ADVANCED METHODS OF NON-LIFE INSURANCE

Aggregate Loss Reserving

Prof. dr. Tim Verdonck

Tim.Verdonck@kuleuven.be

2018 - 2019

ALMA MATER STUDIORUM A.D. 1088
UNIVERSITÀ DI BOLOGNA

Outline

Collective Risk Modeling

Aim: describe distribution of total claim amount S that insurance system (e.g. entire company, a line of business, those covered by group insurance contract, single policy) faces within fixed time period.

- For time period we take one (accounting) year.
- Assume that N **counts all claims** that occur within fixed year.
- Total claim amount S is then

$$S = Y_1 + \ldots + Y_N = \sum_{i=1}^{N} Y_i$$

where $Y_1, \ldots, Y_N$ models individual claim sizes.

$\Rightarrow$ collective risk model because we consider whole portfolio as a collective.
$\Rightarrow$ model is compounded by fact that **number of terms** in sum is **random**.
Hope is to discover a law of large numbers for total insurance portfolio claim so that insurance company can benefit from diversification benefits (between individual risks) that allow to predict possible outcomes of S more accurately.

Collective Risk Modeling

- We investigate various models for risk consisting of total or aggregate amount of claims S payable by company over fixed period of time.
- Such models will inform us and allow us to make decision on amongst others, expected profits, premium loadings, reserves necessary to ensure profitability, as well as impact of reinsurance and deductibles.
- Restricting consideration to shorter periods of time like a few months or a year often allows us to ignore aspects of changing value of money due to inflation.
- In collective risk model multiple claims may result from single policy or policyholder.

Compound distributions

Model assumptions - Compound distribution

The total claim amount S is given by the following compound distribution

$$S = Y_1 + \ldots + Y_N = \sum_{i=1}^{N} Y_i$$

with 3 standard assumptions

1. N is discrete rv which only takes values in $\mathcal{A} \subset \mathbb{N}_0$
2. $Y_1, Y_2, \ldots \overset{iid}{\sim} G$ with $G(0) = 0$.
3. N and $(Y_1, Y_2, \ldots)$ are independent.

If S satisfies these assumptions, then S has compound distribution.
Distribution of S is obtained from distribution of N and distribution of the Y_i's.
Using this approach, frequency and severity of claims are modeled separately.

Compound distributions

Remarks

1. First assumption says that **number of claims** N takes **only non-negative** integer values. The event $\{N = 0\}$ means that no claim occurs which provides a total claim amount of $S = 0$.
2. Second assumption means that **individual claims** Y_i **does not affect each other**, for instance, if we face a large first claim Y_1 this does not give any information for the remaining claims Y_i $(i \geq 2)$. Moreover, we have homogeneity in the sense that all claims have the **same marginal distribution function** G with $0 = G(0) = \mathbb{P}[Y_1 \leq 0]$ i.e. the individual claim sizes Y_i are **strictly positive**, $\mathbb{P}$-a.s. We use synonymously the terminology (individual) claim size, (individual) claim and claims severity for Y_i.
3. Last assumption says that the **individual claim sizes are not affected by the number of claims** and vice versa. For instance, if we observe many claims this does not contain any information whether these claims are of smaller or larger size.

Compound distributions

Compound distribution is base model for collective risk modeling and we are going to describe **different choices for the claims count distribution** of N and for the **individual claim size distribution** of Y_i.
Using both models, we will obtain the characteristics of the distribution of S.

Basic recognition features of compound distributions

Assume S has a **compound distribution**. We have (whenever they exist)

$$
\begin{aligned}
E[S] &= E[N]E[Y_1] \\
\mathrm{Var}(S) &= \mathrm{Var}(N)E[Y_1]^2 + E[N]\,\mathrm{Var}(Y_1) \\
\mathrm{Vco}(S) &= \sqrt{\mathrm{Vco}(N)^2 + \frac{1}{E[N]}\,\mathrm{Vco}(Y_1)^2} \\
M_S(r) &= M_N\left(\log(M_{Y_1(r)})\right) \qquad \text{for } r \in \mathbb{R}.
\end{aligned}
$$

Proof: *Actuarial Mathematics* (Prof. Mulinacci) or lecture notes (Prof. Wüthrich).

Proof

$$
\begin{aligned}
E[S] &= E\left[\sum_{i=1}^{N} Y_i\right] = E\left[E\left[\sum_{i=1}^{N} Y_i \middle| N\right]\right] = E\left[\sum_{i=1}^{N} E[Y_i|N]\right] \\
&= E\left[\sum_{i=1}^{N} E[Y_i]\right] = E[NE[Y_1]] = E[N]E[Y_1]. \\
\mathrm{Var}(S) &= \mathrm{Var}\left(\sum_{i=1}^{N} Y_i\right) = \mathrm{Var}\left(E\left[\sum_{i=1}^{N} Y_i \middle| N\right]\right) + E\left[\mathrm{Var}\left(\sum_{i=1}^{N} Y_i \middle| N\right)\right] \\
&= \mathrm{Var}\left(\sum_{i=1}^{N} E\left[Y_i|N\right]\right) + E\left[\sum_{i=1}^{N} \mathrm{Var}\left(Y_i|N\right)\right] \\
&= \mathrm{Var}\left(\sum_{i=1}^{N} E\left[Y_i\right]\right) + E\left[\sum_{i=1}^{N} \mathrm{Var}\left(Y_i\right)\right] \\
&= \mathrm{Var}(N)E[Y_1]^2 + E[N]\,\mathrm{Var}(Y_1)
\end{aligned}
$$

Proof

$$
\begin{aligned}
M_S(r) &= E\left[e^{rS}\right] = E\left[E\left[e^{rS}|N\right]\right] \\
&= E\left[E\left[e^{r\sum_{i=1}^{N} Y_i}|N\right]\right] = E\left[E\left[\prod_{i=1}^{N} e^{rY_i}\middle|N\right]\right] \\
&= E\left[\prod_{i=1}^{N} E\left[e^{rY_i}|N\right]\right] = E\left[\prod_{i=1}^{N} E\left[e^{rY_i}\right]\right] \\
&= E\left[\prod_{i=1}^{N} M_{Y_i}(r)\right] = E\left[M_{Y_1}(r)^N\right] \\
&= E[e^{\log\left(M_{Y_1}(r)^N\right)}] = E[e^{N\log(M_{Y_1}(r))}] \\
&= M_N(\log[M_{Y_1}(r)]).
\end{aligned}
$$

$\square$

Distribution of S

If assumptions above hold, then distribution function of S can be written as

$$\begin{aligned} F_S(x) &= \mathbb{P}[S \leq x] = \sum_{k \in \mathcal{A}} \mathbb{P}\left[\sum_{i=1}^{N} Y_i \leq x \middle| N = k\right] \mathbb{P}[N = k] \\ &= \sum_{k \in \mathcal{A}} \mathbb{P}\left[\sum_{i=1}^{k} Y_i \leq x\right] \mathbb{P}[N = k] = \sum_{k \in \mathcal{A}} G^{*k}(x)\mathbb{P}[N = k] \end{aligned}$$

G^{*k} denotes the k-th convolution of the distribution function G.
In particular, we have for $Y_1, Y_2 \overset{iid}{\sim} G$

$$\begin{aligned} G^{*2}(x) &= \mathbb{P}[Y_1 + Y_2 \leq x] = \int G(x-y)dG(y) \\ G^{*k}(x) &= \int G^{*(k-1)}(x)dG(y) \end{aligned}$$

We obtain a **closed form solution** for the distribution function of S.

Introduction

- Traditional loss distribution approach to modeling aggregate losses starts by **separately fitting** a frequency (or claims count) distribution to the number of losses and a severity (or claims size) distribution to the size of the losses.
- We first present some distributions that are often used as counting or **frequency** distributions to describe the number of events such as number of accidents to the driver or number of claims to the insurer.
- Next we discuss some popular distributions for describing the claim size of the losses or the **severity**.
- **Aggregate loss distribution** are then estimated by combining the loss frequency and loss severity distribution by **convolution**.
- In general, convolution formulas are not useful due to the **computational complexity**.
- Later we present other solutions for the calculation of the distribution function of S (e.g. using **simulations**, **approximations** and **smart analytic techniques** under additional model assumptions).

Outline

Importance of frequency

- Insurers pay claims in monetary units. Why should they care how much frequently claims occur?
- As mentioned before, one typically starts with an **expected cost** and then adds **margins** for pricing an insurance good.
- We can think of expected cost as the expected number of claims times the expected amount per claims, that is, **expected frequency times expected severity**.
- Other ways that **frequency augments severity information**:
  - **Contractual**: e.g. deductibles and policy limits are often in terms of each occurrence of an insured event.
  - **Behaviorial**: explanatory variables can have different effects on models of how often an event occurs in contrast to the size of the event.
  - **Databases**: most insurers keep separate data files that suggest developing separate frequency and severity models.
  - **Regulatory and Administrative**: regulators routinely require reporting of claims numbers as well as amounts.

Important frequency distributions

Three important **frequency distributions** in insurance

- Binomial
- Poisson
- Negative Binomial

They are important since

- They fit well many insurance data sets of interest
- They provide **basis for more complex distributions** that even better approximate real situations of interest to us.

Explicit claims count distributions

We give **explicit distribution functions** for the **number of claims** N modeling.

- In non-life insurance context, N should always be understood in relation to underlying (deterministic) volume $v > 0$.
- Therefore we consistently use a volume measure to describe N.
- Typical volume measures: number of insured persons, number of policies, number of risks.
- We interpret v as number of risks injured and N counts number of claims.
- Ratio N/v is called claims frequency and expected number of claims is given by $E[N] = \lambda v$, where $\lambda > 0$ denotes expected claims frequency.
- Under these assumptions we would like to describe the **probability weights**

$$p_k = \mathbb{P}[N = k] \qquad \text{for } k \in \mathcal{A} \subset \mathbb{N}_0$$

Binomial distribution

We choose fixed volume $v \in \mathbb{N}$ and fixed default probability $p \in (0,1)$

Binomial distribution

We say N has a binomial distribution $N \sim \text{Binom}(v,p)$ if

$$p_k = \mathbb{P}(N = k) = \binom{v}{k} p^k (1-p)^{v-k} \qquad \forall k \in \{0, \ldots, v\} = \mathcal{A}$$

Proposition

Assume $N \sim \text{Binom}(v,p)$ for fixed $v \in \mathbb{N}$ and $p \in (0,1)$

$$\begin{aligned} E[N] &= vp \\ \text{Var}(N) &= vp(1-p) \\ \text{Vco}(N) &= \sqrt{\frac{1-p}{vp}} \\ M_N(r) &= (pe^r + (1-p))^v \qquad \forall r \in \mathbb{R} \end{aligned}$$

Binomial distribution

Proof

We calculate mgf and then first two moments.

$$\begin{aligned} M_N(r) &= \sum_{k\in\mathcal{A}} e^{rk}\binom{v}{k}p^k(1-p)^{v-k} = \sum_{k\in\mathcal{A}}\binom{v}{k}(pe^r)^k(1-p)^{v-k} \\ &= (pe^r+(1-p))^v \sum_{k\in\mathcal{A}}\binom{v}{k}\left(\frac{pe^r}{pe^r+(1-p)}\right)^k\left(\frac{1-p}{pe^r+(1-p)}\right)^{v-k} \end{aligned}$$

Last sum is again summation over p_k^* of binomial distribution with default probability $p^* = (pe^r)/(pe^r+(1-p)) \in (0,1)$. Therefore it adds up to 1 which completes the proof. Furthermore,

$$\begin{aligned} E[N] &= v(pe^r+(1-p))^{v-1}(pe^r)|_{r=0} = vp \\ E[N^2] &= [v(v-1)(pe^r+(1-p))^{v-2}(pe^r)^2 + v(pe^r+(1-p))^{v-1}(pe^r)]|_{r=0} \\ &= v(v-1)p^2+vp = vp((v-1)p+1) \\ \mathrm{Var}(N) &= vp((v-1)p+1)-(vp)^2 = vp(vp-p+1-vp) \end{aligned}$$

Corollary: Second characterisation of binomial distribution

Assume that $N \sim \text{Binom}(v, p)$ with given $v \in \mathbb{N}$ and $p \in (0, 1)$.

Choose $X_1, \ldots, X_v \overset{iid}{\sim} \text{Bernouilli}(p)$. Then we have

$$N \overset{(d)}{=} \sum_{i=1}^{v} X_i$$

Proof: Let $X = \sum_{i=1}^{v} X_i$, then

$$M_X(r) \overset{iid}{=} \prod_{i=1}^{v} M_{X_i}(r) = \prod_{i=1}^{v} E\left[e^{rX_i}\right] = \prod_{i=1}^{v} (pe^r + (1-p)) = M_N(r)$$

$\square$

Binomial distribution describes physical situation in which v risks are each subject to claim or loss, hence N describes number of defaults within portfolio of fixed size $v \in \mathbb{N}$. Every risk in portfolio has same default probability p and defaults between different risks do not influence each other (are independent). Thus, if N has binomial distribution then every risk in such portfolio can **at most default once**.

Poisson distribution

We choose fixed volume $v > 0$ and fixed expected claims frequency $\lambda > 0$

Poisson distribution

N has Poisson distribution $N \sim Poi(\lambda v)$, if

$$p_k = \mathbb{P}[N = k] = e^{-\lambda v} \frac{(\lambda v)^k}{k!} \qquad \forall k \in \mathcal{A} = \mathbb{N}$$

Proposition

Assume $N \sim Poi(\lambda v)$ for fixed $\lambda, v > 0$. Then

$$\begin{aligned} E[N] &= \lambda v \\ \mathrm{Var}(N) &= \lambda v \\ \mathrm{Vco}(N) &= \sqrt{\frac{1}{\lambda v}} \\ M_N(r) &= e^{\lambda v (e^r - 1)} \qquad \forall r \in \mathbb{R} \end{aligned}$$

Poisson distribution

Proof

Note that $p_k > 0$ is indeed well-defined

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda v} \sum_{k=0}^{\infty} \frac{(\lambda v)^k}{k!} = e^{-\lambda v} e^{\lambda v} = 1.$$

Calculating expectation directly

$$E[N] = \sum_{k=0}^{\infty} k \frac{(\lambda v)^k}{k!} e^{-(\lambda v)} = \sum_{k=1}^{\infty} k \frac{(\lambda v)^k}{k!} e^{-(\lambda v)} = (\lambda v) \sum_{k-1=0}^{\infty} \frac{(\lambda v)^{k-1}}{(k-1)!} e^{-(\lambda v)} = \lambda v$$

$\Rightarrow$ interpretation of λ.

$$E\left[\frac{N}{v}\right] = \lambda$$

Poisson distribution

Calculating variance directly

$$\begin{aligned}
E[X^2] &= \sum_{k=0}^{\infty} k^2 \frac{(\lambda v)^k}{k!} e^{-(\lambda v)} = (\lambda v) e^{-(\lambda v)} \sum_{k=1}^{\infty} \frac{k}{(k-1)!} (\lambda v)^{k-1} \\
&= \lambda v e^{-(\lambda v)} \left(\sum_{k=1}^{\infty} \frac{k-1}{(k-1)!} (\lambda v)^{k-1} + \sum_{k=1}^{\infty} \frac{1}{(k-1)!} (\lambda v)^{k-1} \right) \\
&= \lambda v e^{-(\lambda v)} \left(\lambda v \sum_{k=2}^{\infty} \frac{1}{(k-2)!} (\lambda v)^{k-2} + \sum_{k=1}^{\infty} \frac{1}{(k-1)!} (\lambda v)^{k-1} \right) \\
&= \lambda v e^{-(\lambda v)} \left(\lambda v \sum_{i=0}^{\infty} \frac{1}{i!} (\lambda v)^{i} + \sum_{i=0}^{\infty} \frac{1}{i!} (\lambda v)^{i} \right) \\
&= \lambda v e^{-(\lambda v)} \left(\lambda v e^{(\lambda v)} + e^{(\lambda v)} \right) = (\lambda v)^2 + \lambda v \\
\text{Var}[X] &= E[X^2] - E[X]^2 = \lambda v
\end{aligned}$$

Poisson distribution

Using moment generating function

$$M_N(r) = \sum_{k\geq 0} e^{rk} e^{-\lambda v} \frac{(\lambda v)^k}{k!} = e^{-\lambda v} \sum_{k\geq 0} \frac{(\lambda v e^r)^k}{k!} = e^{-\lambda v + \lambda v e^r} = e^{\lambda v (e^r - 1)}$$

- $M_N'(r) = (\lambda v) e^{(\lambda v)(e^r-1)} e^r$
  $\Rightarrow E(X) = \lambda v.$
- $M_N''(r) = (\lambda v) e^r e^{(\lambda v)(e^r-1)} + (\lambda v)^2 e^{2r} e^{(\lambda v)(e^r-1)}$
  $\Rightarrow E(X^2) = (\lambda v)^2 + \lambda v.$
- $\text{Var}(X) = E(X^2) - (E(X))^2 = \lambda v.$

Mixed Poisson distribution

We have following properties

- binomial distribution $E[N] > \text{Var}(N)$
- Poisson distribution $E[N] = \text{Var}(N)$

whereas insurance data often suggests $E[N] < \text{Var}(N)$!

Mixed Poisson distribution

- Assume $\Lambda \sim H$ with $H(0) = 0, E[\Lambda] = \lambda$ and $\text{Var}(\Lambda) > 0$.
- Conditionally, given Λ, $N \sim Poi(\Lambda v)$ for fixed volume $v > 0$.

If N satisfies this definition, then we have $E[N] < \text{Var}(N)$.

Proof

Tower property implies

- $E[N] = E[E[N|\Lambda]] = E[\Lambda v] = \lambda v$
- $\text{Var}(N) = E[\text{Var}(N|\Lambda)] + \text{Var}(E[N|\Lambda]) = vE[\Lambda] + v^2\,\text{Var}(\Lambda) > \lambda v$

Gamma distribution

Before introducing negative-binomial distribution, we define gamma distribution.

Gamma distribution

$X \sim \Gamma(\gamma, c)$ with shape parameter $\gamma > 0$ and scale parameter $c > 0$ if X is non-negative, absolutely continuous rv with density

$$f(x) = \frac{c^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-cx}$$

with Gamma function $\Gamma(.)$ is defined as

$$\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x} dx \qquad (\gamma > 0)$$

- $\Gamma(\gamma+1) = \gamma\Gamma(\gamma)$
- $\Gamma(1) = \Gamma(2) = 1 \qquad \Gamma(\frac{1}{2}) = \sqrt(\pi)$
- $\Gamma(n) = (n-1)!$

Negative-binomial distribution

Negative-binomial distribution

$X \sim \text{NegBin}(\lambda v, \gamma)$ with volume $v > 0$, expected claims frequency $\lambda > 0$ and dispersion parameter $\gamma > 0$ if

- $\Theta \sim \Gamma(\gamma, \gamma)$
- Conditionally, given Θ, $N \sim Poi(\Theta\lambda v)$

Note that for gamma distributed latent variable $\Lambda = \Theta\lambda$ with first two moments given by $E[\Lambda] = \lambda$ and $\text{Var}(\Lambda) = \lambda^2/\gamma > 0$, we **fulfil mixed Poisson distribution**.

2nd definition (Negative-binomial distribution)

Negative-binomial distribution satisfies for $k \in \mathcal{A} = \mathbb{N}_0$

$$p_k = \mathbb{P}[N = k] = \binom{k+\gamma-1}{k}(1-p)^\gamma p^k$$

where we choose $p = (\lambda v)/(\gamma + \lambda v) \in (0, 1)$

Negative-binomial distribution

Proposition

Assume $N \sim \text{NegBin}(\lambda v, \gamma)$ for fixed $\lambda, v, \gamma > 0$. Then

$$\begin{aligned}
E[N] &= \lambda v \\
\text{Var}(N) &= \lambda v(1 + \lambda v/\gamma) > \lambda v \\
\text{Vco}(N) &= \sqrt{\frac{1}{\lambda v}}\sqrt{1 + \lambda v/\gamma} > \gamma^{-1/2} > 0 \\
M_N(r) &= \left(\frac{1-p}{1-pe^r}\right)^{\gamma} \qquad \forall r < -\log(p)
\end{aligned}$$

where $p = (\lambda v)/(\gamma + \lambda v) \in (0, 1)$.

Proof: Tower property implies

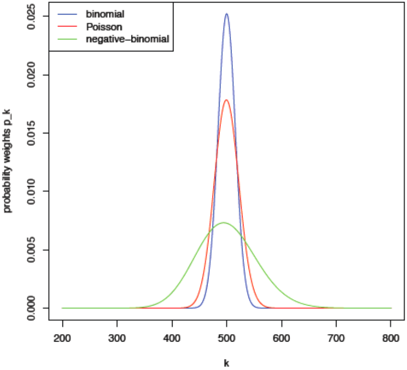
$$\begin{aligned}
M_N(r) &= E[E[e^{rN}|\Theta]] = E[e^{\Theta\lambda v(e^r - 1)}] \\
&= M_\Theta(\lambda v(e^r - 1))
\end{aligned}$$

for which the claims follows for $\Theta \sim \Gamma(\gamma, \gamma)$.

Example

We compare binomial, Poisson and negative-binomial distributions.

- Identical means $E[N] = 500$ with $v = 1000, p = \lambda = 0.5$ and $\gamma = 100$.
- We plot corresponding probability weights p_k



Coefficient of variation is **increasing** from binomial over Poisson to negative-binomial which give successively more **uncertainty** to claims counts.

$(a, b, 0)$ class

- Assume N is claims count distribution that is supported in a possibly infinite interval $\mathcal{A} \subset \mathbb{N}_0$ containing 0.
- Corresponding probability weights are denoted by $p_k = P(N = k)$ for $k = 1, 2, 3, \ldots$ and we set $p_k = 0$ for $k \neq \mathcal{A}$.

$(a, b, 0)$ class or Panjer distribution

N belongs to $(a, b, 0)$ class (or is a Panjer distribution) if there exist constants $a, b \in \mathbb{R}$ such that for all $k = 1, 2, 3, \ldots$ we have the recursion

$$p_k = p_{k-1} \left(a + \frac{b}{k} \right)$$

The recursive expression provides a computationally efficient way to generate probabilities. Note that Panjer distributions require $p_0 > 0$.

Lemma 1

Assume N has a non-degenerate Panjer distribution or belongs to $(a, b, 0)$ class. N is either binomially, Poisson or negative-binomially distributed.

Proof of Lemma 1

- For N to have non-degenerate distribution function we need to have $|\mathcal{A}| > 1$. Hence we choose as initialization for recursion $k = 1 \in \mathcal{A}$ ($\mathcal{A}$ is interval containing at least 0 and 1).
- Panjer distribution then provides for this k the identity $p_1 = p_0(a + b)$.
- To have well defined distribution function we need to have $a + b \geq 0$, otherwise $p_1 < 0$.
- $a + b = 0$ provides degenerate distribution so we focus on $a + b > 0$

1. $a = 0$
   This implies $b > 0$ and

$$p_k = p_{k-1}\frac{b}{k} > 0 \text{ for all } k = 1, 2, \dots.$$

   This is exactly Poisson distribution with parameters $a = 0$ and $b = \lambda v > 0$ for $\mathcal{A} = \mathbb{N}_0$ because for Poisson distribution we have $\frac{p_k}{p_{k-1}} = \frac{\lambda v}{k}$.

