



Stockholms  
universitet

# Extremvärdesteori och POT-modellen

Tove Nilsson

Kandidatuppsats 2009:4  
Matematisk statistik  
Juni 2009

[www.math.su.se/matstat](http://www.math.su.se/matstat)

Matematisk statistik  
Matematiska institutionen  
Stockholms universitet  
106 91 Stockholm

# Extremvärdesteori och POT-modellen

Tove Nilsson\*

Juni 2009

## Sammanfattning

Den här uppsatsen beskriver extremvärdesteorin samt dess koppling till den så kallade POT-modellen. Syftet är att utreda vilka villkor som måste uppfyllas för att dessa modeller ska kunna tillämpas. Det är även av intresse att förstå hur dessa modeller kan användas i försäkringssammanhang. Extremvärdesteorin beskriver vad maximum av ett antal slumpvariabler har för fördelning då de skalas om på ett lämpligt sätt. POT-modellen redogör för fördelningen för en enskild slumpvariabel givet att den överstiger en viss nivå. Slumpvariabler från fyra fördelningar kommer att studeras för att kunna förstå gränsfördelningarna. Extremvärdesteorin visas användbar då maximum kommer från stora stickprov, maximum ska dessutom skalas om på olika sätt beroende på till vilken extremvärdesfördelning konvergens kommer att äga rum. POT-modellen visas användbar då slumpvariablerna betingas överstiga en hög nivå, slumpvariablerna behöver skalas om på ett särskilt sätt även i POT-modellen. POT-modellen är tillämpbar vid prissättning av återförsäkring då återförsäkringskontrakt ofta ger skydd mot skador givet att de överstiger en hög nivå.

---

\*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: [tove@sss.se](mailto:tove@sss.se) . Handledare: Andreas Nordvall Lagerås.

## **Abstract**

This essay describes the extreme value theory and its connection to the POT method. The purpose is to investigate which conditions are necessary for using the two models. It also of interest to understand how these models can be used in insurance mathematics. The extreme value theory describes the distribution of centered and normalized maxima. The POT method gives details about the distribution of a single random variable given that it exceeds a threshold. We will examine the limit distribution of random variables from four distributions. It will be shown that the extreme value theory is useful when maxima comes from large samples. Maxima also have to be centered and normalized in a certain way depending on towards which extreme value distribution convergence will take place. The POT method is useful when the threshold that the random variables will exceed is high. The variables have to be centered and normalized in a certain way even in this model. The POT method is useful for pricing reinsurance since it usually gives protection to damages that exceeds a high threshold.

## Innehåll

<b>1</b>	<b>Introduktion</b>	<b>2</b>
1.1	Inledning . . . . .	2
1.2	Bakgrund och Syfte . . . . .	2
1.3	Metod . . . . .	2
<b>2</b>	<b>Extremvärdesteori</b>	<b>3</b>
<b>3</b>	<b>POT-modellen</b>	<b>8</b>
<b>4</b>	<b>Illustration av teorin</b>	<b>11</b>
4.1	Paretofördelningen . . . . .	12
4.1.1	Konvergens för maxima av Paretovariabler . . . . .	13
4.1.2	POT-modellen för Paretovariabler . . . . .	13
4.2	Lognormalfördelningen . . . . .	14
4.2.1	Konvergens för maxima av lognormalvariabler . . . . .	15
4.2.2	POT-modellen för lognormalvariabler . . . . .	15
4.3	F-fördelningen . . . . .	17
4.3.1	Konvergens för maxima av F-variabler . . . . .	17
4.3.2	POT-modellen för F-variabler . . . . .	18
4.4	U(0,1) . . . . .	20
<b>5</b>	<b>Tillämpningar</b>	<b>21</b>
5.1	Tillämpning av extremvärdesteorin . . . . .	21
5.2	Tillämpning av POT-modellen, återförsäkring . . . . .	21
<b>6</b>	<b>Slutsatser</b>	<b>24</b>
<b>7</b>	<b>Diskussion</b>	<b>25</b>
<b>8</b>	<b>Referenser</b>	<b>26</b>

## 1 Introduktion

### 1.1 Inledning

Extremvärdesteorin beskriver hur maximum av ett antal slumpvariabler konvergerar i fördelning då de skalas om på ett lämpligt sätt. Det finns tre stycken så kallade extremvärdesfördelningar, Fréchet-, Gumbel- och Weibullfördelningen. Dessa tre fördelningar kan modellera för hur stora de största slumpvariablerna kan komma att vara. I försäkringssammanhang skulle dessa slumpvariabler kunna ses som storlekar på skadebelopp som antas komma från en känd fördelning. Extremvärdesteorin gör det då möjligt att modellera hur de största skadorna under de kommande åren kan komma att drabba ett försäkringsbolag.

Den så kallade POT-modellen har en tydlig koppling till extremvärdesteorin. POT står för Peaks over Threshold. Givet att slumpvariabler överstiger en hög nivå, kan de antas ha en generaliserad Paretofördelning. I försäkringssammanhang är det intressant att förstå fördelningen för skadan givet att den överstiger någon hög nivå, eftersom återförsäkringskontrakt ofta ger skydd mot just sådana skador. POT-modellen kan användas för beräkning av förväntade kostnader för olika typer av återförsäkringar.

### 1.2 Bakgrund och Syfte

Extremvärdesteorin är relativt gammal då grunden lades för drygt 80 år sedan. Det finns många verk skrivna om extremvärdesteorin men trots det saknas det i flera av dem kopplingen till POT-modellen. Syftet med detta arbete är att sammanfatta extremvärdesteorin samt dess koppling till POT-modellen, för att förstå hur dessa modeller kan användas i försäkringssammanhang. Vilka villkor som måste uppfyllas för att modellerna ska vara användbara kommer att utredas i den här studien.

### 1.3 Metod

Först kommer extremvärdesteorin att sammanfattas med ord och formler. Extremvärdesteorin sammanfattas till stor del med hjälp av Embrechts et al. För att bevisa satserna i extremvärdesteorin ger Falk et al. bra ledning. Sedan beskrivs kopplingen till POT-modellen med ord och formler. Beviset för POT-modellen kräver ingen kännedom om hur slumpvariabler bör skalas om men beräkningen av dessa normeringskonstanter kommer att göras då det

är nödvändiga för att tillämpa POT-modellen. För att kunna säga något om hur väl extremvärdesteorin fungerar för små och stora antal slumpvariabler från olika fördelningar kommer ett antal Matlabfigurer att illustrera gränsfördelningarna. Eftersom det är svårt att kunna säga något om konvergens är god eller ej kommer ett antal slumpvariabler att simuleras, spridningen av dessa kan då hjälpa till att avgöra om konvergensen är tillräckligt god. Matlab kommer även att användas då villkor för POT-modellen ska redas ut. Avslutningsvis kommer extremvärdesteorin och POT-modellen tillämpas i varsitt exempel. Det ska illustrera vad de två modellerna har för användningsområden i försäkringssammanhang.

## 2 Extremvärdesteori

Extremvärdesteorin förklarar vad maximum av ett antal oberoende och lika fördelade slumpvariabler har för fördelning då de skalas om på lämpligt vis. Slumpvariablerna betecknas här  $(X_1, \dots, X_n)$  och antas vara oberoende och lika fördelade med en gemensam fördelningsfunktion  $F$ . Maximum av alla värden i en viss tidsperiod betecknas  $M_n$  där  $M_n = \max(X_1, \dots, X_n)$ . Fördelningsfunktionen för maximum är

$$P(M_n \leq x) = P(X_i \leq x)^n = F_X^n(x)$$

*Definition 1* (Övre begränsning). För  $X$  från en given fördelning  $F$ , lät  $\omega = \omega(F) = \sup(x : F(x) < 1)$ , så att  $\omega$  är det minsta värde för vilket  $P(X \leq \omega) = 1$ . Om  $X$  inte har någon övre gräns är  $\omega = \infty$ .

*Definition 2* (Långsvansad fördelning). Låt  $X$  komma från en given fördelning  $F$ . Då är  $F$  en långsvansad fördelning om  $E[X^n] = \infty$  för något (litet)  $n$ .

Extrema värden antas i svansen av fördelningsfunktionen.

$$P(M_n \leq x) = F_X^n(x) \rightarrow 0$$

då  $n \rightarrow \infty$  och  $x < \omega$

$$P(M_n \leq x) = F_X^n(x) = 1$$

då  $x \geq \omega$

$M_n$  konvergerar alltså i sannolikhet mot  $\omega$ ,  $(M_n \xrightarrow{p} \omega)$  då  $n \rightarrow \infty$ .

Det är alltså inte särskilt intressant att studera konvergensen för  $M_n$  utan normering, därför är det konvergens för  $\frac{M_n - d_n}{c_n}$  som ska studeras där  $c_n > 0$  och  $d_n$  är några lämpligt valda konstanter som kan bero på antalet  $n$ .

I försäkringssammanhang är extremvärdesteorin tillämpbar då ett försäkringsbolag kan behöva modellera för de största skadebeloppen som kan komma att betalas. Slumpvariablerna ses då som storlekar på skadebelopp och dessa skador inträffar oberoende av varandra med lika sannolikhet. Om försäkringsavtalen täcker översvämningar, stormar eller andra naturkatastrofer är de exempel på sådana försäkringsavtal som kan komma att kosta stora summor. Då är det lämpligt att kunna modellera för hur sannolika dessa stora skador är.

