



Matematisk statistik
Stockholms universitet

Modellering av kostnaden för excess of loss återförsäkring

Andreas Ericsson

Examensarbete 2004:14

Postadress:

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm
Sverige

Internet:

<http://www.math.su.se/matstat>

Modellering av kostnaden för excess of loss återförsäkring

Andreas Ericsson¹

Augusti 2004

¹: andreas.ericsson@folksam.se. Handledare: Esbjörn Ohlsson.

Sammanfattning

Ett försäkringsbolags kostnad för en katastrofskada kan i sällsynta fall bli mycket stor, händelsen innebär därför en stor risk med hänsyn till bolagets resultat och i värsta fall mot fullgörandet av dess åtaganden gentemot försäkringstagarna. Ett sätt att skydda sig mot denna risk är att teckna återförsäkring mot händelsen. En uppskattning av riskpremien fås genom att beräkna väntevärdet för återförsäkrarens kostnad. I excess of loss återförsäkring ges skydd mot att en enskild skada överstiger den s k excesspunkten. För att begränsa återförsäkrarens ansvar anges också en maxgräns för den totala ersättningen. Storleken på återförsäkrarens kostnad beror därför på summan av ersättningen för de skador som går in i skyddet. En vanlig modell, med stöd i extremvärdesteorien, är att ett Poissonfördelat antal skador överstiger excesspunkten och att kostnaden för dessa är Paretofördelade. Här undersöks vad konsekvensen blir då vi istället antar att antalet har en blandad Poissonfördelning, nämligen en Negativ binomialfördelning. Detta prövas på två data-material – ett från Länsförsäkringar och ett från ett franskt bolag. Det visar sig att Negativ Binomial passar bäst för Länsförsäkringar men att konsekvensen för riskpremien inte är så stor. För det franska bolaget är Poisson och Negativ Binomial likvärdiga. Fördelningen för summan beräknas med hjälp av Panjers rekursionsformel. Vi gör också en jämförelse av Panjers- och Ströters rekursionsformler sinsemellan och med en simulering. Vid jämförelsen med simuleringen visar sig precisionen för de båda rekursionerna vara mycket god. Fördelarna med Panjers rekursionsformel är framför allt att den går att använda i samband med excess of loss återförsäkringen, dessutom är metodens beräkningskomplexitet lägre.

Abstract

A catastrophe may cause an insurance company major costs and in the worst case scenario may jeopardize its commitments to policy takers. One way of reducing this risk is for the insurance company to reinsure itself against extreme events. An estimate of the risk premium can be obtained by calculating the expectation of the reinsurers cost. Excess of loss insurance provides protection against the event where the cost of a single claim exceeds the excess limit. In order to limit the reinsurers commitment the policy also specifies a maximum total claim. Therefore the sum of the claims exceeding the excess point will decide the cost of the reinsurer. The model we consider accounts a random number and a random size of the losses exceeding the excess point. A model motivated by extreme value theory is the peaks over threshold model (POT). According to this losses which exceed the excess point come as a Poisson process, and the excesses, i.e. the amount by which the losses exceed the level, follow a Pareto distribution. Here we also consider the consequences of assuming that the number of losses exceeding the level may follow a Negative binomial distribution. Since the Negative binomial model is more complex we compute the expectation of the reinsurers cost to decide if the difference

between the models is large enough to motivate the use of the Negative binomial model. A comparison between Panjers- and Ströters recursion formulas and a simulation is performed. When compared to the simulation it shows that the precision of the recursions is indeed satisfying. The advantages of Panjers recursion formula is foremost that it's applicable for the excess of loss reinsurance problem one of the other main advantages is that it's less complex.

Förord

Detta arbete utgör ett 20 poängs examensarbete i matematisk statistisk och har utförts för Länsförsäkringar AB i Stockholm.

Arbetet påbörjades på initiativ av Slobodanka Mijalkovic, som är verksam som aktuarie på Länsförsäkringar återförsäkring.

Jag vill här passa på att tacka Slobodanka som varit min handledare på Länsförsäkringar och hennes kollega Daniel Rutberg som också deltagit i diskussioner kring arbetet. Jag vill också tacka Esbjörn Ohlsson, min handledare på matematiska institutionen vid Stockholms universitet, som har hjälpt till i processen med detta arbete och som såg till att jag fick chansen att göra arbetet.

Innehåll

1	Inledning	1
1.1	Bakgrund	1
1.2	Metoder	5
1.2.1	POT-modellen	6
1.2.2	Antalet skador	7
1.2.3	Skadekostnad	7
1.2.4	Total skadekostnad	7
1.3	Mål	8
1.4	Datamaterial	8
2	Fördelning för antalet skador per år	11
2.1	Poisson- och blandad Poissonfördelning	11
2.2	Uttunnad Poisson- och uttunnad blandad Poissonfördelning	13
2.3	Data	14
2.4	Anpassning av fördelningar för data	15
2.4.1	Länsförsäkringar	16
2.4.2	Franska bolaget	17
2.4.3	Jämförelse av resultat för respektive datamaterial	17
3	Fördelning för skadekostnaden	20
3.1	Data skadekostnaden	20
3.2	Resultat från anpassning av fördelning för skadekostnaden	22
4	Fördelning för total skadekostnad per år	23
4.1	Data total skadekostnad	23
4.2	Modeller för total skadekostnad per år	23
4.2.1	Panjers rekursionsformel	24
4.2.2	Ströters rekursionsformel	24
4.2.3	Simulering	25
4.3	Beräkning av fördelning för $S' = \sum_{i=1}^N (X'_i - d)$	26
4.3.1	Länsförsäkringar	26
4.3.2	Franska bolaget	27
4.4	Beräkning av fördelning för S	30

A	Fördelning för skadekostnaden	33
A.1	Anpassning av fördelning för skadekostnaden	33
A.1.1	Länsförsäkringar	33
A.1.2	Franska bolaget	36

Figurer

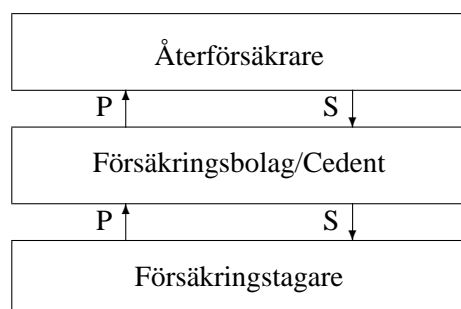
2.1	Skadekostnaders ankomst i tiden	12
2.2	Antal år n_k med k stycken skador	15
2.3	Anpassning av Poisson- och NegBinfördelning, Länsförsäkringar .	16
2.4	Fördelningsfunktioner plottade mot varandra, Länsförsäkringar . .	17
2.5	Anpassning av Poisson- och NegBinfördelning, franska bolaget . .	18
2.6	Fördelningsfunktioner plottade mot varandra, franska bolaget . . .	18
3.1	Enskilda skadors andel av den totala skadekostnaden	21
4.1	Fördelning för S' , Länsförsäkringar	26
4.2	Fördelning för S' , franska bolaget	28
4.3	Fördelning för S' plottad mot $\log(S')$, franska bolaget	28
A.1	Fördelning för skadekostnaden, Länsförsäkringar	34
A.2	Fördelning för skadekostnaden $[0, 1.2E^8]$, Länsförsäkringar . . .	35
A.3	Fördelning för skadekostnaden $[1.2E^8, 8E^8]$, Länsförsäkringar . .	35
A.4	Fördelning för skadekostnaden, franska bolaget	37
A.5	Plot av fördelningsfunktioner, franska bolaget	37
A.6	Fördelning för den logaritmerade skadekostnaden, franska bolaget	38

Kapitel 1

Inledning

1.1 Bakgrund

Vid tecknandet av ett försäkringskontrakt åtar sig försäringsgivaren att ersätta försäkringstagaren för kostnader som uppstått till följd av någon sådan skada som kontraktet avser. I samband med till exempel en storm eller en översvämning drabbas många gånger ett stort antal objekt av skador där objekten skyddas av olika försäkringskontrakt. Exempelvis skyddas bilen av en vagnskadeförsäkring medan villan skyddas av en villaförsäkring. Detta medför att i det område som drabbas kommer ett flertal kontrakt som skyddar mot olika typer av skador att tas i anspråk. För ett försäkringsbolag som har många försäkringstagare i området kan därför totalkostnaden för skadan bli mycket stor. För att garantera fullgörandet av sina åtaganden och för att jämna ut resultaten mellan olika år behöver försäkringsbolaget någon form av skydd mot den här typen av "katastrofhändelser". Ett sådant skydd är återförsäkring, vilken innebär att försäkringsbolaget för vissa risker tecknar försäkring hos en annan försäringsgivare. Försäkringsbolaget kallas i detta sammanhang för cedent och den som försäkrar cedenten kallas återförsäkrare eller reassuradör [4].



FLÖDESSCHEMA FÖR PREMIER (P) OCH SKYDD (S)

Risken i detta sammanhang är alltså att totalkostnaden i samband med en katastrofhändelse kan bli mycket stor. För att skydda sig mot stora kostnader tecknar bolaget/cedenten således en återförsäkring för just denna händelse. Eftersom återförsäkringskontraktet avser totalkostnaden för skadan så är det inte i första hand de skadekostnader som ersätts på grund av kontraktet med den enskilde försäkringstagaren som är det intressanta. Vid prissättningen av återförsäkringskontraktet betraktas istället hela den kostnad som katastrofhändelsen orsakar för försäkringsbolaget. Antag att kontraktstiden för återförsäkringen är ett år, låt X_i vara försäkringsbolagets totalkostnad för katastrofskada i och låt vidare K vara antalet skador per år där $X_i \geq 0$. I ovanstående modell bortses från de enskilda skadekostnaderna istället betraktas följden av totalkostnader X_1, \dots, X_K modellen brukar inom sakförsäkringsmatematiken kallas för den kollektiva modellen [5].

Av anledning som vi återkommer till senare inför vi här också N som är antalet skador per år med en skadekostnad som överstiger d och X'_j , skadekostnaden för den j :te skadan som översiger nivån d .

För att underlätta beskrivningen av återförsäkringskontrakten, börjar vi med att införa beteckningar för ett antal variabler och återkommer med förklaringar till dessa efterhand. Vi låter som tidigare X_i vara cedentens kostnad för skada i dessutom inför vi variablerna Y_i , S och S_n , som definieras enligt nedanstående:

$$Y_i = \begin{cases} 0 & \text{om } X_i \leq b \\ X_i - b & \text{om } b < X_i < b + a \\ a & \text{om } X_i \geq b + a \end{cases} \quad (1.1)$$

$$S = \sum_{i=1}^K Y_i$$

$$S_n = \sum_{i=1}^n Y_i$$

En typ av återförsäkring är excess of loss återförsäkring. Vi kommer här att beskriva några olika varianter av denna. Den enklaste typen är a excess of b vilken vi betecknar a xs b . Med noll så kallade reinstatements ges återförsäkrarens kostnad Z_0 för kontraktet enligt:

$$Z_0 = \begin{cases} S & \text{om } S < a \\ a & \text{om } S \geq a \end{cases}$$

Som vi ser ovan beror storleken på återförsäkrarens kostnad för skada n av hur stor S_{n-1} är. För att ta upp de vanligaste begreppen och för att få en genomgång av vad återförsäkrarens kostnad blir för olika värden på S_{n-1} antas till att börja med att $S_{n-1} = 0$ detta inträffar då $X_i \leq b$, $i = 1, \dots, n-1$. Återförsäkrarens kostnad för den n :te skadan är under dessa förutsättningar: Y_n som är $X_n - b \in (0, a)$ då skadekostnaden är större än b men mindre än $b + a$ och a då skadekostnaden är större än eller lika med $b + a$, annars är kostnaden 0. Intervallet $[b, b + a]$ kallas för en layer,

punkten b där återförsäkringsgivaren träder in kallas excesspunkten och maxbeloppet för ersättningen a kallas layers limit. Om $X_n > b + a$ får cedenten själv stå dels för kostnaden b – detta belopp kallas för cedentens *självbehåll*, dels för kostnaden $X_n - (b + a)$ det vill säga för den del som överstiger $b + a$ – detta belopp kallas *spillover*.

Antag nu istället att $0 < S_{n-1} < a$. Återförsäkrarens kostnad för den n :te skadan är under dessa förutsättningar: minimum av beloppen Y_n och $a - S_{n-1}$ då skadekostnaden är större än b , annars är kostnaden 0. Slutligen om $S_{n-1} \geq a$ har den maximala ersättningen a redan uppnåtts och återförsäkrarens kostnad för skada n är 0. Utan (noll) reinstatements är återförsäkringen förbrukad då $S_n \geq a$.

