alpha^2}{(\gamma-1)(\gamma-2)}$$

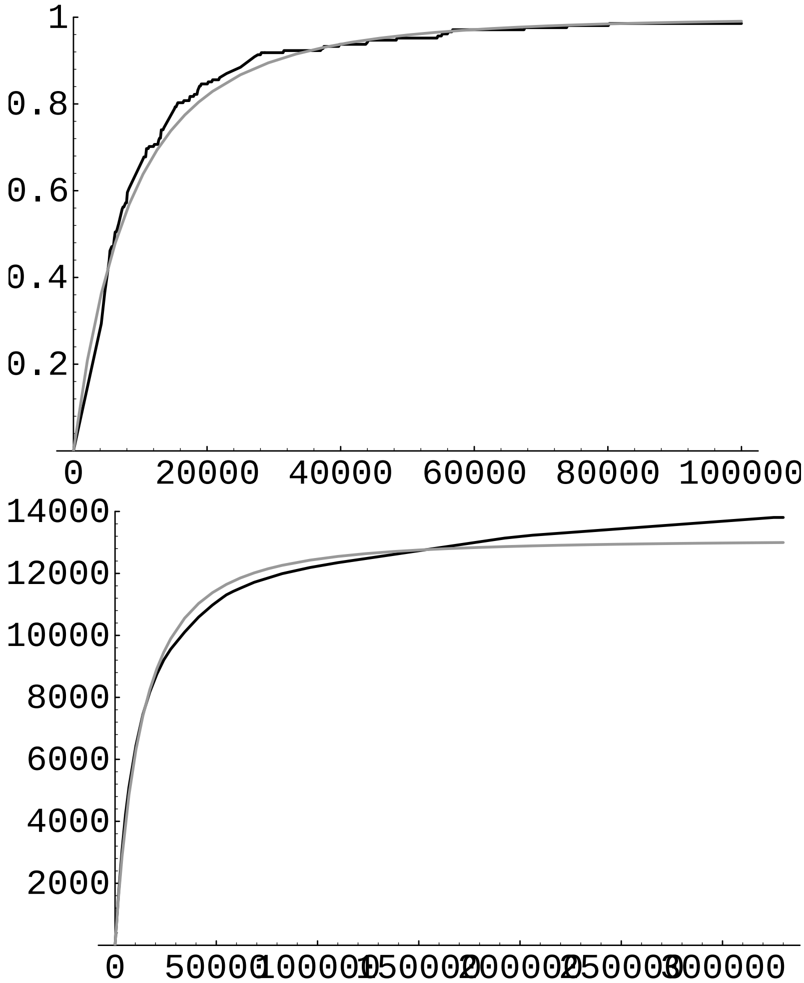
Variansen blir

$$\alpha_2 - \alpha_1^2 = \frac{\gamma\,\alpha^2}{(\gamma-1)^2(\gamma-2)}$$

Momentmetodens skattningar räknas lätt ut till

$$\hat{\alpha} = \bar{x}\,\frac{s^2+\bar{x}^2}{s^2-\bar{x}^2} \qquad \hat{\gamma} = \frac{2\,s^2}{s^2-\bar{x}^2}$$

Observera att eftersom α_2 bara existerar då $\gamma > 2$, är momentmetoden olämplig att använda om man misstänker att mindre värden på γ ger den bästa anpassningen.



Figur 3.1: *Anpassning av Paretofördelning. Empiriska fördelningsfunktionen (mörkast) och fördelningsfunktionen med ML-metodens skattningar av parametrarna instoppade. Under motsvarande väntevärdesfunktioner.*

Så till ML-metoden. Logaritmen av likelihoodfunktionen, baserad på observationerna $x_1, \dots, x_n$, är

$$\ell(\alpha, \gamma) = n \log \gamma + n\gamma \log \alpha - (\gamma + 1) \sum_{k=1}^{n} \log(\alpha + x_k)$$

Vi inför funktionerna

$$a(\alpha) := \frac{1}{n} \sum_{k=1}^{n} \frac{\alpha}{\alpha + x_k}, \qquad b(\alpha) := \frac{1}{n} \sum_{k=1}^{n} \log \frac{\alpha}{\alpha + x_k}$$

ML-ekvationerna

$$\begin{cases} \frac{\partial}{\partial \alpha} \ell(\alpha, \gamma) = 0 \\ \frac{\partial}{\partial \gamma} \ell(\alpha, \gamma) = 0 \end{cases}$$

kan då skrivas

$$\begin{cases} \gamma = a(\alpha)/(1-a(\alpha)) \\ 1/\gamma + b(\alpha) = 0 \end{cases}$$

Stoppar man in den första ekvationen i den andra erhålls följande ekvation för α:

$$\frac{1}{a(\alpha)} + b(\alpha) = 1$$

Denna löses numeriskt och lösningen, dvs ML-skattningen $\hat{\alpha}$, stoppas in i den första ekvationen för att ge ML-skattningen av γ:

$$\hat{\gamma} = a(\hat{\alpha})/(1-a(\hat{\alpha}))$$

För vårt datamaterialet i Tabell 3 finner vi momentskattningarna

$$\hat{\alpha} = 21808.5 \qquad \hat{\gamma} = 2.57962$$

och ML-skattningarna

$$\hat{\alpha} = 24629.3 \qquad \hat{\gamma} = 2.88246$$

I Figur 3.1 ser vi de empiriska fördelnings- och väntevärdesfunktionerna jämförda med motsvarande objekt för Paretofördelningen med ML-skattningarna instoppade.

Övning 3.1 *Visa (3.1).*

Övning 3.2 *Följande data på brandskadebelopp är hämtade från en artikel av J.D. Cummins och L.R. Freifelder:*

Anpassa en Paretofördelning till data och jämför den anpassade fördelningens fördelnings- och väntevärdesfunktioner med den empiriska fördelningens. Hur förändras resultaten om du tar bort den största skadan från datamaterialet?

Övning 3.3 *En fördelning som använts för att modellera storleken på skadebelopp inom sakförsäkring är Weibullfördelningen, som har fördelningsfunktionen*

$$F(x) = 1 - e^{-cx^{\tau}}, \qquad x \geq 0, \qquad c > 0, \qquad \tau > 0$$

Låt $x_1, \ldots, x_n$ *vara* n *oberoende observationer från en Weibullfördelning och sätt*

$$m_1 := \frac{1}{n}\sum_{k=1}^{n} x_k \qquad m_2 := \frac{1}{n}\sum_{k=1}^{n} x_k^2$$

290.40	1248.49	2202.96	3941.30	10560.10
537.19	1268.24	2222.80	4017.01	11179.54
756.80	1284.56	2255.72	4100.00	11461.39
769.19	1363.85	2274.61	4166.98	14538.13
787.69	1436.20	2328.64	4355.02	14789.81
796.18	1445.96	2384.37	5117.93	17186.09
933.62	1469.48	2847.83	5335.96	18582.57
967.97	1507.47	2947.04	5453.02	22857.33
1010.56	1662.36	2948.35	5568.96	23177.85
1017.40	1674.58	3036.51	5761.83	23446.13
1033.49	1690.91	3286.68	6161.81	28409.82
1034.33	1739.96	3331.62	6348.69	57612.82
1056.93	1776.56	3416.67	6859.37	59582.78
1124.09	1932.09	3604.66	7972.20	113164.70
1165.73	1975.89	3671.16	8028.32	123228.90
1217.64	2099.79	3739.30	10047.22	626402.80

Visa att momentmetodens skattningar av parametrarna c och τ kan erhållas genom att först (numeriskt) finna $\hat{\tau}$ som löser ekvationen

$$m_1^2\,\Gamma(1+2/\tau) = m_2\big(\Gamma(1+1/\tau)\big)^2$$

och därefter beräkna

$$\hat{c} = \left(\frac{\Gamma(1+1/\hat{\tau})}{m_1}\right)^{\hat{\tau}}$$

3.2 Burrfördelningen

Paretofördelningen kan generaliseras och göras mer flexibel genom att man inför ytterligare en parameter. Låt X vara Paretofördelad med parametrar α och γ och sätt $Y = X^{1/\tau}$, där $\tau > 0$. Då får Y fördelningsfunktionen

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(X^{1/\tau} \le y) \\ &= P(X \le y^\tau) = 1 - \frac{\alpha^\gamma}{(\alpha + y^\tau)^\gamma} \end{aligned}$$

och tätheten

$$f_Y(y) = F'_Y(y) = \frac{\gamma\tau\alpha^\gamma y^{\tau-1}}{(\alpha + y^\tau)^{\gamma+1}}$$

Denna fördelning kallas för *Burrfördelningen*.

Väntevärdesfunktionen för Burrfördelningen kan uttryckas med hjälp av så kallade betafunktioner, som finns som standardfunktioner i moderna programpaket. Betafunktionen $B(\cdot,\cdot)$

definieras för $\alpha > 0$ och $\gamma > 0$ via

$$B(\alpha, \gamma) = \int_0^1 x^{\alpha-1} (1-x)^{\gamma-1} \, dx$$

Man kan visa att

$$B(\alpha, \gamma) = \frac{\Gamma(\alpha)\,\Gamma(\gamma)}{\Gamma(\alpha+\gamma)}. \tag{3.2}$$

Den ofullständiga betafunktionen $J_{\alpha,\gamma}$ definieras av

$$J_{\alpha,\gamma}(y) := \int_0^y u^{\alpha-1} (1-u)^{\gamma-1} \, du, \qquad 0 \le y \le 1$$

så att $B(\alpha, \gamma) = J_{\alpha,\gamma}(1)$. Med dessa beteckningar får vi att

$$\begin{aligned}
\int_0^x y\, f(y)\, dy = \int_0^x y\, \frac{\gamma\tau\alpha^\gamma y^{\tau-1}}{(\alpha+y^\tau)^{\gamma+1}}\, dy \\
= \int_0^x y\,\gamma \left(\frac{\alpha}{\alpha+y^\tau}\right)^{\gamma-1} \frac{\alpha\tau y^{\tau-1}}{(\alpha+y^\tau)^2}\, dy \\
\{\text{Substituera: } u = \alpha/(\alpha+y^\tau),\ du/dy = -\alpha\tau y^{\tau-1}/(\alpha+y^\tau)^2\} \\
= \int_{\alpha/(\alpha+x^\tau)}^1 \left(\frac{\alpha(1-u)}{u}\right)^{1/\tau} \gamma\, u^{\gamma-1}\, du \\
= \gamma\alpha^{1/\tau} \int_{\alpha/(\alpha+x^\tau)}^1 u^{\gamma-1/\tau-1} (1-u)^{1/\tau}\, du \\
= \gamma\alpha^{1/\tau} \left(B(\gamma-1/\tau, 1+1/\tau) - J_{\gamma-1/\tau,1+1/\tau}\left(\frac{\alpha}{\alpha+x^\tau}\right)\right)
\end{aligned}$$

Väntevärdesfunktionen blir alltså

$$\begin{aligned}
L(x) &= \int_0^x y\, f(y)\, dy + x(1-F(x)) \\
&= \gamma\alpha^{1/\tau} \left(B(\gamma-1/\tau, 1+1/\tau) - J_{\gamma-1/\tau,1+1/\tau}\left(\frac{\alpha}{\alpha+x^\tau}\right)\right) + x\left(\frac{\alpha}{\alpha+x^\tau}\right)^\gamma
\end{aligned}$$

Momenten kring origo blir

$$\begin{aligned}
\int_0^\infty x^k\, f(x)\, dx &= \int_0^\infty x^k\, \frac{\gamma\,\tau\,\alpha^\gamma\, x^{\tau-1}}{(\alpha+x^\tau)^{\gamma+1}}\, dx \\
&= \int_0^\infty x^k\,\gamma \left(\frac{\alpha}{\alpha+x^\tau}\right)^{\gamma-1} \frac{\alpha\,\tau\, x^{\tau-1}}{(\alpha+x^\tau)^2}\, dx \\
&\quad \{\text{ Samma substitution som ovan }\} \\
&= \int_0^1 \left(\frac{\alpha(1-u)}{u}\right)^{k/\tau} \gamma\, u^{\gamma-1}\, du \\
&= \gamma\,\alpha^{k/\tau} \int_0^1 u^{\gamma-k/\tau-1} (1-u)^{k/\tau}\, du \\
&= \gamma\,\alpha^{k/\tau}\, B(\gamma-k/\tau, 1+k/\tau)
\end{aligned}$$

För att α_k ska existera, måste det gälla att $\gamma > k/\tau$.

Vi ska nu se att man kan uppnå en markant förbättring i anpassningen till stormskadematerialet genom att gå från Pareto till dess generalisering Burr. Införandet av ytterligare en parameter ger den flexibilitet som behövs för att få en bra anpassning. Vi nöjer oss med att ta upp ML-metoden.

Tätheten för Burrfördelningen är

$$f(x) = \frac{\gamma\,\tau\,\alpha^\gamma\,x^{\tau-1}}{(\alpha + x^\tau)^{\gamma+1}}$$

så likelihoodfunktionen blir

$$L(\alpha,\gamma,\tau) = \gamma^n\,\tau^n\,\alpha^{n\gamma}\left(\prod_{k=1}^n x_k\right)^{\tau-1}\left(\prod_{k=1}^n (\alpha + x_k^\tau)\right)^{-(\gamma+1)}$$

Den logaritmerade likelihoodfunktionen blir alltså

$$\begin{aligned}\ell(\alpha,\gamma,\tau) &= n\log\gamma + n\log\tau + n\gamma\log\alpha \\ &+ (\tau-1)\sum_{k=1}^n \log x_k - (\gamma+1)\sum_{k=1}^n \log(\alpha + x_k^\tau)\end{aligned}$$

För att finna ML-skattningarna kan man gå tillväga på följande sätt. Välj något startvärde τ_0 på parametern τ, till exempel $\tau_0 = 1$ (dvs Pareto). Sök sedan α och γ som maximerar funktionen $\ell(\alpha,\gamma,\tau_0)$ genom att lösa ekvationssystemet

$$\begin{cases} \frac{\partial \ell(\alpha,\gamma,\tau_0)}{\partial\alpha} = 0 \\ \frac{\partial \ell(\alpha,\gamma,\tau_0)}{\partial\gamma} = 0 \end{cases}$$

Om vi inför beteckningarna

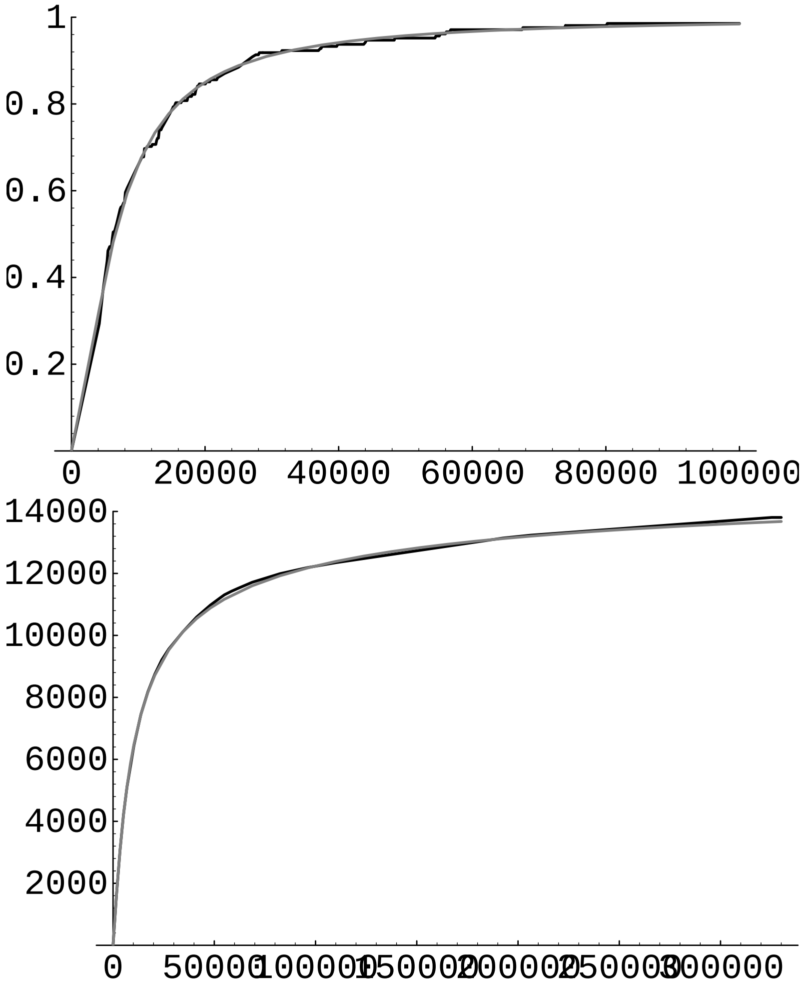
$$a_\tau(\alpha) := \frac{1}{n}\sum_{k=1}^n \frac{\alpha}{\alpha + x_k^\tau} \qquad b_\tau(\alpha) := \frac{1}{n}\sum_{k=1}^n \log\frac{\alpha}{\alpha + x_k^\tau}$$

kan vi göra detta (jämför Pareto) genom att först finna α_0 som löser

$$1/a_{\tau_0}(\alpha) + b_{\tau_0}(\alpha) = 1;$$

och därefter räkna ut

$$\gamma_0 = \frac{a_{\tau_0}(\alpha_0)}{1 - a_{\tau_0}(\alpha_0)}$$



Figur 3.2: *Den empiriska fördelningsfunktionen och väntevärdesfunktionen (mörkast) jämförd med motsvarande objekt för "ML-skattad" Burrfördelning.*

För $\alpha = \alpha_0$ och $\gamma = \gamma_0$ får vi det högsta värdet på $\ell(\alpha, \gamma, \tau_0)$ (om $\tau_0 = 1$ får vi förstås ML-skattningarna för Paretofördelningen). Så prövar vi med ett annat värde τ_1 på τ, erhåller med proceduren ovan α_1 och γ_1 som maximerar $\ell(\alpha, \gamma, \tau_1)$ och undersöker om $\ell(\alpha_1, \gamma_1, \tau_1) > \ell(\alpha_0, \gamma_0, \tau_0)$. På detta sätt fortsätter man att söka sig fram till ett så högt värde på $\ell(\alpha_n, \gamma_n, \tau_n)$ som möjligt. Man har då funnit ML-skattningarna. Om man vill, skriver man ett program som utför allt detta.

För datamaterialet i Tabell 3 finner man med denna metod ML-skattningarna

$$\hat{\alpha} = 1.638 \cdot 10^6 \qquad \hat{\gamma} = 0.8924 \qquad \hat{\tau} = 1.647$$

I Figur 3.2 ser man att anpassningen till den empiriska fördelningen är mycket god.

Övning 3.4 *Låt den stokastiska variabeln Y vara gammafördelad med parametrar γ och λ, dvs ha tätheten*

$$f_Y(y) = \frac{\lambda^\gamma}{\Gamma(\gamma)} y^{\gamma-1} e^{-\lambda y}\,, \qquad y > 0$$

Vidare, låt den betingade fördelningen för variabeln X, givet $Y = y$, vara en gammafördelning med parametrar α och y, dvs

$$f_X\big(x \bigm| Y = y\big) = \frac{y^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-yx}, \qquad x > 0$$

a) Visa att tätheten för X är

$$f_X(x) = \frac{\Gamma(\alpha+\gamma)\,\lambda^\gamma\, x^{\alpha-1}}{\Gamma(\alpha)\,\Gamma(\gamma)\,(\lambda+x)^{\alpha+\gamma}}, \qquad x > 0$$

Denna fördelning kallas för den generaliserade Paretofördelningen i Hogg & Klugman (1984). För $\alpha = 1$ får man Paretofördelningen som specialfall.

b) Låt $G_{\alpha,\gamma}$ vara fördelningsfunktionen för betafördelningen:

$$G_{\alpha,\gamma}(x) = \frac{1}{B(\alpha,\gamma)} \int_0^x u^{\alpha-1}\,(1-u)^{\gamma-1}\,du = \frac{J_{\alpha,\gamma}(x)}{B(\alpha,\gamma)}$$

Denna finns med i moderna programpaket eller räknas lätt ut om betafunktionerna finns. Visa att fördelningsfunktionen respektive väntevärdesfunktionen för den generaliserade Paretofördelningen kan skrivas som

$$\begin{aligned} F(x) &= G_{\alpha,\gamma}\left(\frac{x}{\lambda+x}\right) \\ L(x) &= \frac{\alpha\lambda}{\gamma-1}\, G_{\alpha+1,\gamma-1}\left(\frac{x}{\lambda+x}\right) + x\left(1 - G_{\alpha,\gamma}\left(\frac{x}{\lambda+x}\right)\right) \end{aligned}$$

c) Visa med hjälp av (3.2) att momenten kring origo för den generaliserade Paretofördelningen är

$$\alpha_k = \lambda^k \prod_{i=0}^{k-1}(\alpha+i) \Bigg/ \prod_{i=1}^{k}(\gamma-i), \qquad \gamma > k$$

Övning 3.5 *Anpassa en Burrfördelning till datamaterialet i Övning 3.2 och undersök om anpassningen blir märkbart bättre än för Paretofördelningen. Gör samma sak med den största skadan bortplockad.*

3.3 Lognormalfördelningen

En lognormalfördelad stokastisk variabel kan skrivas som e^X, där X är normalfördelad med väntevärde μ och standardavvikelse σ. Variabeln $(X - \mu)/\sigma$ har en normalfördelning med väntevärde 0 och standardavvikelse 1, vars fördelningsfunktion brukar skrivas $\Phi(\cdot)$. En lognormalfördelnad variabel har alltså fördelningsfunktionen

$$\begin{aligned} F(x) &= P(e^X \leq x) = P(X \leq \log x) \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{\log x - \mu}{\sigma}\right) = \Phi\left(\frac{\log x - \mu}{\sigma}\right) \end{aligned}$$

Tätheten är

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \qquad x > 0$$

Väntevärdesfunktionen är (härledningen lämnas som övning)

$$L(x) = e^{\mu+\sigma^2/2}\,\Phi\left(\frac{\log x - \mu - \sigma^2}{\sigma}\right) + x\left(1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right) \tag{3.3}$$

Momenten kring origo kan erhållas från den momentgenererande funktionen för normalfördelningen:

$$M(t) = E\big[e^{tX}\big] = e^{\mu t + \sigma^2 t^2/2}$$

Man får att

$$\alpha_k = E\big[(e^X)^k\big] = E\big[e^{kX}\big] = M(k) = e^{\mu k + \sigma^2 k^2/2}$$

Så tar vi itu med momentmetoden för skattning av parametrarna μ och σ i lognormalfördelningen. Tidigare har vi "matchat" de två första centralmomenten, men det blir samma resultat om man matchar de två första momenten kring origo. För den empiriska fördelningen är dessa

$$m_1 := \frac{1}{n}\sum_{k=1}^{n} x_k \qquad \text{och} \qquad m_2 := \frac{1}{n}\sum_{k=1}^{n} x_k^2$$

De två första momenten kring origo för lognormalfördelningen har vi räknat ut till

$$\alpha_1 = e^{\mu+\sigma^2/2} \qquad \alpha_2 = e^{2(\mu+\sigma^2)}$$

Momentmetodens skattningar $\hat{\mu}$ och $\hat{\sigma}$ av μ och σ ges av lösningen till ekvationssystemet

$$\begin{cases} m_1 = \alpha_1 \\ m_2 = \alpha_2 \end{cases}$$

dvs

$$\begin{cases} \hat{\mu} = 2\log m_1 - \frac{1}{2}\log m_2 \\ \hat{\sigma} = \sqrt{\log\ m_2 - 2\log m_1} \end{cases}$$

ML-skattningarna är också lätta att räkna ut. Låt $f_{\mu,\sigma}$ beteckna täthetsfunktionen för en normalfördelad variabel med väntevärde μ och varians σ^2. Om $x_1, \ldots, x_n$ är observationer från en lognormalfördelning med parametrar μ och σ, är logaritmen av likelihoodfunktionen

$$\ell(\mu, \sigma) = -\sum_{k=1}^{n} \log x_k + \sum_{k=1}^{n} \log f_{\mu,\sigma}(\log x_k)$$

Att finna μ och σ som maximerar ℓ är alltså detsamma som att finna ML-skattningarna i normalfördelningen givet observationerna $\log x_1, \ldots, \log x_n$. Man får alltså att

$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} \log x_k \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} (\log x_k)^2 - \hat{\mu}^2$$

Läsaren uppmanas att själv studera anpassningen till datamaterialet i Tabell 3.

Övning 3.6 *Visa (3.3).*

Övning 3.7 *Den så kallade inversa normalfördelningen är en fördelning på* $(0, \infty)$ *med täthetsfunktionen*

$$f_{\mu,\beta}(x) = \frac{\mu}{\sqrt{2\pi\beta x^3}} \exp\left\{-\frac{(x-\mu)^2}{2\beta x}\right\}, \qquad x > 0$$

Fördelningens parametrar är $\mu > 0$ *och* $\beta > 0$.

a) Visa att den momentgenererande funktionen är

$$M(t) = \exp\left\{\frac{\mu}{\beta}\left(1 - \sqrt{1 - 2\beta t}\right)\right\}$$

genom att utnyttja att

$$\int_0^\infty f_{\mu,\beta}(x)\, dx = 1$$

för alla $\mu > 0$ *och* $\beta > 0$.

b) Beräkna de tre första centralmomenten och momenten kring origo.

c) Vad kan sägas om summor av oberoende variabler med inversa normalfördelningar?

Övning 3.8 *Anpassa lognormalfördelningar till datamaterialen i Tabell 3 respektive Övning 3.2. Jämför anpassningen med resultaten för Pareto och Burr.*

Övning 3.9 *Den inversa normalfördelningen har tätheten*

$$f(x) = \frac{\mu}{\sqrt{2\pi\beta x^3}} \exp\left\{-\frac{(x-\mu)^2}{2\beta x}\right\}, \qquad x > 0$$

med parametrarna $\mu > 0$ och $\beta > 0$.

a) Visa att ML-skattningarna av μ och β, då man observerat $x_1, \ldots, x_n$ är

$$\hat{\mu} = \bar{x}, \qquad \hat{\beta} = \bar{x}(\bar{x}\tilde{x} - 1)$$

där

$$\tilde{x} = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{x_k}$$

b) Studera anpassningen av en invers normalfördelning till datamaterialet i Övning 3.2. Man kan visa (se Shuster, 1968) att fördelningsfunktionen kan skrivas som

$$F(x) = \Phi\left(\frac{x-\mu}{\sqrt{\beta x}}\right) + e^{2\mu/\beta}\,\Phi\left(-\frac{x+\mu}{\sqrt{\beta x}}\right), \qquad x > 0$$

där $\Phi(\cdot)$ är fördelningsfunktionen för en $N(0,1)$-variabel. Väntevärdesfunktionen kan skrivas som (se ter Berg, 1994):

$$L(x) = x\,(1 - F(x)) - \mu\,F(x) + 2\mu\big(1 - \Phi\big((\mu - x)\,(\beta x)^{-1/2}\big)\big)$$

3.4 Några vanliga komplikationer

Skadeförsäkringsdata behövs ofta justeras på ett eller annat sätt innan de ovan beskrivna skattningsmetoderna kan användas. Det finns också situationer där metoderna behöver utvidgas. I det här avsnittet ger vi exempel på detta.

Inflation

För att få ihop ett tillräckligt stort datamaterial, tvingas man för det mesta använda skadedata från en följd av år. Som bekant varierar ju exempelvis byggkostnader eller sjukvårdskostnader över åren, varför skadedata kan behöva korrigeras med en lämplig kombination av prisindex. Man räknar förslagsvis om priserna i det senaste årets penningvärde. Med hjälp av prognoser för framtida inflation kan man sedan justera den anpassade fördelningen till kommande års penningvärde.

Om X är ett godtyckligt skadebelopp, uttryckt i ett visst års penningvärde, får vi ju beloppet uttryckt i ett annat års penningvärde genom att multiplicera med en lämplig faktor a. Notera att för de fördelningar vi tagit upp gäller att aX har samma fördelningstyp som X, men med modifierade parametrar. Om X exempelvis är Paretofördelad med parametrar α och γ, gäller att

$$\begin{aligned} P(aX \leq x) &= P(X \leq x/a) = 1 - \left(\frac{\alpha}{\alpha + x/a}\right)^{\gamma} \\ &= 1 - \left(\frac{a\alpha}{a\alpha + x}\right)^{\gamma} \end{aligned}$$

dvs aX är Paretofördelad med parametrar $a\alpha$ och γ.

Fördröjda utbetalningar

Det senaste årets data är det mest aktuella då man ska anpassa en skadefördelning. Det händer emellertid, särskilt inom ansvarsförsäkring, att många av det senaste årets skador inte är slutbetalda. För till exempel personskador i samband med trafikolyckor kan det dröja åratal innan skadan slutregleras. Ersättningen kan betalas ut successivt under loppet av flera år. Man är då tvungen att använda en uppskattning av det slutgiltiga skadebeloppet som observation. Denna uppskattning kan antingen bygga på en erfaren skadereglerares bedömning eller en statistisk undersökning av utvecklingsmönstret för enskilda skador (eller en kombination av dessa).

Anhopning av observationer

Det finns en tendens att avrunda kostnader till jämna belopp, varför en del värden (till exempel jämna tusental) kan förekomma åtskilliga gånger i ett datamaterial. Att försöka modellera detta leder dock för det mesta till alltför komplicerade modeller. De fördelningar vi tagit upp i tidigare avsnitt passar ofta bra trots att man har anhopningar i en del punkter.

Självrisker

En självriskklausul innebär att försäkringsbolaget inte ersätter skador vars kostnader understiger en viss nivå s, självrisken. Detta innebär i praktiken att sådana skador inte rapporteras till försäkringsbolaget och datamaterialet innehåller således endast skador som överstiger s. De fördelningar vi hittills diskuterat är ju fördelningar på $[0, \infty)$, varför en viss justering måste till.

Låt X vara en godtycklig skada (det är alltså även möjligt att X hamnar under s), där vi antar att X har fördelningsfunktionen F. Om Y är en godtycklig skada i datamaterialet, som ju endast innehåller de skador som överstiger s, gäller för $y \geq s$ att

$$\begin{aligned}
G(y) &:= P(Y \leq y) = P\big(X \leq y \mid X > s\big) \\
&= 1 - P\big(X > y \mid X > s\big) = 1 - \frac{P(X > y,\, X > s)}{P(X > s)} \\
&= 1 - \frac{P(X > y)}{P(X > s)} = 1 - \frac{1 - F(y)}{1 - F(s)}
\end{aligned}$$

Om $f = F'$ är tätheten för X, gäller alltså att Y har tätheten g, där

$$g(y) = \frac{f(y)}{1 - F(s)}, \qquad y > s$$

Om observationerna är $y_1, \dots, y_n$ blir den logaritmerade likelihoodfunktionen

$$\ell = \sum_{k=1}^{n} \log g(y_k) = \sum_{k=1}^{n} \log f(y_k) - \sum_{k=1}^{n} \log\big(1 - F(s)\big)$$

ML-ekvationerna kan sedan härledas på vanligt sätt. I fallen Pareto eller Burr är dessa ekvationer inte svårare att lösa än vanligt, medan det blir något mer komplicerat då man antar att observationerna är lognormalfördelade.

Svansskattning

I en del situationer är man endast intresserad av hur skadefördelningen ser ut över en viss punkt c, dvs man vill anpassa en fördelningsfunktion F i punkterna $x > c$. Detta kan göras på följande sätt: Om X är en godtycklig observation, gäller att

$$\begin{aligned}
F(x) &:= P(X \leq x) \\
&= P\big(X \leq x \mid X \leq c\big)P\big(X \leq c\big) + P\big(X \leq x \mid X > c\big)P\big(X > c\big)
\end{aligned}$$

Om vi sätter $p := P(X > c)$ och $G(y) := P\big(X - c \le y \bigm| X > c\big)$, gäller att

$$F(x) = 1 - p + p\,G(x - c), \qquad x > c. \tag{3.4}$$

Om vi ur datamaterialet tar bort alla observationer som understiger c, samt drar bort c från de övriga, återstår observationer på $[0, \infty)$ och vi kan försöka finna en fördelningsfunktion G (till exempel Pareto, lognormal, $\ldots$) som passar bra till detta modifierade datamaterial. Sannolikheten p kan helt enkelt skattas som andelen av de ursprungliga observationerna som översteg c, och därefter kan vi beräkna en skattning av $F(x)$ via (3.4).

Blandade fördelningar

Ibland kan det vara svårt att hitta en fördelning som passar bra överallt. Inom brandförsäkring kan det till exempel förekomma ett stort antal småskador (branden släcktes på ett tidigt stadium) och en ansamling av större skador (branden fick fäste och spred sig), så att fördelningen ser ut att ha två “pucklar”. I sådana situationer kan man uppnå en bättre anpassning med en blandning av två fördelningar.

Låt oss tänka på brandexemplet och låt X vara skadebeloppet för en godtycklig skada . Vidare, låt q beteckna sannolikheten för en storskada (dvs att elden spritt sig). Vidare tänker vi oss att småskadorna har fördelningsfunktionen G och att storskadorna har fördelningsfunktionen H. Då har X fördelningsfunktionen

$$\begin{aligned} F(x) &= P(X \le x) \\ &= P\big(\text{småskada}\big)P\big(X \le x \bigm| \text{småskada}\big) + P\big(\text{storskada}\big)P\big(X \le x \bigm| \text{storskada}\big) \\ &= (1-q)\,G(x) + q\,H(x) \end{aligned}$$

För att det ska bli mer konkret tänker vi oss till exempel att småskadorna är gammafördelade och att storskadorna är Paretofördelade, dvs

$$\begin{aligned} g(x) &:= G'(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}\,x^{\alpha-1}\,e^{-\beta x}, \qquad x > 0 \\ h(x) &:= H'(x) = \frac{\gamma\,\tau^\gamma}{(\tau + x)^{\gamma+1}}, \qquad x > 0 \end{aligned}$$

Vi noterar att fördelningen för X har fem parametrar: q, α, β, τ och γ.

Antag nu att vi har observationerna $x_1, \ldots, x_n$, som antas vara oberoende utfall från fördelningen med fördelningsfunktionen F. Innan vi tittar på ML-ekvationerna inför vi storheten

p_k, sannolikheten att den k:te observationen härstammar från en storskada. Då gäller att

$$P\big(\text{storskada} \mid X = x\big) = \frac{h(x)\, q}{f(x)}$$

så att

$$p_k = \frac{h(x_k)\, q}{f(x_k)}$$

På samma sätt gäller att

$$1 - p_k = \frac{g(x_k)\,(1-q)}{f(x_k)}$$

Vi låter f beteckna tätheten för X, dvs

$$f(x) = (1-q)\, g(x) + q\, h(x)$$

Den logaritmerade likelihoodfunktionen är alltså

$$\begin{aligned} \ell(q,\alpha,\beta,\tau,\gamma) &= \sum_{k=1}^{n} \log f(x_k) \\ &= \sum_{k=1}^{n} \log\big\{(1-q)\, g(x_k) + q\, h(x_k)\big\} \end{aligned}$$

Med användning av vad vi fann ovan har vi att

$$\begin{aligned} \frac{\partial \ell}{\partial q} &= \sum_{k=1}^{n} \frac{-g(x_k) + h(x_k)}{f(x_k)} \\ &= \sum_{k=1}^{n} \left\{ -\frac{1-p_k}{1-q} + \frac{p_k}{q} \right\} \end{aligned}$$

ML-ekvationen för q blir alltså, efter förlängning med $q(1-q)$.

$$\sum_{k=1}^{n} \big\{ -q\,(1-p_k) + (1-q)\, p_k \big\} = 0$$

dvs

$$q = \frac{1}{n} \sum_{k=1}^{n} p_k$$

De partiella derivatorna av ℓ med avseende på α och β blir

$$\begin{cases} \frac{\partial \ell}{\partial \alpha} = \sum_{k=1}^{n} \frac{1-q}{f(x_k)} \frac{\partial}{\partial \alpha}\, g(x_k) \\ \frac{\partial \ell}{\partial \beta} = \sum_{k=1}^{n} \frac{1-q}{f(x_k)} \frac{\partial}{\partial \beta}\, g(x_k) \end{cases}$$

Vi kan skriva detta som

$$\begin{cases} \frac{\partial \ell}{\partial \alpha} = \sum_{k=1}^n (1-p_k)\, \frac{\partial}{\partial \alpha}\Big\{\log g(x_k)\Big\} \\ \frac{\partial \ell}{\partial \beta} = \sum_{k=1}^n (1-p_k)\, \frac{\partial}{\partial \beta}\Big\{\log g(x_k)\Big\} \end{cases}$$

ML-ekvationerna för α och β kan alltså skrivas som

$$\begin{cases} \sum_{k=1}^n (1-p_k)\, \frac{\partial}{\partial \alpha}\Big\{\log g(x_k)\Big\} = 0 \\ \sum_{k=1}^n (1-p_k)\, \frac{\partial}{\partial \beta}\Big\{\log g(x_k)\Big\} = 0 \end{cases} \tag{3.5}$$

På samma sätt blir ML-ekvationerna för τ och γ

$$\begin{cases} \sum_{k=1}^n p_k\, \frac{\partial}{\partial \tau}\Big\{\log h(x_k)\Big\} = 0 \\ \sum_{k=1}^n p_k\, \frac{\partial}{\partial \gamma}\Big\{\log h(x_k)\Big\} = 0 \end{cases} \tag{3.6}$$

Givet $p_1, \ldots, p_n$, är de två ekvationssystemen (3.5) och (3.6) inte svårare att lösa än vid vanlig ML-skattning i gamma respektive Pareto.

ML-skattningarna av de fem parametrarna kan beräknas via följande iterativa förfarande:

1. Beräkna startvärden $\hat{q}$, $\hat{\alpha}$, $\hat{\beta}$, $\hat{\tau}$ och $\hat{\gamma}$ (se nedan).
2. Beräkna
$$\hat{p}_k = \frac{\hat{g}(x_k)\,(1-\hat{q})}{\hat{f}(x_k)}$$
där $\hat{g}(x_k)$ och $\hat{f}(x_k)$ betecknar $g(x_k)$ respektive $f(x_k)$ med de senaste värdena på $\hat{\alpha}$, $\hat{\beta}$, $\hat{\tau}$ och $\hat{\gamma}$ instoppade.
3. Beräkna lösningen till (3.5) med p_k ersatt med $\hat{p}_k$ och de senaste värdena på $\hat{\alpha}$ och $\hat{\beta}$ som startvärden. Detta ger nya värden på $\hat{\alpha}$ och $\hat{\beta}$.
4. Beräkna lösningen till (3.6) med p_k ersatt med $\hat{p}_k$ och de senaste värdena på $\hat{\tau}$ och $\hat{\gamma}$ som startvärden. Detta ger nya värden på $\hat{\tau}$ och $\hat{\gamma}$.
5. Beräkna nytt värde på $\hat{q}$:
$$\hat{q} = \frac{1}{n}\sum_{k=1}^n \hat{p}_k$$

Upprepa stegen 2–5 tills $\hat{q}$, $\hat{\alpha}$, $\hat{\beta}$, $\hat{\tau}$ och $\hat{\gamma}$ konvergerat. Konvergensen kan vara långsam.

Här följer ett förslag på ett sätt att räkna ut startvärden för ovanstående iterationsschema.

Först gissar man ett lämpligt startvärde på $\hat{q}$ (andelen storskador) utifrån inspektion av datamaterialet.

Utifrån inspektion av datamaterialet väljer vi därefter a så litet att vi tror att alla (eller så gott som alla) observationer som ligger under a härstammar från småskador. Vi plockar ut alla observationer som ligger under a och kan då anta att dessa är oberoende observationer från fördelningen med täthetsfunktionen

$$\bar{g}(x) = \frac{g(x)}{G(a)}$$

Därefter beräknas ML-skattningarna av α och β utifrån dessa observationer (jämför avsnittet om självrisker ovan). Detta ger startvärdena på $\hat{\alpha}$ och $\hat{\beta}$.

Slutligen väljer vi b så stort att vi tror att alla (så gott som alla) observationer över b härstammar från storskador. Så plockar vi ut alla observationer som ligger över b (det gäller att inte välja b alltför stort, så att datamaterialet blir för litet) och antar att dessa kommer från fördelningen med täthetsfunktionen

$$\bar{h}(x) = \frac{h(x)}{1 - H(x)}$$

samt beräknar ML-skattningarna av τ och γ i denna situation. Detta ger startvärdena på $\hat{\tau}$ och $\hat{\gamma}$.

3.5 Övningscase: Excess of loss-återförsäkring

Beskrivning av problemet

För att minska riskerna för stora förluster kan ett försäkringsbolag återförsäkra en portfölj. Kostnaderna för stora skador delas då mellan försäkringsbolaget och ett eller flera återförsäkringsbolag.

