Mathematics of Non-life Insurance 1
course notes

Lucie Mazurová

December 2017

Contents

Chapter 1

Introduction

The insurance is based on a relationship between an insurer (insurance company) and an insured (policyholder), specified usually by an insurance contract. According to the contract, the insurer idemnifies any financial loss that occurs to the insured in a specified period of time and that is caused by a specified random event. The randomness is crutial for working of the insurance and it means that at the beginning of the period covered by the insurance it is not known whether the event will occur or what the amount of the resulting loss will be. The amount paid by the insurer in case of the random insurance event is called **insurance claim**. The price for the insurance cover, paid by the insured, is **insurance premium**.

The present text is devoted to mathematical metods used in **non-life insurace** (also known as property and liability insurance or general insurance in literature). We distinguish between various **risks** according to the cause of the possible financial loss covered by the insurance. In the field of **life insurance** the risks of death or survival up to a given age are mainly covered. All other risks are included in the field of non-life insurance. The most important branches in this field are property insurance and liability insurance. Less typical representants of non-life insurance are accident and health insurance, credit insurance, legal protection insurance, travel insurance and other products.

Often the insurance claims in different branches of non-life insurance have different features. In property insurance the amount paid by the insurer is often bounded from above by a **sum insured**. The payments for such losses are usually settled quite fast. On the other hand, losses in liability insurance, especially in case of health damage, may have a long process of settlement and also may attain relatively high amounts. Those aspects have to be taken into account when considering appropriate mathematical models.

Mathematical models are used in different areas of non-life insurance:

We need to model and predict sizes of future claims in order to determine the premiums correctly. Another important task for a mathematician is to estimate **technical provisions** - a quantity contained in the liability side of the insurance company's balance sheet, which serves for providing coverage for the liabilities following from the insurance policies contained in the company's portfolio. The most important type of technical provisions in non-life area are **claims reserves**. They are based on the estimates of future payments for the claims covered by the existing policies, that have already ocurred but have not yet been reported or completely settled. There exist many mathematical (statistical) methods that can be applied to solve this problem.

The insurance company also may need calculations connected with **reinsurance**. The reinsurance works on principles similar to insurance. The insurer (cedent) transfers part of the risks covered by the insurance policies to another party - the reinsurer (it can be another insurance company or a specialized reinsurance company). The conditions for such a transfer are specified in a reinsurance contract and the price for it is called **reinsurance premium**.

Mathematical modelling is also applied in the creating of complex internal models that enable to simulate all main processes that influence the working of an insurance company. Those models are then used in the risk management and also may serve for determining the solvency capital (the capital required by the regulator for ensuring the ability of the company to fulfill its insurance liabilities).

The course is organized as follows. Mathematics of Non-Life Insurance 1 covers basic probabilistic models that are used in practice for modelling of the individual claim sizes, of the numbers of claims recorded in a given time period and also of the aggregate losses recorded for a policy or a group of policies in a given period. We also show some applications of those models in the field of reinsurance calculations (after presenting basic forms of reinsurance met in practice). We also deal with some simple models used in claims reserving.

The course is followed by Mathematics of Non-Life Insurance 2 devoted especially to the topic of setting the premium rates in the non-life insurance. Generalized linear models (GLM) and bayesian credibility models are the main tools for this part of the presentation. Possible applications of those models in the field of reserving are also mentioned there.

We do not adress the issues connected with the internal models and solvency capital calculations, since it is beyond the scope of the given course.

Chapter 2

Loss Distributions

The first part of the course is devoted to modelling of random variables representing sizes of individual claims, claim numbers and aggregate claims in non-life insurance.

We start with a brief survey of basic distributional characteristics and their empirical estimation. The main goal of the text is the introduction of suitable parametric distributions and their properties. We mention continuous distributions most frequently used to model non-negative claim sizes and also more general distributional families derived by transforming random varibles with simple parametric distributions. We deal with modelling frequency distributions in the general framework of $(a, b, 0)$ and $(a, b, 1)$ classes. We also mention the derivation of distributions by means of compounding and mixing. Compound distribution is the key model for aggregate loss in a homogeneus portfolio of risks in a given period of time. We concentrate on its useful properties and on possible ways of calculation or approximation of its distribution function.

We note that the statistical methods for fitting distributions to insurance data, e.g. goodness of fit tests and methods for parameter estimation, are not covered by the lectures. They are the subject of the tutorial to the course.

2.1 Basic characteristics of random variables

In this section we deal with basic characteristics of random variables and their empirical estimation.

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with the **distribution function (d.f.)** $F(x)$ (e.g. individual claim sizes incurred in an insurance portfolio, numbers of claims reported for insurance policies in given time period).

We consider two possible forms of data to analyze:

1. Individual data: we work directly with the observed values $X_1 = x_1, \dots, X_n = x_n$.

2. Grouped data: in this case we have boundaries $c_0 < c_1 < \cdots < c_r$ and for each $j = 1, \dots, r$ we know n_j - the number of observations in the interval $(c_{j-1}, c_j]$.

2.1.1 Empirical distribution function

From the individual data, d.f. $F(x)$ is estimated by the **empirical distribution function** defined by

$$F_n(x) = \frac{\sum_{j=1}^{n} I[x_j \le x]}{n}, \tag{2.1}$$

where $I[A] = 1$ if A holds true, $I[A] = 0$ otherwise.

The empirical d.f. is a step function that increases by $1/n$ at each data point. In the case of grouped data, the empirical distribution function can be obtained at the boundaries as

$$F_n(c_j) = \frac{1}{n} \sum_{i=1}^{j} n_i.$$

By connecting those values by straight lines we obtain a piecewise linear function called **ogive**. It is an approximation to the empirical d.f., formally defined by

$$\tilde{F}_n(x) = \begin{cases} 0 & \text{if } x \le c_0, \\ \frac{(c_j - x)\, F_n(c_{j-1}) + (x - c_{j-1})\, F_n(c_j)}{c_j - c_{j-1}}, & \text{if } c_{j-1} < x \le c_j \\ 1 & \text{if } x > c_r. \end{cases} \tag{2.2}$$

The derivative (where it exists) of the ogive (2.2) is an empirical approximation to the probability density function and is called a **histogram**.

$$\tilde{f}_n(x) = \begin{cases} 0 & \text{if } x \le c_0, \\ \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} = \frac{n_j}{n\,(c_j - c_{j-1})} & \text{if } c_{j-1} < x \le c_j, \\ 0 & \text{if } x > c_r. \end{cases} \tag{2.3}$$

2.1.2 Moments

In insurance applications we usually deal with non-negative random variables, we therefore assume $F(x) = 0$ for $x < 0$.

We define:

- **k-th raw moment** (k-th moment about the origin):

$$\mu'_k = \int_0^\infty x^k \, dF(x). \tag{2.4}$$

The estimate of μ'_k based on the individual data is

$$\hat{\mu}'_k = \int x^k \, dF_n(x) = \frac{1}{n} \sum_{j=1}^{n} x_j^k.$$

For grouped data, we obtain using histogram (2.3) as an estimate of the density

$$\hat{\mu}'_k = \sum_{j=1}^{r} \int_{c_{j-1}}^{c_j} x^k \frac{n_j}{n\,(c_j - c_{j-1})} dx = \sum_{j=1}^{r} \frac{n_j\,(c_j^{k+1} - c_{j-1}^{k+1})}{n\,(k+1)\,(c_j - c_{j-1})}.$$

- **k-th central moment**:

$$\mu_k = \mathrm{E}\,(X - \mathrm{E}\,X)^k \,.$$

An important special case is the **variance**, $\mathrm{Var}\,X = \mu_2$.

The **skewness** of r.v. X is defined by

$$\gamma_1 = \frac{\mu_3}{(\mu_2)^{3/2}}.$$

In insurance we often observe positive skewness in the data (an assymetry caused by higher probability of “small” values).

We mention other quantities applicable especially in the modelling of insurance claims.

- **k-th limited moment**:

$$\mathrm{E}(X \wedge u)^k = \mathrm{E}\min(X, u)^k, \tag{2.5}$$

where $u > 0$. $\mathrm{E}(X \wedge u)$ is called **limited expected value** (LEV).

In the context of insurance, u could be a **policy limit** - when the loss exceeds the limit u, only the amount of u is covered by the insurance (similar construct is used in XL-reinsurance dealt with in Section 3.1.2.).

We have

$$\mathrm{E}(X \wedge u)^k = \int_0^u x^k \, dF(x) + u^k \, (1 - F(u)) \,. \tag{2.6}$$

Substituting $k = 1$ in (2.6) and using the integration by parts for Stiltjes integral, we derive the following expression of the LEV of a positive random variable:

$$\mathrm{E}(X \wedge u) = \int_0^u (1 - F(x)) \, dx. \tag{2.7}$$

Note that for $u \to +\infty$ we obtain the well known formula for the expected value of a positive random variable (provided that $\mathrm{E}\, X < \infty$).

Another type of restriction applied sometimes to an insurance cover is a **deductible** - when the loss is less than or equal to the deductible, the amount paid is the loss less the deductible.

Let X be r.v. representing the total loss. With a deductible d and a limit u, the amount paid (per loss) is represented by r.v. Y,

$$Y = \begin{cases} 0 & \text{if } X \leq d, \\ X - d & \text{if } d < X < u, \\ u - d & \text{if } X \geq u. \end{cases}$$

The expected amount paid per loss is

$$\begin{aligned}
\mathrm{E}\, Y &= \int_d^u (x - d) \, dF(x) + (u - d) \, (1 - F(u)) \\
&= \int_0^u x \, dF(x) - \int_0^d x \, dF(x) - d \, (F(u) - F(d)) + (u - d) \, (1 - F(u)) \\
&= \int_0^u x \, dF(x) + u \, (1 - F(u)) - \int_0^d x \, dF(x) - d \, (1 - F(d)) \\
&= \mathrm{E}(X \wedge u) - \mathrm{E}(X \wedge d).
\end{aligned}$$

From here and from (2.7) it follows

$$\mathrm{E}\, Y = \int_d^u (1 - F(x)) \, dx.$$

The **loss elimination ratio** for a deductible d is the relative reduction in the expected payment given the imposition of a deductible.

$$\mathrm{LER}_X(d) = \frac{\mathrm{E}\, [X \wedge d]}{\mathrm{E}\, X}, \tag{2.8}$$

provided that $\mathrm{E}\,X < \infty$.

The **mean excess loss** for a deductible d is the expected loss in excess of d, conditioned on the loss exceeding the deductible,

$$\mathrm{e}_X(d) = \mathrm{E}(X - d|X > d) = \frac{\mathrm{E}\,X - \mathrm{E}(X \wedge d)}{1 - F(d)}.$$

Empirical estimate of the k-th limited moment based on the individual data is

$$\mathrm{E}\widehat{(X \wedge u)}^k = \frac{1}{n}\left(\sum_{x_j<u} x_j^k + \sum_{x_j \geq u} u^k\right).$$

For grouped data with boundaries $c_0 < c_1 < \cdots < c_r$ we assume that u is such that $c_{j-1} \leq u \leq c_j$. Then we use the histogram (2.3) to estimate the k-th limited moment:

$$\begin{aligned}
\mathrm{E}\widehat{(X \wedge u)}^k &= \sum_{i=1}^{j-1} \int_{c_{i-1}}^{c_i} x^k \frac{n_i}{n\,(c_i - c_{i-1})}\,dx + \int_{c_{j-1}}^{u} x^k \frac{n_j}{n\,(c_j - c_{j-1})}\,dx \\
&+ \int_u^{c_j} u^k \frac{n_j}{n\,(c_j - c_{j-1})}\,dx + \sum_{i=j+1}^{r} \int_{c_{i-1}}^{c_i} u^k \frac{n_i}{n\,(c_i - c_{i-1})}\,dx \\
&= \sum_{i=1}^{j-1} \frac{n_i\,(c_i^{k+1} - c_{i-1}^{k+1})}{n\,(k+1)\,(c_i - c_{i-1})} + \frac{n_j\,(u^{k+1} - c_{j-1}^{k+1})}{n\,(k+1)\,(c_j - c_{j-1})} \\
&+ \frac{n_j\,u^k\,(c_j - u)}{n\,(c_j - c_{j-1})} + \sum_{i=j+1}^{r} \frac{n_i\,u^k}{n}.
\end{aligned}$$

2.1.3 Quantiles

Another important type of quantities derived from a probability distribution are **quantiles** - values of the **quantile function**

$$F^{-1}(\alpha) = \inf\{x : \; F(x) \geq \alpha\}. \tag{2.9}$$

(For a strictly increasing continuous d.f. F it is just the ordinary inverse function.)

Remark. ***100p-th percentile*** *(*$0 < p < 1$*) is defined as any number* π_p *satisfying* $F(\pi_p-) \leq p \leq F(\pi_p)$. *Contrary to the quantile function* (2.9), π_p *is not determined uniquely.*

Denote by $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ the order statistics corresponding to the individual data. Then we estimate the α-quantile as the quantile of the empirical distribution function (2.1):

$$\widehat{F^{-1}(\alpha)} = x_{(r)},$$

where $r = [n\,\alpha] + 1$ ($[x]$ is the integer part of x).

2.2 Parametric severity distributions

In this section we deal with continuous parametric distributions suitable for modelling (not only) the size of an individual claim.

In general, we consider a parametric family of distribution functions $\{F(x,\theta), \theta \in \Theta\}$, where θ is a parameter (vector of parameters) and Θ is the set of all possible parameter values.

2.2.1 Basic severity distributions

We summarize some of the most frequently used parametric models and their characteristics.

Gamma distribution

Gamma distribution has the probability density function (p.d.f.)

$$f(x) = \frac{\left(\frac{1}{\theta}\right)^{\alpha}}{\Gamma(\alpha)}\, x^{\alpha-1}\, \mathrm{e}^{-\frac{x}{\theta}},\ x \ge 0,\ \theta > 0,\ \alpha > 0. \tag{2.10}$$

k-th raw moment is

$$\mu'_k = \theta^k\,(k+\alpha-1)\cdots\alpha. \tag{2.11}$$

(2.11) follows from (2.4) and from the properties of the gamma function

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}\,\mathrm{e}^{-y}\,\mathrm{d}\,y.$$

From (2.11) we easily calculate

$$\mu_1 = \alpha\,\theta, \tag{2.12}$$
$$\mu_2 = \alpha\,\theta^2, \tag{2.13}$$
$$\mu_3 = 2\,\alpha\,\theta^3 \tag{2.14}$$

and

$$\gamma_1 = \frac{2}{\sqrt{\alpha}}. \tag{2.15}$$

An important special case is $\alpha = 1$, for which we obtain **exponential distribution** with p.d.f.

$$f(x) = \frac{1}{\theta}\, \mathrm{e}^{-\frac{x}{\theta}} \tag{2.16}$$

and characteristics $\mu_1 = \theta$, $\mu_2 = \theta^2$, $\gamma_1 = 2$.

