UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2017:22

Sakförsäkring och approximation av totalt skadebelopp

Oskar Vedin

Examensarbete i matematik, 15 hp
Handledare: Rolf Larsson
Examinator: Jörgen Östensson
Juni 2017

Department of Mathematics
Uppsala University

Sakförsäkring och approximation av totalt skadebelopp

Oskar Vedin

31 maj 2017

Matematiska Institutionen
Uppsala Universitet

Sammanfattning

Första delen består av teorin bakom den kollektiva modellen inom sakförsäkring. Teorin börjar med att visa hur man kan modellera antalet inträffade skador med en Poissonprocess och hur den kan generaliseras. Sedan presenteras Paretofördelningen samt Lognormal-fördelningen som modeller för de individuella skadebeloppen. Teorin avslutas med den kollektiva modellen för det totala skadebeloppet, samt tre olika metoder för att approximera fördelningen för det totala skadebeloppet.

I andra delen tillämpas sedan teorin på ett datamaterial tillsammans med en jämförelse av de olika approximationsmetoderna.

Innehåll

1 Inledning

En av de viktigaste kvantiteterna för ett sakförsäkringsbolag är storleken av det totala skadebeloppet för en försäkringsportfölj under en viss tidsperiod. Hur mycket kommer företaget att behöva betala kunderna för skadorna dom drabbats av? Vet man hur stora kostnaderna är, vet man även hur mycket kapital man behöver för att täcka kostnaderna. Detta är en av anledningarna till att man behöver veta hur mycket man förväntas betala kunderna.

Hur stora kostnaderna är bestäms såklart av hur många skador som inträffar under tidsperioden och hur mycket varje skada kostar, något som är slumpmässigt och i förväg inte går att bestämma. Det finns två olika typer av modeller man kan använda sig av för att beskriva detta.

Den *individuella modellen* utgår från varje enskilt försäkringskontrakt i en portfölj och summerar sedan kostnaderna för varje försäkringskontrakt till en summa $S = \sum_{i=1}^{n} Y_i$, där Y_i är kostnaden för kontrakt i och n är antalet försäkringskontrakt. Y_i är alltså inte kostnaden för en skada utan den totala kostnaden för alla skador som drabbar försäkringstagare i. I den här modellen kommer många av termerna vara lika med noll och man gör här antagandet att $Y_1, \ldots, Y_n$ är oberoende. [2]

Den *kollektiva modellen* presenterades 1903 av den svenska aktuarien Filip Lundberg och lade grunden för den moderna sakförsäkringsmatematiken. [3] Här väljer man att istället utgå från kostnaderna för varje enskild skada, oberoende av vilken försäkringstagare som drabbas av den.

Om det under en tidsperiod inträffar N stycken skador och kostnaden för skada i är X_i blir det totala skadebeloppet $S = \sum_{i=1}^{N} X_i$. Till skillnad från den individuella modellen kommer ingen av termerna vara lika med noll. Här antar man att N och X_i är oberoende, samt att $X_1, \ldots, X_N > 0$, är oberoende och likafördelade. I praktiken innebär det att man studerar en försäkringsportfölj med liknande risker.

Vad har S för fördelning? Hur kan man approximera S på ett bra sätt?

1.1 Definitioner

N = antalet skador.
X_i = skadebelopp för skada i.
S = det totala skadebeloppet.

Definition 1.1 *Låt X vara en kontinuerlig stokastisk variabel med täthets-*

funktion $f_X(x)$. *Det k:te momentet för X definieras som (om det existerar):*

$$E\left[X^k\right] = \int_{-\infty}^{\infty} x^k f_X(x)dx$$

Och för det diskreta fallet:

$$E\left[X^k\right] = \sum_{-\infty}^{\infty} x^k p_X(x)$$

där $p_X(x)$ *sannolikhetsfunktionen för X.*

Definition 1.2 (Väntevärde, Varians, Standardavvikelse och Skevhet) *Låt X vara en stokastisk variabel. Väntevärdet för X definieras som:*

$$\mu_X = E\left[X\right]$$

Variansen för X definieras som:

$$\sigma_X^2 = E\left[(X - \mu_X)^2\right] = E\left[X^2\right] - \mu_X^2$$

Standardavvilkelsen för X definieras som:

$$\sigma_X = \sqrt{\sigma_X^2}$$

Skevheten för X definieras som:

$$\gamma_X = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^3\right] = \frac{E\left[X^3\right] - 3\mu_X\sigma_X^2 - \mu_X^3}{\sigma_X^3}$$

2 Teori

2.1 Antal skador

Om skador inträffar slumpmässigt vid tidpunkter T_i med $0 < T_1 \leq T_2 \leq \ldots$ kan det totala antalet skador ses som en räkneprocess $\{N(t), t \geq 0\}$, där $N(t) = \#\{i \geq 1\colon T_i \leq t\}\ \ t \geq 0$.

Det går att modellera en sådan räkneprocess med en homogen Poissonprocess, vilken även ligger till grund för mer generella modeller.

Definition 2.1 (Poissonfördelning) *En diskret slumpvariabel Z med sannolikhetsfunktion*

$$p_Z(k) = P(Z = K) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, 2, \ldots$$

sägs vara Poissonfördelad med parameter $\lambda > 0$. $Z \sim Po(\lambda)$.

Definition 2.2 (Exponentialfördelning) *En kontinuerlig slumpvariabel Z med täthetsfunktion*

$$f_Z(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

och fördelningsfunktion

$$F_Z(x) = P(Z \leq x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x} \quad x \geq 0$$

sägs vara Exponentialfördelad med parameter, eller intensitet, $\lambda > 0$. $Z \sim Exp(\lambda)$.

2.1.1 Poissonprocess

En homogen Poissonprocess kan definieras på följande sätt.

Definition 2.3 (Homogen Poissonprocess) *En räkneprocess $\{N(t), t \geq 0\}$ är en homogen Poissonprocess med intensitet λ om:*

1. $N(0) = 0$
2. $N(t)$ *har oberoende ökningar*
3. $N(s, t) = N(t) - N(s) \sim Po(\lambda(t - s)), \quad s < t$

Med oberoende ökningar menas att $N(t_1), N(t_1, t_2), N(t_2, t_3), \cdots, N(t_{n-1}, t_n)$ är oberoende och $0 \leq t_1 < t_2 < \cdots < t_n$. Antalet inträffade händelser i ett tidsintervall $(t_i, t_{i+1}]$ är alltså oberoende av antalet händelser i tidigare tidsintervall.

3. innebär att längden på tidsintervallet, inte dess position, bestämmer fördelningen för $N(t)$. Om man vill studera antalet inträffade händelser under en tidsperiod t räcker det med att studera $N(0,t) = N(t) - N(0) = N(t)$ där t är längden på den tidsintervallet.

Sats 2.1 (Väntevärde, Varians) *Låt N(t) vara en Poissonprocess.*

$$\mu_{N(t)} = \lambda t$$

$$\sigma^2_{N(t)} = \lambda t$$

Låt T_1 vara tiden till den första händelsen inträffar och $s > 0$. Att $T_1 > s$ är detsamma som $N(s) = 0$. Om den första händelsen inträffar efter tidpunkt s måste det totala antalet händelser som inträffat vid tidpunkt s vara 0. Alltså, $\{T_1 > s\} = \{N(s) = 0\}$.

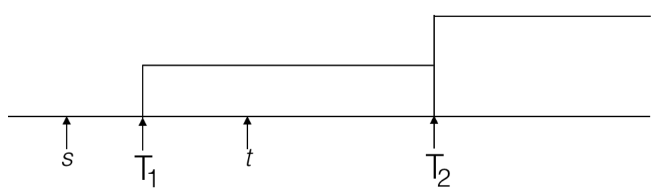
$$P(T_1 > s) = P(N(s) = 0) = \frac{(\lambda s)^0 e^{-\lambda s}}{0!} = e^{-\lambda s} \Rightarrow P(T_1 \leq s) = 1 - e^{-\lambda s}$$

Tiden till den första händelsen inträffar är alltså Exponentialfördelad med parameter λ.