*Definition 3* (Extremvärdesfördelningar). Låt  $\alpha > 0$ . Följande tre fördelningar kallas extremvärdesfördelningar.

$$\begin{aligned} \text{Fréchet: } \Phi_\alpha(x) &= \begin{cases} 0 & \text{om } x \leq 0 \\ e^{-x^{-\alpha}} & \text{om } x > 0 \end{cases} \\ \text{Gumbel: } \Delta(x) &= e^{-e^{-x}} \quad \text{alla } x \\ \text{Weibull: } \Psi_\alpha(x) &= \begin{cases} e^{-(-x)^\alpha} & \text{om } x \leq 0 \\ 1 & \text{om } x > 0 \end{cases} \end{aligned}$$

Kopplingen mellan fördelningarna är

$$Y \sim \Phi_\alpha \iff \log Y^\alpha \sim \Delta \iff -Y^{-1} \sim \Psi_\alpha$$

De tre extremvärdesfördelningarna kan skrivas på en allmän form. Den generaliserade extremvärdesfördelningen har följande utseende:

*Definition 4* (Den generaliserade extremvärdesfördelningen).

$$G_\beta(x) = \begin{cases} e^{-(1+\beta x)^{-1/\beta}} & \text{om } \beta \neq 0 \\ e^{-e^{-x}} & \text{om } \beta = 0 \end{cases}$$

Både läge och skala kan varieras fritt då extremvärdesfördelningen förblir en extremvärdesfördelning. Då  $\beta \neq 0$  används sambandet att om  $X' \stackrel{d}{=} cX + d$  så är

$$F_X(x) = P(X \leq x) = P\left(\frac{X' - d}{c} \leq x\right) = P(X' \leq cx + d) = F_{X'}(cx + d)$$

för att kunna se att en extremvärdesfördelning kan skrivas som den generaliserade extremvärdesfördelningen. Då  $\beta = 0$  är den generaliserade extremvärdesfördelningen definierad som en Gumbelfördelning.



Fréchetfördelningen kan alltså skrivas som den generaliserande extremvärdesfördelningen:

$$\Phi_\alpha(x) = G_\beta(ax + b) \iff e^{-x^{-\alpha}} = e^{-(1+\beta(ax+b))^{-1/\beta}}$$

Detta ger att  $\alpha = 1/\beta$  och  $x = 1 + \beta(ax + b)$  så för att parametrisera  $\Phi_\alpha(x)$  som den generaliserande extremvärdesfördelningen sätts  $\beta = 1/\alpha$ ,  $a = 1/\beta$  och  $b = -1/\beta$ .

Även Weibullfördelningen kan skrivas som den generaliserande extremvärdesfördelningen:

$$\Psi_\alpha(x) = G_\beta(ax + b) \iff e^{-(-x)^\alpha} = e^{-(1+\beta(ax+b))^{-1/\beta}}$$

så för att parametrisera  $\Psi_\alpha(x)$  som den generaliserande extremvärdesfördelningen sätts  $\beta = -1/\alpha$ ,  $a = -1/\beta$  och  $b = -1/\beta$ .

**Sats 1** (Konvergens för maxima). *Antag att  $X_1, \dots, X_n$  är oberoende lika fördelade med fördelningsfunktion  $F$ . Låt  $M_n = \max(X_1, \dots, X_n)$ . Om det finns konstanter  $c_n > 0$  och  $d_n$  sådana att*

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} Y$$

*för någon slumpvariabel  $Y$ , då har  $Y$  en extremvärdesfördelning.*

Följande observation är intressant.  $M_n = \max(X_1, \dots, X_n)$  är maximum av ett  $n$  st slumpvariabler och då  $n$  är tillräckligt stort konvergerar  $\frac{M_n - d_n}{c_n}$  mot en gränsfördelning  $G$ .  $M_{nk}$  kan ses som maximum av  $k$  st maximum, eller maximum av  $nk$  stycken slumpvariabler. Även  $\frac{M_{nk} - d_{nk}}{c_{nk}}$  konvergerar mot gränsfördelningen  $G$  enligt Sats 1 då  $nk$  är tillräckligt stort.

$$P\left(\frac{M_{nk} - d_n}{c_n} \leq x\right) = P\left(\frac{M_n - d_n}{c_n} \leq x\right)^k$$

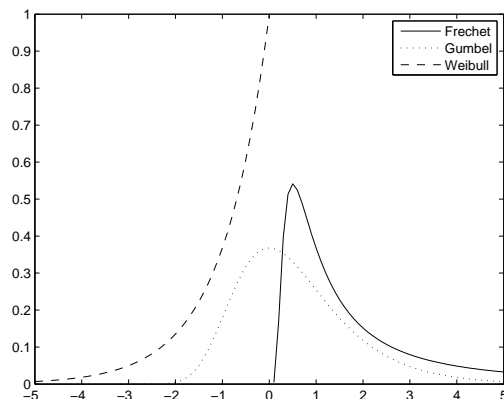
och därför är

$$P(M_{nk} \leq x) \approx G\left(\frac{x - d_{nk}}{c_{nk}}\right)$$

och

$$P(M_{nk} \leq x) \approx G^k\left(\frac{x - d_n}{c_n}\right)$$

Alltså är  $G$  och  $G^k$  lika förutom läge och skalparameter och båda tillhör extremvärdesfördelningarna.



Figur 1: Täthetsfunktioner för de tre extremvärdesfördelningarna,  $\alpha = 1$

Sats 1 ställer inga villkor på varken  $F$  eller normeringskonstanterna. Det kan vara behändigt att kunna säga något om gränsfördelningen utan att ställa några villkor, och Sats 1 kan bevisas. För att tillämpa gränsfördelningarna är det mer användbart att skriva om Sats 1 och bevisa den då villkor ställs på normeringskonstanterna. Beviset är approximativt då ett fullständigt bevis kräver kunskap om så kallad reguljär variation. För att förstå nästa sats behövs ytterligare en definition.

*Definition 5* (Kvantilfunktion). För  $X$  från en given fördelning  $F$ , låt  $q_n = q_n(F)$  vara  $\frac{1}{n}$ -kvantilen, det vill säga  $q_n$  är det minsta värdet som uppfyller att  $P(X > q_n) \leq \frac{1}{n}$ .

**Sats 2** (Konvergens av maxima med villkor). *Med definitioner enligt ovan.*

1.

$$\frac{M_n}{q_n} \xrightarrow{d} \Phi_\alpha \iff \lim_{t \rightarrow \infty} P(X > xt \mid X > t) = x^{-\alpha}$$

$$x > 0, \omega = \infty.$$

2.

$$\frac{M_n - q_n}{E[X - q_n \mid X > q_n]} \xrightarrow{d} \Delta \iff \lim_{t \rightarrow \omega} P\left(\frac{X - t}{E[X - t \mid X > t]} > x \mid X > t\right) = e^{-x}$$

$$\text{alla } x, \omega = \infty.$$

3.

$$\frac{M_n - \omega}{\omega - q_n} \xrightarrow{d} \Psi_\alpha \iff \lim_{t \rightarrow \infty} P\left(X > \omega - \frac{1}{tx} \mid X > \omega - \frac{1}{t}\right) = x^{-\alpha}$$

$$x > 0, \omega = 0.$$

Jämfört med Sats 1 ställs villkor på normeringskonstanterna.

För att maxima ska konvergera till  $\Phi_\alpha$  ska  $c_n = q_n$  och  $d_n = 0$ .

För att maxima ska konvergera till  $\Delta$  ska  $c_n = E[X - q_n \mid X > q_n]$  och  $d_n = q_n$ .

För att maxima ska konvergera till  $\Psi_\alpha$  sätts  $c_n = \omega - q_n$  och  $d_n = \omega$ .

*Bevis.* 1.

$$\lim_{t \rightarrow \infty} P(X > xt \mid X > t) = x^{-\alpha} \Rightarrow P(X > x) \approx ax^{-\alpha}.$$

Enligt definition 5 beräknas kvantilen till

$$q_n \approx (an)^{1/\alpha}.$$

Nu följer det att

$$\begin{aligned} P\left(\frac{M_n}{q_n} \leq x\right) &= P(M_n \leq q_n x) = P(X \leq q_n x)^n = (1 - P(X > q_n x))^n \\ &\approx \left(1 - \frac{x^{-\alpha}}{n}\right)^n \rightarrow e^{-x^{-\alpha}} = \Phi_\alpha(x), \end{aligned}$$

där alltså

$$\begin{aligned} P(X > q_n x) &\approx a(q_n x)^{-\alpha} = a(x(na)^{1/\alpha})^{-\alpha} \\ &= ax^{-\alpha} n^{-1} a^{-1} = \frac{x^{-\alpha}}{n}. \end{aligned}$$

2. Börja med att definiera  $e(t)$  som  $E[X - t \mid X > t]$ . Observera att den betingade sannolikheten

$$\begin{aligned} \lim_{t \rightarrow \omega} P\left(\frac{X - t}{E[X - t \mid X > t]} > x \mid X > t\right) &= e^{-x} \\ \iff P\left(\frac{X - t}{e(t)} > x\right) &= e^{-x} P(X > t) \end{aligned}$$

då  $t \rightarrow \omega$ . Enligt definition 5 konvergerar  $q_n \rightarrow \omega$  då  $n \rightarrow \infty$  samt  $P(X > q_n) \approx \frac{1}{n}$ . Nu följer det att

$$\begin{aligned} P\left(\frac{M_n - q_n}{e(q_n)} \leq x\right) &= P\left(\frac{X - q_n}{e(q_n)} \leq x\right)^n = \left(1 - P\left(\frac{X - q_n}{e(q_n)} > x\right)\right)^n \\ &\approx \left(1 - \frac{e^{-x}}{n}\right)^n \rightarrow e^{-e^{-x}} = \Delta(x). \end{aligned}$$