Remark. *For α integer,* (2.10) *is the p.d.f. of a sum of α i.i.d. exponentially distributed random variables with the density* (2.16)*.*

Lognormal distribution

Lognormal distribution is derived by exponential transformation from normal distribution: Let Y be a random variable with normal distribution $\mathrm{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Then r.v. $X = \mathrm{e}^Y$ has lognormal distribution (denoted by $\mathrm{LN}(\mu, \sigma^2)$) with the d.f.

$$F(x) = \mathrm{P}(Y \le \log x),\ x > 0$$

and the p.d.f.

$$f(x) = \frac{1}{\sqrt{2\,\pi}\,\sigma\, x}\, \mathrm{e}^{-\frac{(\log x - \mu)^2}{2\,\sigma^2}},\ x > 0. \tag{2.17}$$

Moments of lognormal distribution can be expressed using the moment generating function of normal distribution. **Moment generating function (m.g.f.)** of r.v. Y is defined by

$$\mathrm{M}_Y(r) = \mathrm{E}\,\mathrm{e}^{Yr} \tag{2.18}$$

for $r \in \mathbb{R}$ such that the expected value on the right-hand side of (2.18) exists. For Y with $\mathrm{N}(\mu, \sigma^2)$ it holds

$$\mathrm{M}_Y(r) = \mathrm{e}^{r\mu + r^2\sigma^2/2},\ r \in \mathbb{R}\,. \tag{2.19}$$

Inserting $r = 1, 2, 3$ into (2.19) gives the first three raw moments of lognormal distribution $\mathrm{LN}(\mu, \sigma^2)$. The central moments and the skewness are then derived in the form

$$\mathrm{E}\,X = \mathrm{e}^{\mu + \frac{\sigma^2}{2}}, \tag{2.20}$$

$$\mathrm{Var}\,X = \mathrm{e}^{2\,\mu + \sigma^2} \left(\mathrm{e}^{\sigma^2} - 1\right), \tag{2.21}$$

$$\mathrm{E}\,(X - \mathrm{E}\,X)^3 = \mathrm{e}^{3\,\mu + \frac{3}{2}\sigma^2} \left(\mathrm{e}^{3\,\sigma^2} - 3\,\mathrm{e}^{\sigma^2} + 2\right) \tag{2.22}$$

and

$$\gamma_1 = \frac{\mathrm{e}^{2\,\sigma^2} + \mathrm{e}^{\sigma^2} - 2}{\sqrt{\mathrm{e}^{\sigma^2} - 1}} = \left(\mathrm{e}^{\sigma^2} + 2\right) \sqrt{\mathrm{e}^{\sigma^2} - 1}. \tag{2.23}$$

Pareto distribution

Pareto distribution can be defined as a distribution of r.v. X with possible values from the interval $(\theta, +\infty)$, where $\theta > 0$, by the d.f.

$$F(x) = 1 - \left(\frac{x}{\theta}\right)^{-\alpha}, \ x \geq \theta, \ \alpha > 0. \tag{2.24}$$

The corresponding p.d.f. is

$$f(x) = \alpha\,\theta^{\alpha}\,x^{-\alpha-1}, \ x \geq \theta. \tag{2.25}$$

Moments of the distribution with the d.f. (2.24) are

$$\mu'_k = \frac{\alpha\,\theta^k}{\alpha - k} \text{ for } \alpha > k. \tag{2.26}$$

For $\alpha \leq k$ Pareto distribution has not a finite k-th moment. Particularly, the expected value,

$$\mathrm{E}\,X = \frac{\alpha\,\theta}{\alpha - 1}, \tag{2.27}$$

exists only for $\alpha > 1$, and the variance,

$$\mathrm{Var}\,X = \frac{\theta^2\,\alpha}{(\alpha-2)\,(\alpha-1)^2}, \tag{2.28}$$

is finite only for $\alpha > 2$.

2.2.2 Transformed severity distributions

Lognormal distribution with the p.d.f. (2.17) results from a transformation of a normally distributed random variable. In what follows we consider creating more general families of probability distributions by means of a transformation that lies in raising random variable to a power. In such a way we introduce another parameter to the model. We apply the power transformation to a continuously distributed r.v. X with p.d.f. $f_X(x)$ and d.f. $F_X(x), \ x \geq 0$.

Let $Y = X^{1/\tau}$. Then if $\tau > 0$ we obtain

$$F_Y(y) = F_X(y^{\tau}), \ f_X(y) = \tau\,y^{\tau-1}\,f_X(y^{\tau}), \ y > 0, \tag{2.29}$$

and the distribution of Y is called **transformed**. If $\tau < 0$ it holds

$$F_Y(y) = 1 - F_X(y^{\tau}), \ f_Y(y) = -\tau\,y^{\tau-1}\,f_X(y^{\tau}), \ y > 0, \tag{2.30}$$

and the distribution of Y is called **inverse transformed**. In the special case $\tau = -1$ we speak about **inverse** distribution.

Important examples of parametric families are obtained by transforming a gamma distributed random variable.

Let X have gamma distribution (2.10) with $\theta = 1$ and let $\tau > 0$. Then r.v. $Y = X^{1/\tau}$ has the p.d.f.

$$f(y) = \frac{\tau\, y^{\tau\alpha}\, e^{-y^{\tau}}}{y\,\Gamma(\alpha)},\ y > 0. \tag{2.31}$$

After introducing a scale parameter θ we obtain the p.d.f. of **transformed gamma distribution**:

$$f(y) = \frac{\tau\, \left(\frac{y}{\theta}\right)^{\tau\alpha}\, e^{-(\frac{y}{\theta})^{\tau}}}{y\,\Gamma(\alpha)},\ y > 0. \tag{2.32}$$

It is a 3-parameter family with some well known distributions as special cases: gamma ($\tau = 1$), Weibull ($\alpha = 1$), exponential ($\alpha = \tau = 1$).

Moments of the transformed gamma distribution are expressed by the formula

$$\mathrm{E}\, X^k = \frac{\theta^k\, \Gamma(\alpha + k/\tau)}{\Gamma(\alpha)}. \tag{2.33}$$

(It follows from

$$\mathrm{E}\, X^k = \theta^k \int_0^{\infty} \frac{\tau (\frac{y}{\theta})^{\tau\alpha + k}\, e^{-(\frac{y}{\theta})^{\tau}}}{y\,\Gamma(\alpha)}\, dy$$

by substituting $x = (\frac{y}{\theta})^{\tau}$.)

Raising X to a power $\tau < 0$ gives a p.d.f.

$$f_Y(y) = -\tau\, y^{\tau - 1}\, \frac{y^{\tau\alpha}\, e^{-y^{\tau}}}{y^{\tau}\Gamma(\alpha)}. \tag{2.34}$$

We substitute the negative parameter τ by its opposite value and again we introduce a scale parameter θ. The resulting density

$$f(y) = \frac{\tau\, \left(\frac{\theta}{y}\right)^{\alpha\tau}\, e^{-(\frac{\theta}{y})^{\tau}}}{y\,\Gamma(\alpha)},\ y > 0,\ \tau > 0 \tag{2.35}$$

is a p.d.f. of so called **inverse transformed gamma distribution**.

Special cases are: inverse gamma ($\tau = 1$), inverse Weibull ($\alpha = 1$), inverse exponential ($\alpha = \tau = 1$).

Moments of the inverse transformed gamma distribution are given by

$$\mathrm{E}\,X^k = \frac{\theta^k\,\Gamma(\alpha - k/\tau)}{\Gamma(\alpha)},\;\; k < \alpha\tau. \tag{2.36}$$

Pareto distribution is a special case of so called **transformed beta distribution** with p.d.f.

$$f(x) = \frac{\Gamma(\alpha+\tau)}{\Gamma(\alpha)\,\Gamma(\tau)} \frac{\gamma\,(x/\theta)^{\gamma\tau}}{x\;[1+(x/\theta)^{\gamma}]^{\alpha+\tau}}. \tag{2.37}$$

Moments of this distribution are given by

$$\mathrm{E}\,X^k = \frac{\theta^k\,\Gamma(\tau + k/\gamma)\,\Gamma(\alpha - k/\gamma)}{\Gamma(\alpha)\,\Gamma(\tau)} \tag{2.38}$$

and are finite only in case $k < \alpha\gamma$.

Setting $\gamma = 1$, $\tau = 1$ we obtain Pareto distribution with p.d.f.

$$f(x) = \alpha\,\theta^{\alpha}\,(\theta + x)^{-\alpha-1},\;\; x \geq 0. \tag{2.39}$$

Remark. (2.39) *is a density of Pareto distribution with the support* $[0,+\infty)$. *It can be derived from* (2.25) *by changing the location. (*(2.39) *is a density of r.v.* $X - \theta$*, where* X *has the p.d.f.* (2.25)*.)*

2.2.3 Tail behavior

The tail behavior of a random variable is expressed by the **survival function**

$$S(x) = 1 - F(x) = P(X > x) \tag{2.40}$$

considered for $x \to \infty$. When modelling claim sizes by r.v. X, we are usually interested in the shape of (2.40) for high values of x (the right tail of the distribution), since it describes the probability of very high potential losses. We want to choose a model which does not underestimate the values $S(x)$ for high x when compared with the information available from the data on historical losses.

The tail behavior of two probability distributions is similar, if the ratio of their survival functions tends to a constant non-zero limit as $x \to \infty$. The same holds for the ratio of their probability density functions, since

$$\lim_{x\to\infty} \frac{S_X(x)}{S_Y(x)} = \lim_{x\to\infty} \frac{f_X(x)}{f_Y(x)}.$$

We illustrate the comparison of probability distributions according their tail behavior for the following examples:

1) Pareto distribution

$$F(x) = 1 - \left(\frac{x}{a}\right)^{-\beta}, \; x \geq a, \; a > 0, \beta > 0,$$
$$f(x) = a^{\beta} \, \beta \, x^{-\beta-1},$$

2) Gamma distribution

$$f(x) = \frac{\left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{x}{\theta}}}{x \, \Gamma(\alpha)}, x \geq 0, \; \alpha > 0, \theta > 0,$$

3) Lognormal distribution

$$f(x) = \frac{1}{x \, \sigma \sqrt{2\pi}} \exp\Big(-\frac{(\log x - \mu)^2}{2 \, \sigma^2}\Big).$$

We obtain the following comparisons:

1) Gamma vs. Pareto:

$$\lim_{x \to \infty} \frac{x^{\alpha-1} \, e^{-x/\theta}}{x^{-(\beta+1)}} =$$
$$\lim_{x \to \infty} \exp\Big[(\alpha - 1) \, \log x - \frac{x}{\theta} + (\beta + 1) \, \log x\Big] = 0.$$

Pareto distribution has heavier tail than Gamma distribution.

2) lognormal vs. Gamma:

$$\lim_{x \to \infty} \frac{x^{-1} \, \exp\left[-\frac{1}{2\sigma^2} \, (\log x - \mu)^2\right]}{x^{\alpha-1} \, e^{-x/\theta}} =$$
$$\lim_{x \to \infty} \exp\left[-\frac{1}{2\sigma^2} \, (\log x - \mu)^2 - \alpha \, \log x + \frac{x}{\theta}\right] = +\infty.$$

Lognormal distribution has heavier tail than Gamma distribution.

3) Pareto vs. lognormal:

$$\lim_{x \to \infty} \frac{x^{-(\beta+1)}}{x^{-1} \, \exp\left[-\frac{1}{2\sigma^2} \, (\log x - \mu)^2\right]} =$$
$$\lim_{x \to \infty} \exp\left[-\beta \, \log x + \frac{1}{2\sigma^2} \, (\log x - \mu)^2\right] = +\infty.$$

Pareto distribution has heavier tail than lognormal distribution.

2.3 Claim frequency models

We consider discrete distributions defined on non-negative integer values (**counting distributions**), with **probability function (p.f.)**

$$p_k = \mathrm{P}(N = k),\ \ k = 0, 1, \ldots \tag{2.41}$$

N in (2.41) is interpreted as the number of events (losses, claims) in a fixed time period recorded for a homogeneous group of insureds (collective model) or for one insurance policy (individual model).

When dealing with discrete probability distributions, **probability generating function (p.g.f.)** defined by

$$P_N(z) = \mathrm{E}\, z^N = \sum_{k=0}^{\infty} p_k\, z^k, \tag{2.42}$$

is often used. Observe that the power series on the right-hand size of (2.42) converges absolutely at least for z with $|z| \leq 1$.

We first introduce important types of counting distributions together with some of their properties. Subsequently we introduce a general framework for classification of these distributions and their possible modifications.

2.3.1 Basic frequency distributions

Poisson distribution

The Poisson distribution has p.f.

$$p_k = e^{-\lambda}\, \frac{\lambda^k}{k!},\ \ k = 0, 1, \ldots ; \lambda > 0. \tag{2.43}$$

For mean and variance of the Poisson distribution it holds

$$\begin{aligned}
\mathrm{E}\, N &= \sum_{k=0}^{\infty} k\, p_k = \mathrm{e}^{-\lambda}\ \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda, \\
\mathrm{Var}\, N &= \mathrm{E}\left[N\,(N-1)\right] + \mathrm{E}\, N - (\mathrm{E}\, N)^2 \\
&= \sum_{k=0}^{\infty} k\,(k-1)\, p_k + \lambda - \lambda^2 \\
&= \mathrm{e}^{-\lambda}\ \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.
\end{aligned} \tag{2.44}$$

P.g.f. of the Poisson distribution is

$$P(z) = \sum_{k=0}^{\infty} z^k\, p_k = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda\, z)^k}{k!} = e^{\lambda(z-1)}. \tag{2.45}$$

In modeling claim counts the following properties of the Poisson distribution are useful:

Theorem 2.1 ("additivity" for independent random variables)**.** *Let* $N_1,\ldots,$ N_n *be independent Poisson distributed random variables with parameters* $\lambda_1,\ldots,\lambda_n$*. Then* $N = \sum_{i=1}^{n} N_i$ *has Poisson distribution with parameter* $\lambda = \sum_{i-1}^{n} \lambda_i$*.*

Proof.

$$\begin{aligned} P_N(z) &= \mathrm{E}\, z^{\sum_{i=1}^{n} N_i} = \prod_{i=1}^{n} P_{N_i}(z) = \prod_{i=1}^{n} \mathrm{e}^{\lambda_i(z-1)} \\ &= \mathrm{e}^{\sum_{i=1}^{n} \lambda_i(z-1)} = \mathrm{e}^{\lambda\,(z-1)}\,. \end{aligned}$$

The statement follows from the fact that a distribution is uniquely defined by its p.g.f. □

Theorem 2.1 says we can use the well tractable Poisson model for the total number of claims originating from independent contracts, subportfolios or lines of business, where each partial number of claims may be represented by Poisson distribution with different expected value.