Låt $U_2 = T_2 - T_1$ vara tiden mellan den första och den andra händelsen, $t > 0$. Om den första händelsen inträffar vid tidpunkt s innebär det att $N(s) = 1$ och om $U_2 > t$ så är $N(s, s+t) = 0$, ingen händelse inträffar mellan s och $s + t$.

$$P(U_2 > t | T_1 = s) = P(N(s, s+t) = 0 | N(s) = 1) = P(N(s, s+t) = 0) = P(N(t) = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t} \Rightarrow P(U_2 \leq t) = 1 - e^{-\lambda t}.$$

Längden på tidsintervallet mellan den första och den andra händelsen inträffar är Exponentialfördelad med parameter λ.



Figur 1

På samma sätt går det att visa att alla tidsintervall mellan händelserna är Exponentialfördelade med parameter λ genom att betinga på tidigare inträffade händelser och utnyttja att Poissonprocessen har oberoende ökningar.

Det kan finnas situationer när det inte är lämpligt att låta parametern λ vara konstant över hela tidsperioden, utan istället låta den variera med tiden, fördelningen ser då ut på följande sätt:

$$N(s,t) \sim Po\left(\int_s^t \lambda(s)ds\right)$$

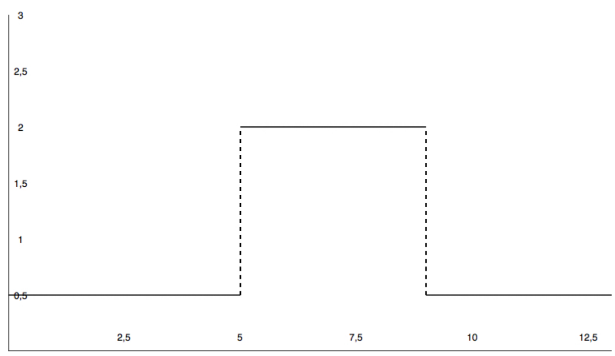
Här går det inte att enbart kolla på $N(t)$ som i den homogena Poissonprocessen eftersom Poissonfördelningen kan se annorlunda ut beroende på tidsintervallets position.

Exempel 2.1 (Skogsbränder) *Ett exempel på när det kanske skulle kunna lämpligt med en inhomogen Poissonprocess, är som modell för antalet skogsbränder under ett år.*

Låt säga att man från historisk data vet att det inträffar i genomsnitt 2 bränder/månad mellan Maj och Augusti och 0.5 bränder/månad resterande månader av året. Vi definierar $\lambda(s)$:

$$\lambda(s) = \begin{cases} 2 & \text{när } 5 \leq s < 9 \\ 0.5 & \text{när } \{0 \leq s < 5\} \cup \{9 \leq s < 13\} \end{cases}$$

Om man är intresserad av antalet bränder mellan första April och sista Juni



Figur 2: $\lambda(s)$

får vi parametern i Poissonfördelningen:

$$\int_4^7 \lambda(s)ds = 0.5\int_4^5 ds + 2\int_5^7 ds = 0.5 * 1 + 2 * 2 = 4.5$$

Det förväntande antalet skogsbränder mellan första April och sista Juni är 4.5 stycken, där $N(4,7) \sim Po(4.5)$.

Om man istället hade använt en homogen Poissonprocess som modell hade man fått att $\lambda(s) = 1/12$ *och att det förväntade antalet bränder skulle vara 0.25, där* $N(4,7) = N(3) \sim Po(0.25)$. *En kraftig underskattning.*

2.1.2 Blandad Poissonprocess

Ett till sätt att generalisera Poissonprocessen är att låta parametern i Poissonfördelningen vara en stokastisk variabel, Λ med täthetsfunktion $g(\lambda) > 0$ för alla $\lambda > 0$. Den betingade fördelningen $(N|\Lambda = \lambda) \sim Po(\lambda)$, men det är inte säkert att den obetingade fördelningen är Poisson som följande exempel visar.

Exempel 2.2 *Låt N(1)=N ha en blandad Poissonfördelning med* $\Lambda \sim \Gamma(\alpha, \beta)$ $\alpha, \beta > 0,\ \alpha \in \mathbb{Z}^+$.
Enligt lagen om total sannolikhet:

$$P(N = n) = \int_0^\infty P(N = n|\Lambda = \lambda) g(\lambda) d\lambda$$

$$= \frac{\beta^\alpha}{n!\Gamma(\alpha)} \frac{\Gamma(n+\alpha)}{(1+\beta)^{n+\alpha}} \int_0^\infty \frac{(1+\beta)^{n+\alpha}}{\Gamma(n+\alpha)} \lambda^{n+\alpha-1} e^{-(\lambda(1+\beta))} d\lambda$$

Om vi utnyttjar det faktum att $\Gamma(\alpha) = (\alpha - 1)!$ *och att den sista integralen är* $\int_0^\infty g(\lambda) d\lambda = 1$ *får vi:*

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

Med $p = \dfrac{1}{1+\beta}$*:*

$$P(N = n) = \binom{n+\alpha-1}{n} p^n (1-p)^\alpha$$

$N \sim NegBin(\alpha, p)$. [2]

2.1.3 Förnyelsprocess

I Poissonprocessen är tidsintervallen mellan händelserna exponentialfördelade som vi såg tidigare. Om man antar att tidsintervallen är oberoende och likafördelade men inte exponentialfördelade får man istället en förnyelseprocess. Det finns situationer då exponentialfördelade tidsintervall inte är ett rimligt antagande och en förnyelseprocess är mer lämpligt. Mikosch ger exempel på stormskador som kan inträffa mellan väldigt långa tidsperioder, då är det mer lämpligt att använda en mer långsvansad fördelning än exponentialfördelningen. [3]

2.2 Skadebelopp

När en skada drabbar försäkringstagaren har den rätt till ersättning från försäkringsbolaget, givet att skadan täcks av villkoren. I den kollektiva modellen är X_i det belopp försäkringsbolget betalar ut för skada i i en portfölj med liknande risker. Man gör antagandet att skadebeloppen, X_i, är oberoende och likafördelade.

Det som karaktäriserar fördelningarna för skadebeloppen är att dom har en tjock högersvans, det finns en inte försumbar sannolikhet att en skada kan ge upphov till en väldigt stor kostnad.

Det finns givetvis en mängd olika fördelningar som kan användas för att modellera skadekostnaden för en enskild skada. Jag har valt att använda mig av två stycken vanliga fördelningar som förekommer i sakförsäkringsmatematiken.

2.2.1 Paretofördelningen

Definition 2.4 (Paretofördelning) *En kontinuerlig slumpvariabel X med täthetsfunktion*

$$f_X(x) = \begin{cases} \dfrac{\alpha\beta^\alpha}{x^{\alpha+1}} & x \geq \beta \\ 0 & x < \beta \end{cases}$$

och fördelningsfunktion

$$F_X(x) = \int_\beta^x \frac{\alpha\beta^\alpha}{y^{\alpha+1}} dy = 1 - \left(\frac{\beta}{x}\right)^\alpha$$

är Paretofördelad med parametrar $\alpha, \beta > 0$. $X \sim Pa(\alpha, \beta)$.

Momenten för paretofördelningen ges av:

$$E\left[X^k\right] = \int_\beta^\infty x^k \frac{\alpha\beta^\alpha}{x^{\alpha+1}} dx = \frac{\alpha\beta^k}{\alpha - k}, \quad \alpha > k$$

Sats 2.2 (Väntevärde, varians) *Låt X vara en paretofördelad stokastisk variabel. Då gäller:*

$$\mu_X = \frac{\alpha\beta}{\alpha - 1}, \quad \alpha > 1$$

$$\sigma_X^2 = \frac{\alpha\beta^2}{(\alpha-2)(\alpha-1)^2}, \quad \alpha > 2$$

Bevis: Appendix.