3. Den övre gränsen  $\omega$  sätts till 0 då det bara handlar om en förskjutning från  $X$  till  $X - \omega$ .

$$\lim_{t \rightarrow \infty} P\left(X > \omega - \frac{1}{tx} \middle| X > \omega - \frac{1}{t}\right) = x^{-\alpha} \Rightarrow P(X > -x) \approx ax^\alpha$$

för små  $x > 0$ . Kvantilen beräknas enligt definition 5 till

$$q_n \approx -(an)^{-1/\alpha}.$$

Detta ger att

$$\begin{aligned} P\left(\frac{M_n - \omega}{\omega - q_n} \leq -x\right) &= P\left(\frac{M_n}{-q_n} \leq -x\right) = P(M_n \leq q_n x) = P(X \leq q_n x)^n \\ &= (1 - P(X > q_n x))^n \approx \left(1 - \frac{x^\alpha}{n}\right)^n \rightarrow e^{-x^\alpha} = \Psi_\alpha(-x), \end{aligned}$$

där alltså

$$\begin{aligned} P(X > q_n x) &\approx a(xq_n)^\alpha = a(-x(an)^{-1/\alpha})^\alpha \\ &= -ax^\alpha a^{-1} n^{-1} = \frac{-x^\alpha}{n}. \end{aligned}$$

Om istället  $x < 0$  så erhålls

$$P\left(\frac{M_n}{-q_n} \leq x\right) \rightarrow e^{-(-x)^\alpha} = \Psi_\alpha(x).$$

□

Standardgränsvärdet  $(1 + \frac{1}{n})^n \rightarrow e$  då  $n \rightarrow \infty$  förekommer i bevisen.

### 3 POT-modellen

I den så kallade Peaks over Threshold-modellen antas att slumpvariablerna, givet att de överstiger en hög nivå, kommer från en generaliserad Paretofördelning. Detta kan motiveras med extremvärdesteori; om maximum konvergerar mot en viss extremvärdesfördelning så är det ekvivalent med att de betingade slumpvariablerna konvergerar mot en viss generaliserad Paretofördelning. I försäkringssammanhang är det intressant att förstå fördelningen för skadebeloppen givet att de överstiger en hög nivå eftersom återförsäkringskontrakt ofta ger skydd mot just sådana stora skadebelopp.

Antalet slumpvariabler som överstiger tröskeln  $t$  antas vara Poissonfördelade med intensitet  $\lambda$ . Detta beror på att en binomialfördelning,  $\text{Bin}(n, p)$  kan approximeras med en Poissonfördelning då sannolikheten  $p$  är liten. För att tillämpa POT-modellen bör tröskeln  $t$  vara hög, så att  $p$  är litet.

*Definition 6* (Den generaliserade Paretofördelningen). Den generaliserade Paretofördelningen har fördelningsfunktionen

$$H_\beta(x) = 1 + \log G_\beta(x) = \begin{cases} 1 - \frac{1}{(1+\beta x)^{1/\beta}} & \text{om } \beta \neq 0 \\ 1 - e^{-x} & \text{om } \beta = 0 \end{cases}$$

*Definition 7* (Mean excess function).  $e(t) = E[X - t \mid X > t]$  kallas för mean excess function,  $e(t)$  beräknar väntevärdet av  $X$  givet att  $X$  är större än  $t$ .

$e(q_n)$  har redan förekommit som en normeringskonstant, men Mean excess function har en speciell betydelse i POT-modellen då det är det sökta väntevärdet.

**Sats 3** (POT).

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} G_\beta \Rightarrow \left( \frac{X - d_n}{c_n} - u \mid \frac{X - d_n}{c_n} > u \right) \xrightarrow{d} (Y - u \mid Y > u)$$

där  $Y \sim H_\beta$ .

*Bevis.* Antag att  $X$  har fördelningsfunktion  $F$  som tidigare. Då har  $\frac{X-d_n}{c_n}$  fördelningen  $F(c_n x + d_n)$  och vidare  $F^n(c_n x + d_n) \rightarrow G_\beta(x)$ . För att följa beviset observeras att  $\log F(z) \approx -(1 - F(z))$  då  $z$  är stort.

$$\begin{aligned} P\left(\frac{X - d_n}{c_n} - u > v \mid \frac{X - d_n}{c_n} > u\right) &= \frac{P\left(\frac{X-d_n}{c_n} > v+u\right)}{P\left(\frac{X-d_n}{c_n} > u\right)} \\ &= \frac{1 - F(c_n(v+u) + d_n)}{1 - F(c_n u + d_n)} = \frac{-\log F(c_n(v+u) + d_n)}{-\log F(c_n u + d_n)} + o(1) \\ &= \frac{-n \log F(c_n(v+u) + d_n)}{-n \log F(c_n u + d_n)} + o(1) = \frac{-\log F^n(c_n(v+u) + d_n)}{-\log F^n(c_n u + d_n)} + o(1) \\ &\rightarrow \frac{-\log G_\beta(v+u)}{-\log G_\beta(u)} = \frac{1 - (1 + \log G_\beta(v+u))}{1 - (1 + \log G_\beta(u))} \\ &= \frac{1 - F_Y(v+u)}{1 - F_Y(u)} = P(Y - u > v \mid Y > u). \end{aligned}$$

□

Sats 3 ställer i likhet med Sats 1 inga krav på normeringskonstanterna. Sats 3 går att bevisa utan att bestämma dessa normeringskonstanter men för att kunna tillämpa Sats 3 är det nödvändigt att veta hur variablerna bör skalas om. I den här studien är det av intresse att ta reda på hur POT-modellen är användbar, därför kommer dessa normeringskonstanter att beräknas. För att göra Sats 3 ännu tydligare kommer  $u$  sättas till 0 i den nya satsen.  $u \neq 0$  säger mer då det betyder att även variabler som betingas på en ännu högre nivå än tröskeln  $t$  följer en generaliserad Paretofördelning. Här kommer  $u = 0$  att studeras och tröskeln som variablerna betingas överstiga är alltså lika med  $t$ .

**Sats 4** (POT med villkor). *Med definitioner enligt ovan,*

$$1. \quad \frac{M_n}{q_n} \xrightarrow{d} \Phi_\alpha \iff \left( \frac{X-t}{t\beta} \middle| X > t \right) \xrightarrow{d} Y$$

$Y \sim H_\beta, \beta = 1/\alpha, n \rightarrow \infty \text{ och } t \rightarrow \infty.$

$$2. \quad \frac{M_n - q_n}{e(q_n)} \xrightarrow{d} \Delta \iff \left( \frac{X-t}{e(t)} \middle| X > t \right) \xrightarrow{d} Y$$

$Y \sim H_\beta, n \rightarrow \infty \text{ och } t \rightarrow \infty.$

$$3. \quad \frac{M_n - \omega}{\omega - q_n} \xrightarrow{d} \Psi_\alpha \iff \left( \frac{X-t}{\beta(t-\omega)} \middle| X > t \right) \xrightarrow{d} Y$$

$Y \sim H_\beta, \beta = -1/\alpha, n \rightarrow \infty \text{ och } t \rightarrow \omega.$

Det är 1. och 2. som är intressant att använda då 3. har övre begränsning. För beräkning av normeringskonstanterna  $c_n$  och  $d_n$  i Sats 3 kommer beräkningarna som gjordes i avsnitt 2 till användning. För att Fréchetfördelningen ska kunna parametreras som den generaliserande extremvärdesfördelningen sätts  $\beta = 1/\alpha$ ,  $a = 1/\beta$  och  $b = -1/\beta$  och för att Weibullfördelningen ska kunna parametreras som den generaliserande extremvärdesfördelningen sätts  $\beta = -1/\alpha$ ,  $a = -1/\beta$  och  $b = -1/\beta$ . Då gäller för 1. att

$$\Phi_\alpha(x) = G_\beta\left(\frac{x-1}{\beta}\right)$$

eller det omvända sambandet

$$\Phi_\alpha(\beta x + 1) = G_\beta(x)$$

Nu är

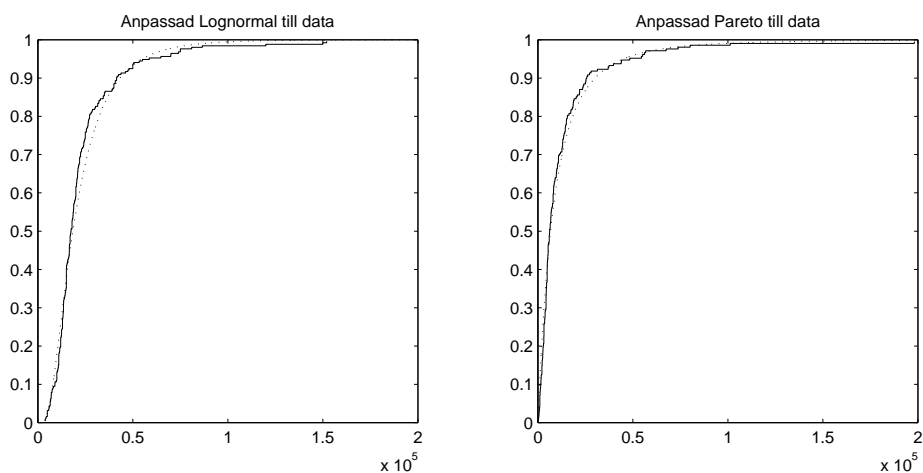
$$\Phi_\alpha(\beta x + 1) = P\left(\frac{M_n}{q_n} \leq \beta x + 1\right) = P\left(\frac{M_n - q_n}{q_n \beta} \leq x\right) = G_\beta(x).$$

Om  $\frac{M_n}{q_n}$  konvergerar mot en Fréchetfördelning konvergerar  $\frac{M_n - q_n}{q_n \beta}$  mot den generaliserade extremvärdesfördelningen och det är det sistnämnda som är av intresse här. Sedan följer Sats 4 om  $u = 0$  och den höga tröskel  $q_n$  betecknas med  $t$ . Beräkning av normeringskonstanterna i  $\beta$ . görs enligt samma mönster, för 2. observeras att en Gumbelfördelning är definierad som den generaliserade extremvärdesfördelning och normeringskonstanterna återfinns i Sats 2.