Proof of Lemma 1

2. $a < 0$

We need to make sure that $a + b/k$ remains positive $\forall k \in \mathcal{A}$. This requires $|\mathcal{A}| < \infty$. We denote maximal value in $\mathcal{A}$ by $v \in \mathbb{N}$ (assuming it has $p_v > 0$). The positivity constraint then provides $b/v > -a$ and $a + \frac{b}{v+1} = 0$. Latter implies that $p_k = 0$ for all $k > v$ and is equivalent to the requirement $v = -(a+b)/a > 0$. We set $p = -a/(1-a) \in (0,1)$ which provides

$$p_k = p_{k-1}\left(a + \frac{b}{k}\right) = p_{k-1}\left(a - \frac{a(v+1)}{k}\right) = p_{k-1}\frac{-p}{1-p}\left(1 - \frac{v+1}{k}\right)$$

For the binomial distribution we have on $\mathcal{A}$

$$\frac{p_k}{p_{k-1}} = \frac{p}{1-p}\frac{v-k+1}{k} = -\frac{p}{1-p} + \frac{p}{1-p}\frac{v+1}{k}$$

This is exactly binomial distribution with parameters $a = -p/(1-p)$ and $b = (v+1)p/(1-p)$ and $\mathcal{A} = \{0, \ldots, v\}$

Proof of Lemma 1

3. $a > 0$
   We define $\gamma = (a+b)/a > 0$. This provides $b = a(\gamma - 1)$ and

$$p_k = p_{k-1}\left(a + \frac{b}{k}\right) = p_{k-1}a\left(1 + \frac{\gamma - 1}{k}\right)$$

   Since latter should be summable in order to obtain well-defined distribution function we need to have $a < 1$.
   For negative-binomial distribution we have

$$\frac{p_k}{p_{k-1}} = \frac{p(k+\gamma-1)}{k} = p + \frac{p(\gamma-1)}{k}$$

   This is exactly negative-binomial distribution with parameters $a = p$ and $b = p(\gamma - 1)$ and $\mathcal{A} = \mathbb{N}_0$

$(a, b, 0)$ class

Lemma 1 shows that the important claims count distributions are Panjer distributions or belong to $(a, b, 0)$ class and the corresponding choices $a, b \in \mathbb{R}$ are provided in the proof.

Corollary

Assume N has non-degenerate Panjer distribution or belongs to $(a, b, 0)$ class.

- For $a = -p/(1-p)$ and $b = (v+1)p/(1-p)$ we have binomial distribution.
- For $a = 0$ and $b = \lambda v$ we have Poisson distribution
- For $a = p$ and $b = p(\gamma - 1)$ we have negative-binomial distribution with $p = \lambda v/(\gamma + \lambda v)$

Exercise 1

The number of dental claims in a year follows a Poisson distribution with a mean of λv. The probability of exactly 6 claims during a year is 40% of the probability that there will be 5 claims. Determine the probability that there will be 4 claims.

Answer of exercise 1

- Let N be the number of dental claims in a year.
- Since N is a Panjer distribution, we can write

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k} = \frac{\lambda v}{k} \qquad k = 1, 2, \ldots$$

- We are told that

$$0.4 = \frac{p_6}{p_5} = \frac{\lambda v}{6} \Rightarrow \lambda v = 2.4$$

- Therefore

$$p_4 = \frac{2.4^4}{4!} e^{-2.4} = 0.1254$$

Exercise 2

X is a discrete random variable with a distribution belonging to the $(a, b, 0)$ class. It is also known that

$$\begin{aligned} P(X=0) &= P(X=1)=0.25 \\ P(X=2) &= 0.1875 \end{aligned}$$

Calculate $P(X=3)$

Answer of exercise 2

- Let N denote the distribution under consideration.
- Since N is a member of $(a, b, 0)$ class, we know

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1} \qquad k = 1, 2, \ldots$$

- From first equation it follows

$$0.25 = (a + b)0.25 \Rightarrow a + b = 1$$

- From second equation it follows

$$0.1875 = (a + b/2)0.25 = (1 - b/2)0.25 \Rightarrow b = 0.5$$

- Hence $a = 0.5$ and

$$P(X = 3) = (0.5 + 0.5/3)(0.1875) = 0.125$$

Other frequency distributions

- Naturally, there are many other count distributions needed in practice.
- For many insurance applications, one can work with one of our three basic distributions (binomial, Poisson, negative binomial) and **allow the parameters to be a function of known explanatory variables**.
  - This allows us to explain claim probabilities in terms of known (to the insurer) variables such as age, sex, territory, ...
  - This field of statistical study is known as **regression modeling** (see later).
- To extend our basis count distributions to alternatives needed in practice, we consider two approaches:
  - Zero truncation or modification
  - Mixing

Zero truncation or modification

- Why truncate or modify zero?
  - If we work with a database of claims, then there are no zeroes!
  - In personal lines (like auto), people may not want to report a certain (first or small) claim because they fear it will increase future insurance rates.
- Therefore we **modify zero probabilities** in terms of the $(a, b, 0)$ class

$(a, b, 1)$ class

A count distribution belongs to $(a, b, 1)$ class if there exist constants $a, b \in \mathbb{R}$ such that for all $k = 2, 3, \ldots$ the probabilities p_k satisfy

$$p_k = p_{k-1}\left(a + \frac{b}{k}\right)$$

Hence recursion starts at p_1, not at p_0

- Therefore all distributions that are members of $(a, b, 0)$ class are also members of $(a, b, 1)$ class. Naturally, there are **additional distributions that are members of this wider class**.

Zero truncation or modification

To see how this works, pick a specific distribution in $(a, b, 0)$ class

- Consider p_k^0 to be a probability for this member of $(a, b, 0)$.
- Let p_k^M be the corresponding probability for a member of $(a, b, 1)$ where M stands for *modified*.
- Pick a new probability of a zero claim, p_0^M, and define $c = \frac{1-p_0^M}{1-p_0^0}$.
- We then calculate zero modified distribution as $p_k^M = cp_k^0$.
- Note that $\sum_{k=0}^{\infty} p_k^M = 1$!

Truncated at Zero

Assume that $p_0^M = 0$, so that probability of $N = 0$ is zero (*truncated at zero*). Then we get zero truncated probabilities (where we use T instead of M now):

$$p_k^T = \begin{cases} 0 & k = 0 \\ \frac{1}{1-p_0^0} p_k^0 & k \geq 1 \end{cases}$$

Exercise 3

- Consider Poisson distribution with parameter $\lambda v = 2$.
- Calculate p_k, $k = 0, 1, 2, 3$ for the usual (unmodified), truncated and a modified version with $p_0^M = 0.6$.

Answer of exercise 3

Solution For the Poisson distribution as a member of the $(a, b, 0)$ class, we have $a = 0$ and $b = \lambda v = 2$. Thus we may use recursion $p_k = \lambda v p_{k-1}/k = 2p_{k-1}/k$ for each type after determining starting probabilities.

k	$p_k^0 = p_k$	p_k^T	p_k^M
0	$p_0 = e^{-\lambda v} = 0.135335$	0	0.6
1	$p_1 = p_0 \left(0 + \frac{\lambda v}{1}\right) = 0.27067$	$\frac{p_1}{1-p_0} = 0.313035$	$\frac{1-p_0^M}{1-p_0} p_1 = 0.125214$
2	$p_2 = p_1 \left(\frac{\lambda v}{2}\right) = 0.27067$	$p_1^T \left(\frac{\lambda v}{2}\right) = 0.313035$	$p_1^M \left(\frac{\lambda v}{2}\right) = 0.125214$
3	$p_3 = p_2 \left(\frac{\lambda v}{3}\right) = 0.180447$	$p_2^T \left(\frac{\lambda v}{3}\right) = 0.208690$	$p_2^M \left(\frac{\lambda v}{3}\right) = 0.083476$

Mixtures of finite populations

- Suppose that our population consists of several subgroups, each having their own distribution.
- We randomly draw an observation from the population, without knowing which subgroup that we are drawing from.
- For example, suppose that N_1 represents claims from *good* drivers and N_2 represents claims from *bad* drivers. We draw

$$N = \begin{cases} N_1 & \text{with probability } \alpha \\ N_2 & \text{with probability } 1-\alpha \end{cases}$$

- Here, α represents the probability of drawing a *good* driver.
- Hence we have a mixture of two subgroups

Exercise 4

- In a certain town the number of common colds an individual will get in a year follows a Poisson distribution that depends on the individual's age and smoking status.
- The distribution of the population and mean number of colds are as follows

	proportion of population	mean number of colds
Children	0.3	3
Adult non-smokers	0.6	1
Adult smokers	0.1	4

Calculate the conditional probability that a person with exactly 3 common colds in a year is an adult smoker.

Answer of exercise 4

Let C denote children; ANS denote Adult non-smokers and AS denote Adult smokers, then

$$
\begin{aligned}
P(AS|N=3) &= \frac{P(AS \cap N=3)}{P(N=3)} = \frac{P(N=3|AS)P(AS)}{P(N=3)} \\
P(N=3) &= P(N=3|c)P(c) + P(N=3|ANS)P(ANS) \\
&\quad + P(N=3|AS)P(AS) \\
P(N=3|c)P(c) &= \frac{3^3 e^{-3}}{3!} 0.3 = 0.067 \\
P(N=3|ANS)P(ANS) &= \frac{1e^{-1}}{3!} 0.6 = 0.037 \\
P(N=3|AS)P(AS) &= \frac{4^3 e^{-4}}{3!} 0.1 = 0.020 \\
P(AS|N=3) &= \frac{0.020}{0.067+0.037+0.020} = 0.16
\end{aligned}
$$

Mixtures of infinitely many populations

- We can extend the mixture idea to an infinite number of populations.
- To illustrate, suppose we have a population of drivers. The ith person has their own (personal) expected number of claims λ_i.
- For some drivers, λ is small (*good* drivers), for others it is high (*bad* drivers). There is a distribution of λ.
- A convenient distribution is to use a gamma distribution with parameters (γ, c).
- This is how we introduced the negative binomial distribution (gamma mixtures of Poisson).

Outline

Modeling loss severity

- Asset values, losses and claim sizes are usually modeled as continuous random variables and as such are modeled using continuous distributions, often referred to as **loss, severity or claims size** distributions.
- For modeling more than one type of claims in liability insurance (small frequent claims and large relatively large claims), we also investigate **mixture** distributions.
- We present also tools to **create new continuous distributions from existing ones**.
- We explore the effect of **coverage modifications** such as applying deductibles and limits on the distribution of individual loss adjustments.

Individual Claim Size Modeling

- We have introduced the **compound distribution**

$$S = Y_1 + Y_2 + \ldots + Y_N = \sum_{i=1}^{N} Y_i$$

  with 3 standard assumptions
  - N is discrete random variable which takes values in $\mathcal{A} \subset \mathbb{N}_0$
  - $Y_1, Y_2, \ldots \overset{iid}{\sim} G$ with $G(0) = 0$
  - N and $(Y_1, Y_2, \ldots)$ are independent.
- We have discussed modeling of claims **count** distribution of N.
- Now we concentrate on **modeling of individual claim sizes** Y_i.
- To get understanding for modeling of G we present analysis based on **two explicit data sets**.
  - first data set is private property (PP) insurance: 72769 claims records
  - second data set commercial property (CP) insurance: 18285 claims records
- We first analyse data sets using tools from descriptive statistics.

Exploratory Data Analysis

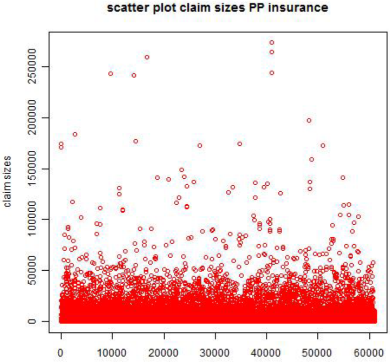
- First step in data analysis should be to **look at the data** and to plot the data in several ways
  - to catch mistakes
  - to see patterns (reveal structure) in the data
  - to find violations of statistical assumptions
  - to generate hypotheses
- Problems such as bad data, outliers, mislabeling of variables, missing elements and an unsuitable model can often be **detected by visual inspection** or examination.
- Bad data (outlying because of errors) should be corrected when possible and otherwise deleted.
- Outliers due, for example, to a stock market crash are *good data* and should be kept in mind (perhaps expand the model to accommodate them). It is important to detect atypical observations and to investigate and understand them so that appropriate action can be taken!

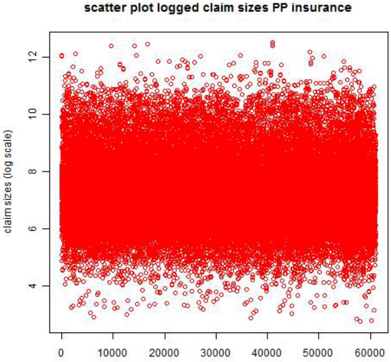
Data analysis and descriptive statistics

- Data sets contain many claims records with **zero claims payments**
  PP insurance: 16% and CP insurance: 21%
- Possible **reasons**:
  - final claim does not exceed deductible
  - insurance company not liable for claim
  - another insurance policy covers claim
  - reporting (small) claim reduces no-claims-benefit too much so that insured decides to withdraw claim.
- Two different ways to **deal with such zero claims**
  - estimate zero probability separately and add probability weight to G at 0.
    $\Rightarrow$ mathematically consistent but contradicts our model assumption $G(0) = 0$.
  - reduce expected claims frequency λ by these zero claims
    $\Rightarrow$ perfectly fits into compound Poisson modeling framework due to disjoint decomposition theorem (see later).
- Therefore we assume $G(0) = 0$ and $E[N] = \lambda v$ where $v > 0$ is portfolio size and N only counts strictly positive claims.
- After **subtracting zero claims**: $n = 61053$ (PP) and $n = 14532$ (CP).

Data analysis and descriptive statistics

- Scatter plots of data



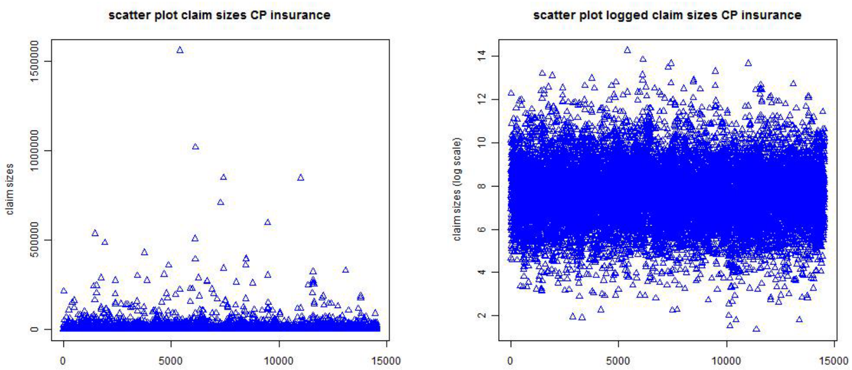


- We calculate sample means and sample variances

$$\text{PP} : \hat{\mu}_n = 3116; \qquad \hat{\sigma}_n = 7534; \qquad \hat{\text{Vco}}_n = 2.42$$

Data analysis and descriptive statistics

- Scatter plots of data

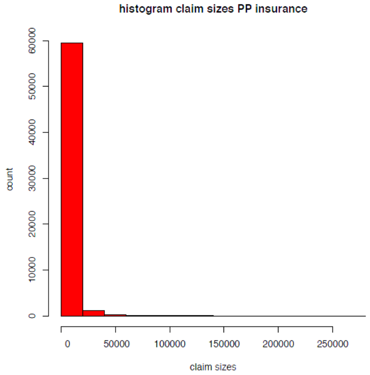


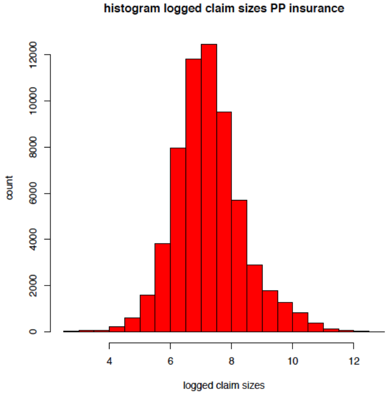
- We calculate sample means and sample variances

$$\text{CP} : \hat{\mu}_n = 6850; \qquad \hat{\sigma}_n = 28505; \qquad \hat{\text{Vco}}_n = 4.16$$

Data analysis and descriptive statistics

- Next we give histogram for PP insurance and for **logged** claim sizes

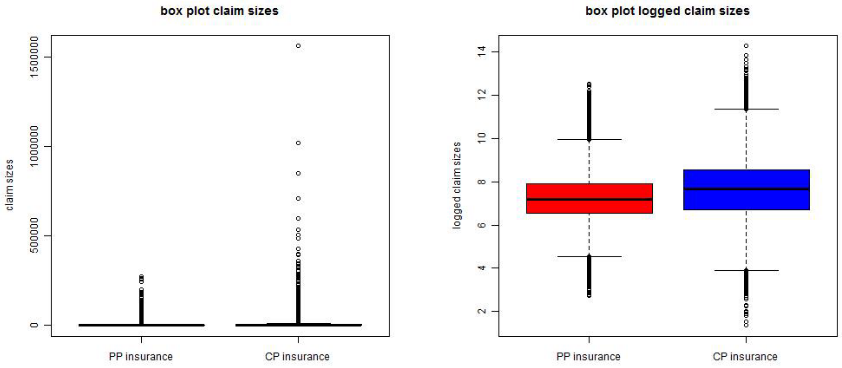




$\Rightarrow$ **Few large claims distort whole picture** so that histogram is not helpful.

Data analysis and descriptive statistics

- Next we give box plots for PP and CP (original and log scale)



$\Rightarrow$ **Positive skewness.**

Data analysis and descriptive statistics

- Goal is to have full distribution functions $G(y) = \mathbb{P}[Y \leq y]$ of both claims classes PP and CP.
- Since we have many observations, we could work with empirical distribution function

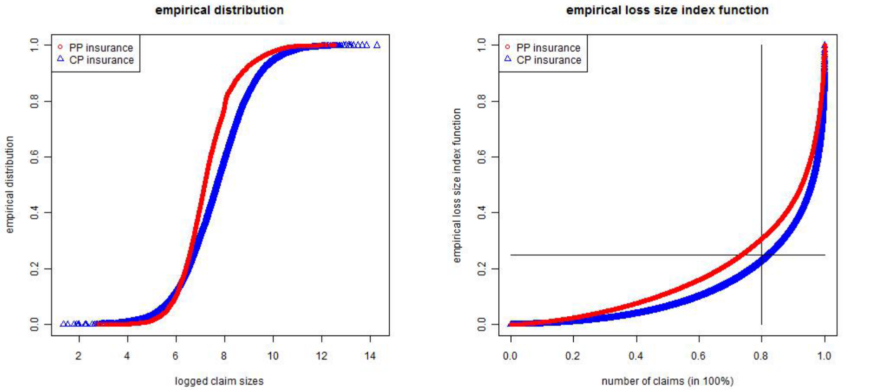
$$\hat{G}_n(y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i \leq y\}}$$

- We define loss size index function and empirical version as

$$\mathcal{I}(G(y)) = \frac{\int_0^y z dG(z)}{\int_0^\infty z dG(z)} \quad \text{and} \quad \hat{\mathcal{I}}_n(\alpha) = \frac{\sum_{n=1}^{\lfloor n\alpha \rfloor} Y_{(i)}}{\sum_{i=1}^{n} Y_i} \quad \text{for } \alpha \in [0,1]$$

loss size index function chooses claim size threshold y and then it evaluates relative expected claim among that can be explained by claim sizes up to this threshold y.

Data analysis and descriptive statistics



⇒ **20% largest claims roughly cause 75%** of total claim size.
⇒ Understand well large claims!

Analysis of tail

For **analysis of tail** of distribution we consider property of regular variation at infinity. Assume that G has **infinite support** and that $\overline{G} = 1 - G$ is survival function.

- $\overline{G} \in \mathcal{R}_{-\alpha}$: $\overline{G} = 1 - G$ is regularly varying at infinity with (tail) index $\alpha > 0$ if

$$\lim_{x\to\infty} \frac{\overline{G}(xt)}{\overline{G}(x)} = \lim_{x\to\infty} \frac{1 - G(xt)}{1 - G(x)} = t^{-\alpha} \qquad \forall t > 0$$

- If the above holds true for $\alpha = 0$ then $\overline{G} \in \mathcal{R}_0$.
- If the above holds true for $\alpha = \infty$ then $\overline{G}$ is rapidly varying at infinity: $\overline{G} \in \mathcal{R}_{-\infty}$.

G with $\overline{G} \in \mathcal{R}_{-\alpha}$ for some $\alpha \in [0, = \infty)$ are dangerous because they have **large potential for big claims**. Therefore it is crucial to know this index of regular variation at infinity.

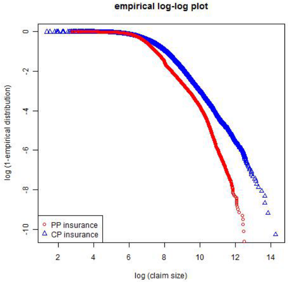
Two plots that especially focus on large claims are:

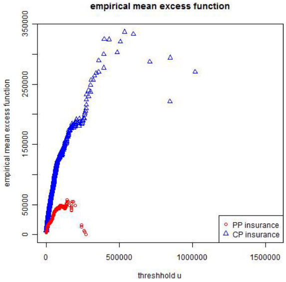
- (empirical) mean excess plot

$$u \mapsto e(u) = E[Y_i - u | Y > u] \quad \text{and} \quad u \mapsto \hat{e}_n(u) = \frac{\sum_{i=1}^n (Y_i - u) 1_{\{Y_i > u\}}}{\sum_{i=1}^n 1_{\{Y_i > u\}}}$$

- (empirical) log-log plot

$$y \mapsto (\log y, \log(1 - G(y))) \quad \text{and} \quad y \mapsto \left(\log y, \log(1 - \hat{G}_n(y)\right)$$





Linear decrease in log-log plot and linear increase in mean excess plot will have interpretation of heavy tailed distributions in the sense that $\overline{G}$ is regularly varying at infinity.

Selected parametric claims size distributions

We introduce **popular parametric claim size distributions**.
We only consider distribution functions G with unbounded support in $\mathbb{R}_+$.
We use following notation for rv $Y \sim G$

- g: density of Y for G absolutely continuous
- $M_Y(r)$: moment generating function of Y in $r \in \mathbb{R}$, where it exists
- μ_Y: expected value of Y, if it exists
- σ_Y^2: variance of Y, if it exists
- $\mathrm{Vco}(Y)$: coefficient of variation of Y, if it exists
- ζ_Y: skewness of Y, if it exists
- $\overline{G} = 1 - G$: survival function of Y, i.e. $\overline{G}(y) = \mathbb{P}[Y > y]$

For the analysis of G also following quantities are of interest

- $E[Y 1_{\{u_1 < Y \le u_2\}}]$: expected value of Y within layer $(u_1, u_2]$
- $I(G(y)) = E[Y 1_{\{Y \le y\}}]/\mu_Y$: loss size index function for level y
- $e(u) = E[Y - u | Y > u]$: mean excess function of Y above u

Gamma distribution

- $Y \sim \Gamma(\gamma, c)$: Gamma distribution with shape parameter $\gamma > 0$ and scale parameter $c > 0$

$$g(y) = \frac{c^\gamma}{\Gamma(\gamma)} y^{\gamma-1} e^{-cy} \qquad \text{for } y \geq 0$$

- No closed form solution for distribution function G

$$G(y) = \int_0^y \frac{c^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-cx} dx = \frac{1}{\Gamma(\gamma)} \int_0^{cy} z^{\gamma-1} e^{-z} dz = \mathcal{G}(\gamma, cy) \qquad y \geq 0$$

  where $\mathcal{G}(.,.)$ is incomplete gamma function.

- Family of gamma distributions is closed towards multiplication with positive constant, that is, for $\rho > 0$ we have

$$\rho Y \sim \Gamma(\gamma, c/\rho)$$

  This property is important when we deal with claims inflation.