Parameterskattning, momentmetoden

Vi använder första momentet och andra momentet från den empiriska fördelningen för att skatta parametrarna med momentmetoden.

$$m_1 = \bar{x} = \frac{1}{n}\sum_1^n x_i \qquad m_2 = \frac{1}{n}\sum_1^n x_i^2$$

Parametrarna i fördelningen fås genom att lösa:

$$\begin{cases} m_1 = \dfrac{\alpha\beta}{\alpha - 1} \\ m_2 = \dfrac{\alpha\beta^2}{\alpha - 2} \end{cases}$$

$$\widehat{\alpha} = \frac{m_1^2}{m_1^2 - m_2 \pm \sqrt{m_2\left(m_2 - m_1^2\right)}} \qquad \widehat{\beta} = \frac{m_2 \pm \sqrt{m_2\left(m_2 - m_1^2\right)}}{m_1}$$

För att denna metod ska vara giltig krävs att $m_1^2 \leq m_2$. Att skatta β på det här sättet är inte att rekommendera då man kan få att $\widehat{\beta} > \min_i x_i$. Men om man på förhand vet vad β är behöver man bara skatta α. Detta kan göras genom att lösa:

$$m_1 = \frac{\alpha\beta}{\alpha - 1} \Rightarrow \widehat{\alpha} = \frac{m_1}{m_1 - \beta}$$

En till nackdel med momentmetoden är att andra moment inte existerar om $\alpha \leq 2$ och metoden går då inte att använda.

Exempel 2.3 *Från 500 genererade tal från en Paretofördelning med $\alpha = 4$ och $\beta = 100$ har vi:*

$$m_1 = \bar{x} \approx 132.6314 \quad m_2 \approx 19630 \quad \min_i x_i \approx 100$$

Om vi nu skattar parametrarna får vi att:

$$\hat{\alpha} \approx 4.1029 \quad \hat{\beta} \approx 195.7010$$

Med den anpassade fördelningen har vi att $P(X \leq 195.7010) = 0$ trots att vi har observationer under $\widehat{\beta}$.

Ett bättre sätt att skatta β är att helt enkelt sätta det till det minsta observerade värdet, detta är också ML-skattningen av β.

Parameterskattning, ML

Med maximum likelihood-metoden skattas parametrarna genom välja det α och β som maximerar likelihoodfunktionen:

$$\mathcal{L}(\alpha, \beta; x) = \prod_1^n \frac{\alpha\beta^\alpha}{x_i^{\alpha+1}} = \alpha^n \beta^{n\alpha} \prod_1^n x_i^{-(\alpha+1)}$$

Log-likelihoodfunktionen:

$$\ell(\alpha, \beta; x) = \log\left(\mathcal{L}(\alpha, \beta; x)\right) = n\log(\alpha) + n\alpha\log(\beta) - (\alpha+1)\sum_1^n \log(x_i)$$

Om man deriverar funktionen med avseende på α och sätter den till noll får man:

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + n\log(\beta) - \sum_1^n \log(x_i) = 0$$

$$\hat{\alpha} = \frac{n}{\sum_1^n \log(x_i) - n\log(\beta)} = \frac{n}{\sum_1^n (\log(x_i) - \log(\beta))}$$

ℓ är strikt växande med β, vi maximerar således ℓ genom att sätta β till det största möjliga värdet. Eftersom $\beta \leq x_i$ för alla i sätter vi β till den minsta observationen:

$$\hat{\beta} = \min_i x_i$$

Exempel 2.4 *Med samma datamaterial som i exempel 2.3 skattas parametrarna till:*

$$\widehat{\alpha} \approx 4.0901 \quad \widehat{\beta} \approx 100$$

ML-skattningarna är i det här fallet bättre än skattningarna med momentmetoden.

ML-metoden är rekommenderad när man vill skatta parametrarna i en Paretofördelning. Det går att visa att ingen av dessa parameterskattningar är väntevärdesriktiga. [5].

2.2.2 Lognormal-fördelning

Definition 2.5 (Lognormal-fördelning) *Låt $Y \sim N\left(\mu_Y, \sigma_Y^2\right)$ och $X = e^Y$. X är då Lognormal-fördelad med täthetsfunktion:*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_Y x} \exp\left\{-\frac{(\ln(x) - \mu_Y)^2}{2\sigma_Y^2}\right\}, \quad x > 0$$

och fördelningsfunktion:

$$F_X(x) = \Phi\left(\frac{\ln(x) - \mu_Y}{\sigma_Y^2}\right), \quad x > 0$$

där Φ är fördelningsfunktionen för en $N(0,1)$-fördelad stokastisk variabel. $X \sim logN(\mu_Y, \sigma_Y^2)$ med parameter $\mu_Y \in \mathbb{R}, \sigma_Y^2 > 0$.

Att täthets- och fördelningsfunktionen ser ut som dom gör följer av:

$$F_X(x) = P(X \leq x) = P(e^Y \leq x) = P(Y \leq \ln(x)) = F_Y(\ln(x))$$

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}F_Y\left(\ln(x)\right) = f_Y\left(\ln(x)\right)\frac{d}{dx}\ln(x) = f_Y\left(\ln(x)\right)\frac{1}{x}$$

I Lognormal-fördelningen är det väntevärdet och variansen för den associerade normalfördelningen som är parametrarna. Dessa ska inte förväxlas med väntevärdet och variansen för den Lognormal-fördelade variabeln ($\mu_X \neq \mu_Y,\ \sigma_X^2 \neq \sigma_Y^2$).

Momenten för en Lognormal-fördelningen ges av:

$$E\left[X^k\right] = E\left[e^{kY}\right] = m_Y(k) = e^{k\mu_Y + (k\sigma_Y)^2/2}$$

Där m_Y är den momentgenerarnade funktionen för $Y \sim N\left(\mu_Y, \sigma_Y^2\right)$.

Sats 2.3 (Väntevärde, Varians) *Låt* $X \sim logN(\mu_Y, \sigma_Y^2)$*. Då gäller:*

$$\mu_X = e^{\mu_Y} e^{\sigma_Y^2/2}$$

$$\sigma_X^2 = e^{2\mu_Y} e^{\sigma_Y^2}\left(e^{\sigma_Y^2} - 1\right)$$

Bevis: Appendix.

Parameterskattning, momentmetoden

Parametrarna fås genom att lösa:

$$\begin{cases} m_1 = e^{\mu_Y + \sigma_Y^2/2} \\ m_2 = e^{2\mu_Y + 2\sigma_Y^2} \end{cases}$$

$$\widehat{\mu_Y} = 2\ln(m_1) - \frac{\ln(m_2)}{2} \qquad \widehat{\sigma_Y^2} = \ln(m_2) - 2\ln(m_1)$$

Parameterskattning, ML

Likelihoodfunktionen för Lognormal-fördelningen:

$$\mathcal{L}(\mu_Y, \sigma_Y^2; x) = \prod_1^n \frac{1}{\sqrt{2\pi}\sigma_Y x_i} \exp\left\{-\frac{(\ln(x_i) - \mu_Y)^2}{2\sigma_Y^2}\right\}$$

Log-likelihood:

$$\ell(\mu_Y, \sigma_Y^2; x) = -\sum_1^n \ln(x_i) - n\ln(\sqrt{2\pi}\sigma_Y) - \frac{1}{2\sigma_Y^2}\sum_1^n (\ln(x_i) - \mu_Y)^2$$