## 4 Illustration av teorin

För att illustrera extremvärdesteorin samt POT-modellen i ett antal grafer kommer Matlab att användas. Idén med att illustrera extremvärdesteorin är att ta reda på hur många observationer varje stickprov bör innehålla för att konvergens ska äga rum. När det gäller POT-modellen är det intressant att studera för vilka nivåer på tröskeln  $t$  som modellen är tillämpbar. Även hur många slumpvariabler som bör överstiga  $t$  för att det ska vara rimligt att använda modellen är intressant att utreda. Det är fyra fördelningar som kommer att studeras: Paretofördelningen, Lognormalfördelningen, F-fördelningen och U(0,1). Både Paretofördelningen och F-fördelningen är långsvansade och saknar övre begränsning. Maxima bör konvergera till en Fréchetfördelning för de två fördelningarna. Anledningen till att studera ytterligare en fördelning, F-fördelningen, är den att i Paretofördelningen kommer konvergensens visas vara lite mer uppenbar. Lognormal saknar även den övre begränsning men definieras inte som långsvansad. Maxima från lognormal kommer att konvergera mot en Gumbelfördelning. Fördelningar med övre begränsning är inte särskilt intressant här, men för att illustrera den tredje extremvärdesfördelningen kommer maxima från U(0,1) att studeras. För att illustrera konvergens mot extremvärdesfördelningarna kommer fördelningsfunktionen för extremvärdesfördelningen, det vill säga  $\Phi_\alpha(x)$ ,  $\Delta(x)$  eller  $\Psi_\alpha(x)$  att plottas mot  $F^n(c_n x + d_n)$  där  $F$  är fördelningsfunktionen för någon av de fyra nämnda fördelningarna. Variabler från fördelningarna kommer även att simuleras då det är intressant att se hur stor spridningen för dessa är. För att illustrera hur väl POT-modellen fungerar för olika värden på  $t$  kommer den generaliserande Paretofördelningen att plottas mot fördelningsfunktionerna från några av de nämnda fördelningarna, även här kommer slumpvariabler att simuleras. Slumpvariablerna från Paretofördelningen, lognormal och F-fördelningen kan ses som storlekar på skadebelopp. Paretofördelningen och lognormal är anpassbara till just skadebelopp där ett fåtal skador står för den största delen av den totala kostnaden. I figur 2 visas fördelningsfunktionen för Paretofördelningen och lognormal som har anpassats till befintliga data över skadebelopp från [4]. Figur 2 visar att de båda fördelningarna anpassas väl till skadebeloppen i de två dataseten. Med stor

sannolikhet är skadebeloppet på en låg nivå, men det finns ingen övre begränsning, därför kan skadebeloppet bli hur stort som helst. Nedan studeras konvergensen för de fyra fördelningarna. För konvergens av maxima kommer stickprovet att innehålla  $n$  stycken slumpvariabler för ett varierande antal  $n$ . Detta kan ses som  $n$  stycken skadebelopp under en tidsperiod. Maxima från 50 stycken tidsperioder kommer att simuleras då det är av intressa att studera spridningen av dessa. För illustration av POT-modellen kommer tröskelt  $t$  studeras för olika värden på  $t$ . Då POT-modellen gäller för enskilda slumpvariabler kan  $n$ , antalet slumpvariabler som simuleras varav några överstiger  $t$  väljas fritt.



Figur 2: Fördelningar anpassade till skadebelopp, Paretofördelningen är anpassad till data över skadebelopp vid stormar.

#### 4.1 Paretofördelningen

Paretofördelningen har täthetsfunktionen:

$$f(x) = \frac{\delta\gamma}{(1 + \delta x)^{1+\gamma}}$$

och fördelningsfunktionen:

$$F(x) = 1 - \left(\frac{1}{1 + \delta x}\right)^\gamma,$$

väntevärdet:

$$E[X] = \frac{1}{\delta(\gamma - 1)}$$

och den  $n$ :te kvantilen beräknas med hjälp av definition 5 till:

$$q_n = \frac{1}{\delta}(n^{\frac{1}{\gamma}} - 1).$$



### 4.1.1 Konvergens för maxima av Paretovariabler

Då slumpvariablerna kommer från en Paretofördelning gäller det att

$$\frac{M_n}{q_n} \xrightarrow{d} \Phi_\alpha(x) = e^{-x^{-\alpha}}$$

enligt Sats 2. Det är alltså en Fréchetfördelningen som bör passa till  $\frac{M_n}{q_n}$ . Konstanten  $\alpha$  i Fréchetfördelningen är lika med  $\gamma$  i Paretofördelningen. Detta kan motiveras med att då  $n \rightarrow \infty$  är

$$1 - F(x) = \frac{1}{(1 + \delta x)^\gamma} \sim \frac{1}{x^\gamma}.$$

I figur 3 illustreras hur väl en Fréchetfördelning passar för några olika stora  $n$ . Konvergensen är tydligast för stora  $n$ , med i jämförelse med spridningen på de simulerade slumpvariablerna är anpassningen god för  $n = 100$

### 4.1.2 POT-modellen för Paretovariabler

Då  $X \sim \text{Pareto}(\delta, \gamma)$  är enligt Sats 4

$$\frac{X - t}{t\beta} \sim H_\beta$$

där den generaliserande Paretofördelningen är

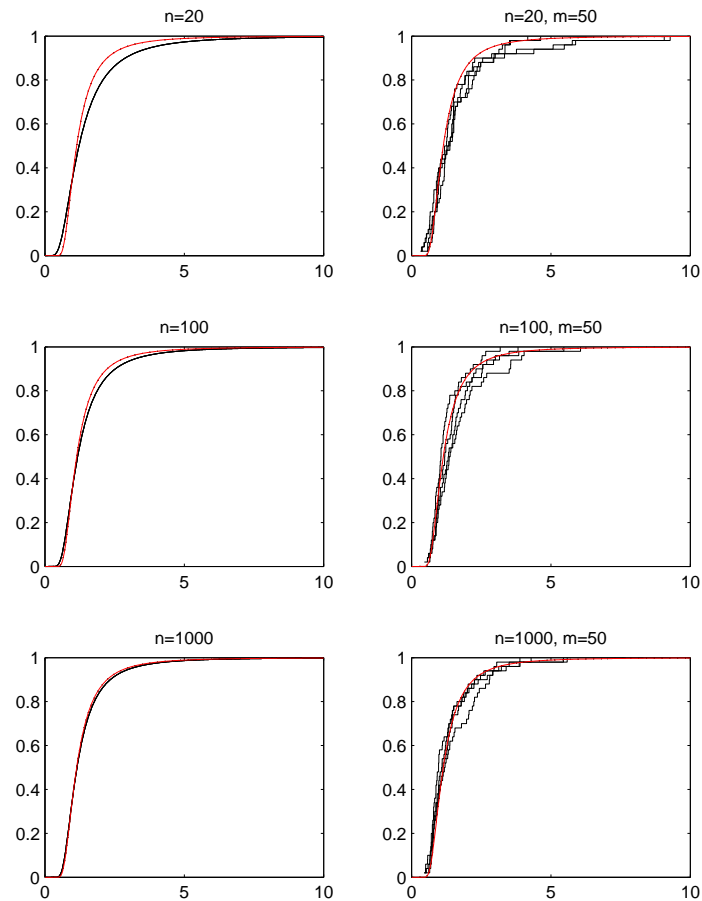
$$H_\beta(x) = 1 - \frac{1}{(1 + \beta x)^{1/\beta}}.$$

Paretofördelningen har den egenskapen att slumpvariabler som överstiger en viss nivå också är Paretofördelade med modifierade parametrar. Den generaliserande Paretofördelningen är just en Paretofördelning och det är likhet i Sats 3 som kommer att råda.

$$\frac{1 - H_\beta(u + v)}{1 - H_\beta(u)} = \frac{1 - F((u + v)c_n + d_n)}{1 - F(uc_n + d_n)}.$$

Mean excess function då  $X$  är Paretofördelad är:

$$e(t) = \frac{\frac{1}{\delta} + t}{\gamma - 1}.$$



Figur 3: Konvergens för maxima, Fréchetfördelningen (röd) närmar sig en Paretofördelning underifrån då  $n$  blir större. Till höger jämförs simulerade variabler från Paretofördelningen med Fréchetfördelningen. Konvergens är tydlig för  $n \geq 100$

## 4.2 Lognormalfördelningen

En slumpvariabel som är lognormalfördelad kan skrivas som  $e^X$  där  $X \sim N(\mu, \sigma^2)$ . En lognormalfördelad variabel har täthetsfunktionen:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$$

och fördelningsfunktionen:

$$F(x) = P(e^X \leq x) = P(X \leq \log x) =$$

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

där nu  $\Phi(\cdot)$  står för fördelningsfunktionen för  $N(0, 1)$ .