En vanlig form är så kallad *excess of loss*-återförsäkring, ofta förkortat XL-återförsäkring. Om kostnaden för en skada överstiger en viss nivå r_1 betalar försäkringsbolaget, som i det här sammanhanget kallas *cedenten*, summan r_1 och ett återförsäkringsbolag betalar kostnaden som överstiger r_1, dock maximalt summan ℓ_1. Om kostnaden för skadan överstiger

$r_2 := r_1 + \ell_1$, betalar ett annat (eller samma) återförsäkringsbolag den del av kostnaden som överstiger r_2, dock maximalt summan ℓ_2. Vid behov träder ytterligare återförsäkrare in i bilden och skadan delas alltså upp i bitar av längden $r_1, \ell_1, \ell_2, \ell_3$, osv. Om skadans storlek är x och x ligger mellan $r_k := r_1 + \ell_1 + \cdots + \ell_{k-1}$ och $r_k + \ell_k$, betalar det sista återförsäkringsbolaget som behöver vara med och dela kostnaden summan $x - r_k$. Varje intervall $[r_j, r_j + \ell_j]$ brukar kallas för en *layer*, punkten r_j där återförsäkringsbolaget träder in kallas ofta layerns *excesspunkt* och maxbeloppet ℓ_j kallas layerns *limit* (det finns flera alternativa benämningar; p g a återförsäkringsbranschens internationella karaktär används ofta engelska termer även i Sverige). Man brukar säga att ett återförsäkringsbolag som betalar den del av kostnaden som ligger mellan r_j och $r_j + \ell_j$ täcker layern ℓ_j xs r_j, vilket utläses "ℓ_j in excess of r_j".

Som redan antytts kan ett återförsäkringsbolag täcka flera layrar. För varje layer avstår cedenten en viss andel av premieinkomsten för portföljen till återförsäkringsbolaget.

Återförsäkringsavtal innehåller ofta begränsningar vad gäller återförsäkrarens åtaganden, som till exempel att det kan finnas ett tak för hur mycket återförsäkringsbolaget totalt betalar ut för skador inträffade under ett år.

Återförsäkringsbolaget Skandia International har en portfölj med återförsäkringskontrakt rörande XL-återförsäkring av motorförsäkringar i ett flertal länder. Portföljen förändras ständigt i och med att nya avtal tecknas och gamla avslutas eller förändras. Varje kontrakt avser täckandet av en viss layer för en motorportfölj.

Prissättningen (ratingen) av ett återförsäkringskontrakt är ofta komplicerad, men ett viktigt underlag är givetvis återförsäkringsbolagets egen skadestatistik. Genom att anpassa en lämplig sannolikhetsfördelning till datamaterialet, får man ett instrument som är till hjälp vid fastställande av återförsäkringspremier.

Datamaterialet utgörs av Skandia Internationals egen skadestatistik för ett visst land. Alla skador som Skandia International varit med att betala och som inträffat 1984 eller senare finns med.

En komplikation med återförsäkringsdata, och inte minst när det gäller ansvarsmomentet inom motorförsäkring, är att det kan dröja flera år från det att skadan inträffar till det att den slutliga ersättningen fastställs. Alla skador i datamaterialet härrör sig inte från avslutade skador och i dessa fall har en uppskattning av det slutliga beloppet gjorts.

Varje post i datamaterialet är av följande typ:

1996	22000	15000

Först står året då skadan inträffade, därefter skadebeloppets storlek (uttryckt i en viss valuta) samt excesspunkten för det berörda kontraktet (uttryckt i samma valuta). Observera att skadebeloppet är den totala kostnaden för skadan (inte bara Skandia Internationals andel).

Ibland har en skada drabbat flera layrar som täckts av Skandia International och i de fallen är det den lägsta excesspunkten som står angiven.

Man vill anpassa en sannolikhetsfördelning för skadebeloppen, vilken sedan kan användas för att beräkna (till exempel) den förväntade kostnaden för ett visst kontrakt. Vilken fördelning ska man välja och hur går anpassningen till?

Förslag till ansats

Problemet är att observationerna inte är likafördelade (men kan antas oberoende). Vi antar att de ursprungliga skadorna är likafördelade med fördelningsfunktionen F och betecknar excesspunkten för den k :te skadan med c_k. Då gäller att den k:te observationen kommer från fördelningen med täthetsfunktionen

$$\bar{f}(x) = \frac{f(x)}{1 - F(c_k)} ,$$

där $f(x) = F'(x)$. Om observationerna betecknas med $x_1, \ldots, x_n$, blir loglikelihoodfunktionen alltså

$$\ell = \sum_{k=1}^{n} \log f(x_k) - \sum_{k=1}^{n} \log(1 - F(c_k)) .$$

Pareto och Burr leder till enkla räkningar och ML-skattningen är likartad med det vanliga fallet med likafördelade observationer.

Eftersom observationerna inte är likafördelade är den empiriska fördelningen inte väldefinierad. Man kan dock räkna ut den empiriska fördelningen för observationerna med den vanligaste excesspunkten och jämföra med denna. Om F är den anpassade fördelningsfunktionen efter ovanstående ML-skattning, och c är den vanligaste excesspunkten, skall denna empiriska fördelning jämföras med fördelningen med fördelningsfunktionen

$$\frac{F(x) - F(c)}{1 - F(c)} .$$

Hur ser väntevärdesfunktionen ut för denna fördelning då F till exempel är Pareto?

4 Modeller för antalet skador

Vi övergår nu till att studera några fördelningar som använts som modeller för antalet skador som drabbar en försäkringsportfölj. I kapitel 5 ska vi sätta samman fördelningen för antalet skador och fördelningen för skadebeloppen till en modell för den totala skadekostnaden för portföljen.

4.1 Poissonfördelningen

Poissonfördelningen är fundamental i detta sammanhang. I grundläggande kurser i sannolikhetsteori visas hur Poissonfördelningen givet vissa antaganden uppstår som en modell för antal händelser i ett tidsintervall. Något vagt uttryckta är antagandena de här:

(A) Antalet händelser i disjunkta intervall är oberoende;

(B) Bara en händelse i taget inträffar;

(C) Sannolikheten för att en händelse ska inträffa vid en viss given tidpunkt är noll.

Om dessa antaganden är realistiska, leds vi till att antalet händelser kan antas vara Poissonfördelat, dvs har sannolikhetsfunktionen

$$p_k = \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad k = 0, 1, 2, \ldots$$

där parametern λ antar ett värde i $(0, \infty)$.

Vi noterar att Poissonsannolikheterna uppfyller det rekursiva sambandet

$$p_k = \frac{\lambda}{k}\, p_{k-1}; \qquad k = 1, 2, \ldots$$

Utgående från startvärdet $p_0 = e^{-\lambda}$ kan man med hjälp av en dator snabbt beräkna sannolikheterna på ett numeriskt stabilt sätt.

Den momentgenererande funktionen för Poissonfördelningen är

$$\begin{aligned} M(t) &= \sum_{k=0}^{\infty} e^{tk}\, \frac{\lambda^k}{k!}\, e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= e^{-\lambda}\, e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

Den kumulantgenererande funktionen är

$$\Psi(t) = \lambda(e^t - 1)$$

så kumulanterna blir

$$\kappa_j = \Psi^{(j)}(0) = \lambda; \qquad j = 1, 2, \ldots$$

Den sannolikhetsgenererande funktionen är

$$Q(s) = M(\log s) = e^{\lambda(s-1)} \tag{4.1}$$

så faktorialmomenten blir

$$\gamma_j = Q^{(j)}(1) = \lambda^j; \qquad j = 1, 2, \ldots \tag{4.2}$$

Låt oss nu titta på parameterskattning i Poissonfördelningen. Poissonfördelningen har bara en parameter, så vid användandet av momentmetoden kan vi endast anpassa väntevärdet till den empiriska fördelningens. Eftersom väntevärdet i Poissonfördelningen är λ, ger detta direkt skattningen $\hat{\lambda} = \bar{x}$.

Så övergår vi till ML-metoden. Vi låter n_k beteckna antalet observationer som är lika med k, vi låter m vara det största värdet på k för vilket $n_k > 0$ samt betecknar antalet observationer med n.

Logaritmen av likelihoodfunktionen är

$$\begin{aligned} \ell(\lambda) &= \sum_{k=0}^{m} n_k \log p_k \\ &= \sum_{k=0}^{m} n_k (k \log \lambda - \log k! - \lambda) \\ &= n\bar{x} \log \lambda - \sum_{k=0}^{m} n_k \log k! - n\lambda \end{aligned}$$

Om man deriverar denna med avseende på λ och sätter derivatan till noll, får man

$$\frac{n\bar{x}}{\lambda} - n = 0 \qquad \text{dvs} \qquad \lambda = \bar{x}$$

Exempel 4.1 Det här exemplet är lånat från Daykin, Pentikäinen & Pesonen (1994) och visar antalet ersättningskrav från trafikförsäkringstagarna hos ett brittiskt försäkringsbolag ett visst år. Med beteckningen n_k för antalet försäkringstagare med k krav, hade man

$n_0 = 370412$	$n_1 = 46545$	$n_2 = 3935$
$n_3 = 317$	$n_4 = 28$	$n_5 = 3$

Tabell 4.1: *Antal ersättningskrav, trafikförsäkring.*

För $k \geq 6$ var $n_k = 0$. Vi låter n beteckna antalet observationer och m det största värde på k för vilket $n_k \geq 1$, dvs $m = 5$ och $n = n_0 + \cdots + n_m = 421240$.

Oavsett vilken av de två metoderna vi använder, får vi för datamaterialet i Tabell 4.1 skattningen

$$\hat{\lambda} = \bar{x} = 0.131737$$

Låt oss jämföra sannolikheterna $\hat{p}_k = n_k/n$ i den empiriska fördelningen med sannolikheterna $p_k(\hat{\lambda})$ i den anpassade Poissonfördelningen: Som synes är anpassningen inte alltför dålig,

k	$\hat{p}_k$	$p_k(\hat{\lambda})$
0	0.879	0.877
1	0.110	0.115
2	0.00934	0.00761
3	0.000753	0.000334
4	0.0000665	0.0000110
5	0.00000712	0.000000290

men kunde vara bättre. □

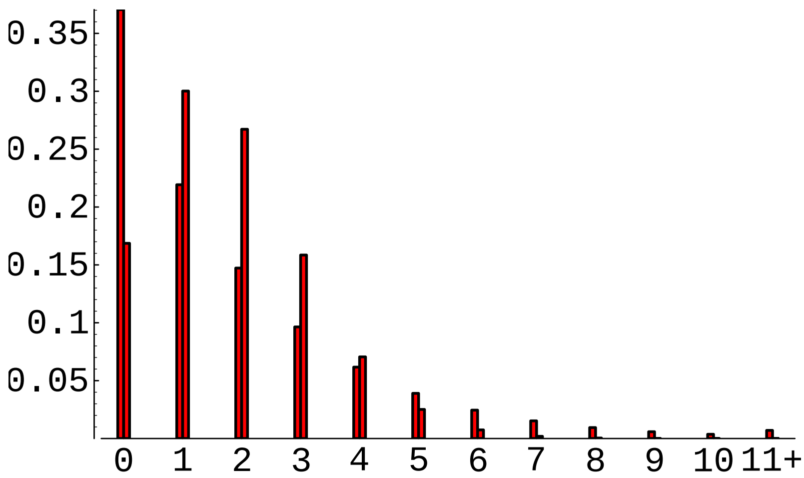
Exempel 4.2 Datamaterialet i det här exemplet kommer från Riksförsäkringsverket. Inom sjukförsäkring är det av intresse att studera antalet sjukfall under ett år. Tabell 4.2 visar sjukfallsstatistiken för år 1991: hur många förvärvsarbetande i Sverige som var sjuka 0 gånger, 1 gång, osv. Det sista värdet, 33078, visar hur många som hade minst 11 sjukfall. Det genomsnittliga antalet sjukfall var 1.78.

0	1	2	3	4	5	6	7
1653334	1031787	693000	453725	290287	183618	115379	71083

8	9	10	11+
44481	27611	17236	33078

Tabell 4.2: *Sjukfallsstatistik från Riksförsäkringsverket.*

Vi antar att observationerna för olika försäkringstagare är oberoende (fundera över relevansen i detta).



Figur 4.1: *Anpassning av Poissonfördelning (högra staplarna) till sjukfallsdata.*

I det här fallet har vi $\hat{\lambda} = \bar{x} = 1.78$. I Figur 4.1 jämförs sannolikheterna $\hat{p}_k$ i den empiriska fördelningen med sannolikheterna $p_k(\hat{\lambda})$. Det framgår att anpassningen är mycket dålig — den empiriska fördelningen är avsevärt mer långsvansad än Poissonfördelningen. I nästa avsnitt ska vi se att man kan få en mycket bättre anpassning med den så kallade negativa binomialfördelningen. □

Övning 4.1 *Anscombes approximation av fördelningsfunktionen F för Poissonfördelningen ser ut som*

$$F(k) \approx \Phi\left(\frac{3}{2}\left(k+\frac{5}{8}\right)^{2/3}\lambda^{-1/6} - \frac{3}{2}\sqrt{\lambda} + \frac{1}{24\sqrt{\lambda}}\right) \tag{4.3}$$

här betecknar $\Phi(\cdot)$ fördelningsfunktionen för en $N(0,1)$-variabel. Studera hur noggrannheten hos approximationen beror av λ.

Övning 4.2 *Ett berömt historiskt exempel utgörs av följande data över antalet dödsfall orsakade av hästsparkar i tio olika regementen i den preussiska armen under vart och ett av åren 1875–1894:*

k	0	1	2	3	4	5+
n_k	109	65	22	3	1	0

Anpassa en Poissonfördelning till detta datamaterial och jämför den anpassade sannolikhetsfunktionen med den empiriska fördelningens.

4.2 Blandade Poissonfördelningar

En vanlig situation är att det råder osäkerhet rörande parametern λ i Poissonfördelningen. Antag t ex att man är intresserad av fördelningen för antalet skogsbränder i ett visst område under kommande år. Antalet skogsbränder beror till stor del på väderförhållandena, men givet dessa verkar en Poissonfördelning rimlig. Problemet är att man inte kan förutse vädret det kommande året. Detta kan attackeras genom att betrakta parametern i Poissonfördelningen som en stokastisk variabel Λ. Modellen är att givet att Λ antar värdet λ, är antalet skogsbränder Poissonfördelat med parameter λ. För att specificera modellen fullständigt, måste en fördelning för Λ ansättas, den så kallade *blandningsfördelningen*. Under dessa antaganden har antalet skogsbränder en *blandad Poissonfördelning*.

Låt oss anta att Λ har en täthetsfunktion $u(\cdot)$, sådan att $u(\lambda) > 0$ för alla $\lambda > 0$. Sannolikheten för n skador blir då

$$\begin{aligned} p_n &:= P(N=n) = \int_0^\infty P(N=n|\Lambda=\lambda)\, u(\lambda)\, d\lambda \\ &= \int_0^\infty \frac{\lambda^n}{n!}\, e^{-\lambda}\, u(\lambda)\, d\lambda \end{aligned}$$

Om M är den momentgenererande funktionen för blandningsfördelningen,

$$M(t) = \int_0^\infty e^{t\lambda}\, u(\lambda)\, d\lambda$$

får man från (4.1) att den sannolikhetsgenererande funktionen för N kan skrivas som

$$\begin{aligned} P(s) &= E[s^N] = \int_0^\infty E\big[s^N \bigm| \Lambda=\lambda\big] u(\lambda)\, d\lambda \\ &= \int_0^\infty e^{\lambda(s-1)}\, u(\lambda)\, d\lambda = M(s-1) \end{aligned} \tag{4.4}$$

Via (4.4) får man följande enkla förhållande mellan faktorialmomenten för N och momenten kring origo för Λ:

$$\begin{aligned} &E\big[N(N-1)\cdots(N-k+1)\big] \\ &= \int_0^\infty E\big[N(N-1)\cdots(N-k+1) \bigm| \Lambda=\lambda\big] u(\lambda)\, d\lambda \\ &= \int_0^\infty \lambda^k\, u(\lambda)\, d\lambda = E\big[\Lambda^k\big] \end{aligned} \tag{4.5}$$

Negativa binomialfördelningen

Den i särklass mest kända modellen får man då man antar att Λ har en gammafördelning:

$$u(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, \lambda^{\alpha-1}\, e^{-\beta\lambda}, \qquad \lambda \geq 0$$

Vi kan uttrycka sannolikheterna $p_n = P(N = n)$ med hjälp av gammafunktionen:

$$\begin{aligned}
p_n &= \int_0^\infty P(N = n|\Lambda = \lambda)\, u(\lambda)\, d\lambda \\
&= \int_0^\infty \frac{\lambda^n}{n!}\, e^{-\lambda}\, \frac{\beta^\alpha}{\Gamma(\alpha)}\, \lambda^{\alpha-1}\, e^{-\beta\lambda}\, d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)\, n!} \int_0^\infty \lambda^{n+\alpha-1}\, e^{-(\beta+1)\lambda}\, d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)\, n!} \int_0^\infty \left(\frac{x}{\beta+1}\right)^{n+\alpha-1} e^{-x}\, \frac{1}{\beta+1}\, dx \\
&= \frac{1}{\Gamma(\alpha)\, n!}\, \frac{\beta^\alpha}{(\beta+1)^{n+\alpha}} \int_0^\infty x^{n+\alpha-1}\, e^{-x}\, dx \\
&= \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\, n!} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^n
\end{aligned} \tag{4.6}$$

Detta är en så kallad negativ binomialfördelning. Ibland är det behändigare att använda andra parametriseringar, t ex $r := 1/(\beta+1)$ i stället för β:

$$p_n = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\, n!}\, (1-r)^\alpha\, r^n$$

Med hjälp av (4.5) och momenten för gammafördelningen får man lätt att de tre första centralmomenten för den negativa binomialfördelningen är α/β, $\alpha(\beta+1)/\beta^2$ och $\alpha(2+3\beta+\beta^2)/\beta^3$.

Den momentgenererande funktionen för gammafördelningen är ju

$$M(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha$$

så enligt (4.4) har N den sannolikhetsgenererande funktionen

$$P(s) = \left(\frac{\beta}{\beta-(s-1)}\right)^\alpha = \big(1+\beta^{-1}(1-s)\big)^{-\alpha}$$

Antag att man har m oberoende negativt binomialfördelade variabler med parametrarna $(\alpha_1, \beta), \ldots, (\alpha_m, \beta)$. Då har summan av dem den sannolikhetsgenererande funktionen

$$\prod_{i=1}^m \big(1+\beta^{-1}(1-s)\big)^{-\alpha_i} = \big(1+\beta^{-1}(1-s)\big)^{-(\alpha_1+\cdots+\alpha_m)}$$

dvs är negativt binomialfördelad med parametrar $(\alpha_1 + \cdots + \alpha_m, \beta)$.

Man verifierar lätt att sannolikheterna $\{p_n\}$ i den negativa binomialfördelningen uppfyller rekursionsformeln

$$p_n = \frac{\alpha + n - 1}{(\beta + 1)n}\, p_{n-1}, \qquad n \geq 1 \tag{4.7}$$

Vi noterar här en gemensam egenskap hos binomial-, Poisson- och negativ binomialfördelning: Sannolikheterna $\{p_n\}$ uppfyller alla ett rekursivt samband av typen

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} \tag{4.8}$$

Detta kommer att visa sig vara av intresse framöver.

Av flera skäl är den negativa binomialfördelningen den blandade Poissonfördelning som använts mest inom försäkringsmatematiken, men på senare år har man även börjat diskutera andra fördelningar (se till exempel Willmot, 1993).

Parameterskattning i negativ binomial

Vi ska nu studera parameterskattning i den negativa binomialfördelningen. Vår modell är alltså att sannolikheten att en observation ska anta värdet k är

$$p_k(\alpha, \beta) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\, k!} \left(\frac{\beta}{\beta + 1}\right)^{\alpha} \left(\frac{1}{\beta + 1}\right)^{k}$$

Den negativa binomialfördelningen har två parametrar, vilket ger oss möjligheten att få de två första momenten att överensstämma med den empiriska fördelningens. Momentmetoden ger

$$\begin{cases} \alpha/\beta = \bar{x} \\ \alpha(\beta + 1)/\beta^2 = s^2 \end{cases} \qquad \text{dvs} \qquad \begin{cases} \widehat{\alpha} = \bar{x}^2/(s^2 - \bar{x}) \\ \widehat{\beta} = \bar{x}/(s^2 - \bar{x}) \end{cases}$$

För trafikskadematerialet i Tabell 4.1 har vi att

$$\bar{x} = 0.131737 \qquad s^2 = 0.138521$$

vilket ger

$$\widehat{\alpha} = 2.558 \qquad \widehat{\beta} = 19.420$$

En alternativ metod när man som här har många värden som är noll, är att välja α och β så att väntevärdet och $p_0(\alpha, \beta)$ överensstämmer med motsvarande värden för den empiriska fördelningen, dvs man löser i stället ekvationen

$$\begin{cases} \alpha/\beta = \bar{x} \\ (\beta/(\beta+1))^\alpha = n_0/n \end{cases}$$

Om vi löser ut α i den andra ekvationen får vi att

$$\alpha = \frac{\log n - \log n_0}{\log(\beta+1) - \log \beta}$$

och stoppar vi in detta i den första får vi

$$\beta(\log(\beta+1) - \log \beta) = \frac{\log n - \log n_0}{\bar{x}}$$

Denna ekvation löses numeriskt och man får då skattningen $\widehat{\beta} = 20.244$, samt slutligen $\widehat{\alpha} = \widehat{\beta}\bar{x} = 2.6669$.

Vi övergår nu till ML-metoden. Likelihoodfunktionen är

$$L(\alpha, \beta) = \prod_{j=1}^{n} p_{x_j}(\alpha, \beta) = \prod_{k=0}^{m} \big(p_k(\alpha, \beta)\big)^{n_k}$$

Vi ska nu härleda ML-ekvationerna

$$\begin{cases} \frac{\partial}{\partial \alpha}\, \ell(\alpha, \beta) = 0 \\ \frac{\partial}{\partial \beta}\, \ell(\alpha, \beta) = 0 \end{cases}$$

där $\ell(\cdot, \cdot)$ är den logaritmerade likelihoodfunktionen.

Genom att upprepade gånger använda att $\Gamma(x+1) = x\,\Gamma(x)$, fås att

$$\begin{aligned} p_k(\alpha, \beta) &= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\,k!}\left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^k \\ &= \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{k!}\left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^k \end{aligned}$$

dvs

$$\begin{aligned}
\ell(\alpha,\beta) &= \sum_{k=0}^{m} n_k \log p_k(\alpha,\beta) = n_0\,\alpha \log\frac{\beta}{\beta+1} \\
&+ \sum_{k=1}^{m} n_k \left[\sum_{i=0}^{k-1} \log(\alpha+i) - \sum_{i=1}^{k} \log i + \alpha\log\frac{\beta}{\beta+1} - k\log(\beta+1)\right] \\
&= \sum_{k=0}^{m} n_k \big[\alpha\log\beta - (\alpha+k)\log(\beta+1)\big] \\
&+ \sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \log(\alpha+i) - \sum_{k=1}^{m} n_k \sum_{i=1}^{k} \log i
\end{aligned}$$

De partiella derivatorna blir

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\,\ell(\alpha,\beta) &= \sum_{k=0}^{m} n_k \log\frac{\beta}{\beta+1} + \sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \frac{1}{\alpha+i} \\
&= -\,n\log(1+\beta^{-1}) + \sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \frac{1}{\alpha+i}
\end{aligned}$$

respektive

$$\begin{aligned}
\frac{\partial}{\partial\beta}\,\ell(\alpha,\beta) &= \sum_{k=0}^{m} n_k \left[\frac{\alpha}{\beta} - \frac{\alpha+k}{\beta+1}\right] \\
&= \left(\frac{\alpha}{\beta} - \frac{\alpha}{\beta+1}\right)\sum_{k=0}^{m} n_k - \frac{1}{\beta+1}\sum_{k=0}^{m} k\,n_k \\
&= \frac{\alpha}{\beta(\beta+1)}\,n - \frac{n\bar{x}}{\beta+1}
\end{aligned}$$

Man får ML-ekvationerna

$$\begin{cases} n\log(1+\beta^{-1}) = \sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \frac{1}{\alpha+i} \\ \alpha = \beta\bar{x} \end{cases}$$

Om man tar uttrycket för β i den andra ekvationen och stoppar in i den första erhålls följande ekvation för α:

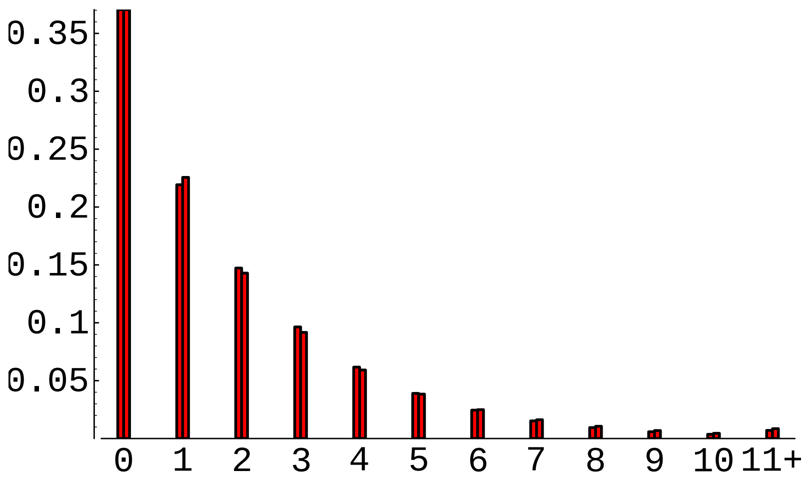
$$n\log\left(1+\frac{\bar{x}}{\alpha}\right) = \sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \frac{1}{\alpha+i}$$

Denna ekvation löses numeriskt. I allmänhet fungerar momentmetodens skattning av α (som ju är lätt att räkna ut) bra som utgångspunkt. För trafikskadeexemplet fås ML-skattningarna $\widehat{\alpha} = 2.6047$, samt $\widehat{\beta} = \widehat{\alpha}/\bar{x} = 19.772$. I följande tabell jämförs den empiriska fördelningen med den ML-anpassade negativa binomialfördelningen:

k	$\hat{p}_k$	$p_k(\hat{\alpha}, \hat{\beta})$
0	0.879	0.879
1	0.110	0.110
2	0.00934	0.00956
3	0.000753	0.000707
4	0.0000665	0.0000477
5	0.00000712	0.00000303

Anpassningen blir lite bättre än i Poissonfallet. Vi ska nu se att man uppnår en avsevärt bättre anpassning för sjukfallsstatistiken i Tabell 4.2 genom att gå från Poisson till negativ binomial.

Ett problem med detta datamaterial är att man slagit ihop alla som hade mer än tio sjukfall. Vi kan inte använda momentmetoden eller ML-metoden rakt upp och ner. Eftersom vi dock känner medelvärdet, kan vi använda metoden med anpassning till väntevärdet och sannolikheten för noll i den empiriska fördelningen.



Figur 4.2: *Anpassning av negativ binomialfördelning (högra staplarna) till sjukfallsdata.*

Med denna metod finner man skattningarna

$$\widehat{\alpha} = 1.01048, \qquad \widehat{\beta} = 0.56769$$

Vi ser i Figur 4.2 att anpassningen är mycket god och är radikalt bättre än vad vi uppnådde med Poissonfördelningen.

ML-metoden går i princip att använda: sannolikheten för det observerade utfallet är

$$\prod_{k=0}^{10} p_k(\alpha, \beta)^{n_k} \left(1 - \sum_{j=0}^{10} p_j(\alpha, \beta)\right)^{n_{11+}}$$

Man kan ge sig på att söka α och β som maximerar (logaritmen av) denna funktion, men det är ganska krångligt.

Övning 4.3 *Antag att man har en blandad Poissonfördelning $\{p_n\}$, där blandningsfördelningen har en täthet u, sådan att $u(\lambda) > 0$ för $0 < \lambda < \infty$. Antag också att*

$$\lim_{\lambda \to 0} \lambda\, u(\lambda) = 0$$

Man kan visa att om u uppfyller

$$\frac{d}{d\lambda} \log u(\lambda) = \frac{c_0 + c_1\,\lambda + c_2\,\lambda^2}{d_1\lambda + d_2\,\lambda^2}$$

där c_0, c_1, c_2, d_1 och d_2 är konstanter, så gäller för $n \geq 0$ att

$$\begin{aligned} &(d_2 - c_2)n(n-1)\,p_n \\ &= \quad (2d_2 + c_1 - d_1 + d_2\,(n-2))(n-1)\,p_{n-1} + (c_0 + d_1(n-1))\,p_{n-2} \end{aligned}$$

Detta är ett specialfall av ett resultat i Willmot (1993).

a) Verifiera att resultatet stämmer för den negativa binomialfördelningen.

b) Den blandade Poissonfördelning som uppstår då man som blandningsfördelning tar en invers normalfördelning,

$$u(\lambda) = \frac{\mu}{\sqrt{2\pi\beta\lambda^3}} \exp\left\{-\frac{(\lambda-\mu)^2}{2\beta\lambda}\right\}, \qquad \lambda > 0$$

*kallas för IGP-fördelningen (*Inverse Gaussian Poisson*).*

Använd resultatet ovan för att härleda en rekursionsformel för sannolikheterna $\{p_n\}$ i IGP-fördelningen. Härled startvärdena p_0 och p_1 med hjälp av den sannolikhetsgenererande funktionen.

Övning 4.4 *Den här uppgiften är också ett specialfall av Willmots (1993) resultat. Låt N ha en blandad Poissonfördelning, där blandningsfördelningen är Pareto:*

$$u(\lambda) = \frac{\gamma\,\alpha^\gamma}{(\alpha+\lambda)^{\gamma+1}}, \qquad \lambda \geq 0$$

Härled följande rekursionsformel för sannolikheterna $p_n := P(N = n)$:

$$p_n = \left(1 - \frac{\gamma + 1 + \alpha}{m}\right) p_{n-1} + \frac{\alpha}{n}\,p_{n-2}, \qquad n \geq 2$$

Ledning: Härled först likheten

$$p_n = p_{n-1} + \int_0^\infty \frac{\lambda^n}{n!}\, e^{-\lambda}\, u'(\lambda)\, d\lambda$$

via partiell integration. Använd sedan att

$$\frac{u'(\lambda)}{u(\lambda)} = \frac{d}{d\lambda}\big\{\log u(\lambda)\big\}$$

har en enkel form.

Övning 4.5 *Följande data är hämtade från Bühlmann (1970) och visar antalet ersättningskrav inom schweizisk trafikförsäkring under en viss period:*

k	0	1	2	3	4	5	6	7+
n_k	103 704	14 075	1 766	255	45	6	2	0

Anpassa en negativ binomialfördelning till detta datamaterial och jämför den anpassade sannolikhetsfunktionen med den empiriska fördelningens. Anpassa även en Poissonfördelning och jämför.

Övning 4.6 *Anpassa en negativ binomialfördelning till datamaterialet i Övning 4.2. Jämför med anpassningen av Poissonfördelningen.*

Övning 4.7 *Använd metoden med matchning av väntevärdet och sannolikheten för noll i den empiriska fördelningen för att anpassa en IGP-fördelning (Övning 4.3) till Riksförsäkringsverkets data i Tabell 4.3. Vilken fördelning passar bäst, negativ binomial eller IGP?*

Ledning: *Man kan visa att väntevärdet i den inversa normalfördelningen är μ.*

4.3 Trunkerade fördelningar

En olycka registreras ej säkert som en sådan (hos försäkringbolaget), såvida ej antalet skador den orsakar är minst en. Ett exempel är antalet skadade personer i bilolyckor: dataunderlaget kan mycket väl bestå av olyckor med personskador (olyckor där ingen skadats finns inte med). Ett annat exempel är stormskador på hus: om man inte får in minst en skadeanmälan, noteras ingen storm.

I sådana sammanhang har man inget underlag för att skatta $p_0 = P(N = 0)$ (och det behövs inte heller), utan endast sannolikheterna

$$q_n := P\big(N = n \mid N \geq 1\big); \qquad n = 1, 2, \ldots$$

Man får att

$$\begin{aligned} q_n &= \frac{P(N = n,\, N \geq 1)}{P(N \geq 1)} \\ &= \frac{P(N = n)}{1 - P(N = 0)} = \frac{p_n}{1 - p_0} \end{aligned}$$

Observera att $\{q_n\}$ är en sannolikhetsfördelning:

$$\sum_{n=1}^{\infty} q_n = \frac{1}{1 - p_0} \sum_{n=1}^{\infty} p_n = \frac{1}{1 - p_0}\,(1 - p_0) = 1$$

Vi noterar följande samband mellan de sannolikhetsgenererande funktionerna $P(\cdot)$ och $Q(\cdot)$ för $\{p_n,\, n \geq 0\}$ och $\{q_n,\, n \geq 1\}$:

$$\begin{aligned} P(s) &= \sum_{n=0}^{\infty} s^n\, p_n \\ &= p_0 + (1 - p_0) \sum_{n=1}^{\infty} s^n\, q_n \\ &= p_0 + (1 - p_0)\, Q(s) \end{aligned}$$

Vi noterar att om sannolikheterna $\{p_n\}$ uppfyller en rekursionsformel av typen (4.8), gäller detta också sannolikheterna $\{q_n\}$: Genom att dela med $1 - p_0$ på båda sidor i (4.8), får vi att

$$q_n = \left(a + \frac{b}{n}\right) q_{n-1}, \qquad n \geq 2$$

Trunkerad negativ binomialfördelning

Låt N vara negativt binomialfördelad, dvs

$$p_n = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)\, n!} \left(\frac{\beta}{\beta + 1}\right)^{\alpha} \left(\frac{1}{\beta + 1}\right)^{n}, \qquad n \geq 0$$

För att få behändigare uttryck, inför vi parametern $r := 1/(\beta + 1)$ i stället för β:

$$p_n = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)\, n!}\, (1 - r)^{\alpha}\, r^n, \qquad n \geq 0$$

Eftersom $0 < \beta < \infty$, har vi att $0 < r < 1$. Vi får att

$$\begin{aligned} q_n = \frac{p_n}{1-p_0} &= \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\, n!\, (1-(1-r)^\alpha)}\, (1-r)^\alpha\, r^n \\ &= \frac{\Gamma(\alpha+n)}{\Gamma(\alpha+1)\, n!}\, \frac{\alpha}{1/(1-r)^\alpha - 1}\, r^n \end{aligned} \tag{4.9}$$

Nu drar vi oss till minnes att $dx^\alpha/d\alpha = x^\alpha \log x$, så l'Hospitals regel ger att

$$\frac{\alpha}{1/(1-r)^\alpha - 1} \to \frac{1}{\log(1/(1-r))} = \frac{1}{-\log(1-r)}$$

då $\alpha \to 0$. Om vi håller $r \in (0,1)$ fixt och låter $\alpha \to 0$, får vi alltså att

$$q_n \to \frac{(n-1)!}{n!}\, \frac{1}{-\log(1-r)}\, r^n = \frac{r^n}{-n \log(1-r)}$$

Detta är en sannolikhetsfördelning, dvs $q_1 + q_2 + \cdots = 1$, eftersom

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = -\log(1-x), \qquad -1 \le x < 1$$

Denna fördelning kallas den *logaritmiska fördelningen*.

Man kan visa att $\{q_n\}$ i (4.9) definierar en sannolikhetsfördelning även då $-1 < \alpha < 0$. Vi kallar fördelningen $\{q_n\}$ i (4.9), med det tillåtna parameterområdet $-1 < \alpha < \infty$, $0 < r < 1$ för den *utvidgade* trunkerade negativa binomialfördelningen. Att α tillåts anta negativa värden är intressant: det kan mycket väl vara ett sådant som ger den bästa anpassningen till data.

Låt oss beräkna de två första faktorialmomenten för trunkerad negativ binomialfördelning. Vi kommer senare att behöva dem när vi ska anpassa modellen till data med hjälp av momentmetoden.

De två första faktorialmomenten för en negativ binomialfördelning är α/β respektive $\alpha(\alpha+1)/\beta^2$. Av detta följer att de två första faktorialmomenten för den trunkerade negativa binomialfördelningen är

$$\gamma_1 = \sum_{n=1}^{\infty} n\, q_n = \frac{1}{1-p_0} \sum_{n=1}^{\infty} n\, p_n = \frac{\alpha}{\beta(1-p_0)}$$

respektive

$$\gamma_2 = \sum_{n=1}^{\infty} n(n-1) q_n = \frac{1}{1-p_0} \sum_{n=1}^{\infty} n(n-1)\, p_n = \frac{\alpha(\alpha+1)}{\beta^2(1-p_0)}$$

Med användande av parametern $r = 1/(\beta + 1)$ blir detta

$$\gamma_1 = \frac{\alpha r}{(1-r)(1-(1-r)^\alpha)}, \qquad \gamma_2 = \frac{\alpha(\alpha+1)r^2}{(1-r)^2(1-(1-r)^\alpha)}. \tag{4.10}$$

Man kan visa att detta även gäller då $-1 < \alpha < 0$. Låter man $\alpha \to 0$, får man de två första faktorialmomenten för den logaritmiska fördelningen.

Parameterskattning i trunkerad negativ binomial

Sannolikhetsfunktionen för den trunkerade negativa binomialfördelningen ges som vi såg ovan av

$$\begin{aligned} q_k &= q_k(\alpha, r) = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha+1)\,k!}\,\frac{\alpha}{(1-r)^{-\alpha}-1}\,r^k \\ &= \prod_{i=0}^{k-1}(\alpha+i)\,\frac{r^k}{k!\,((1-r)^{-\alpha}-1)}, \qquad k \geq 1 \end{aligned}$$

Här använde vi att $\Gamma(x+1) = x\,\Gamma(x)$. Parameterområdet är $-1 < \alpha < \infty$, $0 < r < 1$. För $\alpha = 0$ ska alla uttryck betraktas som gränsvärden då $\alpha \to 0$.

Vi börjar med momentmetoden och söker de parametervärden för vilka de två första faktorialmomenten överensstämmer med motsvarande moment för den empiriska fördelningen:

$$\begin{aligned} \hat{\gamma}_1 &:= \frac{1}{n}\sum_{k=1}^{m} k\,n_k \\ \hat{\gamma}_2 &:= \frac{1}{n}\sum_{k=2}^{m} k\,(k-1)\,n_k \end{aligned}$$

För enkelhets skull bortser vi nu från fallet $\alpha = 0$. Enligt vad vi fann tidigare, ska vi lösa ekvationerna

$$\hat{\gamma}_1 = \tfrac{\alpha\,r}{(1-r)\,(1-(1-r)^\alpha)}, \tag{4.11}$$

$$\hat{\gamma}_2 = \tfrac{\alpha\,(\alpha+1)\,r^2}{(1-r)^2\,(1-(1-r)^\alpha)} \tag{4.12}$$

Från (4.11) och (4.12) får vi

$$\hat{\gamma}_1\,\frac{(\alpha+1)\,r}{1-r} = \hat{\gamma}_2$$

vilket ger

$$r = \frac{\hat{\gamma}_2}{\hat{\gamma}_1\,(\alpha+1)+\hat{\gamma}_2} \qquad 1-r = \frac{\hat{\gamma}_1\,(\alpha+1)}{\hat{\gamma}_1\,(\alpha+1)+\hat{\gamma}_2} \tag{4.13}$$

Stoppar vi in detta i (4.11) och flyttar om litet, får vi följande ekvation för α:

$$\hat{\gamma}_2\,\alpha = \hat{\gamma}_1^2\,(\alpha+1)\left(1-\left(\frac{\hat{\gamma}_1\,(\alpha+1)}{\hat{\gamma}_1\,(\alpha+1)+\hat{\gamma}_2}\right)^{\alpha}\right)$$

Genom att lösa denna ekvation numeriskt, får man fram momentmetodens skattning av α och stoppar man in denna i (4.13) får man skattningen av r.

Härledningen av ML-ekvationerna kräver lite räkningar, vilka lämnas som övning. Ekvationerna blir

$$\frac{1}{n}\sum_{k=1}^{m} n_k \sum_{i=0}^{k-1}\frac{1}{\alpha+i}+\frac{\bar{x}\,(1-r)\log(1-r)}{\alpha\,r}=0 \tag{4.14}$$

$$\bar{x}\,(1-r)(1-(1-r)^{\alpha})-r\alpha=0 \tag{4.15}$$

Tyvärr går det inte att uttrycka den ena variabeln som en enkel funktion av den andra och komma ner till en ekvation.

Ett enkelt sätt att finna ML-skattningarna är det här. För varje fixt värde på α, låt $r_0(\alpha)$ beteckna det värde på r som löser ekvationen (4.15), vilket enkelt kan erhållas numeriskt. Om vi sätter

$$f(\alpha,r):=\frac{1}{n}\sum_{k=1}^{m} n_k \sum_{i=0}^{k-1}\frac{1}{\alpha+i}+\frac{\bar{x}\,(1-r)\log(1-r)}{\alpha\,r}$$

och $g(\alpha):=f(\alpha,r_0(\alpha))$, ska vi alltså lösa ekvationen $g(\alpha)=0$, vilket är en ekvation med en obekant. Vi kan inte skriva upp ett explicit enkelt uttryck för $g(\alpha)$, men vi kan beräkna $g(\alpha)$ för godtyckliga värden på α. Ekvationen $g(\alpha)=0$ kan alltså lösas med någon enkel numerisk metod, som sekantmetoden, eller genom att bara pröva sig fram. Lösningen $\hat{\alpha}$ är ML-skattningen av α och $r_0(\hat{\alpha})$ är ML-skattningen av r. Momentmetodens skattningar är ofta bra att använda som utgångspunkt vid sökandet efter ML-skattningarna.