Gamma distribution

- For mfg and moments we have

$$
\begin{aligned}
M_Y(r) &= \left(\frac{c}{c-r}\right)^{\gamma} \qquad \text{for } r < c \\
\mu_Y &= \frac{\gamma}{c} \\
\sigma_Y^2 &= \frac{\gamma}{c^2} \\
\mathrm{Vco}(Y) &= \gamma^{-1/2} \\
\zeta_Y &= 2\gamma^{-1/2} > 0
\end{aligned}
$$

- For $0 \leq u_1 < u_2$ and $u, y > 0$ we obtain

$$
\begin{aligned}
E[Y 1_{\{u_1 < Y \leq u_2\}}] &= \frac{\gamma}{c}\left[\mathcal{G}(\gamma+1, cu_2) - \mathcal{G}(\gamma+1, cu_1)\right] \\
I(G(y)) &= \mathcal{G}(\gamma+1, cy) \\
e(u) &= \frac{\gamma}{c}\left(\frac{1-\mathcal{G}(\gamma+1, cu)}{1-\mathcal{G}(\gamma, cu)}\right) - u
\end{aligned}
$$

Gamma distribution

Exercise 5

Assume $Y \sim \Gamma(\gamma, c)$

1. Prove the statements of M_Y and $I(G(y))$.
2. Prove statements

$$\begin{aligned} e(u) &= \mu_Y \frac{1 - I(G(u))}{1 - G(u)} - u \\ E[Y 1_{\{u_1 < Y \leq u_2\}}] &= \mu_Y (I(G(u_2)) - I(G(u_1))) \end{aligned}$$

Gamma distribution

Answer of exercise 5

1.

$$\begin{aligned}
M_Y(r) &= \int_{-\infty}^{\infty} e^{ry} \frac{c^\gamma}{\Gamma(\gamma)} y^{\gamma-1} e^{-cy} dy \qquad y \geq 0 \\
&= \frac{c^\gamma}{\Gamma(\gamma)} \int_0^{\infty} e^{-y(c-r)} y^{\gamma-1} dy \qquad z = y(c-r) \text{ and } \frac{dz}{c-r} \\
&= \frac{c^\gamma}{\Gamma(\gamma)} \int_0^{\infty} e^{-z} z^{\gamma-1} \frac{1}{(c-r)^{\gamma-1}} \frac{dz}{c-r} \\
&= \frac{c^\gamma}{(c-r)^\gamma} \frac{1}{\Gamma(\gamma)} \int_0^{\infty} e^{-z} z^{\gamma-1} dz \\
&= \frac{c^\gamma}{(c-r)^\gamma} \frac{1}{\Gamma(\gamma)} \Gamma(\gamma) \\
&= \left(\frac{c}{c-r}\right)^\gamma
\end{aligned}$$

Gamma distribution

2.

$$
\begin{aligned}
I(G(y)) &= \frac{1}{\mu_Y} E\left[Y 1_{Y \leq y}\right] = \frac{c}{\gamma} E\left[Y 1_{Y \leq y}\right] \\
&= \frac{c}{\gamma} \int_0^{\infty} u 1_{u \leq y} \frac{c^{\gamma}}{\Gamma(\gamma)} u^{\gamma-1} e^{-cu} du \\
&= \int_0^{y} \frac{c^{\gamma+1}}{\gamma \Gamma(\gamma)} u^{\gamma} e^{-cu} du \\
&= \int_0^{\gamma} \frac{c^{\gamma+1}}{\Gamma(\gamma+1)} u^{\gamma} e^{-cu} du \\
&= \mathcal{G}(\gamma+1, cy)
\end{aligned}
$$

Gamma distribution

3.

$$
\begin{aligned}
e(u) &= E[Y-u|Y>u] = \frac{1}{P(Y>u)}\int_u^\infty (y-u)\frac{c^\gamma}{\Gamma(\gamma)}y^{\gamma-1}e^{-cy}dy \\
&= \frac{1}{1-G(u)}\left[\int_u^\infty y\frac{c^\gamma}{\Gamma(\gamma)}y^{\gamma-1}e^{-cy}dy - u\int_u^\infty \frac{c^\gamma}{\Gamma(\gamma)}y^{\gamma-1}e^{-cy}dy\right] \\
&= \frac{1}{1-G(u)}\left[\mu_Y - \int_0^u y\frac{c^\gamma}{\Gamma(\gamma)}y^{\gamma-1}e^{-cy}dy - u(1-G(u))\right] \\
&= \frac{1}{1-G(u)}\left[\mu_Y - E\left[Y1_{Y\leq u}\right] - u(1-G(u))\right] \\
&= \frac{\mu_Y - \mu_Y I(G(u))}{1-G(u)} - u \\
&= \mu_Y\frac{1-I(G(u))}{(1-G(u))} - u
\end{aligned}
$$

Gamma distribution

4

$$\begin{aligned} E[Y1_{u_1 \leq Y \leq u_2}] &= E[Y1_{Y \leq u_2}] - E[Y1_{Y \leq u_1}] \\ &= \mu_Y I(G(u_2)) - \mu_Y I(G(u_1)) \\ &= \mu_Y \left(I(G(u_2)) - I(G(u_1)) \right) \end{aligned}$$

Gamma distribution

- Gamma distribution does not have a regularly varying tail at infinity. In fact, $\overline{G}(y) = 1 - G(y)$ decays roughly as e^{-cy} to 0 as $y \to \infty$ because e^{-cy} gives asymptotic lower bound and $e^{-(c-\epsilon)y}$ as an asymptotic upper bound for any $\epsilon > 0$ on $\overline{G}(y)$.
- Generating gamma random numbers in R: rgamma(n,shape=gamma, rate=c)
- Method of moment estimators are given by

$$\hat{c}^{MM} = \frac{\hat{\mu}_n}{\hat{\sigma}_n^2} \quad \text{and} \quad \hat{\gamma}^{MM} = \frac{\hat{\mu}_n^2}{\hat{\sigma}_n^2}$$

- For MLE we have log-likelihood function, set $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$

$$\ell_{\boldsymbol{Y}}(\gamma, c) = \sum_{i=1}^{n} \gamma \log c - \log \Gamma(\gamma) + (\gamma - 1) \log Y_i - cY_i$$

Gamma distribution

Then MLE $\hat{\gamma}^{MLE}$ of γ is solution

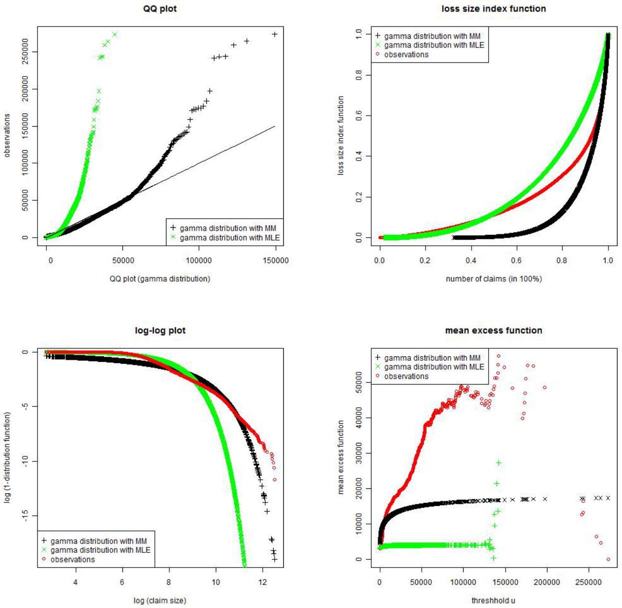
$$\log \gamma - \log \hat{\mu}_n - \frac{\Gamma'(\gamma)}{\Gamma(\gamma)} + \frac{1}{n} \sum_{i=1}^{n} \log Y_i = 0$$

This is solved numerically and MLE for c is then given by

$$\hat{c}^{MLE} = \frac{\hat{\gamma}^{MLE}}{\hat{\mu}_n}$$

- For numerical solution in R: fitdistr(data, "gamma")
- Numerical fitting does not always work when range of observations $\boldsymbol{Y}$ is too large. In such cases it is recommended that in first step the data is scaled by constant factor $\rho > 0$ and parameters are estimated for scaled data. In second step, the constant is scaled back by same factor.
- Special cases
  - $\gamma = 1$ is exponential distribution with parameter $c > 0$.
  - $\gamma = k/2$ and $c = 1/2$ is χ^2_k distribution.

We fit PP insurance data to gamma distribution.



Weibull distribution

- $Y \sim \text{Weibull}(\tau, c)$ Weibull distributed with shape parameter $\tau > 0$ and scale parameter $c > 0$

$$g(y) = (c\tau)(cy)^{\tau-1}e^{-(cy)^\tau}$$

- Survival function does not have regularly varying tail at infinity, but decay of

$$G(y) = 1 - e^{-(cy)^\tau} \qquad \text{for } y \geq 0$$

  is slower than in gamma case for $\tau < 1$. In fact $\overline{G}(y) = 1 - G(y)$ decays as $e^{-(cy)^\tau}$ to 0 for $y \to \infty$.

- Family of Weibull distributions is closed towards multiplication with $\rho > 0$

$$\rho Y \sim \text{Weibull}(\tau, c/\rho)$$

Weibull distribution

- Mgf $M_Y(r)$ does not exist for $\tau < 1$ and $r > 0$ and moments are

$$\begin{aligned}
\mu_Y(r) &= \frac{\Gamma(1+1/\tau)}{c} \\
\sigma_Y^2 &= \frac{\Gamma(1+2/\tau)}{c^2} - \mu_Y^2 \\
\zeta_Y &= \frac{1}{\sigma_Y^3}\left[\frac{\Gamma(1+3/\tau)}{c^3} - 3\mu_Y\sigma_Y^2 - \mu_Y^3\right]
\end{aligned}$$

- For $0 \le u_1 < u_2$ and $u, y > 0$ we obtain

$$\begin{aligned}
E[Y1_{u_1<Y\le u_2}] &= \frac{\Gamma(1+1/\tau)}{c}\left[\mathcal{G}(1+1/\tau,(cu_2)^\tau) - \mathcal{G}(1+1/\tau,(cu_1)^\tau)\right] \\
I(G(y)) &= \mathcal{G}(1+1/\tau,(cy)^\tau) \\
e(u) &= \frac{\Gamma(1+1/\tau)}{c}\left(\frac{1-\mathcal{G}(1+1/tau,(cu)^\tau)}{e^{-(cu)^\tau}}\right) - u
\end{aligned}$$

Weibull distribution

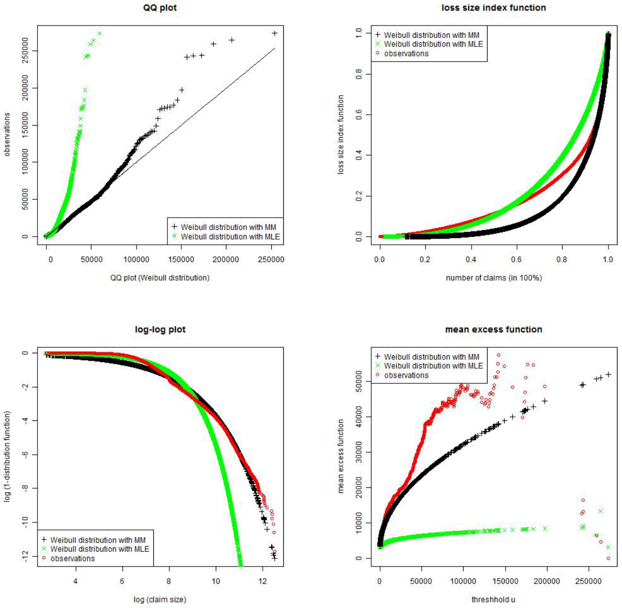
- Generating Weibull random numbers by observing that we have identity $Y \stackrel{(d)}{=} \frac{1}{c} Z^{1/\tau}$ with $Z \sim \exp(1) \stackrel{(d)}{=} \Gamma(1,1)$: rgamma(n,shape=1,rate=1)
- Method of moment estimators are given by

$$
\begin{aligned}
\hat{c}^{MM} &= \frac{\Gamma(1 + 1/\hat{\tau}^{M}M)}{\hat{\mu}_n} \\
\frac{\hat{\sigma}_n^2}{\hat{\mu}_n^2} + 1 &= \frac{1 + 2/\hat{\tau}^{MM}}{\Gamma(1 + 1/\hat{\tau}^{M}M)^2}
\end{aligned}
$$

  which needs to be solved numerically in R
- For MLE we need to solve system of equations

$$
\begin{aligned}
c &= \left(\frac{1}{n} \sum_{i=1}^{n} Y_i^{\tau} \right)^{-1/\tau} \\
\tau \frac{1}{n} \sum_{i=1}^{n} \log(cY_i)((cY_i)^{\tau} - 1) &= 1
\end{aligned}
$$

We fit PP insurance data to Weibull distribution ($\hat{\tau} \in (0.5, 0.75)$).



Log-normal distribution

- $Y \sim \text{LN}(\mu, \sigma^2)$ log-normal distributed with mean parameter $\mu \in \mathbb{R}$ and standard deviation parameter $\sigma > 0$

$$\begin{aligned} g(y) &= \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{y} e^{-\frac{1}{2}\frac{(\log y - \mu)^2}{\sigma^2}} \qquad \text{for } y \geq 0 \\ G(y) &= \Phi\left(\frac{\log y - \mu}{\sigma}\right) \end{aligned}$$

  with $\Phi(.)$ denoting standard Gaussian distribution function.
- Family of log-normal distributions is closed towards multiplication with positive constant, that is, for $\rho > 0$ we have

$$\rho Y \sim \log n(\mu + \log \rho, \sigma^2)$$

Log-normal distribution

- Mgf does not exist for $r > 0$ and we have following moments

$$\begin{aligned}
\mu_Y &= e^{\mu+\sigma^2/2} \\
\sigma_Y^2 &= e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \\
\mathrm{Vco}(Y) &= \left(e^{\sigma^2}-1\right)^{1/2} \\
\zeta_Y &= \left(e^{\sigma^2}+2\right)\left(e^{\sigma^2}-1\right)^{1/2}
\end{aligned}$$

- For $0 \leq u_1 < u_2$ and $u, y > 0$ we obtain

$$\begin{aligned}
E[Y 1_{u_1 < Y \leq u_2}] &= \mu_Y \left[\Phi\left(\frac{\log u_2 - (\mu+\sigma^2)}{\sigma}\right) - \Phi\left(\frac{\log u_1 - (\mu+\sigma^2)}{\sigma}\right)\right] \\
I(G(y)) &= \Phi\left(\frac{\log y - (\mu+\sigma^2)}{\sigma}\right) \\
e(u) &= \mu_Y \left(\frac{1-\Phi\left(\frac{\log u - (\mu+\sigma^2)}{\sigma}\right)}{1-\Phi\left(\frac{\log u - \mu}{\sigma}\right)}\right) - u
\end{aligned}$$

Log-normal distribution

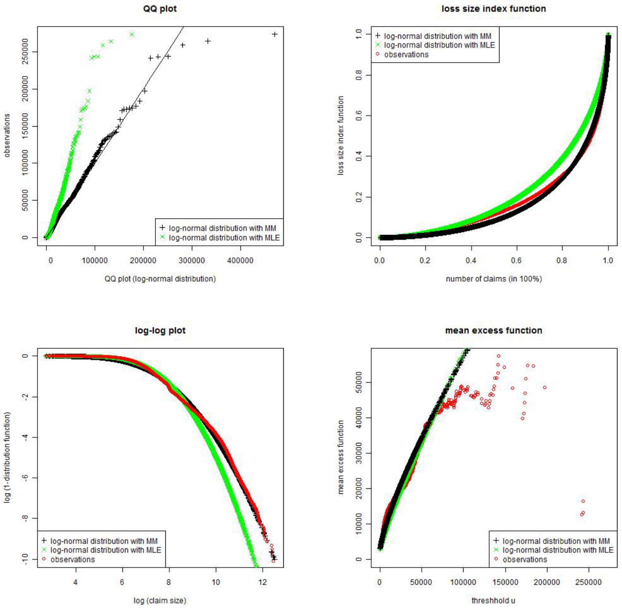
- Log-normal distribution does not have regularly varying survival function at infinity.
- Generating log-normal random numbers
  - Choose standard Gaussian numbers $Z \sim \Phi$
  - Set $Y = e^{\mu+\sigma Z}$
- Method of moment estimators are given by

$$\begin{aligned} \hat{\sigma}^{MM} &= \left[\log\left(\frac{\hat{\sigma}_n^2}{\hat{\mu}_n^2}+1\right)\right]^{1/2} \\ \hat{\mu}^{MM} &= \log \hat{mu}_n - (\hat{\sigma}^{MM})^2/2 \end{aligned}$$

- MLE is given by

$$\begin{aligned} \hat{\mu}^{MLE} &= \frac{1}{n}\sum_{i=1}^{n} \log Y_i \\ (\hat{\sigma}^{MLE})^2 &= \frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n}\left(\log Y_i - \hat{\mu}^{MLE}\right)^2 \end{aligned}$$

We fit PP and observe that log-normal distribution gives quite a good fit.



Log-gamma distribution

- Log-gamma distribution is more heavy tailed than log-normal distribution.
- It is obtained by assuming $\log Y \sim \Gamma(\gamma, c)$ for positive parameters γ and c

$$\begin{aligned} g(y) &= \frac{c^\gamma}{\Gamma(\gamma)}(\log y)^{\gamma-1} y^{-(c+1)} \qquad \text{for } y \geq 1 \\ G(y) &= \mathcal{G}(\gamma, c\log y) \end{aligned}$$

- Mgf does not exist for $r > 0$ and for moments we have

$$\begin{aligned} \mu_Y &= \left(\frac{c}{c-1}\right)^\gamma \qquad \text{for } c > 1 \\ \sigma_Y^2 &= \left(\frac{c}{c-2}\right)^\gamma - \mu_Y^2 \qquad \text{for } c > 2 \\ \zeta_Y &= \frac{1}{\sigma_Y^3}\left[\left(\frac{c}{c-3}\right)^\gamma - 3\mu_Y\sigma_Y^2 - \mu_Y^3\right] \quad \text{for } c > 3 \end{aligned}$$

Log-gamma distribution

- For $0 \leq u_1 < u_2$ and $u, y > 0$ we obtain

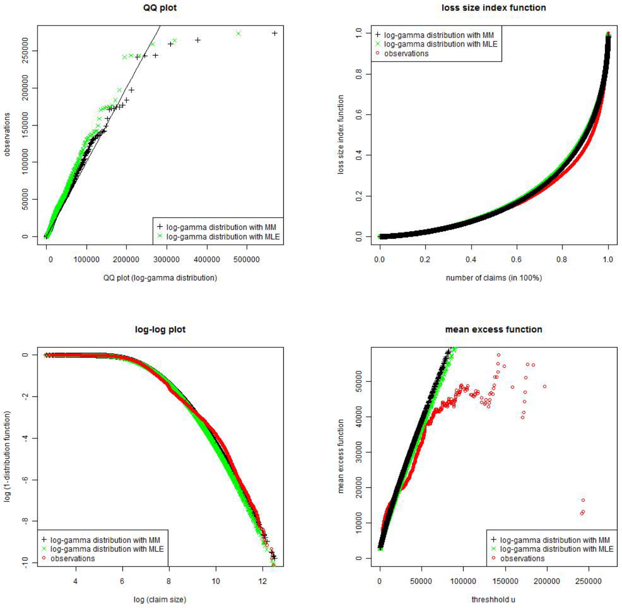
$$
\begin{aligned}
E[Y 1_{u_1 < Y \leq u_2}] &= \left(\frac{c}{c-1}\right)^{\gamma} [\mathcal{G}(\gamma, (c-1)\log u_2) - \mathcal{G}(\gamma, (c-1)\log u_1)] \\
I(G(y)) &= \mathcal{G}(\gamma, (c-1)\log y) \\
e(u) &= \left(\frac{c}{c-1}\right)^{\gamma} \left(\frac{1 - \mathcal{G}(\gamma, (c-1)\log u)}{1 - \mathcal{G}(\gamma, c\log u)}\right) - u
\end{aligned}
$$

- Log-gamma has regularly varying survival function at infinity with $c > 0$
- Method of moment estimators are given by (solved numerically)

$$
\begin{aligned}
\hat{\gamma}^{MM} &= \frac{\log \hat{\mu}_n}{\log \frac{\hat{c}^{MM}}{\hat{c}^{MM}-1}} \\
\frac{\log(\hat{\sigma}_n^2 + \hat{\mu}_n^2)}{\log \hat{\mu}_n} &= \frac{\log \hat{c}^{MM} - \log(\hat{c}^{MM} - 2)}{\log \hat{c}^{MM} - \log(\hat{c}^{MM} - 1)}
\end{aligned}
$$

- MLE is obtained analogously to MLE for gamma observations by simply replacing $_i$ by $\log Y_i$

We fit PP and observe that log-gamma distribution gives best fit ($\hat{c} = 5.8$).



Pareto distribution

- $Y \sim \text{Pareto}(\theta, \alpha)$ with threshold $\theta > 0$ and tail index $\alpha > 0$

$$
\begin{aligned}
g(y) &= \frac{\alpha}{\theta}\left(\frac{y}{\theta}\right)^{-(\alpha+1)} \quad \text{for } y \geq \theta \\
G(y) &= 1 - \left(\frac{y}{\theta}\right)^{-\alpha}
\end{aligned}
$$

- Claims above threshold θ are assumed to have regularly varying tails with $\alpha > 0$.
- We have closedness towards multiplication with a positive constant $\rho > 0$

$$
\rho Y \sim \text{Pareto}(\theta\rho, \alpha)
$$

Pareto distribution

- Mgf does not exist for $r > 0$ and for moments we have

$$\begin{aligned}
\mu_Y &= \theta \frac{\alpha}{\alpha - 1} \qquad \text{for } \alpha > 1 \\
\sigma_Y^2 &= \theta^2 \frac{\alpha}{(\alpha-1)^2(\alpha-2)} \qquad \text{for } \alpha > 2 \\
\zeta_Y &= \frac{2(1+\alpha)}{\alpha - 3} \left(\frac{\alpha - 2}{\alpha} \right)^{1/2} \qquad \text{for } \alpha > 3
\end{aligned}$$

- For $0 \leq u_1 < u_2$ and $u, y > 0$ we obtain

$$\begin{aligned}
E[Y 1_{u_1 < Y \leq u_2}] &= \theta \frac{\alpha}{\alpha - 1} \left[\left(\frac{u_1}{\theta} \right)^{-\alpha+1} - \left(\frac{u_2}{\theta} \right)^{-\alpha+1} \right] \\
I(G(y)) &= 1 - \left(\frac{y}{\theta} \right)^{-\alpha+1} \\
e(u) &= \frac{1}{\alpha - 1} u
\end{aligned}$$

Pareto distribution

- As soon as we only study tails of distributions we should use MLEs for parameter estimation (MM is not sufficiently robust against outliers). Since threshold θ has natural meaning we only need to estimate α

$$\hat{\alpha}^{MLE} = \left(\frac{1}{n} \sum_{i=1}^{n} \log Y_i - \log \theta \right)^{-1}$$

Lemma 2

Assume $Y_1, \ldots, Y_n \overset{iid}{\sim}$ *Pareto*(θ, α), *then*

$$\begin{aligned} E[\hat{\alpha}^{MLE}] &= \frac{n}{n-1}\alpha \\ Var\left(\hat{\alpha}^{MLE}\right) &= \frac{n^2}{(n-1)^2(n-2)}\alpha^2 \end{aligned}$$

Proof: see course notes.