Parametrarna skattas genom att lösa:

$$\begin{cases} \dfrac{\partial \ell}{\partial \mu_Y} = \dfrac{1}{\sigma_Y^2}\left(\displaystyle\sum_1^n \ln(x_i) - n\mu_Y\right) = 0 \\ \dfrac{\partial \ell}{\partial \sigma_Y} = -\dfrac{n}{\sigma_Y} + \dfrac{1}{\sigma_Y^3}\displaystyle\sum_1^n (\ln(x_i) - \mu_Y)^2 = 0 \end{cases}$$

$$\widehat{\mu_Y} = \frac{1}{n}\sum_1^n \ln(x_i) \qquad \widehat{\sigma_Y^2} = \frac{1}{n}\sum_1^n (\ln(x_i) - \widehat{\mu_Y})^2$$

Då $X = e^Y$ får man att:

$$\widehat{\mu_Y} = \frac{1}{n}\sum_1^n y_i \quad \widehat{\sigma_Y^2} = \frac{1}{n}\sum_1^n (y_i - \widehat{\mu_Y})^2$$

Vilket är den vanliga ML-skattningen för normalfördelningen. $\widehat{\mu_Y}$ är en väntevärdesriktig skattning då $E[\widehat{\mu_Y}] = \mu_Y$, men $E[\widehat{\sigma_Y^2}] = \frac{n}{n-1}\sigma_Y^2$ är inte en väntevärdesriktig skattning. Istället kan man använda:

$$\widehat{\sigma_Y^2} = \frac{1}{n-1}\sum_1^n (\ln(x_i) - \widehat{\mu_Y})^2$$

2.3 Sammansatt fördelning

Vi kan nu "sätta ihop" modellerna för antalet skador och de enskilda skadebeloppen till en modell för det totala skadebeloppet för portföljen:

$$S(t) = \sum_{i=1}^{N(t)} X_i$$

En stokastisk variabel som är en summa av ett slumpmässigt antal stokastiska variabler har en *sammansatt fördelning.* I fortsättningen kommer jag anta att antalet skador är Poissonfördelat och S har då en *Sammansatt Poissonfördelning.*

Sats 2.4 (Väntevärde, Varians, Skevhet) *Låt S vara en stokastisk variabel med en Sammansatt Poissonfördelning där N är antalet skador och X_i är skadebeloppen. Då gäller:*

$$\mu_S = \mu_X \mu_N$$

$$\sigma_S^2 = \sigma_N^2 E\left[X^2\right] = \sigma_N^2\left(\sigma_X^2 + \mu_X^2\right)$$

$$\gamma_S = \frac{E\left[X^3\right]}{\sqrt{\mu_N E\left[X^2\right]^3}}$$

Bevis: Appendix.

Fördelningsfunktionen för S kan skrivas:

$$F_S(x) = \sum_{n=0}^{\infty} p_n F_X^{n*}(x)$$

Där $p_n := P(N = n)$ och $F_X^{n*}(x)$ är fördelningsfunktionen för summan av n skador och kan beräknas med faltning. För stora n blir det dock väldigt bökigt att använda sig av faltning.

Det går att numerisk beräkna fördelningen för S med hjälp av Ströters rekursionsformel och Panjers rekursionsformel. Dessa bygger på det rekursiva sambandet i faltningsformlerna, men jag går inte närmare in på dessa.

Sats 2.5 (Momentgenererande funktion) *Låt S vara en stokastisk variabel med en Sammansatt Poissonfördelning. Dess momentgenererande funktion är då:*

$$m_X(t) = m_N(log(m_X(t))$$

Där $m_N(t)$ är den momentgenererande funktionen för antalet skador och $m_X(t)$ är den momentgenererande funktionen för skadebeloppen.

Bevis: Appendix.

Sats 2.6 (Karaktäristisk funktion) *Låt S vara en stokastisk variabel med en Sammansatt Poissonfördelning. Dess karaktäristiska funktion är då:*

$$\chi_S(s) = exp\left[\mu_N \chi_X(s) - \mu_N\right]$$

Där $\chi_X(s)$ är den karaktäristiska funktionen för dom individuella skadebeloppens fördelning.

Bevis: Appendix.

Exempel 2.5 *Antag N är en geometriskt fördelad stokastisk variabel med parameter $0 < p < 1$ och $X_i \sim Exp(1)$. Vi har dom momentgenererande funktionerna för N och X:*

$$m_N(t) = E\left(e^{tN}\right) = \sum_{n=0}^{\infty} e^{tn} p(1-p)^n = p \sum_{n=0}^{\infty} (qe^t)^n = \frac{p}{1-qe^t}, \quad t < -log(q)$$

$$m_X(t) = E\left(e^{tX}\right) = \int_0^{\infty} e^{tx} e^{-x} dx = \int_0^{\infty} e^{x(t-1)} dx = -\frac{1}{t-1} = \frac{1}{1-t}, \quad t < 1$$

Den momentgenererande funktionen för S blir då:

$$m_S(t) = m_N(\log(m_X(t)) = \frac{p}{1-qe^{\log m_X(t)}} = \frac{p}{1-qm_x(t)} = \frac{p-pt}{1-t-q}$$

$$= p + q\frac{p}{p-t}$$

Den momentgenererande funktionen för S är alltså en blandning av den momentgenererande funktionen för konstanten 0 och en exponentialfördelad slumpvariabel med parameter $\lambda = p$. Fördelningsfunktionen blir således $F(x) = p + q(1 - e^{-px}) = 1 - qe^{-px}$.

Det här är det ända fallet då fördelningen för S har en sluten form. [4]

2.3.1 Approximation av S

Sats 2.7 (Centrala gränsvärdessatsen) *Låt $X_1, X_2, \ldots, X_n$ vara oberoende och likafördelade slumpvariabler med $E[X_i] = \mu$, $\sigma^2 < \infty$ och $S = \sum_0^n X_i$. Då gäller att:*

$$\lim_{n\to\infty} P\left[\frac{S - n\mu_X}{\sqrt{n}\sigma_X} \le x\right] = \Phi(x)$$

Där $\Phi(x)$ är fördelningsfunktionen för en normalfördelad stokastisk variabel med väntevärde 0 och varians 1.

Då S är en summa av oberoende och likafördelade stokastiska variabler är det naturligt att i ett första steg använda Φ för att approximera S. Normalapproximationen innebär att $Z = \frac{S-\mu_S}{\sigma_S}$ är approximativt normalfördelad med $\mu_Z = 0\sigma_Z = 1$ när μ_N är "stort".

Fördelningen för skadebeloppen brukar normalt ha en tjock högersvans vilket kan leda till att $\gamma_S > 0$. Centrala gränsvärdessatsen säger att summan av n oberoende och likafördelade slumpvariabler är normalfördelad då $n \to \infty$, men skevheten för en normalfördelad stokastisk variabel är 0. Därför kan det krävas ett väldigt stort n (μ_N) för att $\gamma_S \approx 0$ och normalapproximationen ska ge en tillfredsställande noggrannhet.

Man kan istället använda sig av andra approximationsmetoder som tar hänsyn till skevheten hos S. En vanlig approximationsmetod är NP-approximaition eller Normal Power-approximation. Den bygger på Edgeworthserien, vilken också kan användas som approximation.

Edgeworthexpansion

Vi låter $Z = \frac{S-\mu_S}{\sigma_S}$ och $z = \frac{x-\mu_S}{\sigma_S}$, Edgeworthexpansionen ser då ut på följande sätt:

$$F_S(x) = \Phi(z) - \Phi^{(3)}(z)\frac{\mu_N E\left[X^3\right]}{6\sigma_S^3} + \Phi^{(4)}(z)\frac{\mu_N E\left[X^4\right]}{24\sigma_S^4} + \Phi^{(6)}(z)\frac{\mu_N^2 E\left[X^3\right]^2}{72\sigma_S^6} + \ldots$$

Där $\Phi^{(k)}(z)$ är den k:te derivatan av fördelningsfunktionen för en normalfördelad stokastisk variabel med väntevärde 0 och varians 1.