Följande tabell visar antalet anmälda skador på lantbruksegendomar i Kristianstads län orsakade av stormar under åren 1983–1993.

De två största stormarna, som inträffade 1983 och 1993, orsakade sammanlagt 1572 skadeanmälningar, mer än alla de andra 33 stormarna tillsammans. Här rör det sig alltså om en långsvansad fördelning.

I datamaterialet finns inga nollor: orsakas inga skador, noteras ingen storm. Här har vi en typisk situation där en trunkerad fördelning dyker upp.

1983 :	679	27	1	3	14
1984 :	127				
1985 :	91	11	15	11	
1986 :	1	2	8	106	
1987 :	1				
1988 :	2	1	1	12	
1989 :	6	70	11	1	
1990 :	6	320	139		
1991 :	14	12			
1992 :	4	1	30		
1993 :	893	4	25	34	

Tabell 4.3: Antal stormskador.

För datamaterialet i Tabell 4.3 blir momentmetodens skattningar

$$\hat{\alpha} = -0.0412023, \qquad \hat{r} = 0.998195$$

Det är intressant att notera att skattningen av α är negativ — det var alltså meningsfullt att utvidga parameterområdet från $0 < \alpha < \infty$ till $-1 < \alpha < \infty$.

ML-skattningarna blir

$$\widehat{\alpha} = -0.0675574, \qquad \widehat{r} = 0.99838152$$

I Tabell 4.3 jämförs den empiriska fördelningsfunktionen ($\hat{F}$) med den anpassade fördelningsfunktionen (F), för några utvalda värden.

k	$\hat{F}(k)$	$F(k)$
1	0.2	0.192
2	0.257	0.281
5	0.343	0.413
10	0.429	0.516
20	0.657	0.616
50	0.771	0.740
100	0.829	0.825
200	0.914	0.897
500	0.943	0.964
1000	1	0.991

Tabell 4.4: Fördelningsfunktioner för stormskador.

Vid beräkningen av den anpassade fördelningen använder man givetvis det rekursiva sambandet

$$q_{k+1}(\alpha, r) = \frac{(\alpha + k)\, r}{k+1}\, q_k(\alpha, r), \qquad k \geq 1$$

Övning 4.8 *Räkna ut de två första faktorialmomenten för den logaritmiska fördelningen, dels utifrån sannolikheterna själva,*

$$q_n = \frac{r^n}{-n\log(1-r)}, \qquad n \geq 1$$

dels genom att låta $\alpha \to 0$ i (4.10).

Övning 4.9 *Den trunkerade Poissonfördelningen $\{q_k,\ k \geq 1\}$ med parameter λ ges av*

$$q_k := P\big(N = k \bigm| N \geq 1\big)$$

där N är Poissonfördelad med parameter λ. Beräkna ML-skattningen av λ i den trunkerade Poissonfördelningen då medelvärdet av observationerna är $\bar{x} = 3.1$.

Övning 4.10 *Härled (4.14) och (4.15).*

4.4 Övningscase: Flerpersonskador i bussolyckor

Beskrivning

Trafikförsäkringen för bussar garanterar bland annat ersättning för personskador och dödsfall i händelse av olycka. Om en buss kör av vägen finns det risk för att ett stort antal personer blir skadade eller dödade. För att kunna sätta lämpliga premier behöver man kännedom om fördelningen för antalet personskador/dödsfall vid bussolyckor.

En fördelning som i den försäkringsmatematiska litteraturen föreslagits som modell för antalet dödade och skadade i trafikolyckor är den s k generaliserade Poisson-fördelningen. Denna har två parametrar, θ och λ, där $\theta > 0$ och $\max(-1, -\theta/4) \leq \lambda < 1$. För $\lambda \geq 0$ ges sannolikheten för utfallet n av

$$p_n := \theta(\theta + n\lambda)^{n-1}\,\frac{e^{-\theta-n\lambda}}{n!} \tag{4.16}$$

$n = 0, 1, 2, \ldots$ För $\lambda < 0$ ges sannolikheten p_n av (4.16) för $n = 0, 1, \ldots, m$ och $p_n = 0$ för $n > m$, där m är det största positiva heltal för vilket $\theta + \lambda m > 0$. Fördelningens väntevärde är $\theta/(1-\lambda)$ och variansen $\theta/(1-\lambda)^3$. ML-skattningen $\hat{\lambda}$ av λ utgörs av lösningen till ekvationen

$$\sum_k n_k\,\frac{k(k-1)}{\bar{x} + (k-\bar{x})\lambda} - n\bar{x} = 0$$

där n är antalet olyckor, n_k är antalet olyckor med k skadade/dödade och $\bar{x}$ är det genomsnittliga antalet skadade/dödade per olycka. ML-skattningen $\hat{\theta}$ av θ ges av

$$\hat{\theta} = \bar{x}(1 - \hat{\lambda})$$

På försäkringsbolaget Skandia gjorde man i början på 90-talet en studie av modeller som kunde ligga till grund för tariffbestämningen. Man ville undersöka om den generaliserade Poissonfördelningen var lämplig att använda eller om det fanns andra som passade bättre.

Utgående från uppgifter som lämnats till Försäkringsförbundet sammanställde man en tabell (se nästa sida) över antalet skadade/dödade i trafikolyckor under åren 1987–1993, där en buss varit inblandad. Personer som dödats/skadats, men ej befunnit sig i bussen (fotgängare eller passagerare i andra fordon), har ej tagits med.

Förslag till ansats

Man anar direkt att data är alltför långsvansade för att Poissonfördelningen ska passa bra (men pröva själv). Man bör pröva hur väl generaliserad Poisson passar jämfört med negativ binomial.

Det är tänkbart att det är svårt att få dessa fördelningar att passa bra både i "svansen" och i nollan. Om så är fallet kan man tänka sig följande variant: Låt N vara antalet skador i en godtycklig bussolycka och sätt

$$p_n := P(N = n), \qquad q_n := P\big(N = n \mid N \geq 1\big)$$

Då gäller för $n \geq 1$ att

$$\begin{aligned} P(N = n) &= P\big(N = n \mid N = 0\big)P(N = 0) + P\big(N = n \mid N \geq 1\big)P(N \geq 1) \\ &= q_n(1 - p_0) \end{aligned}$$

Sannolikheten p_0 kan skattas som andelen nollor i datamaterialet. Sannolikheterna $\{q_n\}$ kan skattas genom att anpassa (till exempel) en trunkerad negativ binomialfördelning till det datamaterial man får om man tar bort alla olyckor utan personskador.

Antal dödade/ skadade	Antal olyckor
0	8336
1	542
2	42
3	11
4	4
5	2
6	3
7	1
8	1
9	1
10	1
11	1
12	1
15	1
19	1
22	1
25	1
34	1

5 Sammansatta fördelningar

5.1 De individuella och kollektiva modellerna

Ett försäkringsbolag erbjuder som bekant individer eller företag att mot en avgift (premien) täcka kostnaderna för eventuella skador av något visst slag som drabbar slumpmässigt. Det hela bygger på att kostnaderna kan fördelas över en stor grupp av försäkringstagare. I handhavandet av en försäkringsportfölj dyker ett antal frågor naturligen upp. Här följer några exempel.

(1) Principen är att premieintäkterna ska täcka kostnaderna för inträffade skador, men eftersom premierna betalas i förskott kan man inte vara helt säker på detta. Bolaget behöver alltså en kapitalreserv ifall portföljen skulle drabbas av ovanligt höga kostnader. Man ställs alltså inför två problem: Hur höga bör premierna vara? Hur stora reserver behövs?

(2) För att ytterligare sprida riskerna kan ett försäkringsbolag i sin tur teckna en återförsäkring hos ett återförsäkringsbolag. Mot en andel av premierna åtar sig återförsäkringsbolaget att vara med och betala om portföljen skulle drabbas av ovanligt många eller ovanligt stora skador. Det finns många tänkbara återförsäkringsarrangemang och det är ett svårt och viktigt problem att bestämma sig för ett lämpligt sådant.

(3) Hur försäkringsvillkoren ser ut påverkar förstås också hur mycket försäkringsbolaget får betala ut. Nivån på en självrisk kan varieras, vilken typ av skador försäkringen täcker kan vara olika, osv. Vid valet av försäkringsvillkor är det intressant att studera hur olika alternativ påverkar bolagets kostnader.

Centralt i ovanstående problemställningar är försäkringsbolagets totala kostnader för skador täckta av kontrakten i en viss försäkringsportfölj under en viss tidsperiod. Vi har tidigare studerat modeller för antalet skador som drabbar en portfölj och modeller för de enskilda skadebeloppen. Det är nu dags att studera hur dessa kan sättas samman till en modell för bolagets totala kostnad för en portfölj, samt hur fördelningen för den totala kostnaden kan beräknas.

Det finns två olika sätt att betrakta det hela, vilket leder till två olika modeller: den kollektiva modellen och den individuella modellen.

Den kollektiva modellen uppkommer när man bortser från vilka försäkringstagare som drabbas av skadorna och endast betraktar följden av skador som drabbar portföljen. Man betraktar ett visst tidsintervall, t ex ett år, och låter N beteckna antalet skador som inträffar under detta. Vidare låter man X_k beteckna ersättningsbeloppet som betalas ut för den k:te skadan som inträffar. Det totala beloppet som försäkringsbolaget får betala ut är då $S := X_1 + \cdots + X_N$. Man antar att $X_1, X_2, \ldots$ är oberoende, att de alla har samma fördelning, samt att de är oberoende av N. Givet fördelningarna för N och X_1, som man får skatta med hjälp av data, kan man i princip beräkna fördelningen för S. Vi kommer att studera ett par metoder för detta, dels beräkning via rekursionsformler, dels via approximation.

När man arbetar med den kollektiva modellen är det ofta rimligt att anta att N har en Poissonfördelning eller en blandad Poissonfördelning.

Den individuella modellen uppkommer när man i stället utgår från de enskilda försäkringskontrakten. Antag att portföljen innehåller n kontrakt som gäller under hela eller en del av den tidsperiod vi betraktar och låt Y_i beteckna det belopp försäkringsbolaget får betala ut till försäkringstagare nr i. För många typer av försäkringar kommer de flesta av variablerna $Y_1, \ldots, Y_n$ att anta värdet 0. Bolagets sammanlagda kostnad blir $S := Y_1 + \cdots + Y_n$. Vi antar även här att variablerna $Y_1, \ldots, Y_n$ är oberoende. I det som brukar kallas den individuella modellen gör man normalt inget antagande om att $Y_1, \ldots, Y_n$ har samma fördelning och vi ska återkomma till detta senare. Om vi antar att $Y_1, \ldots, Y_n$ är likafördelade, är $p := P(Y_k > 0)$ samma för alla k och den betingade fördelningen för Y_k givet att $Y_k > 0$, dvs

$$H(x) := P\big(Y_k \leq x \bigm| Y_k > 0\big)$$

är densamma för alla k. Man kan inse (Övning 5.1) att S under detta antagande har samma fördelning som $X_1 + \cdots + X_N$, där $N, X_1, \ldots, X_n$ är oberoende, variablerna $X_1, \ldots, X_n$ alla har fördelningsfunktionen H, och N är binomialfördelad med parametrar n och p:

$$P(N = k) = \binom{n}{k} p^k \, (1 - p)^{n-k}$$

Observera att variablerna $\{X_k\}$ här ej står för samma sak som i den kollektiva modellen. Här handlar det om kostnaden för en enskild försäkringstagare (den sammanlagda kostnaden för alla skador som drabbar försäkringstagaren), medan det i den kollektiva modellen rör sig om kostnaden för en enskild skada. Om sannolikheten för att en försäkringstagare drabbas av mer än en skada är liten, kommer förstås fördelningarna för X_k-variablerna (och N) att vara ungefär desamma i båda fallen.

Vi ser hur som helst att det är av intresse att studera fördelningar för variabler av typen $S = X_1 + \cdots + X_N$, där $N, X_1, X_2, \ldots$ är oberoende och $X_1, X_2, \ldots$ är likafördelade. Man säger att S har en sammansatt fördelning (*compound distribution*).

Det är mycket vanligt att man delat in försäkringstagarna i (mer eller mindre) homogena grupper, säg J stycken, där alla försäkringstagare i en grupp betalar samma premie. Detta leder till att bolagets totala kostnad för portföljen kan skrivas som

$$S = \sum_{j=1}^{J} \sum_{k=1}^{N_j} X_{jk}$$

där X_{jk} är kostnaden för den k:te skadan i grupp j och N_j är antalat skador i grupp j. Fördelningen för S i detta fall kommer vi att studera i nästa kapitel.

Vilka antaganden man gör när man formulerar sin modell styrs naturligtvis av de praktiska omständigheterna. Man måste kunna skatta parametrarna i sina modeller och man måste kunna utföra det numeriska beräkningsarbetet på rimlig tid.

5.2 Några grundläggande resultat

Vi ska nu studera fördelningen för $S = X_1 + \cdots + X_N$, där $N, X_1, X_2, \ldots$ är oberoende och $X_1, X_2, \ldots$ är likafördelade. Vi antar förstås alltid att N är fördelad på de ickenegativa heltalen, men antar för tillfället inget om fördelningen för variablerna $\{X_k\}$. Det vanligaste är att X_k:na står för enskilda skadekostnader och att man antar att variablerna har en täthet, men vi har också sett att blandade fördelningar dyker upp i samband med till exempel ett återförsäkringsbolags kostnader. Vidare dyker sammansatta fördelningar upp för vissa typer av risker, där en olycka kan ge upphov till mer än en skadeanmälan: i en tågolycka skadas eller dödas ofta mer än en människa, en storm orsakar skador på många hus, etc. Om X_k är antalet skador vid olycka k och N är antalet olyckor, blir det totala antalet skador $X_1 + \cdots + X_N$. I detta fall har variablerna $\{X_k\}$ en heltalsfördelning. I det här avsnittet ska vi ta fram några grundläggande resultat som inte har att göra med typen av fördelning för variablerna $\{X_k\}$, för att i kommande avsnitt specialisera oss mer.

Vi låter F beteckna den gemensamma fördelningsfunktionen för $\{X_k\}$-variablerna och sätter $p_k := P(N = k)$. Man brukar beteckna fördelningsfunktionen för variabeln $Y_n := X_1 + \cdots + X_n$ med F^{n*} och dessutom definiera

$$F^{0*}(x) := \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

dvs F^{0*} är fördelningsfunktionen för en variabel som med sannolikheten 1 antar värdet 0. Fördelningsfunktionen G för S kan då skrivas

$$\begin{aligned}
G(x) &= P(S \leq x) = \sum_{k=0}^{\infty} P\big(S \leq x \mid N = k\big) P(N = k) \\
&= P\big(S \leq x \mid N = 0\big)\, p_0 + \sum_{k=1}^{\infty} P\big(X_1 + \cdots + X_k \leq x \mid N = k\big)\, p_k \\
&= F^{0*}(x)\, p_0 + \sum_{k=1}^{\infty} P(X_1 + \cdots + X_k \leq x)\, p_k \\
&= F^{0*}(x)\, p_0 + \sum_{k=1}^{\infty} F^{k*}(x)\, p_k \\
&= \sum_{k=0}^{\infty} F^{k*}(x)\, p_k \qquad (5.1)
\end{aligned}$$

Fördelningsfunktionerna $\{F^{k*}\}$ kan beräknas rekursivt med hjälp av *faltningsformeln*:

$$\begin{aligned}
F^{n*}(y) &= P(Y_n \leq y) = \int P\big(Y_n \leq y \mid X_n = x\big) dF(x) \\
&= \int P\big(X_1 + \cdots + X_{n-1} \leq y - x \mid X_n = x\big) dF(x) \\
&= \int P\big(X_1 + \cdots + X_{n-1} \leq y - x\big) dF(x) \\
&= \int F^{(n-1)*}(y - x)\, dF(x) \qquad (5.2)
\end{aligned}$$

Formlerna (5.1) och (5.2) anger i princip ett sätt att numeriskt beräkna fördelningen för S, men vi kommer att studera mer effektiva metoder för detta. Härledningen av dessa grundar sig dock på ovanstående uttryck.

Låt nu M_S, M_N och M beteckna de momentgenererande funktionerna för S, N respektive

X_1. Vi finner att

$$\begin{aligned}
M_S(t) &= E\big[e^{tS}\big] = \sum_{k=0}^{\infty} E\big[e^{tS} \mid N = k\big] P(N = k) \\
&= P(N = 0) + \sum_{k=1}^{\infty} E\big[e^{t(X_1+\cdots+X_k)} \mid N = k\big] P(N = k) \\
&= P(N = 0) + \sum_{k=1}^{\infty} E\big[e^{t(X_1+\cdots+X_k)}\big] P(N = k) \\
&= P(N = 0) + \sum_{k=1}^{\infty} E\left[\prod_{i=1}^{k} e^{tX_i}\right] P(N = k) \\
&= P(N = 0) + \sum_{k=1}^{\infty} \prod_{i=1}^{k} E\big[e^{tX_i}\big] P(N = k) \\
&= \sum_{k=0}^{\infty} \big(M(t)\big)^k P(N = k) = \sum_{k=0}^{\infty} e^{k \log M(t)} P(N = k) \\
&= M_N\big(\log M(t)\big) \qquad (5.3)
\end{aligned}$$

Logaritmerar vi båda sidor, får vi följande relation mellan de motsvarande kumulantgenererande funktionerna Ψ_S, Ψ_N och Ψ:

$$\Psi_S(t) = \Psi_N(\Psi(t))$$

Genom att derivera detta uttryck en, två, respektive tre gånger och sätta in $t = 0$, erhåller vi följande relationer mellan kumulanterna $\{\kappa_j^S\}$, $\{\kappa_j^N\}$ och $\{\kappa_j\}$ för S, N och X_1:

$$\begin{aligned}
\kappa_1^S &= \kappa_1^N \, \kappa_1 \\
\kappa_2^S &= \kappa_2^N \, (\kappa_1)^2 + \kappa_1^N \, \kappa_2 \\
\kappa_3^S &= \kappa_3^N \, (\kappa_1)^3 + 3 \, \kappa_2^N \, \kappa_1 \, \kappa_2 + \kappa_1^N \, \kappa_3
\end{aligned}$$

För att ovanstående härledning skall vara giltig, måste vi förstås anta att de momentgenererande funktionerna existerar i en omgivning av origo. Uttrycken gäller dock bara man antar att momenten existerar och kan även härledas direkt, dvs utan användning av momentgenererande funktioner.

Om N är Poissonfördelad, säges $S = X_1 + \cdots + X_N$ ha en sammansatt Poissonfördelning (*compound Poisson distribution*). Vi ska nu studera denna fördelning mer i detalj.

Om variablerna $\{X_k\}$ har fördelningsfunktionen F och parametern i Poissonfördelningen för N heter λ, använder vi beteckningen $S \in \text{SaPo}(\lambda, F)$. Enligt vad vi fann ovan blir den kumulantgenererande funktionen för S

$$\Psi_S(t) = \Psi_N(\Psi(t)) = \lambda(e^{\Psi(t)} - 1) = \lambda(M(t) - 1) \tag{5.4}$$

Det följer att den j:te kumulanten för S är

$$\kappa_j^S = \Psi_S^{(j)}(0) = \lambda\, M^{(j)}(0) = \lambda\, \alpha_j$$

där α_j betecknar det j:te momentet kring origo för X_1.

Den sammansatta Poissonfördelningen har en intressant additivitetsegenskap. Antag att $S_1 \in$ SaPo(λ_1, F_1) och att $S_2 \in$ SaPo(λ_2, F_2), samt att S_1 och S_2 är oberoende. Vi låter de momentgenererande funktionerna för F_1 och F_2 betecknas med M_1 respektive M_2. Den kumulantgenererande funktionen för $S_1 + S_2$ blir då

$$\begin{aligned}
&\lambda_1\big(M_1(t) - 1\big) + \lambda_2\big(M_2(t) - 1\big) \\
&\quad = \big(\lambda_1\, M_1(t) + \lambda_2\, M_2(t)\big) - (\lambda_1 + \lambda_2) \\
&\quad = (\lambda_1 + \lambda_2)\bigg(\bigg(\frac{\lambda_1}{\lambda_1 + \lambda_2}\, M_1(t) + \frac{\lambda_2}{\lambda_1 + \lambda_2}\, M_2(t)\bigg) - 1\bigg)
\end{aligned}$$

Definiera

$$G(x) := \frac{\lambda_1}{\lambda_1 + \lambda_2}\, F_1(x) + \frac{\lambda_2}{\lambda_1 + \lambda_2}\, F_2(x)$$

Man förvissar sig lätt om att G är en fördelningsfunktion och att den momentgenererande funktionen för G är

$$\frac{\lambda_1}{\lambda_1 + \lambda_2}\, M_1(t) + \frac{\lambda_2}{\lambda_1 + \lambda_2}\, M_2(t)$$

Det följer att $S_1 + S_2 \in$ SaPo$(\lambda_1 + \lambda_2, G)$. Resultatet kan direkt generaliseras till summor av fler än två oberoende SaPo-variabler.

Ett slags omvändning till denna egenskap är följande. Låt $S, N, X_1, X_2, \ldots$ vara som i definitionen av den kollektiva modellen och låt N vara Poissonfördelad med parameter λ. Låt $a > 0$, låt S_1 vara summan av alla skador som understiger nivån a och S_2 summan av alla skador som överstiger a. Med andra ord, om vi sätter

$$Y_k := \begin{cases} X_k, & \text{om } X_k \le a \\ 0, & \text{om } X_k > a \end{cases} \qquad Z_k := \begin{cases} 0, & \text{om } X_k \le a \\ X_k, & \text{om } X_k > a \end{cases}$$

så är $S_1 = Y_1 + \cdots + Y_N$, $S_2 = Z_1 + \cdots + Z_N$ och $S = S_1 + S_2$. Man kan visa att S_1 och S_2 är oberoende och att $S_1 \in$ SaPo(λ_1, F_1), $S_2 \in$ SaPo(λ_2, F_2), där

$$\begin{aligned}
\lambda_1 &= \lambda P(X_1 \le a) = \lambda F(a) \qquad \lambda_2 = \lambda P(X_1 > a) = \lambda(1 - F(a)) \\
F_1(x) &= P\big(X_1 \le x \mid X_1 \le a\big) = \begin{cases} F(x)/F(a), & 0 \le x \le a \\ 1, & x > a \end{cases} \\
F_2(x) &= P\big(X_1 \le x \mid X_1 > a\big) = \begin{cases} 0, & 0 \le x \le a \\ (F(x) - F(a))/(1 - F(a)), & x > a \end{cases}
\end{aligned}$$

Vi kommer även att använda oss av beteckningarna $S \in \text{SaBi}(n, p, F)$ och $S \in \text{SaNeBi}(\alpha, \beta, F)$ då N är binomialfördelad med parametrar n och p, respektive negativt binomialfördelad med parametrar α och β. Dessa fördelningar har ej den behändiga additivitetsegenskap som kännetecknar den sammansatta Poissonfördelningen.

Exempel 5.1 Antag att N är Poissonfördelad och att $\{X_k\}$-variablerna har en logaritmisk fördelning med parameter r (se avsnitt 4.3). Om P_1 och P_2 betecknar de sannolikhetsgenererande funktionerna för N respektive X_1, får vi enkelt från (5.3) att den sannolikhetsgenererande funktionen för $X_1 + \cdots + X_N$ uppfyller

$$P(s) = P_1(P_2(s))$$

Vi har för $|s| < 1$ att

$$\begin{aligned}
P_1(s) &= e^{-\lambda(1-s)} \\
P_2(s) &= \sum_{n=1}^{\infty} s^n \frac{r^n}{-n \log(1-r)} = \frac{1}{-\log(1-r)} \sum_{n=1}^{\infty} \frac{(rs)^n}{n} \\
&= \frac{\log(1-rs)}{\log(1-r)}
\end{aligned}$$

Man får att

$$\begin{aligned}
P(s) &= P_1(P_2(s)) = \exp\left\{-\lambda\left(1 - \frac{\log(1-rs)}{\log(1-r)}\right)\right\} \\
&= \exp\left\{-\frac{\lambda}{\log(1-r)} \log \frac{1-r}{1-rs}\right\} \\
&= \left(\frac{1-r}{1-rs}\right)^{-\lambda/\log(1-r)} = \left(\frac{1-r}{1-r+r(1-s)}\right)^{-\lambda/\log(1-r)} \\
&= \left(1 + \frac{r}{1-r}(1-s)\right)^{\lambda/\log(1-r)}
\end{aligned}$$

Detta är sannolikhetsgenererande funktionen för en negativ binomialfördelning med parametrar

$$\alpha = \frac{\lambda}{-\log(1-r)} \qquad \beta = \frac{1-r}{r}$$

Den negativa binomialfördelningen kan alltså ses både som en blandad Poissonfördelning och som en sammansatt Poissonfördelning. □

Övning 5.1 *Antag att variablerna* $Y_1, \ldots, Y_n$ *är ickenegativa, oberoende och likafördelade, samt att* $0 < p < 1$*, där* $p := P(Y_k > 0)$*. Visa med hjälp av momentgenererande funktioner att* $Y_1 + \cdots + Y_n$ *har en SaBi*(n, p, H)*-fördelning, där* $H(x) := P(Y_1 \leq x \mid Y_1 > 0)$*.*

Övning 5.2 *För en stormskadeportfölj, låt* K *beteckna antalet stormar under ett år,* N_j *antalet skador orsakade av storm* j *och* X_{jk} *storleken på den* k*:te skadan orsakad av storm* j*. Försäkringsbolagets sammanlagda kostnad blir*

$$S = \sum_{j=1}^{K} \sum_{k=1}^{N_j} X_{jk}$$

Antag att K*,* $\{N_j\}$ *och* $\{X_{jk}\}$ *samtliga är oberoende av varandra. Antag vidare att variablerna* $\{N_j\}$ *är likafördelade och att även variablerna* $\{X_{jk}\}$ *är likafördelade. Visa att*

$$\begin{aligned} E[S] &= E[K]\, E[N_1]\, E[X_{11}], \\ V[S] &= V[K]\, (E[N_1])^2\, (E[X_{11}])^2 + E[K]\, V[N_1]\, (E[X_{11}])^2 \\ &+ E[K]\, E[N_1]\, V[X_{11}] \end{aligned}$$

Antag att alla variablernas momentgenererande funktioner existerar i en omgivning av origo.

Övning 5.3 *Låt* $N = N_1 + \cdots + N_K$*, där* $K, N_1, N_2, \ldots$ *är oberoende och* $N_1, N_2, \ldots$ *är likafördelade. Låt* K *vara Poissonfördelad och låt* N_1 *ha en utvidgad trunkerad negativ binomialfördelning med* $\alpha = -1/2$*. Visa med hjälp av sannolikhetsgenererande funktioner att* N *har en IGP-fördelning (Övning 4.3).*

5.3 Ett exempel: POT-modellen

För vissa försäkringstyper är det ett vanligt fenomen att kostnaderna för ett fåtal stora skador utgör en väsentlig del av kostnaderna för hela portföljen. Typiska exempel är försäkringar som skyddar mot stormskador, översvämningsskador eller industribränder. För att bland annat bestämma ett lämpligt återförsäkringsskydd är det viktigt för försäkringsbolag att försöka skatta hur ofta skador av en viss storleksordning förekommer. Som beslutsunderlag behöver man veta till exempel hur stor sannolikheten är att inga skador över en viss nivå inträffar under den närmaste tioårsperioden. Det är svårt att skatta en sådan sannolikhet med god precision, eftersom stora skador inträffar så sällan att datamaterialet ofta är magert.

En modell som föreslagits i detta sammanhang, och som har sitt ursprung inom andra tillämpningar än försäkringsmatematiken, är den så kallade POT-modellen, där POT står för *peaks over thresholds*. Man väljer en viss nivå eller tröskel u och studerar skador som överstiger denna nivå. Man betraktar ett visst tidsintervall, säg ett år, och låter X_k beteckna storleken på den k:te skadan som överstiger u. Man antar att N, antalet skador som överstiger u, är Poissonfördelat med parameter λ, och att $N, X_1, X_2, \ldots$ alla är oberoende. Vidare antar man att $X_1, X_2, \ldots$ är likafördelade, samt att $X_k - u$ är Paretofördelad med parametrar α och γ, dvs

$$F(x) := P(X_k - u \leq x) = 1 - \left(\frac{\alpha}{\alpha + x}\right)^\gamma, \qquad x \geq 0$$

Antagandena bakom POT-modellen kan motiveras på ett övertygande sätt, se Rootzén & Tajvidi (1995).

Modellen har en tilltalande egenskap som gör den lätt att arbeta med. Antag att vi i stället betraktar en högre tröskel $v > u$, låter M vara antalet skador som överstiger v och Y_k storleken på den k:te skadan som överstiger v. Vilka är relationerna mellan fördelningarna för N och M, respektive $X_k - u$ och $Y_k - v$?

Till att börja med ser vi att variablerna $Y_1, \ldots, Y_M$ utgör en delmängd av variablerna $X_1, \ldots, X_N$. Givet att $N = n$ gäller, eftersom $N, X_1, X_2, \ldots$ är oberoende, att antalet variabler bland $X_1, \ldots, X_n$ som överstiger v, dvs M, är binomialfördelat med parametrar n och $p = P(X_k > v) = 1 - F(v - u)$. Man får att

$$\begin{aligned}
P(M = m) &= \sum_{n=m}^{\infty} P\big(M = m \mid N = n\big) P(N = n) \\
&= \sum_{n=m}^{\infty} \binom{n}{m} p^m (1-p)^{n-m} \frac{\lambda^n}{n!} e^{-\lambda} \\
&= \frac{(\lambda p)^m}{m!} e^{-\lambda} \sum_{n=m}^{\infty} \frac{(\lambda(1-p))^{n-m}}{(n-m)!} \\
&= \frac{(\lambda p)^m}{m!} e^{-\lambda} e^{\lambda(1-p)} = \frac{(\lambda p)^m}{m!} e^{-\lambda p}
\end{aligned}$$

dvs M är Poissonfördelad med parameter $\lambda(1 - F(v-u))$.

Fördelningen för $Y_1 - v$ är densamma som fördelningen för $X_1 - v$ givet att X_1 når över v, dvs, för $x \geq 0$,

$$\begin{aligned} P(Y_1 - v > x) &= P\big(X_1 - v > x \bigm| X_1 > v\big) \\ &= \frac{P(X_1 - v > x, X_1 > v)}{P(X_1 > v)} = \frac{P(X_1 > v + x)}{P(X_1 > v)} \\ &= \frac{P(X_1 - u > v - u + x)}{P(X_1 - u > v - u)} = \frac{(\alpha/(\alpha + v - u + x))^\gamma}{(\alpha/(\alpha + v - u))^\gamma} \\ &= \left(\frac{\alpha + v - u}{\alpha + v - u + x}\right)^\gamma \end{aligned}$$

dvs $Y_1 - v$ har också en Paretofördelning, fast med parametern $\alpha + v - u$ i stället för α. Vi ser alltså att givet en basnivå u, får vi lätt fram fördelningarna för andra tröskelnivåer genom en enkel modifiering av parametrarna.

När ett försäkringsbolag ska bestämma i vilken utsträckning man behöver återförsäkra sig, är man bland annat intresserad av fördelningen för storleken på den största skadan inom de närmaste T åren. På grund av additiviteten hos Poissonfördelningen är antalet skador N över tröskeln u Poissonfördelat med parameter λT. Vi ska studera fördelningen för

$$Z := \max(X_1 - u, \ldots, X_N - u)$$

dvs hur långt över u den största skadan når. Fördelningsfunktionen för Z är

$$\begin{aligned} G(x) &:= P(Z \leq x) = \sum_{n=0}^{\infty} P\big(Z \leq x \bigm| N = n\big) P(N = n) \\ &= P(N = 0) + \sum_{n=1}^{\infty} P\big(X_1 - u \leq x, \ldots, X_n - u \leq x \bigm| N = n\big) P(N = n) \\ &= P(N = 0) + \sum_{n=1}^{\infty} \big(P(X_1 - u \leq x)\big)^n P(N = n) \\ &= \sum_{n=0}^{\infty} (F(x))^n \frac{(\lambda T)^n}{n!} e^{-\lambda T} = e^{\lambda T F(x)} e^{-\lambda T} \\ &= e^{-\lambda T(1 - F(x))} \end{aligned}$$

Följaktligen kan vi enkelt beräkna $G(x)$.

Som beslutsunderlag är man intresserad av kvantilerna för $G(x)$, dvs för ett givet värde på q vill man finna x_q som uppfyller $G(x_q) = 1 - q$. Med andra ord är man intresserad av det x_q

för vilket man kan säga att med sannolikheten $1-q$ inträffar ingen skada över nivån $u+x_q$ inom de närmaste T åren. Man kan enkelt lösa ut x_q ur ekvationen $G(x_q)=1-q$ och finner att

$$x_q = \alpha\left(\left(-\frac{1}{\lambda T}\log(1-q)\right)^{-1/\gamma} - 1\right)$$

Med hjälp av detta kan man enkelt göra en tabell över x_q för olika val av “risknivån” q.

ML-skattningarna av λ respektive α och γ räknas ut precis som vi tidigare gått igenom. Datamaterialet består av alla skador över nivån u under en följd (helst lång) bestående av t år, samt tidpunkterna då skadorna inträffade. Om n_j betecknar antalet (inflationsjusterade) skador över nivån u under år j, är ML-skattningen av λ lika med

$$\hat{\lambda} = \frac{1}{t}\sum_{j=1}^{t} n_j$$

Om $x_1,\ldots,x_n$ är storleken på skadorna, beräknas ML-skattningarna av α och γ på vanligt sätt utifrån observationerna $x_1-u,\ldots,x_n-u$.

Skattningarna av α och γ kan ofta vara ganska osäkra, det är typiskt för storskadeproblematiken att dataunderlaget är litet. För studier av osäkerheten i ML-skattningarna och en jämförelse med andra skattningsmetoder, se Rootzén & Tajvidi (1995), samt Hosking & Wallis (1987).

När det gäller valet av tröskelnivån u, får man prova sig fram. Ju större u väljs, desto mindre blir datamaterialet. Om u å andra sidan väljs för litet, kan det hända att modellen passar dåligt.

Övning 5.4 *Generera ett Poissonfördelat slumptal N med parameter λT, där $\lambda = 3.0$, $T = 15$. Generera därefter N stycken slumptal som är Paretofördelade med parametrarna $\alpha = 900000$ och $\gamma = 1.1$. Beräkna ML-skattningarna $\hat{\lambda}$, $\hat{\alpha}$ och $\hat{\gamma}$ utifrån det så erhållna datamaterialet. Beräkna slutligen skattningen*

$$\hat{x}_q = \hat{\alpha}\left(\left(-\frac{1}{\hat{\lambda} T}\log(1-q)\right)^{-1/\hat{\gamma}} - 1\right)$$

av x_q med $q = 0.99, T = 10$. Upprepa förfarandet några gånger och studera hur $\hat{x}_q$ varierar.

Kommentar: Ett slumptal N som är Poissonfördelat med parameter λ kan erhållas som

$$N = \textit{minsta värdet på } n \textit{ så att } U \le f_n,$$

där $p_n = \lambda^n \, e^{-\lambda}/n!$*,* $f_n = p_0 + \cdots + p_n$ *och* U *är ett slumptal som är likformigt fördelat på* $(0, 1)$.

Ett Paretofördelat slumptal X *kan erhållas som*

$$X = \alpha((1 - U)^{-\gamma} - 1)$$

där U *är likformigt fördelad på* $(0, 1)$.

Övning 5.5 *Bestäm* q*-kvantilen i fördelningen för* $X - (u + v)$*, givet att* $X > u + v$*, dvs bestäm* c *så att*

$$P\big(X - (u + v) > c \;\big|\; X > u + v\big) = q$$

Visa att q*-kvantilen är en växande funktion av* v*. Beräkna* 0.5*-kvantilen då* $\alpha = 900000$*,* $\gamma = 1.1$ *och* $v = 100$ *miljoner.*

5.4 Eliminering av nollskador

Vi antar återigen att $N, X_1, X_2, \ldots$ är oberoende och att $X_1, X_2, \ldots$ är likafördelade, samt sätter $S := X_1 + \cdots + X_N$. Vi gör inga särskilda antaganden om fördelningen för $\{X_k\}$-variablerna, förutom att de är ickenegativa.

Inom försäkringsmatematiken är det inte ovanligt att $P(X_1 = 0) > 0$. Antag exempelvis att X_k är ersättningsbeloppet för den k:te skadan som anmäls till försäkringsbolaget. Om bolaget av någon anledning beslutar att inte betala ut något, blir X_k lika med noll.

Ett annat exempel ges av situationen då ett bolag tecknar ett *excess of loss*-avtal med ett återförsäkringsbolag, så att återförsäkringsbolaget endast är med och betalar de skador som överstiger en viss nivå. Om X_k är återförsäkringsbolagets kostnad för den k:te skadan är $P(X_k = 0)$ i många fall ganska stor.

Ytterligare ett exempel kan vara att man studerar antalet personskador i bilolyckor $X_1 + \cdots + X_N$, där N står för antalet olyckor och X_k för antalet skadade i olycka nummer k. Om det endast uppstår materiella skador i olycka nr k, blir $X_k = 0$.

I detta avsnitt ska vi visa några resultat rörande sådana situationer, som det kan vara praktiskt att känna till. Vi sätter $a := P(X_1 = 0)$ och antar att $0 < a < 1$. Vi låter som tidigare F beteckna fördelningsfunktionen för X_1 och låter H vara fördelningsfunktionen för X_1

betingat av att $X_1 > 0$, dvs

$$H(x) = P\big(X_1 \leq x \mid X_1 > 0\big) \tag{5.5}$$

Om M är den momentgenererande funktionen för X_1 och $\tilde{M}$ är den momentgenererande funktionen för H, gäller att

$$\begin{aligned} M(t) &= E\big[e^{tX_1}\big] \\ &= E\big[e^{tX_1} \mid X_1 = 0\big] P(X_1 = 0) + E\big[e^{tX_1} \mid X_1 > 0\big] P(X_1 > 0) \\ &= a + \tilde{M}(t)\,(1-a) \end{aligned}$$

dvs

$$\tilde{M}(t) = \frac{M(t) - a}{1 - a}$$

Låt oss först anta att $S \in$ SaPo(λ, F), dvs att N är Poissonfördelad med parameter λ. Då kan vi enligt (5.4) skriva den kumulantgenererande funktionen för S som

$$\begin{aligned} \Psi_S(t) &= \log E\big[e^{tS}\big] = \lambda\big(M(t) - 1\big) \\ &= \lambda\big(M(t) - a - (1-a)\big) = \lambda(1-a)\left(\frac{M(t) - a}{1 - a} - 1\right) \\ &= \lambda(1-a)\big(\tilde{M}(t) - 1\big) \end{aligned} \tag{5.6}$$

Det följer att $S \in$ SaPo$\big(\lambda(1-a), H\big)$, dvs fördelningen för $X_1 + \cdots + X_N$ är densamma som för $Y_1 + \cdots + Y_M$, där $M, Y_1, Y_2, \ldots$ är oberoende, variablerna $Y_1, Y_2, \ldots$ alla har fördelningsfunktionen H och M är Poissonfördelad med parameter $\lambda(1-a)$.