Theorem 2.2 (Poisson distribution of "classified" events)**.** *Suppose that the number of events* N *is a Poisson distributed r.v. with mean* λ *and that each event can be classified into one of* m *types with probabilities* $p_1,\ldots,p_m$ *(*$\sum_{i=1}^{m} p_i = 1$*). Then* $N_1,\ldots,N_m$*, representing numbers of events of types* $1,\ldots,m$ *respectively, are mutually independent Poisson distributed random variables with means* $\lambda\, p_1,\ldots,\lambda\, p_m$ *respectively.*

Proof. For fixed $N = n$, the conditional joint distribution of $(N_1,\ldots,N_m)$ is multinomial with parameters $n, p_1,\ldots,p_m$:

$$\mathrm{P}\left(N_1 = n_1,\ldots,N_m = n_m \mid N = n\right) = \frac{n!}{n_1!\ldots n_m!} p_1^{n_1}\ldots p_m^{n_m}.$$

We can also write

$$\mathrm{P}\left(N_1 = n_1,\ldots,N_m = n_m\right) = \mathrm{P}\left(N_1 = n_1,\ldots,N_m = n_m \mid N = n\right)\, \mathrm{P}\left(N = n\right),$$

where $n = n_1 + \cdots + n_m$. We have

$$\begin{aligned}
\mathrm{P}\left(N_1 = n_1, \ldots, N_m = n_m\right) &= \frac{n!}{n_1! \ldots n_m!} p_1^{n_1} \ldots p_m^{n_m} \; \mathrm{e}^{-\lambda} \, \frac{\lambda^n}{n!} \\
&= \prod_{i=1}^{m} \mathrm{e}^{-\lambda \, p_i} \, \frac{(\lambda \, p_i)^{n_i}}{n_i!} .
\end{aligned} \tag{2.46}$$

At the same time, for fixed $N = n$, the conditional marginal distribution of N_i is binomial with parameters n, p_i:

$$\mathrm{P}\left(N_i = n_i \mid N = n\right) = \binom{n}{n_i} p_i^{n_i} \; (1 - p_i)^{n - n_i} \, , \;\; n_i = 0, \ldots, n,$$

from where follows the unconditional Poisson distribution of N_i:

$$\begin{aligned}
\mathrm{P}\left(N_i = n_i\right) &= \sum_{n=n_i}^{\infty} \mathrm{P}\left(N_i = n_i \mid N = n\right) \; \mathrm{P}\left(N = n\right) \\
&= \sum_{n=n_i}^{\infty} \binom{n}{n_i} p_i^{n_i} \; (1 - p_i)^{n - n_i} \; \mathrm{e}^{-\lambda} \, \frac{\lambda^n}{n!} \\
&= \mathrm{e}^{-\lambda} \, \frac{(\lambda \, p_i)^{n_i}}{n_i!} \sum_{n=n_i}^{\infty} \frac{\left[\lambda \, (1 - p_i)\right]^{n - n_i}}{(n - n_i)!} \\
&= \mathrm{e}^{-\lambda} \, \frac{(\lambda \, p_i)^{n_i}}{n_i!} \; \mathrm{e}^{\lambda (1 - p_i)} = \mathrm{e}^{-\lambda \, p_i} \, \frac{(\lambda \, p_i)^{n_i}}{n_i!} .
\end{aligned} \tag{2.47}$$

(2.47) proves Poisson distribution of $N_1, \ldots, N_m$, (2.46) together with (2.47) their independence.

□

Negative binomial distribution

The negative binomial distribution has p.f.

$$p_k = \binom{k + h - 1}{k} \left(\frac{1}{1 + \beta}\right)^h \left(\frac{\beta}{1 + \beta}\right)^k , \;\; k = 0, 1, \ldots ; \;\; h > 0, \beta > 0, \tag{2.48}$$

where we use for $x \in \mathbb{R}$ and $k \in \{0, 1, \ldots\}$ the notation

$$\binom{x}{k} = \frac{x \, (x - 1) \cdots (x - k + 1)}{k!} = \frac{\Gamma(x + 1)}{\Gamma(k + 1) \, \Gamma(x - k + 1)} . \tag{2.49}$$

It is sometimes useful to use another parametrization with parameters $h > 0$ and $0 < p < 1$ with p.f. given by

$$p_k = \binom{k+h-1}{k} p^h (1-p)^k, \; k = 0, 1, \ldots \tag{2.50}$$

(2.50) can be further rewritten in the form

$$p_k = \binom{-h}{k} p^h (p-1)^k, \; k = 0, 1, \ldots, \tag{2.51}$$

when we use notation (2.49) with $x = -h$.

We apply (2.51) to derive expected value and variance of negative binomial distribution.

$$\begin{aligned} \mathrm{E}\, N = \sum_{k=1}^{\infty} k\, p_k &= \frac{h\,(1-p)}{p} \sum_{k=0}^{\infty} \frac{(-h-1)\cdots(-h-1-k+1)}{k!} p^{h+1} (p-1)^k \\ &= \frac{h\,(1-p)}{p}. \end{aligned} \tag{2.52}$$

The sum on the right-hand side of (2.52) equals 1, since it is the sum of p.f. values for negative binomial distribution with parameters $h+1$, p.

Similarly, we have

$$\mathrm{E}\, N\,(N-1) = \sum_{k=2}^{\infty} k\,(k-1)\, p_k = \frac{h(h+1)(1-p)^2}{p^2},$$

so the variance is (see (2.44))

$$\begin{aligned} \mathrm{Var}\, N &= \frac{h(h+1)(1-p)^2}{p^2} + \frac{h(1-p)}{p} - \frac{h^2(1-p)^2}{p^2} \\ &= \frac{h(1-p)}{p^2}. \end{aligned} \tag{2.53}$$

In the parametrization (2.48) we obtain from (2.52) and (2.53)

$$\mathrm{E}\, N = h\,\beta \tag{2.54}$$

$$\mathrm{Var}\, N = h\,\beta\,(1+\beta). \tag{2.55}$$

(2.51) is also suitable for derivation of p.g.f.:

$$\begin{aligned} P(z) &= p^h \sum_{k=0}^{\infty} \binom{-h}{k} [z\,(p-1)]^k \\ &= \left[\frac{p}{1-z\,(1-p)}\right]^h, \end{aligned} \tag{2.56}$$

where $|z| < (1-p)^{-1}$.

Negative binomial distribution with $h = 1$ is called **geometric distribution**. It has p.f.

$$p_k = \frac{1}{1+\beta}\left(\frac{\beta}{1+\beta}\right)^k, \; k = 0, 1, \ldots; \; \beta > 0. \tag{2.57}$$

Binomial distribution

P.f. of binomial distribution has the form

$$p_k = \binom{m}{k} p^k (1-p)^{m-k}, \; k = 0, 1, \ldots, m, \; 0 < p < 1. \tag{2.58}$$

Its expected value and variance are

$$\begin{aligned} \mathrm{E}\, N &= m\, p, \\ \mathrm{Var}\, N &= m\, p\, (1-p). \end{aligned}$$

Binomial distribution is not commonly used in modeling insurance claim counts. We can arrive at the binomial model in the individual model setting, where we consider m mutually independent risks (insurance policies) in the portfolio. We suppose that the number of claims for one policy has an alternative distribution with parameter $0 < p < 1$. Then the total number of claims in the portfolio is a random variable with binomial distribution (2.58).

The number of policies is usually large while the probability of claim occurrence may be quite close to zero. This is why we consider an approximation for the total number of claims distribution resulting from taking the limit of (2.58) as $m \to \infty$ and $p \to 0$ in such a way that the product $m\, p$ remains constant ($m\, p = \lambda$). This gives

$$\begin{aligned} \lim_{\substack{m\to\infty, p\to 0 \\ mp=\lambda}} \mathrm{P}\,(N = k) &= \lim_{\substack{m\to\infty, p\to 0 \\ mp=\lambda}} \frac{m\,(m-1)\cdots(m-k+1)}{k!}\, p^k \left(1 - \frac{\lambda}{m}\right)^{m-k} \\ &= \frac{\lambda^k}{k!}\, \mathrm{e}^{-\lambda}\,. \end{aligned}$$

We thus obtain Poisson distribution as the appropriate approximation for the total claim number when the portfolio size is large and the probability of one individual claim is small.

2.3.2 The $(a, b, 0)$ class

All the distributions mentioned in the previous section belong to a general class of two-parametric distributions, called the $(a, b, 0)$ class.

Definition 2.1. *Discrete random variable distribution with the probability function* $\{p_k,\ k = 0, 1, \dots\}$ *is a member of the* $(\mathbf{a}, \mathbf{b}, \mathbf{0})$ ***class****, provided that there exist constants a, b such that*

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k},\ \ k = 1, 2, \dots. \tag{2.59}$$

Note that the probability p_0 is determined by (2.59) through the condition $\sum_{k=0}^{\infty} p_k = 1$.

For the distributions from section 2.3.1 we obtain the following values of parameters a, b:

Table 2.1: Members of the $(a, b, 0)$ class

Distribution	a	b
Poisson	0	λ
Negative Binomial	$\frac{\beta}{1+\beta}$	$(h-1)\frac{\beta}{1+\beta}$
Binomial	$-\frac{p}{1-p}$	$(m+1)\frac{p}{1-p}$
Geometric	$\frac{\beta}{1+\beta}$	0

It can be shown that the distributions in table 2.1 are the only discrete distributions satisfying (2.59).

Formula (2.59) can be rewritten as

$$k\,\frac{p_k}{p_{k-1}} = a\,k + b,\ \ k = 1, 2, \dots$$

Assume that we observe number of claims during certain period of time for n policies. Let n_k be the number of policies with k recorded claims, $k = 0, 1, \dots$

We can estimate the ratio $\frac{p_k}{p_{k-1}}$ by $\frac{n_k}{n_{k-1}}$. This suggests a graphical way of indicating which of the distributions should be selected: We plot $\left[k, k\,\frac{n_k}{n_{k-1}}\right]$ for $k = 0, 1, \dots$ The points should form a straight line, where the slope is 0 for the Poisson distribution, it is negative for the binomial and positive for the negative binomial distribution.

2.3.3 The $(a, b, 1)$ class

We explain a generalization of the $(a, b, 0)$ class, that enables a better fit of the probability at zero.

Definition 2.2. *Discrete random variable distribution with probability function* $\{p_k,\ k = 0, 1, \dots\}$ *is a member of the* $(\mathbf{a}, \mathbf{b}, \mathbf{1})$ ***class****, provided that there exist constants* a, b *such that*

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k},\ \ k = 2, 3, \dots. \tag{2.60}$$

The distribution for $k = 1, \dots$ has the same shape as the corresponding one from $(a, b, 0)$ class in the sense that the probabilities are the same up to a constant of proportionality. $\sum_{k=1}^{\infty} p_k$ can be set equal to any number in the interval $(0, 1]$. The remaining probability is $p_0 = 1 - \sum_{k=1}^{\infty} p_k$.

When we set $p_0 = 0$, the distribution is called **zero-truncated (ZT)**. When $p_0 > 0$, the distribution is called **zero-modified (ZM)**.

The zero-modified distribution can be viewed as a mixture of a zero-truncated distribution and a degenerate distribution with all the probabilty at zero.

To show this, we denote by $\{p_k,\ k = 0, 1, \dots\}$ the distribution from the $(a, b, 0)$ class and by $\{p_k^M,\ k = 0, 1, \dots\}$ the corresponding distribution from the $(a, b, 1)$ class. The probability generating functions of these distributions are

$$P(z) = \sum_{k=0}^{\infty} p_k\, z^k,\ \ P^M(z) = \sum_{k=0}^{\infty} p_k^M\, z^k.$$

It holds

$$p_k^M = c\, p_k,\ \ k = 1, 2, \dots$$

and p_0^M is an arbitrary number. Then

$$\begin{aligned} P^M(z) &= p_0^M + \sum_{k=1}^{\infty} p_k^M\, z^k \\ &= p_0^M + c \sum_{k=1}^{\infty} p_k\, z^k = p_0^M + c\,[P(z) - p_0]. \end{aligned}$$

Since $P^M(1) = P(1) = 1$,

$$1 = p_0^M + c\,(1 - p_0)$$

resulting in

$$c = \frac{1 - p_0^M}{1 - p_0},$$

hence

$$p_k^M = \frac{1 - p_0^M}{1 - p_0}\, p_k,\ \ k = 1, 2, \dots \tag{2.61}$$

We can write

$$P^M(z) = p_0^M + \frac{1-p_0^M}{1-p_0}\,[P(z)-p_0] = \left[1 - \frac{1-p_0^M}{1-p_0}\right] + \frac{1-p_0^M}{1-p_0}\,P(z),$$

this is a weighted average of the probability generating functions of the degenerate distribution and the corresponding $(a, b, 0)$ member. □

The zero-truncated distribution can be viewed as a special case of the zero-modified distribution with $p_0^M = 0$. Then we obtain

$$p_k^T = \frac{p_k}{1-p_0},\ \ k = 1, 2, \ldots \tag{2.62}$$

We give a summary of the $(a, b, 1)$ class:

1. For $\lambda > 0$ we derive ZT or ZM Poisson distribution by setting $a = 0$, $b = \lambda$ in (2.60) and choosing $p_0 = 0$ or $0 < p_0 < 1$ arbitrary.

2. For $m \in \mathbb{N}$ and $0 < p < 1$ we derive ZT or ZM binomial distribution by setting $a = -\frac{p}{1-p}$, $b = (m+1)\,\frac{p}{1-p}$ in (2.60) and choosing $p_0 = 0$ or $0 < p_0 < 1$ arbitrary.

3. Non-trivial modification is obtained in the case of ZT negative binomial distribution. The set of possible values for the parameter h in this case can be actually extended from $(0, +\infty)$ to $(-1, +\infty)$.

   From (2.60) we derive

   $$p_k = p_1 \left(\frac{\beta}{1+\beta}\right)^{k-1} \left(1 + \frac{h-1}{k}\right) \cdots \left(1 + \frac{h-1}{2}\right),\ \ k = 2, 3, \ldots$$

   It is seen that for any $p_1 > 0$ it holds $p_k > 0$ for all $k = 2, 3, \ldots$, if and only if $h > -1$.

   Also, for $h > -1$

   $$\begin{aligned}\sum_{k=2}^{\infty} p_k &= p_1 \sum_{k=2}^{\infty} \frac{1}{k} \left(\frac{\beta}{1+\beta}\right)^{k-1} \binom{h+1+k-2}{k-1} \\ &< p_1 \sum_{k=2}^{\infty} \left(\frac{\beta}{1+\beta}\right)^{k-1} \binom{-(r+1)}{k-1} (-1)^{k-1} < \infty.\end{aligned}$$

   So it is possible to define p_1 under the condition $\sum_{k=1}^{+\infty} p_k = 1$ by

   $$p_1 = \left[\sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{\beta}{1+\beta}\right)^{k-1} \binom{h+1+k-2}{k-1}\right]^{-1}.$$

We call the resulting distribution **"extended" truncated negative binomial distribution**.

A special case for $r = 0$ is the **logarithmic distribution** with p.f.

$$p_k = \frac{1}{k\,\log(1+\beta)}\left(\frac{\beta}{1+\beta}\right)^k,\ \ k = 1, 2, \ldots \tag{2.63}$$

2.3.4 Compound frequency models

A large class of distributions can be created by the process of compounding any two discrete distributions.

Let N be a r.v. with p.f.

$$p_k = \mathrm{P}(N = k),\ \ k = 0, 1, \ldots \tag{2.64}$$

and let $M_1, M_2, \ldots$ be i.i.d. random variables, independent of N, with p.f.

$$f_k = \mathrm{P}(M = k),\ \ k = 0, 1, \ldots \tag{2.65}$$

We denote the corresponding probability generating functions by $P_1(z)$ and $P_2(z)$:

$$P_1(z) = \mathrm{E}\, z^N = \sum_{k=0}^{\infty} p_k\, z^k, \tag{2.66}$$

$$P_2(z) = \mathrm{E}\, z^M = \sum_{k=0}^{\infty} f_k\, z^k. \tag{2.67}$$

Then

$$S = \sum_{i=1}^{N} M_i$$

has a **compound distribution**. We denote its p.f. by

$$g_k = \mathrm{P}(S = k),\ \ k = 0, 1, \ldots \tag{2.68}$$

Its p.g.f. is

$$P(z) = \mathrm{E}\, z^S = \sum_{k=0}^{\infty} g_k\, z^k. \tag{2.69}$$

For (2.69) we derive

$$P(z) = \mathrm{E}\,\mathrm{E}\left[z^S | N\right] = \sum_{k=0}^{\infty} p_k\ \mathrm{E}\left[z^S | N = k\right]$$
$$= \sum_{k=0}^{\infty} p_k\ \mathrm{E}\left[z^{M_1+\cdots+M_k} | N = k\right] = \sum_{k=0}^{\infty} p_k\, P_2^k(z).$$

So, we can write

$$P(z) = P_1\left(P_2(z)\right) \tag{2.70}$$

for z such that p.g.f. (2.69) exists.