För att beskriva a *xs* b med r reinstatements görs enklast en uppdelning i två fall då $S_n \leq ra$ och då $ra < S_n < (r + 1)a$. Vi inför också en beteckning för återförsäkringspremien P . Till att börja med antar vi att $S_n \leq ra$. Med r reinstatements @ $q\%$ återställs skyddet efter skada n automatiskt till 100%. Cedenten har genom kontraktet förbundit sig att betala en tilläggspremie på $p_n q P$, där $p_n = Y_n/a$ är andelen av skyddet som skada n tog i anspråk. Skyddet för skada $n + 1$ är således hela tiden a . För att se vad återförsäkrarens kostnad för skada n blir hänvisas till ovanstående beskrivning av en a *xs* b utan reinstatements där $S_n = 0$. Den premie cedenten betalar för att påbörja återförsäkringen är P och av ovanstående framgår att ytterligare en premie om $p q P$ erläggs varje gång skyddet återställs. Den totala premien är således $P + S/a q P$ om $S < ra$.

Då $ra < S_n < (r + 1)a$ återställs inte skyddet efter skada n till 100%. Det återstående skyddet är nu $(r + 1)a - S_n$, ingen ytterligare premie erläggs för att få utnyttja detta. Cedenten får sedan erättning tills dess att $S_k \geq (r + 1)a$, $k \geq n + 1$ då den maximala ersättningen $(r + 1)a$ uppnås efter detta utbetalas ingen ytterligare ersättning. Om vi jämför med ovanstående beskrivning av en a *xs* b utan reinstatements kan vi konstatera att för $S_n > ra$ återstår ett skydd motsvarande en $(r + 1)a - S_n$ *xs* b utan reinstatements och därmed framgår vad återförsäkrarens kostnader blir för kommande skador. För $S \geq ra$ har den totala premien uppnått sitt maximum $P + r q P$.

Ovanstående kan på ett mer kärnfullt sätt sammanfattas:

Antag att $S_{n-1} \leq ra$, skyddet för skada n är då a och återförsäkrarens kostnad är Y_n .
Antag att $S_{n-1} > ra$, skyddet för skada n är då $\max(0, (r + 1)a - S_{n-1})$ och återförsäkrarens kostnad är $\min(Y_n, \max(0, (r + 1)a - S_{n-1}))$.

Återförsäkrarens bruttokostnad Z_r för en a *xs* b återförsäkring med r reinstatements ges enligt:

$$Z_r = \begin{cases} S & \text{om } (r + 1)a < S \\ (r + 1)a & \text{om } S \geq (r + 1)a \end{cases}$$

Nettokostnaden Z_r^N ges enligt:

$$Z_r^N = \begin{cases} S - P(1 + \frac{qS}{a}) & \text{om } ra < S \\ S - P(1 + rq) & \text{om } ra \leq S < (r+1)a \\ (r+1)a - P(1 + rq) & \text{om } S \geq (r+1)a \end{cases}$$

Ju högre excesspunkten är desto lägre är frekvensen för antalet skador som träder in i skyddet detta medför att behovet av reinstatements minskar, således gäller generellt att $b \uparrow$ medför $r \downarrow$. Vanligtvis används excess of loss återförsäkring i högre layers det vill säga då b är stort, antalet reinstatements är oftast mindre än 2.

En annan typ av återförsäkring är excess of loss med aggregerat självbehåll som vi ger beteckningen a xs b xs c . För en a xs b xs c med noll reinstatements ges återförsäkraren kostnad Z_0 enligt:

$$Z_0 = \begin{cases} 0 & \text{om } S - c \leq 0 \\ S - c & \text{om } 0 < S - c < a \\ a & \text{om } S - c \geq a \end{cases}$$

Konstruktion medför att cedenten bär hela kostnaden så länge $S \leq c$, c som kallas för *aggregatet* utgör alltså cedentens aggregerade självbehåll. Ersättningen till cedenten är $S - c \in (0, a)$ då summan av de kostnader som träder in i skyddet är större än c men mindre än $c + a$ och når sitt maximum a då summan är större än eller lika med $c + a$, annars är ersättningen 0. Utan reinstatements är återförsäkringen förbrukad då $S \geq c + a$. Vid avtal med r reinstatements återställs skyddet enligt samma princip som beskrivits tidigare och återförsäkringen får utnyttjas även då $S > c + a$. Ytterligare r limiter läggs då till $c + a$ så att återförsäkringen är förbrukad först då $S \geq c + (r+1)a$. För en a xs b xs c med r reinstatements ges återförsäkraren bruttokostnad Z_r av:

$$Z_r = \begin{cases} 0 & \text{om } S - c \leq 0 \\ S - c & \text{om } 0 < S - c < (r+1)a \\ (r+1)a & \text{om } S - c \geq (r+1)a \end{cases}$$

Nettokostnaden Z_r^N ges enligt:

$$Z_r^N = \begin{cases} S - c - P(1 + \frac{qS}{a}) & \text{om } ra < S - c \\ S - c - P(1 + rq) & \text{om } ra \leq S - c < (r+1)a \\ (r+1)a - P(1 + rq) & \text{om } S - c \geq (r+1)a \end{cases}$$

a xs b xs c kontrakt används framförallt i lägre layers där frekvensen för antalet skador som träder in i skyddet är högre det vill säga då b relativt sett är litet. Vi ska dock ha i åtanke att b inte är litet eftersom återförsäkring normalt inte tecknas för små- eller medelstora skador. Om vi i ett a xs b xs c kontrakt låter c vara noll så fås ett vanligt a xs b kontrakt, i ovanstående har uppdelningen i två fall gjorts för att

göra presentationen tydligare.

Då försäkringsbolaget upptäckt ett behov av återförsäkring ges ofta ett flertal presumtiva återförsäkrare möjlighet att inkomma med kontraktförslag. Premien för kontraktet fastställs i slutänden genom förhandling mellan cedenten och återförsäkraren. För att kunna bedömma värdet av ett återförsäkringskontrakt vill cedenten ha reda på om premien motsvarar den kostnad som kan uppstå för återförsäkraren. En riktlinje är således att premien exklusive administrationskostnader och dylikt bör motsvara väntevärdet för återförsäkrarens kostnad för kontraktet. I och med att ett kontrakt om återförsäkring sluts övertar återförsäkraren en del av ansvaret gentemot cedenten. Det bör dock nämnas att kunden aldrig märker av återförsäkrarens roll, försäkringsbolaget sköter alla kontakter med kunden och kundens anspråk kan endast riktas direkt mot försäkringsbolaget och inte mot återförsäkraren.

Med hjälp av den så kallade väntevärdes funktionen: $L_S(x) = E(\min(S, x))$ kan vi beräkna väntevärdet för våra återförsäkringskontrakt. Det är lätt att se att väntevärdet för ett a xs b xs c kontrakt är $L_S(c + a) - L_S(c)$. Väntevärdet för ett a xs b kontrakt är $L_S(a)$ vilket fås genom att sätta $c = 0$.

1.2 Metoder

Vi har tidigare konstaterat att väntevärdet för återförsäkrarens kostnad ger oss en uppfattning om vad storleken på premien bör vara. Vi ska nu se på vad som krävs för att senare kunna beräkna väntevärdet för kostnaden. Vi kommer här att ägna oss åt återförsäkring med noll reinstatements. När vi ovan beskrivit a xs b och a xs b xs c återförsäkring har vi sett att återförsäkrarens kostnad för kontraktet i båda fallen beror på storleken på $S = \sum_{i=1}^K Y_i$. För att kunna beräkna väntevärdet av återförsäkrarens kostnad i samband med de båda försäkringstyperna behöver vi alltså känna till fördelningen för S . Det är lätt att se att storleken på S endast beror på de Y_i som är positiva, $Y_i > 0$. Om vi låter M vara antalet skador per år där $Y_i > 0$ och Y'_j , $j = 1, \dots, M$ vara den j :te skadan sådan att $Y_i > 0$ inses att vi lika gärna kan betrakta S som $S = \sum_{j=1}^M Y'_j$. S har i båda fallen vad man kallar för en sammansatt fördelning. Om vi antar att M, Y'_1, \dots, Y'_M är oberoende stokastiska variabler och att Y'_1, \dots, Y'_M dessutom är likafördelade så kan vi, givet att vi känner till fördelningarna för M och Y'_1 beräkna fördelningen för $S = \sum_{i=1}^M Y'_i$. Då ovanstående villkor gäller för K och Y_1, \dots, Y_K kan vi naturligtvis också beräkna fördelningen för S med hjälp av dessa. I avsnitt 4.4 kommer vi att se att beräkningarna blir enklare genom att betrakta S som $S = \sum_{i=1}^M Y'_i$.

1.2.1 POT-modellen

POT-modellen (Peaks over thresholds) har visat sig vara användbar i samband med återförsäkring. Med denna har vi nämligen, givet vissa antaganden, möjlighet att få en ökad precision i skattningarna av fördelningen dels för antalet skador över en viss nivå och dels för den del skadekostnaden som överstiger samma nivå. Principen går ut på att välja en viss tröskelnivå u och studera de skador som överstiger u . Vi väljer ett visst tidsintervall, säg ett år, och låter V_i vara storleken på den i :te skadan som överstiger u och K vara antalet skador per år som överstiger u . Antag att K, V_1, V_2, \dots är oberoende samt att V_1, V_2, \dots är likafördelade. I POT-modellen antas att K är Poissonfördelad och att $V_i - u$ är Paretofördelade. Låt F vara fördelningsfunktionen för $V_i - u$. Vi väljer den parametrering där $F(x) = P(V_i - u \leq x) = 1 - (\frac{\alpha}{\alpha+x})^\gamma$, $x > 0$, och skriver att $V_i - u \in \text{Pareto}(\alpha, \gamma)$.

Antag nu att vi betraktar en högre tröskelnivå $v > u$, låter L vara antalet skador per år som överstiger v , W_j vara storleken på den j :te skadan som överstiger v och $p = P(V_i > v)$. Det visar sig då att L får en så kallad uttunnad Poissonfördelning med uttunning p och att fördelningen för $W_j - v$ som är samma som för $V_i - v \mid V_i > v$ också är Paretofördelad men med nya parametrar. Om vi har att $K \in \text{Po}(\lambda)$ så gäller att $L \in \text{Po}(p\lambda)$ och med ovanstående parametrering för $V_i - u$ så gäller att $W_j - v \in \text{Pareto}(\alpha + v - u, \gamma)$. För en utförligare beskrivning hänsvisas till [5].

Antagandet att $V_i - u$ är Paretofördelad kan motiveras utifrån extremvärdesteorin (EVT). Antag att fördelningen för $M_n = \max\{U_1, \dots, U_n\}$ av oberoende och likafördelade stokastiska variabler U_i under en linjär normalisering konvergerar enligt: $\lim_{n \rightarrow \infty} P((M_n - b_n)/a_n \leq x) \rightarrow G(x)$ för $a_n, b_n > 0$ där $G(x)$ tillhör familjen av extremvärdesfördelningar. Kortfattat gör EVT gällande att om antagandet för M_n ovan är uppfyllt så gäller för ett tillräckligt stort värde på u att:

$P(U_1 - u \leq x \mid U_1 > u) \approx 1 - (\frac{\alpha}{\alpha+x})^\gamma$, $x > 0$. Om vi låter V_i vara de $\{U_i : U_i > u\}$ gäller således att: $P(V_1 - u \leq x) \approx 1 - (\frac{\alpha}{\alpha+x})^\gamma$, $x > 0$. För en utförligare beskrivning av antagandena i POT-modellen hänsvisas till [2] eller [7].

En av fördelarna med POT-modellen är att vi genom valet av tröskelnivån u kan styra så att vi har tillräckligt med observationer $\{u_i : u_i > u\}$ för att precisionen i skattningarna ska bli godtagbar. Samtidigt undviker vi de små observationer $\{u_i : u_i \leq u\}$ som annars skulle störa anpassningen av fördelningen. Utifrån denna skattning kan vi sedan få fördelningen för en högre tröskelnivå $v > u$ där vi annars hade haft för få observationer $\{u_i : u_i > v\}$ för att precisionen i skattningarna skulle kunna hållas på en godtagbar nivå. Ett motsatsförhållande finns dock mellan att höja nivån u , för att fördelningen för $V - u$ ska vara approximativt Paretofördelad och att ha tillräckligt med observationer för att få tillförlitliga skattningar.