Pareto distribution

- For MLE of α it was assumed that θ is given in natural way. If this threshold needs to be detected from data, Hill estimator and plot can be of help.
- We order claims accordingly to $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)}$ and define Hill estimator

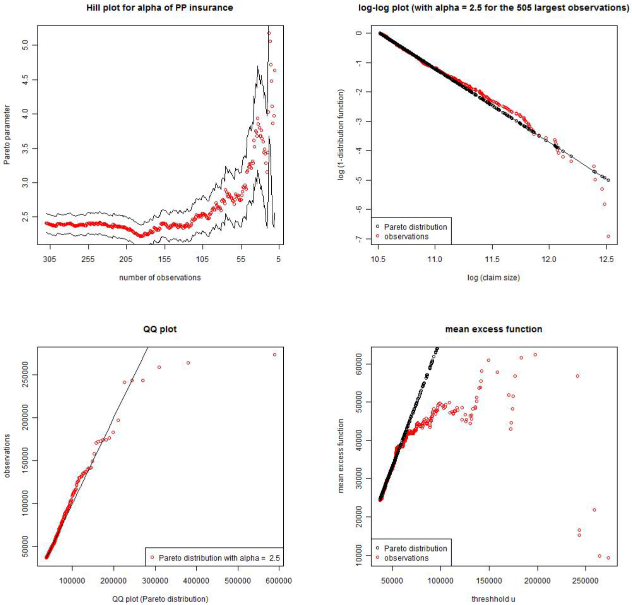
$$\hat{\alpha}^H_{k,n} = \left(\frac{1}{n-k+1} \sum_{i=k}^{n} \log Y_{(i)} - \log Y_{(k)} \right)^{-1} \qquad \text{for } k < n$$

- Hill estimator is based on rationale that Pareto distribution is closed towards increasing thresholds, .i.e. for $Y \sim \text{Pareto}(\theta_0, \gamma)$ and $\theta_1 > \theta_0$ we have

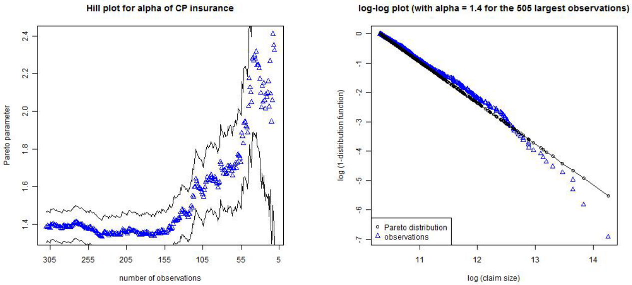
$$\mathbb{P}[Y > y | Y \geq \theta_1] = \frac{\left(\frac{y}{\theta_0}\right)^{-\alpha}}{\left(\frac{\theta_1}{\theta_0}\right)^{-\alpha}} = \left(\frac{y}{\theta_1}\right)^{-\alpha} \quad \text{for } y \geq \theta_1$$

- Therefore if data comes from Pareto distribution we should observe stability in $\hat{\alpha}^H_{k,n}$ for changing k.

We apply Pareto to PP data (only largest claims).

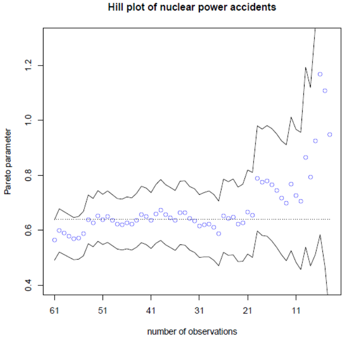


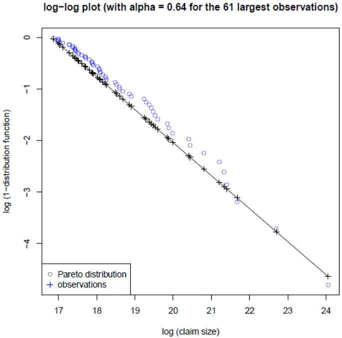
We analyze extremes of CP claims



At first sight they look similar to PP insurance example, i.e. they begin to destabilise between 150 and 100 largest claims. However main difference is that **tail index is much smaller** in CP example. That is, there is a **higher potential for large claims** for this line of business.

Example: Nuclear power accident data.
All nuclear power accidents that have occurred until end of 2011 that have claim size larger than 20 mio. USD.
We provide Hill plot and log-log plot (for $\alpha = 0.64$)





$\Rightarrow$ Data is very heavy tailed and Hill plot suggests to set tail index α around 0.64, which means that we have an infinite mean model.
Log-log plot shows that this tail captures the slope quite well.

Creating new distributions

Goal:

1. understand connections among distributions
2. give insights into when a distribution is preferred when compared to alternatives
3. provide foundations for creating new distributions

Let X be a continuous r.v. with pdf $f_X(x)$ and cdf $F_X(x)$. Apply following transformations to X to create **other, new** distributions:

- Multiplication by a constant
- Raising to a power
- Exponentiation
- Mixing

Multiplication by a constant

Theorem: multiplication by a constant

Let X be a continuous r.v. with pdf $f_X(x)$ and cdf $F_X(x)$. Consider transformation $\boldsymbol{Y = cX}$ with $c > 0$. Then

$$F_Y(y) = F_X\left(\frac{y}{c}\right) \quad \text{and} \quad f_Y(y) = \frac{1}{c} f_X\left(\frac{y}{c}\right)$$

- **Proof:**

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(cX \leq y) = P\left(X \leq \frac{y}{c}\right) = F_X\left(\frac{y}{c}\right) \\ f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{1}{c} f_X\left(\frac{y}{c}\right) \end{aligned}$$

- Equivalent to applying **inflation** uniformly across all loss levels; e.g. this year's losses are represented by X, with uniform inflation of 5% next year's losses are $Y = 1.05X$
- To account for **currency impact** on claim costs we apply currency conversion from a base to a counter currency.

Multiplication by a constant

- If $Y = cX$ $(c > 0)$ belongs to the same set of parametric distributions as X, then the distribution is said to be a **scale distribution**.
- When member of a scale distribution is multiplied by constant c $(c > 0)$, the scale parameter for this scale distribution meets two conditions:
  - the scale parameter is multiplied by the same c
  - all other parameters remain unchanged

Exercise 6

1. The aggregate losses of Eiffel Auto Insurance are denoted in euro currency and follow a Lognormal distribution with $\mu = 8$ and $\sigma = 2$. Given that 1 euro is 1.3 dollars, find the set of lognormal parameters, which describe the distribution of Eiffel's losses in dollars?
2. Demonstrate that the gamma distribution is a scale distribution.

Multiplication by a constant

Answer of exercise 6

1. Let X and Y denote the aggregate losses of Eiffel Auto Insurance in euro currency and dollars respectively. Then $Y = 1.3X$.

$$F_Y(y) = P(Y \leq y) = P(1.3X \leq y) = P(X \leq y/1.3) = F_X(y/1.3)$$

X follows a lognormal distribution with parameters $\mu = 8$ and $\sigma = 2$. The pdf of X is given by

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left[\frac{\log(x)-\mu}{\sigma}\right]^2\right) \qquad \text{for } x > 0.$$

Then the pdf of interest $f_Y(y)$ is

$$\begin{aligned} f_Y(y) &= \frac{1}{1.3} f_X\left(\frac{y}{1.3}\right) = \frac{1}{1.3}\frac{1.3}{y\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left[\frac{\log(y/1.3)-\mu}{\sigma}\right]^2\right) \\ &= \frac{1}{y\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left[\frac{\log y - (\log(1.3)+\mu)}{\sigma}\right]^2\right) \end{aligned}$$

Multiplication by a constant

Then Y follows a lognormal distribution with parameters $\log 1.3 + \mu = 8.26$ and $\sigma = 2.00$. If we let $\mu = \log(m)$ then it can be easily seen that $m = e^{\mu}$ is the scale parameter which was multiplied by 1.3 while σ is the shape parameter that remained unchanged.

2. Let $X \sim \Gamma(\alpha, \theta)$ and $Y = cX$, then

$$f_Y(y) = \frac{1}{c} f_X\left(\frac{y}{c}\right) = \frac{\left(\frac{y}{c\theta}\right)^{\alpha}}{y\Gamma(\alpha)} exp\left(-\frac{y}{c\theta}\right)$$

We can see that $Y \sim \Gamma(\alpha, c\theta)$ indicating that gamma is a scale distribution and θ is a scale parameter.

Raising to a power

Theorem: raising to a power

Let X be a continuous r.v. with pdf $f_X(x)$ and cdf $F_X(x)$.
Let $\boldsymbol{Y = X^\tau}$. Then, if $\boldsymbol{\tau > 0}$,

$$F_Y(y) = F_X(y^{1/\tau}) \qquad \text{and} \qquad f_Y(y) = \frac{1}{\tau} y^{\frac{1}{\tau}-1} f_X(y^{1/\tau})$$

while, if $\boldsymbol{\tau < 0}$,

$$F_Y(y) = 1 - F_X(y^{1/\tau}) \qquad \text{and} \qquad f_Y(y) = \left|\frac{1}{\tau}\right| y^{\frac{1}{\tau}-1} f_X(y^{1/\tau})$$

- **Proof:** if $\tau > 0$ we have

$$F_Y(y) \;=\; P(X \le y^{1/\tau}) = F_X(y^{1/\tau}),$$

while, if $\tau < 0$,

$$F_Y(y) \;=\; P(X \ge y^{1/\tau}) = 1 - F_X(y^{1/\tau})$$

Raising to a power

Exercise 7

We assume that X follows the exponential distribution with mean α, i.e.

$$f_X(x) = \frac{1}{\alpha} e^{-x/\alpha} \qquad x > 0.$$

Consider the transformed variable $Y = X^{\alpha}$. Show that Y follows the Weibull distribution when α is positive and determine the parameters of the Weibull distribution.

Weibull distribution is very flexible to fit reliability data. Looking to the origins of the Weibull distribution, we recognize that the Weibull is a power transformation of the exponential distribution.

Raising to a power

Answer of exercise 7

$$
\begin{aligned}
f_X(x) &= \frac{1}{\alpha} e^{-x/\alpha} \qquad x > 0 \\
f_Y(y) &= \frac{1}{\alpha} y^{1/\alpha - 1} f_X(y^{1/\alpha}) = \frac{1}{\alpha\theta} y^{1/\alpha - 1} e^{-\frac{y^{1/\alpha}}{\theta}} \\
&= (c\tau)(cy)^{\tau - 1} e^{-(cy)^\tau}.
\end{aligned}
$$

Let $\tau = \frac{1}{\alpha}$ and $c^\tau = \frac{1}{\theta}$, then Y follows the Weibull distribution with shape parameter τ and scale parameter c.

Exponentiation

Theorem: exponentiation

Let X be a continuous r.v. with pdf $f_X(x)$ and cdf $F_X(x)$ with $f_X(x) > 0$ for all real x. Let $\boldsymbol{Y = \exp(X)}$. Then, for $y > 0$,

$$F_Y(y) = F_X(\ln y), \qquad \text{and} \qquad f_Y(y) = \frac{1}{y} f_X(\ln y)$$

- **Proof:** we consider

$$F_Y(y) \;=\; P(e^X \leq y) = P(X \leq \ln y) = F_X(\ln y)$$

The pdf follows by differentiation.

Let X have the normal distribution with mean μ and σ^2, then $Y = e^X$ has a lognormal distribution with parameters μ and σ^2. The lognormal random variable has a lower bound of zero, is positively skewed and has a long right tail.
A lognormal distribution is commonly used to describe distributions of financial assets such as stock prices. It is also used in fitting claim amounts for automobile as well as health insurance.

Exponentiation

Exercise 8

X has a uniform distribution on the interval $(0, c)$. $Y = e^X$. Find the distribution of Y.

Exponentiation

Answer of exercise 8

The distribution function of Y is

$$F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \log y) = F_X(\log y)$$

Then

$$f_Y(y) = \frac{1}{y} f_X(\log y) = \frac{1}{cy}$$

Since $0 < x < c$ then $1 < y < e^c$.

Integral transformation

Probability Integral Transformation (PIT)

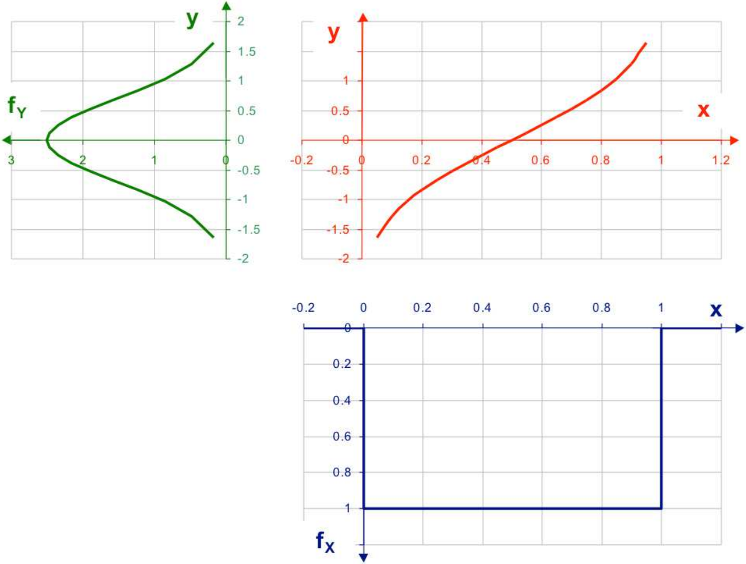
We consider rv X with cdf F, where F is strictly increasing on some interval I, $F = 0$ to the left of I and $F = 1$ to the right of I. F^{-1} is well defined for $x \in I$.

1. Let $Y = F(X)$, then Y has a uniform distribution on $[0, 1]$.
2. Let U be uniform on $[0, 1]$ and let $Z = F^{-1}(U)$. Then the cdf of Z is F.

Proof

1. $P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$.
2. $P(Z \leq z) = P(F^{-1}(U) \leq z) = P(U \leq F(z)) = F(z)$.

Integral transformation



Mixture distributions

- Mixture distributions allow to model data that are drawn from a **heterogeneous population**.
- This parent population can be thought to be divided into **multiple subpopulations with distinct distributions**.
- if underlying phenomenon is diverse and can be described as two phenomena representing **two subpopulations with different modes**, we can construct the 2-point mixture distribution.
- We can generalize this to k-point mixture distributions.
- A mixture with a very large number of subpopulations (k **goes to infinity**) is often referred to as a continuous mixture distribution. Then, subpopulations are not distinguished anymore by a discrete mixing parameter but by a continuous variable.

Two-point mixture

Two-point mixture distribution

Given rvs X_1 and X_2 with pdfs $f_{X_1}(x)$ and $f_{X_2}(x)$ respectively, the pdf of X is the weighted average of $f_{X_1}(x)$ and $f_{X_2}(x)$. For $0 < a < 1$,

$$f_X(x) = af_{X_1}(x) + (1-a)f_{X_2}(x)$$

$$F_X(x) = aF_{X_1}(x) + (1-a)F_{X_2}(x)$$

- The mixing parameters a and $(1-a)$ represent the proportions of data points that fall under each of the two subpopulations respectively.
- This **weighted average** can be applied to a number of other distribution related quantities.

Two-point mixture

- The k-th moment and moment generating function of X are given by

$$\begin{aligned} \mathbb{E}(X^k) &= a\mathbb{E}(X_1^k) + (1-a)\mathbb{E}(X_2^k) \\ M_X(r) &= aM_{X_1}(r) + (1-a)M_{X_2}(r) \end{aligned}$$

Exercise 9

The distribution of the random variable X is an equally weighted mixture of two Poisson distributions with parameters $\lambda_1 v_1$ and $\lambda_2 v_2$ respectively. The mean and variance of X are 4 and 13 respectively. Determine $\mathbb{P}(X > 2)$.

Two-point mixture

Answer of exercise 9

$$
\begin{aligned}
E(X) &= 0.5\lambda_1 v_1 + 0.5\lambda_2 v_2 = 4 \\
E(X^2) &= 0.5\left(\lambda_1 v_1 + (\lambda_1 v_1)^2\right) + 0.5\left(\lambda_2 v_2 + (\lambda_2 v_2)^2\right) = 13 + 16
\end{aligned}
$$

Simplifying both equations gives $\lambda_1 v_1 + \lambda_2 v_2 = 8$ and $(\lambda_1 v_1)^2 + (\lambda_2 v_2)^2 = 50$. Then the parameters of the two Poisson distributions are 1 and 7.

$$P(X > 2) = 0.5P(X_1 > 2) + 0.5P(X_2 > 2) = 0.5235$$

k-point mixture

Let rv X have probability p_i of being drawn from **homogeneous subpopulation** i, where $i = 1, 2, \ldots, k$ and k is the initially specified number of subpopulations in our mixture. The **mixing parameter** p_i represents the proportion of observations from subpopulation i.

k-point mixture distributions

Consider rv X generated from k distinct subpopulations, where subpopulation i is modeled by the continuous distribution $f_{X_i}(x)$, then the pdf of X is given by

$$f_X(x) = \sum_{i=1}^{k} p_i f_{X_i}(x), \quad \text{with } 0 < p_i < 1, \quad \sum_{i=1}^{k} p_i = 1$$

k-point mixture

- The cdf, k-th moment and moment generating function of the k-th point mixture are given as

$$F_X(x) = \sum_{i=1}^{k} p_i F_{X_i}(x)$$

$$\mathbb{E}(X^k) = \sum_{i=1}^{k} p_i \mathbb{E}(X_i^k)$$

$$M_X(r) = \sum_{i=1}^{k} p_i M_{X_i}(r)$$

k-point mixture

Exercise 10

Y_1 is a mixture of X_1 and X_2 with mixing weights a and $(1-a)$.
Y_2 is a mixture of X_3 and X_4 with mixing weights b and $(1-b)$.
Z is a mixture of Y_1 and Y_2 with mixing weights c and $(1-c)$.
Show that Z is a mixture of X_1, X_2, X_3 and X_4, and find the mixing weights.

k-point mixture

Answer of exercise 10

$$
\begin{aligned}
f_{Y_1}(x) &= af_{X_1}(x) + (1-a_1)f_{X_2}(x) \\
f_{Y_2}(x) &= bf_{X_3}(x) + (1-b)f_{X_4}(x) \\
f_Z(x) &= cf_{Y_1}(x) + (1-c)f_{Y_2}(x) \\
f_Z(x) &= c\left[af_{X_1}(x) + (1-a)f_{X_2}(x)\right] + (1-c)\left[bf_{X_3}(x) + (1-b)f_{X_4}(x)\right] \\
&= caf_{X_1}(x) + c(1-a)f_{X_2}(x) + (1-c)bf_{X_3}(x) + (1-c)(1-b)f_{X_4}(x).
\end{aligned}
$$

Then Z is a mixture of X_1, X_2, X_3 and X_4 with mixing weights $ca, c(1-a), (1-c)b$ and $(1-c)(1-b)$.

Continuous mixture

- Extend the previous notion of mixing a finite number of rvs to mixing an uncountable number.
- Continuous mixture: mixture with a very large number of subpopulations (k goes to infinity)
- subpopulations are not distinguished by a discrete mixing parameter but by a **continuous variable** θ (which plays the role of p_i in the finite mixture)
- Consider rv X with a distribution depending on parameter θ, where θ itself is a continuous rv

Continuous mixture

Continuous mixture distributions

Let X have conditional distribution $f_X(x|\theta)$ at a particular value of θ and let $g(\theta)$ be the pdf of the unknown rv θ. The unconditional pdf of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x|\theta)g(\theta)d\theta$$

The pdf $g(\theta)$ is known as the prior distribution of θ (prior information or expert opinion is used in the analysis).

Continuous mixture

- The cdf, k-moment and moment generating function of the continuous mixture are given as

$$F_X(x) = \int_{-\infty}^{\infty} F_X(x|\theta)g(\theta)d\theta$$

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} \mathbb{E}(X^k|\theta)g(\theta)d\theta = \mathbb{E}\left[\mathbb{E}(X^k|\theta)\right]$$

$$M_X(r) = \mathbb{E}(e^{rX}) = \int_{-\infty}^{\infty} \mathbb{E}(e^{tX}|\theta)g(\theta)d\theta$$

- In particular the mean and variance of X are given by

$$\mathbb{E}(X) = \mathbb{E}\left[\mathbb{E}(X|\theta)\right]$$

$$\text{Var}(X) = \mathbb{E}\left[\text{Var}(X|\theta)\right] + \text{Var}\left[\mathbb{E}(X|\theta)\right]$$

Continuous mixture

Exercise 11

1. Claim sizes, X, are uniform on $[\theta, \theta + 10]$ for each policyholder. θ varies by policyholder according to an exponential distribution with mean 5. Find the unconditional distribution, mean and variance of X.

Continuous mixture

Answer of exercise 11

1. The conditional distribution of X is

$$f_X(x|\theta) = \frac{1}{10} \qquad \text{for } \theta \leq x \leq \theta + 10$$

The prior distribution of θ is

$$g(\theta) = \frac{1}{5} e^{-\theta/5} \qquad \text{for } 0 < \theta < \infty$$

The conditional mean and variance of X are given by

$$\begin{aligned} E(X|\theta) &= \frac{\theta + \theta + 10}{2} = \theta + 5 \\ \text{Var}(X|\theta) &= \frac{((\theta + 10) - \theta)^2}{12} = \frac{100}{12} \end{aligned}$$

Continuous mixture

Hence the unconditional mean and variance of X are given by

$$E(X) = E[E(X|\theta)] = E(\theta + 5) = E(\theta) + 5 = 5 + 5 = 10$$

and

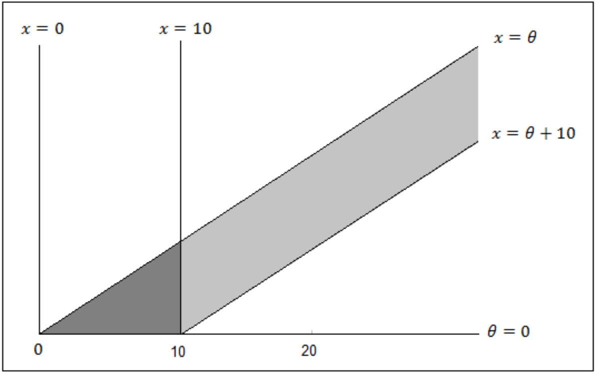
$$\begin{aligned} \mathrm{Var}(X) &= E[\mathrm{Var}(X|\theta)] + \mathrm{Var}(E(X|\theta)) = E\left(\frac{100}{12}\right) + \mathrm{Var}(\theta + 5) \\ &= 8.33 + \mathrm{Var}(\theta) = 33.33 \end{aligned}$$

The unconditional distribution of X is

$$f_X(x) = \int f_X(x|\theta) g(\theta) d\theta$$

$$f_X(x) = \begin{cases} \int_0^x \frac{1}{50} e^{-\frac{\theta}{5}} d\theta = \frac{1}{10}(1 - e^{-x/5}) & 0 \leq x \leq 10 \\ \int_{x-10}^x \frac{1}{50} e^{-\theta/5} d\theta = \frac{1}{10}\left(e^{-(x-10)/5} - e^{-x/5}\right) \end{cases}$$

Continuous mixture



Model selection

- We have presented several claim size distributions and have debated on **which one fits best** the observed data.
- Argumentation was completely **based on graphical tools**.
- When viewing plots it is often difficult to judge **whether any deviation is systematic or merely due to sampling variation**.
- Statistical tests are more methodological tools that consider these questions from a more analytical point of view.
- Two widely used tests to investigate whether given sample $Y_1, \ldots, Y_n$ **fits to particular continuous distribution function** G_0.
  - Kolmogorov-Smirnov (KS) test
  - Anderson-Darling (AS) test
- We also discuss following popular **model comparison techniques**
  - χ^2 goodness of fit test
  - Akaike and Bayesian Information Criteria (AIC and BIC)

Kolmogorov-Smirnov (KS) test

Assume iid sequence $Y_1, Y_2, \ldots$ from unknown G and $\hat{G}_n$ corresponding empirical distribution function of finite sample size n. Non-parametric KS test investigates whether continuous distribution function G_0 fits to given sample $Y_1, Y_2, \ldots, Y_n$.

- Glivenko-Cantelli theorem: empirical distribution function of iid sample converges uniformly to true underlying distribution function $\mathbb{P}$-a.s. $(n \to \infty)$.
- Consider $H_0 : G = G_0$ versus $H_1 : G \neq G_0$
- Define the **KS test statistics**

$$D_n = D_n(Y_1, \ldots, Y_n) = \|\hat{G}_n - G_0\|_\infty = \sup_y |\hat{G}_n(y) - G_0(y)|$$

- $\sqrt{n} D_n \Rightarrow$ Kolmogorov distribution K (as $n \to \infty$)

$$K(y) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j+1} e^{-2j^2y^2} \qquad (y \in \mathbb{R}_+).$$

- H_0 is rejected on significance level $q \in (0,1)$ if

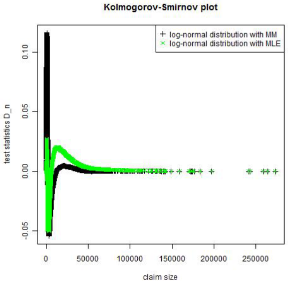
$$D_n > n^{-1/2} K^{\leftarrow}(1-q)$$

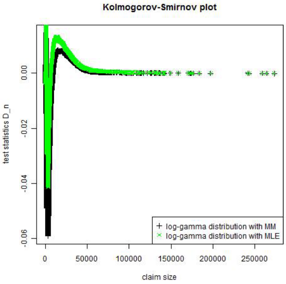
where $K^{\leftarrow}(1-q)$ denotes $(1-q)$-quantile of Kolmogorov distribution K

Kolmogorov-Smirnov (KS) test

q	20%	10%	5%	2%	1%
$K^{\leftarrow}(1-q)$	1.07	1.22	1.36	1.52	1.63

We apply KS test to log-normal and log-gamma fit of PP data.