We call the distribution of N **primary distribution** and the distribution of M **secondary distribution**.

The relationship (2.69) is essential for the derivation of the recursive formula for a compound frequency distribution, usually referred to as Panjer formula.

Theorem 2.3 (Panjer recursive formula for $(a, b, 0)$ class)**.** *If the primary distribution is a member of the $(a, b, 0)$ class, then*

$$g_k = \frac{1}{1 - a\, f_0} \sum_{j=1}^{k} \left(a + \frac{b\, j}{k}\right) f_j\, g_{k-j},\ \ k = 1, 2, \ldots \tag{2.71}$$

Proof. From (2.59) it follows

$$k\, p_k = a\,(k-1)\, p_{k-1} + (a+b)\, p_{k-1},\ \ n = 1, 2, \ldots. \tag{2.72}$$

Multiplying each side of (2.72) by $\left[P_2(z)\right]^{k-1}\ P_2'(z)$ and summing over k yields

$$\sum_{k=1}^{\infty} k\, p_k \left[P_2(z)\right]^{k-1}\ P_2'(z) =$$

$$a \sum_{k=1}^{\infty} (k-1) p_{k-1} \left[P_2(z)\right]^{k-1}\ P_2'(z) + (a+b) \sum_{k=1}^{\infty} p_{k-1} \left[P_2(z)\right]^{k-1}\ P_2'(z).$$

Therefore,

$$P'(z) = a\, P'(z)\, P_2(z) + (a+b)\, P(z)\, P_2'(z).$$

Comparing the coefficients of z^{k-1} we obtain

$$\begin{aligned} k\, g_k &= a \sum_{j=0}^{k} (k-j)\, f_j\, g_{k-j} + (a+b) \sum_{j=0}^{k} j\, f_j\, g_{k-j} \\ &= a\, k\, f_0\, g_k + a \sum_{j=1}^{k} (k-j)\, f_j\, g_{k-j} + (a+b) \sum_{j=1}^{k} j\, f_j\, g_{k-j} \\ &= a\, k\, f_0\, g_k + a\, k \sum_{j=1}^{k} f_j\, g_{k-j} + b \sum_{j=1}^{k} j\, f_j\, g_{k-j}. \end{aligned}$$

Thus,

$$g_k = a\, f_0\, g_k + \sum_{j=1}^{k} \left(a + \frac{bj}{k} \right) f_j\, g_{k-j}.$$

□

Formula (2.71) requires the starting value g_0. This can be computed as

$$\begin{aligned} g_0 &= \sum_{k=0}^{\infty} \mathrm{P}\left(S = 0 | N = k\right)\, \mathrm{P}(N = k) \\ &= \sum_{k=0}^{\infty} \mathrm{P}\left(M_1 + \cdots + M_k = 0\right)\, \mathrm{P}(N = k) \\ &= \sum_{k=0}^{\infty} (f_0)^k\, p_k = P_1(f_0). \end{aligned}$$

Remark. *When $f_0 = 0$, then $g_0 = P_1(0) = p_0$.*

It is possible to derive a similar result in the more general case when the primary distribution is from the $(a, b, 1)$ class.

Theorem 2.4 (Panjer recursive formula for $(a, b, 1)$ class)**.** *If the primary distribution is a member of the $(a, b, 1)$ class, then*

$$g_k = \frac{1}{1 - a\, f_0} \left([p_1 - (a + b)\, p_0]\, f_k + \sum_{j=1}^{k} \left(a + \frac{b\, j}{k} \right) f_j\, g_{k-j} \right), \;\; k = 1, 2, \ldots \tag{2.73}$$

Proof. From (2.60) we have

$$k\, p_k = a\, (k - 1)\, p_{k-1} + (a + b)\, p_{k-1}, \;\; k = 2, 3, \ldots \tag{2.74}$$

Multiplying each side of (2.74) by $[P_2(z)]^{k-1}\; P_2'(z)$ and summing over k yields

$$\sum_{k=2}^{\infty} k\, p_k\, [P_2(z)]^{k-1}\; P_2'(z) =$$
$$a \sum_{k=2}^{\infty} (k-1) p_{k-1}\, [P_2(z)]^{k-1}\; P_2'(z) + (a + b) \sum_{k=2}^{\infty} p_{k-1}\, [P_2(z)]^{k-1}\; P_2'(z).$$

Since

$$P'(z) = \sum_{k=1}^{\infty} k\, p_k\; [P_2'(z)]^{k-1}\; P_2'(z),$$

we obtain

$$P'(z) - p_1 P_2'(z) = a P'(z) P_2(z) + (a+b) P_2'(z) [P(z) - p_0] .$$

After rearranging,

$$P'(z) = a P_2(z) P'(z) + (a+b) P(z) P_2'(z) + [p_1 - (a+b) p_0] P_2'(z).$$

Comparing the coefficients of z^{k-1} we obtain

$$k g_k = a \sum_{j=0}^{k} (k-j) f_j g_{k-j} + (a+b) \sum_{j=0}^{k} j f_j g_{k-j} + [p_1 - (a+b) p_0] k f_k.$$

Therefore,

$$g_k = a f_0 g_k + \sum_{j=1}^{k} \left(a + \frac{bj}{k} \right) f_j g_{k-j} + [p_1 - (a+b) p_0] f_k.$$

□

2.3.5 Mixed frequency models

We arrive at a mixed frequency model in the situation, when the distribution of the counting r.v. N depends on a parameter that is viewed as a realization of a random variable. We start with an illustration of the concept with a simple example.

Example 2.1. *Let N denote the number of accidents recorded by a driver from a group of insureds in an automobile insurance. The drivers are classified as 'good' and 'bad'. We assume that the yearly number of accidents of a good driver follows Poisson distribution with parameter λ_1 and the same variable for a bad driver follows Poisson distribution with parameter λ_2. The distribution of good and bad drivers in the portfolio is modelled by a r.v. Λ with p.f.*

$$\mathrm{P}(\Lambda = \lambda_1) = p,$$
$$\mathrm{P}(\Lambda = \lambda_2) = 1 - p$$

for some $0 < p < 1$. Then the distribution of the number of accidents recorded by a randomly chosen driver from the group is

$$\begin{aligned} \mathrm{P}(N = k) &= \mathrm{P}(N = k|\Lambda = \lambda_1)\ \mathrm{P}(\Lambda = \lambda_1) + \mathrm{P}(N = k|\Lambda = \lambda_2)\ \mathrm{P}(\Lambda = \lambda_2) \\ &= p \frac{\lambda_1^k}{k!} \mathrm{e}^{-\lambda_1} + (1-p) \frac{\lambda_2^k}{k!} \mathrm{e}^{-\lambda_2}, \ k = 0, 1, \ldots \end{aligned}$$

and its p.g.f. is easily derived as

$$P_N(z) = p\ \mathrm{e}^{\lambda_1 (z-1)} + (1-p)\ \mathrm{e}^{\lambda_2 (z-1)} .$$

Example 2.1 is a special case of **Poisson mixture** - mixed model where the distribution of N conditional on $\Lambda = \lambda$ is Poisson distribution (2.43), while the unconditional distribution is in the general case expressed by

$$\mathrm{P}(N=k) = \mathrm{E}\,\mathrm{P}\,(N=k|\Lambda) = \mathrm{E}\,\frac{\Lambda^k}{k!}\,\mathrm{e}^{-\Lambda} = \int \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,dU(\lambda), \tag{2.75}$$

where $U(\lambda)$ is the d.f. of a positive r.v. Λ (with possibly discrete or continuous distribution). P.g.f. of (2.75) is

$$P_N(z) = \mathrm{E}\,z^N = \mathrm{E}\,\mathrm{E}\left[z^N|\Lambda\right] = \mathrm{E}\,\mathrm{e}^{\Lambda\,(z-1)} = \int \mathrm{e}^{\lambda(z-1)}\,dU(\lambda). \tag{2.76}$$

For the expected value and variance of a Poisson mixture it holds

$$\mathrm{E}\,N = \mathrm{E}\,\mathrm{E}\,[N|\Lambda] = \mathrm{E}\,\Lambda, \tag{2.77}$$

$$\mathrm{Var}\,N = \mathrm{E}\,\mathrm{Var}\,[N|\Lambda] + \mathrm{Var}\,\mathrm{E}\,[N|\Lambda] = \mathrm{E}\,\Lambda + \mathrm{Var}\,\Lambda. \tag{2.78}$$

From (2.77) and (2.78) it is seen that Poisson mixture (with a non-degenerate distribution of Λ) has a higher variance in comparison with the Poisson model.

An important example of Poisson mixture is **Poisson-gamma model**. In this case we use gamma distribution with the p.d.f. (2.10) as the distribution of the random parameter Λ. The derivation of the unconditional distribution of N in Poisson-gamma model follows from (2.75):

$$\begin{aligned}\mathrm{P}(N=k) &= \int_0^\infty \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,\frac{\lambda^{\alpha-1}}{\theta^\alpha\,\Gamma(\alpha)}\,\mathrm{e}^{-\frac{\lambda}{\theta}}\,d\lambda \\ &= \frac{1}{k!\,\theta^\alpha\,\Gamma(\alpha)}\int_0^\infty \lambda^{k+\alpha-1}\,\mathrm{e}^{-\lambda\left(1+\frac{1}{\theta}\right)}\,d\lambda \\ &= \frac{1}{k!\,\theta^\alpha\,\Gamma(\alpha)}\,\frac{1}{\left(1+\frac{1}{\theta}\right)^{k+\alpha}}\Gamma(k+\alpha) \\ &= \frac{(k+\alpha-1)\cdots\alpha}{k!}\left(\frac{\theta}{1+\theta}\right)^k\left(\frac{1}{1+\theta}\right)^\alpha.\end{aligned} \tag{2.79}$$

In (2.79) we find out that the unconditional distribution of N in Poisson-gamma model is negative binomial distribution (2.48) with parameters $h = \alpha$ and $\beta = \theta$.

Remark. *By introducing parameter $\nu = \alpha\,\theta$ we obtain p.f. of N in the form*

$$\mathrm{P}(N=k) = \binom{k+\alpha-1}{k}\left(\frac{\nu}{\alpha+\nu}\right)^k\left(\frac{\alpha}{\alpha+\nu}\right)^\alpha,\ k=0,1,\ldots, \tag{2.80}$$

which represents a negative binomial distribution with $\mathrm{E}\,N = \nu$ *and* $\mathrm{Var}\,N = \nu + \frac{\nu^2}{\alpha}$. *It can be shown that the limit of* (2.80) *for* α *tending to* $+\infty$ *is Poisson distribution p.f. with parameter* ν.

Remark. *Another example of a mixed discrete model is* ***binomial-beta*** *model, based on the assumption of* N *having conditionally (for given* p*) binomial distribution* (2.58)*, while the parameter* p *is considered as a realization of a r.v. with beta distribution, i.e. a continuous distribution on* $(0,1)$ *with p.d.f.*

$$u(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, p^{a-1}\,(1-p)^{b-1},\ \ 0 < p < 1;\ \ a > 0,\ \ b > 0.$$

2.4 Aggregate loss models

We are interested in the total amount of claims incurred in a specific period of time (for one insurance policy or the whole portfolio of risks with similar characteristics). First, we summarize some properties of compound distributions. The content of this section then lies mainly in the description of selected methods used in practice for computing the d.f. of the aggregate loss random variable.

2.4.1 Compound model for aggregate claims

In section 2.3.4 we dealt with compound frequency models. The concept of compounding is applicable also in the situation, where the secondary distribution is not discrete. Let N denote the number of claims occurred in a given period, let $X_1, X_2, \ldots$ be the sizes of individual claims. The aggregate loss

$$S = \sum_{i=1}^{N} X_i \tag{2.81}$$

has a **compound distribution**, if $\{X_1, X_2, \ldots\}$ are i.i.d. random variables independent on N. (It means we have to work with a homogeneous group of risks, where the individual claim sizes can be considered as identically distributed.) The expected value and variance expressed by means of corresponding moments of the primary and the secondary distributions are

$$\mathrm{E}\,S = \mathrm{E}\,\mathrm{E}[S|N] = \mathrm{E}\,N\ \mathrm{E}\,X, \tag{2.82}$$

$$\begin{aligned}\mathrm{Var}\,S &= \mathrm{E}\,\mathrm{Var}[S|N] + \mathrm{Var}\,\mathrm{E}[S|N] \\ &= \mathrm{E}\,N\ \mathrm{Var}\,X + (\mathrm{E}\,X)^2\ \mathrm{Var}\,N. \end{aligned} \tag{2.83}$$

It is sometimes convenient to use the expression of moments based on the m.g.f. of S. By arguments analogical to (2.70) it is deduced

$$M_S(z) = P_N\left(M_X(z)\right) \tag{2.84}$$

for appropriate values of argument z.

M.g.f. (2.84) can be used to calculate $\mathrm{E}\,S$, $\mathrm{Var}\,S$, $\mathrm{E}(S-\mathrm{E}\,S)^3$ by differentiating the logarithm of (2.84) at point $z=0$:

$$\left.\frac{\mathrm{d}\log M_S(z)}{\mathrm{d}\,z}\right|_{z=0} = \left.\frac{M'_S(z)}{M_S(z)}\right|_{z=0} = M'_S(0) = \mathrm{E}\,S, \tag{2.85}$$

$$\left.\frac{\mathrm{d}^2\log M_S(z)}{\mathrm{d}\,z^2}\right|_{z=0} = \left.\frac{M''_S(z)\,M_S(z) - \left[M'_s(z)\right]^2}{M^2_S(z)}\right|_{z=0}$$

$$= \mathrm{E}\,S^2 - \left(\mathrm{E}\,S\right)^2 = \mathrm{Var}\,S, \tag{2.86}$$

$$\left.\frac{\mathrm{d}^3\log M_S(z)}{\mathrm{d}\,z^3}\right|_{z=0}$$

$$= \left[\frac{\left[M^{(3)}_S(z)\,M_S(z) + M''_S(z)\,M'_S(z) - 2\,M'_S(z)\,M''_S(z)\right]\,M^2_S(z)}{M^4_S(z)}\right.$$

$$\left. - \frac{2\,M_S(z)\,M'_S(z)\,\left[M''_S(z)\,M_S(z) - \left[M'_S(z)\right]^2\right]}{M^4_S(z)}\right]_{z=0}$$

$$= \mathrm{E}\,S^3 - 3\;\mathrm{E}\,S\;\mathrm{E}\,S^2 + 2\left(\mathrm{E}\,S\right)^3 = \mathrm{E}\left(S - \mathrm{E}\,S\right)^3. \tag{2.87}$$

In (2.85)-(2.87) we use well known property of the moment generating function,

$$\left.M^{(k)}_S(z)\right|_{z=0} = \mathrm{E}\,S^k,\ k = 0, 1, \ldots$$

whenever $\mathrm{E}\,S^k < +\infty$.