1.2.2 Antalet skador

Av de anledningar som angetts i avsnitt 1.2.1 väljer vi en tröskelnivå d och studerar de skador som överstiger denna. I avsnitt 1.1 införde vi variabeln N som är antalet skador över d . Enligt POT-modellen kan vi anta att N är Poissonfördelad. I försäkringssammanhang är det känt att man ofta får en betydligt bättre anpassning till data genom att anta att N följer en så kallad blandad Poissonfördelning. Den mest använda blandade Poissonfördelningen är Negativ binomialfördelning (NegBin) vilken också är den som kommer att användas här. Tidigare konstaterades att antalet skador per år där $Y_i > 0$ är samma som antalet skador per år där $X_i > b$ således gäller enligt POT-modellen att M får en uttunnad Poissonfördelning då N antas Poissonfördelad. Vi kommer i avsnitt 2.2 att visa att motsvarande förhållande mellan N och M finns även då N antas ha en blandad Poissonfördelning.

NegBin-fördelningen är något mer komplicerad vad gäller till exempel skattning av parametrar och fördelningsfunktionens utseende. En fråga vi söker svaret på är om fördelarna är så pass stora att det väger upp den mer komplicerade modellen. Till att börja med undersöks respektive fördelnings anpassning till de två datamaterialen. För att på något sätt svara på om det är värt besväret att använda NegBin-fördelningen ska vi sedan jämföra vad skillnaden blir när vi använder den ena eller andra fördelningen för att beräkna väntevärdet för återförsäkrarens kostnad.

1.2.3 Skadekostnad

Från avsnitt 1.1 har vi att X' är skadekostnaden för de skador som överstiger tröskelnivån d . I POT-modellen antas att $X' - d$ är Paretofördelad, och vi såg att fördelningen för $X' - b \mid X' > b$ där $b > d$ också är Paretofördelad och att denna fördelning erhålls från fördelningen för $X' - d$ med hjälp av en enkel omparametrisering. I avsnitt 4.4 kommer vi se att vi kan använda fördelningen för $X' - b \mid X' > b$ för att få fördelningen för Y' . Vi har alltså stor anledning att välja Paretofördelning för $X' - d$, dels säger EVT att för ett tillräckligt stort värde på d så är $X' - d$ approximativt Paretofördelad vilket teoretiskt motiverar att Paretofördelningen bör ge god anpassning, dessutom är den enkel att arbeta med när vi vill ha fördelningen för Y' . Vi kommer dock även att anpassa Lognormalfördelningen för $X' - d$, en jämförelse av de två anpassningarna ska framförallt betraktas som en kontroll av att anpassningen för Paretofördelningen är tillräckligt god.

1.2.4 Total skadekostnad

Inom sakförsäkringsmatematiken förekommer det i olika sammanhang att man är intresserad av att känna till fördelningen för summan av ett stokastiskt antal oberoende och likafördelade stokastiska variabler. Fördelningen för en sådan summa kan beräknas med Panjers- eller Ströters rekursionsformler. Ett delmål som ligger något utanför den försäkringstekniska ramen, men som är av eget intresse är att jämföra de två rekursionerna. För att peka på för- och nackdelar med respektive

metod görs enklast en jämförelse av fördelningen för $S' = \sum_{i=1}^N X'_i - d$. Fördelningen som beräknas med de två rekursionerna kommer att jämföras inbördes, dessutom kommer de att jämföras med en simulering av samma fördelning.

Den summa som är intressant för återförsäkringen är $S = \sum_{i=1}^K Y_i = \sum_{i=1}^M Y'_i$. Då Y'_i (och även Y_i) har en blandad fördelning kan endast Panjers rekursionsformel användas för att beräkna fördelningen för S . För att kunna använda Ströters rekursionsformel krävs nämligen att Y'_i (eller Y_i) har en kontinuerlig fördelning. Jämförelsen av Panjers och Ströters rekursionsformel kommer därför också att tjäna som en kontroll av att Panjers rekursionsformel ger tillräckligt god precision då fördelningen för S ska modelleras.

1.3 Mål

Vad vi avser att undersöka med denna uppsats är:

- Om vi ska beräkna väntevärdet för återförsäkrarens kostnad, är i så fall den skillnad som finns mellan Poisson- och NegBin-fördelningen tillräckligt stor för att motivera användandet av den mer komplicerade NegBin-fördelningen?
- Vilka är för- och nackdelarna med Panjers respektive Ströters rekursionsformler då vi ska beräkna en sammansatt fördelning som S' ? Finns det någon anledning att föredra den ena rekursionen framför den andra?
- Eftersom endast Panjers rekursionsformel fungerar i samband med ovanstående återförsäkring, är precisionen i Panjers rekursionsformel tillräckligt god i detta sammanhang?

1.4 Datamaterial

I denna uppsats används två datamaterial: ett från Länsförsäkringar, och ett från ett franskt bolag, båda dessa har tillhandahållits av Länsförsäkringar AB. Den återförsäkring som behandlas är återförsäkring mot katastrofskador, detta medför att de intressanta observationerna är hela den kostnad bolaget haft för att ersätta sina kunder för en och samma katastrofhändelse. Låt x_i = den observerade kostnaden för katastrofhändelse i , den relevanta följderna av observationer är således x_1, x_2, \dots . Inom Länsförsäkringar finns 24 stycken lokala länsbolag. I datamaterialet från Länsförsäkringar finns observationer y_{ij} där y_{ij} är kostnaden för katastrof i hos bolag j . För Länsförsäkringar fås kostnaden x_i genom att summera kostnaderna för samma katastrof hos respektive länsbolag, dvs $x_i = \sum_{j=1}^{24} y_{ij}$. Genom att summera kostnaden på ovanstående sätt fås den kostnad som återförsäkras, samtidigt undviks det beroende som finns mellan olika kostnader i olika länsbolag som uppstått till

följd av samma händelse. I datamaterialet från det franska bolaget ges kostnaden per katastrofhändelse i det ursprungliga datamaterialet. Kostnaderna i respektive datamaterialet kan betraktas som observationer från oberoende och likafördelade stokastiska variabler.

För datamaterialet från Länsförsäkringar gäller att skador som uppstått på grund av händelsen storm, frost, snötryck, översvämning, skogsbrand, åska och landglidning eller av kombinationer av dessa betraktas som en katastrofhändelse oavsett vilken eller vilka av ovanstående händelse/er som orsakat den. I datamaterialet från det franska bolaget ingår endast stormskador men vi väljer för enkelhetens skull att kalla även dessa för katastrofskador. Återförsäkringskontrakten innehåller i regel en tidsklausul som reglerar den tidsrymd för vilken försäkringsbolaget kan räkna skadekostnader till en och samma katastrofhändelse. Olika katastrofhändelser har olika tidsklausuler, för t ex en stormskada gäller det att kontraktet avser alla skador orsakade av kraftig vind under en tidsperiod om 72 timmar. Cedenten väljer själv i vilken tidpunkt perioden börjar vilket medför att cedenten väljer den tidsperiod som maximerar kostnaden för stormen eftersom det är denna kostnad cedenten använder då anspråk på ersättning från återförsäkraren görs. Enligt definitionen av katastrofhändelserna har ingen händelse inträffat om ingen kostnad har uppstått.

Datamaterialet från Länsförsäkringar består av de kostnader som observerats under en tioårsperiod 1992-2001 och datamaterialet från det franska bolaget består av de kostnader som observerats under en tjugoförårsperiod 1982-2001. I båda datamaterialen har kostnaderna indexerats för att vara jämförbara över tiden dels med hänsyn till inflation och dels med hänsyn till portföljförändringar. Eftersom anpassningar av fördelningar görs för $X'_i - d$ och N har de båda datamaterialen begränsats till att omfatta de $\{x_i : x_i > d\}$. Tröskelvärde d brukar i detta sammanhang kallas för capture limit för datamaterialet. För att inte lämna ut känslig information för bolagen har capture limit för de båda datamaterialen utelämnats. Det bör också nämnas att storleken på capture limit i förhållande till medelskadan (medelvärde av skadekostnaderna) är i stort sett samma för de båda datamaterialen.

Låt $k = 1, 2, \dots, 10$ för åren 1992, \dots , 2001, w_k vara premievolymen för år k och $r_k = w_{10}/w_k$. I datamaterialet från Länsförsäkringar har en kostnad x_i från år k indexerats så att den nya observationen är $r_k * x_i$. Motsvarande indexering har gjorts för datamaterialet från det franska bolaget med skillnaden att $k = 1, 2, \dots, 20$ för åren 1982, \dots , 2001 och $r_k = w_{20}/w_k$.

De anpassningar av fördelningar som refereras till i denna uppsats har gjorts till det datamaterial som återstår då kostnaderna indexerats och då de indexerade kostnader som understiger capture limit har tagits bort. Från Länsförsäkringar består detta datamaterial av femtiotre observationer på kostnader angivna i miljoner kronor och tio observationer på antal skador under ett år, för det franska bolaget finns fyrtio observationer på kostnader angivna i miljoner Euro och tjugo observationer

på antal skador under ett år.

Kapitel 2

Fördelning för antalet skador per år

2.1 Poisson- och blandad Poissonfördelning

Den kanske vanligaste modellen för fördelningen för antalet skador per år N inom sakförsäkringsmatematiken är Poissonfördelningen. Att denna modell ska ge en bra anpassning är rimligt då följande villkor är uppfyllda [5]:

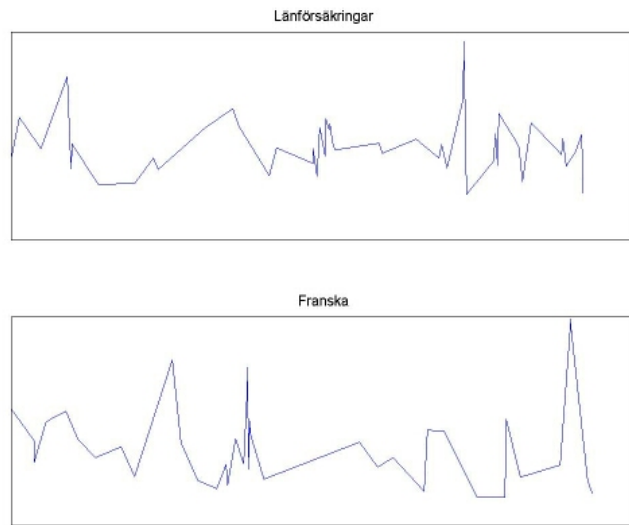
- (i) antalet händelser i disjunkta tidsintervall är oberoende
- (ii) bara en händelse i taget inträffar
- (iii) sannolikheten för att en händelse ska inträffa vid en viss given tidpunkt är noll

Om dessa villkor är uppfyllda ges sannolikhetsfunktionen av:

$$p_k = P(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots \quad \lambda \in (0, \infty)$$

Med ovanstående parametrisering skriver vi $N \in Po(\lambda)$. Att dessa villkor i alla fall är approximativt uppfyllda är rimligt att utgå ifrån för båda de datamaterial vi har att tillgå här. Det villkor som vi har störst anledning att ifrågasätta är (i) att antalet händelser i disjunkta tidsintervall är oberoende. I figur 2.1 kan vi nämligen se en antydning till att ibland följer en ny katastrofhändelse tätt inpå den föregående. I figuren plottas logritmen av skadekostnaden på y-axeln mot tiden på x-axeln. Skalan på axlarna har plockats bort för att inte utelämna känslig information för företagen. Logaritmering av skadekostnaden gjordes enbart för att figuren blev tydligare på detta sätt.

Det är rimligt att antalet katastrofskador per år och därmed fördelningen för N påverkas av t ex rådande väderförhållanden. Därför kan man tänka sig att givet ett visst väderförhållande är den korrekta modellen $N \in Po(\lambda_1)$ och givet ett annat är den korrekta modellen istället $N \in Po(\lambda_2)$. Ovanstående illustrerar en osäkerhet kring värdet på parametern λ . För att fånga upp denna osäkerhet kan vi betrakta



Figur 2.1: Skadekostnaders ankomst i tiden

parametern som en stokastisk variabel Λ . Antag att $\lambda_1 =$ ett visst väderförhållande, då kan ovanstående formaliseras genom att skriva $N | \Lambda = \lambda_1 \in Po(\lambda_1)$. Genom att ansätta en fördelning för Λ kan den obetingade fördelningen för N erhållas. Antag att Λ har en täthet $u(\cdot)$ sådan att $u(\lambda) > 0, \forall \lambda > 0$. Fördelningen för Λ kallas blandningsfördelning och N får en så kallad blandad Poissonfördelning. Vi skriver $N \in Po(\Lambda)$. Med ovanstående villkor ges sannolikhetsfunktionen för N av [5]:

$$p_k = \int_0^{\infty} P(N = k | \Lambda = \lambda) u(\lambda) d\lambda = \int_0^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} u(\lambda) d\lambda.$$

Genom att se på sannolikhetsfunktionen för N ovan framgår det att i blandad Poisson tillåts flera Poissonfördelningar viktat med den sannolikhet som fördelningen för Λ ger för att det är just denna Poissonfördelning som är den riktiga. Skattningen av parametervärdet i Poissonfördelningen erhålls som medelvärdet av antalet skador per år i datamaterialet. Att ett och samma parametervärde ska kunna ge en god anpassning är mindre troligt om vi ser en antydning till att en uppdelning av datamaterialet skulle kunna göras, sådan att respektive del ser ut att ha en egen Poissonfördelning. Om datamaterialet uppvisar sådana tendenser bör detta naturligtvis också visa sig genom att datamaterialet har en varians större än väntevärdet. I så fall kan vi förvänta oss att en blandad Poissonfördelning skulle kunna vara en bättre modell för fördelningen.