Exempel 5.2 Antag att N är Poissonfördelad med parameter $\lambda = 20$ och att X_1 antar värdena 0, 1, 2 eller 3 med sannolikheterna $f_k := P(X_1 = k)$, där

$$f_0 = 0.1 \qquad f_1 = 0.2 \qquad f_2 = 0.3 \qquad f_3 = 0.4$$

Enligt ovanstående har $X_1 + \cdots + X_N$ samma fördelning som $Y_1 + \cdots + Y_M$, där M är Poissonfördelad med parameter $\lambda(1 - f_0) = 18$ och $h_k := P(Y_1 = k)$ ges av

$$\begin{aligned} h_0 &= 0 \\ h_1 &= 0.2/0.9 = 2/9 \\ h_2 &= 0.3/0.9 = 1/3 \\ h_3 &= 0.4/0.9 = 4/9 \end{aligned}$$

□

Antag nu i stället att N har en blandad Poissonfördelning: Givet $\Lambda = \lambda$ är N Poissonfördelad med parameter λ, där Λ är en stokastisk variabel med täthet u. Enligt (5.6) kan den momentgenererande funktionen skrivas som

$$\begin{aligned} E\big[e^{tS}\big] &= \int_0^\infty E\big[e^{tS} \mid \Lambda = \lambda\big] u(\lambda)\, d\lambda \\ &= \int_0^\infty \exp\big\{\lambda\,(1-a)\,(\tilde{M}(t)-1)\big\}\, u(\lambda)\, d\lambda \end{aligned}$$

Om vi substituerar $\theta = \lambda(1-a)$ i den sista integralen ovan och sätter

$$v(\theta) := u\left(\frac{\theta}{1-a}\right)\frac{1}{1-a} \tag{5.7}$$

får vi att

$$E\big[e^{tS}\big] = \int_0^\infty \exp\big\{\theta(\tilde{M}(t)-1)\big\} v(\theta)\, d\theta$$

Men detta är den momentgenererande funktionen för $Y_1 + \cdots + Y_M$, där variablerna $M, Y_1, Y_2, \ldots$ är oberoende, variablerna $\{Y_k\}$ alla har fördelningsfunktionen i (5.5) och M har en blandad Poissonfördelning med blandningstäthet v. Detta är för övrigt tätheten för den stokastiska variabeln $\Theta := \Lambda(1-a)$. Genom att missbruka notationen för sammansatt Poissonfördelning skulle vi kunna uttrycka detta som

$$\text{SaPo}(\Lambda, F) = \text{SaPo}(\Lambda(1-a), H)$$

Exempel 5.3 Antag att skadebeloppen $Z_1, Z_2, \ldots$ för en portfölj är Paretofördelade med parametrar τ respektive γ, dvs har fördelningsfunktionen

$$F_0(x) := 1 - \left(\frac{\tau}{\tau + x}\right)^\gamma, \qquad x \geq 0$$

och att antalet skador N är negativt binomialfördelat med parametrar α och β, dvs

$$P(N = n) = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)\, n!}\left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^n$$

Antag vidare att ett återförsäkringsbolag betalar beloppet $X_k := h(Z_k)$ i samband med skada nr k, där

$$h(x) = \begin{cases} 0, & \text{om } x \leq C \\ x - C, & \text{om } x > C \end{cases}$$

dvs man har en *excess of loss*-återförsäkring utan övre gräns. Återförsäkringsbolagets sammanlagda kostnad är $X_1 + \cdots + X_N$, vilken har en SaNeBi(α, β, F)-fördelning, där

$$F(x) = \begin{cases} 0, & \text{om } x < 0 \\ F_0(x + C), & \text{om } x \geq 0 \end{cases}$$

Vi noterar speciellt att

$$a = P(X_k = 0) = F_0(C) = 1 - \left(\frac{\tau}{\tau + C}\right)^{\gamma}$$

Den negativa binomialfördelningen är en blandad Poissonfördelning med blandningstäthet

$$u(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\,\lambda^{\alpha-1}\,e^{-\beta\lambda}, \qquad \lambda > 0$$

dvs en gammafördelning. Tätheten v i (5.7) blir

$$\begin{aligned} v(\theta) &= \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(\frac{\theta}{1-a}\right)^{\alpha-1} e^{-\beta\theta/(1-a)}\,\frac{1}{1-a} \\ &= \frac{(\beta/(1-a))^{\alpha}}{\Gamma(\alpha)}\,\theta^{\alpha-1}\,e^{-\beta\theta/(1-a)} \end{aligned}$$

Detta är tätheten för en gammafördelning med parametrar α och $\beta/(1-a)$.

Fördelningsfunktionen H i (5.5) blir, för $x \geq 0$,

$$\begin{aligned} H(x) &= P\big(X \leq x \mid X_1 > 0\big) = \frac{P(X_1 > 0,\, X_1 \leq x)}{P(X_1 > 0)} \\ &= \frac{P(0 < X_1 < x)}{P(X_1 > 0)} = \frac{F(x) - F(0)}{1 - F(0)} \\ &= \frac{F_0(x + C) - F_0(C)}{1 - F_0(C)} = \frac{(\tau/(\tau + C))^{\gamma} - (\tau/(\tau + x + C))^{\gamma}}{(\tau/(\tau + C))^{\gamma}} \\ &= 1 - \left(\frac{\tau + C}{\tau + C + x}\right)^{\gamma} \end{aligned}$$

vilket är fördelningsfunktionen för en Paretofördelning med parametrar $\tau + C$ och γ. Enligt vad vi fann ovan, har alltså återförsäkringsbolagets kostnad en SaNeBi$(\alpha, \beta/(1-a), H)$-fördelning. $\square$

Avslutningsvis gäller för den sammansatta binomialfördelningen att (lämnas som övning)

$$\text{SaBi}(n, p, F) = \text{SaBi}(n, p(1-a), H). \tag{5.8}$$

Övning 5.6 *Visa (5.8).*

Övning 5.7 *Antag att antalet skador, inklusive nollskador, för en portfölj är IGP-fördelat (se övning 4.3). Vad har antalet strikt positiva skador för fördelning?*

5.5 Panjers rekursionsformel

Vi fortsätter med samma beteckningar som tidigare: $N, X_1, X_2, \ldots$ är oberoende, N är en heltalsvariabel och $X_1, X_2, \ldots$ likafördelade. Vi ska nu studera hur man kan beräkna fördelningen för $S = X_1 + \cdots + X_N$. I det här avsnittet kommer vi att härleda *Panjers rekursionsformel*, som behandlar fallet då variablerna $\{X_k\}$ är heltalsvärda. Nästa avsnitt handlar om *Ströters rekursionsformel*, för fallet då variablerna $\{X_k\}$ har en täthet.

Antag alltså att variablerna $\{X_k\}$ är heltalsvärda. En sådan situation uppstår till exempel då en olycka kan ge upphov till flera skador (såsom vid naturkatastofer eller trafikolyckor). Om N betecknar antalet olyckor och X_j antalet skador orsakade av olycka j, får det totala antalet skador $S = X_1 + \cdots + X_N$ en sammansatt fördelning.

Ibland förekommer det också att skadebelopp är av typen nc, där n är ett heltal och $c > 0$ ett reellt tal. Exempelvis kan n vara antal förlorade arbetsdagar och c ersättningen per dag. Samma situation uppstår också om man avrundar alla skadebelopp till till exempel närmaste 1000-tal kronor. Om variablerna $\{X_k\}$ endast antar något av värdena $0, c, 2c, 3c, \ldots$, gäller detsamma för $S = X_1 + \cdots + X_N$. Det är då naturligt att använda c som enhet, dvs om skadekostnad j är nc, antar X_j värdet n.

Antag alltså att variablerna $\{X_j\}$ är icke-negativa och heltalsvärda och sätt, för $k = 0, 1, \ldots$,

$$p_k := P(N = k) \qquad f_k := P(X_1 = k) \qquad g_k := P(S = k)$$

Vi antar nu att sannolikheterna $\{p_k\}$ uppfyller

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \qquad n \geq 1 \tag{5.9}$$

för några konstanter a och b. Vi har tidigare noterat att detta är uppfyllt i de viktiga fallen då N har en Poisson-, binomial- eller negativ binomialfördelning. Vi ska nu se hur detta leder till ett samband mellan sannolikheterna $\{g_k\}$ och $\{f_k\}$, som kan användas för beräkning av $\{g_k\}$.

Vi använder oss av de sannolikhetsgenererande funktionerna: Sätt

$$\begin{aligned} P_1(t) &:= E\big[t^N\big] = \sum_{n=0}^{\infty} t^n\, p_n \\ P_2(t) &:= E\big[t^{X_1}\big] = \sum_{k=0}^{\infty} t^k\, f_k \\ P(t) &:= E\big[t^S\big] = \sum_{k=0}^{\infty} t^k\, g_k \end{aligned}$$

Mellan de sannolikhetsgenererande funktionerna gäller sambandet

$$\begin{aligned} P(t) &= E\big[t^S\big] = \sum_{n=0}^{\infty} E\big[t^S \mid N = n\big]\, p_n \\ &= p_0 + \sum_{n=1}^{\infty} E\big[t^{X_1+\cdots+X_n} \mid N = n\big]\, p_n = p_0 + \sum_{n=1}^{\infty} E\big[t^{X_1+\cdots+X_n}\big]\, p_n \\ &= p_0 + \sum_{n=1}^{\infty}\prod_{k=1}^{n} E\big[t^{X_k}\big]\, p_n = p_0 + \sum_{n=1}^{\infty}\big(P_2(t)\big)^n\, p_n \\ &= \sum_{n=0}^{\infty}\big(P_2(t)\big)^n\, p_n = P_1(P_2(t)) \end{aligned} \tag{5.10}$$

Vi ska härnäst se att rekursionssambandet (5.9) implicerar en viss egenskap hos P_1. Vi skriver om (5.9):

$$n\, p_n = (an + b)\, p_{n-1} = an\, p_{n-1} + b\, p_{n-1}$$

Om vi multiplicerar båda sidor med t^{n-1} och summerar över $n \geq 1$, får vi

$$\sum_{n=1}^{\infty} n\, t^{n-1}\, p_n = \sum_{n=1}^{\infty} a\, n\, t^{n-1}\, p_{n-1} + \sum_{n=1}^{\infty} b\, t^{n-1}\, p_{n-1}$$

Vi kan skriva detta som

$$\frac{d}{dt}\sum_{n=0}^{\infty} t^n\, p_n = a\,\frac{d}{dt}\sum_{n=1}^{\infty} t^n\, p_{n-1} + b\sum_{n=1}^{\infty} t^{n-1}\, p_{n-1}$$

eller

$$\frac{d}{dt}\sum_{n=0}^{\infty} t^n\, p_n = a\,\frac{d}{dt}\Big\{t\sum_{n=1}^{\infty} t^{n-1}\, p_{n-1}\Big\} + b\sum_{n=1}^{\infty} t^{n-1}\, p_{n-1}$$

Vi får alltså att

$$\frac{d}{dt}\, P_1(t) = a\,\frac{d}{dt}\Big\{t\, P_1(t)\Big\} + b\, P_1(t)$$

dvs

$$P_1'(t) = a\big(P_1(t) + t\,P_1'(t)\big) + b\,P_1(t)$$

Vi skriver denna ekvation som

$$P_1'(t)\,(1 - at) = (a + b)\,P_1(t). \tag{5.11}$$

Vi såg i (5.10) att

$$P(t) = P_1(P_2(t))$$

Om vi deriverar båda sidor, får vi att

$$P'(t) = P_1'(P_2(t))\,P_2'(t)$$

Om vi multiplicerar båda sidor med $(1 - a\,P_2(t))$ och använder (5.11), får vi

$$\begin{aligned}(1 - a\,P_2(t))\,P'(t) &= (1 - a\,P_2(t))\,P_1'(P_2(t))\,P_2'(t)\\ &= (a + b)\,P_1(P_2(t))\,P_2'(t)\\ &= (a + b)\,P(t)\,P_2'(t)\end{aligned} \tag{5.12}$$

Vi ser alltså att (5.9) leder fram till ett samband mellan P och P_2, uttryckt i termer av a och b. Vi ska nu använda detta till att härleda ett samband mellan sannolikheterna $\{g_k\}$ och $\{f_k\}$.

Vi skriver sambandet i (5.12) som

$$P'(t) = a\,P_2(t)\,P'(t) + (a + b)\,P_2'(t)\,P(t). \tag{5.13}$$

Låt $P^{(n)}$ beteckna n:te derivatan av funktionen P och motsvarande för P_2. Om vi deriverar båda sidor i (5.13) $n - 1$ gånger, får vi via Leibniz regel för derivatan av en produkt,

$$\begin{aligned}P^{(n)}(t) &= a\sum_{k=0}^{n-1}\binom{n-1}{k}P_2^{(k)}(t)\,P^{(n-1-k+1)}(t)\\ &+ (a + b)\sum_{k=0}^{n-1}\binom{n-1}{k}P_2^{(k+1)}(t)\,P^{(n-1-k)}(t)\end{aligned}$$

Sätter vi $t = 0$ i detta uttryck och använder att

$$P^{(k)}(0) = k!\,g_k, \qquad P_2^{(k)}(0) = k!\,f_k$$

får vi

$$\begin{aligned}n!\,g_n &= a\sum_{k=0}^{n-1}\frac{(n-1)!}{k!\,(n-1-k)!}\,k!\,f_k\,(n-k)!\,g_{n-k}\\ &+ (a + b)\sum_{k=0}^{n-1}\frac{(n-1)!}{k!\,(n-1-k)!}\,(k+1)!\,f_{k+1}\,(n-1-k)!\,g_{n-1-k}\end{aligned}$$

dvs

$$
\begin{aligned}
n\, g_n &= a \sum_{k=0}^{n-1} (n-k)\, f_k\, g_{n-k} + (a+b) \sum_{k=0}^{n-1} (k+1)\, f_{k+1}\, g_{n-(k+1)} \\
&= a \sum_{k=0}^{n-1} (n-k)\, f_k\, g_{n-k} + (a+b) \sum_{k=1}^{n} k\, f_k\, g_{n-k} \\
&= a \sum_{k=0}^{n} (n-k)\, f_k\, g_{n-k} + (a+b) \sum_{k=0}^{n} k\, f_k\, g_{n-k} \\
&= \sum_{k=0}^{n} (an+bk)\, f_k\, g_{n-k}
\end{aligned}
$$

Om vi delar med n på båda sidor, får vi att

$$
\begin{aligned}
g_n &= \sum_{k=0}^{n} \left(a + \frac{bk}{n}\right) f_k\, g_{n-k} \\
&= \sum_{k=1}^{n} \left(a + \frac{bk}{n}\right) f_k\, g_{n-k} + a\, f_0\, g_n
\end{aligned}
$$

Löser vi ut g_n, får vi Panjers rekursionsformel:

$$
g_n = \frac{1}{1-af_0} \sum_{k=1}^{n} \left(a + \frac{bk}{n}\right) f_k\, g_{n-k}, \qquad n \geq 1 \tag{5.14}
$$

Som startvärde för rekursionerna får vi från (5.10) att

$$
g_0 = P(0) = P_1(P_2(0)) = P_1(f_0)
$$

Om N exempelvis är Poissonfördelad med parameter λ, är ju $P_1(t) = e^{-\lambda(1-t)}$, så $g_0 = e^{-\lambda(1-f_0)}$.

Formeln ser något enklare ut då $f_0 = 0$:

$$
g_n = \sum_{k=1}^{n} \left(a + \frac{bk}{n}\right) f_k\, g_{n-k}\,, \qquad n \geq 1
$$

och $g_0 = p_0$. Observera att vi alltid kan erhålla denna situation med hjälp av resultaten i avsnitt 5.4.

Notera att om $f_k = 0$ för $k > m$, får vi att

$$
g_n = \frac{1}{1-af_0} \sum_{k=1}^{m} \left(a + \frac{bk}{n}\right) f_k\, g_{n-k}\,, \qquad \text{för } m \leq n
$$

dvs för att beräkna g_n behövs endast $g_{n-m}, \ldots, g_{n-1}$. Detta reducerar tiden för beräkningarna.

Låt oss sammanställa en tabell som kan användas vid beräkningar med Panjers rekursionsformel då N har fördelningen Poisson:

$$p_n = \frac{\lambda^n}{n!}\, e^{-\lambda}$$

binomial:

$$p_k = \binom{n}{k} q^k\, (1-q)^{n-k}, \qquad 0 \le k \le n$$

eller negativ binomial:

$$p_k = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\, k!}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{k}, \qquad k \ge 0$$

Har man ett färdigt program för Panjerrekursion, behöver man endast mata in g_0, a, b samt $\{f_k\}$.

Fördelning	g_0	a	b
Poisson	$e^{-\lambda(1-f_0)}$	0	λ
Binomial	$(1-q(1-f_0))^n$	$-\frac{q}{1-q}$	$\frac{(n+1)q}{1-q}$
Neg Bin	$(1+\beta^{-1}(1-f_0))^{-\alpha}$	$\frac{1}{\beta+1}$	$\frac{\alpha-1}{\beta+1}$

Vi ska nu titta på ett alternativt bevis av Panjers rekursionsformel. Man kan med rätta fråga sig varför vi tar upp två bevis. Orsakerna är följande. Beviset ovan är konstruktivt och ledde oss fram till rekursionsformeln. Samma idéer kan användas för att härleda rekursionsformler för andra situationer. Det bevis vi nu ska presentera förutsätter, kan man säga, att man vet vad svaret ska bli. Det är dock en aning elegantare och går framför allt att generalisera utan vidare till fallet då X_1 har en täthet, vilket vi tar upp i nästa avsnitt.

Vi antar att $f_0 = 0$ (vi har redan kommenterat att detta inte innebär någon inskränkning). Sätt

$$f_k^{n*} := P(X_1 + \cdots + X_n = k), \qquad n \ge 1$$

Speciellt har vi alltså att $f_k^{1*} = f_k$. På grund av antagandet $f_0 = 0$, gäller att $f_k^{n*} = 0$ för $k \le n-1$.

Vi kommer att behöva *faltningsformeln*:

$$\begin{aligned}
f_k^{(n+1)*} &= P(X_1 + \cdots + X_{n+1} = k) \\
&= \sum_{j=1}^{k-n} P\big(X_1 + \cdots + X_{n+1} = k \mid X_{n+1} = j\big) P\big(X_{n+1} = j\big) \\
&= \sum_{j=1}^{k-n} P\big(X_1 + \cdots + X_n = k - j \mid X_{n+1} = j\big) P\big(X_{n+1} = j\big) \\
&= \sum_{j=1}^{k-n} P\big(X_1 + \cdots + X_n = k - j\big) P\big(X_{n+1} = j\big) \\
&= \sum_{j=1}^{k-n} f_{k-j}^{n*} f_j
\end{aligned}$$

Vi noterar också att $g_0 = p_0$ och att, för $k \geq 1$,

$$\begin{aligned}
g_k &= P(S = k) = \sum_{n=0}^{\infty} P\big(S = k \mid N = n\big) P(N = n) \\
&= \sum_{n=1}^{k} P\big(X_1 + \cdots + X_n = k\big) p_n = \sum_{n=1}^{k} f_k^{n*} p_n \qquad (5.15)
\end{aligned}$$

Beviset bygger på följande resultat, som vi visar först:

$$\sum_{j=1}^{k-n+1} j \, \frac{f_j \, f_{k-j}^{(n-1)*}}{f_k^{n*}} = E\big[X_1 \mid X_1 + \cdots + X_n = k\big] = \frac{k}{n}. \qquad (5.16)$$

När det gäller den första likheten i (5.16), notera att

$$\begin{aligned}
&P\big(X_1 = j \mid X_1 + \cdots + X_n = k\big) \\
&\quad = \frac{P(X_1 = j, X_1 + \cdots + X_n = k)}{P(X_1 + \cdots + X_n = k)} \\
&\quad = \frac{P(X_1 = j, X_2 + \cdots + X_n = k - j)}{P(X_1 + \cdots + X_n = k)} \\
&\quad = \frac{P(X_1 = j) \, P(X_2 + \cdots + X_n = k - j)}{P(X_1 + \cdots + X_n = k)} \\
&\quad = \frac{P(X_1 = j) \, P(X_1 + \cdots + X_{n-1} = k - j)}{P(X_1 + \cdots + X_n = k)} \\
&\quad = \frac{f_j \, f_{k-j}^{(n-1)*}}{f_k^{n*}}
\end{aligned}$$

Man får att

$$\begin{aligned}
&E\big[X_1 \,\big|\, X_1+\cdots+X_n=k\big] \\
&\quad= \sum_{j=0}^{k-n+1} j\,P\big(X_1=j \,\big|\, X_1+\cdots+X_n=k\big) \\
&\quad= \sum_{j=0}^{k-n+1} j\,\frac{f_j\,f_{k-j}^{(n-1)*}}{f_k^{n*}}.
\end{aligned}$$

Den andra likheten i (5.16) är en följd av att $X_1, X_2, \ldots$ är oberoende och likafördelade:

$$\begin{aligned}
&E\big[X_1 \,\big|\, X_1+\cdots+X_n=k\big] \\
&\quad= \frac{1}{n}\sum_{j=1}^{n} E\big[X_1 \,\big|\, X_1+\cdots+X_n=k\big] \\
&\quad= \frac{1}{n}\sum_{j=1}^{n} E\big[X_j \,\big|\, X_1+\cdots+X_n=k\big] \\
&\quad= \frac{1}{n}\,E\big[X_1+\cdots+X_n \,\big|\, X_1+\cdots+X_n=k\big] = \frac{k}{n}
\end{aligned}$$

Därmed är (5.16) visad.

Vi skriver (5.16) som

$$\frac{1}{n}\,f_k^{n*} = \frac{1}{k}\sum_{j=1}^{k-n+1} j\,f_j\,f_{k-j}^{(n-1)*}$$

Med användning av detta och faltningsformeln, fås att

$$\begin{aligned}
g_k &= \sum_{n=1}^{k} p_n\, f_k^{n*} = \sum_{n=1}^{k}\left(a+\frac{b}{n}\right)p_{n-1}\, f_k^{n*} \\
&= (a+b)\,p_0\,f_k + \sum_{n=2}^{k} a\,p_{n-1}\,f_k^{n*} + \sum_{n=2}^{k} b\,p_{n-1}\,\frac{1}{n}\,f_k^{n*} \\
&= (a+b)\,p_0\,f_k + \sum_{n=2}^{k} a\,p_{n-1}\sum_{j=1}^{k-n+1} f_{k-j}^{(n-1)*}\,f_j \\
&+ \sum_{n=2}^{k} b\,p_{n-1}\,\frac{1}{k}\sum_{j=1}^{k-n+1} j\,f_j\,f_{k-j}^{(n-1)*} \\
&= (a+b)\,p_0\,f_k + \sum_{j=1}^{k-1}\sum_{n=2}^{k-j+1} a\,f_j\,p_{n-1}\,f_{k-j}^{(n-1)*} \\
&+ \sum_{j=1}^{k-1}\sum_{n=2}^{k-j+1}\frac{bj}{k}\,f_j\,p_{n-1}\,f_{k-j}^{(n-1)*} \\
&= (a+b)\,p_0\,f_k + \sum_{j=1}^{k-1} a\,f_j\sum_{m=1}^{k-j} p_m\,f_{k-j}^{m*} \\
&+ \sum_{j=1}^{k-1}\frac{bj}{k}\sum_{m=1}^{k-j} p_m\,f_{k-j}^{m*} \\
&= (a+b)\,g_0 f_k + \sum_{j=1}^{k-1} a\,f_j\,g_{k-j} + \sum_{j=1}^{k-1}\frac{bj}{k}\,f_j\,g_{k-j} \\
&= \sum_{j=1}^{k}\left(a+\frac{bj}{k}\right)f_j\,g_{k-j}
\end{aligned}$$

Därmed är Panjers rekursionsformel visad.

Det finns andra, något mer komplicerade rekursionsformler för situationer där Panjers rekursionsformel inte kan användas, se till exempel Hesselager (1996).

Övning 5.8 *Kontrollera att tabellen över g_0, a och b för olika val av fördelningen för N stämmer.*

Övning 5.9 *Vi har nämnt att man genom att modifiera parametrarna i fördelningen för N kan erhålla situationen $f_0 = 0$. Men då måste ju Panjers rekursionsformel se likadan ut* *vare sig man gör denna förändring eller inte. Kontrollera detta då N är Poisson-, binomial- respektive negativt binomialfördelad.*

Övning 5.10 *Låt beteckningarna vara desamma som i detta avsnitt och antag att sannolikheterna $\{p_n\}$ uppfyller sambandet*

$$p_n = \left(a_1 + \frac{b_1}{n}\right)p_{n-1} + \left(a_2 + \frac{b_2}{n}\right)p_{n-2}, \qquad n \geq 2$$

Antag att $f_0 = 0$. Visa Sundts rekursionsformel*:*

$$g_k = \sum_{j=1}^{k}\left[\left(a_1 + \frac{b_1 j}{k}\right)f_j + \left(a_2 + \frac{b_2 j}{2k}\right)f_j^{2*}\right]g_{k-j} + (p_1 - (a_1 + b_1)p_0)f_k, \quad k \geq 1$$

Ledning: Modifiera det andra beviset för Panjers rekursionsformel. Visa först att, för $1 \leq m \leq n-1$,

$$f_k^{n*} = \frac{n}{mk}\sum_{j=1}^{k} j\, f_j^{m*}\, f_{k-j}^{(n-m)*}$$

samt följande generalisering av faltningsformeln:

$$f_k^{n*} = \sum_{j=1}^{k} f_j^{m*}\, f_{k-j}^{(n-m)*}, \qquad 1 \leq m \leq n-1$$

5.6 Ströters rekursionsformel

Vi fortsätter med beteckningarna i föregående avsnitt och ska nu se hur man numeriskt kan beräkna fördelningen för S i det fall då variablerna $\{X_k\}$ har en täthet.

Antag alltså att fördelningsfunktionen F för X_k:na är (kontinuerligt) deriverbar, samt att $F(0) = 0$, och sätt $f(x) := F'(x)$, $x > 0$. Låt som tidigare $p_n := P(N = n)$ och låt G beteckna fördelningsfunktionen för S. Då gäller att, för $y \geq 0$,

$$\begin{aligned} G(y) &= \sum_{n=0}^{\infty} p_n\, F^{n*}(y) = p_0 + \sum_{n=1}^{\infty} p_n\, F^{n*}(y) \\ &= p_0 + \sum_{n=1}^{\infty} p_n \int_0^y f^{n*}(x)\, dx = p_0 + \int_0^y \sum_{n=1}^{\infty} p_n\, f^{n*}(x)\, dx \end{aligned}$$

Sätt

$$g(x) := \sum_{n=1}^{\infty} p_n\, f^{n*}(x), \qquad x > 0$$

så att

$$G(y) = p_0 + \int_0^y g(x)\,dx$$

Vi ska härleda en formel för beräkning av g, varefter G kan beräknas.

Genom att följa beviset för Panjers rekursionsformel, får man (Övning 5.11) att

$$g(y) = p_1\, f(y) + \int_0^y \left(a + \frac{bx}{y}\right) f(x)\, g(y-x)\,dx \tag{5.17}$$

Med beteckningarna

$$c(y) := p_1\, f(y) \qquad d(x,y) := \left(a + \frac{bx}{y}\right) f(x)$$

ser vi alltså att g uppfyller integralekvationen

$$g(y) = c(y) + \int_0^y d(x,y)\, g(y-x)\,dx \tag{5.18}$$

Det finns en hel del skrivet om hur man löser integralekvationer numeriskt och vi ska nu beskriva en förhållandevis enkel metod.

Första steget är att approximera integralen i (5.18) enligt någon metod för numerisk integration. En vanlig metod utgår från den s k *Simpsons formel*:

$$\int_y^{y+2h} \xi(x)\,dx \approx \frac{h}{3}\big[\xi(y) + 4\,\xi(y+h) + \xi(y+2h)\big]$$

Om ξ är ett polynom av ordningen 3 eller mindre, har man likhet i Simpsons formel och för godtyckliga funktioner med fyra kontinuerliga derivator är approximationsfelet av storleksordningen h^5 då $h \to 0$.

Vi kommer också att använda en variant som brukar kallas Simpsons $3/8$ -formel:

$$\int_y^{y+3h} \xi(x)\,dx \approx \frac{3h}{8}\big(\xi(y) + 3\,\xi(y+h) + 3\,\xi(y+2h) + \xi(y+3h)\big]$$

Låt $h > 0$ vara en given steglängd och låt k vara ett heltal större än eller lika med 1. Då beräknas $g(kh)$ approximativt enligt följande.

Om $k \geq 2$ är jämnt, delar man upp integralen i (5.17) i bitar av längden $2h$ och tillämpar

Simpsons formel på var och en av dem:

$$\begin{aligned}
g(kh) &= c(kh) + \int_0^{kh} d(x,kh)\,g(kh-x)\,dx \\
&= c(kh) + \int_0^{2h} d(x,kh)\,g(kh-x)\,dx + \cdots + \int_{kh-2h}^{kh} d(x,kh)\,g(kh-x)\,dx \\
&\approx c(kh) + \frac{h}{3}\big[d(0,kh)\,g(kh) + 4\,d(h,kh)\,g((k-1)h) + d(2h,kh)\,g((k-2)h)\big] \\
&+ \cdots + \frac{h}{3}\big[d((k-2)h,kh)\,g(2h) + 4\,d((k-1)h,kh)\,g(h) + d(kh,kh)\,g(0)\big] \\
&= c(kh) + \frac{h}{3}\big\{d(0,kh)\,g(kh) + 4\big[d(h,kh)\,g((k-1)h) + d(3h,kh)\,g((k-3)h) \\
&+ \cdots + d((k-1)h,kh)\,g(h)\big] \\
&+ 2\big[d(2h,kh)\,g((k-2)h) + d(4h,kh)\,g((k-4)h) + \ldots \\
&+ d((k-2)h,kh)\,g((k-2)h)\big] + d(kh,kh)\,g(0)\big\}
\end{aligned}$$

Man får alltså approximationen

$$g(kh) \approx \big(1 - w_{k0}\,d(0,kh)\big)^{-1}\Big[c(kh) + \sum_{j=1}^{k} w_{kj}\,d(jh,kh)\,g((k-j)h)\Big], \tag{5.19}$$

där $w_{k0} = w_{kk} = h/3$ och i övrigt $w_{kj} = 4h/3$ för udda j samt $w_{kj} = 2h/3$ för jämna j.

Om i stället $k \geq 3$ är udda, använder man Simpsons formel som ovan, utom på de tre sista delintervallen, där man i stället använder Simpsons $3/8$-formel:

$$\begin{aligned}
\int_{(k-3)h}^{kh} d(x,kh)\,g(kh-x)\,dx &\approx \frac{3h}{8}\big(d((k-3)h,kh)\,g(3h) \\
&+ 3\,d((k-2)h,kh)\,g(2h) + 3\,d((k-1)h,kh)\,g(h) + d(kh,kh)\,g(0)\big)
\end{aligned}$$

Man får då igen formeln (5.19), fast med skillnaden

$$\begin{aligned}
w_{k,k-3} &= \frac{h}{3} + \frac{3h}{8} = \frac{17\,h}{24} \\
w_{k,k-2} &= w_{k,k-1} = \frac{9h}{8} \\
w_{kk} &= \frac{3h}{8}
\end{aligned}$$

förutom för $k = 3$, där $w_{30} = 3/8$.

I fallet $k = 1$, slutligen, använder man den enkla *trapetsregeln*:

$$\begin{aligned}
g(h) &= c(h) + \int_0^h d(x,h)\,g(h-x)\,dx \\
&\approx c(h) + \frac{h}{2}\,d(0,h)\,g(h) + \frac{h}{2}\,d(h,h)\,g(0)
\end{aligned}$$

dvs formeln (5.19) gäller med $w_{10} = w_{11} = h/2$.

Följande tabell visar koefficienterna för h i w_{kj} för $k = 0, 1, \ldots, 8$:

$k \backslash j$	0	1	2	3	4	5	6	7	8
1	1/2	1/2							
2	1/3	4/3	1/3						
3	3/8	9/8	9/8	3/8					
4	1/3	4/3	2/3	4/3	1/3				
5	1/3	4/3	17/24	9/8	9/8	3/8			
6	1/3	4/3	2/3	4/3	2/3	4/3	1/3		
7	1/3	4/3	2/3	4/3	17/24	9/8	9/8	3/8	
8	1/3	4/3	2/3	4/3	2/3	4/3	2/3	4/3	1/3

Sätter man likhet i (5.19), får man en rekursionsformel för (approximativ) beräkning av $g(0), g(h), g(2h), \ldots$ Genom att välja h allt mindre, får man ett allt mindre approximationsfel. Å andra sidan ökar exekveringstiden när h blir mindre.

Ett startvärde för rekursionerna erhålls genom att låta $y \to 0$ i (5.17):

$$g(0) = p_1\, f(0+) := p_1 \lim_{h \downarrow 0} f(h)$$

Med denna rekursionsmetod erhålls $g(0), g(h), g(2h), \ldots$ Med ovanstående metoder för numerisk integralberäkning kan man slutligen beräkna fördelningsfunktionen G:

$$G(kh) = p_0 + \int_0^{kh} g(x)\, dx \approx p_0 + \sum_{j=0}^{k} w_{kj}\, g(jh)$$

där $\{w_{kj}\}$ är desamma som ovan.

Man får pröva sig fram till ett lämpligt värde på h: När resultatet inte längre förändras nämnvärt när man minskar h, kan man stanna. Man bör verifiera att fördelningsfunktionen växer upp mot 1 — detta är ett tecken på att avrundningsfelen är försumbara.

I samband med så kallad *stop loss*-återförsäkring är man intresserad av den stokastiska variabeln

$$S' := \begin{cases} S, & \text{om } S \le M \\ M, & \text{om } S > M \end{cases} \tag{5.20}$$

där $M > 0$. Om k och h väljs så att $M = kh$, kan väntevärdet för S' beräknas approximativt som

$$\begin{aligned} E[S'] &= 0 \cdot p_0 + \int_0^M x\, g(x)\, dx + \int_M^\infty M\, g(x)\, dx \\ &= \int_0^M x\, g(x)\, dx + M(1 - G(M)) \\ &\approx \sum_{j=0}^{k} w_{kj}\, jh\, g(jh) + M\Big(1 - p_0 - \sum_{j=0}^{k} w_{kj}\, g(jh)\Big) \\ &= M(1 - p_0) - \sum_{j=0}^{k} w_{kj}\, (k - j)h\, g(jh) \end{aligned}$$

Högre moment för S kan beräknas på liknande sätt.

I Ströter (1985) beskrivs också en mer avancerad metod för numerisk lösning av (5.17); se även Panjer & Willmot (1992), Appendix D.

Övning 5.11 *Härled (5.17).*

Övning 5.12 *Antag att N är Poissonfördelad med parameter λ och att X_1 är lognormalfördelad med parametrar $\mu = 0$ och $\sigma = 1$. Beräkna fördelningsfunktionen för S med Ströters rekursionsmetod för $\lambda = 10$ och $\lambda = 20$ samt $\lambda = 100$.*

5.7 Approximationer

Även med dagens datorer kan "exakt"beräkning av fördelningen för S vara alltför tidskrävande. Det är därför av intresse att ha tillgång till bra approximationer som går snabbt att räkna ut. Vi ska ta upp två av de mest använda, NP-approximation och gammaapproximation. Den som vill läsa mer om approximationsmetoder kan studera t ex Buchwalder, Chevalier & Klüppelberg (1993).

Låt oss som inledning anta att antalet skador är Poissonfördelat med parameter λ och låt $\alpha_k = E\big[X_1^k\big]$. Då gäller ju att

$$E[S] = \lambda\alpha_1, \qquad V[S] = \lambda\alpha_2$$

Låt M_S och M beteckna de momentgenererande funktionerna för S respektive X_1. Vi utgår från den normerade variabeln

$$Z := \frac{S - E[S]}{\sqrt{V[S]}} = \frac{S - \lambda\alpha_1}{\sqrt{\lambda\alpha_2}}$$

som ju har väntevärde 0 och varians 1. Den momentgenererande funktionen för Z är

$$\begin{aligned} E[e^{tZ}] &= E\left[\exp\left\{t\,\frac{S-\lambda\alpha_1}{\sqrt{\lambda\alpha_2}}\right\}\right] \\ &= E\left[\exp\left\{\frac{t}{\sqrt{\lambda\alpha_2}}\,S - \frac{\lambda\alpha_1}{\sqrt{\lambda\alpha_2}}\,t\right\}\right) \\ &= M_S\left(\frac{t}{\sqrt{\lambda\alpha_2}}\right)\exp\left\{-\frac{\lambda\alpha_1}{\sqrt{\lambda\alpha_2}}\,t\right\} \\ &= \exp\left\{\lambda\left[M\left(\frac{t}{\sqrt{\lambda\alpha_2}}\right) - 1\right] - \frac{\lambda\alpha_1}{\sqrt{\lambda\alpha_2}}\,t\right\} \end{aligned}$$

Om X_1 har fördelningsfunktionen F, gäller att

$$\begin{aligned} M(t) &= \int e^{tx}\,dF(x) = \int \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}\,dF(x) \\ &= \sum_{k=0}^{\infty} \int x^k\,dF(x)\,\frac{t^k}{k!} = \sum_{k=0}^{\infty} \alpha_k\,\frac{t^k}{k!} \end{aligned}$$

Den momentgenererande funktionen för Z kan alltså skrivas som

$$\begin{aligned} &\exp\left\{\lambda\left[1 + \alpha_1\,\frac{t}{\sqrt{\lambda\alpha_2}} + \frac{\alpha_2}{2}\left(\frac{t}{\sqrt{\lambda\alpha_2}}\right)^2 + \frac{\alpha_3}{6}\left(\frac{t}{\sqrt{\lambda\alpha_2}}\right)^3 + \cdots - 1\right] - \frac{\lambda\alpha_1}{\sqrt{\lambda\alpha_2}}\,t\right\} \\ &= \exp\left\{\frac{1}{2}\,t^2 + \frac{1}{6}\frac{1}{\sqrt{\lambda}}\,\frac{\alpha_3}{\alpha_2^{3/2}}\,t^3 + \ldots\right\} \end{aligned} \tag{5.21}$$

Låter man $\lambda \to \infty$ i ovanstående uttryck, får man att

$$E[e^{tZ}] \to e^{t^2/2}$$

vilket är den momentgenererande funktionen för en $N(0,1)$-fördelning. Om Φ är fördelningsfunktionen för en $N(0,1)$-fördelning, skulle vi alltså ha att

$$P(S \le x) = P\left(Z \le \frac{x-\lambda\alpha_1}{\sqrt{\lambda\alpha_2}}\right) \approx \Phi\left(\frac{x-\lambda\alpha_1}{\sqrt{\lambda\alpha_2}}\right)$$

då λ är stort.

Tyvärr kan konvergensen mot normalfördelningen vara rätt långsam och approximationen därför inte särskilt skarp. Problemet är att fördelningen för Z ofta är skev och inte symmetrisk som normalfördelningen. Skevheten hos en fördelning definieras som tredje centralmomentet dividerat med variansen upphöjd till $3/2$. Skevheten hos fördelningen för Z är

$$\begin{aligned}\frac{E[(Z-E[Z])^3]}{(V[Z])^{3/2}} &= E[Z^3] = \frac{E[(S-E[S])^3]}{(V[S])^{3/2}} \\ &= \frac{\lambda\alpha_3}{(\lambda\alpha_2)^{3/2}} = \frac{\alpha_3}{\sqrt{\lambda}\,\alpha_2^{3/2}} > 0\end{aligned}$$

medan skevheten hos normalfördelningen är 0. Vi noterar att skevheten för Z dyker upp i koefficienten för t^3 i (5.21).

Vi ska nu presentera NP-approximationen och antar inte längre att antalet skador är Poisson-fördelat. Vi använder beteckningarna

$$\mu := E[S] \qquad \sigma^2 := V[S] \qquad \gamma := \frac{E[(S-\mu)^3]}{\sigma^3} \qquad Z := \frac{S-\mu}{\sigma}$$

γ är som vi såg ovan skevheten för såväl Z som S.

Idén bakom NP-approximationen (NP står för *normal power*) är att “vrida till”$N(0,1)$-fördelningen så att den passar bättre till fördelningen för Z. Med detta avses att man söker en transformation $p(Y)$ av en $N(0,1)$-variabel Y som har en fördelning som är mer lik fördelningen för Z. Om p har en invers p^{-1} skulle man då få approximationen

$$P(Z \le x) \approx P\big(p(Y) \le x\big) = P\big(Y \le p^{-1}(x)\big) = \Phi\big(p^{-1}(x)\big)$$

Om man som p tar ett andragradspolynom, kan man via diverse serieutvecklingar resonera sig fram till att p bör väljas som

$$p(x) = x + \frac{\gamma}{6}(x^2-1)$$

Andragradspolynomet p har en entydig invers för $x > -3/\gamma$:

$$p^{-1}(x) = \sqrt{9/\gamma^2 + 6x/\gamma + 1} - 3/\gamma$$

så approximationen blir

$$P(Z \le x) \approx \Phi\Big(\sqrt{9/\gamma^2 + 6x/\gamma + 1} - 3/\gamma\Big), \qquad x > -3/\gamma$$

eller

$$P(S \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right) \approx \Phi\left(\sqrt{\frac{9}{\gamma^2} + \frac{6(x-\mu)}{\sigma\gamma} + 1} - \frac{3}{\gamma}\right), \qquad x > \mu - \frac{3\sigma}{\gamma}$$

Detta är NP-approximationen av andra ordningen (syftar på att vi valde p som ett andragradspolynom), men den kan utvecklas vidare. För detta och en detaljerad härledning av NP-approximationen, se Buchwalder & Co.

Vid så kallad stop loss-återförsäkring är det av intresse att beräkna väntevärden av typen $E[h(S)]$, där

$$h(x) = \begin{cases} x - a, & \text{om } x > a \\ 0, & \text{om } x \leq a \end{cases} \tag{5.22}$$

Idén bakom NP-approximationen är ju att approximera högra svansen av fördelningen för Z med fördelningen för $p(Y)$. Väntevärdet $E[h(S)]$ beror bara på högra svansen av fördelningen för S, så om a är tillräckligt stort kan man hoppas att approximationen

$$E[h(S)] = E[h(\mu + \sigma Z)] \approx E[h(\mu + \sigma p(Y))]$$

skall fungera. Vi ska nu reda ut hur approximationen ser ut i detalj.