Example 2.2. *For Poisson distribution of N with $\mathrm{E}\,N = \lambda$ we have compound Poisson model for the aggregate loss with m.g.f.*

$$M_S(z) = \mathrm{e}^{\lambda\,(M_X(z)-1)}$$

and

$$\begin{aligned}
&\mathrm{E}\,S = \lambda\,M'_X(0) = \lambda\,p_1,\\
&\mathrm{Var}\,S = \lambda\,M''_X(0) = \lambda\,p_2,\\
&\mathrm{E}\left(S - \mathrm{E}\,S\right)^3 = \lambda\,M^{(3)}_X(0) = \lambda\,p_3,
\end{aligned}$$

where $p_k = \mathrm{E}\,X^k$.

In some applications we need to work not only with moments of the aggregate loss variable S, but also with its distribution function. From the definition of the compound distribution we have

$$F_S(x) = \sum_{k=0}^{\infty} \mathrm{P}\left(S \le x | N = k\right) \, \mathrm{P}(N = k) = \sum_{k=0}^{\infty} p_k \, F_X^{k*}(x), \qquad (2.88)$$

where F_X^{k*} denotes k-fold convolution of d.f. F_X.

Formula (2.88) is not suitable for evaluation of d.f. F_S, since the expression of convolutions F_X^{k*} can be rather complicated. For this reason different approaches are used for calculation of the aggregate loss distribution.

In the following subsections we describe examples of approaches widely used in practice. They either consist of calculation of a discrete compound distribution, which requires a discrete model for individual claim sizes, or of an approximation of the aggregate loss by some known parametric distribution with given characteristics.

2.4.2 Panjer recursive formula

Theorems 2.3 and 2.4 give recursive calculation of a compound distribution in which the primary distribution is a member of $(a, b, 0)$ or $(a, b, 1)$ class and the secondary distribution is a counting distribution (on values 0,1,...). The same formula is applicable in calculation of the compound distribution of (2.81), under slight modification in the interpretation of the secondary distribution. Suppose that the severity distribution is defined by

$$f_k = \mathrm{P}(X = k\,h), \ k = 0, 1, \ldots; h > 0. \qquad (2.89)$$

The claim sizes thus attain values that are multiples of some unit h. Such a discrete distribution is called **arithmetic distribution**. Then the aggregate loss variable S also attains values that are multiples of h and its distribution is described by p.f.

$$g_k = \mathrm{P}(S = k\,h), \ k = 0, 1, \ldots,$$

which can be obtained by applying (2.71) resp. (2.73).

In reality, we often want to model the individual claim sizes by a continuous parametric distribution (see section 2.2), justified by data from past observations. In the next subsection we describe the construction of an arithmetic distribution preserving some properties of the original continuous model.

2.4.3 Constructing arithmetic distributions

We deal with possible ways of derivation a discrete distribution (2.89) from given continuous d.f. $F_X(x)$, defined for $x > 0$.

Method of rounding (mass dispersal)

For chosen $h > 0$ we set

$$\begin{aligned} f_0 &= \mathrm{P}\left(X < \frac{h}{2}\right) = F_X\left(\frac{h}{2}\right), \\ f_k &= \mathrm{P}\left(k\,h - \frac{h}{2} \le X < k\,h + \frac{h}{2}\right) \\ &= F_X\left(k\,h + \frac{h}{2}\right) - F_X\left(k\,h - \frac{h}{2}\right), \quad k = 1, 2, \ldots \end{aligned}$$

Method of local moment matching

In this method we construct an arithmetic distribution that matches p moments of the arithmetic and the true continuous severity distribution.

Consider an arbitrary interval of length ph, denoted by $[x_k, x_k + ph)$. We will locate point masses $m_0^k,\ m_1^k, \ldots, m_p^k$ at points $x_k,\ x_k + h, \ldots, x_k + ph$ so that

$$\sum_{j=0}^{p} (x_k + jh)^r\, m_j^k = \int_{x_k}^{x_k+ph} x^r dF_X(x), \ \ r = 0, 1, \ldots, p. \tag{2.90}$$

Arrange the intervals so that $x_{k+1} = x_k + ph$. Then the point masses at the endpoints are added together. With $x_0 = 0$, the resulting discrete distribution has successsive probabilities

$$\begin{aligned} &f_0 = m_0^0,\ \ f_1 = m_1^0,\ \ f_2 = m_2^0, \ldots, \\ &f_p = m_p^0 + m_0^1,\ \ f_{p+1} = m_1^1,\ \ f_{p+2} = m_2^1, \ldots \end{aligned}$$

By summing (2.90) for all possible values of k, with $x_0 = 0$, it is clear that p moments are preserved for the entire distribution and that the probabilities add to one exactly.

The solution of (2.90) is

$$m_j^k = \int_{x_k}^{x_k+ph} \prod_{i \ne j} \frac{x - x_k - ih}{(j-i)\,h}\, dF_X(x), \ \ j = 0, 1, \ldots, p.$$

The proof is based on the Lagrange formula for collocation of a polynomial $f(y)$ at points $y_0, y_1, \dots, y_n$:

$$f(y) = \sum_{j=0}^{n} f(y_j) \prod_{i \neq j} \frac{y - y_i}{y_j - y_i}.$$

Applying this formula to the polynomial $f(y) = y^r$ over the points $x_k,\ x_k + h, \dots, x_k + ph$ yields

$$x^r = \sum_{j=0}^{p} (x_k + jh)^r \prod_{i \neq j} \frac{x - x_k - ih}{(j-i)\, h},\ \ r = 0, 1, \dots, p.$$

2.4.4 Fast Fourier transform

The fast Fourier transform (FFT) is an algorithm that can be used for inverting characteristic functions to obtain probability functions of discrete random variables.

Assume that the distribution of the aggregate loss is arithmetic with finite support, given by

$$g_k = \mathrm{P}(S = k\, h),\ \ k = 0, \dots, n-1. \tag{2.91}$$

The discrete **Fourier transform** of (2.91) is

$$\tilde{g}_z = \sum_{k=0}^{n-1} g_k \, \exp\left\{2\pi i \, \frac{zk}{n}\right\},\ \ z = 0, \dots, n-1. \tag{2.92}$$

It has the inversion formula

$$g_k = \frac{1}{n} \sum_{z=0}^{n-1} \tilde{g}_z \, \exp\left\{-2\pi i \, \frac{zk}{n}\right\},\ \ k = 0, \dots, n-1. \tag{2.93}$$

From (2.92) we observe

$$\begin{aligned} \tilde{g}_z &= \mathrm{E} \exp\left\{2\pi i \, \frac{z(S/h)}{n}\right\} = \mathrm{E}\, \mathrm{E}\left[\exp\left\{2\pi i \, \frac{z}{hn}\, S\right\} \big| N\right] \\ &= \sum_{k=0}^{\infty} p_k \left[\mathrm{E} \exp\left\{2\pi i \, \frac{z}{hn}\, X\right\}\right]^k = P_N\left(\tilde{f}_z\right), \end{aligned} \tag{2.94}$$

where $\tilde{f}_z$ is discrete Fourier transform of the arithmetic severity distribution (with the same support $\{0, \dots, n-1\}$).

The algorithm for calculation of (2.81) is the following:

1. Discretize the severity distribution (see section 2.4.3) to obtain the distribution
$$f_k = \mathrm{P}(X = k\,h),\ \ k = 0, \ldots, n-1. \tag{2.95}$$

2. Apply the discrete Fourier transform to (2.95).

3. Calculate the Fourier transform $\tilde{g}_z$ by (2.94).

4. Apply the inverse Fourier transform (2.93) to obtain the distribution of aggregate loss for the dicretized severity model.

The **fast Fourier transform (FFT)** is an algorithm for an efficient calculation of the discrete Fourier transform (2.92). We choose $n = 2^d$ for some integer d. Then we write

$$\begin{aligned}
\tilde{g}_z &= \sum_{k=0}^{2^d-1} g_k \exp\left\{2\pi i\,\frac{zk}{2^d}\right\} \\
&= \sum_{k=0}^{2^{d-1}-1} g_{2k} \exp\left\{2\pi i\,\frac{2zk}{2^d}\right\} + \sum_{k=0}^{2^{d-1}-1} g_{2k+1} \exp\left\{2\pi i\,\frac{z(2k+1)}{2^d}\right\} \\
&= \sum_{k=0}^{2^{d-1}-1} g_{2k} \exp\left\{2\pi i\,\frac{zk}{2^{d-1}}\right\} + \exp\left\{2\pi i\,\frac{z}{2^d}\right\} \sum_{k=0}^{2^{d-1}-1} g_{2k+1} \exp\left\{2\pi i\,\frac{zk}{2^{d-1}}\right\} \\
&= \tilde{g}_z^{(1)} + \exp\left\{2\pi i\,\frac{z}{2^d}\right\} \tilde{g}_z^{(2)},
\end{aligned}$$

where $\tilde{g}_z^{(1)}$ and $\tilde{g}_z^{(2)}$ are discrete Fourier transforms of two sequences of the length 2^{d-1}: $\{g_0, g_2, \ldots, g_{2^d-2}\}$ and$\{g_1, g_3, \ldots, g_{2^d-1}\}$. Iterating the above indicated approach until we have reduced the length of the transformed sequences to $2^0 = 1$ calculates the discrete Fourier transform efficiently.

2.4.5 Approximations for aggregate loss distribution

In this subsection we present examples of approximations of an unknown aggregate loss distribution by some well known parametric model.

In the case of a compound Poisson model for S, where the secondary distribution has a finite variance, we might consider a normal approximation based on a version of the central limit theorem holding for the sum of random number of summands with Poisson distribution with $\mathrm{E}\,N = \lambda$. It can be shown, using the convergence of characteristic functions, that

$$\lim_{\lambda\to\infty} \mathrm{P}\left(\frac{S - \lambda p_1}{\sqrt{\lambda p_2}} \le x\right) = \Phi(x),\ \ x \in \mathbb{R}\,.$$

Thus S can be approximated by a normally distributed variable, when the expected number of claims is high enough. However, such an approximation gives poor results when there is still a substantial skewness in the aggregate loss data. We usually want to employ not only the estimated expected value and variance, but also the skewness in the resulting approximation. We denote those values by

$$\begin{aligned} \mathrm{E}\, S &= m \\ \mathrm{Var}\, S &= s^2 \\ \frac{\mathrm{E}(S - \mathrm{E}\, S)^3}{\sqrt{\mathrm{Var}\, S}} &= \gamma_1. \end{aligned} \tag{2.96}$$

In the insurance applications usually $\gamma_1 > 0$.

In what follows we present examples of two approaches to the approximation.

Approximations by translated gamma and lognormal distributions

We describe an approximation of a variable S with the characteristics (2.96) (not necessarily with compound distribution) by a distribution with positive skewness. Since we have three input values in (2.96) we need a distribution with three parameters. We construct such a distribution by translating a gamma or a lognormal random variable by a value $k \in \mathbb{R}$. We consider r.v. $X = k + Z$, where a) Z has gamma distribution $\Gamma(\alpha, \theta)$ (see (2.10)) or b) Z has lognormal distribution $LN(\mu, \sigma^2)$ (see (2.17)). Then in case a) we have (see (2.12)-(2.15))

$$\begin{aligned} \mathrm{E}\, X &= k + \alpha\,\theta, \\ \mathrm{Var}\, X &= \alpha\,\theta^2, \\ \frac{\mathrm{E}(X - \mathrm{E}\, X)^3}{\sqrt{\mathrm{Var}\, X}} &= \frac{2}{\sqrt{\alpha}}. \end{aligned} \tag{2.97}$$

In case b) those characteristics are (see (2.20)-(2.23))

$$\begin{aligned} \mathrm{E}\, X &= k + \mathrm{e}^{\mu + \frac{\sigma^2}{2}}, \\ \mathrm{Var}\, X &= \mathrm{e}^{2\mu + \sigma^2} \left(\mathrm{e}^{\sigma^2} - 1 \right), \\ \frac{\mathrm{E}(X - \mathrm{E}\, X)^3}{\sqrt{\mathrm{Var}\, X}} &= \left(\mathrm{e}^{\sigma^2} + 2 \right) \sqrt{\mathrm{e}^{\sigma^2} - 1}. \end{aligned} \tag{2.98}$$

For the approximation we use the distribution of X with parameters a) k, α, θ or b) k, μ, σ^2 which fulfill the system of equations

$$\begin{aligned} \operatorname{E} X &= m, \\ \operatorname{Var} X &= s^2, \\ \frac{\operatorname{E}(X - \operatorname{E} X)^3}{\sqrt{\operatorname{Var} X}} &= \gamma_1. \end{aligned} \tag{2.99}$$

For the left-hand sides of (2.99) we substitute a) (2.97) or b) (2.98).

Edgeworth approximation

Let us denote

$$Z = \frac{S - m}{s}$$

the standardized aggregate loss random variable. It holds

$$\operatorname{E} Z = 0, \ \operatorname{Var} Z = 1, \ \frac{\operatorname{E}(Z - \operatorname{E} Z)^3}{\sqrt{\operatorname{Var} Z}} = \gamma_1.$$

The Edgeworth approximation is based on the Taylor expansion of the logarithm of m.g.f. M_Z around the origin. We write

$$\log M_Z(r) = \sum_{k=0}^{n} a_k\, r^k + o\left(r^k\right), \ r \to 0.$$

Here

$$a_k = \frac{1}{k!} \left(\frac{\mathrm{d}^k}{\mathrm{d}\, r^k} \log M_Z(r) \Big|_{r=0} \right).$$

Especially,

$$\begin{aligned} a_0 &= \log M_Z(0) = 0, \\ a_1 &= \operatorname{E} Z = 0, \\ a_2 &+ \frac{1}{2} \operatorname{Var} Z = \frac{1}{2}, \\ a_3 &= \frac{1}{6} \operatorname{E}(Z - \operatorname{E} Z)^3 = \frac{\gamma_1}{6}. \end{aligned}$$

We use the approximation

$$\begin{aligned} M_Z(r) &\doteq \exp\left(\frac{1}{2} r^2\right) \exp\left(\sum_{k=3}^{n} a_k r^k\right) \\ &\doteq \exp\left(\frac{1}{2} r^2\right) \left[1 + \sum_{k=3}^{n} a_k r^k + \frac{1}{2} \left(\sum_{k=3}^{n} a_k r^k\right)^2 + \dots \right]. \end{aligned} \tag{2.100}$$

For an approximation based on the knowledge of (2.96), we use just terms up to order r^3 on the right-hand side of (2.100). We have

$$M_Z(r) \doteq \mathrm{e}^{r^2/2}\left[1+\frac{\gamma_1}{6}\,r^3\right]. \tag{2.101}$$

For inverting (2.101) in terms of the distribution of r.v. Z we use the following lemma.

Lemma 2.1. *For $k \in \{0, 1, \dots\}$ and $r \in \mathbb{R}$ it holds*

$$r^k\ \mathrm{e}^{r^2/2} = (-1)^k \int_{-\infty}^{+\infty} \mathrm{e}^{rx}\ \Phi^{(k+1)}(x)dx, \tag{2.102}$$

where Φ is the standard normal d.f. and $\Phi^{(k)}$ is its k-derivative.