Antag att Λ följer en Gammafördelning så att för $\lambda > 0$: $u(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$

$\alpha, \beta > 0$ och vi skriver $\Lambda \in \text{Gamma}(\alpha, \beta)$. Med dessa förutsättningar ges sannolikhetsfunktionen för N av:

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

För en härledning av p_k se s.52 i [5]. Ovanstående visar att N har en Negativ binomialfördelning och med denna parametrisering skriver vi $N \in \text{NegBin}(\alpha, \beta)$.

Vi ska nu se på några egenskaper hos Poisson- respektive NegBin-fördelningen som kan vara intressanta då vi ska skatta parametrar och anpassa fördelningsfunktioner. Antag att $L \in \text{Po}(\lambda)$ och att $K \in \text{NegBin}(\alpha, \beta)$ då gäller att $E(L) = \text{Var}(L) = \lambda$, $E(K) = \frac{\alpha}{\beta}$ och $\text{Var}(K) = \frac{\alpha(\beta+1)}{\beta^2}$ och vi har att $\text{Var}(L)/E(L) = 1$ och $\text{Var}(K)/E(K) = 1 + 1/\beta$. Låt s^2 vara stickprovsvariansen och \bar{x} vara stickprovets medelvärde. För att en fördelning ska ge en bra anpassning krävs att datamaterialet har de egenskaper som fördelningen har. Antag att $\frac{s^2}{\bar{x}} \approx 1$, i så fall uppfylls villkoren i de båda fördelningarna empiriskt och med hänsyn till detta kan vi inte förvänta oss någon större skillnad i anpassningen mellan de två modellerna. Har vi istället att $\frac{s^2}{\bar{x}} \gg 1$ uppfylls inte villkoret i Poissonfördelningen medan NegBin-fördelningen fortfarande kan passa bra.

2.2 Uttunnad Poisson- och uttunnad blandad Poissonfördelning

Om vi för något $b > d > 0$ låter N vara antalet skador som överstiger d , X_i' vara storleken på den i :te skadan som överstiger d , $p = P(X_i' > b)$ och M antalet skador som överstiger b . Antag att $N \in \text{Po}(\lambda)$ då gäller enligt känd egenskap hos Poissonfördelningen att $M \in \text{Po}(p\lambda)$ och vi säger att M har en uttunnad Poissonfördelning. Vi ska här visa att motsvarande egenskap gäller även då N har en blandad Poissonfördelning. Med samma beteckningar som ovan antar vi nu istället att $N \in \text{Po}(\Lambda)$ då gäller att $M | N = k \in \text{Bin}(k, p)$ se s.77 [5] och sannolikhetsfunktionen för M är:

$$\begin{aligned} p_m &= \sum_{k=m}^{\infty} P(M=m | N=k)P(N=k) = \sum_{k=m}^{\infty} \binom{k}{m} p^m (1-p)^{k-m} \int_0^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} u(\lambda) d\lambda = \\ &= \int_0^{\infty} \left(\sum_{k=m}^{\infty} \binom{k}{m} p^m (1-p)^{k-m} \frac{\lambda^k}{k!} e^{-\lambda} \right) u(\lambda) d\lambda = \int_0^{\infty} e^{-\lambda} \frac{(p\lambda)^m}{m!} \left(\sum_{k=m}^{\infty} \frac{((1-p)\lambda)^{k-m}}{(k-m)!} \right) u(\lambda) d\lambda = \\ &= \int_0^{\infty} \frac{(p\lambda)^m}{m!} e^{-p\lambda} u(\lambda) d\lambda = \{x = p\lambda \quad dx = pd\lambda\} = \int_0^{\infty} \frac{x^m}{m!} e^{-x} \frac{u(\frac{x}{p})}{p} dx \end{aligned}$$

Antag att Λ har fördelningsfunktion $F_\Lambda(x)$ och täthet $u(x) = F'_\Lambda(x)$ då har $p\Lambda$ täthet

$$F'_{p\Lambda}(x) = \frac{d}{dx} P(p\Lambda \leq x) = \frac{d}{dx} P(\Lambda \leq x/p) = F'_\Lambda(x/p) = \frac{u(\frac{x}{p})}{p}$$

Ovanstående visar att M har en blandad Poissonfördelning med blandningsfördelning $p\Lambda$. Om Λ antas Gammafördelad så gäller att:

$$\frac{u(\frac{\lambda}{p})}{p} = \frac{(\beta/p)^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta/p}$$

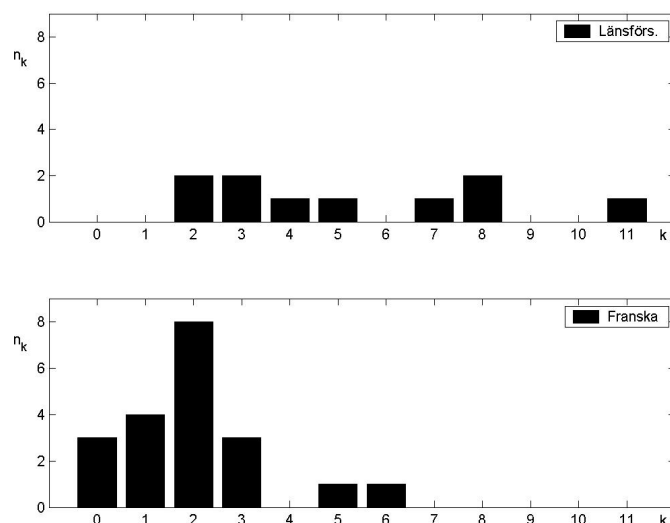
Dvs blandningsfördelningen är $\text{Gamma}(\alpha, \beta/p)$ och vi får att $M \in \text{NegBin}(\alpha, \beta/p)$. Att kunna ange fördelningen för M för olika värden på b är värdefullt i samband med vår tillämpning inom återförsäkring, med hänsyn till just denna egenskap innebär valet av blandad Poisson- istället för Poissonfördelning ingen nackdel.

2.3 Data

I datamaterialet från det franska bolaget är medelvärdet för antalet skador per år över nivån d $\bar{x} = 2.0$, varians $s^2 = 2.3$ och $s^2/\bar{x} = 1.2$. I datamaterialet från Länsförsäkringar har vi $\bar{x} = 5.3$, $s^2 = 9.3$ och $s^2/\bar{x} = 1.8$. Villkoret: $E(N)/\text{Var}(N) = 1$ är således empiriskt ungefärligen uppfyllt för datamaterialet från det franska bolaget medan avvikelserna från villkoret är större för datamaterialet från Länsförsäkringar. Därför kan vi förvänta oss att den största skillnaden i anpassning mellan Poisson- och NegBin-fördelningen ska framträda för datamaterialet från Länsförsäkringar.

Om vi ser på hur data ser ut för Länsförsäkringar och jämför detta med hur det ser ut för det franska bolaget så framgår det i figur 2.2 att antalet skador per år antar värden i ett större intervall. Det framgår dessutom att den relativa frekvensen för observationer som i absoluta termer ligger längre ifrån medelvärdet är större, vilket förklarar den högre variansen för detta datamaterial. Figuren visar också att antalet skador per år är högre för datamaterialet från Länsförsäkringar. Eftersom skadefrekvensen vanligtvis är högre för de skador där skadekostnaden är liten hade en möjlig förklaring till den högre frekvensen varit att storleken på capture limit i förhållande till medelskadan varit större för det franska bolaget, men som vi har konstaterat tidigare är så inte fallet (se avsnitt 1.4). En möjlig förklaring är istället att skillnaden beror på att i det franska datamaterialet förekommer endast stormskador medan i datamaterialet från Länsförsäkringar ingår ett antal olika skadehändelser. En annan faktor som påverkar frekvensen, i synnerhet frekvensen för skador över en viss nivå är exponeringen, dvs hur många och hur stora objekt som ingår och om de finns spridda över stora geografiska områden. För att få reda på om det finns några regionala skillnader mellan svenska och franska förhållande hade det varit önskvärt med två datamaterial som innehöll samma typer av skador.

Med tanke på att antalet observationer i datamaterialet är ganska litet så kommer stegen i den empiriska fördelningsfunktionen ibland att bli ganska stora, därför kan man knappast förvänta sig att de parametriska fördelningarna ska ge någon perfekt anpassning.



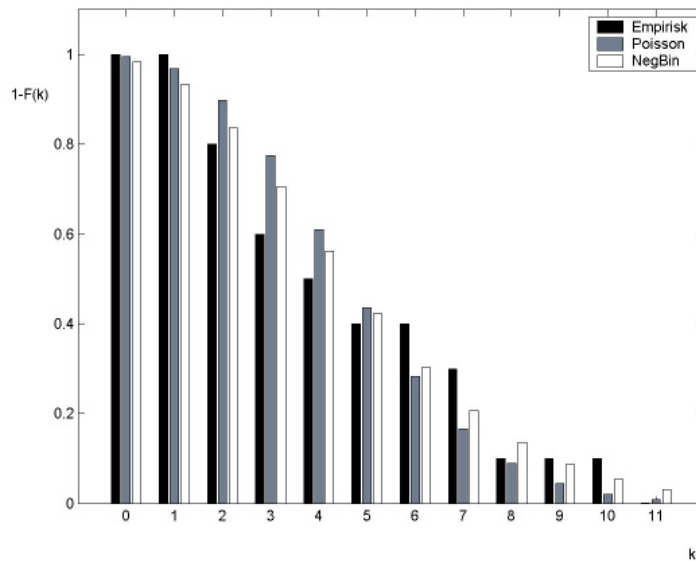
Figur 2.2: Antal år n_k med k stycken skador

Ett alternativ till att använda sig av parametriska fördelningar är att använda sig av den empiriska fördelningen. En uppenbar nackdel med detta är att sannolikheten för de utfall som inte förekommer i datamaterialet är noll. Detta innebär bland annat att sannolikheten för utfall större än den största observationen är noll vilket vi egentligen inte har någon anledning att tro på. Med de parametriska fördelningarna får vi sannolikheter även för utfall som inte har observerats vilket har betydelse då vi senare ska anpassa en fördelning för den totala kostnaden per år.

2.4 Anpassning av fördelningar för data

Anpassning av fördelningar görs här för respektive datamaterial. Efter en separat analys av anpassningarna görs en jämförelse av anpassningarna för de båda datamaterialen. De fördelningar som anpassas är alltså Poisson- och Negativ binomialfördelning. Skattningar av parametrar görs med maximumlikelihood- (ML) och momentmetoden (mom). I figurerna redovisas de resultat som erhållits med ML-metoden.

I Poissonfördelningen finns som vi tidigare sett en okänd parameter λ . I såväl ML- som momentmetoden skattas λ med hjälp av medelvärdet av antalet skador per år dvs $\hat{\lambda} = \bar{x}$. I NegBinfördelningen finns två okända parametrar α och β . ML- och momentmetoden ger olika skattningar av de två parametrarna. För detaljer angående skattningar hänvisas till s.48-56 i [5].