Väntevärdet är:

$$E[X] = e^{\mu + \frac{\sigma^2}{2}}.$$

och kvantilen  $q_n$  löses ur:

$$\frac{1}{n} = 1 - F(q_n) \Rightarrow 1 - \frac{1}{n} = \Phi\left(\frac{\log q_n - \mu}{\sigma}\right).$$

Kvantiler,  $z_{1/n}$  utläses ur normalfördelningsstabellen och finns dessutom inbyggda i Matlab.

$$\frac{\log q_n - \mu}{\sigma} = z_{1/n} \Rightarrow q_n = e^{\sigma z_{1/n} + \mu}.$$

#### 4.2.1 Konvergens för maxima av lognormalvariabler

Enligt Sats 2 ska

$$\frac{M_n - q_n}{e(q_n)} \xrightarrow{d} \Delta(x) = e^{-e^{-x}}.$$

Då  $q_n \rightarrow \infty$  kan  $e(q_n)$  enligt [2] approximeras med

$$\frac{\sigma^2 q_n}{\log q_n - \mu}.$$

I figur 4 visas hur väl maxima från lognormalfördelningen konvergerar mot en Gumbelfördelning för några olika värden på  $n$ . Den här lognormalfördelningen verkar konvergera snabbt till sin extremvärdesfördelning.

#### 4.2.2 POT-modellen för lognormalvariabler

Då  $X \sim \text{LogN}(\mu, \sigma^2)$  gäller det enligt Sats 4 att

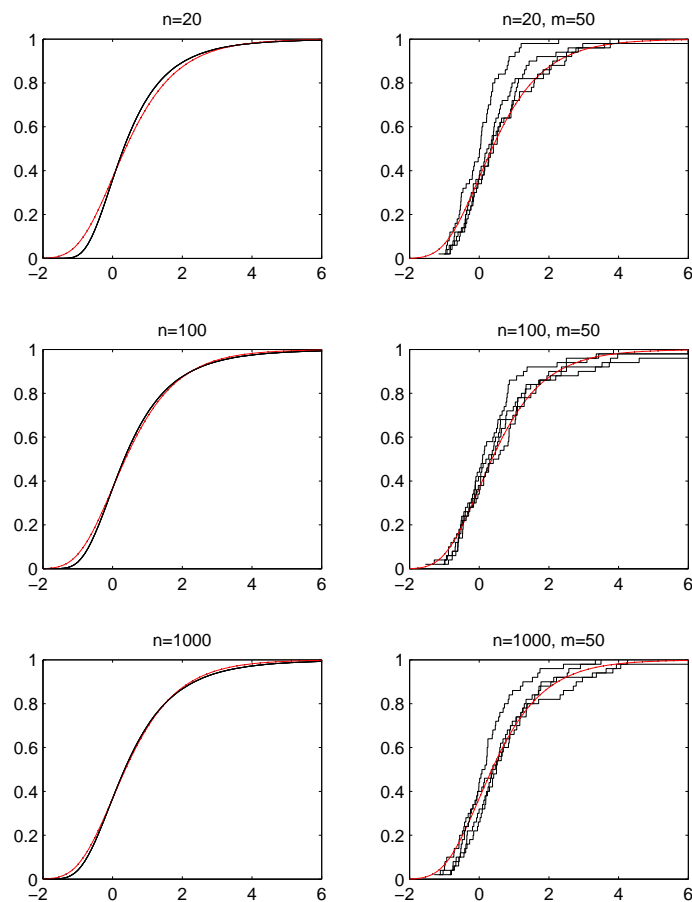
$$\frac{X - t}{e(t)} \xrightarrow{d} H_\beta.$$

Då maxima konvergerar mot en Gumbelfördelning har den generaliserade Paretofördelningen fördelningsfunktion:

$$H_\beta(x) = 1 - e^{-x}.$$

Då  $X$  är Lognormalfördelat och  $t$  är stort approximeras alltså Mean excess function enligt:

$$e(t) \approx \frac{\sigma^2 t}{\log t - \mu}$$



Figur 4: Konvergens för maxima. Gumbelfördelningen (röd) och Lognormalfördelningen till vänster, konvergens tydligt redan då  $n = 20$ , och i princip lika god för  $n = 100$  som  $n = 1000$ . Till höger visas Gumbelfördelningen och de simulerade variabler från lognormalfördelningen.

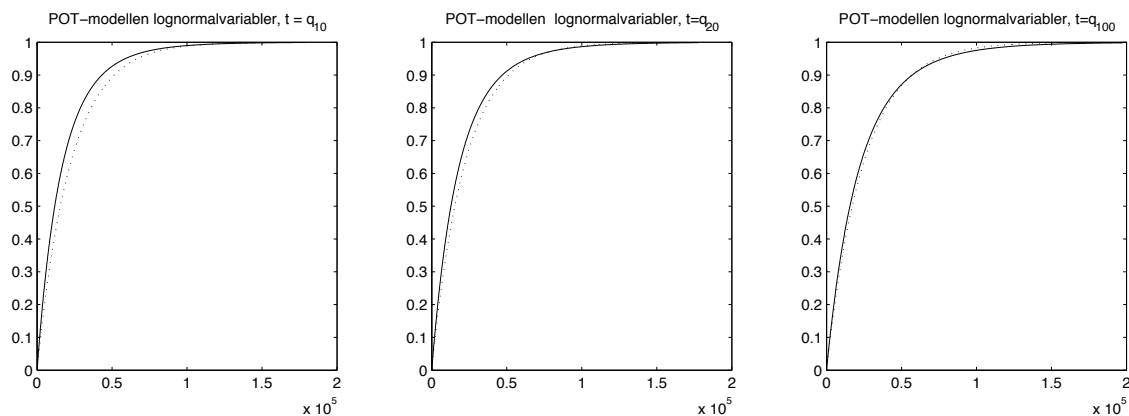
Sats 4 gör det enkelt att illustrera hur väl lognormalfördelningen konvergerar till den generaliserande Paretofördelningen, detta visas genom att jämföra

$$P\left(\frac{X-t}{e(t)} < v \mid \frac{X-t}{e(t)} > 0\right)$$

där  $\frac{X-t}{e(t)} \sim H_\beta$  med

$$P(X < ve(t) + t \mid X > t)$$

där  $X \sim \text{log}N(\mu, \sigma^2)$ . I figur 5 illustreras gränsfördelningen för olika nivåer på  $t$ . Trösklarna  $t$  är valda så att 1, 5 respektive 10 procent av slumpvariablerna bör överstiga  $t$ . Konvergensen är bättre då  $t$  är på den högre nivån men modellen bör fungera väl då  $t \geq q_{20}$ .



Figur 5: Den generaliserande Paretofördelningen är den prickade linjen och lognormal är heldragen. Tre nivåer på  $t$  illustreras.

### 4.3 F-fördelningen

Att maxima från en Paretofördelning konvergerar mot en Fréchetfördelning är kanske inte så överraskande med tanke på fördelningsfunktionens utseende. Därför är det intressant att titta på ytterligare en fördelning, nämligen F-fördelningen som även den är långsvansad och saknar övre begränsning. F-fördelningens maximum bör enligt Sats 2 konvergera mot en Fréchetfördelning. F-fördelningen är egentligen en hel familj av fördelningar då den beror av ett antal frihetsgrader i både täljaren och nämnaren.

Täthetsfunktionen för  $F(n, m)$  är:

$$f(x) = \frac{\Gamma(\frac{n+m}{2}) (\frac{n}{m})^{n/2}}{\Gamma(\frac{n}{2}) \Gamma(\frac{m}{2})} \frac{x^{n/2-1}}{(1 + \frac{nx}{m})^{(n+m)/2}}$$

och dess väntevärde är:

$$E[X] = \frac{m}{m-2}.$$

#### 4.3.1 Konvergens för maxima av F-variabler

Även maxima från F-fördelningen kommer alltså att konvergera mot en Fréchetfördelning. För beräkning av konstanten  $\alpha$  i Fréchetfördelningen är

fördelningsfunktionen för  $F(n, m)$  av intresse. Denna kan erhållas med integrering av täthetsfunktionen.

$$f(x) \sim \frac{1}{x^{m/2+1}} \Rightarrow 1 - F(x) \sim \frac{1}{x^{m/2}}.$$

Konstanten  $\alpha$  i Fréchetfördelningen är därmed lika med  $\frac{m}{2}$ . För att skala om maxima behövs kvantiler från F-fördelningen. Dessa kvantiler finns i tabeller, men det finns även inbyggda i Matlab. I figur 6 jämförs en Fréchetfördelning med F-fördelningen för varierande  $n$ . F-fördelningen konvergerar långsammare än de redan studerade fördelningarna. Anpassningen för  $n = 20$  är sämre här. Konvergensen blir tydligare för större  $n$  med kan med tanke på slumpvariablernas spridning ändå inte anses som tillräcklig god för  $n = 100$ . Dessa slumpvariabler ser inte ut att kunna komma från en Fréchetfördelning.