Proof. We prove (2.102) by induction. For $k = 0$ both sides of (2.102) are equal to the m.g.f. of $N(0, 1)$, $M_X(r) = \mathrm{e}^{r^2/2}$.

Using integration by parts we obtain

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

The first term on the right-hand side is equal to zero and the second term can be rewritten using the induction assumption. We have

$$\begin{aligned}
(-1)^{k+1} \int_{-\infty}^{+\infty} \mathrm{e}^{rx}\,\Phi^{(k+2)}(x)dx &= r\,(-1)^k \int_{-\infty}^{+\infty} \mathrm{e}^{rk}\ \Phi^{(k+1)}(x)dx \\
&= r\,r^k\ \mathrm{e}^{r^2/2}\,.
\end{aligned}$$

□

Lemma 2.1 together with (2.101) provides approximation

$$M_Z(r) \doteq \int_{-\infty}^{+\infty} \mathrm{e}^{rx}\left[\phi(x) - \frac{\gamma_1}{6}\phi^{(3)}(x)\right]dx, \tag{2.103}$$

where ϕ is the p.d.f. of standard normal distribution. Comparing the right-hand side of (2.103) with

$$M_Z(r) = \int_{-\infty}^{+\infty} \mathrm{e}^{rx}\,dF_Z(x)$$

we arrive at the approximation of the p.d.f. of Z:

$$f_Z(x) \doteq \phi(x) - \frac{\gamma_1}{6}\phi^{(3)}(x). \tag{2.104}$$

In comparison with the standard normal approximation of a standardized distribution by $N(0, 1)$, (2.104) contains a term correcting the result with respect to positive skewness of r.v. Z.

Chapter 3

Reinsurance

Reinsurance is a tool used by an insurance company to reduce the risk stemming from its insurance policies. It works on principles analogous to those of insurance - one party (reinsurer) indemnifies the other party (insurance company, cedant) for a specified share of its financial loss for a fixed price - reinsurance premium.

The main function of reinsurance is increasing the underwriting capacity of the insurer. It can also help to stabilize the insurer's underwriting and financial results over time and it provides protection against large losses. Reinsurers often provide consulting services to their cedants.

3.1 Forms of reinsurance

There are two types of reinsurance: facultative and obligatory. In **facultative reinsurance** each individual risk is presented to the reinsurer. Both the insurer and the reinsurer are free to present or accept the risk. In **obligatory reinsurance** every claim within the specifications of the reinsurance treaty is ceded and accepted by the reinsurer.

In what follows we will deal with basic structures used in obligatory reinsurance. We use the classification with respect to the way in which losses and premiums are divided between the cedant and the reinsurer.

3.1.1 Proportional reinsurance

In proportional reinsurance claims that occur during a given period are shared in an agreed proportion by the cedant and the reinsurer.

In **quota share reinsurance** the proportion is the same for all insured risks covered by the reinsurance treaty.

Let us denote the proportion by q, $q \in (0,1)$. Let us consider a risk (insurance policy) with the **sum insured** Z. It can be the total value of the insured property or a limit for the claim size paid by the insurer. In some lines of business insurers work with **probable maximum loss** - the estimated maximal value of loss that can result from one insured event.

If a loss of size X occurs, $q\,X$ is the ceded part (paid by the reinsurer) and $(1-q)\,X$ is the retained part (paid by the cedant).

There can be agreed a limit (**capacity**) C applied to the sum insured: for risks with sum insured greater than the capacity ($Z > C$), the proportion of losses ceded is reduced to $\frac{C}{Z}\,q$. Only the proportion $\frac{C}{Z}$ of each loss is for such a risk covered by the reinsurance with capacity C. The remaining share $\frac{Z-C}{Z}$ has to be paid by the cedant or from another reinsurance arrangement.

In **surplus reinsurance** the proportion ceded is not the same for all reinsured risks, it depends on the sum insured. The cedant determines the maximum loss that it can retain for each risk in the portfolio - **retention (line)**. The capacity of reinsurer is usually expressed by the number of lines.

Let us denote the retention by R, the number of lines by L. For a risk with sum insured Z the proportion ceded q is

- $q = 0$ for $Z \le R$,
- $q = 1 - \frac{R}{Z}$ for $R < Z \le (L+1)\,R$.

In case $Z \le (1+L)\,R$, the proportion ceded is expressed as

$$q = \left(1 - \frac{R}{Z}\right)_+$$

and the proportion retained as

$$1 - q = \frac{R}{Z} \wedge 1.$$

For a risk with the sum insured greater than the capacity of the reinsurance, $Z > (L+1)\,R$, the retained proportion is $\frac{R}{Z}$, the ceded proportion is $\frac{LR}{Z}$. The remainder $1 - \frac{(L+1)R}{Z}$ has to be covered by the cedant or by another reinsurance policy.

3.1.2 Non-proportional reinsurance

The mathematical description of all common forms of non-proportional reinsurance is based on the same formula with two parameters - **priority (retention)** r and **layer** l. For loss variable X the ceded part is

$$X^{(C)} = (X - r)_+ \wedge l \tag{3.1}$$

and the retained part is

$$X^{(R)} = X \wedge r. \tag{3.2}$$

For $X > r + l$ there remains part $X - (r + l)$ to be paid from other sources.

Most frequently used form of non-proportional reinsurance is **XL-reinsurance (excess of loss reinsurance)**, in which X in (3.1) and (3.2) is either the size of one insurance claim (XL-reinsurance **per risk**) or the aggregate loss caused by one event (XL-reinsurance **per event**). A special kind of the second type is **cat XL** (catastrophic XL-reinsurance), where the event considered is some natural or man-made disaster (windstorm, flood etc.).

Similar forms as (3.1) and (3.2) can be used for **SL-reinsurance (stop loss reinsurance)**, where X is taken as the aggregate loss in given time period (one year). Parameters r and l are then usually expressed as multiples of total yearly premium corresponding to reinsured risks.

3.2 Reinsurance pricing

In proportional reinsurance the premium is ceded in the same proportion as claim payments. On the other hand, the reinsurer often pays to the cedant a commission as a share on the acquisition costs and a share on the profit from the reinsurance treaty. The commission is either a **fixed** percentage of premium ceded or it can be decreased or increased depending on the final loss ratio (**sliding scale commission**). Another possible form is a **profit commission** - an agreed percentage of the profit from the reinsurance treaty is paid back to the cedant.

The pricing of non-proportional reinsurance is not as straightforward as for the proportional case. Several approaches are used in practice depending on information and data available for risks concerned. In the following subsection we explain them for the simple form of XL-reinsurance with unlimited cover.

3.2.1 XL-reinsurance with unlimited cover

Let us assume that the aggregate loss for a group of reinsured risks is

$$S = \sum_{i=1}^{N} X_i,$$

where N is the total number of claims, X_i, $i = 1, 2, \ldots,$ the individual claims, are independent and identically distibuted, independent on N (S has a compound distribution).

For an unlimited XL cover we suppose $l = +\infty$, so for each claim X_i the ceded part is

$$X_i^{(C)} = (X_i - r)_+$$

and the retained part is

$$X_i^{(R)} = X_i \wedge r.$$

The ceded part of the aggregate loss,

$$S^{(C)} = \sum_{i=1}^{N} X_i^{(C)}, \tag{3.3}$$

has again a compound distribution, its expected value is

$$\begin{aligned} \mathrm{E}\, S^{(C)} &= \mathrm{E}\, N \ \mathrm{E}\, X^{(C)} = \mathrm{E}\, N \ \mathrm{E}\, X \, \frac{\mathrm{E}\, X^{(C)}}{\mathrm{E}\, X} \\ &= \mathrm{E}\, S \, \frac{\mathrm{E}\, X^{(C)}}{\mathrm{E}\, X}. \end{aligned} \tag{3.4}$$

Since

$$\mathrm{E}\, X = \mathrm{E}\, X^{(C)} + \mathrm{E}\, X^{(R)},$$

we have

$$\frac{\mathrm{E}\, X^{(C)}}{\mathrm{E}\, X} = 1 - \frac{\mathrm{E}\, X^{(R)}}{\mathrm{E}\, X} = 1 - \mathrm{LER}_X(r),$$

where

$$\mathrm{LER}_X(r) = \frac{\mathrm{E}(X \wedge r)}{\mathrm{E}\, X}$$

is the loss elimination ratio introduced in (2.8). We rewrite (3.4) as

$$P^{(C)} = P \ (1 - \mathrm{LER}_X(r)) \,. \tag{3.5}$$

Thus the pure reinsurance premium $P^{(C)} = \mathrm{E}\, S^{(C)}$ is obtained by multiplying the total pure premium $P = \mathrm{E}\, S$ by a coefficient $1 - \mathrm{LER}_X(r)$ depending on the claim size distribution and on the chosen priority r.

Experience rating approach (known also as **burning cost** method) determines the reinsurance premium as the total pure premium multiplied by a rate computed as an average of

$$\frac{S_t^{(C)}}{P_t}, \ t = 1, \ldots, T.$$

Here $S_t^{(C)}$ is ceded loss in past year t, possibly reevaluated according to inflation, changes in underwriting or reserving practice. P_t is premium of year t adjusted to current tariffs.

When the loss experience from past years is limited or none, the approach of **exposure rating** is popular.

When X is a loss occurred on a risk with sum insured Z, we denote by

$$Y = \frac{X}{Z}$$

loss degree. We have

$$\mathrm{LER}_X(r) = \frac{\mathrm{E}(X \wedge r)\,\frac{1}{Z}}{\mathrm{E}\,\frac{X}{Z}} = \frac{\mathrm{E}(Y \wedge r')}{\mathrm{E}\,Y} = \mathrm{LER}_Y(r'),$$

where $r' = \frac{r}{Z}$ in **normalized priority** with values in $[0, 1]$.

Exposure rating approach uses instead of (3.5)

$$P^{(C)} = P\,\left(1 - \mathrm{LER}_Y(r')\right),$$

where $1 - \mathrm{LER}_Y(r')$, $r' \in [0, 1]$, is **market exposure curve** applied to the risk profile of the cedant.

In the most demanding approach from the point of view of the necessary data, **frequency and severity modeling**, we use models described in Chapter 2.

We can use suitable parametric models for total claim count N and individual claim size X, then derive the expected value of $\mathrm{E}\,X^{(C)}$ for given priority r and so come to an expression of pure reinsurance premium in the form

$$P^{(C)} = \mathrm{E}\,N\;\mathrm{E}\,X^{(C)}. \tag{3.6}$$

We illustrate the technique with an example of compound Poisson distribution with Pareto distributed claims.

Let N have Poisson distribution with $\mathrm{E}\,N = \lambda$. For the total claim size we use Pareto distribution with finite expectation, i.e.

$$F_X(x) = 1 - \left(\frac{x}{a}\right)^{-\alpha},\; x \ge a,\; \alpha > 1,\; a > 0.$$

In case we have data comprising all claims (including those with size less than the priority r), we use (3.6), where

$$\mathrm{E}\,X^{(C)} = \mathrm{E}(X - r)_+ = \int_r^\infty (x - r) a^\alpha\,\alpha\,x^{-\alpha-1} dx.$$

We obtain

$$\mathrm{E}\,X^{(C)} = \frac{a^\alpha}{\alpha - 1}\,r^{-\alpha+1}$$

and

$$P^{(C)} = \frac{\lambda\, a^{\alpha}}{\alpha - 1}\, r^{-\alpha+1}. \tag{3.7}$$

We insert in (3.7) estimates of λ and α based on past observations of total claim numbers and all claim sizes.

Alternatively, we can rewrite (3.3) as

$$S^{(C)} = \sum_{i=1}^{\bar{N}} \left(\bar{X}_i - r\right),$$

where $\bar{N}$ is the number of losses exceeding given priority r and $\bar{X}_i$, $i = 1, 2. \ldots$ are i.i.d. random variables distributed as X conditionally, given $X > r$. From Theorem 2.2 we know that $\bar{N}$ has Poisson distribution with

$$\mathrm{E}\,\bar{N} = \lambda\,\mathrm{P}(X > r) = \lambda\,\left(\frac{r}{a}\right)^{-\alpha},$$

and for $x \geq r$

$$\mathrm{P}\left(\bar{X} > x\right) = \mathrm{P}(X > x | X > r) = \left(\frac{x}{r}\right)^{-\alpha},$$

so $\bar{X}$ has Pareto distribution with the same parameter α as X.

We now write instead of (3.6)

$$\begin{aligned} P^{(C)} &= \mathrm{E}\,\bar{N}\left(\mathrm{E}\,\bar{X} - r\right) \qquad (3.8)\\ &= \lambda\left(\frac{r}{a}\right)^{-\alpha}\left(r\frac{\alpha}{\alpha - 1} - r\right) = \lambda\frac{a^{\alpha}}{\alpha - 1}\, r^{-\alpha+1}. \end{aligned}$$

Naturally, we arrived at the same result as in (3.7). In the situation when we have only information concerning claims exceeding the priority, we can base our estimate of $P^{(C)}$ on (3.8).

3.2.2 XL-reinsurance with reinstatements

Commonly used form of XL-reinsurance has finite size of layer l. Except from the limit l for each individual claim payment there is usually also aggregate limit for the sum of all reinsurer's payments during the period covered. It is again defined by the size of the layer l.

In XL-reinsurance with paid **reinstatements** the cover for loss exceeding the layer is bought for additional reinsurance premium. There is maximum number of possible reinstatements denoted by K. Then the maximum capacity of such a cover is $(1 + K)\, l$.

Denote now by S the sum of all parts of individual claims hitting the layer l,

$$S = \sum_{i=1}^{N} \left((X_i - r)_+ \wedge l \right). \tag{3.9}$$

In XL-reinsurance with paid reinstatements not only the final amount of claims paid by reinsurer, but also the final amount of premium paid by cedant, are random variables (not known at the beginning of the cover).

There is the initial premium P, that covers the 0th reinstatement which is

$$R_0 = S \wedge l.$$

Let $c_k\, P$ be the premium for k-th reinstatement. It is the premium paid in case the total loss S exceeds the value $k\, l$.

In practice the premium for reinstatements is paid on pro rata basis - whenever a claim hits the layer and the total loss is between $(k-1)\, l$ and $k\, l$, proportional part of the reinstatement premium given by

$$\frac{(X_i - r)_+ \wedge l}{l}\, c_k\, P$$

is paid. Thus the total premium paid for k-th reinstatement is

$$\frac{R_{k-1}}{l}\, c_k\, P,$$

where

$$R_k = (S - k\, l)_+ \wedge l$$

is the part of the total loss covered by k-th reinstatement.