Figur 2.3: Anpassning av Poisson- och NegBinfördelning, Länsförsäkringar

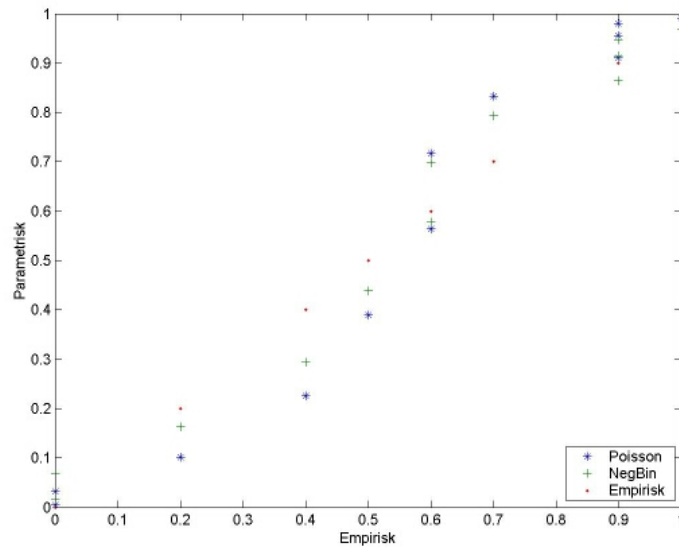
2.4.1 Länsförsäkringar

Följande paramterskattningar har erhållits för Poisson- respektive NegBinfördelningen:

	λ	α	β	α/β	$1 + 1/\beta$
ML	5.3	8.9451	1.6877	5.3	1.59
mom	5.3	9.0322	1.7042	5.3	1.58

Figur 2.3 visar plottar av överlevnadsfunktionen ($1-F(k)$) för $k = 0, 1, \dots, 11$. Figuren visar de två parametriska fördelningarnas anpassning parallellt med den empiriska fördelningen. I figuren framgår det att NegBinfördelningen ger en något bättre anpassning i de flesta punkter.

I figur 2.4 plottas värdet på fördelningsfunktionerna för de parametriska fördelningarna mot fördelningsfunktionen för den empiriska fördelningen i punkterna $k = 0, 1, \dots, 11$. Som referens visas de punkter som innebär en perfekt anpassning till den empiriska fördelningsfunktionen. Från denna figur framgår kanske något tydligare att det finns en viss skillnad i anpassning mellan de två parametriska fördelningarna.



Figur 2.4: Fördelningsfunktioner plottade mot varandra, Länsförsäkringar

2.4.2 Franska bolaget

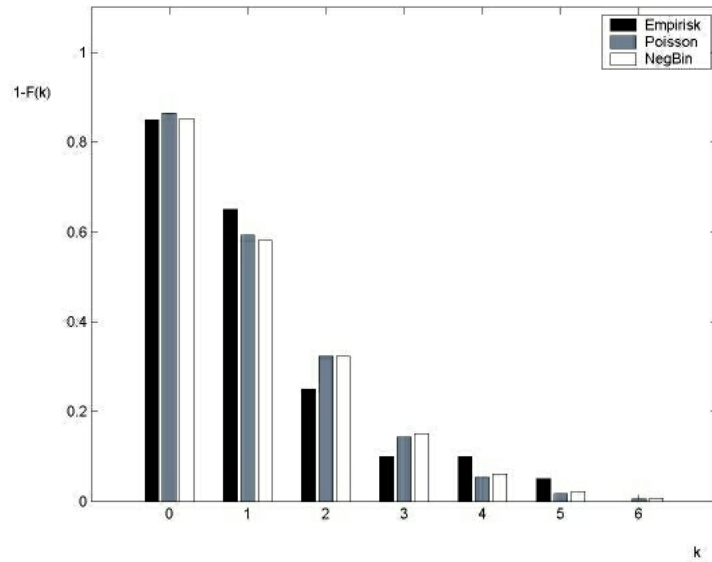
Följande paramterskattningar har erhållits för Poisson- respektive NegBinfördelningen:

	λ	α	β	α/β	$1 + 1/\beta$
ML	2.0	22.0702	11.0351	2.0	1.09
mom	2.0	20.0	10.0	2.0	1.1

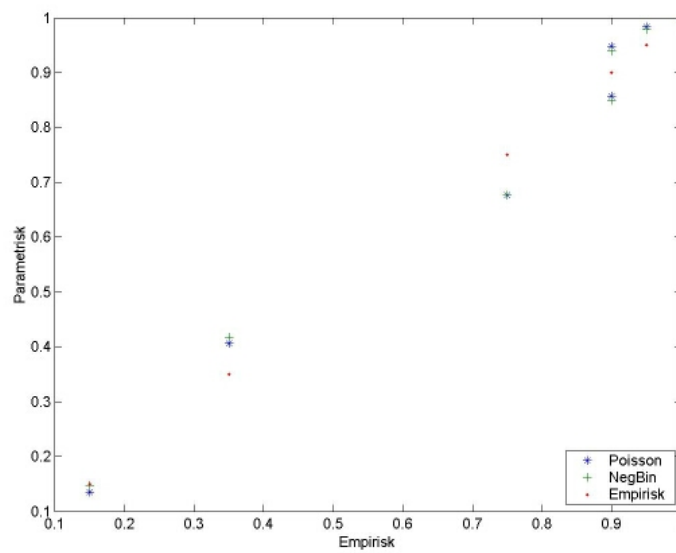
I detta avsnitt redogörs för de figurer som erhållits med datamaterialet för det franska bolaget, figurerna i detta avsnitt motsvarar figurerna från avsnitt 2.4.1 men med $k = 0, 1, \dots, 6$. I figur 2.5 är det svårt att se att det skulle vara någon skillnad mellan Poisson- och Negativ binomialfördelningen. Från föregående avsnitt såg vi att skillnaden möjligtvis framgick lite tydligare då fördelningsfunktionerna plottades mot varandra, men i figur 2.6 ser vi att inte heller denna plot tyder på att det skulle vara någon skillnad mellan de två fördelningarna.

2.4.3 Jämförelse av resultat för respektive datamaterial

Det har alltså visat sig att skillnaden mellan Poissonfördelningen och NegBin-fördelningen är större för datamaterialet från Länsförsäkringar än den är för det franska bolaget vilket vi förväntat oss med tanke på vad vi tidigare observerat för s^2 och \bar{x} , se avsnitt 2.3. Vi har tidigare visat att vi kan använda NegBinfördelningen



Figur 2.5: Anpassning av Poisson- och NegBinfördelning, franska bolaget



Figur 2.6: Fördelningsfunktioner plottade mot varandra, franska bolaget

på samma sätt som vi kan använda Poissonfördelningen se avsnitt 2.1 och 2.3. Då skillnaden i anpassningen är så pass stor som den visade sig vara för datamaterialet från Länsförsäkringar kan det synas välmotiverat att använda Negativ binomialfördelningen. När skillnaden mellan de två anpassningarna är så liten som den visade sig vara för det franska bolaget tyder detta på att Poissonfördelningen verkligen är rätt modell för detta datamaterial. För att slutligen svara på om användandet av NegBin-fördelningen kan motiveras trots att den ger omständligare skattningar och fördelningsfunktion ska vi i kapitel 4.4 undersöka om det gör någon nämnvärd skillnad i samband med vår tillämpning.

Kapitel 3

Fördelning för skadekostnaden

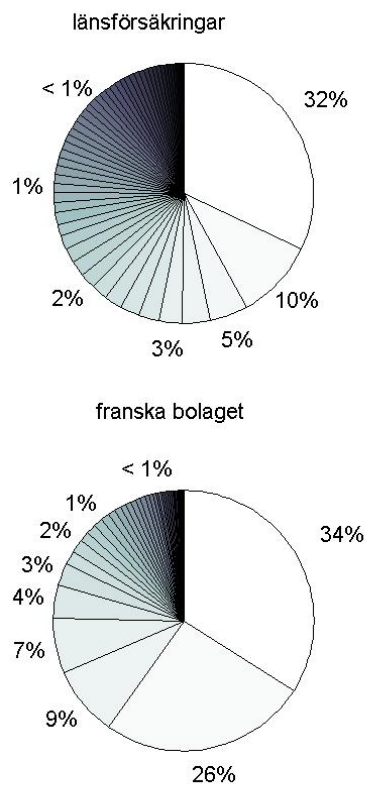
3.1 Data skadekostnaden

Då man vill anpassa en fördelning till de utbetalade ersättningsbeloppen, är för vissa försäkringstyper ett vanligt förekommande fenomen att ett relativt litet antal stora skador står för en avsevärd andel av den totala skadekostnaden. Då vi som här har ersättningsbelopp för katastrofskador syns fenomenet mycket tydligt.

Vi ska börja med att skapa oss en uppfattning om hur respektive datamaterial ser ut. Figur 3.1 visar de enskilda skadornas andel av den totala skadekostnad. I denna figur framgår det tydligt att de största skadorna svarar för en (mycket) stor andel av den totala kostnaden för perioden. I de båda datamaterialen står den största skadan för ca en tredjedel av den totala kostnaden. För det franska bolaget står de 4 största skadorna tillsammans för ca 75 % av den totala kostnaden under perioden – motsvarande siffra för Länsförsäkringar är ca 50 %.

Att ett litet antal skador står för en betydande andel av den totala skadekostnaden medför att den fördelning som ska användas för att modellera fördelningen för skadekostnaden måste vara av så kallad tjocksvansad typ. Tjocksvansad innebär att fördelningen ska ha en relativt sett stor andel av den totala sannolikhetsmassan i svansen det vill säga viss sannolikhet ska tillåtas även för stora eller mycket stora värden på den stokastiska variabeln i fråga. Två sådana fördelningar är Pareto- och Lognormalfördelningen. Dessa två fördelningar är de kanske två mest använda tjocksvansade fördelningarna inom sakförsäkringsmatematiken. Vi kommer att anpassa dessa fördelningar till den del av skadekostnaden som överstiger capture limit d dvs de kommer att anpassas till observationerna $x_1 - d, \dots, x_n - d$.

Det finns många tänkbara fördelningar för ovanstående ändamål, här görs ingen ansats att utreda huruvida de fördelningar som används här är de bästa möjliga för ändamålet. Snarare kommer intresset att fokuseras på metoder för att erhålla återförsäkringkostnaden och på metoder för att erhålla fördelningen för totalkostnaden.



Figur 3.1: Enskilda skadors andel av den totala skadekostnaden

På grund av POT-modellen föredrar vi att arbeta med Paretofördelningen, men en anpassning av Lognormalfördelningen görs för att kontrollera att Paretofördelningen verkligen ger en god anpassning till data.

3.2 Resultat från anpassning av fördelning för skadekostnaden

Eftersom själva anpassningen av fördelningarna till skadekostnaderna inte är det viktiga i denna uppsats kommer endast en kortfattad sammanfattning av resultaten att redovisas här. I appendix ges en noggrann redogörelse av anpassningen för Pareto- och Lognormalfördelningen för de båda datamaterialen.

I datamaterialet från det franska bolaget gav Paretofördelningen den klart bästa anpassningen över hela det intervall som observationerna låg i. För detta datamaterial är det således ingen tvekan om att Paretofördelningen av dessa två är den bästa modellen för skadekostnaden. För datamaterialet från Länsförsäkringar var det inte lika tydligt vilken fördelning som gav den bästa anpassningen. Det framgick dock att Lognormalfördelningen gav bäst anpassning för de mindre skadorna medan fördelningen för de stora skadorna modellerades bäst mha Paretofördelningen. Eftersom ett a vs b kontrakt skyddar mot skador som överstiger excesspunkten b är det anpassningen av fördelningen för dessa skador som ska betraktas som viktigast. Detta talar för att använda Paretofördelningen även för datamaterialet från Länsförsäkringar. För Lognormalfördelningen såg vi att $F_{X'-d}(x) \rightarrow 1$ då $x \rightarrow \max(x_1, \dots, x_n)$ vilket innebär väldigt liten sannolikhet ges för händelsen $\{X' - d > x\}$. Paretofördelningen gav större sannolikhet för denna händelse vilket innebär att anpassningen av Paretofördelningen blev mer långsvansad. Detta måste betraktas som en tilltalande egenskap eftersom vi egentligen inte har någon anledning att tro att vi i framtiden inte kommer att observera någon skadekostnad som är större än den största i datamaterialet.

Kapitel 4

Fördelning för total skadekostnad per år

4.1 Data total skadekostnad

För att jämföra de två rekursionerna kommer vi att inleda detta kapitel med att studera fördelningen för $S' = \sum_{i=1}^N (X'_i - d)$, där d är vår capture limit. Därför görs här en genomgång av vad våra datamaterial innehåller i detta avseende. I våra datamaterial från Länsförsäkringar och det franska bolaget har vi som vi tidigare sett varje år k en observation på antalet katastrofskador över nivån d , som vi kan kalla n_k , och observationer på de tillhörande skadekostnaderna x_{k1}, \dots, x_{kn_k} där $x_{ki} > d$, $i = 1, \dots, n_k$. Den observerade årliga totalkostnaden över nivån d som vi kan kalla s_k fås genom att summera den för d överskridande delen av kostnaden så att $s_k = \sum_{i=1}^{n_k} (x_{ki} - d)$. Vi har med dessa beteckningar för de båda datamaterialen tio respektive tjugo observationer på s_k . Om vi med hjälp av dessa totalkostnader anpassar en fördelning för den årliga totalkostnaden över nivån d skulle, med tanke på den begränsade omfattningen i datamaterialet, precisionen i anpassningarna knappast kunna anses godtagbar.