Till att börja med har vi att

$$E[h(S)] = E[h(\mu + \sigma Z)] = E[g(Z)]$$

där

$$\begin{aligned} g(z) := h(\mu + \sigma z) &= \begin{cases} \mu + \sigma z - a, & \text{om } \mu + \sigma z > a \\ 0, & \text{om } \mu + \sigma z \leq a \end{cases} \\ &= \begin{cases} \mu + \sigma z - a, & \text{om } z > (a - \mu)/\sigma \\ 0, & \text{om } z \leq (a - \mu)/\sigma \end{cases} \end{aligned}$$

Approximationen kan alltså skrivas som

$$E[h(S)] \approx E[g(p(Y))]$$

Vi noterar att

$$g(p(y)) = \begin{cases} \mu + \sigma p(y) - a, & \text{om } p(y) > (a - \mu)/\sigma \\ 0, & \text{annars} \end{cases}$$

Funktionen p antar sitt minimum i punkten $-3/\gamma$ och i denna punkt antar funktionen värdet

$$p(-3/\gamma) = -\frac{3}{\gamma} + \frac{\gamma}{6}\left(\frac{9}{\gamma^2} - 1\right) = -\frac{3}{2\gamma} - \frac{\gamma}{6}$$

Detta innebär att om

$$\frac{a - \mu}{\sigma} > -\frac{3}{2\gamma} - \frac{\gamma}{6}$$

dvs om

$$a > \mu - \frac{3\sigma}{2\gamma} - \frac{\sigma\gamma}{6}$$

så finns ett unikt tal $y_0 > -3/\gamma$ som löser ekvationen

$$p(y) = \frac{a-\mu}{\sigma}$$

Till höger om y_0 är p strikt växande.

Om vi låter φ beteckna $N(0,1)$-tätheten,

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

har vi alltså att

$$\begin{aligned} E[g(p(Y))] &= \int_{y_0}^{\infty} \big(\mu + \sigma p(y) - a\big)\, \varphi(y)\, dy \\ &= \left(\mu - \frac{\sigma\gamma}{6} - a\right) \int_{y_0}^{\infty} \varphi(y)\, dy + \sigma \int_{y_0}^{\infty} y\, \varphi(y)\, dy + \frac{\sigma\gamma}{6} \int_{y_0}^{\infty} y^2\, \varphi(y)\, dy \end{aligned}$$

Man har att

$$\begin{aligned} \int_{y_0}^{\infty} y\, \varphi(y)\, dy &= \int_{y_0}^{\infty} y\, \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy = \left[-\frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\right]_{y_0}^{\infty} \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-y_0^2/2} \end{aligned}$$

och att

$$\begin{aligned} &\int_{y_0}^{\infty} y^2\, \varphi(y)\, dy = \int_{y_0}^{\infty} y^2\, \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy \\ &= \left[-y\, \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\right]_{y_0}^{\infty} + \int_{y_0}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy \\ &= y_0\, \frac{1}{\sqrt{2\pi}}\, e^{-y_0^2/2} + 1 - \Phi(y_0) \end{aligned}$$

Man får alltså att

$$\begin{aligned} E\big[g(p(Y))\big] &= \left(\mu - \frac{\sigma\gamma}{6} - a\right)\big(1 - \Phi(y_0)\big) \\ &+ \sigma\, \frac{1}{\sqrt{2\pi}}\, e^{-y_0^2/2} + \frac{\sigma\gamma}{6}\left(y_0\, \frac{1}{\sqrt{2\pi}}\, e^{-y_0^2/2} + 1 - \Phi(y_0)\right) \\ &= (\mu - a)\big(1 - \Phi(y_0)\big) + \sigma\left(1 + \frac{\gamma y_0}{6}\right) \frac{1}{\sqrt{2\pi}}\, e^{-y_0^2/2} \end{aligned}$$

Vår approximation blir alltså

$$E\big[h(S)\big] \approx (\mu - a)\big(1 - \Phi(y_0)\big) + \sigma\left(1 + \frac{\gamma y_0}{6}\right) \frac{1}{\sqrt{2\pi}}\, e^{-y_0^2/2} \tag{5.23}$$

då

$$a > \mu - \frac{3\sigma}{2\gamma} - \frac{\sigma\gamma}{6}$$

och y_0 är den största roten till ekvationen

$$y + \frac{\gamma}{6}\,(y^2 - 1) = \frac{a-\mu}{\sigma}$$

Observationen att fördelningen för Z har en positiv skevhet leder till tanken att man kanske skulle använda en fördelning med samma egenskap att approximera med. En ofta omnämnd metod som bygger på denna idé är följande. Låt Y vara gammafördelad med parametrar α och 1, och sätt

$$W := \frac{Y - E[Y]}{\sqrt{V[Y]}} = \frac{Y-\alpha}{\sqrt{\alpha}}$$

Då har W väntevärde 0 och varians 1, precis som Z, oavsett värdet på parametern α. Idén är nu att välja α så att även skevheten för W, dvs

$$\gamma_W := \frac{E\big[(W - E[W])^3\big]}{(V[W])^{3/2}}$$

överensstämmer med skevheten för Z, dvs γ. Vi har att

$$\gamma_W = E[W^3] = E\left[\left(\frac{Y-\alpha}{\sqrt{\alpha}}\right)^3\right] = \frac{2\alpha}{\alpha^{3/2}} = \frac{2}{\sqrt{\alpha}}$$

så för att uppnå $\gamma_W = \gamma$, ska vi alltså välja $\alpha = 4/\gamma^2$.

Om G_α betecknar fördelningsfunktionen för en gamma fördelning med parametrar $\alpha = 4/\gamma^2$ och $\beta = 1$, skulle vi alltså få approximationen

$$\begin{aligned}
P(S \le x) &= P\left(Z \le \frac{x-\mu}{\sigma}\right) \approx P\left(W \le \frac{x-\mu}{\sigma}\right) \\
&= P\left(\frac{Y-\alpha}{\sqrt{\alpha}} \le \frac{x-\mu}{\sigma}\right) = P\left(Y \le \alpha + \sqrt{\alpha}\,\frac{x-\mu}{\sigma}\right) \\
&= G_\alpha\left(\alpha + \sqrt{\alpha}\,\frac{x-\mu}{\sigma}\right) = G_\alpha\left(\frac{4}{\gamma^2} + \frac{2}{\gamma}\,\frac{x-\mu}{\sigma}\right)
\end{aligned}$$

Med precis samma idé kan man härleda en approximation av väntevärdet (och högre moment) för $h(S)$, där h definierades i (5.22). Approximationen blir (Övning 5.13)

$$E\big[h(S)\big] \approx \sigma\sqrt{\alpha}\,\big(G_\alpha(\xi) - G_{\alpha+1}(\xi)\big) - (a-\mu)\big(1 - G_\alpha(\xi)\big), \tag{5.24}$$

där $\xi := \alpha + \sqrt{\alpha}(a-\mu)/\sigma$, och det förutsätts att $\xi > 0$.

Det har gjorts en del undersökningar av noggrannheten hos approximationerna i detta avsnitt, men det är inte så enkelt att den ena alltid är bättre än den andra. En tumregel som angivits för NP-approximationen är att noggrannheten är godtagbar om $\gamma < 1.2$ (se Daykin, Pentikäinen & Pesonen, 1994).

Generellt kan man säga att rekursionsmetoderna å ena sidan och approximationerna å den andra kompletterar varandra. För små portföljer kan approximationerna vara dåliga, men rekursionsberäkningarna gå snabbt att utföra. För stora portföljer kan de senare bli alltför tidskrävande, men approximationerna fungera bra.

Övning 5.13 *Härled (5.24).*

Övning 5.14 *Antag att N är Poissonfördelad med parameter λ och att X_1 är lognormalfördelad med parametrar $\mu = 0$ och $\sigma = 1$. Beräkna fördelningsfunktionen för S med NP-approximation respektive gamma-approximation, för $\lambda = 10$ och $\lambda = 20$, samt $\lambda = 100$. Jämför med vad du fick i Övning 5.12.*

Övning 5.15 *För en portfölj ville man beräkna hur stor premieinkomsten måste vara för att sannolikheten att den räcker till för att ersätta skadorna är minst 0.99. Antalet skador antas Poissonfördelat med parameter λ. Härled via NP-approximation ett uttryck för den nödvändiga premieinkomsten C i termer av λ, momenten α_k kring origo för skadefördelningen, samt 0.99-kvantilen q för $N(0,1)$-fördelningen (dvs $\Phi(q) = 0.99$).*

5.8 Övningscase: Sjuklöneförsäkring

Beskrivning

I proposition 1990/91:181 föreslog den dåvarande regeringen att fr o m 1.1.1992 ett sjuklönesystem skulle införas. Detta innebar att för varje sjukperiods första 14 dagar en anställd skulle erhålla sjuklön från arbetsgivaren i stället för sjukpenning från den allmänna försäkringskassan. I propositionen stod bl a:

Enligt förslaget ges arbetstagare en lagstadgad rätt att under de första 14 dagarna av varje sjukdomsfall (sjuklöneperioden) behålla viss del av den lön som han skulle ha fått om han fullgjort sina arbetsuppgifter. Andelen skall vara 75% de första tre dagarna med sjuklön och 90% återstående dagar av sjuklöneperioden.”

Sjuklön utbetalas endast för dagar då man skulle förvärvsarbetat. Om en anställd arbetar måndag–fredag och är sjuk från en torsdag till påföljande onsdag, får han/hon alltså 75% av lönen för torsdagen, fredagen och måndagen, samt 90% av lönen för tisdagen och onsdagen. Om den anställde är sjuk längre än 14 dagar, dvs 10 arbetsdagar, tar försäkringskassan över.

Skandia såg ett försäkringsbehov för mindre företag, där sjukfrånvaron p g a det relativt ringa antalet anställda skulle kunna variera avsevärt från år till år. Man beslöt att erbjuda en försäkring som trädde in om antalet sjukdagar under ett år översteg en viss gräns. Försäkringsbolaget skulle då betala kostnaderna för de överskjutande sjuklönedagarna. Man tänkte sig att gränsen skulle ligga i trakten av företagets förväntade totala sjuklönekostnad per år.

Den fråga vi söker ett svar på är hur stor Skandias förväntade årskostnad är för ett sådant kontrakt. Kostnaden kommer att bero på bl a antalet anställda, så man får räkna på lite olika alternativ.

Som underlag rekvirerade man från Riksförsäkringsverkets statistikenhet aktuella data rörande sjukfrånvaron i landet. Ändrade ersättningsregler från och med december 1987 innebar att endast data från de senaste åren, 1988 och 1989, var relevanta. Det framgick att det genomsnittliga antalet sjukfall per individ under 1988 och 1989 var 2.23 respektive 2.18.

I nedanstående tabell visas antalet avslutade sjukfall under 1988 respektive 1989 för vilka det utgick sjukpenning under k dagar, för $k = 1, 2, \ldots, 9$, samt $k \geq 10$. Sjukpenning utbetalades endast för dagar under vilka man skulle förvärvsarbetat.

I datamaterialet återfinns vidare hur många individer i olika åldersklasser som var sjuka k gånger under 1988 respektive 1989, för $k = 0, 1, \ldots, 10$, samt $k \geq 11$. Raderna i filen är åldersklasser. Uppgifterna avser antalet avslutade sjukfall.

Antal sjp-dgr	Antal fall	Antal sjp-dgr	Antal fall
1	2 222 945	1	2 332 215
2	2 064 541	2	2 066 031
3	1 428 585	3	1 391 967
4	1 101 532	4	1 023 987
5	1 397 779	5	1 347 307
6	248 457	6	221 689
7	335 404	7	316 271
8	138 297	8	129 957
9	129 530	9	116 161
≥ 10	1 364 498	≥ 10	1 297 095
	1988		1989

Förslag till ansats

Det första problemet är vilka faktorer premien ska bero av, dvs vilka tariffvariabler man ska välja. Antalet anställda och lönenivån på företaget är givna faktorer, men många andra är tänkbara: bransch, län, åldersstruktur, könssammansättning, etc. En viktig möjlighet är att låta företagets egen sjukfrånvarostatistik påverka premien, för att förhindra *moturval*, dvs att företag med hög sjukfrånvaro i högre grad tecknar en sådan här försäkring. Vi ska kort återkomma till detta nedan.

En tänkbar strategi är att starta med ett mindre antal variabler, som man tror är de viktigaste, för att sedan i mån av tid utvidga modellen. Enklast möjliga ansats skulle kunna se ut som följer.

Vi antar att alla anställda har samma lön ($=$ genomsnittslönen på företaget). I ett företag med stora löneskillnader kan detta behöva modifieras, till exempel genom att man delar in de anställda i löneklasser. Vi låter L beteckna full (100%) ersättning för en sjukdag och mäter alla kostnader i enheten L. Sålunda skulle premien kunna vara till exempel $120\ L$.

Vi låter J beteckna antalet anställda och M den årskostnad, mätt i L, över vilken Skandia går in och betalar återstoden. Vidare inför vi följande beteckningar:

$$
\begin{aligned}
N &= \text{totala antalet sjukfall under året} \\
X_i &= \text{kostnaden för sjukfall nr } i\text{, mätt i } L \\
Y_i &= \text{antal sjukdagar för sjukfall nr } i \\
S &= \text{totala sjuklönekostnaden under året, mätt i } L
\end{aligned}
$$

I definitionen av Y_i avser vi med "antalet sjukdagar" antalet dagar för vilka sjuklön utgår. Om vi sätter $C = 15/100$, får vi att

$$
\begin{aligned}
X_i &= \begin{cases} 0.75\,Y_i, & \text{om } Y_i \leq 3 \\ 0.75 \cdot 3 + 0.9\,(Y_i - 3), & \text{om } 4 \leq Y_i \leq 9 \\ 0.75 \cdot 3 + 0.9 \cdot 7, & \text{om } Y_i \geq 10 \end{cases} \\
&= \begin{cases} 5\,Y_i\,C, & \text{om } Y_i \leq 3 \\ (15 + 6(Y_i - 3))\,C, & \text{om } 4 \leq Y_i \leq 9 \\ 57\,C, & \text{om } Y_i \geq 10 \end{cases}
\end{aligned}
$$

Den stokastiska variabeln X_i antar något av värdena

$$5\,C,\ 10\,C,\ 15\,C,\ 21\,C,\ 27\,C,\ 33\,C,\ 39\,C,\ 45\,C,\ 51\,C,\ 57\,C$$

Man har alltså att

$$S = \sum_{i=1}^{N} X_i$$

Om vi antar att variablerna $N, X_1, X_2, \ldots$ är oberoende och att $X_1, X_2, \ldots$ är likafördelade, kan vi räkna på fördelningen för S med de metoder vi gått igenom i det här kapitlet (man måste dock fundera över hur pass realistiska dessa antaganden är). Eftersom variablerna $\{X_i\}$ antar något värde av typen kC, där k är ett heltal, gäller detsamma för S. Vi sätter nu

$$
\begin{aligned}
g_k &:= P(S = kC), \qquad k \geq 0 \\
f_k &:= P(X_1 = kC), \qquad k \geq 0
\end{aligned}
$$

Skandias förväntade kostnad blir

$$\begin{aligned}
E\big[\max(S-M,0)\big] &= \sum_{k:kC>M}(k\,C-M)\,g_k \\
&= \sum_k (k\,C-M)\,g_k - \sum_{k\le M/C}(k\,C-M)\,g_k \\
&= \sum_k k\,C\,g_k - M\sum_k g_k - \sum_{k\le M/C} k\,C\,g_k + M\sum_{k\le M/C} g_k \\
&= E[S] - M - C\sum_{k\le M/C} k\,g_k + M\sum_{k\le M/C} g_k
\end{aligned}$$

Vi vill få fram sannolikheterna $\{g_k\}$ och behöver för detta fördelningen för N och X_1.

När det gäller fördelningen för N, studerade vi ett likadant exempel i kapitel 4, fast för ett annat år, och fann att den negativa binomialfördelningen passade bra. Vi kan på precis samma sätt anpassa en negativ binomialfördelning till det datamaterial vi har här, och betecknar den anpassade fördelningen med NegBin(α,β). Om vi inte vet någonting om de anställda på företaget, är det rimligt att anta att antalet sjukfall per år för en enskild anställd har denna fördelning. Om man vidare antar att de anställda blir sjuka oberoende av varandra (detta kan förstås diskuteras: blir man ofta smittad på arbetsplatsen eller snarare av sjuka barn, på tunnelbanan, etc), får N fördelningen NegBin$(J\alpha,\beta)$.

När det gäller fördelningen för X_1, är det naturligt att först skatta fördelningen för Y_1 med den empiriska fördelningen. Man skattar alltså $P(Y_1=1)$ med $\hat{p}_1$, andelen sjukfall med en sjukdag, $P(Y_1=2)$ med $\hat{p}_2$, andelen sjukfall med två sjukdagar, osv, samt $P(Y_1\ge 10)$ med $\hat{p}_{10}$, andelen sjukfall med 10 eller fler sjukdagar. Därefter erhåller man sannolikheterna f_k som $\hat{f}_5=\hat{p}_1$,$\hat{f}_{10}=\hat{p}_2$, etc., för k lika med 5, 10, 15, 21, 27, 33, 39, 45, 51 eller 57. I övriga fall är f_k lika med noll.

Man kan nu beräkna sannolikheterna $\{g_k\}$ med hjälp av Panjers rekursionsformel. (Att S och X_i inte är heltalsvariabler är inget problem: S/C och X_i/C är ju heltalsvariabler och $g_k=P(S/C=k)$, $f_k=P(X_1/C=k)$.)

Det är tänkbart att låta försäkringstagaren själv välja M. Studera vad som händer med kvoten $E[\max(S-M,0)]/E[S]$ (dvs Skandias förväntade kostnad i förhållande till den förväntade totalkostnaden) då J ökar, om man väljer $M=E[S]$.

Det vi vill beräkna är av precis samma typ som i (5.22), så vi kan även tänka oss att använda till exempel NP-approximation. Studera hur bra approximationen är för olika värden på J, genom att jämföra med det resultat som Panjerrekursionen ger.

En intressant tanke är att modifiera fördelningen för N och X utifrån företagets egen sjukfrånvarostatistik, till exempel med hjälp av bayesianska metoder. Vi har inte tagit upp sådana i det här kompendiet, men den intresserade kan rota vidare i detta.

5.9 Övningscase: Återförsäkring av stormskador

Beskrivning

En storm kan orsaka skador på ett stort antal egendomar som alla är försäkrade hos samma bolag, och den totala kostnaden för försäkringsbolaget kan bli mycket stor. För att garantera bolagets fortsatta solvens och jämna ut resultatet mellan olika år tecknar man avtal med återförsäkringsbolag, som vanligtvis löper över ett år. Återförsäkringsbolagen åtar sig att mot en andel av premieintäkterna vara med och betala stormskadorna under året. Det är vanligt att många återförsäkringsbolag är inblandade och att en återförsäkringsmäklare sköter kontakten med dessa.

Ett vanligt sätt att utforma ett så kallat katastrofskydd är i form av XL-kontrakt (se avsnitt 3.5), dvs kostnaden delas mellan ett antal återförsäkringsbolag som täcker olika layrar. Kontraktet avser dock inte enskilda skador, utan den totala kostnaden som orsakas av en storm. Skulle kostnaden för en storm överstiga den översta layern, får cedenten själv betala det överstigande beloppet.

Eftersom det kan vara svårt att exakt säga när en storm börjar och slutar har kontrakten en tidsklausul. Kontraktet avser alla skador orsakade av kraftig vind, som inträffat under en viss tidsperiod, vanligtvis tre dygn (den så kallade 72-timmarsregeln). Cedenten bestämmer själv när perioden börjar.

Risken finns att återförsäkringsbolaget drabbas av stora förluster ifall flera kraftiga stormar inträffar under samma år, och för att skydda mot detta innehåller kontraktet ofta en klausul om så kallad *reinstatements*. Det innebär att om återförsäkringsbolaget täcker layern l xs C, så betalar man sammanlagt högst beloppet $(K+1)l$ för skador inträffade under kontraktstiden, där K är ett icke-negativt heltal. Man säger att kontraktet medger K reinstatements. Ofta kan det vara så att cedenten måste betala in ytterligare premier ifall återförsäkrarens ackumulerade kostnad X överstiger l, till exempel som följer. Sätt

$$Y_j := \begin{cases} 0, & \text{om } X \leq jl \\ (X - jl)/l, & \text{om } jl < X \leq (j+1)l \\ 1, & \text{om } X > (j+1)l \end{cases}$$

Cedenten betalar premien P innan kontraktet träder i kraft och därutöver ytterligare $(q_1 Y_1 + \cdots + q_K Y_K)P$ i tilläggspremie, där talen $q_1, \ldots, q_K$ är angivna i kontraktet. Några vanliga värden på $q_1, \ldots, q_k$ är

a) $q_1 = q_2 = \cdots = q_k = 0$ (s k fria reinstatements),

b) $q_1 = q_2 = \cdots = q_k = 1$,

c) $q_1 = 0$, $q_2 = \cdots = q_k = 1$.

Länsförsäkringsbolagen i Skåne försäkrar en mycket stor andel av lantbruksegendomarna mot stormskador. Man har tillgång till ett datamaterial bestående av samtliga utbetalda ersättningsbelopp för stormskador under åren 1982–1993. Under denna period har försäkringsbeståndet varit tämligen konstant.

Vid sammanställningen av datamaterialet har man utgått från 72-timmarsregeln. Om de sammanlagda kostnaderna för stormskador under en tredygnsperiod översteg en viss nivå, rubricerade man det som en storm.

Datamaterialet är uppdelat på Kristianstads län och Malmöhus län. En post i datamaterialet ser ut som “8804 35742”. Talet 8804 är numret på stormen enligt länsförsäkringsbolagens gemensamma numrering, de två första siffrorna anger vilket år stormen inträffade. Talet 35742 är det utbetalda ersättningsbeloppet, omräknat med hjälp av byggnadskostnadsindex för lantbruksbyggnader till 1994 års penningvärde.

Vi kommer att använda materialet till att studera katastrofskydd av den typ som beskrevs ovan. Exempel på intressanta frågeställningar är:

a) För en given layer och givna värden på K och $q_1, \ldots, q_K$, vad är ett rimligt värde på P (undantaget provisioner etc)?

b) Hur ska ett tillfredställande återförsäkringsskydd utformas?

Förslag till ansats

Sammanställ utifrån data en tabell över hur mycket varje storm kostade totalt, med de två länen sammanslagna. Tag också med vilket år stormen inträffade. Anpassa en POT-modell till dessa data. Antag till att börja med att man har ett katastrofskydd med excesspunkt (till exempel) 5 miljoner och obegränsad limit, samt inga reinstatements eller andra begränsningar. Beräkna med hjälp av Ströters rekursionsformel fördelningen för återförsäkringbolagens kostnad, samt deras förväntade kostnad. Pröva med steglängden $h = 100\,000$. När detta fungerar, försök modifiera det hela till fallet då limiten är begränsad (räkna på lite olika alternativ) och fundera sedan över hur man ska göra när man har reinstatements.

6 Den individuella modellen

6.1 Inledning

I den individuella modellen utgår man från hur skador drabbar de enskilda försäkringskontrakten, i stället för portföljen som helhet. I det här kapitlet ska vi studera hur man i detta fall med hjälp av rekursionsformler och approximationer kan räkna på fördelningen för bolagets totala skadekostnad för portföljen.

En mycket vanlig situation är att försäkringstagarna är indelade i (mer eller mindre) homogena grupper, för vilka man antar att alla i en grupp har (ungefär) samma fördelning för skadekostnaderna. Vi kommer att utgå från denna situation.

Vi antar att försäkringstagarna är indelade i a st grupper och låter n_i beteckna antalet försäkringstagare i grupp nr i. Vi betraktar ett visst tidsintervall, till exempel ett år eller ett kvartal, och låter X_{ij} vara (det totala) ersättningsbeloppet till försäkringstagare j i grupp i, till följd av skador inträffade under tidsintervallet. Vi antar att alla variablerna $\{X_{ij}\}$ är oberoende, samt att $X_{i1}, \ldots, X_{in_i}$ är likafördelade för $i = 1, \ldots, a$.

Vi inför också beteckningarna

$$
\begin{aligned}
p_i &:= P(X_{i1} > 0) \\
q_i &:= P(X_{i1} = 0) = 1 - p_i, \\
G_i(x) &:= P\big(X_{i1} \le x \bigm| X_{i1} > 0\big)
\end{aligned}
$$

Om F_i är (den obetingade) fördelningsfunktionen för X_{i1}, gäller att

$$
\begin{aligned}
G_i(x) &= \frac{P(X_{i1} > 0, X_{i1} \le x)}{P(X_{i1} > 0)} = \frac{P(0 < X_{i1} \le x)}{P(X_{i1} > 0)} \\
&= \frac{P(X_{i1} \le x) - P(X_{i1} = 0)}{P(X_{i1} > 0)} = \frac{F_i(x) - q_i}{p_i}
\end{aligned}
$$

Låt S_i beteckna försäkringsbolagets sammanlagda ersättningskostnader till försäkringstagarna i grupp i, dvs

$$S_i := X_{i1} + \cdots + X_{in_i}$$

Om vi låter M_i beteckna den momentgenererande funktionen för G_i, får vi att

$$\begin{aligned} E\big[e^{tS_i}\big] &= \big(E[e^{tX_{i1}}]\big)^{n_i} \\ &= \big(E\big[e^{tX_{i1}} \mid X_{i1} = 0\big]P(X_{i1} = 0) + E\big[e^{tX_{i1}} \mid X_{i1} > 0\big]P(X_{i1} > 0)\big)^{n_i} \\ &= \big(q_i + p_i\, M_i(t)\big)^{n_i} \end{aligned}$$

Det följer av (5.3) att S_i har samma fördelning som $Y_1 + \cdots + Y_N$, där $N, Y_1, Y_2, \ldots$ är oberoende, N är binomialfördelad med parametrar n_i och p_i, och alla Y_k:na har fördelningsfunktionen G_i. Alltså har S_i en sammansatt binomialfördelning och vi sammanfattar detta som $S_i \in \text{SaBi}(n_i, p_i, G_i)$. Vi noterar att fördelningen för S_i kan beräknas med hjälp av metoderna i föregående avsnitt.

Tyvärr har den sammansatta binomialfördelningen inte samma trevliga faltningsegenskap som den sammansatta Poissonfördelningen; $S_1 + \cdots + S_a$ har inte en sammansatt binomialfördelning, såvida inte $F_1 = F_2 = \cdots = F_a$ (men i så fall hade vi också bara en grupp).

I nästa avsnitt ska vi se hur man kan härleda rekursionsformler för beräkning av fördelningen för den sammanlagda ersättningskostnaden för portföljen, dvs fördelningen för

$$S := \sum_{i=1}^{a} S_i = \sum_{i=1}^{a} \sum_{j=1}^{n_i} X_{ij}$$

I det efterföljande avsnittet ska vi studera approximativa metoder för beräkning av fördelningen för S.

6.2 Rekursionsformler

I detta avsnitt antar vi att vi använder en monetär enhet av lämplig storleksordning, säg c kr, och avrundar alla ersättningsbelopp till närmaste c-tal kr. Vi räknar i denna enhet, så ersättningsbeloppet j svarar mot jc kr. Naturligtvis får man ett visst fel när man avrundar på detta sätt, och man får pröva sig fram till ett lämpligt värde på c. Ju mindre värde på c, desto större noggrannhet men också, som vi ska se, desto mer tidskrävande beräkningar.

Vi använder samma beteckningar som i föregående avsnitt: a är antalet grupper, n_i är antalet försäkringstagare i grupp i och

$$p_i = P(X_{ij} > 0) = P(X_{ij} \geq 1) = 1 - q_i$$

Vi inför också beteckningarna

$$\begin{aligned} f_{ik} &:= P\big(X_{ij} = k \mid X_{ij} \geq 1\big) \\ g_k &:= P(S = k) \end{aligned}$$

Vi antar att det för $i = 1, \ldots, a$ finns heltal m_i så att $f_{ik} = 0$ för $k > m_i$. Vi noterar att $g_k = 0$ för $k > n_1 m_1 + \cdots + n_a m_a$.

Följande sats innehåller en algoritm av de Pril (1989) för rekursiv beräkning av sannolikheterna $\{g_k\}$.

Sats 6.1 *Sannolikheterna $\{g_k\}$ kan beräknas enligt rekursionsformeln*

$$\begin{cases} g_0 = \prod_{i=1}^{a} q_i^{n_i} \\ g_{k+1} = \frac{1}{k+1} \sum_{i=1}^{a} n_i \sum_{j=0}^{k} w_{ij}\, g_{k-j} \end{cases}$$

där talen $\{w_{ik}\}$ kan beräknas rekursivt enligt

$$\begin{cases} w_{i0} = \frac{p_i}{q_i} f_{i1} \\ w_{ik} = \frac{p_i}{q_i} \left((k+1)\, f_{i,k+1} - \sum_{j=1}^{k} f_{ij}\, w_{i,k-j} \right), & k < m_i \\ w_{ik} = -\frac{p_i}{q_i} \sum_{j=1}^{m_i} f_{ij}\, w_{i,k-j}, \qquad k \geq m_i \end{cases}$$

Anmärkning: de Prils algoritm kan ses som en tvåstegsprocedur: Talen $\{w_{ik}\}$ kan beräknas rekursivt och givet dessa kan sannolikheterna $\{g_k\}$ beräknas rekursivt. Talen $\{w_{ik}\}$ är inte intressanta i sig, utan är ett slags hjälpvariabler i algoritmen. □

Bevis. Den sannolikhetsgenererande funktionen för S är

$$P(u) := \sum_{k=0}^{m} g_k\, u^k$$

Den sannolikhetsgenererande funktionen för X_{ij} ges av

$$\begin{aligned} E\big[u^{X_{ij}}\big] &= E\big[u^{X_{ij}} \mid X_{ij} = 0\big] P(X_{ij} = 0) + E\big[u^{X_{ij}} \mid X_{ij} \geq 1\big] P(X_{ij} \geq 1) \\ &= q_i + p_i\, Q_i(u) \end{aligned}$$

där

$$Q_i(u) := \sum_{j=1}^{m_i} f_{ij}\, u^j$$

är den sannolikhetsgenererande funktionen för fördelningen $\{f_{ij},\ 1 \leq j \leq m_i\}$. Vi får alltså att

$$P(u) = \prod_{i=1}^{a} \prod_{j=1}^{n_i} E\big[u^{X_{ij}}\big] = \prod_{i=1}^{a} \big(q_i + p_i\, Q_i(u)\big)^{n_i} \tag{6.1}$$

Sätter vi $u = 0$ i (6.1) får vi startvärdet för de Prils algoritm:

$$g_0 = \prod_{i=1}^{a} q_i^{n_i}$$

(att g_0 ser ut så inser man givetvis ändå).

Om man logaritmerar och deriverar båda sidor i (6.1), får man att

$$P'(u) = P(u) \sum_{i=1}^{a} n_i\, W_i(u), \tag{6.2}$$

där

$$W_i(u) := \frac{p_i\, Q_i'(u)}{q_i + p_i\, Q_i(u)}. \tag{6.3}$$

Vi söker nu en serieutveckling av $W_i(u)$:

$$W_i(u) = \sum_{j=0}^{\infty} w_{ij}\, u^j$$

Om vi sätter $u = 0$ i (6.3), får vi till att börja med att

$$w_{i0} = \frac{p_i}{q_i}\, f_{i1}$$

Vi har att

$$\big(q_i + p_i\, Q_i(u)\big) W_i(u) = p_i\, Q_i'(u)$$

Om vi deriverar detta uttryck k gånger på båda sidor, får vi med hjälp av Leibniz regel:

$$\begin{aligned}&\big(q_i + p_i\, Q_i(u)\big) W_i(u)^{(k)}(u) + \sum_{l=1}^{k} \binom{k}{l} p_i\, Q_i^{(l)}(u)\, W_i^{(k-l)}(u) \\ &= p_i\, Q_i^{(k+1)}(u)\end{aligned} \tag{6.4}$$

där $Q_i^{(l)}$ betecknar l:te derivatan av Q_i och motsvarande för W_i ($W_i^{(0)} := W_i$). Vi har att

$$Q_i^{(l)}(0) = f_{il}\, l! \qquad W_i^{(l)}(0) = w_{il}\, l!$$

så om vi sätter $u = 0$ i (6.4), får vi

$$q_i\, w_{ik}\, k! + \sum_{l=1}^{k} \binom{k}{l} p_i\, f_{il}\, l!\, w_{i,k-l}\, (k-l)! = p_i\, f_{i,k+1}\, (k+1)!$$

dvs

$$w_{ik} = \frac{p_i}{q_i}\left((k+1)\, f_{i,k+1} - \sum_{l=1}^{k} f_{il}\, w_{i,k-l}\right)$$

För $l > m_i$ gäller att $f_{il} = 0$, så för $k \geq m_i$ får vi att

$$w_{ik} = -\frac{p_i}{q_i}\sum_{l=1}^{m_i} f_{il}\, w_{i,k-l}$$

Om vi slutligen deriverar båda sidorna i (6.2) k gånger och sätter $u = 0$, får vi

$$g_{k+1}\,(k+1)! = \sum_{l=0}^{k}\binom{k}{l} g_l\, l! \sum_{i=1}^{a} n_i\, w_{i,k-l}\,(k-l)!$$

dvs

$$\begin{aligned} g_{k+1} &= \frac{1}{k+1}\sum_{i=1}^{a} n_i \sum_{l=0}^{k} g_l\, w_{i,k-l} \\ &= \frac{1}{k+1}\sum_{i=1}^{a} n_i \sum_{l=0}^{k} w_{il}\, g_{k-l} \qquad (6.5) \end{aligned}$$

Därmed är rekursionsformeln härledd. □

Beräkningar med algoritmen i Sats 6.1 blir lätt tidskrävande, och de Pril föreslår också en approximativ, men snabbare algoritm, som vi nu ska motivera.

Talen $\{w_{ik}\}$ går i själva verket att skriva upp explicit. Vi återvänder till (6.3):

$$W_i(u) = \frac{p_i\, Q_i(u)}{q_i + p_i\, Q_i(u)} = \frac{p_i}{q_i}\,\frac{Q_i'(u)}{1 - \frac{p_i}{q_i}\, Q_i(u)}$$

Eftersom

$$1 + s + s^2 + \cdots = \frac{1}{1-s}, \qquad |s| < 1$$

och $Q_i(0) = 0$, gäller i en omgivning till 0 (för $|u| < 1$ om $p_i < 1/2$) att

$$\frac{1}{1 - \frac{p_i}{q_i}\, Q_i(u)} = \sum_{j=0}^{\infty}(-1)^j\left(\frac{p_i}{q_i}\, Q_i(u)\right)^j$$

Vi får alltså att

$$
\begin{aligned}
W_i(u) &= \frac{p_i}{q_i}\,Q_i'(u)\sum_{j=0}^{\infty}(-1)^j\left(\frac{p_i}{q_i}\right)^j\big(Q_i(u)\big)^j \\
&= \sum_{j=0}^{\infty}(-1)^j\left(\frac{p_i}{q_i}\right)^{j+1}Q_i'(u)\big(Q_i(u)\big)^j \\
&= \sum_{j=0}^{\infty}(-1)^j\left(\frac{p_i}{q_i}\right)^{j+1}\frac{d}{du}\left\{\frac{(Q_i(u))^{j+1}}{j+1}\right\} \\
&= \sum_{l=1}^{\infty}\frac{(-1)^{l+1}}{l}\left(\frac{p_i}{q_i}\right)^l\frac{d}{du}\big\{(Q_i(u))^l\big\} \qquad (6.6)
\end{aligned}
$$

Det gäller ju att $(Q_i(u))^l$ är genererande funktionen för summan av l st oberoende variabler som alla har fördelningen $\{f_{ij},\ j\geq 0\}$, dvs

$$(Q_i(u))^l = \sum_{j=l}^{lm_i} f_{ij}^{l*}\,u^j$$

Deriverar vi båda sidor i (6.6) k gånger och sätter $u=0$, får vi

$$w_{ik}\,k! = \sum_{l=1}^{k+1}\frac{(-1)^{l+1}}{l}\left(\frac{p_i}{q_i}\right)^l f_{i,k+1}^{l*}\,(k+1)!$$

($f_{i,k+1}^{l*}=0$ om $l>k+1$), dvs

$$w_{ik} = \sum_{l=1}^{k+1}\frac{(-1)^{l+1}}{l}\left(\frac{p_i}{q_i}\right)^l(k+1)\,f_{i,k+1}^{l*}$$

Vi sätter in detta i (6.5) och byter summationsordning:

$$
\begin{aligned}
g_{k+1} &= \frac{1}{k+1}\sum_{i=1}^{a}n_i\sum_{j=0}^{k}w_{ij}\,g_{k-j} \\
&= \frac{1}{k+1}\sum_{i=1}^{a}n_i\sum_{j=0}^{k}\sum_{l=1}^{k+1}\frac{(-1)^{l+1}}{l}\left(\frac{p_i}{q_i}\right)^l(j+1)\,f_{i,j+1}^{l*}\,g_{k-j} \\
&= \frac{1}{k+1}\sum_{i=1}^{a}n_i\sum_{l=1}^{k+1}\frac{(-1)^{l+1}}{l}\left(\frac{p_i}{q_i}\right)^l\sum_{j=0}^{k}(j+1)\,f_{i,j+1}^{l*}\,g_{k-j} \\
&= \frac{1}{k+1}\sum_{i=1}^{a}n_i\sum_{l=1}^{k+1}\frac{(-1)^{l+1}}{l}\left(\frac{p_i}{q_i}\right)^l\sum_{j=1}^{k+1}j\,f_{ij}^{l*}\,g_{k-j+1} \\
&= \frac{1}{k+1}\sum_{i=1}^{a}n_i\sum_{l=1}^{k+1}\frac{(-1)^{l+1}}{l}\left(\frac{p_i}{q_i}\right)^l\sum_{j=l}^{\min(k+1,m_il)}j\,f_{ij}^{l*}\,g_{k-j+1}
\end{aligned}
$$

I den sista likheten har vi använt att $f_{ij}^{l*} = 0$ om $j < l$ eller $j > m_i l$.

Idén är nu att om p_i:na är små (eller åtminstone mindre än $1/2$), blir termen $(p_i/q_i)^l$ snabbt liten när l växer. de Prils approximativa algoritm av ordningen r går ut på att approximera g_k med $g_k^{(r)}$, där $\{g_k^{(r)},\ k \geq 0\}$ beräknas enligt rekursionsformeln

$$g_{k+1}^{(r)} = \frac{1}{k+1} \sum_{i=1}^{a} n_i \sum_{l=1}^{\min(k+1,r)} \frac{(-1)^{l+1}}{l} \left(\frac{p_i}{q_i}\right)^l \sum_{j=l}^{\min(k+1,m_i l)} j\, f_{ij}^{l*}\, g_{k-j+1}^{(r)}$$

Oftast (se Dhaene & Vandebroek, 1995) räcker det med ett litet r, säg $r = 5$, för att erhålla praktiskt taget samma resultat som med algoritmen i Sats 6.1. Sannolikheterna $\{f_{ij}^{k*}\}$ kan beräknas via upprepad användning av faltningsformeln.

I Dhaene & Vandebroek (1995) finns mer att läsa om rekursionsformler för den individuella modellen.

Övning 6.1 *Med beteckningarna i ovanstående avsnitt, antag att man har följande situation:*

$$a = 4$$

$n_1 = 400$	$n_2 = 600$	$n_3 = 800$	$n_4 = 1000$
$p_1 = 0.01$	$p_2 = 0.02$	$p_3 = 0.005$	$p_4 = 0.008$
$f_{11} = 0.1$	$f_{12} = 0.5$	$f_{13} = 0.1$	$f_{14} = 0.3$
$f_{21} = 0.3$	$f_{22} = 0.2$	$f_{23} = 0.2$	$f_{24} = 0.3$
$f_{31} = 0.5$	$f_{32} = 0.4$	$f_{33} = 0.1$	

$$f_{41} = f_{42} = f_{43} = f_{44} = f_{45} = 0.1 \qquad f_{46} = 0.5.$$

Beräkna $P(S \leq k)$ för $k = 0, 1, \ldots, 150$, dels med hjälp av de Prils exakta algoritm, dels med hjälp av de Prils approximativa algoritm.

Övning 6.2 *Låt $X_1, \ldots, X_n$ vara oberoende och likafördelade stokastiska variabler som antar värden i mängden $\{0, 1, 2, \ldots\}$. Sätt $p_k := P(X_1 = k)$ och antag att $p_0 > 0$. Låt $Y := X_1 + \cdots + X_n$ och sätt $h_k := P(Y = k)$. Visa med hjälp av sannolikhetsgenererande funktioner rekursionsformeln*

$$h_k = \frac{1}{p_0} \sum_{j=1}^{k} \left(\frac{n+1}{k} j - 1\right) p_j\, h_{k-j}, \qquad k \geq 1$$

med startvärdet $h_0 = p_0^n$.