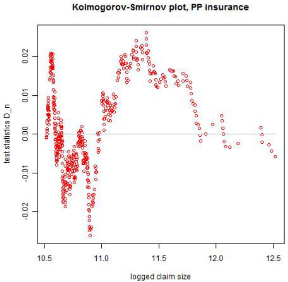


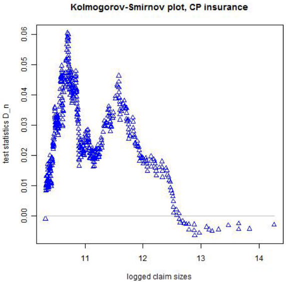
- Log-normal: $D_n = 0.05$ for MLE and $D_n = 0.12$ for MM
- Log-gamma: $D_n = 0.04$ for MLE and $D_n = 0.06$ for MM

Conclusion: We should split claim size modeling into different layers.

Kolmogorov-Smirnov (KS) test

We apply KS test to Pareto fit of the tails (505 largest) of PP and CP data.





- PP: $D_n = 0.027$ for $\alpha = 2.5$
  $\Rightarrow$ sufficiently small so that H_0 cannot be rejected on 5% significance level.
- CP: $D_n = 0.04$ for $\alpha = 1.4$ $\Rightarrow$ resulting p-value is just about 5%.

Anderson-Darling (AD) test

- KS is often **not very powerful** and **not very good in detecting particular properties** such as tail behaviour.
- AD-test is modification of KS-test which gives **more weight to tail** of distributions and is therefore more sensitive in detecting tail fits.
  $\Rightarrow$ **not non-parametric** anymore and calculate critical values for every chosen distribution function.
- **Test statistic**

$$\sup_y |\hat{G}_n(y) - G_0(y)| \sqrt{\psi(G_0(y))}$$

  where $\psi : [0,1] \to \mathbb{R}_+$ is **weight function**.
- Different choices of ψ allow to weight different regions of the support of the distribution function differently.
- In AD test: $\psi(t) = (t(1-t))^{-1}$ to investigate tails.
- In contrast to maximal difference between $\hat{G}_n$ and G_0, we could also consider weighted L^2 distance. This leads to AD modification of Cramer-von Mises test.

χ^2 Goodness-of-fit or Pearson's χ^2 test

- Based on asymptotic normality.
- Splits support of G_0 into K disjoint intervals $I_k = [c_k, c_{k+1})$ and groups data accordingly.
- O_k counts **number of observed realisations** $Y_1, \ldots, Y_n$ in I_k and E_k denotes **expected number of observations** in I_k according to G_0.
- **Test statistic** of n observations

$$X^2_{n,K} = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k}$$

- If d parameters were estimated, then $X^2_{n,K}$ is compared to χ^2_{K-1-d} distribution.
- Rule of thumb: $E_k > 4$

Information criteria

Assume we want to compare different densities g_1 and g_2 that were fitted to $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$

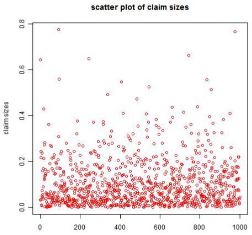
- Akaike and Bayesian Information Criterion

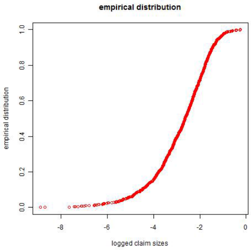
$$\begin{aligned} AIC^{(i)} &= -2\ell_{\boldsymbol{Y}}^{(i)} + 2d^{(i)} \\ BIC^{(i)} &= -2\ell_{\boldsymbol{Y}}^{(i)} + \log(n)d^{(i)} \end{aligned}$$

  where $\ell_{\boldsymbol{Y}}^{(i)}$ is **log-likelihood function** of density g_i for data $\boldsymbol{Y}$ and $d^{(i)}$ denotes **number of estimated parameters** in g_i.
- Model with **smallest** AIC/BIC should be preferred.
- Within framework of MLE methods AIC and BIC are very popular.

Exercise 12

Assume we have iid claim sizes $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ with $n = 1000$ which were generated by a gamma distribution.





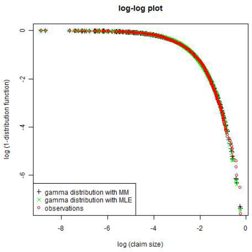
- Sample mean and standard deviation are

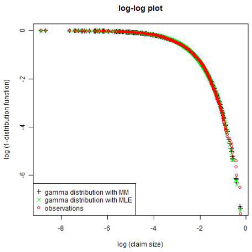
$$\hat{\mu}_n = 0.1039 \quad \text{and} \quad \hat{\sigma}_n = 0.1050$$

- Fitting parameters of gamma distribution yields MM and ML estimators

$$\hat{\gamma}^{MM} = 0.9794; \hat{c}^{MM} = 9.4249 \quad \text{and} \quad \hat{\gamma}^{MLE} = 1.0013; \hat{c}^{MLE} = 9.6360$$

- This provides the fitted distributions





- Fits look perfect and corresponding log-likelihoods are

$$\ell_{\mathbf{Y}}(\hat{\gamma}^{MM}, \hat{c}^{MM}) = 1264.013 \quad \text{and} \quad \ell_{\mathbf{Y}}(\hat{\gamma}^{MLE}, \hat{c}^{MLE}) = 1264.171$$

1. Why is $\ell_{\mathbf{Y}}(\hat{\gamma}^{MLE}, \hat{c}^{MLE}) > \ell_{\mathbf{Y}}(\hat{\gamma}^{MM}, \hat{c}^{MM})$ and which fit should be preferred according to AIC?
2. Estimates of γ are very close to 1 and we could also use an exponential distribution function. For the exponential distribution function we obtain MLE $\hat{c}^{MLE} = 9.6231$ and $\ell_{\mathbf{Y}}(\hat{c}^{MLE}) = 1264.169$. Which model (gamma or exponential) should be preferred according to the AIC and the BIC?

Answer of exercise 12

1. By definition, the MLE looks for values of $\hat{\gamma}$ and $\hat{c}$ that maximizes the likelihood function. Hence any other estimators will provide estimates with likelihood values smaller than or equal to the former. Here we have strict inequality because the MM estimator does not equal the maximizer of the likelihood function (defined differently).

$$\begin{aligned} AIC^{MLE} &= -2\ell_Y^{MLE} + 2d^{MLE} = -2524.342 \qquad (d^{MLE} = 2) \\ &< -2524.026 = -2\ell_Y^{MM} + 2d^{MM} = AIC^{MM} \qquad (d^{MM} = 2) \end{aligned}$$

The MLE have smallest AIC value and should thus be preferred.

2.

$$\begin{aligned} AIC^{gamma} &= -2\ell_Y^{MLE} + 2d^{MLE} = -2524.342 \\ AIC^{exp} &= -2 * 1264.169 + 2 * 1 = -2526.338 \\ BIC^{gamma} &= -2\ell_Y^{MLE} + \log(1000) d^{MLE} = -2514.526 \\ BIC^{exp} &= -2 * 1264.169 + \log(1000)1 = -2521.43 \end{aligned}$$

Hence prefer exponential model on AIC and BIC.

Calculating within layers for claim sizes

- It is difficult to fit one parametric distribution function to entire range of possible outcomes of claim sizes.
  $\Rightarrow$ Consider claim sizes in different layers.
- Re-insurance can also often be bought for different layers.
- Therefore investigate how claim sizes behave in different layers.
- We first discuss the modeling issue and then we describe modeling of re-insurance layers.

Claim size modeling using layers

- KS test rejects most popular parametric examples.
- Assume $Y_1, Y_2, \ldots \overset{iid}{\sim} G$ and we like to split G into different layers.
- Simplest case is two layers: choose large claims threshold $M > 0$ such that $G(M) \in (0, 1)$
- We define disjoint decomposition $\{Y_1 \leq M\}$ and $\{Y_1 > M\}$
- We can model large claims (lc) layer and small claims (sc) layer separately

$$\begin{aligned} G_{sc}(y) &= P[Y_1 \leq y | Y_1 \leq M] \\ G_{lc}(y) &= P[Y_1 \leq y | Y_1 > M] \end{aligned}$$

- We have following decomposition

$$\begin{aligned} G(y) &= P[Y_1 \leq y | Y_1 \leq M] G(M) + P[\leq Y_1 \leq y | Y_1 > M](1 - G(M)) \\ &= G_{sc}(y) G(M) + G_{lc}(y)(1 - G(M)) \end{aligned}$$

Claim size modeling using layers

- Often a successful modeling approach involves 3 steps.

Modeling approach

1. Choose threshold $M > 0$ sufficiently large so that many of the observations fall into lower layer $(0, M]$. In this lower layer one either fits parametric distribution function to the data or one directly works with empirical distribution function (due to Glivenko-Cantelli theorem). If a distribution function is fitted one should ensure that this distribution function has compact support $(0, M]$, for instance, by choosing a truncated gamma distribution.
2. Estimate probability $G(M)$ of the event $\{Y_1 \leq M\}$ which is typically large
3. Fit a Pareto distribution to G_{lc} for threshold $\theta = M$, i.e. estimate tail index $\alpha > 0$ from the observations exceeding this threshold M.

Example

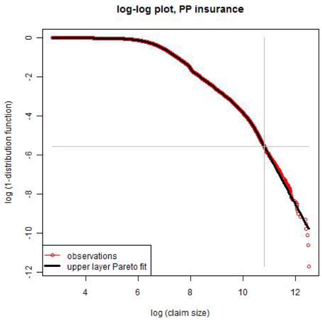
- We revisit PP and CP insurance data set
- We choose large claims threshold $M = 50000$ in both cases
- In PP we have 237 observations above M, which provides $1 - \hat{G}(M) = 237/61053 = 0.39\%$
- For CP we have 272 claims above M, which provides $1 - \hat{G}(M) = 1.87\%$
- For small claims layer $\{Y_i \leq M\}$ we calculate

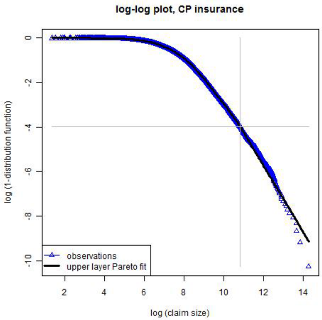
PP	$\hat{\mu}_{\{Y_i \leq M\}} = 2805$	$\hat{Vco}_{\{Y_i \leq M\}} = 1.80$
CP	$\hat{\mu}_{\{Y_i \leq M\}} = 4377$	$\hat{Vco}_{\{Y_i \leq M\}} = 1.51$

- Substantial reduction of sample coefficient of variation in small claims layer compared to entire range of possible outcomes. Not surprising since large claims drive coefficient of variation.
- For CP: sample mean in lower claims layer is substantially reduced. This is due to fact that 1.87% of claims exceed M and these claims may get very large and drive mean.

Example

- Finally we fit G to data. We choose empirical distribution functions below M and Pareto distributions for the tail fit in the large claims layer, having tail parameters α as estimated before.





- For PP they look convincing whereas CP is not entirely satisfactory in large claims layer (maybe bigger large claims threshold M and a slightly bigger tail parameter α).

Re-insurance layers and deductibles

- We have calculated expected values in claims layers $E[Y 1_{\{u_1 < Y \leq u_2\}}]$ for various parametric distribution functions. This is of interest for several reasons
- First reason is that insurance contracts often have deductibles. On the one hand small claims often cause too much administrative costs and on the other hand deductibles are also an instrument to prevent from fraud.
- For instance, it can become quite expensive for an insurance company if every insured claims that his umbrella got stolen. Therefore a deductible $d > 0$ of say 200 euro is introduced and insurance company only covers claim $(Y - d)_+$ that exceeds this deductible d.

$$(Y - d)_+ = \begin{cases} 0 & Y \leq d \\ Y - d & Y > d \end{cases}$$

$(Y - d)_+$ is often referred to as left censored and shifted variable since the values below d are ignored and all losses are shifted by a value d.

Re-insurance layers and deductibles

- In this case pure risk premium for claim $Y \sim G$ is given by

$$\begin{aligned} E[(Y-d)_+] &= \int_d^\infty (y-d)dG(y) = E[Y1_{\{Y>d\}}] - dP[Y>d] \\ &= P[Y>d](E[Y|Y>d]-d) = P[Y>d]e(d) \end{aligned}$$

  under assumption that $P[Y>d]>0$ and that mean excess function $e(.)$ of Y exists.
- Second reason is that insurance company may have maximal insurance cover per claim, i.e. it covers claims only up to maximal size of $M>0$ and exceedances need to be paid by the insured; or similarly, it may cover claims exceeding M but has a re-insurance cover for these exceedances.

$$Y \wedge M = \begin{cases} Y & Y < M \\ M & Y \geq M \end{cases}$$

Re-insurance layers and deductibles

- In that case, insurance company covers $(Y \wedge M)$ and pure risk premium for this (bounded) claim is given by

$$\begin{aligned} E[Y \wedge M] &= \int_0^M y dG(y) + MP[Y > M] = E[Y1_{\{Y \leq M\}}] + MP[Y > M] \\ &= E[Y] - \left(E[Y1_{\{Y > M\}}] - MP[Y > M]\right) \\ &= E[Y] - P[Y > M]e(M) = E[Y] - E[(Y - M)_+] \end{aligned}$$

- If we combine deductibles with maximal covers we obtain excess-of-loss (XL) (re-)insurance treaties.
- Assume we have deductible $u_1 > 0$ (in re-insurance terminology this also called **priority** or **retention**).
- Insurance treaty "$u_2 XL u_1$" covers claims layer $(u_1, u_1 + u_2]$ that is, this contract covers maximal excess of u_2 above priority u_1.

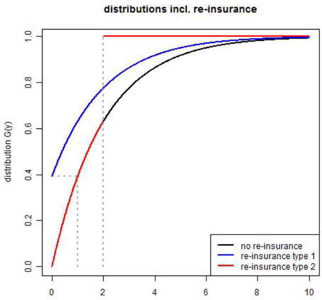
Re-insurance layers and deductibles

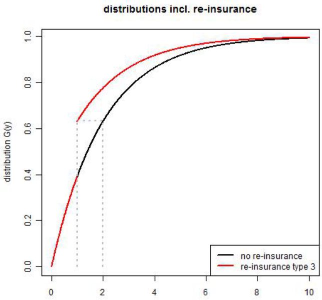
- Pure risk premium for such contracts is then given by

$$E[((Y - u_1)_+) \wedge u_2] = E[(Y - u_1)_+] - E[(Y - u_1 - u_2)_+]$$

Exercise 13

We display the distribution function of loss $Y \sim G$ and the distribution function of the loss after applying different re-insurance covers to Y. Can you explicitly determine the re-insurance covers from the graphs?





Re-insurance layers and deductibles

Answer of exercise 13

1. Reinsurance type 1: $\tilde{Y} = (Y - d)_+$ is the new loss variable. a deductible of $d = 1$ is used. We observe that

$$P(\tilde{Y} \leq y) = P(Y \leq y + d) \qquad \forall y > 0$$

2. Reinsurance type 2: $\tilde{Y} = Y \wedge M$ with $M = 2$ is the new loss variable. We see that all probability mass of claim sizes larger than $M = 2$ moves to the point $y = M$. For values $y < M$, the distribution function remains unchanged. Hence a maximal insurance cover of $M = 2$ per claim is introduced.
3. Reinsurance type 3: $\tilde{Y} = Y 1_{Y<1} + (Y - 2)_+ 1_{Y \geq 1}$ is the new loss variable. A deductible of $d = 2$ is used for all claims larger than $y = 1$. We observe that all probability mass between $y = 1$ and $y = 2$ is moved to $y = 1$. Claims with size $y < 1$ are treated normally.

Outline

Compound binomial model

The total claim amount S has a compound binomial distribution, write

$$S \sim \mathrm{CompBinom}(v, p, G)$$

if S has a compound distribution with $N \sim \mathrm{Binom}(v, p)$ for given $v \in \mathbb{N}$ and $p \in (0, 1)$ and individual claim size distribution G.

Proposition

Assume $S \sim \mathrm{CompBinom}(v, p, G)$. We have

$$\begin{aligned} E[S] &= vpE[Y_1] \\ \mathrm{Var}(S) &= vp\left(E[Y_1^2] - pE[Y_1]^2\right) \\ \mathrm{Vco}(S) &= \sqrt{\frac{1}{vp}}\sqrt{1 - p + \mathrm{Vco}(Y_1)^2} \\ M_S(r) &= (pM_{Y_1}(r) + (1-p))^v \qquad r \in \mathbb{R} \end{aligned}$$

whenever they exist.

Compound binomial distribution

Exercise 14

Consider a collection of 5000 policies each of which has probability 0.002 of giving rise to a claim in a given year. Assume all policies are for a fixed amount Y_i of $K = 400$. Calculate mean, variance and skewness of this collection.

Compound binomial distribution

Answer of exercise 14

Since $E(Y_1^n) = (400)^n$ we obtain

$$\begin{aligned} E(S) &= vpE(Y_1) = 4000 \\ \text{Var}(S) &= vp\left(E[Y_1^2] - pE[Y_1]^2\right) = 1596800 \end{aligned}$$

For skewness we calculate cgf

$$\log M_S(r) = v \log(pM_{Y_1}(r) + q)$$

and find 3rd derivative wrt r and evaluate at $r = 0$

$$\mu_3 = vpE[Y_1^3] - 3vp^2 E[Y_1^2]E[Y_1] + 2vp^3(E[Y_1])^3$$

Hence

$$\begin{aligned} \zeta_S &= \left[\frac{vpE[Y_1^3] - 3vp^2 E[Y_1^2]E[Y_1] + 2vp^3(E[Y_1])^3}{[vp(E[Y_1^2] - pE[Y_1])]^{3/2}}\right] \\ &= \dots \\ &= 0.31527 \end{aligned}$$

Corollary - Aggregation property

Assume that $S_1, \ldots, S_n$ are independent with $S_j \sim \text{CompBinom}(v_j, p, G)$ for all $j = 1, \ldots, n$. The aggregated claim has a compound binomial distribution with

$$S = \sum_{j=1}^{n} S_j \sim \text{CompBinom}\left(\sum_{j=1}^{n} v_j, p, G\right)$$

Proof:

$$\begin{aligned} M_S(r) &= \prod_{i=1}^{n} M_{S_i}(r) \\ &= \prod_{i=1}^{n} (pM_{Y_1}(r) + (1-p))^{v_i} \\ &= (pM_{Y_1}(r) + (1-p))^{\sum_{i=1}^{n} v_i} \end{aligned}$$

Exercise 15

Assume $S \sim \text{CompBinom}(v, p, G)$ and choose $M > 0$ such that $G(M) \in (0, 1)$. Define the compound distribution of claims Y_i exceeding threshold M by

$$S_{lc} = \sum_{i=1}^{N} Y_i 1_{\{Y_i > M\}}$$

Then we have $S_{lc} \sim \text{CompBinom}(v, p(1 - G(M)), G_{lc})$ where the large claims size distribution satisfies $G_{lc}(y) = \mathbb{P}[Y_1 \leq y | Y_1 > M]$.

Answer of exercise 15

$$
\begin{aligned}
M_{S_{lc}}(r) &= E\left[e^{rS_{lc}}\right] \\
&= E\left[e^{r\sum_{i=1}^{N} Y_i 1_{Y_i>M}}\right] \\
&= E\left[\prod_{i=1}^{N} e^{rY_i 1_{Y_i>M}}\right] \\
&= E\left[E\left[\prod_{i=1}^{N} e^{rY_i 1_{Y_i>M}} \middle| N\right]\right] \\
&= E\left[\prod_{i=1}^{N} E\left[e^{rY_i 1_{Y_i>M}} \middle| N\right]\right] \\
&= E\left[\prod_{i=1}^{N} E\left[e^{rY_i 1_{Y_i>M}}\right]\right]
\end{aligned}
$$

We solve inner part

$$\begin{aligned} E\left[e^{rY_i 1_{Y_i > M}}\right] &= E\left[e^{rY_i 1_{Y_i > M}} | Y_i > M\right] P(Y_i > M) \\ &\quad + E\left[e^{rY_i 1_{Y_i > M}} | Y_i \leq M\right] P(Y_i \leq M) \\ &= E\left[e^{rY_i} | Y_i > M\right] P(Y_i > M) + E\left[e^{0} | Y_i \leq M\right] P(Y_i \leq M) \\ &= M_{\tilde{Y}_i}(r)(1 - G(M)) + G(M) \end{aligned}$$

with $\tilde{Y}_i \sim G_{lc}$.

We can then further calculate (where Y_i are iid).

$$
\begin{aligned}
M_{S_{lc}}(r) &= E\left[\prod_{i=1}^{N} E\left[e^{rY_i 1_{Y_i>M}}\right]\right] \\
&= E\left[\prod_{i=1}^{N}\left(M_{\tilde{Y}_1}(r)(1-G(M))+G(M)\right)\right] \\
&= E\left[\left(M_{\tilde{Y}_1}(r)(1-G(M))+G(M)\right)^N\right] \\
&= E\left[\exp\left[\log\left(M_{\tilde{Y}_1}(r)(1-G(M))+G(M)\right)^N\right]\right] \\
&= E\left[\exp\left[N\log\left(M_{\tilde{Y}_1}(r)(1-G(M))+G(M)\right)\right]\right] \\
&= M_N\left(\log\left[M_{\tilde{Y}_1}(r)(1-G(M))+G(M)\right]\right) \\
&= \left[pM_{\tilde{Y}_1}(r)(1-G(M))+pG(M)+(1-p)\right]^v \\
&= \left[p(1-G(M))M_{\tilde{Y}_1}(r)+1-p(1-G(M))\right]^v
\end{aligned}
$$

Compound Poisson model

Total claim amount S has compound Poisson distribution

$$S \sim \text{CompPoi}(\lambda v, G)$$

if S has compound distribution with $N \sim \text{Poi}(\lambda v)$ for given $\lambda, v > 0$ and individual claim size distribution G.

Proposition

Assume $S \sim \text{CompPoi}(\lambda v, G)$. We have

$$\begin{aligned}
E[S] &= \lambda v E[Y_1] \\
\text{Var}(S) &= \lambda v E[Y_1^2] \\
\text{Vco}(S) &= \sqrt{\frac{1}{\lambda v}} \sqrt{1 + \text{Vco}(Y_1)^2} \\
M_S(r) &= e^{\lambda v (M_{Y_1}(r) - 1)} \qquad \text{for } r \in \mathbb{R}
\end{aligned}$$

whenever they exist.

Compound Poisson distribution

A very useful property of compound Poisson distribution is that it is preserved under convolutions. Given that one often wants to bring together claims from different portfolios or companies, this can be useful in studying the distribution of the aggregate claims from different risks.