Edgeworthexpansionen kommer från den karaktäristiska funktionen för den Sammansatta Poissonfördelningen $\chi(s) = e^{\mu_N \psi_X(s) - \mu_N}$, där man serieutvecklar $\psi_X(s)$ som är den karaktäristiska funktionen för X. Full härledning finns i appendix. Genom att endast använda första termen i Edgeworthserien för att approximera $F_S(x)$ får man den vanliga normalapproximationen. Använder man de två första termerna i serien får man en Edgeworthapproximation av andra ordningen:

$$F_S(x) \approx \Phi(z) - \frac{\gamma_S}{6}\Phi^{(3)}(z)$$

Edgeworth ger en bra approximation upp till två standardavvikelser från medelvärdet. Men är man intresserad av svansen på fördelningen för S är inte det här den bästa metoden. [1]

NP-approximation

Ett annat sätt att approximera fördelningsfunktionen för S är med hjälp av NP-approximationen. Tanken är att man vill justera argumentet i $F_Z(z)$ så att $F_Z(z + \Delta(z)) \approx \Phi(z)$. I appendix finns en härledning som visar att man med hjälp av Edgeworthserien kommer man fram till att:

$$\Delta(z) = \frac{\gamma_S}{6}(z^2 - 1)$$

Och NP-approximationen blir:

$$F_S(x) \approx \Phi\left(\sqrt{\frac{9}{\gamma_S^2} + \frac{6(x - \mu_S)}{\sigma_S \gamma_S} + 1} - \frac{3}{\gamma_S}\right), \quad x > \mu_S - \frac{3\sigma_S}{\gamma_S}$$

Villkoret kommer egentligen aldrig ställa till problem om man är intresserad av värden på $F(x)$ för stora x. Det beror på att $\lim_{\gamma_S \to \infty} \mu_S - \frac{3\sigma_S}{\gamma_S} = \mu_S$ och vi kan som "sämst" använda approximationen för $x > \mu_S$. I fallet då γ_S är väldigt liten kan vi lika väl använda normalapproximationen.

3 Exempel

Här ges nu ett exempel samt en jämförelse mellan normalapproximationen, Edgeworthapproximationen och NP-approximationen på verklig data. Jag antar i fortsättningen att antalet skador är Poissonfördelat.

3.1 Data

Det datamaterial jag valde att använda mig av bestod av historisk data över fordonsförsäkringar från ett amerikanskt försäkringsbolag. Det var 6773 observationer över de 5 variablerna:

- *State*: Vilken stat fordon var registrerat i.
- *Class*: En typ av riskklassificering över försäkringen baserat på ålder, kön, civilstånd, användningsområde för fordonet.
- *Gender*: Kön på försäkringstagaren.
- *Age*: Ålder på försäkringstagaren.
- *Paid*: Kostnaden för en skada.

Jag är egentligen bara intresserad av två variabler, *Class* och *Paid*. *Class* delar in försäkringarna i olika risktyper och jag antar därför att skadekostnaderna (*Paid*) för en viss risktyp kommer från samma fördelning. Jag valde en risktyp med 157 observationer.

Datamaterialet är hämtat från en bok som behandlar regression med tillämpningar inom försäkring och är därför inte optimalt, men då det var svårt att hitta bra data till mitt exempel valde jag ändå detta.

Den största nackdelen med detta datamaterial är att det inte innehåller någon information om vilken tidsperiod datamaterialet kommer ifrån. Då jag inte vet om skadorna kommer från flera år/månader etc. har jag antagit att datamaterialet är hämtat under ett års tid.

Jag väljer därför tidsperioden för det totala skadebeloppet till ett år. Låt säga att datamaterialt kommer från 2016 så är S det totala skadebeloppot för 2017.

3.2 Simulering

När jag har fördelningen för antalet skador och fördelningen för skadebeloppen simulerar jag sedan 1 000 000 totala skadebelopp för att få en fördelning att jämföra approximationerna mot.

Jag börjar med att simulera 1 000 000 tal från Poissonfördelningen, för var och ett av dessa tal simulerar jag sedan det antalet skadebelopp från Lognormal-fördelningen och summerar dessa skadebelopp.

Om det första Poissonfördelade talet är 157 simulerar jag 157 skadebelopp från Lognormal-fördelningen och summerar sedan dessa till ett totalt skadebelopp. På så sätt får jag 1 000 000 stycken totala skadebelopp. Fördelningen för dessa simulerade totala skadebeloppen kan antas ligga nära den sanna fördelninen för S och används därför till att jämföra de olika approximationsmetoderna mot.

3.3 Parameterskattning

Poisson

Då jag bara har skador från en observation av antalet skador skattas parametern i Poissonfördelningen till:

$$\widehat{\mu_N} = \widehat{\sigma_N^2} = 157$$

Pareto

Med *ML-metoden* skattas parametrarna till:

$$\widehat{\alpha} \approx 0.3334 \quad \widehat{\beta} \approx 49.95$$

Eftersom $\widehat{\alpha} < 1$ existerar inga moment för den anpassade fördelningen och kan därför inte användas.

Med *momentmetoden* skattas parametrarna till:

$$\widehat{\alpha} \approx 1.0276 \quad \widehat{\beta} \approx 49.95$$

Här har jag använt *ML-skattningen* av β. Då $\hat{\alpha} < 3$ existerar inte det tredje momentet för den anpassade Paretofördelningen. Det tredje momentet behövs för att kunna beräkna skevheten för S som sedan används i de olika approximationsmetoderna. Därför kommer jag inte att använda Paretofördelningen som modell för de individuella skadebeloppen.

Lognormal

I Lognormal-fördelningen skattas parametrarna med *ML-metoden* till:

$$\widehat{\mu_Y} \approx 6.910 \quad \widehat{\sigma_Y} \approx 1.193$$

Där μ_Y och σ_Y är parametrarna i den associerade normalfördelningen.

Med *momentmetoden* skattas parametrarna till:

$$\widehat{\mu_Y} \approx 7.0927 \quad \widehat{\sigma_Y} \approx 0.9314$$

Det är ingen större skillnad mellan dessa olika metoder, jag valde att använda ML-metoden.

För att NP-approximationen ska vara giltig måste $x > \mu_S - \frac{3\sigma_S}{\gamma_S}$ och vi har att:

$$\mu_S - \frac{3\sigma_S}{\gamma_S} \approx 89017$$

Så den går bara att använda för $x > 89017$.

3.4 Resultat

Teori mot simulering
För den Sammansatta Poissonfördelningen beräknas det teoretiska väntevärdet, teoretiska variansen och teoretiska skevheten till:

$$\mu_S \approx 3.2077 * 10^5 \quad \sigma_S^2 \approx 2.7213 * 10^9 \quad \gamma_S \approx 0.6753$$

Från den simulerade fördelningen för de totala skadebeloppen beräknas väntevärde, varians och skevhet till:

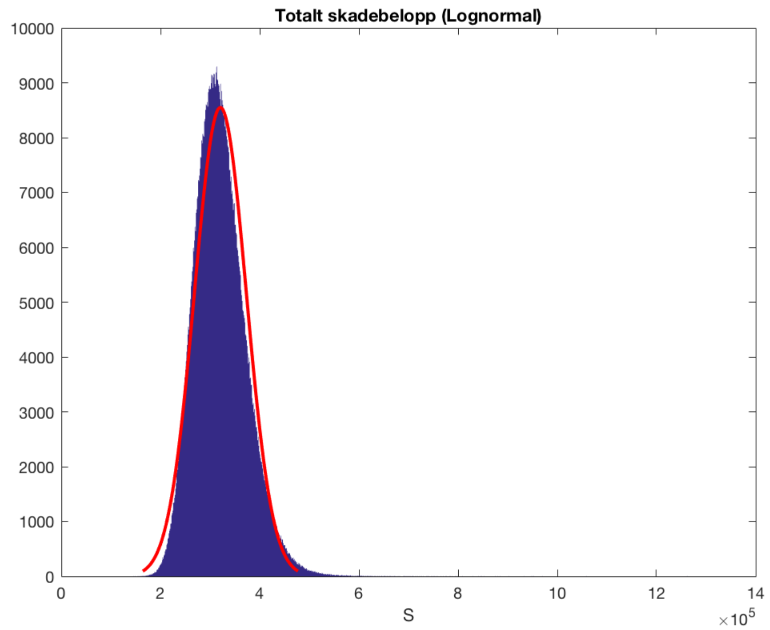
$$\widetilde{\mu_S} \approx 3.2074 * 10^5 \quad \widetilde{\sigma_S^2} \approx 2.7178 * 10^9 \quad \widetilde{\gamma_S} \approx 0.6720$$

Väntevärdet, variansen och skevheten från simuleringen ligger väldigt nära de teoretiska värdena vilket visar på att fördelningen för de simulerade totala skadebeloppen ligger nära den sanna fördelningen.