Övning 6.3 *Antag att* $Y_1, \ldots, Y_m$ *är oberoende och att* Y_i *är binomialfördelad med parametrar* n_i *och* p_i*. Sätt* $Z := Y_1 + \cdots + Y_m$ *och* $g_j := P(Z = j)$*. Härled rekursionsformeln*

$$g_k = \sum_{j=1}^{k} \frac{1}{k}\, w_j\, g_{k-j}$$

där

$$w_j := (-1)^{j+1} \sum_{i=1}^{m} n_i \left(\frac{p_i}{q_i}\right)^j$$

6.3 Approximationer

Vi ska nu ta upp hur fördelningen för S kan approximeras med en sammansatt binomialfördelning. Vi inför beteckningarna

$$\begin{aligned} n &:= n_1 + \cdots + n_a, \\ p &:= \frac{1}{n} \sum_{i=1}^{a} n_i\, p_i, \\ G(x) &:= \frac{1}{np} \sum_{i=1}^{a} n_i\, p_i\, G_i(x) \end{aligned}$$

(Om alla G_i:na är diskreta eller alla absolutkontinuerliga, gäller detsamma för G och sannolikhetsfunktionen/tätheten fås genom samma linjärkombination som för fördelningsfunktionerna.) Idén är att approximera fördelningen för S med en SaBi(n, p, G)-fördelning och vi ska nu motivera detta genom att studera de momentgenererande funktionerna.

Den momentgenererande funktionen för S är

$$M_S(t) = \prod_{i=1}^{a} \big(1 + p_i(M_i(t) - 1)\big)^{n_i}$$

där M_i är den momentgenererande funktionen för G_i. Den momentgenererande funktionen för SaBi(n, p, G)-fördelningen är

$$\tilde{M}(t) := \big(1 + p(M(t) - 1)\big)^n$$

där

$$M(t) := \frac{1}{np} \sum_{i=1}^{a} n_i p_i\, M_i(t)$$

är den momentgenererande funktionen för fördelningen G.

Vi ska använda en olikhet som är en generalisering av olikheten mellan de geometriska och aritmetiska medelvärdena. Om $x_1, \dots, x_m$ är icke-negativa reella tal och $\lambda_1, \dots, \lambda_m$ är tal mellan 0 och 1 sådana att $\lambda_1 + \cdots + \lambda_m = 1$, gäller att

$$\prod_{i=1}^{m} x_i^{\lambda_i} \le \sum_{i=1}^{m} \lambda_i \, x_i, \tag{6.7}$$

med likhet om och endast om alla x_i:na är lika (se Övning 6.5). Om man tar $m = a$ och

$$x_i = 1 + p_i(M_i(t) - 1), \qquad \lambda_i = n_i/n$$

ger olikheten att

$$\prod_{i=1}^{a} \bigl(1 + p_i(M_i(t) - 1)\bigr)^{n_i/n} \le \sum_{i=1}^{a} \frac{n_i}{n} \bigl(1 + p_i(M_i(t) - 1)\bigr)$$

vilket efter en enkel omskrivning blir

$$M_S(t) \le \tilde{M}(t) \tag{6.8}$$

(Om $M(t)$ existerar för $t \le u > 0$, gäller olikheten (6.8) för $t \le v$, där $v > 0$.) Approximationen är alltså exakt om alla grupperna är lika, och man kan hoppas att den är god så länge grupperna inte är alltför olika, eller (vilket är vanligt) om alla p_i:na är nära noll.

En klassisk metod är att approximera fördelningen för S med en SaPo(np, G)-fördelning. Studier har dock visat att SaBi-approximationen är bättre (se till exempel Kuon, Radtke & Reich, 1993) och man kan förstå detta om man studerar de momentgenererande funktionerna. Låt $\bar{M}$ vara den momentgenererande funktionen för en SaPo(np, G)-fördelning, dvs

$$\bar{M}(t) = \exp\bigl\{np(M(t) - 1)\bigr\}$$

En annan olikhet från analysen säger att

$$\log(1 + x) \le x, \qquad \text{om } x > -1, \tag{6.9}$$

med likhet om och endast om $x = 0$ (se Övning 6.5). Denna ger att

$$\begin{aligned} \tilde{M}(t) &= \bigl(1 + p(M(t) - 1)\bigr)^n \\ &= \exp\bigl\{n \log(1 + p(M(t) - 1))\bigr\} \\ &\le \exp\bigl\{np(M(t) - 1)\bigr\} = \bar{M}(t) \end{aligned} \tag{6.10}$$

(Om $M(t)$ existerar för $t \le u$, där $u > 0$, gäller olikheten för $t \le v$, där $v > 0$.) Detta ger tillsammans med olikheten (6.8) att $\tilde{M}(t)$ ligger närmare $M_S(t)$ än $\bar{M}(t)$ gör, varför man kan förvänta sig att SaBi-approximationen är bättre. Vi noterar också att olikheten i (6.10) närmar sig likhet då p blir litet, men (nästan) likhet i (6.8) gäller under svagare villkor. Eftersom Panjers/Ströters rekursionsformler gäller både för SaBi- och SaPo-fördelningarna, är SaBi-approximationen att föredra. En jämförelse av fördelningarnas moment återfinns i Övning 6.6.

Man kan tänka sig följande stegvisa förfarande när man vill räkna ut fördelningen för S. Först prövar man med de “exakta” algoritmerna i föregående avsnitt. Om det är för tidskrävande, tar man till SaBi-approximationen ovan och använder Panjers alternativt Ströters rekursionsformel. Skulle detta också gå för tungt, approximerar man i sin tur SaBi-fördelningen med någon av metoderna i avsnitt 5.7. Låt oss räkna ut de μ, σ och γ som skall användas i dessa approximationsformler.

Den kumulantgenererande funktionen för SaBi-fördelningen är

$$\Psi(t) = n \log\bigl(1 + p(M(t) - 1)\bigr)$$

De tre första derivatorna av denna är

$$\begin{aligned}
\Psi'(t) &= \frac{npM'(t)}{1 + p(M(t) - 1)}, \\
\Psi''(t) &= \frac{npM''(t)}{1 + p(M(t) - 1)} - n\left(\frac{p\,M'(t)}{1 + p(M(t) - 1)}\right)^2 \\
\Psi'''(t) &= \frac{np\,M'''(t)}{1 + p(M(t) - 1)} - \frac{3np^2\,M'(t)\,M''(t)}{(1 + p(M(t) - 1))^2} + 2n\left(\frac{p\,M'(t)}{1 + p(M(t) - 1)}\right)^3
\end{aligned}$$

Om α_1, α_2 och α_3 betecknar de tre första momenten kring origo för fördelningen G, får vi alltså att

$$\begin{aligned}
\mu &= \Psi'(0) = np\alpha_1 \\
\sigma^2 &= \Psi''(0) = np\alpha_2 - np^2\alpha_1^2 \\
\gamma &= \frac{\Psi'''(0)}{\sigma^3} = \frac{\alpha_3 - 3p\alpha_1\alpha_2 + 2p^2\alpha_1^3}{\sqrt{np}(\alpha_2 - p\alpha_1^2)^{3/2}}
\end{aligned} \tag{6.11}$$

Approximationerna i avsnitt 5.7 blir normalt bättre ju mindre γ är och vi noterar att γ är omvänt proportionell mot roten ur np, som är det förväntade antalet skador för hela portföljen. Ju större portfölj, desto bättre approximation (som väntat). Vi noterar slutligen att α_j är en linjärkombination av de j:te momenten α_{ji} kring origo för fördelningarna G_i:

$$\alpha_j = \frac{1}{np}\sum_{i=1}^{a} n_i\,p_i\,\alpha_{ji}$$

Den approximerande SaBi-fördelningen har samma väntevärde som S, men momenten av ordning 2 och högre är ej desamma (se Övning 6.6 nedan). Detta leder till tanken att man i NP- (gamma-)approximationen skulle använda variansen respektive skevheten för S i stället för σ^2 och γ i (6.11).

Det gäller ju att $S = S_1 + \cdots + S_a$, där $S_i = X_{i1} + \cdots + X_{in_i}$. Eftersom $S_i \in$ SaBi(n_i, p_i, G_i) och S_i:na är oberoende, fås att

$$\begin{aligned} V[S] &= \sum_{i=1}^{a} n_i\, p_i\, (\alpha_{2i} - p_i\, \alpha_{1i}^2) \\ E\big[(S - E[S])^3\big] &= \sum_{i=1}^{a} n_i\, p_i\, (\alpha_{3i} - 3\, p_i\, \alpha_{1i}\, \alpha_{2i} + 2\, p_i^2\, \alpha_{1i}^3) \end{aligned}$$

Övning 6.4 *a) För situationen i Övning 6.1, beräkna approximativt $P(S \le k)$, $k = 0, 1, \ldots, 150$ med hjälp av SaBi- respektive SaPo-approximationerna. Jämför med resultatet i Övning 6.1.*

b) Beräkna approximativt $P(S \le k)$ via NP-approximation av SaBi-approximationen i a). Gör även samma sak då σ^2 och γ i (6.11) ersatts med variansen respektive väntevärdet för S. Jämför med tidigare resultat.

Övning 6.5 *a) Visa olikheten*

$$x^\lambda\, y^{1-\lambda} \le \lambda\, x + (1 - \lambda)\, y, \qquad x \ge 0, \quad y \ge 0, \quad 0 < \lambda < 1$$

genom att studera $\varphi(t)$ och $\varphi'(t)$ för $t \ge 0$, där

$$\varphi(t) := 1 - \lambda + \lambda t - t^\lambda$$

b) Använd resultatet i a) och induktion för att visa olikheten (6.7).

c) Visa olikheten (6.9).

Övning 6.6 *Låt S vara som i avsnitt 6.1, låt S' vara SaBi(n, p, G)-fördelad och låt S'' vara SaPo(np, G)- fördelad, där n, p och G definierades i inledningen till detta avsnitt. Visa att*

$$E[S] = E[S'] = E[S'']$$

och att

$$\begin{aligned} V[S'] &= V[S] + \frac{1}{2n} \sum_{i \neq j} n_i \, n_j \, (p_i \, \alpha_{1i} - p_j \, \alpha_{1j})^2, \\ V[S''] &= V[S'] + \frac{1}{n} \left(\sum_{i=1}^{a} n_i \, p_i \, \alpha_{1i} \right)^2 \end{aligned}$$

(Dvs speciellt att $V[S] \leq V[S'] \leq V[S'']$.)

6.4 Övningscase: Avtalsgruppsjukförsäkring

Beskrivning

Detta övningscase, som handlar om så kallad avtalsgruppsjukförsäkring (AGS), har ställts till förfogande av AMF och Folksam.

En arbetstagare som blir långvarigt eller kroniskt sjuk kan av försäkringskassan beviljas sjukbidrag eller förtidspension. Försäkringskassan betalar då varje månad ut en ersättning som baserar sig på arbetstagarens inkomst vid insjuknandet. Ersättningen är densamma vid sjukbidrag som vid förtidspensionering, skillnaden är att sjukbidraget årligen omprövas.

Som ett komplement till sjukbidraget/förtidspensionen tecknar de företag som omfattas av kollektivavtal mellan arbetsmarknadens parter trygghetsförsäkringar för sina anställda, till exempel avtalsgruppsjukförsäkringen AGS. En anställd som insjuknar och beviljas sjukbidrag/förtidspension får utöver ersättningen från försäkringskassan en tilläggsersättning från försäkringsbolaget. I det fall det här handlar om kan han/hon på så sätt behålla ca 80–85% av sin inkomst.

Man kan också beviljas sjukbidrag/förtidspension till 25%, 50% eller 75%. Tilläggsersättningen reduceras då i motsvarande grad. Fram t o m 1993 fanns även möjligheten att beviljas 67% förtidspension.

Avtalsgruppsjukförsäkringen omfattar också ett tillägg till sjukpenningen.

Det belopp som försäkringsbolaget betalar ut varje månad baserar sig alltså på inkomsten vid insjuknandetidpunkten. Ersättningen blev skattepliktig från och med 1991, och man höjde då ersättningsbeloppen för att kompensera för detta. Av det skälet finns det två tabeller: en för fallet då den försäkrade insjuknat före den 1 januari 1991 och en då insjuknandedatumet är 1 januari 1991 eller senare. Försäkringsvillkoren har ej ändrats efter 1991.

Om sjukperioden varar mer än två år, utgår ett värdesäkringstillägg, ifall det allmänna prisläget höjts sedan sjukperiodens början. Värdesäkringstillägget räknas ut med hjälp av basbeloppen (se tabell nedan). Tillägget är den summa som behövs för att uppnå samma procentuella ökning för ersättningen som för basbeloppet under sjukperioden. Dock lämnas ej ett större tilläggsbelopp än vad man får om man räknar med 4% ökning per år.

Ett problem när det gäller AGS är att det kan dröja lång tid, ibland upp till tio år, från insjuknandet till beviljandet av sjukbidrag/förtidspension. Det kan också inträffa att en anställd inte känner till att han/hon är berättigad till ersättning från försäkringsbolaget, utan hör av sig först flera år efter kassans beviljande. Även i ett sådant fall utgår ersättning för hela sjukdomstiden.

Det datamaterial vi ska räkna på gäller ett visst fackförbund och består av dels statistik över utbetalningar gjorda av försäkringsbolaget under åren 1990–95, dels statistik över antalet medlemmar i fackförbundet under samma period.

Statistiken över utbetalningarna består av 2219 poster, var och en innehållande 14 värden. Här är ett exempel på hur en post kan se ut, tillsammans med vad de olika värdena kallas. Termerna förklaras närmare nedan.

Värde	Exempel	Benämning
1	61	Insjuknandeålder
2	0	Kön
3	900605	Insjuknandedatum
4	920301	Beviljandedatum
5	9010	Registreringsmånad
6	1	Ersättningsgrad
7	1492	Månadsbelopp
8	3	Slutmarkering
9	2856	Belopp 1
10	37300	Belopp 2
11	0	Belopp 3
12	6204	Belopp 4
13	0	Belopp 5
14	940228	T o m-datum

Här följer nu närmare förklaringar av benämningarna ovan.

Insjuknandeålder: Den försäkrades ålder vid insjuknandet (fyllda år).

Kön: Den försäkrades kön; 0 för man, 1 för kvinna.

Insjuknandedatum: Det datum då den försäkrade anmälde sig sjuk till försäkringskassan.

Beviljandedatum: Datum för försäkringskassans beviljande av sjukbidrag/förtidspension.

Registreringsmånad: Den månad då sjukfallet registrerades av Folksam.

Ersättningsgrad: Som tidigare nämnts, kan man även erhålla halvt sjukbidrag / halv förtidspension, etc. Följande beteckningar används: hel ersättning - 1, 50% ersättning - 2, 67% ersättning - 3, 25% ersättning - 4, samt 75% ersättning - 5.

Månadsbelopp: Grundbeloppet per månad före värdesäkring; bestäms enligt tabell A eller B längst bak i försäkringsvillkoren (bilaga 1).

Slutmarkering: Siffra som anger varför ersättning upphört, om den gjort det. Om ersättning fortfarande utgår står det 0. För övrigt förekommer följande varianter i datamaterialet: Den försäkrade är friskförklarad - 1, den försäkrade har avlidit - 2, den försäkrade har fyllt 65 - 3.

Belopp 1: För tid då den försäkrade erhåller sjukpenning, får han/hon en tilläggsersättning från försäkringsbolaget. Belopp 1 anger hur mycket tilläggsersättning av denna typ som betalts ut till den försäkrade exklusive värdesäkring.

Belopp 2: Det totala utbetalade beloppet, minus värdesäkringen, utgörande tilläggsersättning till sjukbidrag/förtidspension.

Belopp 3: Totalt utbetalt värdesäkringsbelopp för sjukpenningstillägget.

Belopp 4: Totalt utbetalt värdesäkringsbelopp för tillägget till sjukbidrag/förtidspension.

Belopp 5: Försäkringsbolaget lämnar också ersättning för journal- och utredningskostnader, och kostnader för läkarintyg. Belopp 5 är den totala ersättning av denna typ som betalts ut till den försäkrade.

T o m-datum: Datum för sista utbetalningen om ersättning upphört, annars senaste utbetalningsdatum.

I datamaterialet ingår också uppgifter om antalet medlemmar i förbundet under åren 1989-

1995.

Din uppgift är att beräkna fördelningen för försäkringsbolagets totala kostnad för de försäkringsfall som inträffar under ett år. Med försäkringsfall avses att en medlem blir så sjuk att det leder till sjukbidrag eller förtidspension. I det sammanhangen är det av intresse att känna till att så gott som alla som beviljas sjukbidrag sedan förtidspensioneras, samt att det stora flertalet som beviljas förtidspension lever tills de fyller 65 år.

1985	21 800
1986	23 300
1987	24 100
1988	25 800
1989	27 900
1990	29 700
1991	32 200
1992	33 700
1993	34 400
1994	35 200

Tabell 6.1: Basbeloppets utveckling under åren 1985-94.

Förslag till ansats

Börja med att bortse från säkerhetstillägget. I mån av tid kan du studera även detta.

Bortse från kön, dvs slå ihop män och kvinnor i antalsstatistiken.

Låt varje åldersklass utgöra en grupp och låt n_i beteckna antalet medlemmar i åldersgrupp i 1996. Tag med åldersklasserna 17 till 65. Skatta $n_{17}, \ldots, n_{65}$ genom att studera tabellen över antalet medlemmar genom åren. Använd ingen statistisk metod, utan endast "ögonmått".

Räkna på fördelningen för den totala kostnaden med hjälp av den individuella modellen. Vi behöver då vidare p_i, sannolikheten för en individ i åldersklass i att bli så sjuk att det leder till förtidspension. Ett sätt att angripa detta är som följer.

Antag att man väljer en person på måfå ur hela populationen. Låt

$$\begin{aligned} W &= \begin{cases} 1, & \text{om personen blir sjuk under året} \\ 0, & \text{annars} \end{cases} \\ R &= \text{personens åldersklass} \end{aligned}$$

Låt vidare

$$\bar{p} := P(W = 1) \qquad \bar{q}_i := P\big(R = i \;\big|\; W = 1\,\big)$$

dvs $\bar{p}$ är sannolikheten att en godtyckligt vald medlem blir sjuk och $\bar{q}_i$ sannolikheten att en insjuknad tillhör klass i. Vidare har vi att

$$P(R = i) = n_i/(n_{17} + \cdots + n_{65})$$

Detta ger att

$$\begin{aligned} p_i &= P\big(W = 1 \;\big|\; R = i\big) = \frac{P(R = i|W = 1)\,P(W = 1)}{P(R = i)} \\ &= \bar{p}\,\bar{q}_i\,\frac{n_{17} + \cdots + n_{65}}{n_i} \end{aligned}$$

Sannolikheten $\bar{p}$ skattas med andelen som blir sjuka under ett år. Problemet är att det kan dröja flera år innan alla försäkringsfall är kända. Om du känner till reservsättningsmetoder, använd till exempel Chain Ladder-metoden och Craigheads kurvanpassningsmetod för att uppskatta det slutliga antalet försäkringsfall för de olika åren. Varning: År 1995 är ofullständigt registrerat. Tag sedan medelvärdet av andelen sjuka över åren för att skatta $\bar{p}$. Eftersom denna skattning kan vara ganska osäker, är det intressant att studera hur fördelningen för det sammanlagda skadebeloppet varierar då $\bar{p}$ varierar. Om du inte är bekant med reservsättningsmetoder eller har ont om tid, använd $\bar{p} = 0.005$.

Sannolikheten $\bar{q}_i$ skattas helt enkelt med andelen av de sjuka som tillhör åldersklass i. Det är lämpligt att gruppera åldersklasserna får att få tillräckligt med observationer. Man kan till exempel anta att sannolikheten är densamma i åldrarna 17–25, 26–35, 36–45 och 46–55. Över 55 är insjuknandefrekvensen tillräckligt hög för att skatta $\bar{q}_i$ utan gruppering. Använd ditt eget omdöme för att göra gruppindelningen. Så få grupper som möjligt är förstås önskvärt.

Använd den enklaste beräkningsmetoden: NP- eller gammaapproximation. (För att använda en rekursionsmetod krävs att fördelningen för de individuella skadekostnaderna tas fram. Detta kan vara rätt krångligt, men den intresserade kan fundera vidare själv.) Observera att om man först approximerar med en SaBi-fördelning, behöver man inte skatta $\bar{q}_i$:na. Det p

som man använder i SaBi-fördelningen blir nämligen

$$
\begin{aligned}
p &= \frac{1}{n_{17}+\cdots+n_{65}} \sum_{i=17}^{65} n_i\, p_i \\
&= \frac{1}{n_{17}+\cdots+n_{65}} \sum_{i=17}^{65} n_i\, \bar{p}\, \bar{q}_i\, \frac{n_{17}+\cdots+n_{65}}{n_i} \\
&= \bar{p} \sum_{i=1}^{65} q_i = \bar{p}
\end{aligned}
$$

dvs p kan skattas direkt med skattningen av $\bar{p}$.

Det vi behöver för att kunna räkna på fördelningen med NP- eller gammametoden är momenten för X_i, $i = 17, \ldots, 65$, där X_i är kostnaden för en godtycklig insjuknad i åldersklass i. Vi betraktar en godtycklig insjuknad i åldersklass i och inför följande beteckningar:

Y_i	$=$	personens ersättningsbelopp per år
Z_i	$=$	andelen förtidspension
V_i	$=$	antalet år från beviljandet av förtidspension till det att förtidspension beviljas

Om vi antar att alla som beviljas förtidspension lever tills de är 65 (borde inte ge upphov till något större fel) och att andelen förtidspension inte ändras (har vi inte data för att undersöka), har vi att

$$X_i = Y_i\, Z_i\, (65 - i - V_i)$$

Om vi vidare antar att Y_i, Z_i och V_i är oberoende (diskutera om detta är rimligt), gäller att

$$\alpha_{ki} = E\big[X_i^k\big] = E\big[Y_i^k\big] E\big[Z_i^k\big] E\big[(65 - i - V_i)^k\big]$$

Momenten för Y_i och Z_i skattas lämpligen med de empiriska momenten. Även här är det lämpligt att slå ihop åldersklasser till grupper för att få tillräckligt med observationer.

När det gäller fördelningen för V_i krävs en viss försiktighet. Använd det tidigaste året för att skatta sannolikheterna $r_j^i := P(V_i = j)$ som relativa frekvenser (gruppera även här). Antag att det inte kommer att beviljas några fler förtidspensioner för sjukfall inträffade detta år. Det stämmer inte riktigt, men vi har inte data för att studera problemet. Den som vill kan dock försöka “skatta” svansen i fördelningen. Utifrån skattningarna $\hat{r}_j^i$ skattas sedan $E[(65 - i - V_i)^k]$ med

$$\widehat{E}[(65 - i - V_i)^k] = \sum_j (65 - i - j)^k \hat{r}_j^i$$

När alla momenten för variablerna Y_i, Z_i och $65-i-V_i$ är skattade, multipliceras dessa ihop för att ge skattningarna av momenten för X_i. Så är allt redo för beräkningen med NP- eller gammaapproximation.

Den här beskrivna ansatsen är den grövsta tänkbara. För entusiasten finns många alternativ att jobba vidare på.

7 Lösningar

Kapitel 3: Modeller för skadebeloppen

3.1. Det finns ett enkelt samband mellan α_k och α_{k-1}. För $k < \gamma$ har man att

$$\begin{aligned}
\alpha_k &= \int_0^\infty x^k\, f(x)\, dx = \int_0^\infty x^k\, \frac{\gamma\, \alpha^\gamma}{(\alpha+x)^{\gamma+1}}\, dx \\
&= \left[-x^k\, \frac{\alpha^\gamma}{(\alpha+x)^\gamma}\right]_0^\infty + \int_0^\infty k\, x^{k-1}\, \frac{\alpha^\gamma}{(\alpha+x)^\gamma}\, dx \\
&= 0 + \int_0^\infty k\, x^{k-1}\, \frac{\alpha+x}{\gamma}\, \frac{\gamma\, \alpha^\gamma}{(\alpha+x)^{\gamma+1}}\, dx \\
&= \frac{k}{\gamma}\left(\alpha \int_0^\infty x^{k-1}\, f(x)\, dx + \int_0^\infty x^k\, f(x)\, dx\right) \\
&= \frac{k}{\gamma}\big(\alpha\alpha_{k-1} + \alpha_k\big)
\end{aligned}$$

dvs

$$\alpha_k = \frac{k\alpha}{\gamma - k}\, \alpha_{k-1}.$$

För $k = 3$ får man

$$\alpha_3 = \frac{3\alpha}{\gamma-3}\, \alpha_2 = \frac{6\alpha^3}{(\gamma-1)(\gamma-2)(\gamma-3)}.$$

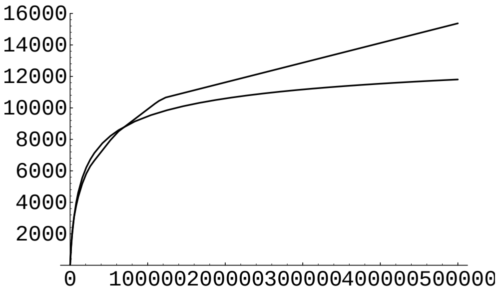
3.2. ML-skattningarna baserade på hela datamaterialet blev

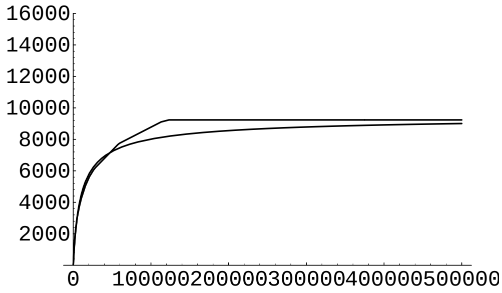
$$\hat{\alpha} = 4640.15, \qquad \hat{\gamma} = 1.29413.$$

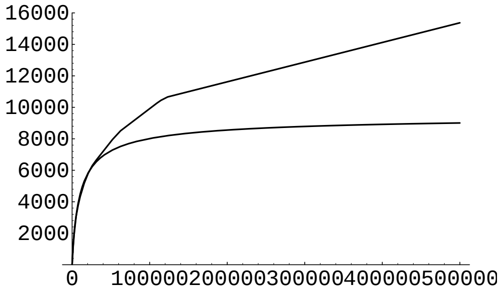
ML-skattningarna baserade på datamaterialet med den största skadan borttagen blev

$$\hat{\alpha} = 6037.39, \qquad \hat{\gamma} = 1.62909.$$

De översta två figurerna nedan visar anpassningen av de skattade väntevärdesfunktionerna tillsammans med de empiriska väntevärdesfunktionerna i de två fallen. Den sista figuren visar väntevärdesfunktionen som anpassades till datamaterialet utan den största skadan jämförd med den empiriska fördelningsfunktionen för datamaterialet där den största skadan ingår. För detta datamaterial tycks Paretomodellen underskatta potentialen för stora skador. Jämför motsvarande figur för Burrfördelningen i övning 3.5.







3.3. Fördelningsfunktionen är

$$F(x) = 1 - e^{-cx^\tau}, \qquad x \geq 0,$$

så tätheten blir

$$f(x) = c\tau\, x^{\tau-1}\, e^{-cx^\tau}, \qquad x > 0.$$

Momenten kring origo är

$$\begin{aligned}
\alpha_k &= \int_0^\infty x^k\, c\tau\, x^{\tau-1}\, e^{-cx^\tau}\, dx \\
&= (\text{Sätt } y = cx^\tau,\ dy = c\tau\, x^{\tau-1}\, dx) \\
&= \int_0^\infty \left(\frac{y}{c}\right)^{k/\tau} e^{-y}\, dy = \frac{\Gamma(1+k/\tau)}{c^{k/\tau}}.
\end{aligned}$$

Momentmetodens skattningar erhålls som lösning till ekvationerna

$$m_1 = \frac{\Gamma(1+1/\tau)}{c^{1/\tau}}, \qquad m_2 = \frac{\Gamma(1+2/\tau)}{c^{2/\tau}}.$$

Enligt den första ekvationen gäller att

$$c^{1/\tau} = \frac{\Gamma(1+1/\tau)}{m_1}, \qquad \text{så att} \qquad c^{2/\tau} = \left(\frac{\Gamma(1+1/\tau)}{m_1}\right)^2,$$

och enligt den andra gäller att

$$c^{2/\tau} = \frac{\Gamma(1+2/\tau)}{m_2}.$$

Man får alltså att

$$\frac{\Gamma(1+2/\tau)}{m_2} = \frac{(\Gamma(1+1/\tau))^2}{m_1^2},$$

dvs
$$m_1^2\,\Gamma(1+2/\tau) = m_2\,(\Gamma(1+1/\tau))^2.$$
Denna ekvation löses numeriskt och ger $\hat{\tau}$. Insättning i den första ekvationen ovan ger skattningen av c:
$$\hat{c} = \left(\frac{\Gamma(1+1/\hat{\tau})}{m_1}\right)^{\hat{\tau}}.$$

3.4. a) Tätheten för X blir
$$\begin{aligned}
f_X(x) &= \int_0^\infty f_X(x|Y=y)\,f_Y(y)\,dy \\
&= \int_0^\infty \frac{y^\alpha}{\Gamma(\alpha)}\,x^{\alpha-1}\,e^{-yx}\,\frac{\lambda^\gamma}{\Gamma(\gamma)}\,y^{\gamma-1}\,e^{-\lambda y}\,dy \\
&= \frac{\lambda^\gamma\,x^{\alpha-1}}{\Gamma(\alpha)\,\Gamma(\gamma)}\int_0^\infty y^{\alpha+\gamma-1}\,e^{-(\lambda+x)y}\,dy \\
&= \frac{\lambda^\gamma\,x^{\alpha-1}}{\Gamma(\alpha)\,\Gamma(\gamma)}\int_0^\infty \left(\frac{u}{\lambda+x}\right)^{\alpha+\gamma-1} e^{-u}\,\frac{1}{\lambda+x}\,du \\
&= \frac{\lambda^\gamma\,x^{\alpha-1}}{\Gamma(\alpha)\,\Gamma(\gamma)\,(\lambda+x)^{\alpha+\gamma}}\int_0^\infty u^{\alpha+\gamma-1}\,e^{-u}\,du \\
&= \frac{\Gamma(\alpha+\gamma)\,\lambda^\gamma\,x^{\alpha-1}}{\Gamma(\alpha)\,\Gamma(\gamma)\,(\lambda+x)^{\alpha+\gamma}}.
\end{aligned}$$

b) Fördelningsfunktionen för X kan skrivas som
$$\begin{aligned}
F(x) &= \int_0^x f(y)\,dy \\
&= \frac{1}{B(\alpha,\gamma)}\int_0^x \left(\frac{y}{\lambda+y}\right)^{\alpha-1}\left(\frac{\lambda}{\lambda+y}\right)^{\gamma-1}\frac{\lambda}{(\lambda+y)^2}\,dy \\
&\quad \{\text{Substituera } u = y/(\lambda+y) \text{ i integralen}\} \\
&= \frac{1}{B(\alpha,\gamma)}\int_0^{x/(\lambda+x)} u^{\alpha-1}\,(1-u)^{\gamma-1}\,du \\
&= G_{\alpha,\gamma}\!\left(\frac{x}{\lambda+x}\right).
\end{aligned}$$
Använd att $\Gamma(x+1) = x\,\Gamma(x)$ för att erhålla att
$$\begin{aligned}
\int_0^x y\,f(y)\,dy &= \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\,\Gamma(\gamma)}\int_0^x \left(\frac{y}{\lambda+y}\right)^{\alpha}\left(\frac{\lambda}{\lambda+y}\right)^{\gamma-2}\frac{\lambda^2}{(\lambda+y)^2}\,dy \\
&= \frac{\Gamma(\alpha+\gamma)\,\alpha\,\lambda}{\Gamma(\alpha+1)\,(\gamma-1)\,\Gamma(\gamma-1)}\int_0^{x/(\lambda+x)} u^{\alpha}\,(1-u)^{\gamma-2}\,du \\
&= \frac{\alpha\,\lambda}{\gamma-1}\,G_{\alpha+1,\gamma-1}\!\left(\frac{x}{\lambda+x}\right).
\end{aligned}$$

Väntevärdesfunktionen blir alltså

$$
\begin{aligned}
L(x) &= \int_0^x y\, f(y)\, dy + x(1 - F(x)) \\
&= \frac{\alpha\,\lambda}{\gamma - 1}\, G_{\alpha+1,\gamma-1}\left(\frac{x}{\lambda + x}\right) + x\left(1 - G_{\alpha,\gamma}\left(\frac{x}{\lambda + x}\right)\right).
\end{aligned}
$$

c) Momenten kring origo blir, via (3.2),

$$
\begin{aligned}
\alpha_k &= \int_0^\infty x^k\, f(x)\, dx \\
&= \frac{1}{B(\alpha,\gamma)} \int_0^\infty x^{\alpha+k-1}\, \frac{\lambda^\gamma}{(\lambda + x)^{\alpha+\gamma}}\, dx \\
&= \frac{\lambda^k}{B(\alpha,\gamma)} \int_0^\infty \left(\frac{x}{\lambda + x}\right)^{\alpha+k-1} \left(\frac{\lambda}{\lambda + x}\right)^{\gamma-k-1} \frac{\lambda}{(\lambda + x)^2}\, dx \\
&= \frac{\lambda^k}{B(\alpha,\gamma)} \int_0^1 u^{\alpha+k-1}\, (1 - u)^{\gamma-k-1}\, du \\
&= \frac{\lambda^k\, B(\alpha + k, \gamma - k)}{B(\alpha,\gamma)} = \frac{\lambda^k\, \Gamma(\alpha + k)\, \Gamma(\gamma - k)}{\Gamma(\alpha)\, \Gamma(\gamma)} \\
&= \lambda^k \prod_{i=0}^{k-1} (\alpha + i) \Bigg/ \prod_{i=1}^{k} (\gamma - i).
\end{aligned}
$$

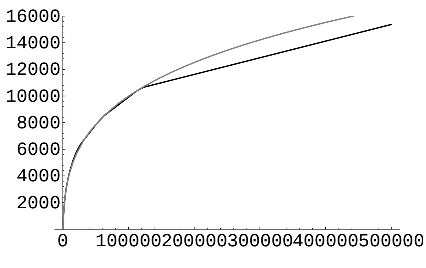
3.5. ML-skattningarna baserade på hela datamaterialet blev

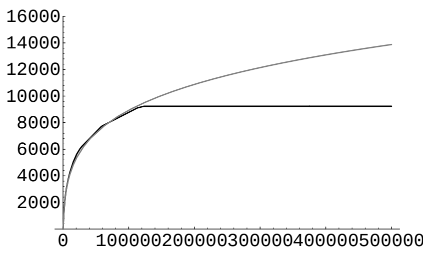
$$\hat{\alpha} = 4.81 \cdot 10^{10}, \qquad \hat{\gamma} = 0.212, \qquad \hat{\tau} = 3.53.$$

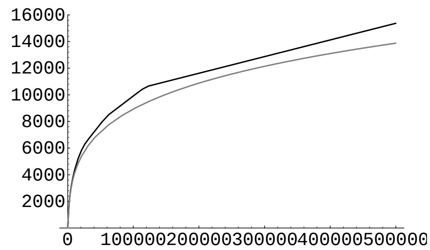
ML-skattningarna baserade på datamaterialet med den största skadan borttagen blev

$$\hat{\alpha} = 1.74 \cdot 10^{10}, \qquad \hat{\gamma} = 0.241, \qquad \hat{\tau} = 3.36.$$

De tre figurerna nedan motsvarar figurerna i lösningen till övning 3.2. Vi noterar att Burrfördelningen passar bra i området där data ligger, och en tolkning av den sista figuren är att modellen indikerar möjligheten för stora skador (som också inträffade). Burrfördelningen verkar mindre känslig för inverkan av stora skador i datamaterialet än Pareto, åtminstone i detta exempel.







3.6. Man har att

$$
\begin{aligned}
\int_0^x y\,f(y)\,dy &= \int_0^x y\,\frac{1}{\sqrt{2\pi}\sigma y}\exp\left\{-\frac{(\log y-\mu)^2}{2\sigma^2}\right\}dy \\
&= \{\text{ Sätt } u=\log y,\, du=y^{-1}\,dy\ \} \\
&= \int_{-\infty}^{\log x} e^u\,\frac{1}{\sqrt{2\pi}\sigma}\,\exp\left\{-\frac{(u-\mu)^2}{2\sigma^2}\right\}du \\
&= \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi}\sigma}\,\exp\left\{-\frac{(u-\mu)^2-2\sigma^2 u}{2\sigma^2}\right\}du \\
&= \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi}\sigma}\,\exp\left\{-\frac{(u-(\mu+\sigma^2))^2-2\mu\sigma^2-\sigma^4}{2\sigma^2}\right\}du \\
&= e^{\mu+\sigma^2/2}\int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi}\sigma}\,\exp\left\{-\frac{(u-(\mu+\sigma^2))^2}{2\sigma^2}\right\}du \\
&= \{\text{ Sätt } v=(u-(\mu+\sigma^2))/\sigma,\, dv=\sigma^{-1}\,du\ \} \\
&= e^{\mu+\sigma^2/2}\int_{-\infty}^{(\log x-\mu-\sigma^2)/\sigma} \frac{1}{\sqrt{2\pi}}e^{-v^2/2}\,dv \\
&= e^{\mu+\sigma^2/2}\,\Phi\left(\frac{\log x-\mu-\sigma^2}{\sigma}\right)
\end{aligned}
$$

Man får alltså att

$$
\begin{aligned}
L(x) &= \int_0^x y\,f(y)\,dy + x(1-F(x)) \\
&= e^{\mu+\sigma^2/2}\,\Phi\left(\frac{\log x-\mu-\sigma^2}{\sigma}\right) + x\left(1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right).
\end{aligned}
$$

3.7. a) Man har att

$$
\begin{aligned}
M(t) &= \int_0^\infty e^{tx}\,\frac{\mu}{\sqrt{2\pi\beta x^3}}\,\exp\left\{-\frac{(x-\mu)^2}{2\beta x}\right\}dx \\
&= \int_0^\infty \frac{\mu}{\sqrt{2\pi\beta x^3}}\exp\left\{-\frac{(1-2\beta t)x^2-2\mu x+\mu^2}{2\beta x}\right\}dx.
\end{aligned}
$$

Om man sätter $\beta_1 := \beta/(1-2\beta t)$, $\mu_1 := \mu/\sqrt{1-2\beta t}$, kan man skriva M som

$$\begin{aligned} M(t) &= \exp\left\{\frac{\mu}{\beta} - \frac{\mu_1}{\beta_1}\right\} \int_0^\infty \frac{\mu_1}{\sqrt{2\pi\beta_1 x^3}} \exp\left\{-\frac{(x-\mu_1)^2}{2\beta_1 x}\right\} dx \\ &= \exp\left\{\frac{\mu}{\beta}\big(1-\sqrt{1-2\beta t}\big)\right\}; \end{aligned}$$

integralen blir 1 eftersom integranden är täthetsfunktionen för en invers normalfördelning med parametrar μ_1 och β_1. Den momentgenererande funktionen existerar för $t \leq 1/2\beta$.

b) De tre första derivatorna av den kumulantgenererande funktionen $\Psi(s) = \log M(s)$ är

$$\begin{aligned} \Psi'(t) &= \mu(1-2\beta t)^{-1/2}, \\ \Psi''(t) &= \mu\beta(1-2\beta t)^{-3/2}, \\ \Psi'''(t) &= 3\mu\beta^2(1-2\beta t)^{-5/2}. \end{aligned}$$

De tre första kumulanterna blir alltså

$$\kappa_1 = \Psi'(0) = \mu, \qquad \kappa_2 = \Psi''(0) = \mu\beta, \qquad \kappa_3 = \Psi'''(0) = 3\mu\beta^2.$$

Från (1.1) får vi att

$$\begin{aligned} \alpha_1 &= \mu_1 = \mu, \\ \alpha_2 &= \mu_2 + \alpha_1^2 = \mu\beta + \mu^2, \\ \alpha_3 &= \mu_3 + 3\alpha_2\alpha_1 - 2\alpha_1^2 = 3\mu\beta + 3(\mu\beta+\mu^2)\mu - 2\mu_1^3 \\ &= 3\mu\beta^2 + 3\mu^2\beta + \mu_1^3. \end{aligned}$$

c) Om $X_1, \ldots, X_n$ är oberoende och X_k har en invers normalfördelning med parametrar μ_k och β, är den momentgenererande funktionen för $X_1 + \cdots + X_k$

$$\prod_{k=1}^n \exp\left\{\frac{\mu_k}{\beta}\Big(1-\sqrt{1-2\beta t}\Big)\right\} = \exp\left\{\frac{\mu_1+\cdots+\mu_n}{\beta}\Big(1-\sqrt{1-2\beta t}\Big)\right\},$$

dvs $X_1 + \cdots + X_n$ har en invers normalfördelning med parametrar $\mu_1 + \cdots + \mu_n$ och β.