Theorem - Aggregation of compound Poisson distributions

Assume $S_1, \ldots, S_n$ are independent with $S_j \sim \text{CompPoi}(\lambda_j v_j, G_j)$ $\forall j = 1, \ldots, n$. The aggregated claim has compound Poisson distribution

$$S = \sum_{j=1}^{n} S_j \sim \text{CompPoi}(\lambda v, G)$$

with

$$v = \sum_{j=1}^{n} v_j \qquad \lambda = \sum_{j=1}^{n} \frac{v_j}{v} \lambda_j \qquad G = \sum_{j=1}^{n} \frac{\lambda_j v_j}{\lambda v} G_j.$$

Compound Poisson distribution

Proof

Since $G_j(0) = 0$ for all $j = 1, \ldots, n$ it holds that $S \geq 0$ $\mathbb{P}$-a.s.

$$\begin{aligned} M_S(r) &= E\left[e^{r\sum_{j=1}^n S_j}\right] = E\left[\prod_{j=1}^n e^{rS_j}\right] = \prod_{j=1}^n E\left[e^{rS_j}\right] \\ &= \prod_{j=1}^n e^{\lambda_j v_j \left(M_{Y_1^{(j)}}(r) - 1\right)} = e^{\lambda v \left(\sum_{j=1}^n \frac{\lambda_j v_j}{\lambda v} M_{Y_1^{(j)}}(r) - 1\right)} \end{aligned}$$

where we assumed $Y_1^{(j)} \sim G_j$.
$\Rightarrow$ Compound Poisson distribution with expected number of claims λv and claim size distribution G obtained from the mgf $\sum_{j=1}^n \frac{\lambda_j v_j}{\lambda v} M_{Y_1^{(j)}}(r)$.

Compound Poisson distribution

Note that $G = \sum_{j=1}^{n} \frac{\lambda_j v_j}{\lambda v} G_j$ is distribution function (non-decreasing, right-continuous, $\lim_{x \to -\infty} G(x) = 0$ and $\lim_{x \to \infty} G(x) = 1$).
Choose $Y \sim G$ and then

$$\begin{aligned} M_Y(r) &= \int_0^\infty e^{ry} dG(y) = \int_0^\infty e^{ry} d\left(\sum_{j=1}^{n} \frac{\lambda_j v_j}{\lambda v} G_j(y)\right) \\ &= \sum_{j=1}^{n} \frac{\lambda_j v_j}{\lambda v} \int_0^\infty e^{ry} dG_j(y) = \sum_{j=1}^{n} \frac{\lambda_j v_j}{\lambda v} M_{Y_1^{(j)}}(r) \end{aligned}$$

Since mgf determines distribution uniquely, this proves theorem.
□

Compound Poisson distribution

Exercise 16

Claims in a company are grouped into two portfolios and modeled by compound Poisson distributions. Those in portfolio 1 are modeled by a compound Poisson distribution with parameter $\lambda_1 v_1 = 3$ per month and claims are exponentially distributed with mean 500. The rate parameter for those in portfolio 2 is $\lambda_2 v_2 = 7$ per month and claims are exponentially distributed with mean 300. Calculate expectation, variance and mgf of S which is total **annual** claims in both portfolios.

Compound Poisson distribution

Answer of exercise 16

Total annual claims S in both portfolios are modeled by compound Poisson distribution with parameters $\lambda v = 12(3+7)$ per year and claim distribution G is a 30% : 70% mixture of exponential distributions with mean 500 and 300 respectively. Hence mgf of S is given by

$$M_S(r) = e^{120[M_Y(r)-1]}$$

where

$$M_Y(r) = 0.3\frac{1}{1-500r} + 0.7\frac{1}{1-300r}$$

Therefore

$$\begin{aligned} E[S] &= \lambda v E[Y] = 120[(0.3)500 + (0.7)300] = 432000 \\ \mathrm{Var}[S] &= \lambda v E[Y^2] = 120[(0.3)2(500)^2 + (0.7)2(300)^2] = 33120000 \end{aligned}$$

Compound Poisson distribution

We now slightly extend compound Poisson model.

- Let $(p_j^+)_{j=1,\ldots,m}$ be discrete probability distribution on finite set $\{1,\ldots,m\}$. Assume $p_j^+ > 0$ for all j.
- Assume G_j corresponding claim size distributions of the sub-portfolios with $G_j(0) = 0$.
- Define mixture distribution

$$G(y) = \sum_{j=1}^{m} p_j^+ G_j(y) \qquad \text{for } y \in \mathbb{R}$$

  Former theorem exactly provides such a mixture distribution with $p_j^+ = \frac{\lambda_j v_j}{\lambda v}$ if we aggregate sub-portfolios.

- Define discrete random variable I which indicates to which sub-portfolio particular claim Y belongs

$$\mathbb{P}[I = j] = p_j^+ \qquad \text{for all } j \in \{1,\ldots,m\}$$

Compound Poisson distribution

Definition - Extended compound poisson model

The total claim amount $S = \sum_{i=1}^{N} Y_i$ has compound Poisson distribution as defined before. In addition, we assume that $(Y_i, I_i)_{i \geq 1}$ are iid and independent of N with Y_i having marginal distribution function G with $G(0) = 0$ and I_i having marginal distribution function given before.

(Y_1, I_1) takes values in $\mathbb{R}_+ \times \{1, \ldots, m\}$ and let $A_1, \ldots, A_n$ be a measurable disjoint decomposition of $\mathbb{R}_+ \times \{1, \ldots, m\}$, i.e.

- $A_k \cap A_l = \emptyset$ for all $k \neq l$
- $\cup_{i=1}^{n} A_k = \mathbb{R}_+ \times \{1, \ldots, m\}$

This measurable disjoint decomposition is called admissible for (Y_1, I_1) if for all $k = 1, \ldots, n$

$$p^{(k)} = \mathbb{P}\left[(Y_1, I_1) \in A_k\right] > 0.$$

Note that $\sum_{k=1}^{n} p^{(k)} = 1$.

Compound Poisson distribution

Theorem - Disjoint decomposition property

Assume that S fulfils extended compound Poisson model assumptions. We choose an admissible, measurable disjoint decomposition $A_1, \ldots, A_n$ for (Y_1, I_1). Define for $k = 1, \ldots, n$ the random variables

$$S_k = \sum_{i=1}^{N} Y_i 1_{\{(Y_i, I_i) \in A_k\}}$$

S_k are independent and $\text{CompPoi}(\lambda_k v_k, G_k)$ distributed for $k = 1, \ldots, n$ with

$$\lambda_k v_k = \lambda v p^{(k)} > 0 \text{ and } G_k(y) = \mathbb{P}[Y_1 \leq y | (Y_1, I_1) \in A_k]$$

Compound Poisson distribution

Proof

We prove theorem using multivariate version of mgf. Choose $\boldsymbol{r} = (r_1, \ldots, r_n)' \in \mathbb{R}^n$. Multivariate mgf of $\boldsymbol{S} = (S_1, \ldots, S_n)'$ is given by

$$\begin{aligned}
M_{\boldsymbol{S}}(\boldsymbol{r}) &= E[e^{\boldsymbol{r}'\boldsymbol{S}}] = E\left[e^{\sum_{k=1}^n r_k S_k}\right] \\
&= E\left[e^{\sum_{k=1}^n r_k \sum_{i=1}^N Y_i 1_{\{(Y_i, I_i) \in A_k\}}}\right] \\
&= E\left[\prod_{i=1}^N E\left[e^{\sum_{k=1}^n r_k Y_i 1_{\{(Y_i, I_i) \in A_k\}}} \middle| N\right]\right] \\
&= E\left[\prod_{i=1}^N E\left[e^{\sum_{k=1}^n r_k Y_i 1_{\{(Y_i, I_i) \in A_k\}}}\right]\right]
\end{aligned}$$

Compound Poisson distribution

We calculate inner expected values

$$
\begin{aligned}
E\left[e^{\sum_{k=1}^{n} r_k Y_i 1_{\{(Y_i,I_i)\in A_k\}}}\right] &= \sum_{\ell=1}^{n} E\left[e^{\sum_{k=1}^{n} r_k Y_i 1_{\{(Y_i,I_i)\in A_k\}}} 1_{\{(Y_i,I_i)\in A_\ell\}}\right] \\
&= \sum_{\ell=1}^{n} E\left[e^{\sum_{k=1}^{n} r_k Y_i 1_{\{(Y_i,I_i)\in A_k\}}} \middle| (Y_i, I_i) \in A_\ell\right] \mathbb{P}\left[(Y_i, I_i) \in A_\ell\right] \\
&= \sum_{\ell=1}^{n} E\left[e^{r_\ell Y_i} | (Y_i, I_i) \in A_\ell\right] p^{(\ell)} \\
&= \sum_{\ell=1}^{n} p^{(\ell)} M_{Y_1^{(\ell)}}(r_\ell)
\end{aligned}
$$

where we assume $Y_1^{(\ell)} \sim G_\ell$.

Collecting all items we obtain

$$
\begin{aligned}
M_{\boldsymbol{S}}(\boldsymbol{r}) &= E\left[\left(\sum_{\ell=1}^{n} p^{(\ell)} M_{Y_1^{(\ell)}}(r_\ell)\right)^N\right] \\
&= E\left[e^{N \log\left(\sum_{\ell=1}^{n} p^{(\ell)} M_{Y_1^{(\ell)}}(r_\ell)\right)}\right] \\
&= e^{\lambda v\left(\sum_{\ell=1}^{n} p^{(\ell)} M_{Y_1^{(\ell)}}(r_\ell)-1\right)} \\
&= e^{\lambda v \sum_{\ell=1}^{n} p^{(\ell)}\left(M_{Y_1^{(\ell)}}(r_\ell)-1\right)} \\
&= \prod_{\ell=1}^{n} e^{\lambda v p^{(\ell)}\left(M_{Y_1^{(\ell)}}(r_\ell)-1\right)} \\
&= \prod_{\ell=1}^{n} M_{S_\ell}(r_\ell)
\end{aligned}
$$

This proves theorem because we have obtained product (i.e. independence) of mgf's of compound Poisson distributed rv's S_ℓ with $\ell = 1, \ldots, n$.
$\square$

Remarks

- Aggregation property implies that we can follow **bottom-up modeling** approach for entire insurance business. Thus, we model each sub-portfolio S_j independently with compound Poisson distribution. Total portfolio is then easily obtained by aggregation theorem and we stay in same family of distributions.
- Disjoint decomposition property implies that we can also **follow top-down modeling** approach. Thus we model overall portfolio by compound Poisson distribution and by disjoint decomposition property we can easily allocate total claim to subportfolios. Crucial result here is that this allocation results in independent compound poisson distributions for S_j.
- For I we have chosen finite (discrete) indicator. Of course, this model can easily be extended to other indicators. Crucial property is the iid assumption on random vectors (Y_i, I_i).

Remarks

Choice of appropriate volume on sub-portfolios depends on choice of indicator I.

- If $m = 1$ (we only consider 1 portfolio), but we apply disjoint decomposition as follows
$$Y_i = Y_i 1_{\{Y_i \in A_1\}} + \ldots + Y_i 1_{\{Y_i \in A_n\}}$$
then it is natural to set $v_k = v$ and $\lambda_k = \lambda p^{(k)}$ for $k = 1, \ldots, n$.
$\Rightarrow$ volume $v > 0$ remains constant but expected claims frequencies λ_k change accordingly to A_k.
$\Rightarrow$ thinning of Poisson point process.
- Second extreme case is $m = n > 1$ and disjoint decomposition is given by
$$\{(Y_i, I_i) \in A_k\} = \{I_i = k\}$$
i.e. we only consider decomposition according to different sub-portfolios $k = 1, \ldots, m$. In this case we would rather define $v_k > 0$ by volume of portfolio k and $\lambda_k = \lambda p^{(k)} v / v_k$.

Example: large claims separation

Important application: separation of large from small claims.

- Individual claim sizes are **divided into different layers**:
  - large claims threshold $M > 0$ such that $G(M) \in (0, 1)$
- Define disjoint decomposition A_1, A_2 of $\mathbb{R}_+$ by

$$A = A_1 = \{Y_1 \leq M\} \quad \text{and} \quad A^c = A_2 = \{Y_1 > M\}.$$

- Assume $S \sim \text{CompPoi}(\lambda v, G)$ and define

$$S_{sc} = \sum_{i=1}^{N} Y_i 1_{\{Y_i \leq M\}} \quad \text{and} \quad S_{lc} = \sum_{i=1}^{N} Y_i 1_{\{Y_i > M\}}$$

  Theorem implies that S_{sc} and S_{lc} independent and compound Poisson

$$\begin{aligned} S_{sc} &\sim \text{CompPoi}(\lambda_{sc} v = \lambda G(M) v); \qquad G_{sc}(y) = \mathbb{P}[Y_1 \leq y | Y_1 \leq M] \\ S_{lc} &\sim \text{CompPoi}(\lambda_{lc} v = \lambda(1 - G(M)) v); \qquad G_{lc}(y) = \mathbb{P}[Y_1 \leq y | Y_1 \geq M] \end{aligned}$$

- **Model small and large claims layers separately** and obtain total claim distribution using convolution.

Compound negative-binomial model

Total claim amount S has compound Negative-binomial distribution

$$S \sim \text{CompNB}(\lambda v, \gamma, G)$$

if S has compound distribution with $N \sim \text{NegBin}(\lambda v, \gamma)$ for given $\lambda, v, \gamma > 0$ and individual claim size distribution G.

Proposition

Assume $S \sim \text{CompNB}(\lambda v, \gamma, G)$. We have, whenever they exist

$$\begin{aligned}
E[S] &= \lambda v E[Y_1] \\
\text{Var}(S) &= \lambda v E[Y_1^2] + (\lambda v)^2 E[Y_1]^2/\gamma \\
\text{Vco}(S) &= \sqrt{\frac{1}{\lambda v}}\sqrt{1 + \text{Vco}(Y_1)^2 + \lambda v/\gamma} > \gamma^{-1/2} \\
M_S(r) &= \left(\frac{1-p}{1-pM_{Y_1}(r)}\right)^{\gamma} \quad \text{for } r \in \mathbb{R} \text{ such that } M_{Y_1}(r) < 1/p
\end{aligned}$$

with $p = (\lambda v)/(\gamma + \lambda v) \in (0, 1)$.

Compound negative-binomial model

Exercise 17

Consider compound negative-binomial model for aggregate claims where N follows negative-binomial distribution with $p = 0.02$ and $\gamma = 800$ and typical claim Y_i is exponential with mean 400. Compute expectation and variance of total claim amount S.

Compound negative-binomial model

Answer of exercise 17

First three moments of Y_1 are given by $E[Y_1] = 400, E[Y_1^2] = 2(400)^2$ and $E[Y_1^3] = 6(400)^3$. Furthermore, $\lambda v = \frac{p\gamma}{1-p} = 16.32653$. Hence

$$
\begin{aligned}
E(S) &= \lambda v E[Y_1] = 6530.612 \\
\text{Var}(S) &= \lambda v E[Y_1^2] + (\lambda v)^2 E[Y_1]^2/\gamma = 5277801
\end{aligned}
$$

Compound negative-binomial model

Exercise 18

Assume $S \sim \text{CompNB}(\lambda v, \gamma, G)$ and choose $M > 0$ such that $G(M) \in (0, 1)$. Define compound distribution of claims Y_i exceeding threshold M by

$$S_{lc} = \sum_{i=1}^{N} Y_i 1_{\{Y_i > M\}}$$

Then we have $S_{lc} \sim \text{CompNB}(\lambda(1 - G(M))v, \gamma, G_{lc})$ where large claims size distribution satisfies $G_{lc}(y) = \mathbb{P}[Y_1 \leq y | Y_1 > M]$.

Outline

Parameter estimation

Once we have specified distribution functions for N and Y_i we still need to **determine their parameters**. In case of claims count distribution for N

- default probability p for binomial distribution
- expected claims frequency λ for Poisson distribution
- expected claims frequency λ and dispersion parameter γ for negative-binomial distribution

We discuss **two popular ways** to estimate parameters

1. Method of moments (MM)
2. Maximum Likelihood Estimation (MLE) method
3. Bayesian inference method (inverse probability method).

Note that we have a vector of observations $\boldsymbol{N} = (N_1, \ldots, N_T)'$ where N_t denotes number of claims in accounting year t. Difficulty is that N_t, $t = 1, \ldots, T$ are not iid because they depend on different volumes $v_t \Rightarrow$ portfolio changes over accounting years.

Method of moments (in general)

We start with simple example (not suitable for our situation)

- Assume $X_1, \ldots, X_T \overset{iid}{\sim} F$ where F is parametric distribution function that depends on two dimensional (real valued) parameter $(\vartheta_1, \vartheta_2)$. Assume first two moments of X_1 are **finite**, and thus, for all $t = 1, \ldots, T$

$$\mu = \mu(\vartheta_1, \vartheta_2) = E[X_t] < \infty \qquad \sigma^2 = \sigma^2(\vartheta_1, \vartheta_2) = \mathrm{Var}(X_t) < \infty$$

  For general d-dimensional parameters we extend argument to first d moments.
- We define sample mean and sample variance by

$$\hat{\mu}_T = \frac{1}{T}\sum_{t=1}^{T} X_t \qquad \hat{\sigma}_T^2 = \frac{1}{T-1}\sum_{t=1}^{T}(X_t - \hat{\mu}_T)^2$$

  $\Rightarrow$ unbiased estimators for μ and σ^2
- Moment estimator $(\hat{\vartheta}_1, \hat{\vartheta}_2)$ for $(\vartheta_1, \vartheta_2)$ solves system of equations

$$\hat{\mu}_T = \mu(\vartheta_1, \vartheta_2) \qquad \hat{\sigma}_T^2 = \sigma^2(\vartheta_1, \vartheta_2)$$

We need to slightly modify this framework.

Method of moments (for our situation)

Assumption

Assume that there exist strictly positive volumes $v_1, \ldots, v_T$ such that the components of $\boldsymbol{F} = (N_1/v_1, \ldots, N_T/v_T)'$ are **independent** with

$$\lambda = E[N_t/v_t] \qquad \text{and} \qquad \tau_t^2 = \text{Var}(N_t/v_t) \in (0, \infty)$$

Lemma 3

(assumption above holds) Unbiased linear (in $\boldsymbol{F}$) estimator for λ with minimal variance is given by

$$\hat{\lambda}_T^{MV} = \left(\sum_{t=1}^{T} \frac{1}{\tau_t^2} \right)^{-1} \sum_{t=1}^{T} \frac{N_t/v_t}{\tau_t^2}$$

Variance of this estimator is given by

$$Var(\hat{\lambda}_T^{MV}) = \left(\sum_{t=1}^{T} \frac{1}{\tau_t^2} \right)^{-1}$$

Proof uses method of Lagrange multipliers (see course notes).

Method of moments: binomial and Poisson distribution

We apply lemma 3 to case of binomial and Poisson distributions

- $N_t, t = 1, \ldots, T$ are independent with $N_t \sim \text{Binom}(v_t, p)$ or $Poi(\lambda v_t)$.
  - In binomial case
$$E[N_t/v_t] = p \qquad \text{and} \qquad \text{Var}(N_t/v_t) = p(1-p)/v_t = \tau_t^2$$
  - In Poisson case
$$E[N_t/v_t] = \lambda \qquad \text{and} \qquad \text{Var}(N_t/v_t) = \lambda/v_t = \tau_t^2$$
- Unknown parameter p and λ respectively appears in variance.
- Appearance is of multiplicative type which implies that it cancels in weights $w_t = \tau_t^{-2} \left(\sum_{s=1}^{T} \tau_s^{-2} \right)^{-1}$.

Moment estimator

- In binomial case

$$\hat{p}_T^{MV} = \frac{1}{\sum_{s=1}^T v_s} \sum_{t=1}^T N_t = \sum_{t=1}^T \frac{v_t}{\sum_{s=1}^T v_s} \frac{N_t}{v_t}$$

- In Poisson case

$$\hat{\lambda}_T^{MV} = \frac{1}{\sum_{s=1}^T v_s} \sum_{t=1}^T N_t = \sum_{t=1}^T \frac{v_t}{\sum_{s=1}^T v_s} \frac{N_t}{v_t}$$

Variances are given by

$$\mathrm{Var}(\hat{p}_T^{MV}) = \frac{p(1-p)}{\sum_{s=1}^T v_s} \qquad \text{and} \qquad \mathrm{Var}(\hat{\lambda}_T^{MV}) = \frac{\lambda}{\sum_{s=1}^T v_s}$$

We can explicitly give distributions of the estimators because

$$\sum_{t=1}^T N_t \sim \mathrm{Binom}(\sum_{t=1}^T v_t, p) \qquad \text{and} \qquad \sum_{t=1}^T N_t \sim Poi(\lambda \sum_{t=1}^T v_t)$$

Method of moments: negative-binomial distribution

Negative-binomial is more complex

- Assume that $N_t, t = 1, \ldots, T$ are independent with $N_t \sim \text{NegBin}(\lambda v_t, \gamma)$.
- For first two moments we have

$$E[N_t/v_t] = \lambda \qquad \text{and} \qquad \text{Var}(N_t/v_t) = \lambda/v_t + \lambda^2/\gamma = \tau_t^2$$

- Variance term has two unknown parameters λ and γ and we lose nice multiplicative structure from binomial and Poisson case which has allowed to apply lemma 3.
- If we drop condition minimal variance we obtain following unbiased linear estimator

Method of moments: negative-binomial distribution

Moment estimator in negative-binomial case

We have following unbiased linear estimator for λ

$$\hat{\lambda}_T^{NB} = \frac{1}{\sum_{s=1^T} v_s} \sum_{t=1}^{T} N_t = \sum_{t=1}^{T} \frac{v_t}{\sum_{s=1}^{T} v_s} \frac{N_t}{v_t}$$

In last formula we could also take other volume weighted averages (respecting unbiasedness). Variance of estimator is given by

$$\mathrm{Var}\left(\hat{\lambda}_T^{NB}\right) = \left(\frac{1}{\sum_{s=1}^{T} v_s}\right)^2 \sum_{t=1}^{T} \mathrm{Var}(N_t) = \frac{\sum_{t=1}^{T} \lambda v_t + (\lambda v_t)^2/\gamma}{\left(\sum_{s=1}^{T} v_s\right)^2}$$

There remains estimate of γ. Therefore we define

$$\hat{V}_T^2 = \frac{1}{T-1} \sum_{t=1}^{T} v_t \left(\frac{N_t}{v_t} - \hat{\lambda}_T^{NB}\right)^2$$

Method of moments: negative-binomial distribution

Lemma 4 (Proof in course notes)

In negative-binomial model $\hat{V}_T^2$ satisfies

$$E\left[\hat{V}_T^2\right] = \lambda + \frac{\lambda^2}{T-1}\left(\sum_{t=1}^{T} v_t - \frac{\sum_{t=1}^{T} v_t^2}{\sum_{t=1}^{T} v_t}\right)\frac{1}{\gamma}$$

This motivates following estimator

Moment estimator in negative-binomial case

Method of moments suggests following estimator for γ

$$\hat{\gamma}_T^{NB} = \frac{(\hat{\lambda}_T^{NB})^2}{\hat{V}_T^2 - \hat{\lambda}_T^{NB}}\frac{1}{T-1}\left(\sum_{t=1}^{T} v_t - \frac{\sum_{t=1}^{T} v_t^2}{\sum_{t=1}^{T} v_t}\right)$$

for $\hat{V}_T^2 > \hat{\lambda}_T^{NB}$ otherwise use Poisson or binomial model (no over-dispersion in data $N_1, \ldots, N_T$).

Maximum Likelihood Estimators

- For MLE first objective is not unbiasedness but **maximising probability of given observations**.
- Can be done for densities or for probability weights.
- Assume that components of $\boldsymbol{N} = (N_1, \ldots, N_T)'$ are independent with probability weights $p_k^{(t)}(\vartheta) = \mathbb{P}_\vartheta[N_t = k] = \mathbb{P}[N_t = k]$ which depend on common unknown parameter ϑ.
- Independence property of $N_1, \ldots, N_T$ implies that their joint likelihood function for observation $\boldsymbol{N}$ is given by

$$\mathcal{L}_{\boldsymbol{N}}(\vartheta) = \prod_{t=1}^{T} p_{N_t}^{(t)}(\vartheta)$$

and their joint log-likelihood function is given by

$$\ell_{\boldsymbol{N}}(\vartheta) = \sum_{t=1}^{T} \log p_{N_t}^{(t)}(\vartheta)$$

Maximum Likelihood Estimators

- MLE for ϑ is based on rationale that ϑ should be chosen such that probability of observing $\boldsymbol{N} = (N_1, \ldots, N_T)'$ is maximised.
- $\hat{\vartheta}_T^{MLE}$ for ϑ based on observation $\boldsymbol{N}$ is given by

$$\hat{\vartheta}_T^{MLE} = \underset{\vartheta}{\operatorname{argmax}}\, \mathcal{L}_{\boldsymbol{N}}(\vartheta) = \underset{\vartheta}{\operatorname{argmax}}\, \ell_{\boldsymbol{N}}(\vartheta)$$

  This is solved by root search algorithm.
- Under suitable regularity properties and real valued parameter, $\hat{\vartheta}_T^{MLE}$ is solution of

$$\frac{\partial}{\partial \vartheta} \ell_{\boldsymbol{N}}(\vartheta) = \frac{\partial}{\partial \vartheta} \sum_{t=1}^{T} \log p_{N_t}^{(t)}(\vartheta) = 0$$

- If $p_k^{(t)}(\vartheta)$ are sufficiently regular as function of ϑ in regular domain which contains true parameter ϑ then $\hat{\vartheta}_T^{MLE}$ is **asymptotically unbiased** for $T \to \infty$ and under appropriate scaling it has **asymptotic Gaussian distribution** with inverse Fisher's information as covariance matrix.