Grafer
Figur 3 visar en anpassad normalfördelning över ett histogram med de simulerade totala skadebeloppen. Även om man inte ser det så tydligt i figur 3 så finns det totala skadebelopp som ligger långt till höger om medelvärdet. Detta är tydligare i figur 4 och visar på att fördelningen är skev.

I figur 5 har jag jämfört värden för $1 - F_S(x)$ från 0.1 till 0.001 för de olika approximationsmetoderna mot den simulerade fördelningsfunktionen. Det går inte att urskilja Edgeworthapproximationen i den här figuren, men som figur 7 visar är Edgeworthapproximationen i princip identisk med normalapproximationen (notera skalan på y-axeln).

En till sak man ser från figur 5 är att NP-approximationen överskattar $1 - F_S(x)$ fram till och med ungefär $1 - F_S(487730) \approx 0.005$. I ett val mellan en metod som överskattar respektive underskattar risker skulle den metoden som överskattar risken vara att föredra, NP-approximationen i det här fallet.



Figur 3: Histogram med totala skadebelopp mot normalfördelning

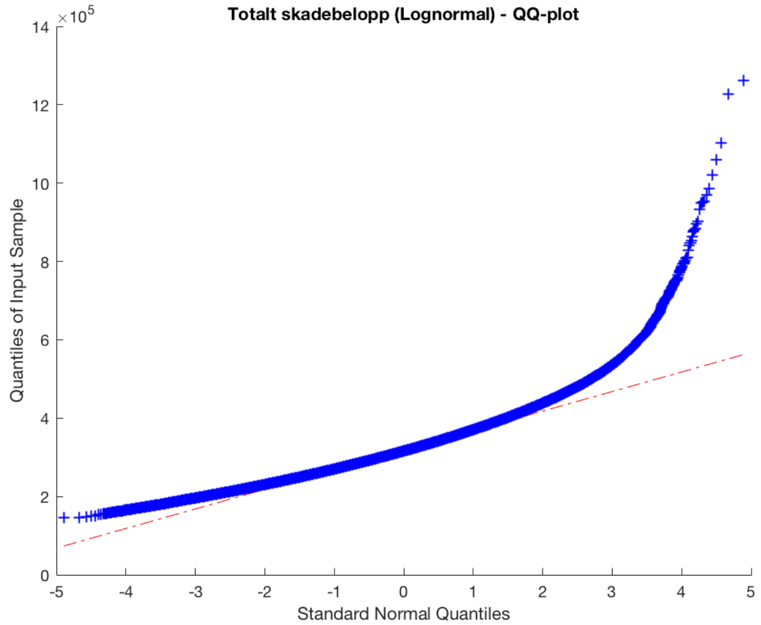
I figur 6 kan man se att normal- och Edgeworthapproximationerna ger ett bättre resultat än NP-approximationen fram till ungefär $1 - F_S(403670) \approx 0.064$.

Diskussion

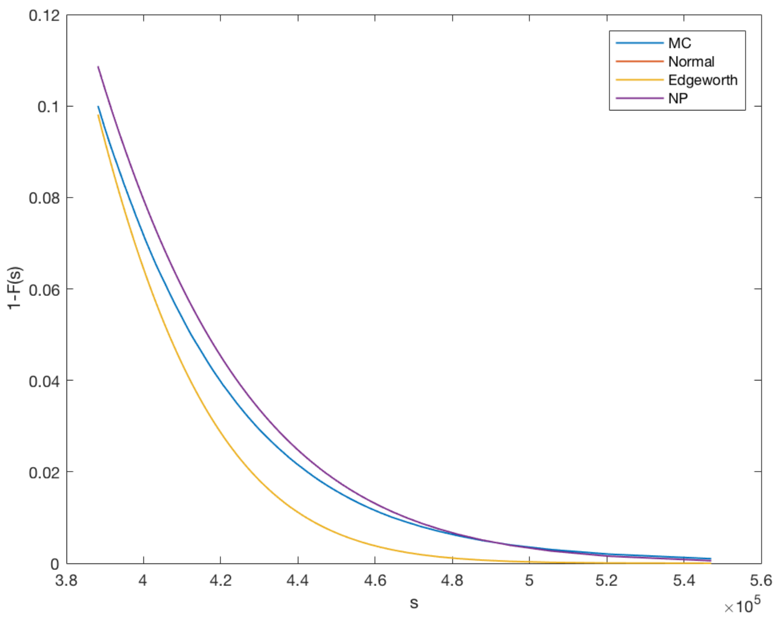
En fördel med NP-approximationen är att den är väldigt enkel att använda och kan användas istället för den ”vanliga” normalapproximationen. Den är egentligen inte mycket svårare att använda än normalapproximationen, det enda man behöver göra är att beräkna argumentet för Φ och sedan beräkna sannolikheten med hjälp av normalfördelningen som finns innbyggt i de flesta dataprogram (eller slå upp i en tabell).

Att normalapproximationen och Edgeworthapproximationen ligger så pass nära varandra i det här området beror på att den andra termen i Edgeworthapproximationen blir väldigt liten när x växer vilket man ser om man skriver om approximationen med hjälp av $\Phi^{(3)}(z) = (z^2 - 1)f(x)$, där $f(x)$ är täthetsfunktionen för normalfördelningen N(0,1).

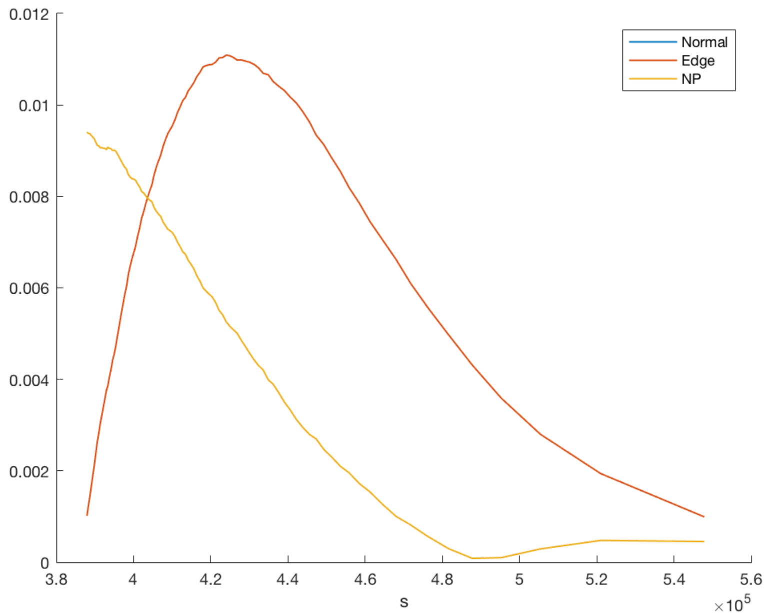
$$F_S(x) \approx \Phi(z) - \frac{\gamma_S}{6}\Phi^{(3)}(z) = \Phi(z) - \frac{\gamma_S(z^2 - 1)}{6} * \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$