4.2 Modeller för total skadekostnad per år

Givet fördelningarna för $X'_i - d$ och N samt att X'_1, X'_2, \dots är oberoende och likafördelade stokastiska variabler som dessutom är oberoende av N kan i princip fördelningen för $S' = \sum_{i=1}^N (X'_i - d)$ beräknas. Fördelningen kan beräknas med hjälp av rekursionsformler eller via approximation.

I detta avsnitt görs en genomgång av hur Panjers- och Ströters rekursionsformler används. För att kontrollera de båda rekursionernas precision görs en jämförelse med en simulering. En inbördes jämförelse mellan de båda rekursionerna görs för att peka på för- och nackdelar med respektive metod. Det är inte S' som används då vi ska bestämma återförsäkrarens kostnad, men eftersom syftet i första hand är att

jämföra de två rekursionerna är valet av den summa som ska studeras av underordnad betydelse. Att vi valt att studera fördelningen för S' beror på att det är enklast då vi från föregående kapitel har skattningar av fördelningarna för N och $X'_1 - d$.

4.2.1 Panjers rekursionsformel

Antag att vi ska beräkna fördelningen för $T = \sum_{i=1}^K V_i$. De antaganden som gäller för Panjers rekursionsformel är att K, V_1, V_2, \dots är oberoende samt att V_1, \dots, V_K är ickenegativa, heltalsvärda och likafördelade. Låt: $p_k = P(K = k)$, $f_k = P(V_i = k)$ och $g_k = P(T = k)$, $k = 0, 1, 2, \dots$. Antag att p_k uppfyller rekursionen $p_k = (a + b/k)p_{k-1}$ för $k \geq 1$. Då ges sannolikhetsfunktionen för T enligt Panjers rekursionsformel av [5]:

$$g_k = \sum_{j=1}^k \left(a + \frac{bj}{k}\right) f_j g_{k-j}, \quad k \geq 1$$

För rekursionens startvärdet g_0 gäller att $g_0 = \phi(f_0)$, där ϕ är den sannolikhetsgenererande funktionen för K . De fördelningar vi tidigare studerat för N , Poisson- och NegBin-fördelningen, uppfyller villkoren på p_k . Vidare är det rimligt att anta att N, X'_1, X'_2, \dots är oberoende samt att X'_1, X'_2, \dots är likafördelade.

Med X' menas nedan, såvida inget annat anges, ett godtyckligt X'_i . Uppenbarligen är villkoret att $X' - d$ är ickenegativt uppfyllt, däremot har vi ju tidigare modellerat fördelningen med en kontinuerlig fördelning, trots att kostnaden egentligen inte är kontinuerlig med tanke på att den måste anges i till exempel heltal kronor. Detta innebär i praktiken att $X' - d$ inte är kontinuerlig, det får dock anses som troligt att fördelningen för $X' - d$ modelleras bäst med någon kontinuerlig fördelning, vilket innebär att fördelningen för $X' - d$ inte uppfyller villkoren för Panjers rekursionsformel. Detta kan vi råda bot på genom att diskretisera fördelningen och därmed få sannolikheter av typen $f_k = P(X' - d = kc) = F((k + 1/2) * c) - F((k - 1/2) * c)$, $k = 0, 1, 2, \dots$ se s.122 i [3] c anger hur fin diskretiseringen görs, tex $c = 1$ miljon. Enligt ovanstående antar $X' - d$ värdena $0, c, 2c, 3c, \dots$, detsamma gäller i så fall även för S' . Nu är villkoren för att kunna använda Panjers rekursionsformel uppfyllda.

4.2.2 Ströters rekursionsformel

Ströters rekursionsformel skulle kunna sägas vara en motsvarighet till Panjers rekursionsformel då V har en kontinuerlig fördelning. Om vi tror på att V har en kontinuerlig fördelning så innebär diskretiseringen av fördelningen för V en approximation. I Ströters rekursionsformel förutsätts att V har en kontinuerlig fördelning och vid beräkning används tätheten för V istället för f_k som vi såg tidigare. Således görs ingen approximation i detta avseende. Valet av steglängd i den metod för numerisk integration som används för att beräkna integralen i Ströters rekursionsformel innebär dock en approximation. Med samma beteckningar som tidigare men där $f(x)$

är tätheten för V ges fördelningsfunktionen för T enligt [5]:

$$G(y) = \sum_{k=0}^{\infty} p_k F^{k*}(y) = p_0 + \sum_{k=1}^{\infty} p_k F^{k*}(y) =$$

$$p_0 + \sum_{k=1}^{\infty} p_k \int_0^y f^{k*}(x) dx = p_0 + \int_0^y \sum_{k=1}^{\infty} p_k f^{k*}(x) dx$$

om vi låter

$$g(x) = \sum_{k=1}^{\infty} p_k f^{k*}(x), \quad x > 0,$$

fås det behändigare uttrycket:

$$G(y) = p_0 + \int_0^y g(x) dx$$

Genom att följa beviset för Panjers rekursionsformel fås att:

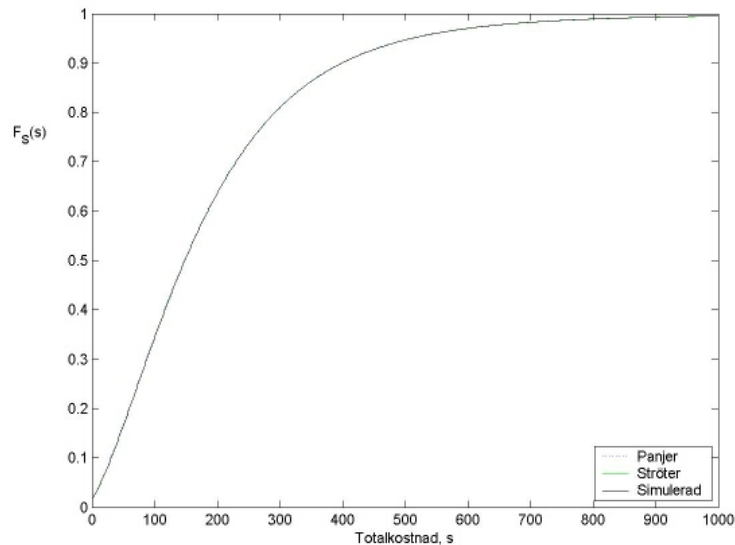
$$g(x) = p_1 f(x) + \int_0^x \left(a + \frac{bt}{x}\right) f(t) g(x-t) dt$$

Med hjälp av Simpsons formel fås en rekursionsformel för (approximativ) beräkning av $g(0), g(h), g(2h), \dots$ dessa värden används sedan i en ny numerisk integration som ger värden på fördelningsfunktionen $G(\cdot)$.

4.2.3 Simulering

Vi behöver en referens för att kontrollera precisionen för de två rekursionerna. En sådan referens kan skapas genom att simulera ett datamaterial med ett antal observationer på den årliga totalkostnaden s_1, s_2, \dots . En sådan observation s_i fås genom att först simulera antalet skador år i n_i och sedan simulera storleken på lika många skadekostnader $x_{i1} - d, \dots, x_{in_i} - d$ för att sedan beräkna summan av av dessa $s_i = \sum_{j=1}^{n_i} (x_{ij} - d)$. Genom att upprepa förfarande n gånger erhålls ett simulerat datamaterial med n observationer s_1, \dots, s_n på årliga totalkostnader. För att få tillförlitliga resultat med simulering bör n väljas stort.

I Monte Carlo metoden simuleras antalet skador per år och kostnaden för dessa genom att utgå från den fördelning som anpassats för respektive stokastisk variabel, dessa sätts sedan enligt ovanstående samman till simulerade totalkostnader se s.138 i [3]. Med hjälp av de simulerade observationerna på totalkostnaden kan en empirisk fördelningsfunktion skattas för den totala skadekostnaden per år, vilken kan användas som referens till de två rekursionerna.



Figur 4.1: Fördelning för S' , Länsförsäkringar

4.3 Beräkning av fördelning för $S' = \sum_{i=1}^N (X'_i - d)$

I avsnitt 4.3.1 och 4.3.2 antas att $N \in NegBin(\alpha, \beta)$ och att $X'_1 - d \in Pareto(\alpha', \gamma)$, dessa antaganden gäller för såväl simuleringarna som för beräkningarna med rekursionerna. Fördelningsfunktionen för simuleringen i nedanstående figurer har beräknats med hjälp av en miljon simulerade observationer. Då antalet simulerade värden är så pass stort kan denna fördelningsfunktion betraktas som den sanna förutsatt att fördelningarna för N och $X'_1 - d$ är de rätta. Huruvida fördelningarna verkligen är de rätta har i detta sammanhang ingen betydelse eftersom syftet i första hand är att undersöka hur väl de två rekursionen stämmer överens med simuleringen, samma förutsättningar gäller ju även för rekursionerna. För att få motsvarande precision i de båda rekursionerna är steglängden i diskretiseringen respektive stegen i integrationen desamma för Panjer respektive Ströter.

4.3.1 Länsförsäkringar

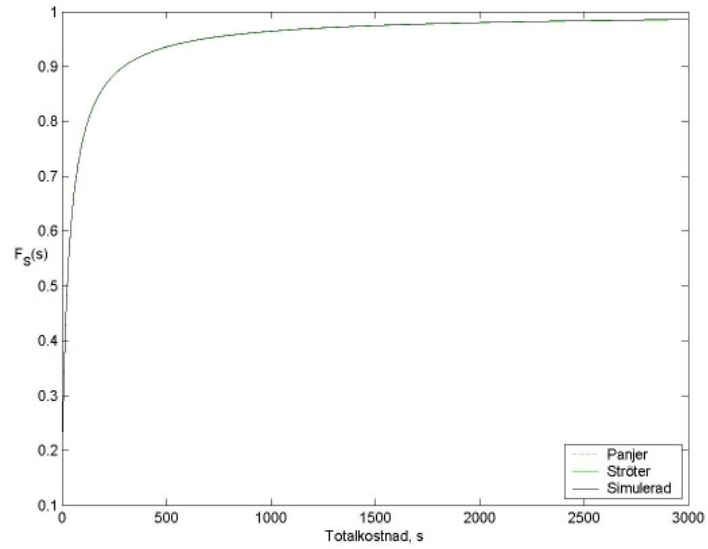
I figur 4.1 plottas fördelningsfunktionen för S' . Fördelningsfunktionerna i figuren är de som erhållits med hjälp av simuleringen respektive med Panjers- och Ströters rekursionsformler. I figuren är det svårt att se någon skillnad mellan de olika anpassningarna, men att de båda rekursioner ligger så nära simuleringen tyder på att båda rekursionerna modellerar fördelningen mycket bra. Om man studerar kortare delintervall kan man se att fördelningsfunktionen för Panjer ofta ligger något över simuleringen medan för Ströter ligger den ofta något under. För en vidare

kontroll av anpassning beräknas skillnaden mellan fördelningsfunktionerna för respektive anpassning i det intervall som visas i figuren. Låt P_k , S_k och Q_k vara fördelningsfunktionen $P(S' \leq kc)$, $k = 0, 1, \dots, 1000$ för Panjer, Ströter och simuleringen. Där $c =$ en miljon i kronor för Länsförsäkringar och i Euro för franska bolaget. Vi beräknar dessutom den totala avvikelsen mot simuleringen. Låt dessutom p_k , s_k och q_k vara sannolikhetsfunktionen $P(S' = kc)$, $k = 0, 1, \dots, 1000$. Det visar sig att $\sum_{k=1}^{1000} |P_k - S_k| = \sum_{k=1}^{1000} P_k - S_k > 0$ eftersom Panjer i varje punkt ligger över Ströter. Vidare har vi att $\sum_{k=1}^{1000} |P_k - Q_k| > \sum_{k=1}^{1000} P_k - Q_k > 0$ där summan av beloppet är ungefär dubbelt så stor som summan vilket visar att Panjer för det mesta ligger över men ganska ofta ligger under simuleringen. Slutligen har vi att $\sum_{k=1}^{1000} |Q_k - S_k| \approx \sum_{k=1}^{1000} Q_k - S_k > 0$ då Ströter ligger under simuleringen i nästan alla punkter. Vi får att $\sum_{k=1}^{1000} |p_k - q_k| = 0.022$ $\sum_{k=1}^{1000} |s_k - q_k| = 0.021$ således är den totala avvikelse något lägre för Ströter rekursionen, men skillnaden är marginell.