#### 4.3.2 POT-modellen för F-variabler

Då  $X \sim F(n, m)$  gäller det enligt Sats 4 att

$$\frac{X - t}{t\beta} \xrightarrow{d} H_\beta$$

där den generaliserande Paretofördelningen är

$$H_\beta(x) = 1 - \frac{1}{(1 + \beta x)^{1/\beta}}.$$

Enligt Sats 3 gäller det att

$$\frac{1 - H_\beta(u + v)}{1 - H_\beta(u)} \xrightarrow{d} \frac{1 - F((u + v)c_n + d_n)}{1 - F(uc_n + d_n)}.$$

Mean excess function då  $X$  är F-fördelad erhålls som alltid genom integrering:

$$e(t) = \frac{1}{1 - F(t)} \int_t^\omega (1 - F(x)) dx.$$

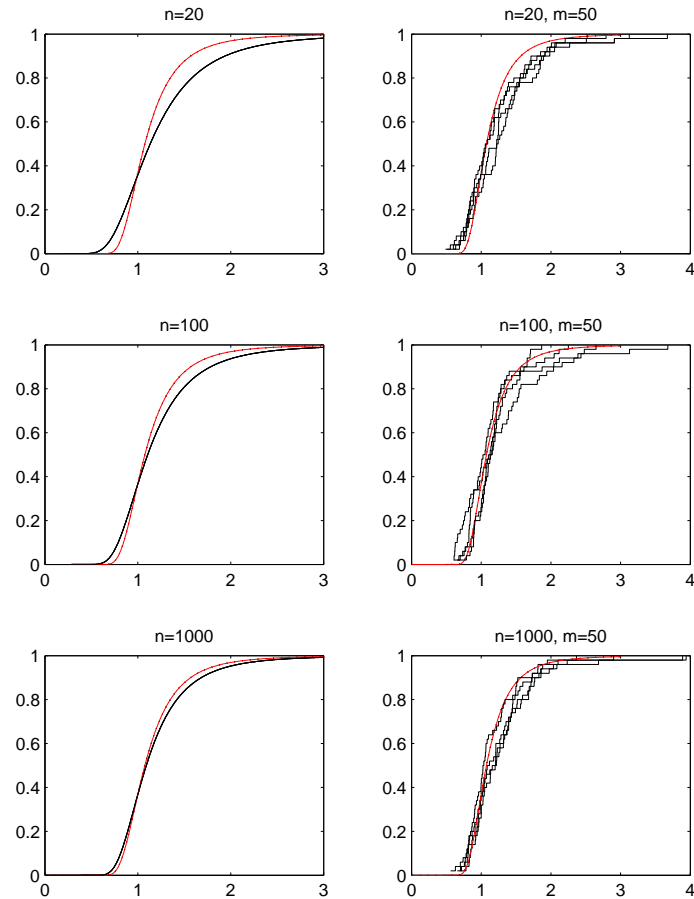
Med hjälp av Sats 4 framgår det att det är

$$P\left(\frac{X - t}{t\beta} < v \mid \frac{X - t}{t\beta} > 0\right)$$

där  $\frac{X-t}{t\beta} \sim H_\beta$  och

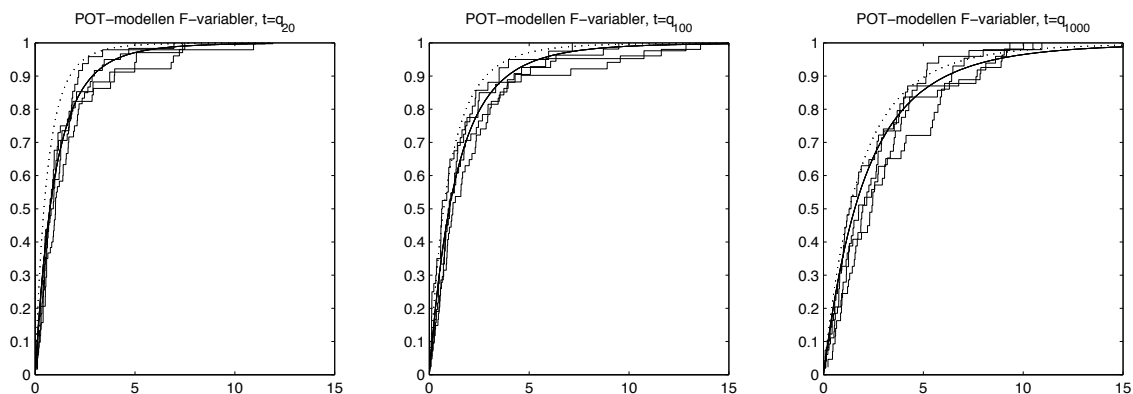
$$P(X < vt\beta + t \mid X > t)$$

där  $X \sim F(n, m)$  som ska studeras för att kunna säga något om konvergens för olika nivåer på  $t$ .  $\beta = 2/m$  där  $m$  är antalet frihetsgrader i nämnaren för



Figur 6: Konvergens för maxima, Fréchetfördelningen (röd) närmar sig långsamt F-fördelningen underifrån. Konvergensen är tydligare för större värden på  $n$ . Fréchetfördelning och de simulerade F-variablerna jämförs till höger.

F-fördelningen. Figur 7 visar att en hög nivå på  $t$  krävs för att konvergens ska äga rum. I figur 7 är  $t$  bestämt enligt att 0.1, 1 respektive 5 procent av slumpvariablerna bör överstiga  $t$ . Slumpvariabler från F-fördelningen har även simulerats för att kunna avgöra om modellen är lämplig för dessa nivåer på  $t$ . I simuleringen väljs antalet slumpvariabler  $n$  så att ungefär 50 stycken överstiger  $t$ . Konvergensen är klart sämre än för lognormalvariabler, här krävs en högre tröskel, då  $t = q_{1000}$  är modellen fortfarande inte tillfredsställande.



Figur 7: Den generaliserande Paretofördelningen är den prickade linjen och F-fördelningen är den heldragna. Tre nivåer på  $t$  illustreras.

#### 4.4 $U(0,1)$

Weibullfördelningen är inte särskilt intressant i detta sammanhang eftersom skadebelopp inte har någon övre begränsning, men ett exempel får illustrera den tredje extremvärdesfördelningen. Den likformiga fördelningen  $U(0,1)$  har täthetsfunktion:

$$f(x) = 1$$

och fördelningsfunktionen:

$$F(x) = x$$

då  $0 < x < 1$ . Väntevärdet är:

$$E[X] = \frac{1}{2}$$

och kvantiler är enkla att beräkna då:

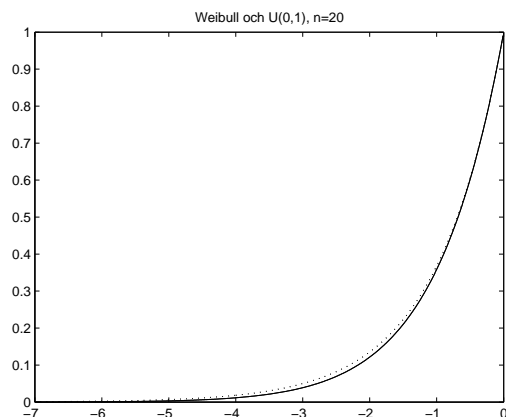
$$q_n = 1 - \frac{1}{n}.$$

Den övre begränsningen  $\omega = 1$  och  $P(X \leq x) = x$ . Låt nu  $x < 0$ .

$$\begin{aligned} P\left(\frac{M_n - \omega}{\omega - q_n} \leq x\right) &= P(n(M_n - 1) \leq x) = \left(M_n \leq 1 + \frac{x}{n}\right) \\ &= P\left(X \leq 1 + \frac{x}{n}\right)^n = P\left(X \leq 1 - \frac{-x}{n}\right)^n \rightarrow e^{-(-x)}. \end{aligned}$$

Detta är alltså en Weibullfördelningen med parametern  $\alpha = 1$ . Weibullfördelningen illustreras i figur 8.





Figur 8: Konvergens för maxima, Weibullfördelningen är den prickade linjen och  $U(0,1)$  den heldragna. Konvergensen är tydlig redan för  $n=20$ .