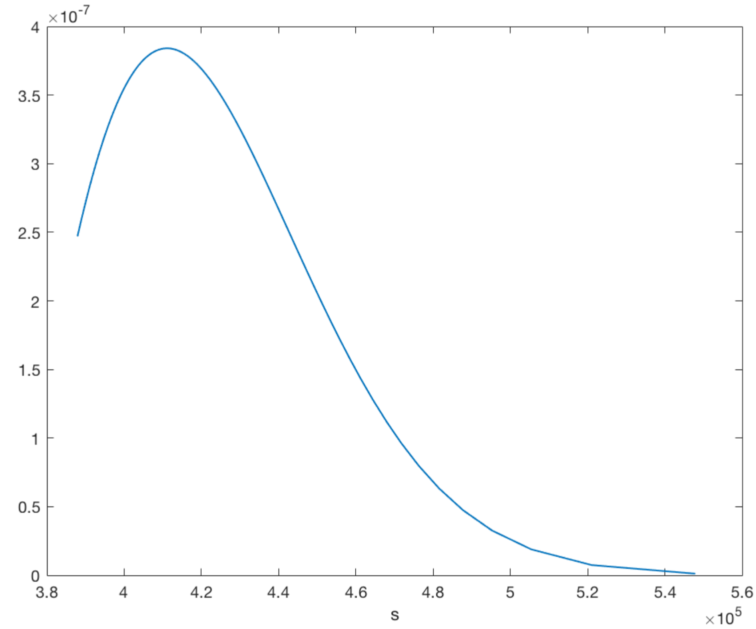
Figur 4: QQ-plot med totala skadebelopp mot normalfördelning



Figur 5: Approximation av $1 - F_S(x)$ för de olika metoderna



Figur 6: Absolut värde av skillnaden mellan den simulerade fördelningsfunktionen och approximationerna



Figur 7: Absolut värde av skillnaden mellan Edgeworthapproximationen och normalapproximationen

När z växer går $e^{-z^2/2}$ mot 0 betydligt snabbare än $(z^2 - 1)$ och för väldigt stora z blir den andra termen försumbar.

Ett av antagandena för den kollektiva modellen är att de individuella skadebeloppen ska vara likafördelade. Detta skulle kunna innebära att man tittar på en portfölj med samma typ av försäkringar, till exempel fordonsförsäkringar.

I mitt exempel var skadebeloppen ytterligare indelade i olika riskklasser och antagandet att skadebelopp från samma riskklass var likafördelade känns rimligt. Utan en sån klassificering skulle man istället kunna anta att alla skadebeloppen från fordonsförsäkringarna är likafördelade. Detta är kanske inte lika troligt, utan man bör nog ha någon typ av ”finare” indelning av försäkringarna (fordonsförsäkringar i det här fallet).

Man skulle sedan kunna använda approximationen av S på följande förenklade sätt.

Exempel 3.1 *Låt säga att ett taxibolag vill försäkra 100 fordon hos oss, historisk data från liknande försäkringar ger oss (tidsperiod 1 år för 100 fordon):*

$$\mu_S = 300000 \quad \sigma_S^2 = 30000 \quad \gamma_S = 0.7$$

Med NP-approximation har vi:

x	$1 - F_S(x)$
300 000	*0.454*
400 000	*0.068*
500 000	*0.0047*
600 000	*0.00019*

Sannolikheten att det totala skadebeloppet för dessa 100 fordon ska överstiga 600 000 kronor är ungefär 0.0002=0.02%. Om vi tar 600 000 kronor betalt för att försäkra dessa 100 fordon kommer vi med hög sannolikhet ha råd att betala skadorna under nästa år.

Slutsats

NP-approximationen är den bättre av de tre olika metoderna i det här fallet. Det är också vad jag förväntade mig då den tar hänsyn till skevheten hos fördelningen. I det här exemplet var den bättre än de två andra metoderna i det område som är av störst intresse, $P(S > x) = 1 - \alpha$ för $\alpha > 0.94$.

Av dessa tre approximationsmetoder är NP att föredra så länge den går att använda. Eftersom γ_S minskar när antalet skador växer så går normalapproximation att använda om antalet skador är stort och γ_S är nära 0.

Jag ser ingen anledning att använda Edgeworth för att approximera fördelningen då den inte ger ett bättre resultat än normalapproximationen i svansen och är bökigare att använda än NP. Det verkar inte heller som att Edgeworth är en etablerad approximationsmetod utan används endast för att få fram NP-approximationen.

4 Referenser

[1] Robert E. Beard, Teivo Pentikäinen, and Erkki Pesonen. *Risk theory.* Methuen, 1969.

[2] Björn Johansson. Matematiska modeller inom sakförsäkring, 1997.

[3] Thomas Mikosch. *Non-life insurance mathematics : an introduction with the Poisson process.* Springer, 2009.

[4] Kaas Rob, Goovaerts Marc, Dhaene Jan, and Denuit Michel. *Modern Actuarial Risk Theory: Using R.* Springer, 2008.

[5] Mette Rytgaard. Estimation in the pareto distribution. *Astin Bulletin*, 20(2), 1990.

5 Appendix

5.1 Bevis

5.1.1 Sats 2.2 (Paretofördelning)

Väntevärdet ges av:

$$\mu_X = E[X] = \frac{\alpha\beta}{\alpha - 1}$$

Variansen ges av:

$$\sigma_X^2 = E[X^2] - \mu_X^2 = \frac{\alpha\beta^2}{\alpha - 2} - \frac{\alpha^2\beta^2}{(\alpha-1)^2} = \frac{\alpha\beta^2(\alpha-1)^2 - \alpha^2\beta^2(\alpha-2)}{(\alpha-2)(\alpha-1)^2}$$

$$= \frac{\alpha\beta^2}{(\alpha-2)(\alpha-1)^2}$$

5.1.2 Sats 2.3 (Lognormal-fördelning)

Väntevärdet ges av:

$$\mu_X = E[X] = E[e^Y] = m_Y(1) = exp\left[\mu_Y + \frac{1}{2}\sigma_Y^2\right]$$

Där $m_Y(t) = exp\left[\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2\right]$ är den momentgenererande funktionen för en normalfördelad variabel Y.

Variansen ges av:

$$\sigma_X^2 = E[X^2] - (\mu_X)^2 = E[e^{2Y}] = m_Y(2) = exp\left[2\mu_Y - 2\sigma_Y^2\right]$$

5.1.3 Sats 2.5 (Sammansatt Poissonfördelning)

Den *momentgenererande* funktionen för S ges av:

$$m_S(t) = E\left[e^{tS}\right] = E\left[E\left[e^{tS}|N\right]\right] = E\left[E\left[e^{tX_1}\right]E\left[e^{tX_2}\right]\cdots E\left[e^{tX_N}\right]|N\right]$$

$$= E\left[(m_X(t))^N\right] = E\left[e^{Nlog(m_X(t))}\right] = m_N\left(log\left(m_X(t)\right)\right)$$

Där $m_N(t)$ är den momentgenererande funktionen för N och $m_X(t)$ är den momentgenererande funktionen för X.