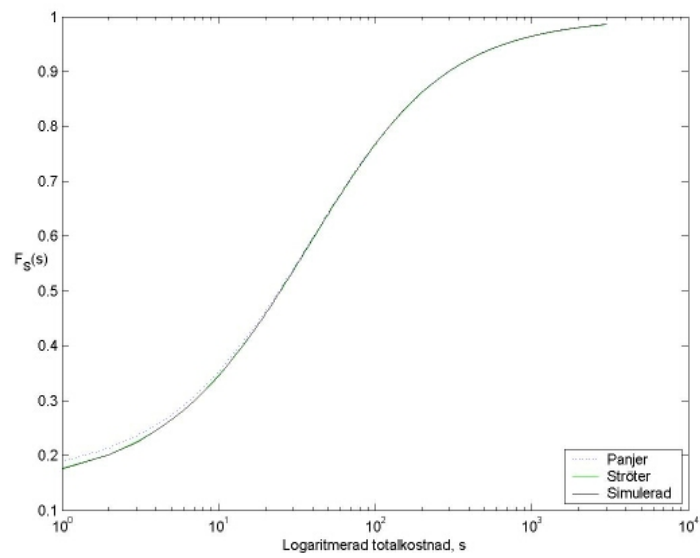
En annan kontroll av anpassningen fås genom att beräkna väntevärdet $E(S') = \sum_{k=0}^{1000} kP(S' = k)$ för respektive rekursion och jämföra detta med det teoretiska väntevärdet $E(S') = E(N)E(X' - d) = \alpha/\beta * (\alpha'/(\gamma - 1))$ som vi känner till. Denna jämförelse ska dock resultera i att $E(N)E(X' - d)$ är större än $\sum_{k=0}^{1000} kP(S' = k)$ eftersom vi stoppar rekursionerna efter 1000 steg, det vill säga de kan inte summeras till oändligheten som de teoretiskt sett ska. Därför är också en jämförelse av $\sum_{k=0}^{1000} kP(S' = k)$ för simuleringen med motsvarigheten för respektive rekursion intressant eftersom detta ger möjligheten att beräkna väntevärdet med samma antal iterationer. Det visar sig att väntevärdet för båda rekursionerna ligger under det teoretiska väntevärdet som sig bör och att väntevärdet för Panjer ligger under och för Ströter ligger över det för simuleringen. Avvikelsen från det simulerade väntevärdet är något lägre för Panjer än vad det är för Ströter.

4.3.2 Franska bolaget

Figur 4.2 motsvarar figur 4.1 men för datamaterialet för det franska bolaget. Även i denna figur är det svårt att se någon skillnad mellan de olika anpassningarna, men att de båda rekursioner ligger så nära simuleringen tyder ånyo på att båda rekursionerna modellerar fördelningen mycket bra. I figur 4.3 visas plotten av fördelningsfunktionen mot logaritmen av S' , här kan man se att för små värden på S' ligger Panjer något över den simulerade fördelningen medan Ströter ligger mycket nära. Om man studerar kortare delintervall framgår det Panjer ligger något över simuleringen för små värden på S' , men att anpassningen är mycket bra för större värden på S' . Ströter ligger för nästan alla värden på S' något under simuleringen. För en vidare kontroll av anpassning beräknas den totala avvikelsen mellan respektive anpassning i det intervall som visas i figuren. Med samma beteckningar som i föregående avsnitt fås även här att $\sum_{k=1}^{3000} |P_k - S_k| = \sum_{k=1}^{3000} P_k - S_k > 0$ vilket visar att Panjer i varje punkt ligger över Ströter. Vidare har vi att $\sum_{k=1}^{3000} |P_k - Q_k| > \sum_{k=1}^{3000} P_k - Q_k > 0$ där summan av beloppet är ganska mycket större vilket visar att



Figur 4.2: Fördelning för S' , franska bolaget



Figur 4.3: Fördelning för S' plottad mot $\log(S')$, franska bolaget

Panjer för det mesta ligger över men ganska ofta ligger under simuleringen. Slutligen har vi att $\sum_{k=1}^{3000} |Q_k - S_k| \approx \sum_{k=1}^{3000} Q_k - S_k > 0$ vilket visar att Ströter ligger under simuleringen i nästan hela intervallet. Vi får att $\sum_{k=1}^{1000} |p_k - q_k| = 0.50$ och $\sum_{k=1}^{1000} |s_k - q_k| = 0.55$ således är den totala avvikelse något lägre för Panjer rekursionen.

Vi jämför också $E(S') = \sum_{k=0}^{3000} kP(S' = k)$ för respektive rekursion med det teoretiska väntevärdet $E(S')$ som vi känner. Det visar sig att väntevärdet för båda rekursionerna ligger under det teoretiska väntevärdet som sig bör samt att väntevärdet både för Panjer och för Ströter ligger under väntevärdet för simuleringen. Avvikelsen från det simulerade väntevärdet är något lägre för Ströter än vad det är för Panjer.

De för- och nackdelar som har upptäckts genom att arbeta med de två rekursionerna sammanfattas nedan utan inbördes ordning.

Fördelar med Panjers rekursionsformel:

- Lättare att skriva program för beräkningar;
- Enkel beräkningsprocedur som medför korta beräkningstider för programmet.

Nackdelar med Panjers rekursionsformel:

- Val av metod för diskretisering ej entydigt given;
- $P(X' - d = 0) \neq 0$ vilket den borde vara här, detta genererar ett fel i g_0 som följer med i varje iteration. Felet minskas genom att använda kortare steglängden i diskretiseringen.

Fördelar med Ströters rekursionsformel:

- Fördelningen för en kontinuerlig variabel går att använda direkt, ur denna aspekt är den enkel att arbeta med.

Nackdelar med Ströters rekursionsformel:

- Mer omständligt att skriva program för beräkningar;
- Mer komplicerad beräkningsprocedur vilket medför längre beräkningstider för programmet;
- Eftersom V se 4.2.1 förutsätts ha en kontinuerlig fördelning går metoden ej att använda då V har en diskret- eller blandad fördelning, vilket innebär att den inte går att använda för excess of loss.

Vi har i ovanstående sett att avvikelsen mot simuleringen förefaller vara i stort sett lika för de båda rekursionerna. Dessutom såg vi att Panjers rekursionsformel har lägre beräkningskomplexitet och att rekursionen dessutom går att använda för såväl diskreta-, blandade- som kontinuerliga fördelningar. Mot bakgrund av ovanstående blir slutsatsen att fördelarna med Panjers rekursionsformel överväger.

4.4 Beräkning av fördelning för S

Det som avgör storleken på återförsäkrarens kostnad är i båda fallen ($a \leq b$ och $a \leq b \leq c$) storleken på S . Som vi har sett tidigare kan vi betrakta S som $S = \sum_{i=1}^K Y_i$ eller som $S = \sum_{i=1}^M Y'_i$. Med hjälp av Panjers rekursionsformel kan vi beräkna fördelningen för S antingen genom att känna till fördelningen för K och Y_1 eller genom att känna till fördelningen för M och Y'_1 . Eftersom vi kan anta att Y_i respektive Y'_i är likafördelade för alla i försätter vi i nedanstående utan index. Vi ska här se hur vi får fördelningar för Y och Y' . För att sedan beräkna fördelningen för S använder vi Panjers rekursionsformel på samma sätt som beskrivits tidigare. Om vi låter F_Y vara fördelningsfunktionen för Y så gäller enligt tidigare definition av Y se (1.1) att:

$$F_Y(y) = \begin{cases} 0 & \text{om } y < 0 \\ F_X(y+b) & \text{om } 0 \leq y < a \\ 1 & \text{om } y \geq a \end{cases}$$

Tillsammans med fördelningen för N kan vi efter diskretisering av fördelningen för Y beräkna fördelningen för S .

Vi har tidigare konstaterat att Y' har samma fördelning som $Y \mid Y > 0$, fördelningen för Y' ges därför av:

$$P(Y' \leq y) = P(Y \leq y \mid Y > 0) = \begin{cases} 0 & \text{om } y \leq 0 \\ \frac{P(b < X \leq y+b)}{P(X > b)} & \text{om } 0 < y < a \\ 1 & \text{om } y \geq a \end{cases}$$

Om vi låter X''_j vara den j :te skadan sådan att $X_i > b$ så gäller att $X'' - b$ har samma fördelning som $X - b \mid X > b$. Fördelningen för $X'' - b$ ges av:

$$P(X'' - b \leq x) = P(X - b \leq x \mid X > b) = \frac{P(b < X \leq x+b)}{P(X > b)}, \quad x > 0$$

Ur definition av Y' följer att Y' har samma fördelning som $\min\{X'' - b, a\}$, fördelningen för $\min\{X'' - b, a\}$ ges av:

$$P(\min\{X'' - b, a\} \leq x) = \begin{cases} 0 & \text{om } x \leq 0 \\ \frac{P(b < X \leq x+b)}{P(X > b)} & \text{om } 0 < x < a \\ 1 & \text{om } x \geq a \end{cases}$$

Eftersom vi från POT-modellen känner till fördelningen för $X' - b$ kan vi enkelt få fördelningen för Y' . Tillsammans med fördelningen för M kan vi efter en diskretisering av fördelningen för Y' beräkna fördelningen för S . Den uttunning som används för att få fördelningen för M är $P(X' > b) = P(X' - d > b - d) = \left(\frac{\alpha}{\alpha + b - d}\right)$. Fördelen med att använda Y' är att diskretiseringen blir enklare då vi betingat bort nollan dessutom framstår det som mer intuitivt riktigt och kanske elegantare lösning att inte summera ett antal variabler med värdet noll. Om vi ska använda oss

av fördelningen för Y måste vi dessutom ta hänsyn till att vi har fördelningen för $X' - d$ och inte för X vilket krånglar till det hela något. Väntevärdet för återförsäkrarens kostnad fås enligt:

$$E(Z_0) = \sum_{k=0}^{a-1} k P(S = k) + a \sum_{k=a}^{\infty} P(S = k)$$

för a xs b med noll reinstamants och

$$E(Z_0) = \sum_{k=1}^{a-1} k P(S = c + k) + a \sum_{k=a}^{\infty} P(S = c + k)$$

för a xs b xs c med noll reinstamants.

Tabellerna nedan visar väntevärden beräknade för återförsäkrarens kostnad tillsammans med limiter och excesspunkter. Värdena har beräknats med NegBin- respektive Poissonfördelning som modell för M .

Tabell med resultaten för datamaterialet från det franska bolaget:

Limit	Excesspunkt	NegBin	Poisson	Avvikelse i %
130	175	13.613	13.645	0.2
145	305	10.569	10.586	0.2
150	450	8.378	8.389	0.1
200	600	8.866	8.875	0.1
200	800	7.220	7.226	<0.1
200	1000	6.127	6.131	<0.1

Tabell med resultaten för datamaterialet från det Länsförsäkringar:

Limit	Excesspunkt	Aggregat	NegBin	Poisson	Avvikelse i %
50	50	0	24.556	25.204	2.6
50	50	25	15.217	15.362	1.0
500	3000	0	0.034	0.034	0

Ovanstående visar att skillnaden mellan de två modellerna är liten. För det franska bolaget måste skillnaden anses vara för liten för att motivera användandet av NegBin som modell för antalet skador per år. För Länsförsäkringar kan NegBin modellen vara ett alternativ. Skillnaden mätt i pengar är i vissa fall inte försumbar, beloppen ovan är ju angivna i miljoner Euro respektive kronor, men vi ska komma ihåg att ovanstående väntevärden endast är ett riktmärke för vad en rättvis premie kan vara. Premien fastställs som tidigare nämnts genom anbudsförfarande och förhandling. Detta innebär dock inte att metoden för att beräkna väntevärdet är av mindre värde eftersom det oavsett hur premien fastställs är viktigt att känna till vad

en rimlig premie för återförsäkringen är. Vi har heller ingen anledning att förkasta NegBinfördelningen som modell, för andra datamaterial kan det visa sig att NegBinfördelningen ger en klart bättre anpassning till data än vad Poissonfördelningen gör.

Bilaga A

Fördelning för skadekostnaden

A.1 Anpassning av fördelning för skadekostnaden

Anpassning av fördelningar görs här först separat för respektive datamaterial. De fördelningar som anpassas är Pareto- och Lognormalfördelningen. De parameter-skattningar som redovisas görs med maximumlikelihoodmetoden (ML-metoden).