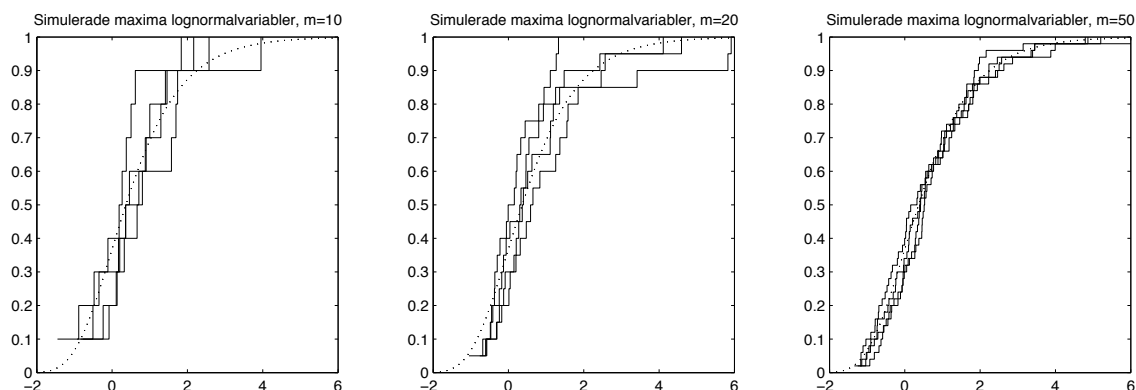
## 5 Tillämpningar

### 5.1 Tillämpning av extremvärdesteorin

I avsnitt 4 illustrerades hur många observationer ett stickprov bör innehålla för att maxima ska konvergera till någon extremvärdesfördelning. I detta avsnitt ska värden från lognormalfördelningen simuleras för att studera konvergens av maxima. Idén är att simulera  $n$  stycken slumpvariabler från  $m$  stycken stickprov som då kan ses som  $n$  stycken skadebelopp från varje tidsperiod  $m$ . Det är alltså  $m$  stycken maximum som kommer att studeras. Syftet med illustrationen är att undersöka om det är rimligt att med hjälp av extremvärdesteorin modellera för hur stort det största skadebeloppet för varje år kan komma att bli de kommande åren. Maxima kommer från stickprov som innehåller 100 stycken slumpvariabler, då det enligt avsnitt 4 är tillräckligt. Detta skadebolag drabbas alltså av 100 stycken utbetalningar per år. I figur 9 visas fördelningen för maxima då värden för 10, 20 och 50 år framåt simuleras. Anpassningen blir bättre desto fler maxima som studeras men redan då det är 50 stycken maxima fungerar modellen tillfredställande.

### 5.2 Tillämpning av POT-modellen, återförsäkring

Återförsäkring betyder att försäkringsbolagen försäkrar sina egna åtaganden hos en större aktör, ett återförsäkringsbolag. Försäkringsbolaget är alltså försäkringstagare hos ett återförsäkringsbolag medan de är försäkringsgivare till sina kunder (försäkringstagarna). Genom att försäkringsgivarna i sin tur



Figur 9: Simulering av maxima från Lognormalfördelningen. Gumbelfördelningen är den prickade linjen.

är försäkrade är den enskilde försäkringstagaren skyddad även om ett enskilt försäkringsbolag drabbas av många och stora skadebelopp, se [5]. Detta skulle kunna hända om exempelvis en översvämning drabbar ett område där majoriteten har tecknat försäkring i samma försäkringsbolag.

Ett exempel på återförsäkringsavtal skulle kunna vara att om en skada överstiger en viss nivå,  $t$ , går återförsäkringsbolaget in och betalar det överstigande beloppet. Med hjälp av POT-modellen kan återförsäkringsavtal prissättas.

Det första antaganden i detta exempel är att skadebeloppen  $X$  är Paretofördelade med parametrar enligt ovan, ( $\delta = \frac{1}{24000}, \gamma = 2.8$ ). Det ger att  $E[X] \approx 13000$ . Det andra antagandet är att försäkringsbolagen har 1000 eller 5000 skadeutbetalningar per år. Det tredje antagandet är att antalet skador  $N$  som överstiger en viss nivå är Poissonfördelade med intensitet  $\lambda$ . Det är fyra olika fall som kommer att studeras över en 10 års period och återförsäkringsbolaget kommer här i exemplet att stå för hela kostnaden som överstiger  $t$ .

1. Då skadebelopp överstiger 46000 kr betalas den överstigande summan av återförsäkringsbolaget.  $n = 10000$  och  $t = 46000$ .
2. Då skadebelopp överstiger 46000 kr betalas den överstigande summan av återförsäkringsbolaget.  $n = 50000$  och  $t = 46000$ .
3. Då skadebelopp överstiger 100000 kr betalas den överstigande summan av återförsäkringsbolaget.  $n = 10000$  och  $t = 100000$ .
4. Då skadebelopp överstiger 100000 kr betalas den överstigande summan av återförsäkringsbolaget.  $n = 50000$  och  $t = 100000$ .

I fall 1 och 2 är  $\lambda = 0.05$  då  $F(46000) = 0.95$  och i fall 3 och 4 är  $\lambda = 0.01$  då  $F(100000) = 0.99$ .

Väntevärdet för återförsäkringsbolagets åtaganden i de fyra fallen beräknas med hjälp av

$$nE[N]e(t)$$

där

$$N \sim Po(\lambda) \Rightarrow E[N] = \lambda$$

och

$$e(t) = \frac{\frac{1}{\delta} + t}{(\gamma - 1)}.$$

För att illustrera utfallet av de fyra fallen har  $n$  stycken slumpvariabler simulerats från Paretofördelningen. Dessa kan ses som skadebeloppen under de kommande åren. För att få en bättre bild av hur väl modellen fungerar i de fyra fallen upprepas försöken fyra gånger. Det kan ses som att fyra försäkringsbolag med samma förutsättningar har tecknat samma typ av återförsäkring.

Resultat av simuleringen. Tabell 1 redogör för den totala kostnaden i tusentals kronor för det fyra återförsäkringsbolagen.

	Förväntad kostnad	ÅFB 1	ÅFB 2	ÅFB 3	ÅFB 4
fall 1	19444	20690	17591	19243	18360
fall 2	97222	97488	94448	92305	95365
fall 3	6889	5784	5554	10396	8240
fall 4	34444	35375	37103	35698	35511

Tabell 2 redogör för antalet skadebelopp som drabbar återförsäkringsbolagen.

	$nE[N]$	ÅFB 1	ÅFB 2	ÅFB 3	ÅFB 4
fall 1	500	533	492	454	518
fall 2	2500	2506	2512	2443	2467
fall 3	100	104	88	110	120
fall 4	500	503	490	484	489

Tabell 3 redogör för det största skadebeloppet i tusentals kronor under tidsperioden

	ÅFB 1	ÅFB 2	ÅFB 3	ÅFB 4
fall 1	434	772	3600	403
fall 2	1660	1637	3688	777
fall 3	638	542	2226	568
fall 4	1259	7508	2399	2387

Priset på återförsäkringen kommer sättas högre än den förväntade kostnaden då återförsäkringsbolaget vill ha betalt för sitt risktagande. I fall 3 skulle det betyda att återförsäkringsbolaget kräver mer än 6888900 kronor under de 10 åren. Ju fler försäkringstagare återförsäkringsbolaget har, desto mindre blir risken och de skulle kunna sätta en lägre premie. Även om återförsäkring bör vara dyrare än förväntad kostnad är det ändå en poäng för ett försäkringsbolag att köpa återförsäkring då risktagandet skjuts över på en annan part. I fall 4 drabbas återförsäkringsbolag nr 2 av större utgifter under tidsperioden än väntat och det finns ett skadebelopp som överstiger 7000000 kronor. Detta är nu återförsäkringsbolagets åtaganden. Ytterligare en motivering till återförsäkring är att det blir ett jämnare flöde av utgifter då de stora skadebeloppen får ett fast pris. I detta exempel står återförsäkringsbolaget för den del av skadebeloppet som överstiger en viss nivå. Andra tänkbara scenarion är att endast en del av den överstigande summan tillfaller återförsäkringsbolaget eller att det finns en övre begränsning även för återförsäkringsbolaget, detta är mycket vanligt. Det är även troligt att det är fler återförsäkringsbolag som är med och delar på de stora skadorna. För att beräkna den förväntade kostnaderna i ett intervall, då återförsäkringsbolaget står för kostnaden som överstiger  $t_1$  upp till den övre gränsen  $t_2$ , subtraheras  $e(t_2)$  från  $e(t_1)$ .

Av simuleringen framgår att det är ett stort  $n$  som gör att kostnaden kommer nära sitt väntevärde. Om  $t$  är på en så hög nivå att det är få observationer som överstiger blir det osäkrare. Även om  $n$  är litet men  $t$  är på en rimlig nivå blir det osäkert. Den generaliserande Paretofördelningen fungerar för alla värden på  $t$  men  $t$  bör ändå vara stort så att sannolikheten att överstiga  $t$  blir liten. Då kan Poissonfördelningen användas för att modellera hur många skador som överstiger  $t$ .

## 6 Slutsatser

Extremvärdesteorin är tillämpbar för de fyra fördelningar som har studerats. Maximum med lämplig omskalning konvergerar väl mot någon av extremvärdesfördelningarna då  $n$  närmar sig 100 för tre av de fyra studerade fördelningar. F-fördelningen är den som konvergerar långsammast. För att kunna säga något om när konvergensen är god simulerades även variabler från fördelningarna. Då kunde spridningen av variablerna tas med i resonemanget. Då det inte går att avgöra från vilken fördelning slumpvariablerna kommer ifrån kan konvergensen ses som god. Konvergensen studeras enbart genom att titta på figurer, det är inget exakt mått. För att kunna använda extremvärdesteorin är Sats 2 nödvändig. Utan att känna till normeringskonstanterna eller till vilken extremvärdesfördelning som konvergens kommer äga rum blir inte extremvärdesteorin användbar. I avsnitt 5.1 illustreras i ett exempel att det är flera maxima som bör studeras för att teorin ska vara

användbar. Även ett enskilt maxima följer fördelningen men det är mycket osäkert att modellera för ett enstaka maxima. I exemplet framgår att då 50 maxima studeras fungerar modellen tillfredställande.