5.1.4 Sats 2.6 (Sammansatt Poissonfördelning)

Den *karaktäristiska* funktionen för S ges av:

$$\chi_S(k) = E\left[e^{ikS}\right] = E\left[E\left[e^{ikS}|N\right]\right] = E\left[e^{Nlog(\chi_X(k))}\right] = m_N(log\left(\chi_X(k)\right))$$

Med den momentgenerererande funktionen för Poissonfördelningen $m_N(t) = exp\left[\mu_N(e^t - 1)\right]$ kan funktionerna skrivas:

$$m_S(t) = exp\left[\mu_N(m_X(t) - 1)\right] \quad \chi_S(k) = exp\left[\mu_N(\chi_X(k) - 1)\right]$$

5.1.5 Sats 2.4 (Sammansatt Poissonfördelning)

Väntevärde för S ges av:

$$\mu_S = E\left[E\left[S|N\right]\right] = E\left[N\mu_X\right] = \mu_X\mu_N$$

Variansen för S ges av:

$$\sigma_S^2 = E\left[V\left[S|N\right]\right] + V\left[E\left[S|N\right]\right] = E\left[N\sigma_X^2\right] + V\left[N\mu_X\right] = \mu_N\sigma_X^2 + \sigma_N^2\mu_X^2$$

För en Poissonfördelningen har vi att $\mu_N = \sigma_N^2$ så $\sigma_S^2 = \mu_N(\sigma_X^2 + \mu_X^2)$.

Skevheten för S ges av:

$$\gamma_S = \frac{\Psi_S^{(3)}(0)}{\sigma_S^3} = \frac{E\left[X^3\right]}{\sqrt{\mu_N E\left[X^2\right]}}$$

Där $\Psi_S^{(3)}(0) = \frac{E[X^3]}{\sqrt{\mu_N}}$ är tredje derivatan av den kumulantgenererande funktionen för S, $\Psi_S(t) = log(m_S(t))$ utvärderad i 0.

5.2 Edgeworth

För en oändligt deriverbar fördelningsfunktion $F(x)$ med $F^{(k)}(x) = 0$ för $\pm\infty$ och $k \geq 1$ kan den karaktäriska funktionen för den k:te derivatan av fördelningsfunktionen skrivas:

$$\chi^{(k)}(s) = \int_{-\infty}^{\infty} e^{isx} dF^{(k)}(x) = -is\int_{-\infty}^{\infty} e^{isx} dF^{(k-1)}(x) = \ldots = (-is)^k\chi(s)$$

Normalfördelningen är en sådan funktion där den karaktäristiska funktionen är:

$$\chi_N(s; \mu_X, \sigma_X^2) = exp\left[is\mu_X - \frac{1}{2}s^2\sigma_X^2\right]$$

Den karaktäristiska funktionen för en Sammansatt Poissonfördelning:

$$\chi_S(s) = exp\left[\mu_N\chi_X(s) - \mu_N\right]$$

Där $\chi_X(s)$ är den karaktäristiska funktionen för skadebeloppen, μ_X väntevärdet för skadebeloppen och μ_N väntevärdet för antalet skador.

Om man taylorutvecklar $\chi_X(s)$ och använder $\chi_X^{(k)}(0) = i^k E\left[X^k\right]$ får man att:

$$\chi_S(s) = exp\left[\mu_N \sum_{k=0}^{4} \frac{\chi_X^{(k)}(0)}{k!} s^k + \mu_N O(s^5) - \mu_N\right]$$

$$= exp\left[is\mu_N\mu_X - \frac{1}{2}s^2\mu_N(\sigma_X^2 + \mu_X^2)\right] exp\left[\mu_N\chi_X^{(3)}(0)\frac{s^3}{3!} + \mu_N\chi_X^{(4)}(0)\frac{s^4}{4!} + \mu_N O(s^5)\right]$$

$$= \chi_N(s;\mu_S,\sigma_S^2)\left(1 + \frac{\mu_N E\left[X^3\right]}{6}(is)^3 + \frac{\mu_N E\left[X^4\right]}{24}(is)^4 + \frac{\mu_N^2 E\left[X^3\right]^2}{72}(is)^6 + \ldots\right)$$

$$= \chi_N(s;\mu_S,\sigma_S^2) - \chi_N^{(3)}(s)\frac{\mu_N E\left[X^3\right]}{6} + \chi_N^{(4)}(s)\frac{\mu_N E\left[X^4\right]}{24} + \chi_N^{(6)}(s)\frac{\mu_N^2 E\left[X^6\right]^2}{72} + \ldots$$

Låt $Z = \frac{S-\mu_S}{\sigma_S}$, $z = \frac{x-\mu_S}{\sigma_S}$ och utnyttja det faktum att $\Phi^k(x;\mu_S,\sigma_S^2) = \sigma_S^{-k}\Phi^k(z)$, då kan fördelningsfunktionen för en Sammansatt Poissonfördelning skrivas som:

$$F_S(x) = F_Z(z) = \Phi(z) - \Phi^{(3)}(z)\frac{\mu_N E\left[X^3\right]}{6\sigma_S^3} + \Phi^{(4)}(x)\frac{\mu_N E\left[X^4\right]}{24\sigma_S^4} + \Phi^{(6)}(x)\frac{\mu_N^2 E\left[X^3\right]^2}{72\sigma_S^6} + \ldots$$

Genom att använda de två första termerna och att $\sigma_S^2 = \mu_N(\sigma_X^2 + \mu_X^2) = \mu_N E\left[X^2\right]$ och $\gamma_S = \frac{E\left[X^3\right]}{\sqrt{\mu_N E\left[X^2\right]^3}}$ får man approximationen:

$$F_S(x) \approx \Phi(z) - \frac{\gamma_S}{6}\Phi^{(3)}(z)$$

5.3 NP-approximation

För Edgeworthapproximationen hade vi:

$$F_S(x) = F_Z(z) = \Phi(z) - \Phi^{(3)}(z)\frac{\mu_N E\left[X^3\right]}{6\sigma_S^3} + \cdots$$

Där $Z = \frac{S-\mu_S}{\sigma_S}$, $z = \frac{x-\mu_S}{\sigma_S}$.

För NP-approximationen vill vi justera argumentent i $F_Z(s)$ med $\Delta = \Delta(s)$ så att $F_Z(s+\Delta) \approx \Phi(s)$. Vi bestämmer Δ genom att hitta vart funktionen $h(\Delta) = 0$:

$$h(\Delta) = \Phi(s) - \Phi(s+\Delta) + \frac{\gamma_S}{6}\Phi^{(3)}(s+\Delta)$$

Det kan göras med hjälp av en Taylorexpansion i origo:

$$h(\Delta) \approx h(0) + \Delta h'(0)$$

Med $h(0) = \frac{\gamma_S}{6}\Phi^{(3)}(s)$, $h'(0) = -\Phi'(s) + \frac{\gamma_S}{6}\Phi^{(4)}(s)$, $\Phi'(s) = \phi(s)$ och $\Phi^{(4)}(s) = -(s^3 - 3s)\phi(s)$, där $\phi(s)$ är täthetsfunktionen för N(0,1) får vi att:

$$\Delta \approx -\frac{h(0)}{h'(0)} = -\frac{\frac{\gamma_S}{6}\Phi^{(3)}(s)}{-\Phi'(s) + \frac{\gamma_S}{6}\Phi^{(4)}(s)} = \frac{\frac{\gamma_S}{6}(s^2-1)\phi(s)}{\phi(s) + \frac{\gamma_S}{6}(s^3 - 3s)\phi(s)}$$

$$\Delta \approx \frac{\gamma_S}{6}(s^2 - 1)$$

Vi har nu $F_X(x) = F_Z\left(\frac{x-\mu_S}{\sigma_S}\right) = F_Z\left(s + \frac{\gamma_S}{6}(s^2-1)\right) \approx \Phi(s)$ och att:

$$\frac{x - \mu_S}{\sigma_S} = s + \frac{\gamma_S}{6}(s^2 - 1) \Leftrightarrow s = \sqrt{\frac{9}{\gamma_S^2} + \frac{6(x-\mu_S)}{\gamma_S \sigma_S} + 1} - \frac{3}{\gamma_S}$$

Så för att approximera $F_X(x)$ används:

$$F_X(x) \approx \Phi\left(\sqrt{\frac{9}{\gamma_S^2} + \frac{6(x-\mu_S)}{\gamma_S \sigma_S} + 1} - \frac{3}{\gamma_S}\right)$$