The total payment of ceded claims is

$$S_K = S \wedge (K+1)\, l = \sum_{k=0}^{K} R_k. \tag{3.10}$$

The total premium paid to the end of the period of cover is

$$P_K = P \left(1 + \frac{1}{l} \sum_{k=1}^{K} c_k\, R_{k-1} \right). \tag{3.11}$$

The basic premium P is determined from (3.10) and (3.11) based on the assumption that the expected value of the total reinsurer's payments should equal to the expected value of the total reinsurance premium paid by the cedant:

$$\mathrm{E}\, S_K = \mathrm{E}\, P_K.$$

We further assume $c_k = 1$, $k = 1, \ldots, K$, and write

$$\mathrm{E}\, S_K = P\left(1 + \frac{1}{l}\, \mathrm{E}\, S_{K-1}\right). \qquad (3.12)$$

The expected value $\mathrm{E}\, S_K$ is in fact the limited expected value LEV introduced in (2.5),

$$\mathrm{E}\, S_K = \mathrm{E}\left[S \wedge (K+1)\, l\right].$$

Using (2.6) we deduce from (3.12) an expression for the basic premium P:

$$P = \frac{\int_0^{(K+1)l} x dF_S(x) + (K+1)l\left[1 - F_S\left((K+1)l\right)\right]}{1 + \frac{1}{l}\int_0^{Kl} x dF_S(x) + K\left[1 - F_S(Kl)\right]}. \qquad (3.13)$$

To evaluate (3.13) we need the d.f. of total loss (3.9). It is usually assumed that S has a compound distribution. For N being a member of the $(a, b, 0)$ or $(a, b, 1)$ class we can use for example Panjer recursive formula. Then we need an arithmetic distribution for r.v.

$$Y = (X - r)_+ \wedge l,$$

which can be derived from an appropriate continuous model for individual claim X.

When

$$g_k = \mathrm{P}(S = k\, h),\ \ k = 0, 1, \ldots$$

is the distribution resulting from the application of Panjer formula, we obtain from (3.13) the basic premium in the form

$$P = \frac{h \sum_{k=0}^{\frac{(K+1)l}{h}} k\, g_k + (K+1)l\left(1 - \sum_{k=0}^{\frac{(K+1)l}{h}} g_k\right)}{1 + \frac{h}{l}\sum_{k=0}^{\frac{KL}{h}} k\, g_k + K\left(1 - \sum_{k=0}^{\frac{Kl}{h}} g_k\right)}.$$

Chapter 4

Basic Methods of Claims Reserving

4.1 Notation

We separate loss data according to the **accident year** (year of ocurrence) i and the **development year** j. We assume $i \in \{0, \ldots, I\}$, $j \in \{0, \ldots, J\}$, where I denotes the most recent accident year and J denote the last development year. We further assume that $I = J$.

We denote the **incremental** data as X_{ij}: X_{ij} stands for payments for claims in accidental year i made in year $i + j$ (alternatively, X_{ij} may denote the number of reported claims with delay j or the change of the reported claims amount).

The **cumulative** amount C_{ij} for accident year i after j development years is then given by

$$C_{ij} = \sum_{k=0}^{j} X_{ik}. \tag{4.1}$$

The observations available at time I are represented by the set

$$D_I = \{X_{ij} : i + j \leq I,\ 0 \leq j \leq J\}. \tag{4.2}$$

They are usually represented in the form of **development triangle**. Data in (4.2) arranged in rows according to accident years and in columns according to development years represent incremental development triangle. The information available at time I is equivalently described by the cumulative development triangle formed by data

$$\{C_{ij} : i + j \leq I,\ 0 \leq j \leq J\}.$$

The values X_{ij} (resp. C_{ij}) for $i+j > I$ need to be estimated or predicted.

We suppose that there is no further development after development year J, so C_{iJ} represents aggregate loss from all claims ocurred in accident year i. Then the **outstanding loss liabilities** (claims reserve) for accident year i at time I are given by

$$R_i = \sum_{j=I-i+1}^{J} X_{ij} = C_{iJ} - C_{i,I-i}. \tag{4.3}$$

In the following sections we introduce basic methods used most frequently for estimating claims reserves. These methods consist in simple algorithms applied to claims development triangles. We present stochastic models that can be used to justify the corresponding algorithms and to quantify the uncertainty in the resulting estimates.

4.2 Chain-ladder method

There are several stochastic models that justify the CL method. Well known is the model introduced by Thomas Mack in [2]. We start with first two assumptions from Mack's model.

Model assumptions (CL)

CL1 Cumulative claims C_{ij} of different accident years are independent.

CL2 There exist **development factors** $f_0, \dots, f_{J-1} > 0$ such that for all $0 \le i \le I$ and all $1 \le j \le J$ we have

$$\mathrm{E}\left[C_{ij} | C_{i0}, \dots, C_{i,j-1}\right] = f_{j-1}\, C_{i,j-1}.$$

CL2 is an assumptions on the first moments - it is used together with CL2 for deriving estimates of (conditionally) expected future claims.

Lemma 4.1. *Under Model Assumptions (CL) we have*

$$\mathrm{E}\left[C_{iJ} | D_I\right] = \mathrm{E}\left[C_{iJ} | C_{i,I-i}\right] = C_{i,I-i}\, f_{I-i} \cdots f_{J-1} \tag{4.4}$$

for all $1 \le i \le I$.

Proof. From the independence of accident years we have

$$\mathrm{E}\left[C_{iJ} | D_I\right] = \mathrm{E}\left[C_{iJ} | C_{i0}, \dots, C_{i,I-i}\right].$$

We rewrite the last expression as an iterated expected value and apply CL2:

$$\begin{aligned}\mathrm{E}\left[C_{iJ}|C_{i0},\dots,C_{i,I-i}\right] &= \mathrm{E}\left[\mathrm{E}\left[C_{iJ}|C_{i0},\dots,C_{i,J-1}\right]\big|C_{i0},\dots,C_{i,I-i}\right]\\ &= \mathrm{E}\left[f_{J-1}\,C_{i\,J-1}|C_{i0},\dots,C_{i,I-i}\right]\\ &= f_{J-1}\ \mathrm{E}\left[C_{i\,J-1}|C_{i0},\dots,C_{i,I-i}\right].\end{aligned}$$

We repeat the last argument until the expected value is conditioned by values of $C_{i,0}$,…,$C_{i,I-i}$, which gives (4.4). □

Chain-ladder estimator

The CL estimator for $\mathrm{E}\left[C_{ij}|D_I\right]$ motivated by (4.4) is

$$\hat{C}_{ij}^{CL} = \mathrm{E}\,\widehat{\left[C_{ij}|D_I\right]} = C_{i,I-i}\,\hat{f}_{I-i}\cdots\hat{f}_{j-1} \tag{4.5}$$

for $i+j>I$, where Cl estimates of development factors are given by

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{ij}} \tag{4.6}$$

Remark. *The development factors f_j are estimated in* (4.6) *by a weighted average of individual development factors*

$$F_{i,j+1} = \frac{C_{i,j+1}}{C_{ij}}, \tag{4.7}$$

$$\hat{f}_j = \sum_{i=0}^{I-j-1} \frac{C_{ij}}{\sum_{k=0}^{I-j-1} C_{kj}}\,F_{i,j+1}.$$

Lemma 4.2 (Properties of CL estimators)**.** *Denote $B_k = \{C_{ij}:\ i+j\le I,\ 0\le j\le k\}$. ($B_k$ is a part of cumulative development triangle containing all columns up the one corresponding to the development year k.)*

The following statements hold

(a) *given B_j, $\hat{f}_j$ is an unbiased estimator for f_j, i.e.*

$$\mathrm{E}\left[\hat{f}_j|B_j\right] = f_j,$$

(b) *$\hat{f}_j$ is unconditionally unbiased, i.e.* $\mathrm{E}\,\hat{f}_j = f_j$,

(c) $\mathrm{E}\left[\hat{f}_0\cdots\hat{f}_j\right] = \mathrm{E}\,\hat{f}_0\cdots\mathrm{E}\,\hat{f}_j$, $j=1,\dots,J-1$. *It means $\hat{f}_0,\cdots,\hat{f}_{J-1}$ are uncorrelated,*

(d) *given* $C_{i,I-i}$, $\hat{C}_{iJ}^{CL}$ *is an unbiased estimator for*

$$\mathrm{E}\left[C_{iJ}|D_I\right] = \mathrm{E}\left[C_{iJ}|C_{i,I-i}\right],$$

i.e.

$$\mathrm{E}\left[\hat{C}_{iJ}^{CL}|C_{i,I-i}\right] = \mathrm{E}\left[C_{iJ}|D_I\right],$$

(e) $\hat{C}_{iJ}^{CL}$ *is unconditionally unbiased, i.e.* $\mathrm{E}\,\hat{C}_{iJ}^{CL} = \mathrm{E}\,C_{iJ}$.

Proof. To show (a), we substitute the expression (4.6) for $\hat{f}_j$ and we use the fact, that the sum in the denominator on the right-hand side of (4.6) contains only variables that are all included in B_j. We obtain

$$\mathrm{E}\left[\hat{f}_j|B_j\right] = \frac{\sum_{i=0}^{I-j-1}\mathrm{E}\left[C_{i,j+1}|B_j\right]}{\sum_{i=0}^{I-j-1}C_{ij}} = \frac{\sum_{i=0}^{I-j-1}C_{ij}\,f_j}{\sum_{i=0}^{I-j-1}C_{ij}} = f_j.$$

The second equality above follows from assumptions CL1 and CL2.

(b) is a direct consequence of (a).

For (c) we show with the help of (a)

$$\begin{aligned}\mathrm{E}\left[\hat{f}_0\cdots\hat{f}_j\right] &= \mathrm{E}\left[\mathrm{E}\left[\hat{f}_0\cdots\hat{f}_j|B_j\right]\right] = \mathrm{E}\left[\hat{f}_0\cdots\hat{f}_{j-1}\,\mathrm{E}\left[\hat{f}_j|B_j\right]\right]\\ &= \mathrm{E}\left[\hat{f}_0\cdots\hat{f}_{j-1}\right]\,f_j = \mathrm{E}\left[\hat{f}_0\cdots\hat{f}_{j-1}\right]\,\mathrm{E}\,\hat{f}_j.\end{aligned}$$

We repeat the same steps with $\mathrm{E}\left[\hat{f}_0\cdots\hat{f}_{j-1}\right]$ and proceed in the same way until we get the product of expected values $\mathrm{E}\,\hat{f}_0\cdots\mathrm{E}\,\hat{f}_j$ on the right-hand side.

In the proof of (d) we use (4.5) and rewrite

$$\begin{aligned}\mathrm{E}\left[\hat{C}_{ij}^{CL}|C_{i,I-i}\right] &= \mathrm{E}\left[C_{i,I-i}\,\hat{f}_{I-i}\cdots\hat{f}_{j-2}\,\hat{f}_{j-1}|C_{i,I-i}\right]\\ &= \mathrm{E}\left[\mathrm{E}\left[C_{i,I-i}\,\hat{f}_{I-i}\cdots\hat{f}_{j-2}\,\hat{f}_{j-1}|B_{J-1}\right]|C_{i,I-i}\right]\\ &= \mathrm{E}\left[C_{i,I-i}\,\hat{f}_{I-i}\cdots\hat{f}_{j-2}\,\mathrm{E}\left[\hat{f}_{j-1}|B_{J-1}\right]|C_{i,I-i}\right]\\ &= f_{J-1}\,\mathrm{E}\left[C_{i,I-i}\,\hat{f}_{I-i}\cdots\hat{f}_{j-2}|C_{i,I-i}\right].\end{aligned}$$

We again repeat the same procedure with the expected value on the right-hand side. Now we use conditioning with respect to the set B_{J-2} and so on. We finaly have

$$\mathrm{E}\left[\hat{C}_{ij}^{CL}|C_{i,I-i}\right] = C_{i,I-i}\,f_{I-i}\cdots f_{J-1},$$

which proves (d) thanks to (4.4).

(e) follows directly from (d).

□

Mack's stochastic model is specified by three assumptions. In addition to CL1 and CL2 we also assume

CL3 There exist parameters $\sigma_0^2, \ldots, \sigma_{J-1}^2 > 0$ such that for all $0 \leq i \leq I$ and all $1 \leq j \leq J$ we have

$$\mathrm{Var}\left[C_{ij}|C_{i0}, \ldots, C_{i,j-1}\right] = \sigma_{j-1}^2\, C_{i,j-1}.$$

From CL3 we deduce for the conditional variance of an individual development factor (4.7)

$$\mathrm{Var}\left[F_{ij}|C_{i0}, \ldots, C_{i,j-1}\right] = \frac{\sigma_{j-1}^2}{C_{i,j-1}}. \tag{4.8}$$

(4.8) is based on the observation, that the variance of an individual development factor F_{ij} should be inversely proportional to the cumulative claim amount C_{ij}.

Assumption CL3 is essential for the derivation of an expression for the mean square error of the chain-ladder estimate (4.5) defined by

$$\mathrm{mse}(\hat{C}_{iJ}) = \mathrm{E}\left[\left(\hat{C}_{ij} - C_{ij}\right)^2 |D_I\right]. \tag{4.9}$$

The question of the mean squared error of the outstanding claim liabilities estimates obtained from the application of various stochastic models and methods of estimation is the subject of further explanation in the framework of the course Mathematics of Non-life insurance 2.

4.3 Bornhuetter-Ferguson method

We present two possibilities of a stochastic model which motivates the BF method.

Model assumptions (BFI)

BFI1 Cumulative claims C_{ij} of different accident years i are independent.

BFI2 There exist parameters $\mu_0, \ldots, \mu_I > 0$ and a pattern $\beta_0, \ldots, \beta_J > 0$ with $\beta_J = 1$ such that for all $0 \le i \le I$, $0 \le j \le J-1$ and $1 \le k \le J-j$ we have

$$\mathrm{E}\, C_{i0} = \beta_0\, \mu_i, \tag{4.10}$$

$$\mathrm{E}\left[C_{i,j+k} | C_{i0}, \ldots C_{ij}\right] = C_{ij} + (\beta_{j+k} - \beta_j)\, \mu_i. \tag{4.11}$$

Under (BFI) it holds

$$\mathrm{E}\, C_{ij} = \beta_j\, \mu_i,\ 0 \le i \le I,\ 0 \le j \le J. \tag{4.12}$$

Indeed, from (4.11) it is deduced for the unconditional expectation of C_{ij}

$$\begin{aligned}\mathrm{E}\, C_{ij} &= \mathrm{E}\, C_{i,j-1} + (\beta_j - \beta_{j-1})\mu_i \\ &= \mathrm{E}\, C_{i,j-2} + (\beta_{j-1} - \beta_{j-2})\mu_i + (\beta_j - \beta_{j-1})\mu_i \\ &= \cdots = \mathrm{E}\, C_{i,0} + (\beta_j - \beta_0)\mu_i.\end{aligned}$$

This together with (4.10) proves (4.12).

Note that $\mathrm{E}\, C_{iJ} = \mu_i$. The sequence $(\beta_j)_{j=0,\ldots,J}$ denotes the **claims development pattern**.

We summarize the weaker set of assumptions implied by (BFI) as

Model Assumptions (BFII)

BFII1 Cumulative claims C_{ij} of different accident years i are independent.

BFII2 There exist parameters $\mu_0, \ldots, \mu_I > 0$ and a pattern $\beta_0, \ldots, \beta_J > 0$ with $\beta_J = 1$ such that for all $0 \le i \le I$, $0 \le j \le J$ we have

$$\mathrm{E}\, C_{ij} = \beta_j\, \mu_i.$$

We again use the model assumptions to derive an expression of the expected value of aggregate loss for given accident year, conditional on the history known at the end of year I.