POT-modellen fungerar bäst då slumpvariabler kommer från en Paretofördelning. Den generaliserande Paretofördelningen är just en Paretofördelning och i Sats 3 och 4 är det likhet och inte konvergens som råder. Paretofördelningen har den egenskapen att givet en basnivå  $t$  erhålls enkelt fördelningen för andra nivåer då det också är en Paretofördelning med modifierade parametrar. POT-modellen fungerar även för lognormalvariabler och F-variabler då  $t$  är på en hög nivå. I de två fallen bör tröskeln  $t$  vara på en hög nivå av två anledningar, dels för att Sats 4 kräver det och dels för att då  $t$  är på en hög nivå blir sannolikheten att överstiga  $t$  liten och skador som överstiger  $t$  kan då modelleras med en Poissonfördelning. För Paretofördelningen är det bara det sistnämnda som gör att  $t$  bör vara på en hög nivå. Figur 5 och 7 säger att POT-modellen för lognormalvariabler fungerar för en lägre nivå på  $t$  än vad POT-modellen för F-variabler gör. Det beror på att F-fördelningen inte ser ut som en Paretofördelning förrän långt ute i svansen. Beräkningarna av normeringskonstanterna i Sats 4 var viktiga för att förstå hur den generaliserande Paretofördelningen och därmed POT-modellen kan användas. Om inte beräkningen av dessa hade gjorts skulle det inte vara möjligt att anpassa en generaliserad Paretofördelning. Sats 3 säger enbart att det finns konstanter så att konvergens äger rum, men den är alltså inte tillämpbar.

I avsnitt 5.2 illustreras POT-modellen i ett exempel. Där visas att spridningen är relativt stor. Då många observationer hamnar över  $t$  blir spridningen mindre. Ett lägre  $t$  ger fler värden som överstiger  $t$  men då  $t$  ska vara stort behöver  $n$  vara så stort att flera variabler överstiger  $t$ . I exemplet har fall 3 stor spridning mellan återförsäkringsbolagen trots att det är runt 100 observationer som överstiger  $t$ . I detta perspektiv framgår det att modellen kräver stora tal för bättre säkerhet.

## 7 Diskussion

För att tillämpa extremvärdesteorin och POT-modellen i försäkringsmanhang måste anpassningen av skadebelopp till en känd fördelning vara god. Om det finns knapphändigt med data över skadebelopp för en viss försäkringstyp blir det större osäkerhet i modellen för dessa skadebelopp. Detta medför även större osäkerhet vid tillämpning av extremvärdesteorin och POT-modellen. En annan aspekt att ta hänsyn till vid anpassning av skadebelopp är att tiderna förändras. Exempel på det skulle kunna vara växthuseffekten. Växthuseffekten verkar påverka naturfenomen, gör den att vi drabbas av fler naturkatastrofer i framtiden? Ett annat exempel är villain-

brott, gör nya och säkrare larm att villainbrotten minskar? Sådana frågor försvårar anpassning av skadebelopp även om tillräckligt med data finns att tillgå.

Det hade varit intressant att studera modellerna med riktigt data, men simulering var nödvändigt då det är svårt att få tillräckligt med data för att studera gränsfördelningar. Extra svårt blir det när det handlar om extrema händelser.

Det är intressant att reflektera över att det krävs en djup förståelse för ämnet då flera beräkningar av normeringskonstanter krävs för tillämpning. Det som i teorin ser behändigt ut kräver alltså en del omformuleringar. En annan komplikation för att kunna tillämpa modellerna är att det krävs enorma mängder med data för att få någon säkerhet i modellerna. För att tillämpa POT-modellen måste det även tas hänsyn till en rimlig nivå på  $t$ . I simuleringen kan önskat antal variabler erhållas men det går att diskutera dess rimlighet.

## 8 Referenser

### Referenser

- [1] COLES. S (2001). An introduction to statistical modeling of extreme values, Springer.
- [2] EMBRECHTS. P, KLÜPPELBERG. C OCH MIKOSCH. T (1997). Modelling extremal events, Springer.
- [3] FALK. M, HÜSLER. J OCH REISS. R-D (2004). Laws of small numbers: Extremes and rare events, Birkhäuser.
- [4] OLSSON. E (2008). Matematiska metoder inom sakförsäkring, Stockholms universitet.
- [5] <http://partner.insplanet.com/ordlista/aterforsakring.asp> (2009-05-05)

## 9 Bilagor

Matlab kod

```
% m-fil 1 pareto extremvärde  
gamma=2.8;  
alpha=1/24000;
```

```

m=100;
n=100;
%Simulering från Pareto(alpha,gamma)
S = ((1-rand(n, m)).^(-1/gamma)-1)./alpha;
%cdf:  $F(x) = 1 - (1 + \alpha x)^{-\gamma}$ 
M=sort(max(S));
x=[0.01:0.01:10];
%kvantilen q
q=(1/alpha)*n^(1/gamma)-(1/alpha);
%Fördelningsfunktion  $F^n(qx)$ 
Fpareto=(1-(1./(1+alpha*q*x)).^gamma).^n;
%Simulerad fördelningsfunktion
EmpF=[1:m]/m;
%Fördelningsfunktion Frechet
Ffrechet=exp(-(x).^(-gamma));

% m-fil 2 Lognormal extremvärde
mu=9.8236;
sigma2=0.4407;
sigma=sqrt(sigma2);
m=100;
n=100;
% Simulerade värden från logn(mu,sigma2)
L=exp(mu+sigma.*randn(n,m));
M=sort(max(L));
%kvantilen q = konstanten d
q=exp(sigma*norminv(1-(1/n))+mu);
%aux = konstanten d
aux=sigma2*q/(log(q)-mu);
x= [-2:0.01:5];
%x=[0:100:100000];
%Simulerad fördelningsfunktion
EmpF=[1:m]/m;
T=(M-q)./aux;
%Fördelningsfunktion Gumbel
G=exp(-exp(-x));
%Fördelningsfunktionen  $F^n(x*aux+q)$ 
Flogn=normcdf((log(x*aux+q)-mu)/sigma).^n;

% m-fil 3 F-fördelningen extremvärde
n=10; %här antal frihetsgrader
m=10; %här antal frihetsgrader
j=100; % "m" antal maxima
i=100; % "n" antal i varje stickprov

```

```

% i*j stycken simulerade F(n,m)-variabler
Fnm=[];
for t=1:i
Xn=sum((randn(n,j)).^2);
Xm=sum((randn(m,j)).^2);
Fnm(t,1:j) =(Xn/n)./(Xm/m);
end
F=sort(max(Fnm));
%kvantilen q = konstanten c
q=finv(1-1/i,n,m);
%Simulerad fördelning
EmpF=[1:j]/j;
%Frechetfördelningens fördelningsfunktion
x=[0 0.01:0.01:3];
alpha=m/2;
Ffrechet=exp(-x.^(-alpha));
%Fördelningen F^n(cx+d)
FF=(fcdf(q*x,n,m)).^i;

% m-fil 4 Likformigafördelningen Extremvärde
n=100;
m=100;
%Simulerade variabler från U(0,1)
UU=rand(n,m);
U=sort(max(UU));
%övre gräns w = konstanten d
w=1;
%kvantilen q , konstanten c = w - q
q=1-(1/n);
%simulerad fördelningsfunktion
EmpF=[1:m]/m;
T=(U-w)./(w-q);
%Wiebullfördelningens fördelningsfunktion
x=-[0.01:0.01:7];
alpha=1;
W=exp(-(-x).^alpha);
% F^n((w-q)x+w)^n
Funi=(x*(w-q)+w).^n;

% m-fil 5 Pareto pot-modell
phi=1/24000;
gamma=2.8;
alpha=2.8;
beta=1/alpha;

```



```

u=100000;
v=[0:500000];
genpar=((1+beta*u)./(1+beta*u+beta*v)).^(1/beta);
n=1000;
q=(n^(1/gamma)-1)/phi;
pareto= ((1-phi*q+phi*u*q/alpha)./(1-phi*q+phi*q*(u+v)/alpha)).^gamma;

% m-fil 6 lognormal pot-modell
mu=9.8236;
sigma2=0.4407;
sigma=sqrt(sigma2);
n=10000;
u=55000;
L=exp(mu+sigma.*randn(n,1));
V=sort(L(L>u));
v=length(V);
q=u;
eq=sigma2*q/(log(q)-mu);
EmpF=[1:v]/v;
x=[0:0.01:8];
y=x*eq;
genpar=1-exp(-x);
t1=u+y;
t2=u;
F=1-(1-normcdf((log(t1)-mu)./sigma))./(1-normcdf((log(t2)-mu)./sigma));

% m-fil 7 F-pot-modell
t=8.75; %F(2.98)=0.95
n=223; % antal
v=[0:0.1:20];
i=10; %frihetsgrader
j=10; %frihetsgrader
q=finv(1-1/n,i,j); %kvantil
beta=2/j;
alpha=1/beta;
genpar=1-(1./(1+beta*v)).^(1/beta);
F=(1-fcdf(v*t*beta+t,i,j))./(1-fcdf(t,i,j));
V=frnd(i,j,n); %ger n*n värden
V1=sort(V(V>t));
x=length(V1);
EmpF=[1:x]/x;

```