3.8. För materialet i tabell 3.1 fick jag ML-skattningarna till

$$\hat{\mu} = 8.807, \qquad \hat{\sigma}^2 = 1.299,$$

och för materialet i övning 3.2,

$$\hat{\mu} = 8.215, \qquad \hat{\sigma} = 1.820.$$

3.9. a) Likelihoodfunktionen är

$$L(x) = \prod_{k=1}^n f(x_k) = \mu^n\,(2\pi\beta)^{-n/2}\left(\prod_{k=1}^n x_k\right)^{3/2} \exp\left\{-\sum_{k=1}^n \frac{(x_k-\mu)^2}{2\beta x_k}\right\}.$$

Loglikelihoodfunktionen blir

$$\begin{aligned} \ell(x) &= n\log\mu - \frac{n}{2}\log 2\pi - \frac{n}{2}\log\beta - \frac{3}{2}\sum_{k=1}^n \log x_k - \frac{1}{2\beta}\sum_{k=1}^n\left(x_k - 2\mu + \frac{\mu^2}{x_k}\right) \\ &= n\log\mu - \frac{n}{2}\log 2\pi - \frac{n}{2}\log\beta - \frac{3}{2}\sum_{k=1}^n \log x_k - \frac{1}{2\beta}\sum_{k=1}^n x_k + \frac{n\mu}{\beta} - \frac{\mu^2}{2\beta}\sum_{k=1}^n \frac{1}{x_k}. \end{aligned}$$

De partiella derivatorna av loglikelihoodfunktionen är

$$\begin{cases} \frac{\partial \ell}{\partial \mu} = \frac{n}{\mu} + \frac{n}{\beta} - \frac{\mu}{\beta} \sum_{k=1}^{n} \frac{1}{x_k}, \\ \frac{\partial \ell}{\partial \beta} = -\frac{n}{2\beta} + \frac{1}{2\beta^2} \sum_{k=1}^{n} x_k - \frac{n\mu}{\beta^2} + \frac{\mu^2}{2\beta^2} \sum_{k=1}^{n} \frac{1}{x_k}. \end{cases}$$

ML-ekvationerna blir, efter en enkel omskrivning,

$$\begin{cases} \frac{\beta}{\mu^2} + \frac{1}{\mu} = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{x_k}, \\ -\frac{\beta}{2} + \frac{1}{2} \frac{1}{n} \sum_{k=1}^{n} x_k - \mu + \frac{\mu^2}{2} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{x_k} = 0. \end{cases}$$

Löser man ut β i den första ekvationen får man

$$\beta = \mu^2 \, \tilde{x} - \mu. \tag{7.1}$$

Sätter man in detta i den andra ekvationen får man

$$-\frac{\mu^2}{2} \, \tilde{x} + \frac{\mu}{2} + \frac{1}{2}\bar{x} - \mu + \frac{\mu^2}{2} \, \tilde{x} = 0,$$

dvs ML-skattningen av μ är

$$\hat{\mu} = \bar{x}.$$

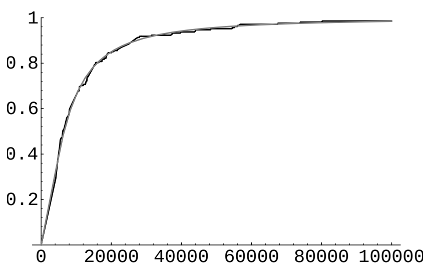
Sätter man in detta i (7.1), får man ML-skattningen av β:

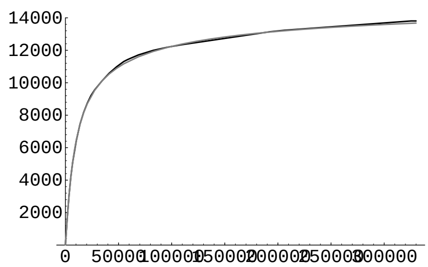
$$\hat{\beta} = \bar{x}^2 \, \tilde{x} - \bar{x} = \bar{x}(\bar{x} \, \tilde{x} - 1).$$

b) ML-skattningarna av μ och β blir

$$\hat{\mu} = 13806.2, \qquad \hat{\beta} = 40398.1.$$

Nedanstående figurer visar anpassningen av fördelnings- och väntevärdesfunktionen med dessa parametervärden, jämförda med empiriska fördelnings- och väntevärdesfunktionerna.





Kapitel 4: Modeller för antalet skador

4.1. Som ett exempel att jämföra med återges för $\lambda = 10$ en tabell över fördelningsfunktionen F:s värden i punkterna $0, 1, 2, \ldots, 20$ tillsammans med Anscombes approximation A i samma punkter. (Vid beräkningen av F utnyttjades det rekursiva sambandet $p_k = \lambda \, p_{k-1}/k$, $F(k) = F(k-1) + p_k$.)

k	$F(k)$	$A(k)$
0	0.0000454	0.000033997
1	0.0004994	0.00045377
2	0.0027694	0.0026716
3	0.010336	0.010206
4	0.029253	0.029151
5	0.067086	0.067069
6	0.13014	0.13020
7	0.22022	0.22028
8	0.33282	0.33277
9	0.45793	0.45773
10	0.58304	0.58270
11	0.69678	0.69639
12	0.79156	0.79121
13	0.86446	0.86422
14	0.91654	0.91642
15	0.95126	0.95123
16	0.97296	0.97298
17	0.98572	0.98577
18	0.99281	0.99286
19	0.99655	0.99659
20	0.99841	0.99844

4.2. Moment- och ML-skattningarna av λ är $\hat{\lambda} = 0.610$. Följande tabell visar relativa frekvensen $\hat{p}_k$ av k hästsparkar tillsammans med $p_k(\hat{\lambda}) = \hat{\lambda}\, e^{-\hat{\lambda}}/k!$:

k	$\hat{p}_k$	$p_k(\hat{\lambda})$
0	0.545	0.543
1	0.325	0.331
2	0.110	0.101
3	0.015	0.020
4	0.005	0.003
5	0	0.000

En god anpassning, således.

4.3. a) Den negativa binomialfördelningen uppkommer när vi tar gammafördelningen som blandningsfördelning:

$$u(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, \lambda^{\alpha-1}\, e^{-\beta\lambda}.$$

Vi noterar att

$$\lim_{\lambda \downarrow 0} \lambda u(\lambda) = \lim_{\lambda \downarrow 0} \frac{\beta^\alpha}{\Gamma(\alpha)}\, \lambda^\alpha\, e^{-\beta\lambda} = 0,$$

eftersom $\alpha > 0$.

Vidare har vi att

$$\begin{aligned}\frac{d}{d\lambda}\log u(\lambda) &= \frac{d}{d\lambda}\big\{\alpha\log\beta - \log\Gamma(\alpha) + (\alpha-1)\log\lambda - \beta\lambda\big\}\\ &= \frac{\alpha-1}{\lambda} - \beta = \frac{\alpha-1-\beta\lambda}{\lambda}.\end{aligned}$$

Förutsättningarna för resultatet är alltså uppfyllt, med

$$\begin{gathered}c_0 = \alpha-1, \quad c_1 = -\beta, \quad c_2 = 0,\\ d_1 = 1, \quad d_2 = 0\end{gathered}$$

Man får alltså rekursionsformeln

$$0 = -(\beta+1)(n-1)\,p_{n-1} + (\alpha-1+n-1)\,p_{n-2},$$

eller, om vi byter ut $n-1$ mot n,

$$p_n = \frac{\alpha+n-1}{(\beta+1)n}\,p_{n-1}.$$

b) Vi har blandningstätheten

$$u(\lambda) = \frac{\mu}{\sqrt{2\pi\beta\lambda^3}}\exp\bigg\{-\frac{(\lambda-\mu)^2}{2\beta\lambda}\bigg\}.$$

Man har att

$$\lim_{\lambda\to 0}\lambda u(\lambda) = \lim_{\lambda\to 0}\frac{\mu}{\sqrt{2\pi\beta\lambda}}\exp\left\{-\frac{1}{2\beta}\left(\lambda - 2\mu + \frac{\mu^2}{\lambda}\right)\right\} = 0.$$

Vidare,

$$\begin{aligned}&\frac{\partial}{\partial\lambda}\quad\log u(\lambda)\\ &\quad= \frac{\partial}{\partial\lambda}\bigg\{\log\mu - \frac{1}{2}\log 2\pi\beta - \frac{3}{2}\log\lambda - \frac{1}{2\beta}\bigg(\lambda - 2\mu + \frac{\mu^2}{\lambda}\bigg)\bigg\}\\ &\quad= -\frac{3}{2\lambda} - \frac{1}{2\beta} + \frac{\mu^2}{2\beta\lambda^2} = \frac{\mu^2 - 3\beta\lambda - \lambda^2}{2\beta\lambda^2}.\end{aligned}$$

Vi ser att förutsättningarna är uppfyllda med

$$\begin{gathered}c_0 = \mu^2, \quad c_1 = -3\beta, \quad c_2 = -1,\\ d_1 = 0, \quad d_2 = 2\beta\end{gathered}$$

Man får alltså rekursionsformeln

$$\begin{aligned}&(2\beta+1)n(n-1)\,p_n\\ &\quad= (4\beta - 3\beta - 0 + 2\beta(n-2))(n-1)\,p_{n-1} + (\mu^2 + 0\cdot(n-1))\,p_{n-2},\end{aligned}$$

dvs

$$p_n = \frac{\beta(2n-3)}{(2\beta+1)n}\,p_n + \frac{\mu^2}{(2\beta+1)n(n-1)}\,p_{n-2}, \qquad n \geq 2.$$

Enligt (4.4) och övning 3.7 gäller att den sannolikhetsgenererande funktionen för IGP-fördelningen är

$$P(s) = \exp\bigg\{\frac{\mu}{\beta}\Big(1 - \sqrt{1+2\beta(1-s)}\Big)\bigg\},$$

så att

$$P'(s) = P(s)\,\mu(1+2\beta(1-s))^{-1/2}.$$

Man får alltså att

$$\begin{aligned}
p_0 &= P(0) = \exp\left\{\frac{\mu}{\beta}\Big(1-\sqrt{1+2\beta}\Big)\right\},\\
p_1 &= P'(0) = p_0\,\frac{\mu}{\sqrt{1+2\beta}}.
\end{aligned}$$

4.4. Man har att

$$u(\lambda) = \frac{\gamma\,\alpha^\gamma}{(\alpha+\lambda)^{\gamma+1}},$$

så att

$$\log u(\lambda) = \log\gamma + \gamma\log\alpha - (\gamma+1)\log(\alpha+\lambda).$$

Det följer att

$$\frac{u'(\lambda)}{u(\lambda)} = \frac{d}{dx}\log u(\lambda) = -\frac{\gamma+1}{\alpha+\lambda},$$

dvs

$$(\alpha+\lambda)\,u'(\lambda) = -(\gamma+1)\,u(\lambda).$$

Partiell integration ger

$$\begin{aligned}
p_n &= \int_0^\infty \frac{\lambda^n}{n!}\,e^{-\lambda}\,u(\lambda)\,d\lambda\\
&= \left[-\frac{\lambda^n}{n!}\,e^{-\lambda}\,u(\lambda)\right]_0^\infty + \int_0^\infty \frac{n\,\lambda^{n-1}}{n!}\,e^{-\lambda}\,u(\lambda)\,d\lambda + \int_0^\infty \frac{\lambda^n}{n!}\,e^{-\lambda}\,u'(\lambda)\,d\lambda\\
&= p_{n-1} + \int_0^\infty \frac{\lambda^n}{n!}\,e^{-\lambda}\,u'(\lambda)\,d\lambda\\
&= p_{n-1} + \int_0^\infty \frac{\lambda^{n-1}}{n!}\,e^{-\lambda}\,\lambda\,u'(\lambda)\,d\lambda
\end{aligned}$$

Detta ger dels

$$\alpha\,p_n = \alpha\,p_{n-1} + \int_0^\infty \frac{\lambda^n}{n!}\,e^{-\lambda}\,\alpha\,u'(\lambda)\,d\lambda \tag{7.2}$$

och dels

$$p_{n+1} = p_n + \int_0^\infty \frac{\lambda^n}{(n+1)!}\,e^{-\lambda}\,\lambda\,u'(\lambda)\,d\lambda,$$

dvs

$$(n+1)\,p_{n+1} = (n+1)\,p_n + \int_0^\infty \frac{\lambda^n}{n!}\,e^{-\lambda}\,\lambda\,u'(\lambda)\,d\lambda. \tag{7.3}$$

Adderar man (7.2) och (7.3) får man

$$\begin{aligned}
(n+1)\,p_{n+1} + \alpha\,p_n &= (n+1)\,p_n + \alpha\,p_{n-1} + \int_0^\infty \frac{\lambda^n}{n!}\,e^{-\lambda}\,(\alpha+\lambda)\,u'(\lambda)\,d\lambda\\
&= (n+1)\,p_n + \alpha\,p_{n-1} - \int_0^\infty \frac{\lambda^n}{n!}\,e^{-\lambda}\,(\gamma+1)\,u(\lambda)\,d\lambda\\
&= (n+1)\,p_n + \alpha\,p_{n-1} - (\gamma+1)\,p_n.
\end{aligned}$$

Sätter vi $m = n + 1$, får vi

$$m\, p_m = (m - \gamma - 1 - \alpha)\, p_{m-1} + \alpha\, p_{m-2},$$

dvs

$$p_m = \left(1 - \frac{\gamma + 1 + \alpha}{m}\right) p_{m-1} + \frac{\alpha}{m}\, p_{m-2}.$$

4.5. Momentmetodens skattningar blev

$$\hat{\alpha}_1 = 0.9956, \qquad \hat{\beta}_1 = 6.418.$$

Metoden med anpassning till $\bar{x}$ och $\hat{p}_0$ gav skattningarna

$$\hat{\alpha}_2 = 1.0532, \qquad \hat{\beta}_2 = 6.789.$$

ML-skattningarna blev

$$\hat{\alpha}_3 = 1.0327, \qquad \hat{\beta}_3 = 6.656.$$

En beräkning av ML-skattningen av parametern λ i Poissonfördelningen utifrån datamaterialet gav $\hat{\lambda} = 0.1551$. Sätt

$$p_k(\alpha, \beta) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\, k!} \left(\frac{\beta}{\beta + 1}\right)^{\alpha} \left(\frac{1}{\beta + 1}\right)^{k}, \qquad p_k(\lambda) = \frac{\lambda^k}{k!}\, e^{-\lambda}.$$

Nedanstående tabell sammanfattar anpassningen till den empiriska fördelningen ($\hat{p}_k$ betecknar som vanligt relativa frekvensen av observationen k i datamaterialet).

k	$\hat{p}_k$	$p_k(\hat{\alpha}_1, \hat{\beta}_1)$	$p_k(\hat{\alpha}_2, \hat{\beta}_2)$	$p_k(\hat{\alpha}_3, \hat{\beta}_3)$	$p_k(\hat{\lambda})$
0	0.86526	0.865734	0.86526	0.865424	0.856295
1	0.117436	0.116203	0.117002	0.116726	0.132846
2	0.0147347	0.0156316	0.0154214	0.0154946	0.0103048
3	0.00212761	0.00210428	0.00201503	0.00204579	0.000532898
4	0.00037546	0.000283377	0.000262145	0.000269384	0.0000206685
5	0.0000500613	0.0000381697	0.0000340141	0.0000354141	$6.41301 \cdot 10^{-7}$

4.6. För detta material gäller att $\bar{x} > s^2$, dvs väntevärdet i den empiriska fördelningen är större än variansen. Detta är aldrig uppfyllt för den negativa binomialfördelningen, så momentmetoden fungerar inte här. Metoden med anpassning till $\hat{p}_0$ och $\bar{x}$ gav skattningarna

$$\hat{\alpha}_2 = 60.986, \qquad \hat{\beta}_2 = 99.977$$

och ML-skattningarna blev

$$\hat{\alpha}_3 = 984.44, \qquad 1613.83.$$

Följande tabell visar resultatet, jämfört med vad vi fick i övning 4.2.

k	$\hat{p}_k$	$p_k(\hat{\lambda})$	$p_k(\hat{\alpha}_2, \hat{\beta}_2)$	$p_k(\hat{\alpha}_3, \hat{\beta}_3)$
0	0.545	0.543	0.545	0.543
1	0.325	0.331	0.329	0.331
2	0.110	0.101	0.101	0.101
3	0.015	0.020	0.021	0.020
4	0.005	0.003	0.003	0.003
5	0	0.000	0.000	0.000

4.7. Väntevärdet i den inversa normalfördelningen är alltså μ och enligt (4.5) är detta då också väntevärdet i IGP-fördelningen. I Övning 4.2 fann vi att sannolikheten för 0 i IGP-fördelningen är

$$p_0 = \exp\Big\{\mu\beta^{-1}\Big(1 - \sqrt{1+2\beta}\Big)\Big\}.$$

Om $\bar{x}$ är medelvärdet av observationerna, n antalet observationer och n_0 antalet observationer som är 0, ska vi alltså finna μ och β som löser

$$\begin{cases} \mu = \bar{x}, \\ \exp\Big\{\mu\beta^{-1}\Big(1 - \sqrt{1+2\beta}\Big)\Big\} = n_0/n \end{cases}$$

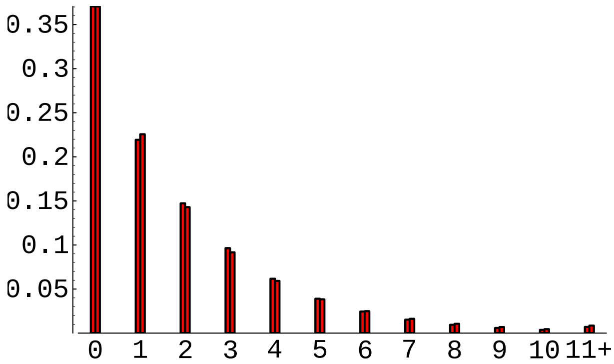
Följdaktligen gäller att

$$\hat{\mu} = \bar{x} = 1.78.$$

Ersätter vi μ med $\bar{x}$ i den andra ekvationen och löser ut β, får vi att

$$\hat{\beta} = 2\bar{x}(\bar{x} + \log\{n_0/n\})/(\log\{n_0/n\})^2 = 2.84389.$$

Nedanstående figur visar de sålunda skattade IGP-sannolikheterna (de högra staplarna) jämförda med de empiriska sannolikheterna. Vid beräkningen av IGP-sannolikheterna användes rekursionsformeln som härleddes i Övning 4.2. Anpassningen är sämre än för negativ binomial.



4.8. Med hjälp av att

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = -\log(1-x), \qquad -1 \le x \le 1,$$

får man

$$\begin{aligned} \gamma_1 &= \sum_{n=1}^{\infty} n\, q_n = \sum_{n=1}^{\infty} n \frac{r^n}{-n\log(1-r)} \\ &= \frac{1}{-\log(1-r)}\left(\sum_{n=0}^{\infty} r^n - 1\right) = \frac{1}{-\log(1-r)}\left(\frac{1}{1-r} - 1\right) \\ &= \frac{1}{-\log(1-r)}\,\frac{r}{1-r}. \end{aligned}$$

Vidare har vi att

$$\gamma_2 = \sum_{n=1}^{\infty} n(n-1)\, q_n = \frac{1}{-\log(1-r)} \sum_{n=1}^{\infty} (n-1)\, r^n.$$

Eftersom

$$\begin{aligned}\sum_{n=1}^{\infty} n\,r^n &= r\sum_{n=1}^{\infty} n\,r^{n-1} = r\sum_{n=1}^{\infty}\frac{d}{dr}\,r^n \\ &= r\,\frac{d}{dr}\sum_{n=1}^{\infty} r^n = r\,\frac{d}{dr}\sum_{n=0}^{\infty} r^n \\ &= r\,\frac{d}{dr}\,\frac{1}{1-r} = r\,\frac{1}{(1-r)^2},\end{aligned}$$

får vi att

$$\begin{aligned}\sum_{n=1}^{\infty}(n-1)\,r^n &= \sum_{n=1}^{\infty} n\,r^n - \sum_{n=1}^{\infty} r^n = \\ &= \frac{r}{(1-r)^2} - \frac{r}{1-r} = \frac{r-r(1-r)}{(1-r)^2} = \left(\frac{r}{1-r}\right)^2.\end{aligned}$$

Vi får alltså att

$$\gamma_2 = \frac{1}{-\log(1-r)}\left(\frac{r}{1-r}\right)^2.$$

Så ska vi se om vi får samma sak om vi låter $\alpha \to 0$ i (4.10). Först hade vi att

$$\gamma_1 = \frac{\alpha r}{(1-r)(1-(1-r)^{\alpha})}.$$

Om vi sätter $f(\alpha) = \alpha r$, $g(\alpha) = (1-r)(1-(1-r)^{\alpha})$, får vi med hjälp av L'Hospitals regel att

$$\begin{aligned}\lim_{\alpha\to 0}\gamma_1 &= \lim_{\alpha\to 0}\frac{f(\alpha)}{g(\alpha)} = \lim_{\alpha\to 0}\frac{f'(\alpha)}{g'(\alpha)} \\ &= \lim_{\alpha\to 0}\frac{r}{-(1-r)\,(1-r)^{\alpha}\log(1-r)} = \frac{1}{-\log(1-r)}\,\frac{r}{1-r}.\end{aligned}$$

På samma sätt får vi

$$\lim_{\alpha\to 0}\gamma_2 = \lim_{\alpha\to 0}\frac{r^2(2\alpha+1)}{-(1-r)^2\,(1-r)^{\alpha}\log(1-r)} = \frac{1}{-\log(1-r)}\left(\frac{r}{1-r}\right)^2.$$

4.9. Man har att

$$q_k = P\big(N = k \mid N \geq 1\big) = \frac{P(N=k)}{1-P(N=0)} = \frac{\lambda^k}{k!\,(e^{\lambda}-1)}.$$

Likelihoodfunktionen blir

$$L(\lambda) = \prod_{k=1}^{n}\frac{\lambda^{x_k}}{x_k!\,(e^{\lambda}-1)} = \lambda^{n\bar{x}}\left(\prod_{k=1}^{n} x_k!\right)(e^{\lambda}-1)^{-n}.$$

Den logaritmerade likelihoodfunktionen är

$$\ell(\lambda) = n\bar{x}\log\lambda - \sum_{k=1}^{n}\log(x_k!) - n\log(e^{\lambda}-1),$$

så att
$$\ell'(\lambda) = \frac{n\bar{x}}{\lambda} - \frac{ne^{\lambda}}{e^{\lambda}-1}.$$
ML-ekvationen $\ell'(\lambda) = 0$ blir
$$\bar{x}(1-e^{-\lambda}) = \lambda.$$
Vi löser denna med hjälp av Newton-Raphson. Sätt
$$f(\lambda) := \bar{x}(1-e^{-\lambda}) - \lambda, \qquad f'(\lambda) = \bar{x}e^{-\lambda} - 1.$$
Via iterationsformeln
$$\lambda_{n+1} = \lambda_n - \frac{f(\lambda_n)}{f'(\lambda_n)}$$
finner vi snabbt lösningen $\hat{\lambda} = 2.935$.

4.10. Antag först att $\alpha > 0$. Logaritmen av likelihoodfunktionen är
$$\ell(\alpha, r) = \sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \log(\alpha+i) - \sum_{k=1}^{n} n_k \sum_{i=1}^{k} \log i - n\log\big\{(1-r)^{-\alpha} - 1\big\} + n\bar{x}\log r.$$
Man får att
$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \frac{1}{\alpha+i} + n\frac{(1-r)^{-\alpha}\log(1-r)}{(1-r)^{-\alpha}-1} \\ &= \sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \frac{1}{\alpha+i} + \frac{n\log(1-r)}{1-(1-r)^{\alpha}},\end{aligned}$$
så den första ML-ekvationen blir
$$\sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \frac{1}{\alpha+i} + \frac{n\log(1-r)}{1-(1-r)^{\alpha}} = 0. \tag{7.4}$$
Vidare får man att
$$\frac{\partial \ell}{\partial r} = -n\,\frac{\alpha(1-r)^{-\alpha-1}}{(1-r)^{-\alpha}-1} + \frac{n\bar{x}}{r} = -\frac{n\alpha}{(1-r)(1-(1-r)^{\alpha})} + \frac{n\bar{x}}{r},$$
så den andra ML-ekvationen blir
$$\frac{\bar{x}}{r} = \frac{\alpha}{(1-r)(1-(1-r)^{\alpha})},$$
dvs
$$\bar{x}(1-r)(1-(1-r)^{\alpha}) - r\alpha = 0. \tag{7.5}$$
Detta är ekvationen (4.15). Av ekvationen (7.5) ser vi också att
$$1-(1-r)^{\alpha} = \frac{r\alpha}{\bar{x}(1-r)},$$
och stoppar vi in detta i (7.4), får vi
$$\sum_{k=1}^{m} n_k \sum_{i=0}^{k-1} \frac{1}{\alpha+i} + \frac{n\bar{x}(1-r)\log(1-r)}{r\alpha} = 0.$$

Delar vi med n, får vi (4.14).

Om $-1 < \alpha < 0$, skriver vi i stället

$$q_1 = \frac{-\alpha\, r}{1-(1-r)^{-\alpha}},$$

och, för $k \geq 2$,

$$\begin{aligned} q_k &= \prod_{i=0}^{k-1}(\alpha+i)\,\frac{r^k}{k!\,((1-r)^{-\alpha}-1)} \\ &= \prod_{i=1}^{k-1}(\alpha+i)\,\frac{\alpha\, r^k}{k!\,((1-r)^{-\alpha}-1)} \\ &= \prod_{i=1}^{k-1}(\alpha+i)\,\frac{-\alpha\, r^k}{k!\,(1-(1-r)^{-\alpha})}. \end{aligned}$$

Logaritmen av likelihoodfunktionen kan då skrivas som

$$\begin{aligned} \ell(\alpha,r) &= \sum_{k=1}^{n} n_k \log q_k = n_1 \log q_1 + \sum_{k=2}^{m} n_k \log q_k \\ &= n_1 \log \frac{-\alpha\, r}{1-(1-r)^{-\alpha}} + \sum_{k=2}^{m} n_k \log\left\{\prod_{i=1}^{k-1}(\alpha+i)\,\frac{-\alpha\, r^k}{k!\,(1-(1-r)^{-\alpha})}\right\} \\ &= n_1\big[\log(-\alpha)+\log r-\log(1-(1-r)^{-\alpha})\big] \\ &+ \sum_{k=2}^{m} n_k\left[\sum_{i=1}^{k-1}\log(\alpha+i)+\log(-\alpha)+k\log r-\log k!-\log(1-(1-r)^{-\alpha})\right] \\ &= \sum_{k=1}^{m} n_k\log(-\alpha)+\sum_{k=1}^{m} n_k\, k\log r-\sum_{k=1}^{m} n_k\log k!-\sum_{k=1}^{m} n_k\log(1-(1-r)^{-\alpha})+\sum_{k=2}^{m} n_k\sum_{i=1}^{k-1}\log(\alpha+i) \\ &= n\log(-\alpha)+n\bar{x}\log r-\sum_{k=1}^{m} n_k\log k!-n\log(1-(1-r)^{-\alpha})+\sum_{k=2}^{m} n_k\sum_{i=1}^{k-1}\log(\alpha+i). \end{aligned}$$

De partiella derivatorna är

$$\begin{aligned} \frac{\partial\ell}{\partial\alpha} &= \frac{n}{\alpha}-\frac{n(1-r)^{-\alpha}\log(1-r)}{1-(1-r)^{-\alpha}}+\sum_{k=2}^{m} n_k\sum_{i=1}^{k-1}\frac{1}{\alpha+i} \\ &= -\frac{n(1-r)^{-\alpha}\log(1-r)}{1-(1-r)^{-\alpha}}+\sum_{k=1}^{m} n_k\,\frac{1}{\alpha}+\sum_{k=2}^{m} n_k\sum_{i=1}^{k-1}\frac{1}{\alpha+i} \\ &= -\frac{n(1-r)^{-\alpha}\log(1-r)}{1-(1-r)^{-\alpha}}+\sum_{k=1}^{m} n_k\sum_{i=0}^{k-1}\frac{1}{\alpha+i}, \end{aligned}$$

respektive

$$\frac{\partial\ell}{\partial r} = \frac{n\bar{x}}{r}+\frac{n\alpha(1-r)^{-\alpha-1}}{1-(1-r)^{-\alpha}}.$$

ML-ekvationerna blir alltså

$$\begin{aligned} -\frac{n(1-r)^{-\alpha}\log(1-r)}{1-(1-r)^{-\alpha}}+\sum_{k=1}^{m} n_k\sum_{i=0}^{k-1} &= 0, &(7.6)\\ \frac{n\bar{x}}{r}+\frac{n\alpha(1-r)^{-\alpha-1}}{1-(1-r)^{-\alpha}} &= 0. &(7.7) \end{aligned}$$

Om vi multiplicerar (7.7) med $r(1-(1-r)^{-\alpha})(1-r)^{\alpha+1}$, får vi

$$n\bar{x}((1-r)^{\alpha+1}-(1-r))+nr\alpha=0,$$

eller

$$\bar{x}(1-r)(1-(1-r)^{-\alpha})-r\alpha=0.$$

Detta är (4.14). Vidare gäller enligt (7.7) att

$$1-(1-r)^{-\alpha}=\frac{r\alpha(1-r)^{-\alpha-1}}{\bar{x}},$$

så (7.6) kan skrivas

$$\frac{n(1-r)^{-\alpha}\log(1-r)\bar{x}}{r\alpha(1-r)^{-\alpha-1}}+\sum_{k=1}^{m}n_k\sum_{i=0}^{k-1}\frac{1}{\alpha+i}=0,$$

dvs

$$\frac{\bar{x}(1-r)\log(1-r)}{r\alpha}+\frac{1}{n}\sum_{k=1}^{m}n_k\sum_{i=o}^{k-1}\frac{1}{\alpha+i}=0,$$

vilket är (4.15). Man får alltså samma ML-ekvationer som för $\alpha>0$.

Kapitel 5: Sammansatta fördelningar

5.1. Låt M vara den momentgenererande funktionen för H, dvs

$$M(t)=E\big[e^{tY_1}\bigm| Y_1>0\big].$$

Den momentgenererande funktionen för $Y_1+\cdots+Y_n$ är

$$\begin{aligned}
E\big[e^{t(Y_1+\cdots+Y_n)}\big] &= \prod_{k=1}^{n}E\big[e^{tY_k}\big]=\big(E\big[e^{tY_1}\big]\big)^n\\
&= \big(E\big[e^{tY_1}\bigm| Y_1=0\big]\,P(Y_1=0)+E\big[e^{tY_1}\bigm| Y_1>0\big]\,P(Y_1>0)\big)^n\\
&= \big(1-p+M(t)\,p\big)^n=\big(1+p(M(t)-1)\big)^n.
\end{aligned}$$

Den momentgenererande funktionen för binomialfördelningen är $(1+p(e^t-1))^n$, så enligt (5.3) har $Y_1+\cdots+Y_n$ en SaBi(n,p,H)-fördelning.

5.2. Sätt $M_1(t)=E[e^{tK}]$, $M_2(t)=E[e^{tN_1}]$, $M_3(t)=E[e^{tX_{11}}]$. Den momentgenererande funktionen för S

kan då skrivas

$$\begin{aligned}
E[e^{tS}] &= \sum_{r=0}^{\infty} E\Big[\exp\Big\{t\sum_{j=1}^{r}\sum_{k=1}^{N_j} X_{jk}\Big\}\ \Big|\ K=r\Big] P(K=r) \\
&= \sum_{r=0}^{\infty}\prod_{j=1}^{r} E\Big[\exp\Big\{t\sum_{k=1}^{N_j} X_{jk}\Big\}\Big] P(K=r) \\
&= \sum_{r=0}^{\infty}\prod_{j=1}^{r}\sum_{s=0}^{\infty} E\Big[\exp\Big\{t\sum_{k=1}^{s} X_{jk}\Big\}\Big] P(N_j=s)\, P(K=r) \\
&= \sum_{r=0}^{\infty}\prod_{j=1}^{r}\sum_{s=0}^{\infty}\prod_{k=1}^{s} E\big[e^{tX_{11}}\big]\, P(N_j=s)\, P(K=r) \\
&= \sum_{r=0}^{\infty}\prod_{j=1}^{r}\sum_{s=0}^{\infty}\big(M_3(t)\big)^s P(N_j=s)\, P(K=r) = \sum_{r=0}^{\infty}\prod_{j=1}^{r} M_2(\log M_3(t))\, P(K=r) \\
&= \sum_{r=0}^{\infty}\big(M_2(\log M_3(t))\big)^r P(K=r) = M_1\big(\log M_2(\log M_3(t))\big).
\end{aligned}$$

Om vi sätter $\Psi_k(t) := \log M_k(t)$, blir den kumulantgenererande funktionen för S

$$\Psi(t) = \Psi_1(\Psi_2(\Psi_3(t))).$$

Vi deriverar denna två gånger för att finna $E[S]$ och $V[S]$:

$$\begin{aligned}
\Psi'(t) &= \Psi_1'(t)(\Psi_2(\Psi_3(t)))\,\Psi_2'(\Psi_3(t))\,\Psi_3'(t), \\
\Psi''(t) &= \Psi_1''(\Psi_2(\Psi_3(t)))\,(\Psi_2'(\Psi_3(t)))^2\,(\Psi_3'(t))^2 + \Psi_1'(\Psi_2(\Psi_3(t)))\,\Psi_2''(\Psi_3(t))\,(\Psi_3'(t))^2 \\
&+ \Psi_1'(\Psi_2(\Psi_3(t)))\,\Psi_2'(\Psi_3(t))\,\Psi_3''(t).
\end{aligned}$$

Man får alltså att

$$\begin{aligned}
E[S] &= \Psi'(0) = \Psi_1'(0)\,\Psi_2'(0)\,\Psi_3'(0) = E[K]\,E[N_1]\,E[X_{11}], \\
V[S] &= \Psi''(0) = V[K]\,\big(E[N_1]\big)^2\,\big(E[X_{11}]\big)^2 + E[K]\,V[N_1]\,\big(E[X_{11}]\big)^2 + E[K]\,E[N_1]\,V[X_{11}].
\end{aligned}$$

5.3. Vi behöver först den sannolikhetsgenererande funktionen för den utvidgade trunkerade negativa binomialfördelningen:

$$\begin{aligned}
Q(s) &= \sum_{n=1}^{\infty} q_n\, s^n = \sum_{n=1}^{\infty}\frac{\Gamma(\alpha+n)\, r^n}{\Gamma(\alpha)\, n!\,((1-r)^{-\alpha}-1)}\, s^n \\
&= \frac{1}{\Gamma(\alpha)\,((1-r)^{-\alpha}-1)}\sum_{n=1}^{\infty}\frac{\Gamma(\alpha+n)\,(rs)^n}{n!}.
\end{aligned}$$

Man har att

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{\Gamma(\alpha+n)\,(rs)^n}{n!} &= \sum_{n=1}^{\infty}\int_0^{\infty} x^{\alpha+n-1}\,e^{-x}\,dx\,\frac{(rs)^n}{n!} = \int_0^{\infty} x^{\alpha-1}\,e^{-x}\sum_{n=1}^{\infty}\frac{(rsx)^n}{n!}\\ &= \int_0^{\infty} x^{\alpha-1}\,e^{-x}\big(e^{rsx}-1\big)dx = \int_0^{\infty} x^{\alpha-1}\,e^{-(1-rs)x}\,dx - \int_0^{\infty} x^{\alpha-1}\,e^{-x}\,dx\\ &= \int_0^{\infty}\left(\frac{y}{1-rs}\right)^{\alpha-1} e^{-y}\,\frac{1}{1-rs}\,dy - \Gamma(\alpha) = (1-rs)^{-\alpha}\int_0^{\infty} y^{\alpha-1}\,e^{-y}\,dy - \Gamma(\alpha)\\ &= \Gamma(\alpha)\big((1-rs)^{-\alpha}-1\big).\end{aligned}$$

Detta ger att

$$Q(s) = \frac{\Gamma(\alpha)((1-rs)^{-\alpha}-1)}{\Gamma(\alpha)\,((1-r)^{-\alpha}-1)} = \frac{1-(1-rs)^{-\alpha}}{1-(1-r)^{-\alpha}}.$$

Man får samma sak om man utgår från den momentgenererande funktionen P för den negativa binomialfördelningen och utnyttjar att

$$Q(s) = \frac{P(s)-P(0)}{1-P(0)}.$$

Om vi nu sätter $\alpha = -1/2$, får vi

$$Q(s) = \frac{1-\sqrt{1-rs}}{1-\sqrt{1-r}} = P_2(s).$$

Den momentgenererande funktionen för N är

$$\begin{aligned}P_1(P_2(s)) &= \exp\left\{-\lambda\left(1-\frac{1-\sqrt{1-rs}}{1-\sqrt{1-r}}\right)\right\} = \exp\left\{-\lambda\,\frac{1-\sqrt{1-r}-1+\sqrt{1-rs}}{1-\sqrt{1-r}}\right\}\\ &= \exp\left\{\frac{\lambda\sqrt{1-r}}{1-\sqrt{1-r}}\left(1-\frac{\sqrt{1-rs}}{\sqrt{1-r}}\right)\right\}\\ &= \exp\left\{\frac{\lambda\sqrt{1-r}}{1-\sqrt{1-r}}\left(1-\sqrt{\frac{1-r+r(1-s)}{1-r}}\right)\right\}\\ &= \exp\left\{\frac{\lambda\sqrt{1-r}}{1-\sqrt{1-r}}\left(1-\sqrt{1+\frac{r}{1-r}\,(1-s)}\right)\right\}.\end{aligned}$$

Den sannolikhetsgenererande funktionen för en IGP-fördelning med parametrar μ och β är (se övning 4.3)

$$P(s) = \exp\Big\{\mu\beta^{-1}\Big(1-\sqrt{1+2\beta(1-s)}\Big)\Big\}.$$

Följdaktligen är N IGP-fördelad med parametrar

$$\beta = \frac{r}{2(1-r)}, \qquad \mu = \beta\,\frac{\lambda\sqrt{1-r}}{1-\sqrt{1-r}} = \frac{\lambda r}{2\sqrt{1-r}(1-\sqrt{1-r})}.$$

5.4. Följande tabell visar 20 utfall av $\hat{x}_q$, uttryckta i miljoner, som jag själv fick.

3.926	4.885	4.553	3.471	4.020	2.974	8.731	4.844	6.021	10.022
4.159	7.746	7.127	9.192	6.569	2.737	2.915	9.817	7.428	4.128

Dessa kan jämföras med det sanna värdet: $x_q = 4.045$.

5.5. Till att börja med har vi att

$$\begin{aligned} P\big(X-(u+v)>c \mid X>u+v\big) &= P\big(X>u+v+c \mid X>u+v\big) \\ &= \frac{P(X>u+v+c,\, X>u+v)}{P(X>u+v)} \\ &= \frac{P(X>u+v+c)}{P(X>u+v)} \\ &= P(X-u>v+c)P(X-u>v) \\ &= \frac{\big(\alpha/(\alpha+v+c)\big)^\gamma}{\big(\alpha/(\alpha+v)\big)^\gamma} \\ &= \left(\frac{\alpha+v}{\alpha+v+c}\right)^\gamma . \end{aligned}$$

Vi söker q-kvantilen, dvs det c som uppfyller

$$P\big(X-(u+v)>c \mid X>u+v\big) = q.$$

Enligt vad vi fann ovan kan denna ekvation skrivas

$$\left(\frac{\alpha+v}{\alpha+v+c}\right)^\gamma = q.$$

Löser man ut c, får man

$$c = (\alpha+v)\big(q^{-1/\gamma}-1\big).$$

Eftersom $q^{-1/\gamma} > 1$ då $0 < q < 1$ och $\gamma > 0$, är kvantilen växande som funktion av v. (Är detta vad man väntar sig?)

För $\alpha = 900\,000$, $\gamma = 1.1$ och $v = 100$ miljoner får man 0.5-kvantilen 88 miljoner.

5.6. Den momentgenererande funktionen för SaBi(n, p, F)-fördelningen är $(1 + p(M(t) - 1))^n$. Men eftersom $M(t) = a + \tilde{M}(t)(1 - a)$, kan denna skrivas

$$(1 + p(a + \tilde{M}(t)(1 - a)) - 1)^n = (1 + p(1 - a)(\tilde{M}(t) - 1))^n,$$

vilket är momentgenererande funktionen för SaBi$(n, p(1 - a), H)$-fördelningen.

5.7. IGP-fördelningen är en blandad Poissonfördelning med blandningstätheten

$$u(\lambda) = \frac{\mu}{\sqrt{2\pi\beta\lambda^3}} \exp\left\{-\frac{(\lambda-\mu)^2}{2\beta\lambda}\right\}.$$

Men då har antalet strikt positiva skador en blandad Poissonfördelning med blandningstätheten i (5.7):

$$v(\theta) = u\left(\frac{\theta}{1-a}\right)\frac{1}{1-a}.$$

Men

$$u\left(\frac{\theta}{1-a}\right)\frac{1}{1-a} = \frac{\mu}{\sqrt{2\pi\beta\theta^3/(1-a)^3}}\,\frac{1}{1-a}\,\exp\left\{-\frac{(\theta/(1-a)-\mu)^2}{2\beta\theta/(1-a)}\right\}$$
$$= \frac{\mu(1-a)}{\sqrt{2\pi\beta(1-a)\theta^3}}\,\exp\left\{-\frac{(\theta-\mu(1-a))^2}{2\beta(1-a)\,\theta}\right\},$$

vilket är tätheten för en invers normalfördelning med parametrar $\mu(1-a)$ och $\beta(1-a)$. Det följer att antalet strikt positiva skador är IGP-fördelat med samma parametrar.