Lemma 4.3. *Under assumptions (BFI) it holds*

$$\mathrm{E}\left[C_{iJ} | D_I\right] = C_{i,I-i} + (1 - \beta_{I-i})\, \mu_i. \tag{4.13}$$

Proof. From BFI1 we have

$$\mathrm{E}\left[C_{iJ} | D_I\right] = \mathrm{E}\left[C_{iJ} | C_{i,0}, \ldots, C_{i,J-1}\right].$$

Then the statement of the lemma is proved by direct application of (4.11), where we insert $j = J - 1$ and $k = 1$ and we take into account $\beta_J = 1$. □

Remark. *The same result we obtain from (BFII) when we assume the independence of $C_{iJ} - C_{i,I-i}$ of $C_{i0}, \ldots, C_{i,I-i}$.*

BF estimator

The BF estimator for $\mathrm{E}\left[C_{iJ}|D_I\right]$ based on (4.13) is given by

$$\hat{C}_{iJ}^{BF} = \mathrm{E}\,\widehat{\left[C_{iJ}|D_I\right]} = C_{i,I-i} + (1 - \hat{\beta}_{I-i})\,\hat{\mu}_i \tag{4.14}$$

for $1 \leq i \leq I$, where $\hat{\beta}_{I-i}$ is an appropriate estimate for β_{I-i} and $\hat{\mu}_i$ is a prior estimate for the expected ultimate claim $\mathrm{E}\left[C_{iJ}\right]$ (i.e. an estimate that is not based on the observed data in the development triangle).

Comparison of BF and CL models

From assumptions (CL) it follows for $1 \leq j \leq J$

$$\mathrm{E}\left[C_{ij}\right] = \mathrm{E}\left[\mathrm{E}\left[C_{ij}|C_{i,j-1}\right]\right] = f_{j-1}\ \mathrm{E}\,C_{i,j-1} = \cdots = \mathrm{E}\,C_{i,0}\prod_{k=0}^{j-1} f_k. \tag{4.15}$$

We insert in (4.15) for $\mathrm{E}\,C_{i,0}$ from

$$\mathrm{E}\left[C_{iJ}\right] = \mathrm{E}\left[C_{i0}\right]\prod_{k=0}^{J-1} f_k,$$

and obtain

$$\mathrm{E}\left[C_{ij}\right] = \prod_{k=j}^{J-1} f_k^{-1}\ \mathrm{E}\left[C_{iJ}\right]. \tag{4.16}$$

It corresponds to model assumptions (BFII) with

$$\mu_i = \mathrm{E}\,C_{iJ},\ \beta_j = \prod_{k=j}^{J-1} f_k^{-1},\ j = 0, \ldots, J-1,\ \beta_J = 1.$$

Although the BF estimator (4.14) does not prescribe any specific method for estimation of the claim development pattern parameters $\beta_0 \ldots, \beta_{J-1}$, in practice we often derive the estimates from data using estimated CL factors $\hat{f}_j$:

$$\hat{\beta}_j^{CL} = \prod_{k=j}^{J-1} \frac{1}{\hat{f}_k}. \tag{4.17}$$

We denote the BF estimator of $\mathrm{E}\left[C_{iJ}|D_I\right]$ with this special choice of $\hat{\beta}_j$ as $\hat{\hat{C}}_{ij}^{BF}$:

$$\hat{\hat{C}}_{iJ}^{BF} = C_{i,I-i} + \left(1 - \hat{\beta}_{I-i}^{CL}\right)\,\hat{\mu}_i.$$

Comparison of CL and BF estimators

For the CL estimator we have

$$\begin{aligned}
\hat{C}_{iJ}^{CL} &= C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j = C_{i,I-i} + C_{i,I-i} \left(\prod_{j=I-i}^{J-1} \hat{f}_j - 1 \right) \\
&= C_{i,I-i} + \frac{\hat{C}_{iJ}^{CL}}{\prod_{j=I-i}^{J-1} \hat{f}_j} \left(\prod_{j=I-i}^{J-1} \hat{f}_j - 1 \right) \\
&= C_{i,I-i} + \left(1 - \hat{\beta}_{I-i}^{CL} \right) \hat{C}_{iJ}^{CL}.
\end{aligned}$$

Hence, if we identify the claims development pattern from the CL and BF methods, the difference between the BF and CL estimators is that for the BF method we use a prior estimate $\hat{\mu}_i$, in the CL method the prior estimate is replaced by the estimate $\hat{C}_{iJ}^{CL}$ which is based only on the observations.

4.4 Poisson model for the number of claims

In this chapter we consider the incremental variable X_{ij} as the number of claims in accident year i reported in year $i+j$.

The Poisson model is an example of claims reserving model that is based on explicit distributional assumptions.

Model assumptions (Po)

There exist parameters $\mu_0, \ldots, \mu_I > 0$ and $\gamma_0, \ldots, \gamma_J > 0$ such that the incremental claims X_{ij} are independent and Poisson distributed with

$$\mathrm{E}\left[X_{ij}\right] = \mu_i \, \gamma_j,$$

for all $0 \le i \le I$, $0 \le j \le J$, and $\sum_{j=0}^{J} \gamma_j = 1$.

Note the following implications of assumptions (Po):

1. The assumptions (Po) imply that the increments X_{ij} are non-negative.

2. The cumulative claim in accident year i, C_{iJ} is Poisson distributed with $\mathrm{E}\, C_{iJ} = \mu_i$. (See theorem 2.1 of section 2.3.)

3. We have $\frac{\mathrm{E}[X_{ij}]}{\mathrm{E}[X_{i0}]} = \frac{\gamma_j}{\gamma_0}$, which is independent of i.

Lemma 4.4. *The Poisson model satisfies model assumptions (BF1).*

Proof. The independence of the cumulative claims of different accident years follows from the independence of X_{ij}.

We have $\mathrm{E}[C_{i0}] = \mathrm{E}[X_{i0}] = \mu_i\,\beta_0$, with $\beta_0 = \gamma_0$, and

$$\begin{aligned}\mathrm{E}[C_{i,j+k}|C_{i0},\ldots,C_{ij}] &= C_{ij} + \sum_{l=1}^{k} \mathrm{E}[X_{i,j+k}|C_{i0},\ldots,C_{ij}] \\ &= C_{ij} + \mu_i \sum_{l=1}^{k} \gamma_{j+k} = C_{ij} + \mu_i\ (\beta_{j+k} - \beta_j)\end{aligned}$$

with $\beta_j = \sum_{l=0}^{j} \gamma_l$. □

According to the previous lemma we could use BF estimator for the determination of claims reserves in the Poisson model. The distributional assumption however allows for estimation of the parameters by means of the maximum likelihood method.

The likelihood function on the set of observations D_I is given by

$$L(\mu_0,\ldots,\mu_I,\gamma_0,\ldots,\gamma_I) = \prod_{i+j\le I} \left(\exp(-\mu_i\,\gamma_j)\,\frac{(\mu_i\,\gamma_j)^{X_{ij}}}{X_{ij}!} \right).$$

The log-likelihood function is then

$$l(\mu_0,\ldots,\mu_I,\gamma_0,\ldots,\gamma_I) = -\sum_{i+j\le I} \mu_i\,\gamma_j + \sum_{i+j\le I} X_{ij}\ \log(\mu_i\,\gamma_j) - \sum_{i+j\le I} \log(X_{ij}!)\,. \tag{4.18}$$

Equating to zero the partial derivatives of (4.18) with respect to the unknown parameters μ_i and γ_j leads to a system of equations

$$\begin{aligned} -\sum_{j=0}^{I-i} \hat{\gamma}_j + \sum_{j=0}^{I-i} \frac{X_{ij}}{\hat{\mu}_i} &= 0,\ i = 0,\ldots,I, \\ -\sum_{i=0}^{I-j} \hat{\mu}_i + \sum_{i=0}^{I-j} \frac{X_{ij}}{\hat{\gamma}_j} &= 0,\ j = 0,\ldots,J, \end{aligned}$$

which can be rewritten as

$$\sum_{j=0}^{I-i} \hat{\mu}_i\,\hat{\gamma}_j = \sum_{j=0}^{I-i} X_{ij} = C_{i,I-i},\ i = 0,\ldots,I, \tag{4.19}$$

$$\sum_{i=0}^{I-j} \hat{\mu}_i\,\hat{\gamma}_j = \sum_{i=0}^{I-j} X_{ij},\ j = 0,\ldots,J. \tag{4.20}$$

For the solution of (4.19) and (4.20) there is a constraint

$$\sum_{j=0}^{J} \hat{\gamma}_j = 1. \tag{4.21}$$

Poisson ML estimator

The estimators for $\mathrm{E}[X_{ij}]$ and $\mathrm{E}[C_{iJ}|D_I]$ for $i+j>I$ are given by

$$\hat{X}_{ij}^{Po} = \widehat{\mathrm{E}[X_{ij}]} = \hat{\mu}_i\,\hat{\gamma}_j \tag{4.22}$$

$$\hat{C}_{iJ}^{Po} = \widehat{\mathrm{E}[C_{iJ}|D_I]} = C_{i,I-i} + \sum_{I-i+1}^{J} \hat{\mu}_i\,\hat{\gamma}_j, \tag{4.23}$$

where $\hat{\mu}_i,\ i=0,\dots,I$, and $\hat{\gamma}_j,\ j=0,\dots,J$, are ML estimates given by (4.19) and (4.20) with the constraint (4.21).

Remark. *It holds*

$$\hat{C}_{iJ}^{Po} = C_{i,I-i} + \left(1 - \sum_{j=0}^{I-i} \hat{\gamma}_j\right) \hat{\mu}_i, \tag{4.24}$$

hence the Poisson ML estimator has the same form as the BF estimator (4.14)*. However, in* (4.24) *we use the ML estimates of μ_i and γ_j that depend on the data in the development triangle.*

There is an interesting connection between the Poisson distributional model and the Mack stochastic model for chain-ladder method. In fact, the CL estimator and the Poisson ML estimator for $\mathrm{E}[C_{iJ}|D_I]$ are the same. To show this, we assume that there is a positive solution to (4.19)-(4.21), $\hat{\mu}_i$, $\hat{\gamma}_j$, $i=0,\dots,I$, $j=0,\dots,J$. We need the following result.

Lemma 4.5. *Under model assumptions (Po) it holds*

$$\sum_{i=0}^{I-j} C_{ij} = \sum_{i=0}^{I-j} \hat{\mu}_i \sum_{k=0}^{j} \hat{\gamma}_k. \tag{4.25}$$

Proof. (4.25) is proved by induction. Using (4.19) for $i=0$ and $I=J$ gives

$$C_{0J} = \sum_{j=0}^{J} X_{oj} = \hat{\mu}_0 \sum_{j=0}^{J} \hat{\gamma}_j.$$

For induction step $j \longrightarrow j-1$ we use the relation between the set of observed incremental claims $\{X_{il}, i = 0, \ldots, I-j, l = 0, \ldots, j\}$ and the set $\{X_{il}, i = 0, \ldots, I-j+1, l = 0, \ldots, j-1\}$. The second set is obtained from the first one by adding the accident year $I-j+1$ and subtracting the development year j:

$$\sum_{i=0}^{I-(j-1)} C_{i,j-1} = \sum_{i=0}^{I-j} C_{ij} - \sum_{i=0}^{I-j} X_{ij} + \sum_{k=0}^{j-1} X_{I-j+1,k}. \tag{4.26}$$

Using (4.19)-(4.20) and the induction assumption we obtain

$$\begin{aligned}
\sum_{i=0}^{I-(j-1)} C_{i,j-1} &= \sum_{i=0}^{I-j} \hat{\mu}_i \sum_{k=0}^{J} \hat{\gamma}_k - \hat{\gamma}_j \sum_{i=0}^{I-j} \hat{\mu}_i + \hat{\mu}_{I-j+1} \sum_{k=0}^{j-1} \hat{\gamma}_k \\
&= \sum_{i=0}^{J-j} \hat{\mu}_i \sum_{k=0}^{j-1} \hat{\gamma}_k + \hat{\mu}_{I-j+1} \sum_{k=0}^{j-1} \hat{\gamma}_k \\
&= \sum_{i=0}^{I-(j-1)} \hat{\mu}_i \sum_{k=0}^{j-1} \hat{\gamma}_k.
\end{aligned}$$

□

From lemma 4.5 it follows the expression of the CL development factor estimate for $j \leq J$,

$$\begin{aligned}
\hat{f}_{j-1} &= \frac{\sum_{i=0}^{I-j} C_{ij}}{\sum_{i=0}^{I-j} C_{i,j-1}} = \frac{\sum_{i=0}^{I-j} C_{ij}}{\sum_{i=0}^{I-j} (C_{ij} - X_{ij})} = \frac{\sum_{i=0}^{I-j} \hat{\mu}_i \sum_{k=0}^{j} \hat{\gamma}_k}{\sum_{i=0}^{I-j} \hat{\mu}_i \sum_{k=0}^{j-1} \hat{\gamma}_k} \\
&= \frac{\sum_{k=0}^{j} \hat{\gamma}_k}{\sum_{k=0}^{j-1} \hat{\gamma}_k}.
\end{aligned} \tag{4.27}$$

We write the Poisson estimator of the aggregate claims from year $i > 0$ as

$$\hat{C}_{iJ}^{Po} = \hat{\mu}_i \sum_{j=I-i+1}^{J} \hat{\gamma}_j + C_{i,I-i} = \hat{\mu}_i \sum_{j=0}^{J} \hat{\gamma}_j. \tag{4.28}$$

We insert in (4.28) for $\hat{\mu}_i$ from

$$C_{i,I-i} = \hat{\mu}_i \sum_{j=0}^{I-i} \hat{\gamma}_j$$

(see (4.19)). Since $I = J$ we have

$$\hat{C}_{iJ}^{Po} = C_{i,I-i} \frac{\sum_{j=0}^{I} \hat{\gamma}_j}{\sum_{j=0}^{I-i} \hat{\gamma}_j},$$

which can be rewritten to

$$\hat{C}_{iJ}^{Po} = C_{i,I-i} \frac{\sum_{j=0}^{I-i+1} \hat{\gamma}_j}{\sum_{j=0}^{I-i} \hat{\gamma}_j} \cdots \frac{\sum_{j=0}^{I} \hat{\gamma}_j}{\sum_{j=0}^{I-1} \hat{\gamma}_j}. \tag{4.29}$$

(4.29) together with (4.27) prove the equivalence of $\hat{C}_{iJ}^{Po}$ with CL estimator (4.5).

Bibliography

[1] S.A.Klugman, H.H.Panjer, G.E.Willmot *Loss Models: From Data to Decisions*, John Wiley & Sons, 1998.

[2] T. Mack *Distribution-free calculation of the standard error of chain ladder reserve estimates*, ASTIN Bulletin 23, 213-225, 1993.

[3] P. Mandl, L. Mazurová *Matematické základy neživotního pojištění*, Matfyzpress, Praha, 1999.

[4] A.J. Mata *Pricing Excess of Loss Reinsurance with Reinstatements*, ASTIN Bulletin 30, 349-368, 2000.

[5] M.Merz, M.V.Wüthrich *Stochastic claims reserving methods in insurance*, John Wiley & Sons, 2008.

[6] P. Mlej *Proces zajištění a oceňování zajistných produktů v praxi*, Aktuárský seminář, MFF UK a ČSpA, 5.12.2014, http://www.actuaria.cz/sdeleni.asp?ID=671.