Examples

MLE in binomial case

Assume that $N_1, \ldots, N_T$ are independent and $\text{Binom}(v_t, p)$. MLE is given by

$$\hat{p}_T^{MLE} = \frac{1}{\sum_{s=1}^T v_s} \sum_{t=1}^T N_t = \sum_{t=1}^T \frac{v_t}{\sum_{s=1}^T v_s} \frac{N_t}{v_t} = \hat{p}_T^{MV}$$

Proof

$$\ell_{\boldsymbol{N}}(p) = \sum_{t=1}^T \left[\log \binom{v_t}{N_t} + N_t \log p + (v_t - N_t) \log(1-p) \right]$$

Calculating derivative wrt p provides requirement

$$\frac{\partial}{\partial p} \ell_{\boldsymbol{N}}(p) = \sum_{t=1}^T \left[\frac{N_t}{p} - \frac{v_t - N_t}{1-p} \right] = 0$$

Examples

MLE in Poisson case

Assume that $N_1, \ldots, N_T$ are independent and $Poi(\lambda v_t)$. MLE is given by

$$\hat{\lambda}_T^{MLE} = \frac{1}{\sum_{s=1}^T v_s} \sum_{t=1}^T N_t = \sum_{t=1}^T \frac{v_t}{\sum_{s=1}^T v_s} \frac{N_t}{v_t} = \hat{\lambda}_T^{MV}$$

Log likelihood is given by

$$\ell_{\mathbf{N}}(\lambda) = \sum_{t=1}^T \left(N_t \log(\lambda v_t) - \lambda v_t - \log(N_t!)\right)$$

Setting 1st derivative w.r.t. λ equal to zero

$$\frac{\partial}{\partial \lambda} \ell_{\mathbf{N}}(\lambda) = \sum_{t=1}^T \left(\frac{N_t}{\lambda} - v_t\right) = 0$$

Solving this for λ provides proof. (Check that 2nd derivative is negative!)

$\square$

Examples

MLE in negative-binomial case

Assume that $N_1, \ldots, N_T$ are independent and $\mathrm{NegBin}(\lambda v_t, \gamma)$. MLE $(\hat{\lambda}^{MLE}, \hat{\gamma}^{MLE})$ is solution of

$$\frac{\partial}{\partial(\lambda, \gamma)} \sum_{t=1}^{T} \log \binom{N_t + \gamma - 1}{N_t} + \gamma \log(1 - p_t) + N_t \log p_t = 0$$

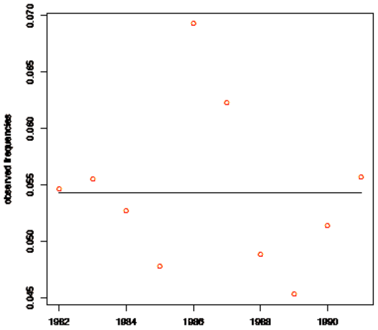
with $p_t = (\lambda v_t)/(\gamma + \lambda v_t) \in (0, 1)$

Unfortunately this last system of equations does not have closed form solution, and root search algorithm is needed to find MLE solution in negative-binomial case.

Example (dataset provided in Gisler)

- Number of claims in insurance portfolio that protect private households against water claims.
- Observed yearly claims frequencies N_t/v_t from $t = 1982$ to 1991 compared to overall average frequency of 5.43%.

year t	volume v_t	number of claims N_t	frequency N_t/v_t
1982	240'755	13'153	5.46%
1983	255'571	14'186	5.55%
1984	269'739	14'207	5.27%
1985	281'708	13'461	4.78%
1986	306'888	21'261	6.93%
1987	320'265	19'934	6.22%
1988	323'481	15'796	4.88%
1989	334'753	15'157	4.53%
1990	340'265	17'483	5.14%
1991	344'757	19'185	5.56%
total	3'018'182	163'823	5.43%



Poisson $N_t \overset{iid}{\sim} Poi(\lambda v_t)$

- Linear minimal variance estimator and MLE for λ

$$\hat{\lambda}_T^{MV} = \hat{\lambda}_T^{MLE} = \frac{1}{\sum_{s=1982}^{1991} v_s} \sum_{t=1982}^{1991} N_t = 5.43\%$$

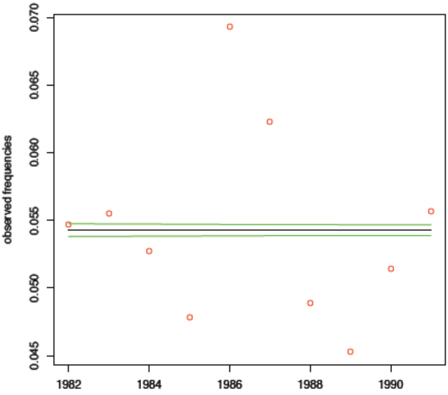
- Coefficient of variation ($\text{Vco}(N_t/v_t) = (\lambda v_t)^{-1/2}$) is estimated

$$\hat{\text{Vco}}(N_t/v_t) = (\hat{\lambda}_T^{MV} v_t)^{-1/2} \approx 0.8\%$$

- We consider confidence interval $CI_t = (\lambda \pm \lambda(\lambda v_t)^{-1/2})$ we obtain

$$\hat{CI}_t = (5.39\%, 5.47\%)$$

Poisson $N_t \overset{iid}{\sim} Poi(\lambda v_t)$



$\Rightarrow$ Very narrow CI and most N_t/v_t lie outside of these confidence bounds.
$\Rightarrow$ Reject assumption of Poisson distributions for number of claims.

Negative-binomial

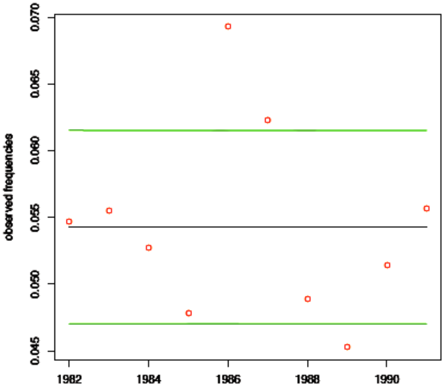
- This model is able to model heterogeneity over different accounting years t.
- We obtain $\hat{V}_T^2 = 15.84 > 5.43\%$ and thus we have clear over-dispersion which results in estimate

$$\begin{aligned} \hat{\gamma}_T^{NB} &= 56.23 \\ \hat{\text{Vco}}(N_t/v_t) &= \sqrt{(\hat{\lambda}_T^{NB})^{-1} + (\hat{\gamma}_T^{NB})^{-1}} \approx 13\% \end{aligned}$$

- 1 standard deviation confidence bounds

$$\hat{CI}_t = (4.70\%, 6.15\%)$$

Negative-binomial



$\Rightarrow$ 7/10 of observations are within confidence bounds.
$\Rightarrow$ This makes much more sense in view of observed frequencies N_t/v_t.

Statistical test

χ^2 goodness of fit test to see whether certain model fits to the data.

- $H_0 : N_t \overset{iid}{\sim} Poi(\lambda v_t)$ for $t = 1, \ldots, T$.
- For test statistic, we define

$$\chi^* = \chi^*(\boldsymbol{N}) = \sum_{t=1}^{T} \frac{(N_t/v_t - \lambda)^2}{\lambda/v_t}$$

- Not straightforward to determine explicit distribution function of χ^*.
- Aggregation and disjoint decomposition theorems imply that $N_t \sim Poi(\lambda v_t)$ can be understood as a sum of v_t iid random variables $X_i \sim Poi(\lambda)$. Hence

$$N_t \overset{(d)}{=} \sum_{i=1}^{v_t} X_i$$

with $E[X_1] = \lambda$ and $\mathrm{Var}(X_1) = \lambda$. But then CLT applies as $v_t \to \infty$

$$\tilde{Z}_t = \frac{N_t/v_t - \lambda}{\sqrt{\lambda/v_t}} = \frac{N_t - \lambda v_t}{\sqrt{\lambda v_t}} \overset{(d)}{=} \frac{\sum_{i=1}^{v_t} X_i - \lambda v_t}{\sqrt{\lambda v_t}} \xrightarrow{D} Z_t \sim \mathcal{N}(0,1)$$

Statistical test

- Approximate $\tilde{Z}_t$ by $Z_t \sim \mathcal{N}(0,1)$ for v_t sufficiently large.
- If $Z_1, \ldots, Z_n \overset{iid}{\sim} \mathcal{N}(0,1)$ then $\sum_{t=1}^{T} Z_t^2$ has χ_T^2 distribution. Therefore

$$\chi^* = \chi^*(\boldsymbol{N}) = \sum_{t=1}^{T} \frac{(N_t/v_t - \lambda)^2}{\lambda v_t} = \sum_{t=1}^{T} \tilde{Z}_t^2 \overset{(d)}{\approx} \sum_{t=1}^{T} Z_t^2 \sim \chi_T^2$$

- We replace unknown λ by $\hat{\lambda}_T^{MLE}$ and lose 1 df.

$$\hat{\chi}^* = \sum_{t=1}^{T} v_t \frac{\left(N_t/v_t - \hat{\lambda}_T^{MLE}\right)^2}{\hat{\lambda}_T^{MLE}} \overset{(d)}{\approx} \chi_{T-1}^2$$

- We revisit previous example. For data we obtain $\hat{\chi}^* = 2627$.
- 99% quantile of χ^2 distribution with $T-1=9$ df is given by 21.67. This is by far smaller than $\hat{\chi}^*$, so we reject H_0 on significance level 1%.

Exercise 19

Consider data given in following table

t	1	2	3	4	5	6	7	8	9	10
N_t	1000	997	985	989	1056	1070	994	986	1093	1054
v_t	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000

Estimate parameters for the Poisson and the negative-binomial models. Which model is preferred? Does a χ^2-goodness-of-fit test reject the null hypothesis on the 5% significance level of having Poisson distributions?

Answer of exercise 19

We use ML estimation for the parameter λ in the Poisson distribution

$$\hat{\lambda}_{Pois}^{MLE} = \frac{1}{\sum_{t=1}^{10} v_t} \sum_{t=1}^{10} N_t = 0.10224$$

The coefficient of variation equals $Vco(N_t/v_t) = (\lambda v_t)^{-1/2}$ this is estimated by

$$\hat{Vco}(N_t/v_t) = (\hat{\lambda} v_t)^{-1/2} = 0.03127$$

A confidence interval (with 1 s.d.) can be estimated by

$$\hat{CI}_t = (\hat{\lambda} \pm \hat{\lambda}(\hat{\lambda} v_t)^{1/2}) = (0.0990425; 0.10540375)$$

The observed claims frequencies N_t/v_t are

t	1	2	3	4	5	6	7	8	9	10
N_t/v_t	0.1	0.0997	0.0985	0.0989	0.1056	0.1070	0.0994	0.0986	0.1093	0.1054

We observe that these lie within the confidence bounds. Hence the Poisson model is probably a good model for this data set.
Under the negative bindomial model the estimate for λ is also

$$\hat{\lambda}_{NB}^{MLE} = \frac{1}{\sum_{t=1}^{10} v_t} \sum_{t=1}^{10} N_t = 0.10224$$

using the moment estimator.
Next we need to estimate γ, but first

$$\hat{V}_T^2 = \frac{1}{9} \sum_{t=1}^{10} v_t \left(\frac{N_t}{v_t} - \hat{\lambda}_{NB} \right)^2 = 0.16856$$

This value is only slightly larger than $\hat{\lambda}_{NB}$. Hence there is no clear sign for overdispersion. However, we continue to estimate γ by the moment estimator 2.30.

$$\begin{aligned} \hat{\gamma}^{NB} &= 1576.149 \\ \hat{Vco}\left(N_t/v_t\right) &= \sqrt{(\hat{\lambda} v_t)^{-1} + \hat{\gamma}^{-1}} = 0.04015655. \end{aligned}$$

This gives a confidence interval (1 s.d.) for N_t/v_t that is a little wider than under the Poisson model. All observations also lie within this interval.
Because there is no clear sign of overdispersion and the Poisson model seems to fit well, we believe the Poisson model is the preferred model.
The test statistic for a χ^2 goodness-of-fit test equals

$$\chi^* = \sum_{t=1}^{10} v_t \frac{(N_t/v_t - \hat{\lambda}_{Pois}^{MLE})^2}{\hat{\lambda}_{Pois}^{MLE}} = 14.83803$$

The 95% quantile of the χ^2 distribution with 9 degrees of freedom is $\chi^2_{0.95,9} = 16.91898$. We have $\chi^* < \chi^2_{0.95,9}$ and hence we cannot reject the null hypothesis that the Poisson model fits well on a 5% significance level.

Exercise 20

An insurance company decides to offer a no-claims bonus to good car drivers, namely

- a 10% discount after 3 years of no claim and
- a 20% discount after 6 years of no claim.

How does the base premium need to be adjusted so that this no-claims bonus can be financed? For simplicity we assume that all risks have been insured for at least 6 years. Answer the question in the following situation: Homogeneous portfolio with iid risks having iid Poisson claim counts with frequency parameter $\lambda = 0.2$.

Answer of exercise 20

Denote by v the total volume of contracts. Write v_0, v_{10}, v_{20} to be the volumes in the raw contract with $0\%, 10\%$ and 20% discount. Of course it holds that

$$v = v_0 + v_{10} + v_{20}$$

We are given

$$\begin{aligned} S_0 &\sim \text{CompPoi}(\lambda v_0, G) \\ S_{10} &\sim \text{CompPoi}(\lambda v_{10}, G) \\ S_{20} &\sim \text{CompPoi}(\lambda v_{20}, G) \end{aligned}$$

Originally the insurance company had as total claim $\tilde{S} \sim \text{CompPoi}(\lambda v, G)$. Under a homogeneous portfolio with iid risks having iid claims with frequency parameter $\lambda = 0.2$, the total expected claim amount will still be

$$\begin{aligned} E[S] &= E[S_0] + E[S_{10}] + E[S_{20}] \\ &= \lambda v_0 E[Y_1] + \lambda v_{10} E[Y_1] + \lambda v_{20} E[Y_1] \\ &= \lambda v E[Y_1] = E[\tilde{S}] \end{aligned}$$

Hence the expected claim amount remains unchanged. Suppose a late premium of π per contract is asked to have a total reserved premium of $v\pi$ to cover all claims. Denote the new base premium by $\tilde{\pi}$. Because the expected claim amount did not change, we should obtain the same reserve, hence

$$v_0\tilde{\pi} + v_{10}0.9\tilde{\pi} + v_{20}0.8\tilde{\pi} = v\pi$$

Therefore

$$\tilde{\pi} = \frac{v}{v_0 + 0.9v_{10} + 0.8v_{20}}\pi$$

Exercise 21

Natural hazards in Switzerland are covered by so-called ES-Pool that organise diversification of these natural hazards. Natural hazards exceeding CHF 50 millions are considered large events. Following 15 large storm and flood events (claim amounts in CHF millions) are observed in years 1986-2005:

52.8	135.2	55.9	138.6	122.9
55.8	368.2	83.8	78.5	75.3
178.3	182.8	54.4	365.3	1051.1

1. Fit Pareto distribution with parameters $\theta = 50$ and $\alpha > 0$ to the observed claim sizes. Estimate parameter α using unbiased version of MLE.
2. We introduce maximal claims cover of $M = 2$ billions CHF per event, i.e. the individual claims are given by $Y_i \wedge M = \min(Y_i, M)$. For yearly claim amount of storm and flood events we assume compound Poisson distribution with Pareto claim sizes Y_i. What is expected total yearly claim amount?
3. What is probability that we observe a storm and flood event next year which exceeds level of $M = 2$ billions CHF?

Answer of exercise 21

1. The unbiased version of the MLE for α is given by

$$\hat{\alpha} = \frac{n-1}{n} \left(\frac{1}{n} \sum_{i=1}^{n} \log Y_i - \log \theta \right)^{-1}$$

Using the given data and $\theta = 50$ we obtain

$$\hat{\alpha} = \frac{14}{15} \left(\frac{1}{15} 72.92974 - \log 50 \right)^{-1} = 0.9824978$$

Using the biased estimator we obtain $\hat{\alpha}_b = 1.052676$

2. Define $\tilde{Y}_i = Y_i \wedge M$ with $M = 2$ billion CHF.

$$
\begin{aligned}
S &= \sum_{i=1}^{N} \tilde{Y}_i \sim CompPois(\lambda v, \tilde{G}) \\
G(y) &= 1 - \left(\frac{y}{\theta}\right)^{-\alpha} \\
E[S] &= \lambda v E[\tilde{Y}_1] = \lambda v E[Y_1 \wedge M] \\
&= \lambda v \left[\int_0^M y dG(y) + MP(Y > M)\right] \\
&= \lambda v \left[E[Y 1_{Y \leq M}] + M(1 - G(M))\right] \\
&= \lambda v \left[I(G(M))\mu_Y + M(1 - G(M))\right] \\
&= \lambda v \left[\left(1 - \left(\frac{M}{\theta}\right)^{-\alpha+1}\right)\frac{\theta\alpha}{\alpha - 1} + M\left(\frac{M}{\theta}\right)^{-\alpha}\right]
\end{aligned}
$$

In the example we estimate λv by $\frac{15}{20}$, the total number of events divided by the observation period. This is the MLE estimation for the mean of a Poisson distribution. Note that we cannot use the unbiased estimate because it is smaller than 1. Thus

$$
\begin{aligned}
E[S] &= \frac{15}{20}\left[\left(1-\left(\frac{2000}{50}\right)^{-\hat{\alpha}_b+1}\right)\frac{50\hat{\alpha}_b}{\hat{\alpha}_b-1}+2000\left(\frac{2000}{50}\right)^{-\hat{\alpha}_b}\right] \\
&= 163.2227 \text{ million CHF}
\end{aligned}
$$

3

$$P(Y > M) = \left(\frac{M}{\theta}\right)^{-\alpha}$$

Using biased estimate for α:

$$P(Y > M) = 0.02058$$

Using unbiased estimate for α:

$$P(Y > \alpha) = 0.026667$$

Outline

Compound distributions

Compound distribution is base model for collective risk modeling and we have discussed **different choices for the claims count distribution** of N and for the **individual claim size distribution** of Y_i.
Using both models, we obtained the characteristics of the distribution of S.

Basic recognition features of compound distributions

Assume S has a **compound distribution**. We have (whenever they exist)

$$\begin{aligned} E[S] &= E[N]E[Y_1] \\ \operatorname{Var}(S) &= \operatorname{Var}(N)E[Y_1]^2 + E[N]\operatorname{Var}(Y_1) \\ \operatorname{Vco}(S) &= \sqrt{\operatorname{Vco}(N)^2 + \frac{1}{E[N]}\operatorname{Vco}(Y_1)^2} \\ M_S(r) &= M_N\left(\log(M_{Y_1(r)})\right) \qquad \text{for } r \in \mathbb{R}. \end{aligned}$$

Distribution of S

If assumptions above hold, then distribution function of S can be written as

$$\begin{aligned} F_S(x) &= \mathbb{P}[S \leq x] = \sum_{k \in \mathcal{A}} \mathbb{P}\left[\sum_{i=1}^{N} Y_i \leq x \middle| N = k\right] \mathbb{P}\left[N = k\right] \\ &= \sum_{k \in \mathcal{A}} \mathbb{P}\left[\sum_{i=1}^{k} Y_i \leq x\right] \mathbb{P}\left[N = k\right] = \sum_{k \in \mathcal{A}} G^{*k}(x) \mathbb{P}\left[N = k\right] \end{aligned}$$

G^{*k} denotes the k-th convolution of the distribution function G.
In particular, we have for $Y_1, Y_2 \overset{iid}{\sim} G$

$$\begin{aligned} G^{*2}(x) &= \mathbb{P}[Y_1 + Y_2 \leq x] = \int G(x-y) dG(y) \\ G^{*k}(x) &= \int G^{*(k-1)}(x) dG(y) \end{aligned}$$

We obtain a **closed form solution** for the distribution function of S.

Distribution of S

- **Aim**: calculate compound distribution S
  $\Rightarrow$ we can easily calculate moments and mgf of S
- Distribution of S is difficult since it involves in general **(too) many convolutions** of claim size distribution function G.
- We discuss several methods to circumvent this difficulty.
- Typically Monte Carlo simulations are applied and then the resulting empirical distribution function is considered as a sufficiently good **approximation** to the true one (based on Glivenko-Cantelli theorem).
- It is however often unclear to asses what *sufficiently good* means, i.e. the rates of convergence of the Monte Carlo samples may be very poor which results in a **huge amount of simulations** (especially for **heavy tailed** distribution functions of regularly varying type).
- Therefore we present alternatives, such as approximations, Panjer algorithm and fast Fourier transform.

Approximations

- In many situations approximations to S are used, which are justified by Central Limit Theorem (CLT) if expected number of claims is large.
- Compound distributions may have two different **risk drivers** in the tail of the distribution function
  - number of claims N
  - single large claims in $Y_1, \ldots, Y_N$
- We focus on compound Poisson model and **separate small from large claims** resulting in the independent decomposition

$$S = S_{sc} + S_{lc}.$$

- If expected number of small claims $v\lambda_{sc}$ is large, S_{sc} can be approximated by parametric distribution function.
- S_{lc} should be modeled explicitly.

Normal approximation

- Classical CLT holds for fixed number of claims.
- In our approach number of claims is not fixed, hence we need refinement.
- For a Poissonian number of claims N we do this by keeping expected claims frequency λ fixed and by sending $v \to \infty$

Theorem 5

Assume $S \sim CompPoi(\lambda v, G)$ with G having a finite second moment. We have

$$\frac{S - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}} \Rightarrow N(0,1) \text{ as } v \to \infty$$

- We consider special class of distribution functions G having finite second moment. As long as we work in set-up of S_{sc} this is not a restriction since claim sizes are bounded by large claims threshold M and therefore have finite variance.

Proof

- We choose $v \in \mathbb{N}$ and then disjoint decomposition theorem provides

$$S = \sum_{i=1}^{N} Y_i \stackrel{(d)}{=} \sum_{\ell=1}^{v} \sum_{i=1}^{N_\ell} Y_i^{(\ell)} = \sum_{\ell=1}^{v} S_\ell$$

  where $S_\ell \stackrel{iid}{\sim} \text{CompPoi}(\lambda, G)$
- First two moments of these compound Poisson distributions are given by

$$\begin{aligned} E[S_1] &= \lambda E[Y_1] \\ \text{Var}[S_1] &= \lambda E[Y_1^2] \end{aligned}$$

- Therefore, assumptions of CLT are fulfilled and theorem is proven.

Normal approximation

- Theorem 5 is motivation for following approximation of distribution function of S

$$P[S \leq x] = P\left[\frac{S - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}} \leq \frac{x - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}}\right] \approx \Phi\left(\frac{x - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}}\right) \quad (1)$$

- Approximation works well when v is large and if claim sizes Y_i do not have heavy tailed distribution functions G.
- Otherwise it under-estimates true potential of large outcomes of S
- Normal approximation allows for negative claims S, which under our model assumptions is excluded.