I Paretofördelningen finns två parametrar, värdet på dessa erhålls genom att lösa den sk ML-ekvationen vilket görs numeriskt, som startvärde används här momentmetodens skattning som enkelt löses analytiskt. Även i Lognormalfördelningen finns två parametrar. För Lognormalfördelningen erhålls skattningen analytiskt med ML-metoden.

Vi har fördelningsfunktion

$$F(x) = P(X \leq x) = 1 - \left(\frac{\alpha}{\alpha + x}\right)^\gamma$$

där α och γ är parametrar för Paretofördelningen och fördelningsfunktion

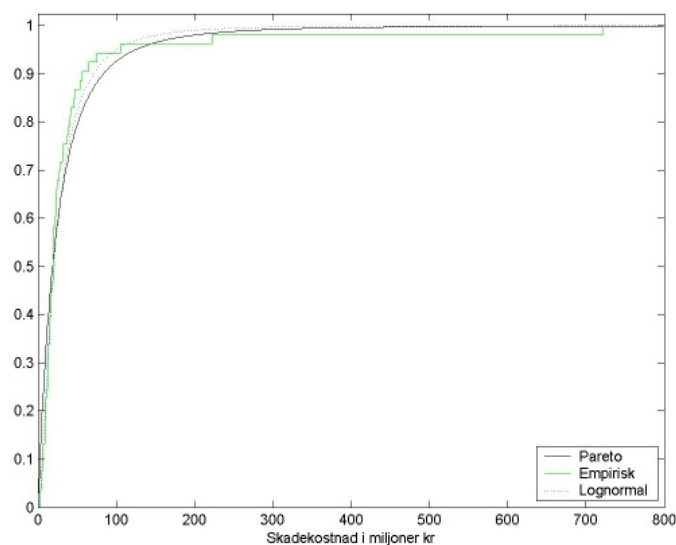
$$F(x) = P(e^X \leq x) = P(X \leq \log(x)) = P\left(\frac{X - \mu}{\sigma} \leq \frac{\log(x) - \mu}{\sigma}\right) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right)$$

där μ och σ är parametrar i Lognormalfördelningen.

A.1.1 Länsförsäkringar

Följande paramterskattningar har erhållits för Pareto- respektive Lognormalfördelningen:

	α	γ	μ	σ
ML	$8.2655 * 10^7$	3.1811	16.9276	0.9016

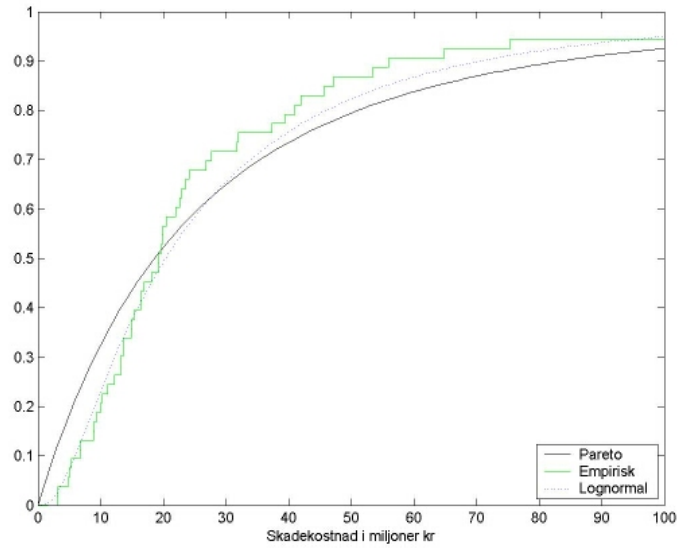


Figur A.1: Fördelning för skadekostnaden, Länsförsäkringar

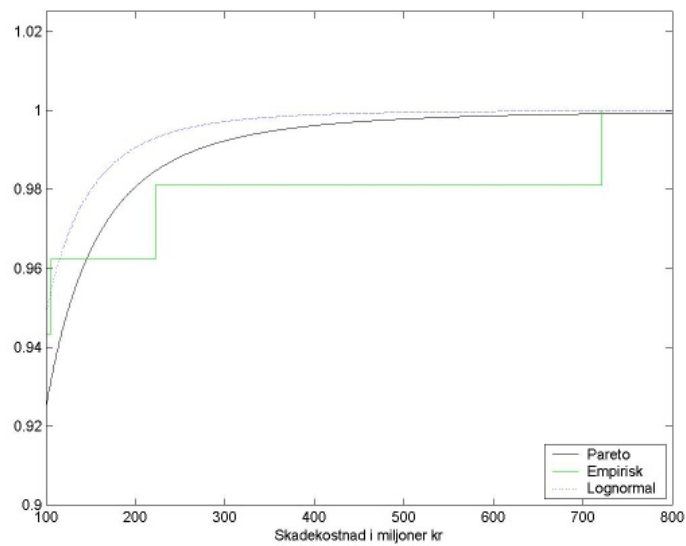
I figur A.1 som visar de anpassade fördelningsfunktionerna plottade mot skadekostnaden är det, för värden på skadekostnaden under ca 40 miljoner, svårt att se någon större skillnad i anpassningen mellan de två fördelningarna. För skadekostnader mellan ca 40 och 120 miljoner är anpassningen för Lognormalfördelningen bättre, men för värden över 120 miljoner verkar anpassningen för Paretofördelningen vara bättre.

Figur A.2 visar samma plottar som i A.1 men för skadekostnader i intervallet 0 till 120 miljoner. I denna figur framgår det att Lognormalfördelningen på det hela taget ger en bättre anpassning i intervall, undantaget är ett litet intervall under ca 30 miljoner där kurvorna för Pareto- och Lognormalfördelningen skär varandra. I figur A.3 visas slutligen plotten i det återstående intervallet mellan 120 och 800 miljoner här framgår det tydligt vilket vi delvis redan konstaterat att Paretofördelningen ger den bästa anpassningen.

Som en kontroll av ovanstående resultat har avvikelsen mot den empiriska fördelningsfunktionen för respektive fördelning beräknats i samma intervall som figurerna visar. Avvikelsen bekräftar det vi kunnat observera i figurerna nämligen att Lognormalfördelningen ger en klart bättre anpassning i det nedre intervallet medan Paretofördelningen ger en klart bättre anpassning i det övre intervallet. Då vi beräknar avvikelsen över hela intervallet fås en något lägre avvikelse för Lognormalfördelningen vilket är svårt att se genom att endast studera figurerna.



Figur A.2: Fördelning för skadekostnaden $[0, 1.2E^8]$, Länsförsäkringar



Figur A.3: Fördelning för skadekostnaden $[1.2E^8, 8E^8]$, Länsförsäkringar

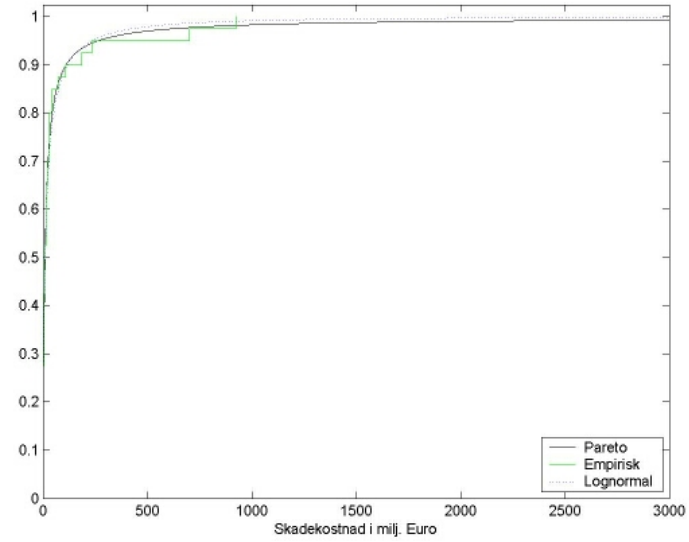
A.1.2 Franska bolaget

Följande parameterskattningar har erhållits för Pareto- respektive Lognormalfördelningen:

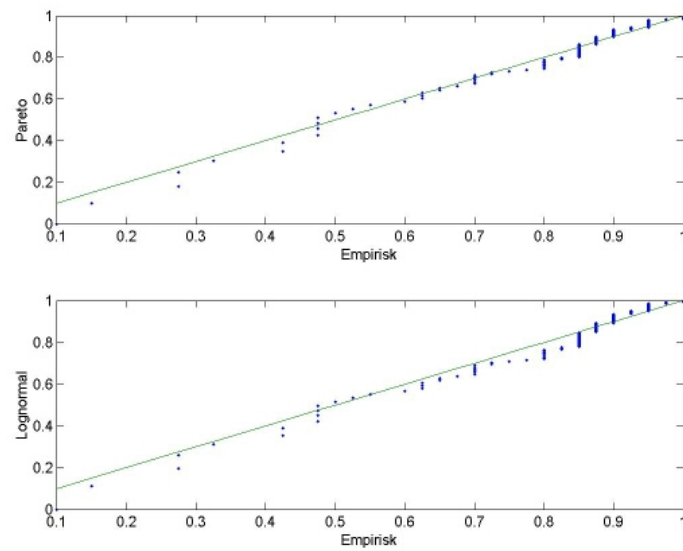
	α	γ	μ	σ
ML	16.3593	1.1145	2.7779	1.4469

I figur A.4 som visar fördelningsfunktionen plottad mot skadekostnaden ser vi att anpassningen av Pareto- och Lognormalfördelning ser likvärdiga ut för skadekostnader upp till ca 100 miljoner Euro. För skadekostnader mellan 100 och 900 miljoner Euro kan vi i samma figur se att Paretofördelningen ger en bättre anpassning än vad Lognormalfördelningen ger.

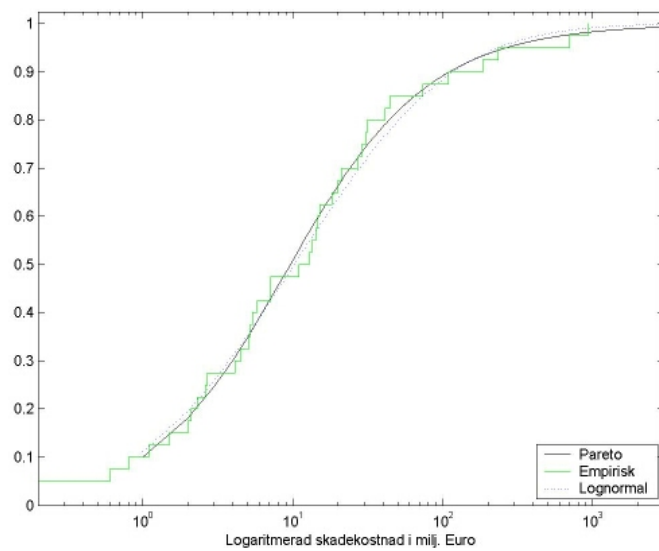
I figur A.5 som visar de skattade fördelningsfunktionerna plottade mot den empiriska fördelningsfunktionen framgår det att Paretofördelningen över så gott som hela intervallet ger en något bättre anpassning än vad Lognormalfördelningen gör. Den största skillnaden mot föregående figur är att den senare figuren visar att Paretofördelningen verkar ge en något bättre anpassning även för de mindre skadekostnaderna. Ovanstående framgår kanske ännu tydligare i figur A.6 som visar samma plot som figur A.4 med den skillnad att det är den logaritmerade skadekostnaden som plottas mot fördelningsfunktionen. En kontroll av de resultat vi observerat i figurerna får vi genom att beräkna den totalavvikelsen som ger en betydligt lägre avvikelse för Paretofördelningen.



Figur A.4: Fördelning för skadekostnaden, franska bolaget



Figur A.5: Plot av fördelningsfunktioner, franska bolaget



Figur A.6: Fördelning för den logaritmerade skadekostnaden, franska bolaget

Litteraturförteckning

- [1] Blom, G. *Sannolikhets teori och statistikteori med tillämpningar*. Studentlitteratur. 2000.
- [2] Coles, S. *An introduction to statistical modeling of extreme values*. Springer. 2000.
- [3] Daykin, C, Pentikäinen, T & Pesonen, M. *Practical risk theory for actuaries*. Chapman & Hall 1994.
- [4] Gustafsson, B. *Återförsäkring* 2000.
- [5] Johansson, B. *Matematiska modeller inom sakförsäkring* 1997.
- [6] Persson, C. *Distributing reinsurance costs within the LF insurance group* 2002.
- [7] Rootzen, H & Tajvidi, N. *Extreme value statistics and wind storm losses: a case study* 1995.