5.8. Värdena på g_0, a och b då N är Poissonfördelat finns i texten.

Om N är binomialfördelad med parametrar n och q, gäller att $P_1(t) = (1 + q(t-1))^n$, så

$$g_0 = P_1(f_0) = (1 - q(1-f_0))^n.$$

Vidare har man att

$$\begin{aligned} p_k &= \binom{n}{k} q^k\,(1-q)^{n-k} = \frac{n!}{k!\,(n-k)!}\,q^k\,(1-q)^{n-k} \\ &= \frac{n-k+1}{k}\,\frac{q}{1-q}\,\frac{n!}{(k-1)!\,(n-(k-1))!}\,q^{k-1}\,(1-q)^{n-(k-1)} \\ &= \frac{n-k+1}{k}\,\frac{q}{1-q}\,p_{k-1} = \left(-\frac{q}{1-q} + \frac{(n+1)q}{(1-q)k}\right)p_{k-1}, \qquad 1 \le k \le n. \end{aligned}$$

Man har alltså att

$$a = -\frac{q}{1-q}, \qquad b = \frac{(n+1)q}{1-q}.$$

Om N är negativt binomialfördelad, har vi att $P_1(t) = (1 + \beta^{-1}(1-s))^{-\alpha}$, så

$$g_0 = P_1(f_0) = (1 + \beta^{-1}(1-f_0))^{-\alpha}.$$

Vidare har vi att

$$p_k = \frac{\alpha+k-1}{(\beta+1)k}\,p_{k-1} = \left(\frac{1}{\beta+1} + \frac{\alpha-1}{(\beta+1)k}\right)p_{k-1}, \qquad k \ge 1,$$

dvs

$$a = \frac{1}{\beta+1}, \qquad b = \frac{\alpha-1}{\beta+1}.$$

5.9. Vi antar alltså att $f_0 > 0$ och sätter $h_k := f_k/(1-f_0)$, $k \ge 1$, samt $h_0 := 0$. Låt F och H vara fördelningsfunktionerna hörande till sannolikheterna $\{f_k\}$ respektive $\{h_k\}$.

Poisson: Vi ska visa att Panjers rekursionsformel ser likadan ut då $S \in \mathrm{SaPo}(\lambda, F)$ som då $S \in \mathrm{SaPo}(\lambda(1-f_0), H)$. I det första fallet får vi enligt (5.14) rekursionsformeln

$$g_n = \sum_{k=1}^{n} \frac{\lambda k}{n}\,f_k\,g_{n-k}, \tag{7.8}$$

och startvärdet $g_0 = e^{-\lambda(1-f_0)}$.

I det andra fallet får vi

$$\begin{aligned} g_n &= \sum_{k=1}^{n} \frac{\lambda(1-f_0)k}{n} h_k\, g_{n-k} = \sum_{k=1}^{n} \frac{\lambda(1-f_0)k}{n} \frac{f_k}{1-f_0} g_{n-k} \\ &= \sum_{k=1}^{n} \frac{\lambda k}{n} f_k\, g_{n-k}, \end{aligned}$$

dvs samma som i (7.8), och startvärdet

$$g_0 = e^{-\lambda(1-f_0)(1-h_0)} = e^{-\lambda(1-f_0)}.$$

Binomial: Vi ska nu verifiera att vi får samma formler då $S \in \mathrm{SaBi}(m, q, F)$, som för fallet $S \in \mathrm{SaBi}(m, q(1-f_0), H)$. I det första fallet har vi rekursionsformeln

$$\begin{aligned} g_n &= \frac{1}{1+qf_0/(1-q)} \sum_{k=1}^{n} \left(-\frac{q}{1-q} + \frac{(m+1)qk}{(1-q)n} \right) f_k\, g_{n-k} \\ &= \sum_{k=1}^{n} \left(-\frac{q}{1-q+qf_0} + \frac{(m+1)qk}{(1-q+qf_0)n} \right) f_k\, g_{n-k} \\ &= \sum_{k=1}^{n} \left(-\frac{q}{1-q(1-f_0)} + \frac{(m+1)qk}{(1-q(1-f_0))n} \right) f_k\, g_{n-k}. \end{aligned} \tag{7.9}$$

Startvärdet är

$$P_1(f_0) = (1-q(1-f_0))^n.$$

Då $S \in \mathrm{SaBi}(m, q(1-f_0), H)$, får vi rekursionsformeln

$$\begin{aligned} g_n &= \sum_{k=1}^{n} \left(-\frac{q(1-f_0)}{1-q(1-f_0)} + \frac{(m+1)q(1-f_0)k}{(1-q(1-f_0))n} \right) h_k\, g_{n-k} \\ &= \sum_{k=1}^{n} \left(-\frac{q(1-f_0)}{1-q(1-f_0)} + \frac{(m+1)q(1-f_0)k}{(1-q(1-f_0))n} \right) \frac{f_k}{1-f_0} g_{n-k} \\ &= \sum_{k=1}^{n} \left(-\frac{q}{1-q(1-f_0)} + \frac{(m+1)qk}{(1-q(1-f_0))n} \right) f_k\, g_{n-k}, \end{aligned}$$

dvs samma som i (7.9). Startvärdet är

$$(1-q(1-f_0)(1-h_0))^n = (1-q(1-f_0))^n,$$

dvs samma.

Negativ binomial: Slutligen ska vi visa att vi får samma sak då $S \in \mathrm{SaNeBi}(\alpha, \beta, F)$, som då $S \in \mathrm{SaNeBi}(\alpha, \beta/(1-f_0), H)$. I det första fallet har vi rekursionsformeln

$$\begin{aligned} g_n &= \frac{1}{1-f_0/(\beta+1)} \sum_{k=1}^{n} \left(\frac{1}{\beta+1} + \frac{(\alpha-1)k}{(\beta+1)n} \right) f_k\, g_{n-k} \\ &= \sum_{k=1}^{n} \left(\frac{1}{\beta+1-f_0} + \frac{(\alpha-1)k}{(\beta+1-f_0)n} \right) f_k\, g_{n-k} \end{aligned} \tag{7.10}$$

och startvärdet $P_1(f_0) = (1+\beta^{-1}(1-f_0))^{-\alpha}$.

I det andra fallet får vi

$$\begin{aligned} g_n &= \sum_{k=1}^{n}\left(\frac{1}{1+\beta/(1-f_0)}+\frac{(\alpha-1)k}{(1+\beta/(1-f_0))n}\right)h_k\,g_{n-k} \\ &= \sum_{k=1}^{n}\left(\frac{1}{1+\beta/(1-f_0)}+\frac{(\alpha-1)k}{(1+\beta/(1-f_0))n}\right)\frac{f_k}{1-f_0}\,g_{n-k} \\ &= \sum_{k=1}^{n}\left(\frac{1}{\beta+1-f_0}+\frac{(\alpha-1)k}{(\beta+1-f_0)n}\right)f_k\,g_{n-k}, \end{aligned}$$

dvs samma som i (7.10). Startvärdet är

$$(1+(\beta/(1-f_0))^{-1}(1-h_0))^{-\alpha}=(1+\beta^{-1}(1-f_0))^{-\alpha},$$

dvs samma.

5.10. Vi ska först visa formeln

$$f_k^{n*}=\frac{n}{mk}\sum_{j=1}^{k} j\,f_j^{m*}\,f_{k-j}^{(n-m)*}. \tag{7.11}$$

Om $X_1, X_2, \ldots$ är oberoende och likafördelade med $f_k = P(X_1 = k)$, gäller för det första att, för $1 \leq m < n$,

$$\begin{aligned} &P\big(X_1+\cdots+X_m=j \mid X_1+\cdots+X_n=k\big)=\frac{P(X_1+\cdots+X_m=j,\,X_1+\cdots+X_n=k)}{P(X_1+\cdots+X_n=k)} \\ &= \frac{P(X_1+\cdots+X_m=j,\,X_{m+1}+\cdots+X_n=k-j)}{P(X_1+\cdots+X_n=k)} \\ &= \frac{P(X_1+\cdots+X_m=j)\,P(X_{m+1}+\cdots+X_n=k-j)}{P(X_1+\cdots+X_n=k)} \\ &= \frac{P(X_1+\cdots+X_m=j)\,P(X_1+\cdots+X_{n-m}=k-j)}{P(X_1+\cdots+X_n=k)}=\frac{f_j^{m*}\,f_{k-j}^{(n-m)*}}{f_k^{n*}}, \end{aligned}$$

från vilket det följer att

$$\begin{aligned} E\big[X_1+\cdots+X_m \mid X_1+\cdots+X_n=k\big] &= \sum_{j=0}^{k} j\,P\big(X_1+\cdots+X_m=j \mid X_1+\cdots+X_n=k\big) \\ &= \sum_{j=0}^{k} j\,\frac{f_j^{m*}\,f_{k-j}^{(n-m)*}}{f_k^{n*}}. \end{aligned} \tag{7.12}$$

För det andra gäller att

$$\begin{aligned} E\big[X_1+\cdots+X_m \big| X_1+\cdots+X_n=k\big] &= \sum_{j=1}^{m} E\big[X_j \mid X_1+\cdots+X_n=k\big] \\ =\sum_{j=1}^{m} E\big[X_1 \mid X_1+\cdots+X_n=k\big] &= m\,E\big[X_1 \mid X_1+\cdots+X_n=k\big] \\ =\frac{m}{n}\,n\,E\big[X_1 \mid X_1+\cdots+X_n=k\big] &= \frac{m}{n}\sum_{j=1}^{n} E\big[X_1 \mid X_1+\cdots+X_n=k\big] \\ =\frac{m}{n}\sum_{j=1}^{n} E\big[X_j \mid X_1+\cdots+X_n=k\big] &= \frac{m}{n}E\big[X_1+\cdots+X_n \mid X_1+\cdots+X_n=k\big] \\ &= \frac{m}{n}\,k. \end{aligned} \tag{7.13}$$

Formeln (7.11) följer av (7.12) och (7.13). När det gäller faltningsformeln har vi att

$$
\begin{aligned}
f_k^{n*} &= P(X_1,\ldots,X_n=k)=\sum_{j=1}^{k} P(X_1+\cdots+X_m=j,\,X_1+\cdots+X_n=k)\\
&= \sum_{j=1}^{k} P(X_1+\cdots+X_m=j,\,X_{m+1}+\cdots+X_n=k-j)\\
&= \sum_{j=1}^{k} P(X_1+\cdots+X_m=j)\,P(X_{m+1}+\cdots+X_n=k-j)\\
&= \sum_{j=1}^{k} P(X_1+\cdots+X_m=j)\,P(X_1+\cdots+X_{n-m}=k-j)=\sum_{j=1}^{k} f_j^{m*}\,f_{k-j}^{(n-m)*}. \qquad (7.14)
\end{aligned}
$$

Via (7.11) och (7.14) får vi nu, för $k\ge 1$,

$$
\begin{aligned}
g_k &= \sum_{n=1}^{\infty} p_n\,f_k^{n*} = p_1\,f_k+\sum_{n=2}^{\infty} p_n\,f_k^{n*}\\
&= p_1\,f_k+\sum_{n=2}^{\infty}\Big(a_1+\frac{b_1}{n}\Big)p_{n-1}\,f_k^{n*}+\sum_{n=2}^{\infty}\Big(a_2+\frac{b_2}{n}\Big)p_{n-2}\,f_k^{n*}\\
&= p_1\,f_k+\sum_{n=2}^{\infty} a_1\,p_{n-1}\,f_k+\sum_{n=2}^{\infty} b_1\,p_{n-1}\,\frac{1}{n}\,f_k^{n*}+\sum_{n=2}^{\infty} a_2 p_{n-2}\,f_k^{n*}+\sum_{n=2}^{\infty} b_2\,p_{n-2}\,\frac{1}{n} f_k^{n*}\\
&= p_1\,f_k+\sum_{n=2}^{\infty} a_1\,p_{n-1}\sum_{j=1}^{k} f_j\,f_{k-j}^{(n-1)*}+\sum_{n=2}^{\infty} b_1\,p_{n-1}\sum_{j=1}^{k}\frac{j}{k}\,f_j\,f_{k-j}^{(n-1)*}\\
&+ \sum_{n=2}^{\infty} a_2\,p_{n-2}\sum_{j=1}^{k} f_j^{2*}\,f_{k-j}^{(n-2)*}+\sum_{n=2}^{\infty} b_2\,p_{n-2}\sum_{j=1}^{k}\frac{j}{2k}\,f_j^{2*}\,f_{k-j}^{(n-2)*}\\
&= p_1\,f_k+\sum_{j=1}^{k} a_1\,f_j\sum_{n=2}^{\infty} p_{n-1}\,f_{k-j}^{(n-1)*}+\sum_{j=1}^{k}\frac{b_1\,j}{k}\,f_j\sum_{n=2}^{\infty} p_{n-1}\,f_{k-j}^{(n-1)*}\\
&+ \sum_{j=1}^{k} a_2\,f_j^{2*}\sum_{n=2}^{\infty} p_{n-2}\,f_{k-j}^{(n-2)*}+\sum_{j=1}^{k}\frac{b_2\,j}{2k}\,f_j^{2*}\sum_{n=2}^{\infty} p_{n-2}\,f_{k-j}^{(n-2)*}\\
&= p_1\,f_k+\sum_{j=1}^{k} a_1\,f_j\sum_{m=0}^{\infty} p_m\,f_{k-j}^{m*}+\sum_{j=1}^{k}\frac{b_1 j}{k}\sum_{m=0}^{\infty} p_m\,f_{k-j}^{m*}-\sum_{j=1}^{k} a_1\,f_j\,p_0\,f_{k-j}^{0*}-\sum_{j=1}^{k} b_1\,f_j p_0\,f_{k-j}^{0*}\\
&+ \sum_{j=1}^{k} a_2\,f_j^{2*}\sum_{m=0}^{\infty} p_m\,f_{k-j}^{m*}+\sum_{j=1}^{k}\frac{b_2 j}{2k}\,f_j^{2*}\sum_{m=0}^{\infty} p_m\,f_{k-j}^{m*}\\
&= \sum_{j=1}^{k}\bigg[\Big(a_1+\frac{b_1 j}{k}\Big)f_j+\Big(a_2+\frac{b_2 j}{2k}\Big)f_j^{2*}\bigg]g_{k-j}+(p_1-(a_1+b_1)p_0)\,f_k.
\end{aligned}
$$

5.11. Tätheten för X_1 givet $X_1 + \cdots + X_n = y$ ges av

$$\begin{aligned} f_{X_1}\big(x \mid X_1 + \cdots + X_n = y\big) &= \frac{f_{X_1}(x)\, f_{X_1+\cdots+X_n}\big(y \mid X_1 = x\big)}{f_{X_1+\cdots+X_n}(y)} \\ &= \frac{f_{X_1}(x)\, f_{X_1+\cdots+X_{n-1}}(y-x)}{f_{X_1+\cdots+X_n}(y)} \\ &= \frac{f(x)\, f^{(n-1)*}(y-x)}{f^{n*}(y)}. \end{aligned}$$

Det gäller alltså att

$$E\big[X_1 \mid X_1 + \cdots + X_n = y\big] = \int x\, \frac{f(x)\, f^{(n-1)*}(y-x)}{f^{n*}(y)}\, dx.$$

Precis som i det diskreta fallet har man att

$$\begin{aligned} E\big[X_1 \mid X_1 + \cdots + X_n = y\big] &= \frac{1}{n} \sum_{j=1}^{n} E\big[X_j \mid X_1 + \cdots + X_n = y\big] \\ &= \frac{1}{n} E\big[X_1 + \cdots + X_n \mid X_1 + \cdots + X_n = y\big] = \frac{y}{n}. \end{aligned}$$

Man får alltså att

$$\int x\, \frac{f(x)\, f^{(n-1)*}(y-x)}{f^{n*}(y)}\, dx = \frac{y}{n},$$

eller

$$\frac{1}{n} f^{n*}(y) = \frac{1}{y} \int x\, f(x)\, f^{(n-1)*}(y-x)\, dx.$$

Det följer att g ska uppfylla

$$\begin{aligned} g(y) &= \sum_{n=1}^{\infty} p_n\, f^{n*}(y) = p_1\, f(y) + \sum_{n=2}^{\infty} \Big(a + \frac{b}{n}\Big) p_{n-1}\, f^{n*}(y) \\ &= p_1\, f(y) + a \sum_{n=2}^{\infty} p_{n-1}\, f^{n*}(y) + b \sum_{n=2}^{\infty} p_{n-1}\, \frac{1}{n}\, f^{n*}(y) \\ &= p_1\, f(y) + a \sum_{n=2}^{\infty} p_{n-1} \int f(x)\, f^{(n-1)*}(y-x)\, dx + b \sum_{n=2}^{\infty} p_{n-1}\, \frac{1}{y} \int x\, f(x)\, f^{(n-1)*}(y-x)\, dx \\ &= p_1\, f(y) + a \int f(x) \sum_{n=2}^{\infty} p_{n-1}\, f^{(n-1)*}(y-x)\, dx + \frac{b}{y} \int x\, f(x) \sum_{n=2}^{\infty} p_{n-1}\, f^{(n-1)*}(y-x)\, dx \\ &= p_1\, f(y) + a \int f(x) \sum_{n=1}^{\infty} p_n\, f^{n*}(y-x)\, dx + \frac{b}{y} \int x\, f(x) \sum_{n=1}^{\infty} p_n\, f^{n*}(y-x)\, dx \\ &= p_1\, f(y) + \int \Big(a + \frac{bx}{y}\Big) f(x)\, g(y-x)\, dx. \end{aligned}$$

5.12. Se lösningen till uppgift 5.14.

5.13. Sätt

$$h(x) = \begin{cases} x - a, & \text{om} \quad x > a, \\ 0, & \text{om} \quad x \le a. \end{cases}$$

Approximationen är

$$E\big[h(S)\big] = E\big[h(\mu+\sigma Z)\big] \approx E\big[h(\mu+\sigma W)\big] = E\left[h\left(\mu+\sigma\,\frac{Y-\alpha}{\sqrt{\alpha}}\right)\right].$$

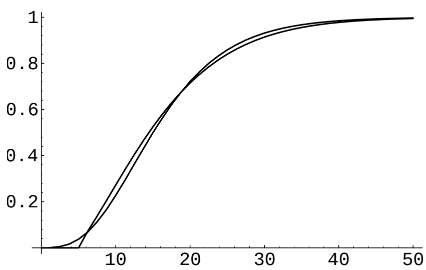
Vidare gäller att

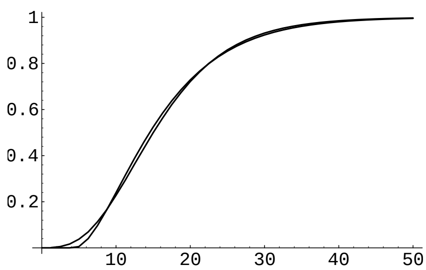
$$\begin{aligned} h\left(\mu+\sigma\,\frac{y-\alpha}{\sqrt{\alpha}}-a\right) &= \begin{cases} \mu+\sigma\,(y-\alpha)/\sqrt{\alpha}-a, & \text{om} \quad \mu+\sigma(y-\alpha)/\sqrt{\alpha} > a, \\ 0, & \text{om} \quad \mu+\sigma(y-\alpha)/\sqrt{\alpha} \le a \end{cases} \\ &= \begin{cases} \mu+\sigma(y-\alpha)/\sqrt{\alpha}-a, & \text{om} \quad y > \alpha+\sqrt{\alpha}(a-\mu)/\sigma, \\ 0, & \text{om} \quad y \le \alpha+\sqrt{\alpha}(a-\mu)/\sigma. \end{cases} \end{aligned}$$

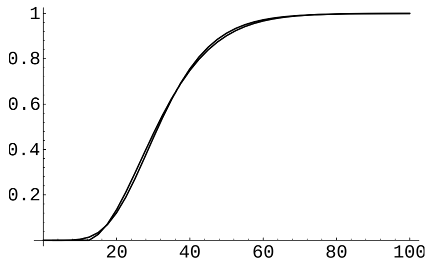
Antag att $\xi := \alpha+\sqrt{\alpha}(a-\mu)/\sigma > 0$. Om vi låter g_α beteckna tätheten för en gammafördelning med parametrar α och 1, får vi att

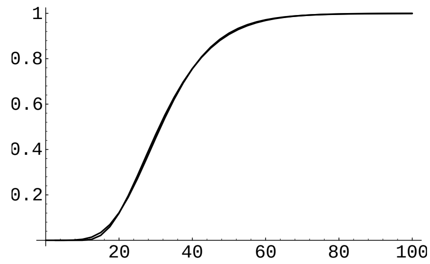
$$\begin{aligned} E\left[h\left(\mu+\sigma\,\frac{Y-\alpha}{\sqrt{\alpha}}\right)\right] &= \int_\xi^\infty \left(\mu+\sigma\,\frac{y-\alpha}{\sqrt{\alpha}}-a\right) g_\alpha(y)\,dy \\ &= \frac{\sigma}{\sqrt{\alpha}}\int_\xi^\infty y\,g_\alpha(y)\,dy - \big(a+\sigma\sqrt{\alpha}-\mu\big)\int_\xi^\infty g_\alpha(y)\,dy \\ &= \frac{\sigma}{\sqrt{\alpha}}\int_\xi^\infty y\,\frac{1}{\Gamma(\alpha)}\,y^{\alpha-1}\,e^{-y}\,dy - \big(a+\sigma\sqrt{\alpha}-\mu\big)\big(1-G_\alpha(\xi)\big) \\ &= \sigma\sqrt{\alpha}\int_\xi^\infty \frac{1}{\Gamma(\alpha+1)}\,y^{\alpha}\,e^{-y}\,dy - \big(a+\sigma\sqrt{\alpha}-\mu\big)\big(1-G_\alpha(\xi)\big) \\ &= \sigma\sqrt{\alpha}\big(1-G_{\alpha+1}(\xi)\big) - \big(a+\sigma\sqrt{\alpha}-\mu\big)\big(1-G_\alpha(\xi)\big) \\ &= \sigma\sqrt{\alpha}\big(G_\alpha(\xi)-G_{\alpha+1}(\xi)\big) - (a-\mu)\big(1-G_\alpha(\xi)\big). \end{aligned}$$

5.14. I nedanstående figurer jämförs NP-approximationen (vänstra figurerna) och gamma-approximationen (högra figurerna) med det “exakta” resultatet av Ströters metod. De övre figurerna visar resultaten för $\lambda = 10$ och de undre resultaten för $\lambda = 20$.









Så följer en tabell med resultaten för $\lambda = 100$. I tabellen betecknar F_1, F_2 och F_3 fördelningsfunktionen för S beräknad med Ströters metod, NP-approximation, respektive gamma-approximation. För Ströters metod har steglängden $h = 0.025$ använts.

x	$F_1(x)$	$F_2(x)$	$F_3(x)$
100	0.002	0.001	0.002
105	0.005	0.004	0.004
110	0.011	0.009	0.009
115	0.020	0.019	0.018
120	0.035	0.034	0.033
125	0.057	0.057	0.056
130	0.088	0.090	0.088
135	0.128	0.132	0.130
140	0.179	0.184	0.181
145	0.238	0.244	0.242
150	0.305	0.312	0.310
155	0.378	0.384	0.383
160	0.453	0.458	0.458
165	0.528	0.531	0.532
170	0.600	0.602	0.603
175	0.668	0.668	0.669
180	0.729	0.727	0.729
185	0.782	0.780	0.781
190	0.828	0.825	0.827
195	0.866	0.863	0.865
200	0.897	0.894	0.896
205	0.922	0.920	0.921
210	0.941	0.940	0.941
215	0.956	0.956	0.956
220	0.968	0.968	0.968
225	0.976	0.977	0.977
230	0.982	0.983	0.984
235	0.987	0.988	0.989
240	0.990	0.992	0.992
245	0.993	0.994	0.995
250	0.994	0.996	0.996

5.15. Låt S vara den totala skadekostnaden för portföljen: $S = X_1 + \cdots + X_N$, där N är antalet skador och

$\{X_k\}$ de enskilda skadebeloppen. Då gäller att

$$\mu = E[S] = \lambda\,\alpha_1, \qquad \sigma^2 = V[S] = \lambda\,\alpha_2, \qquad \gamma = \frac{E[(S-E[S])^3]}{\sigma^{3/2}} = \frac{\alpha_3}{\sqrt{\lambda}\,\alpha_2}.$$

NP-approximationen är

$$P(S \le x) = P\Big(Z \le \frac{x-\mu}{\sigma}\Big) \approx \Phi\Big(\sqrt{\frac{9}{\gamma^2} + \frac{6(x-\mu)}{\gamma\sigma} + 1} - \frac{3}{\gamma}\Big).$$

Vi söker C så att $P(S \le C) = 0.99$, så vi får att

$$\sqrt{\frac{9}{\gamma^2} + \frac{6(C-\mu)}{\gamma\sigma} + 1} - \frac{3}{\gamma} \approx q.$$

Löser man ut C, får man

$$C \approx \mu + q\sigma + \frac{\gamma\sigma(q^2-1)}{6},$$

dvs

$$C \approx \lambda\alpha_1 + q\sqrt{\lambda\alpha_2} + \frac{\alpha_3(q^2-1)}{6\alpha_2}.$$

Kapitel 6: Den individuella modellen

6.1. Se lösningen till uppgift 6.4.

6.2. Låt P och H vara de sannolikhetsgenererande funktionerna för X_1 respektive Y, dvs

$$P(u) = \sum_{k=0}^{\infty} p_k\,u^k, \qquad H(u) = \sum_{k=0}^{\infty} h_k\,u^k.$$

Då gäller ju att

$$H(u) = (P(u))^n.$$

Om vi logaritmerar och deriverar på båda sidor av detta uttryck, får vi att

$$P(u)\,H'(u) = n\,P'(u)\,H(u).$$

Om vi nu deriverar detta uttryck $k-1$ gånger på båda sidor, får vi, via Leibniz regel,

$$P(u)\,H^{(k)}(u) + \sum_{j=1}^{k-1}\binom{k-1}{j}P^{(j)}(u)\,H^{(k-j)}(u) = n\sum_{j=1}^{k-1}\binom{k-1}{j}P^{(j+1)}(u)\,H^{(k-1-j)}(u).$$

Sätter man $u = 0$, får man

$$p_0\,h_k\,k! + \sum_{j=1}^{k-1}\binom{k-1}{j}p_j\,j!\,h_{k-j}\,(k-j)! = n\sum_{j=0}^{k-1}\binom{k-1}{j}p_{j+1}\,(j+1)!\,h_{k-1-j}\,(k-1-j)!,$$

dvs

$$p_0\,h_k\,k + \sum_{j=1}^{k-1}p_j\,h_{k-j}\,(k-j) = n\sum_{j=0}^{k-1}p_{j+1}\,(j+1)\,h_{k-1-j},$$

dvs

$$p_0\, h_k\, k = n \sum_{j=1}^{k} j\, p_j\, h_{k-j} - \sum_{j=1}^{k} (k-j)\, p_j\, h_{k-j} = \sum_{j=1}^{k} \big((n+1)j - k\big) p_j\, h_{k-j}.$$

Delar man båda sidor med $p_0 k$, får man

$$h_k = \frac{1}{p_0} \sum_{j=1}^{k} \left(\frac{(n+1)j}{k} - 1 \right) p_j\, h_{k-j}.$$

6.3. Den sannolikhetsgenererande funktionen för S är

$$P(s) = \prod_{i=1}^{m} (q_i + p_i s)^{n_i}.$$

Om vi logaritmerar båda sidor och deriverar, får vi

$$\frac{P'(s)}{P(s)} = \sum_{i=1}^{m} n_i\, \frac{p_i}{q_i + p_i s}.$$

Högerledet kan skrivas som

$$\begin{aligned}
\sum_{i=1}^{m} n_i\, \frac{p_i}{q_i}\, \frac{1}{1 + \frac{p_i}{q_i}\, s} &= \sum_{i=1}^{m} n_i\, \frac{p_i}{q_i} \sum_{j=0}^{\infty} \left(-\frac{p_i}{q_i}\, s \right)^j \\
&= \sum_{i=1}^{m} n_i \sum_{j=0}^{\infty} (-1)^j \left(\frac{p_i}{q_i} \right)^{j+1} s^j \\
&= \sum_{j=0}^{\infty} \left((-1)^j \sum_{i=1}^{m} n_i \left(\frac{p_i}{q_i} \right)^{j+1} \right) s^j.
\end{aligned}$$

Om vi sätter

$$v_j := (-1)^j \sum_{i=1}^{m} n_i \left(\frac{p_i}{q_i} \right)^{j+1}$$

och

$$V(s) := \sum_{j=0}^{\infty} v_j\, s^j,$$

har vi alltså att

$$P'(s) = P(s)\, V(s).$$

Om vi deriverar båda sidor av detta uttryck $k-1$ gånger, ger Leibniz formel att

$$P^{(k)}(s) = \sum_{j=0}^{k-1} \binom{k-1}{j} V^{(j)}(s)\, P^{(k-1-j)}(s).$$

Sätter vi $s = 0$, får vi

$$g_k\, k! = \sum_{j=0}^{k-1} \frac{(k-1)!}{j!\,(k-j-1)!}\, v_j\, j!\, g_{k-1-j}\, (k-1-j)!,$$

dvs

$$g_k\,k = \sum_{j=0}^{k-1} v_j\, g_{k-1-j} = \sum_{j=1}^{k} v_{j-1}\, g_{k-j}.$$

Om vi sätter

$$w_j := v_{j-1} = (-1)^{j+1} \sum_{i=1}^{m} n_i \left(\frac{p_i}{q_i}\right)^j,$$

får vi den givna rekursionsformeln.

6.4. Här är mina beräkningar av $P(S \leq k)$ för några utvalda värden på k, via de Prils exakta algoritm, SaBi-approximationen, SaPo-approximationen, NP-approximationen av den approximerande SaBi-fördelningen (NP1), samt NP-approximationen med de rätta momenten för S (NP2).

k	de Pril	SaBi	SaPo	NP1	NP2
30	0.000338	0.000342	0.000356	0.000348	0.000344
40	0.004322	0.004354	0.004474	0.004126	0.004097
50	0.027213	0.027335	0.027785	0.025658	0.025542
60	0.101475	0.101718	0.102611	0.096511	0.096270
70	0.253527	0.253783	0.254719	0.244409	0.244143
80	0.465581	0.465673	0.465999	0.454482	0.454343
90	0.678482	0.678359	0.677898	0.668611	0.668729
100	0.838954	0.838727	0.837893	0.832194	0.832425
110	0.932788	0.932589	0.931864	0.929052	0.929257
120	0.976486	0.976367	0.975935	0.974772	0.974895
130	0.993046	0.992992	0.992798	0.992382	0.992438
140	0.998246	0.998227	0.998157	0.998026	0.998046

6.5. a) Vi ska visa att

$$x^\lambda\, y^{1-\lambda} \leq \lambda\, x + (1-\lambda)\, y,$$

för $x \geq 0$, $y \geq 0$, $0 < \lambda < 1$, med likhet då och endast då $x = y$. För $y = 0$ är olikheten uppenbar, så anta $y > 0$. Om vi delar på båda sidor med y och sätter $t := x/y$, kan vi skriva olikheten som

$$t^\lambda \leq \lambda t + 1 - \lambda.$$

Om vi sätter $\varphi(t) := 1 - \lambda + \lambda t - t^\lambda$, ska vi alltså visa att $\varphi \geq 0$, med likhet om och endast om $t = 1$. Man har att

$$\varphi'(t) = \lambda(1 - (1/t)^{1-\lambda}).$$

Eftersom $0 < \lambda < 1$, har vi att $\varphi'(t) < 0$ för $t < 1$ och $\varphi'(t) > 0$ för $t > 1$. Detta innebär att funktionen φ har ett strikt minimum i punkten $t = 1$, dvs $\varphi(t) \geq \varphi(1) = 0$, med likhet endast för $t = 1$.

b) Vi visade i a) att olikheten gäller för $m = 2$. Vi antar nu att olikheten gäller för $m = n - 1$ och ska visa att den då gäller för $m = n$:

$$\begin{aligned}
\prod_{i=1}^{n} x_i^{\lambda_i} &= \left(\prod_{i=1}^{n-1} x_i^{\lambda_i}\right) x_n^{\lambda_n} = \left(\prod_{i=1}^{n-1} x_i^{\lambda_i/(\lambda_1+\cdots+\lambda_{n-1})}\right)^{\lambda_1+\cdots+\lambda_{n-1}} x_n^{\lambda_n} \\
&\le \text{(enligt a))} \le (\lambda_1+\cdots+\lambda_{n-1}) \prod_{i=1}^{n-1} x_i^{\lambda_i/(\lambda_1+\cdots+\lambda_{n-1})} + \lambda_n\, x_n \\
&\le \text{(enligt induktionsantagandet)} \\
&\le (\lambda_1+\cdots+\lambda_{n-1}) \sum_{i=1}^{n-1} \frac{\lambda_i}{\lambda_1+\cdots+\lambda_{n-1}}\, x_i + \lambda_n\, x_n = \sum_{i=1}^{n} \lambda_i\, x_i.
\end{aligned}$$

c) Vi sätter $\varphi(s) := s - \log(1+s)$. Då gäller att

$$\varphi'(s) = 1 - \frac{1}{1+s} = \frac{s}{1+s}.$$

Vi ser att $\varphi'(s) < 0$ om $-1 < s < 0$ och $\varphi'(s) > 0$ om $s > 0$, så det följer att $\varphi(s) \ge \varphi(0) = 0$ om $s > -1$, med likhet om och endast om $s = 0$.

6.6. Vi räknar först ut väntevärdet och variansen för S. Man har att

$$E\big[X_{i1}^k\big] = E\big[X_{i1}^k \bigm| X_{i1} = 0\big] P(X_{i1} = 0) + E\big[X_{i1}^k \bigm| X_{i1} > 0\big] P(X_{i1} > 0) = \alpha_{ki}\, p_i.$$

Detta ger att

$$\begin{aligned}
E[S] &= \sum_{i=1}^{a} E[S_i] = \sum_{i=1}^{a}\sum_{j=1}^{n_i} E[X_{ij}] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i} \alpha_{1i}\, p_i = \sum_{i=1}^{a} n_i\, p_i\, \alpha_{1i},
\end{aligned}$$

och

$$\begin{aligned}
V[S] &= \sum_{i=1}^{a} V[S_i] = \sum_{i=1}^{a}\sum_{j=1}^{n_i} V[X_{ij}] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i} \big(E[X_{ij}^2] - (E[X_{ij}])^2\big) \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i} (p_i\, \alpha_{2i} - p_i^2\, \alpha_{1i}^2) = \sum_{i=1}^{a} n_i\, p_i\, (\alpha_{2i} - p_i\, \alpha_{1i}^2).
\end{aligned}$$

Vidare gäller att

$$E[S'] = n\, p\, \alpha_1 = \sum_{i=1}^{a} n_i\, p_i\, \alpha_{1i},$$

$$V[S'] = n\, p\, \alpha_2 - n\, p^2\, \alpha_1^2 = \sum_{i=1}^{a} n_i\, p_i\, \alpha_{2i} - \frac{1}{n}\left(\sum_{i=1}^{a} n_i\, p_i\, \alpha_{1i}\right)^2.$$

Enligt resultaten om den sammansatta Poissonfördelningen i avsnitt 5.2, har vi att

$$E[S''] = n\,p\,\alpha_1 = \sum_{i=1}^{a} n_i\,p_i\,\alpha_i,$$
$$V[S''] = n\,p\,\alpha_2 = \sum_{i=1}^{a} n_i\,p_i\,\alpha_{2i}.$$

Av ovanstående ser vi att

$$E[S] = E[S'] = E[S''].$$

Vidare får vi att

$$\begin{aligned}
V[S'] \quad - \quad V[S] &= \sum_{i=1}^{a} n_i\,p_i\,\alpha_{2i} - \frac{1}{n}\left(\sum_{i=1}^{a} n_i\,p_i\,\alpha_{1i}\right)^2 - \sum_{i=1}^{a} n_i\,p_i\,\alpha_{2i} + \sum_{i=1}^{a} n_i\,p_i^2\,\alpha_{1i}^2 \\
&= \sum_{i=1}^{a} n_i\,p_i^2\,\alpha_{1i}^2 - \frac{1}{n}\left(\sum_{i=1}^{a} n_i\,p_i\,\alpha_{1i}\right)^2 \\
&= n\left(\sum_{i=1}^{a} \frac{n_i}{n}\,p_i^2\,\alpha_{1i}^2 - \left(\sum_{i=1}^{a} \frac{n_i}{n}\,p_i\,\alpha_{1i}\right)^2\right).
\end{aligned}$$

Sätt $\lambda_i := n_i/n$, $x_i := p_i\,\alpha_{1i}$. Eftersom $\lambda_1 + \cdots + \lambda_a = 1$, gäller att

$$\begin{aligned}
\sum_{i=1}^{a} \lambda_i\,x_i^2 \quad - \quad \left(\sum_{i=1}^{a} \lambda_i\,x_i\right)^2 & \\
&= \sum_{i=1}^{a} \lambda_i\,x_i^2 \sum_{j=1}^{a} \lambda_j - \sum_{i=1}^{a}\sum_{j=1}^{a} \lambda_i\,\lambda_j\,x_i\,x_j \\
&= \sum_{i=1}^{a}\sum_{j=1}^{a} \lambda_i\,\lambda_j\,x_i^2 - \sum_{i=1}^{a}\sum_{j=1}^{a} \lambda_i\,\lambda_j\,x_i\,x_j \\
&= \sum_{i=1}^{a} \lambda_i^2\,x_i^2 + \sum_{i<j} \lambda_i\,\lambda_j\,(x_i^2 + x_j^2) - \sum_{i=1}^{a} \lambda_i^2\,x_i^2 - \sum_{i<j} 2\,\lambda_i\,\lambda_j\,x_i\,x_j \\
&= \sum_{i<j} \lambda_i\,\lambda_j\,(x_i - x_j)^2 \\
&= \frac{1}{2}\sum_{i\neq j} \lambda_i\,\lambda_j\,(x_i - x_j)^2
\end{aligned}$$

Man får alltså att

$$\begin{aligned}
V[S'] \quad - \quad V[S] &= \frac{n}{2}\sum_{i\neq j} \frac{n_i}{n}\,\frac{n_j}{n}\big(p_i\,\alpha_{1i} - p_j\,\alpha_{1j}\big)^2 \\
&= \frac{1}{2n}\sum_{i\neq j} n_i\,n_j\big(p_i\,\alpha_{1i} - p_j\,\alpha_{1j}\big)^2.
\end{aligned}$$

Slutligen ser vi direkt att

$$V[S''] - V[S'] = \frac{1}{n}\left(\sum_{i=1}^{a} n_i\,p_i\,\alpha_{1i}\right)^2.$$

8 Referenser

ter Berg, P. (1994): Deductibles and the Inverse Gaussian Distribution, *Astin Bulletin* **24**, 319–323.

Buchwalder, M., Chevalier, E. & Klüppelberg, C. (1993): Approximation methods for the total claimsize distribution — an algorithmic and graphical presentation, *Mitteilungen der Schweiz. Vereinigung der Versicherungsmathematiker*, 187–227.

Bühlmann, H. (1970): *Mathematical Methods in Risk Theory*, Springer-Verlag.

Daykin, C., Pentikäinen, T. & Pesonen, M. (1994): *Practical Risk Theory for Actuaries*, Chapman & Hall.

Dhaene, J. & Vandebroek, M. (1994): Recursions for the individual model, *Insurance: Mathematics and Economics* **16**, 31–38.

Hesselager, O. (1996): A recursive procedure for calculation of some mixed compound Poisson distributions, *Scandinavian Actuarial Journal*, 54–63.

Hogg, R. & Klugman, S. (1984): *Loss Distributions*, Wiley.

Hosking, J.R.M. & Wallis, J.R. (1987): Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.

Kuon, S., Radtke, M. & Reich, A. (1993): An appropriate way to switch from the individual risk model to the collective one, *Astin Bulletin*, **23**, 23–54.

van der Laan, B.S. & Hop, J.P. (1989): Probability distributions for the amount of damage, *Proceedings of the XIIth ASTIN Colloquium in New York*, 465–511.

Panjer, H. & Willmot, G. (1992): *Insurance Risk Models*, Society of Actuaries.

Rootzén, H. & Tajvidi, N. (1995): Extreme value statistics and wind storm losses: a case study, *Technical report 1995:5*, Avd för matematisk statistik, Chalmers Tekniska Högskola, Göteborg.

Shuster, J.J. (1968): On the inverse Gaussian distribution function, *J. Amer. Statist. Ass.* **63**, 1514–1516.

Ströter (1985): The numerical evaluation of the aggregate claim density function via integral equations, *Blätter der Deutsche Gesellschaft für Versicherungsmathematik*, **17**, 1–14.

Willmot, G. (1993): On recursive evaluation of mixed Poisson probabilities and related quantities, *Scandinavian Actuarial Journal*, 114–133.