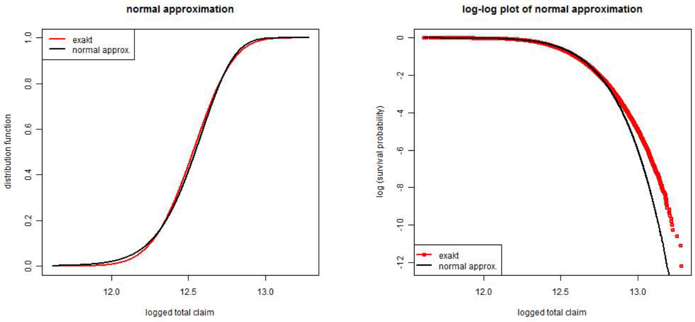
Example (PP Insurance)

We consider 3 different examples

1. Only small claims: $G(y) = P[Y \leq y | Y \leq M]$. As explicit claim size distribution function we choose empirical distribution of example 3.13 with $M = 50000$. We choose portfolio size v such that $\lambda v = 100$
2. Same G as before, but now v is chosen such that $\lambda v = 1000$
3. In addition to 2. we add large claims layer modeled by Pareto distribution with $\theta = M = 50000$ and $\alpha = 2.5$ and for expected number of large claims we set $\lambda_{lc} v = 3.9$.

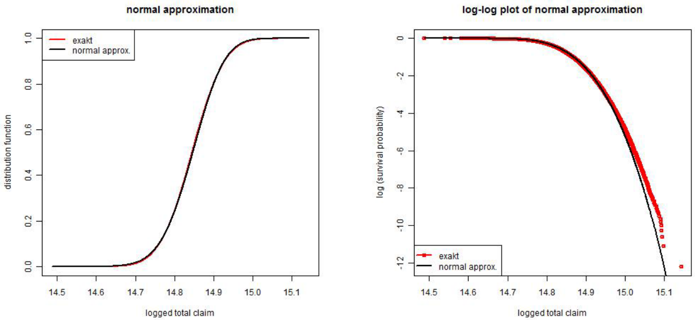
True distribution function is evaluated by Monte Carlo simulation. (100000 simulations)

Example (PP Insurance): Case a)



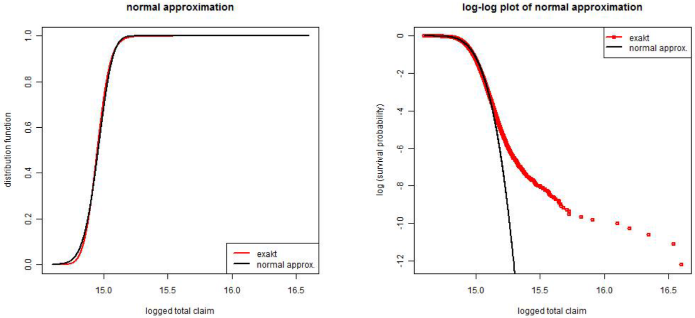
- Appropriately good fit around the mean but normal approximation clearly underestimates tails of true distribution.
- True distribution has positive skewness $\zeta_S = 0.43$ whereas normal approximation has 0 skewness.
- Probability mass $\Phi\left(-\lambda v E[Y_1]/\sqrt{\lambda v E[Y_1^2]}\right) = 6 \cdot 10^{-7}$ for negative total claim amount (fairly small).

Example (PP Insurance): Case b)



- Better approximation properties due to fact that we have convergence in distribution for portfolio size $v \to \infty$
- Lower skewness $\zeta_S = 0.15$ which improves normal approximation (also in the tails).

Example (PP Insurance): Case c)



- Including large claims has an expected number of large claims of 3.9 and Pareto tail parameter of $\alpha = 2.5$.
- Normal approximation is useless in the tails which strongly favors large claims separation (as suggested before).

Translated gamma and log-normal approximation

- Normal approximation can be useful for large portfolio sizes v and under exclusion of large claims.
- For small portfolio sizes approximation may be bad because true distribution has substantial skewness.
  $\Rightarrow$ approximate small claims layer by other distribution functions **enjoying positive skewness**.
- We choose $k \in \mathbb{R}$ and define translated or shifted random variables

$$X = k + Z \qquad \text{where } Z \sim \Gamma(\gamma, c) \quad \text{or} \quad Z \sim LN(\mu, \sigma^2)$$

Translated gamma case	Translated log-normal case
$E[X] = k + \gamma/c$	$E[X] = k + e^{\mu+\sigma^2/2}$
$\mathrm{Var}(X) = \gamma/c^2$	$\mathrm{Var}(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$
$\zeta_X = 2\gamma^{-1/2}$	$\zeta_X = \left(e^{\sigma^2} + 2\right)\left(e^{\sigma^2} - 1\right)^{1/2}$

Translated gamma and log-normal approximation

Do **fit of moments** between S and X.

- Assume S has **finite third moment** and then we choose

$$X = k + Z \qquad \text{where } Z \sim \Gamma(\gamma, c) \quad \text{or} \quad Z \sim LN(\mu, \sigma^2)$$

such that three parameters of X fulfill

$$E[X] = E[S] \qquad \text{Var}(X) = \text{Var}(S) \qquad \zeta_X = \zeta_S \tag{2}$$

- This fitted random variable X is chosen as approximation to S.

Exercise 22

Given that mean and variance of the number of claims equals respectively 6.7 and 2.3, whereas the mean and variance of individual losses is respectively 179247 and 52141.

1. Determine mean and variance of aggregate claims per month
2. Using normal and lognormal distributions as approximating distributions for aggregate claims, calculate probability that claims will exceed 140% of expected costs.

Answer of exercise 22

1.

$$\begin{aligned} E(S) &= 6.7(179247) = 1200955 \\ \text{Var}(S) &= 6.7(52141)^2 + 2.3^2(179247)^2 = 1.88189 \times 10^{11} \end{aligned}$$

2.

$$\mathbb{P}[S > 1.40 \times 1200955] = \mathbb{P}[S > 1681337]$$

- For normal distribution

$$\begin{aligned} \mathbb{P}[S > 1681337] &= \mathbb{P}\left(\frac{S - E(S)}{\sqrt{\text{Var}(S)}} > \frac{1681337 - 1200955}{433797}\right) \\ &= 1 - \Phi(1.107) = 0.134 \end{aligned}$$

- For lognormal distribution

$$\begin{aligned} \mathbb{P}[S > 1681337] &= \mathbb{P}\left(Z > \frac{\log(1681337) - 13.93731}{(0.1226415)^{1/2}}\right) \\ &= 1 - \Phi(1.135887) = 0.128 \end{aligned}$$

by using $\sigma^2_{\log S} = \log[1 + Vco^2(S)]$ and $\mu_{\log S} = \log(\mu_S) - 0.5\sigma^2_{\log S}$ (check this).

Exercise 23

Assume that S has compound Poisson distribution with expected number of claims $\lambda v > 0$ and claim size distribution G having finite third moment

1. Prove that the fit of moments approximation (2) for a translated gamma distribution for X provides the following system of equations

$$
\begin{aligned}
\lambda v E[Y_1] &= k + \gamma/c \\
\lambda v E[Y_1^2] &= \gamma/c^2 \\
\frac{E[Y_1^3]}{(\lambda v)^{1/2} E[Y_1^2]^{3/2}} &= 2\gamma^{-1/2}
\end{aligned}
$$

2. Solve this system of equations for $k \in \mathbb{R}$, $\gamma > 0$ and $c > 0$ and prove that it has a well-defined solution for $G(0) = 0$.
3. Why should this approximation not be applied to case (c) of former Example?

Answer of exercise 23

1. From the setting

$$\begin{aligned} E[X] &= E[S] \\ \mathrm{Var}(X) &= \mathrm{Var}(S) \\ \zeta_X &= \zeta_S \end{aligned}$$

we immediately obtain

$$\begin{aligned} k + \frac{\gamma}{c} &= \lambda v E[Y_1] \\ \frac{\gamma}{c^2} &= \lambda v E[Y_1^2] \\ 2\gamma^{-1/2} &= \frac{1}{(\lambda v E[Y_1^2])^{3/2}} E\left[(S - E(S))^3\right] \end{aligned}$$

Hence we only need to simplify $E\left[(S - E(S))^3\right]$

$$\begin{aligned} E\left[(S - E(S))^3\right] &= \frac{\partial^3}{\partial r^3} \log M_S(r)|_{r=0} \\ &= \lambda v M_{Y_1}'''(r)|_{r=0} = \lambda v E[Y_1^3] \end{aligned}$$

Hence we obtain the system

$$
\begin{aligned}
k+\frac{\gamma}{c} &= \lambda v E[Y_1] \\
\frac{\gamma}{c^2} &= \lambda v E[Y_1^2] \\
2\gamma^{-1/2} &= \frac{E[Y_1^3]}{(\lambda v)^{1/2} E[Y_1^2]^{3/2}}
\end{aligned}
$$

2. From

$$
\begin{aligned}
k+\frac{\gamma}{c} &= \lambda v E[Y_1] \\
\frac{\gamma}{c^2} &= \lambda v E[Y_1^2] \\
2\gamma^{-1/2} &= \frac{E[Y_1^3]}{(\lambda v)^{1/2} E[Y_1^2]^{3/2}}
\end{aligned}
$$

it follows

$$
\begin{aligned}
\gamma &= \left(\frac{E[Y_1^3]}{2(\lambda v)^{1/2}E[Y_1^2]^{3/2}}\right)^{-2} = \frac{4\lambda v E[Y_1^2]^3}{E[Y_1^3]^2} \\
c^2 &= \frac{\gamma}{\lambda v E[Y_1^2]} = \frac{4E[Y_1^2]^2}{E[Y_1^3]^2} \\
k &= \lambda v E[Y_1] - \frac{\gamma}{c}
\end{aligned}
$$

Hence

$$
\begin{aligned}
\gamma &= \frac{4\lambda v E[Y_1^2]^3}{E[Y_1^3]^2} \\
c &= \frac{2E[Y_1^2]}{E[Y_1^3]} \\
k &= \lambda v E[Y_1] - \frac{2\lambda v E[Y_1^2]^2}{E[Y_1^3]}
\end{aligned}
$$

Because $\lambda, v > 0$ and $E[Y_1^2] > 0$ we only need $E[Y_1^3] > 0$ to have $\gamma, c > 0$.

We have

$$E[Y_1^3] = \int_{-\infty}^{+\infty} y^3 dG(y) = \int_0^{+\infty} y^3 dG(y) > 0$$

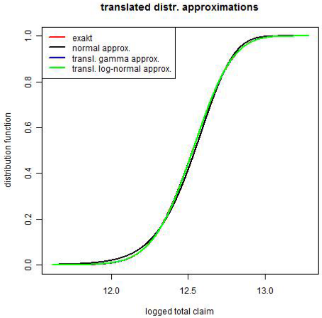
because Y_1 is not degenerate in 0 and $E[Y_1^3]$ is finite because it is given.

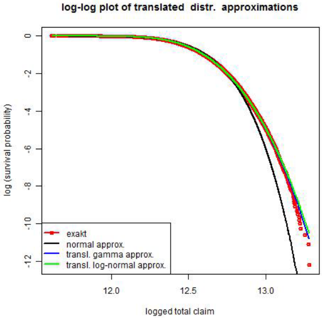
3. Becasue when $\alpha = 2.5$ tjem $\int_{\theta}^{\infty} dG_{lc} = \infty$ and thus the tail part (large claims part) of the model has no finite third moment.

$$E(X^3) < \infty \qquad \text{for} \qquad \alpha > 3$$

Example (PP insurance): Case a)

We only consider small claims layer and expected number of claims is $\lambda v = 100$

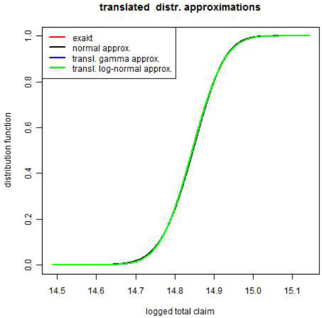


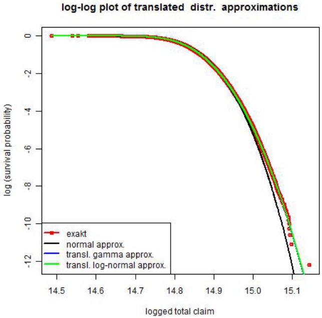


- Both fits are remarkably good.
- Small claims layer is often approximated by one of these two parametric distribution functions.

Example (PP insurance): Case b)

We only consider small claims layer and expected number of claims is $\lambda v = 1000$





- Both fits are remarkably good.
- Small claims layer is often approximated by one of these two parametric distribution functions.

Example (PP insurance): Case a) and b)

- KS test rejects H_0 on 5% significance level for normal approximation in both cases, whereas this is not the case for translated gamma and log-normal approximation, see p-values

approximation	case (a)	case (b)
normal	0%	0%
translated gamma	51%	57%
translated log-normal	8%	59%

- Case a) favors translated gamma approximations, whereas in case b) log-normal approximation is slightly preferred.

Edgeworth approximation

- Until now we have just chosen a simple distribution function that enjoys skewness and then we have done a fit of moments.
- Edgeworth approximation starts from CLT and then tries to adjust higher order terms by evaluation of mgf in terms of Taylor expansions.
- Assume S is compound Poisson distributed with claim size distribution G having a positive radius of convergence $\rho_0 > 0$.
- Define normalized random variable

$$Z = \frac{S - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}}$$

$\Rightarrow E[Z] = 0, \mathrm{Var}(Z) = 1$ and $\zeta_Z = \zeta_S$.

- Aim is to approximate mgf of Z by comparable terms coming from normal distributions.

Edgeworth approximation

- Consider following Taylor expansion around origin, choose $n \geq 3$

$$\log M_Z(r) = \sum_{k=0}^{n} \frac{\frac{d^k}{dr^k} \log M_Z(r)|_{r=0}}{k!} r^k + o(r^n) \qquad \text{as } r \to 0$$

- Set $a_k = \frac{\frac{d^k}{dr^k} \log M_Z(r)|_{r=0}}{k!}$
$\Rightarrow a_0 = \log M_Z(0) = 0, a_1 = E[Z] = 0$ and $a_2 = \text{Var}(Z)/2! = 1/2$.
- This provides approximation

$$M_Z(r) \approx e^{\frac{1}{2}r^2 + \sum_{k=3}^{n} a_k r^k} = e^{\frac{1}{2}r^2} e^{\sum_{k=3}^{n} a_k r^k}$$

- Using second Taylor expansion for $e^x = 1 + x + x^2/2! + \ldots$ applied to latter exponential function in last expression, mgf of Z is approximated by

$$M_Z(r) \approx e^{r^2/2} \left[1 + \sum_{k=3}^{n} a_k r^k + \frac{\left(\sum_{k=3}^{n} a_k r^k\right)^2}{2!} + \ldots \right]$$

Edgeworth approximation

- Depending on required precision as $r \to 0$ we can choose more terms in the bracket and we can take more terms in the summation reflected by the upper index n in the summation.
- For appropriate constants $b_k \in \mathbb{R}$ we get approximation (for small r)

$$M_Z(r) \approx e^{r^2/2}\left[1 + a_3 r^3 + \sum_{k\geq 4} b_k r^k\right] \tag{3}$$

Lemma 6

Let Φ denote standard Gaussian distribution function and $\Phi^{(k)}$ its k-th derivative. For $k \in \mathbb{N}_0$ and $r \in \mathbb{R}$

$$r^k e^{r^2/2} = (-1)^k \int_{-\infty}^{\infty} e^{rx}\Phi^{(k+1)}(x)dx$$

Edgeworth approximation

Proof

- Proof goes by induction. Choose $k = 0$ then

$$\int_{-\infty}^{\infty} e^{rx}\Phi'(x)dx = \int_{-\infty}^{\infty} e^{rx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = M_X(r) = e^{r^2/2}$$

which is mgf of $X \sim N(0,1)$.

- Induction step $k \to k+1$. Using integration by parts we have

$$\begin{aligned}
&(-1)^{k+1}\int_{-\infty}^{\infty} e^{rx}\Phi^{(k+2)}(x)dx \\
&= \left[(-1)^{k+1}e^{rx}\Phi^{(k+1)}(x)\right]_{-\infty}^{\infty} - (-1)^{k+1}\int_{-\infty}^{\infty} re^{rx}\Phi^{(k+1)}(x)dx \\
&= 0 + r(-1)^{k}\int_{-\infty}^{\infty} e^{rx}\Phi^{(k+1)}(x)dx \\
&= rr^{k}e^{r^2/2}
\end{aligned}$$

Edgeworth approximation

- Set $X \sim N(0,1)$ and rewrite approximation (3) as (using Lemma 6):

$$\begin{aligned} M_Z(r) &\approx E[e^{rX}] - a_3 \int_{-\infty}^{\infty} e^{rx}\Phi^{(4)}(x)dx + \sum_{k\geq 4} b_k(-1)^k \int_{-\infty}^{\infty} e^{rx}\Phi^{(k+1)}(x)dx \\ &= \int_{-\infty}^{\infty} e^{rx}\left[\Phi'(x) - a_3\Phi^{(4)}(x) + \sum_{k\geq 4} b_k(-1)^k\Phi^{(k+1)}(x)\right] dx \end{aligned}$$

- Let Z have distribution function F_Z, then the latter suggests approximation

$$dF_Z(z) \approx \left[\Phi'(z) - a_3\Phi^{(4)}(z) + \sum_{k\geq 4} b_k(-1)^k\Phi^{(k+1)}(z)\right] dz$$

Edgeworth approximation

- Integration provides Edgeworth approximation ($x = \sqrt{\lambda v E[Y_1^2]}z + \lambda v E[Y_1]$)

$$P[S \leq x] = F_Z(z) \approx EW(z) \stackrel{def}{=} \Phi(z) - a_3\Phi^{(3)}(z) + \sum_{k\geq 4} b_k(-1)^k\Phi^{(k)}(z) \qquad (4)$$

- First order approximation Φ is corrected by higher order terms involving skewness and other higher order terms reflected by a_S and b_k.
- Consider derivatives $\Phi^{(k)}$ for $k \geq 1$

$$\begin{aligned} \Phi'(z) &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \\ \Phi^{(k)}(z) &= \frac{d^{k-1}}{dz^{k-1}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \mathcal{O}\left(z^{k-1}e^{-z^2/2}\right) \qquad \text{for } |z| \to \infty; k \geq 2 \end{aligned}$$

- From this it follows that

$$\lim_{z\to-\infty} EW(z) = 0 \qquad \text{and} \qquad \lim_{z\to\infty} EW(z) = 1$$

Edgeworth approximation

Edgeworth approximation is elegant but its use requires some care:

Edgeworth approximation $EW(z)$ is NOT necessarely a distribution function because it does NOT need to be monotone in z

Example 4.5

- We only take into account skewness, i.e. $a_3 = \zeta_Z \sigma_Z^3/6 = \zeta_S/6$ and approximation $e^z \approx 1 + z$ in (4)

$$\begin{aligned}
\Phi'(z) &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \\
\Phi^{(2)}(z) &= -z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \\
\Phi^{(3)}(z) &= -\frac{1}{\sqrt{2\pi}} e^{-z^2/2} + z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \\
\Phi^{(4)}(z) &= z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} + 2z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} - z^3 \frac{1}{\sqrt{2\pi}} e^{-z^2/2}
\end{aligned}$$

Edgeworth approximation

- This implies

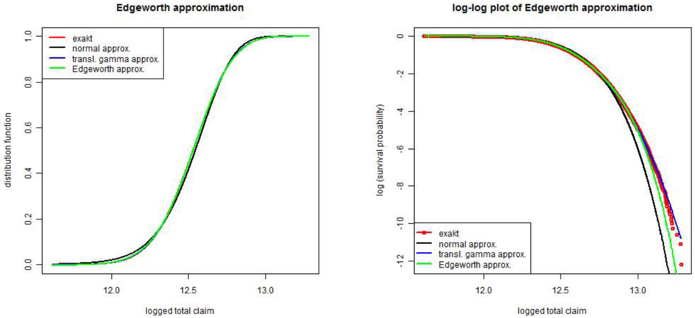
$$\frac{d}{dz}EW(z) = \Phi'(z) - a_3\Phi^{(4)}(z) = \Phi'(z)(1 - 3a_3 z + a_3 z^3)$$

- Consider function $h(z) = 1 - 3a_3 z + a_3 z^3$ for positive skewness $\zeta_S > 0$.
- Then we have

$$\lim_{z\to-\infty} h(z) = -\infty \quad \text{and} \quad \lim_{z\to\infty} h(z) = \infty$$

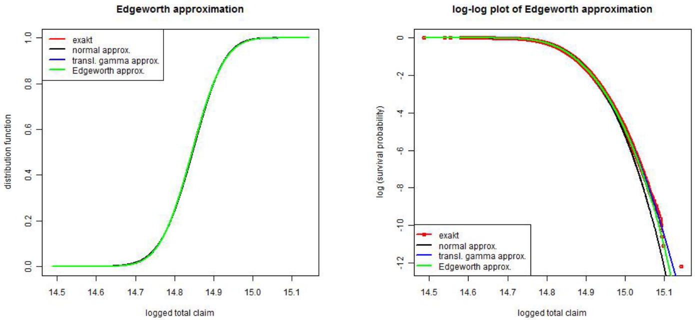
  $\Rightarrow$ derivative of $EW(z)$ has both signs and hence $EW(z)$ is not monotone.
- However, in the upper tail of the distribution of S, that is, for z sufficiently large, the Edgeworth approximation is monotone and can be used as an appropriate approximation.
- Always check these monotonicity properties in the Edgeworth approximation.

Example (PP insurance): Case a)



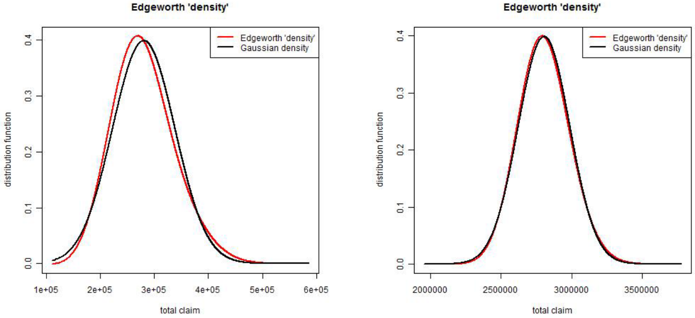
- We only choose next additional moment, which is skewness and refers to term a_3 and we use approximation $e_z \approx 1 + z$
- EW clearly outperforms Gaussian approximation
- EW is still light-tailed which can be seen by comparing it to translated gamma approximation.

Example (PP insurance): Case b)



- We only choose next additional moment, which is skewness and refers to term a_3 and we use approximation $e_z \approx 1 + z$
- EW clearly outperforms Gaussian approximation
- EW is still light-tailed which can be seen by comparing it to translated gamma approximation.

EW density versus Gaussian density



- We see influence of skewness parameter a_3 and $\zeta_S > 0$ respectively.
- Influence of skewness parameter is decreasing with higher expected number of claims (which reflects the CLT).

Example (PP insurance): Case a) and b)

- If we calculate minimum value of Edgeworth density we obtain

$$(a) - 9.8 \cdot 10^{-4} \quad \text{and} \quad (b) - 4.1 \cdot 10^{-5}.$$

  $\Rightarrow$ EW density is not a proper probability density because it violates the positivity property.
- However, this only occurs in range of very small claims and therefore it can be used as an approximation in range of large claims.
- Table shows p-values from KS test of different approximations.

approximation	case (a)	case (b)
normal	0%	0%
translated gamma	51%	57%
translated log-normal	8%	59%
Edgeworth	13%	58%

Example (PP insurance): Case a) and b)

- In this particular case we see that translated gamma distribution is preferred in (a), whereas in (b) approximations are very similar.
- For this reason one often chooses the translated gamma distribution in practice (and also because it can easily be handled).
- Note that Edgeworth approximation can be refined and improved by considering more terms in Taylor expansion.
- There exist similar approximations as Edgeworth approximation, for instance Gram-Charlier expansion, Laguerre-gamma expansion or Jacobi-beta expansion.
- These expansions have similar weaknesses as Edgeworth approximation and we will not further discuss them.