Risk Theory
lecture notes

Lucie Mazurová

December 2018

Contents

Chapter 1

Counting Processes

Motivation: We want to study a stochastic process $\{S(t),\ t \geq 0\}$ representing the total insurance loss incurred in the time interval $(0, t]$. For this we need to model the occurrence of claims in time. We will deal with the compound model of the form

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

where N_t denotes the number of claims in the interval $(0, t]$.

An appropriate model for the number of claims (events) occurring over time is a **counting process**.

Definition 1.1. *A* ***counting process*** *satisfies the following properties:*

a) The state space of $\{N_t,\ t \geq 0\}$ *is* $\{0, 1, \ldots\}$.

b) All trajectories of $\{N_t,\ t \geq 0\}$ *are non-decreasing and right-continuous.*

Remark. A counting process can be defined by association with a **point process** $\{\sigma_k,\ k = 1, 2, \ldots\}$ representing time points of occurrence of events.

When σ_k denotes the time of occurrence of the k-th event, the corresponding counting process is

$$N_t = \sum_{k=1}^{\infty} \mathrm{I}\left[\sigma_k \leq t\right],\ t \geq 0,$$

where $\mathrm{I}[A]$ is the indicator variable, i.e. it has the value 1 or 0 according to whether the event A occurs or not.

The next section deals with the most important example of counting process which will be used later in developing the collective risk model in continuous time.

1.1 Poisson Process

Definition 1.2. *A counting process* $\{N_t,\ t \geq 0\}$ *is called* ***Poisson process*** *with intensity* $\lambda > 0$*, if it satisfies the following conditions:*

a) $N_0 = 0$ *a.s.*

b) $\{N_t,\ t \geq 0\}$ *has independent increments,*

c) for all $t \geq 0$ *it holds* $\mathrm{P}\left(N_{t+h} - N_t = 1\right) = \lambda\, h + o(h),\ h \to 0+$,

d) for all $t \geq 0$ *it holds* $\mathrm{P}\left(N_{t+h} - N_t > 1\right) = o(h),\ h \to 0+$.

c), d) means that in sufficiently short intervals, no more than one event can occur. It eliminates the possibility of multiple events at a single point in time and the possibility of an infinite number of events in a finite time interval.

From the above definition the distribution of increments of Poisson process can be derived:

For $0 \leq s \leq t$ denote

$$p_n(s,t) = \mathrm{P}\left(N_t - N_s = n\right),\ n = 0, 1, \ldots$$

Let $h > 0$. For the probability of no claims in $(s, t+h]$ it holds

$$\begin{aligned} p_0(s,t+h) &= \mathrm{P}\left(N_t - N_s = 0\right)\ \mathrm{P}\left(N_{t+h} - N_t = 0\right) \\ &= p_0(s,t)\ \left(1 - \lambda\, h + o(h)\right),\ h \to 0+. \end{aligned}$$

Subtracting $p_0(s,t)$ from both sides, dividing by h and letting $h \to 0+$ yields

$$\frac{\partial p_0(s,t)}{\partial t} = -\lambda\, p_0(s,t). \tag{1.1}$$

Similarly, for $n \geq 1$:

$$\begin{aligned} p_n(s,t+h) &= \sum_{m=0}^{n} \mathrm{P}\left(N_t - N_s = m\right)\ \mathrm{P}\left(N_{t+h} - N_t = n - m\right) \\ &= \sum_{m=0}^{n-2} p_m(s,t)\, o(h) + p_{n-1}(s,t)\,(\lambda\, h + o(h)) \\ &\quad + p_n(s,t)\ \left(1 - \lambda\, h + o(h)\right),\ h \to 0+. \end{aligned}$$

Subtracting $p_n(s,t)$ from both sides, dividing by h and letting $h \to 0+$ yields

$$\frac{\partial p_n(s,t)}{\partial t} = \lambda\, p_{n-1}(s,t) - \lambda\, p_n(s,t),\ n = 1, 2, \ldots \tag{1.2}$$

(1.1) and (1.2) is a system of differential equations with initial conditions

$$p_0(s,s) = 1, \; p_n(s,s) = 0 \text{ for } n = 1, 2, \ldots \tag{1.3}$$

From (1.1) and (1.3) it follows

$$p_0(s,t) = \mathrm{e}^{-\lambda\,(t-s)}, \; t \geq s.$$

To derive the solution of (1.2) satisfying (1.3), we write

$$\frac{\partial}{\partial t}\left(\mathrm{e}^{\lambda\,(t-s)}\; p_n(s,t)\right) = \lambda\left(\mathrm{e}^{\lambda\,(t-s)}\; p_{n-1}(s,t)\right), n = 1, 2, \ldots$$

from which we have

$$\frac{\partial^n}{\partial t^n}\left(\mathrm{e}^{\lambda\,(t-s)}\; p_n(s,t)\right) = \lambda^n.$$

Using (1.3), we obtain

$$\mathrm{e}^{\lambda\,(t-s)}\; p_n(s,t) = \frac{\lambda^n}{n!}\,(t-s)^n.$$

We thus conclude that the increment $N_t - N_s$ has Poisson distribution with mean $\lambda\,(t-s)$:

$$p_n(s,t) = \frac{[\lambda\,(t-s)]^n}{n!}\;\mathrm{e}^{-\lambda\,(t-s)}, \; n = 0, 1, \ldots \tag{1.4}$$

We see that (1.4) depends on s and t only through the difference $t-s$. We reformulate the result: the number of events in any interval of length h is Poisson distributed with mean $\lambda\,h$.

We say that Poisson process is a process with **stationary increments**.

Remark. Poisson process can be defined equivalently as a process with independent increments where the increment in an interval of length h has Poisson distribution with mean $\lambda\,h$.

Let $\tau_1 = \sigma_1, \; \tau_k = \sigma_k - \sigma_{k-1}, \; k = 2, 3, \ldots$ denote the sequence of **interoccurrence times** (times between events) of the Poisson process.

Theorem 1.1. *The interoccurrence times of the Poisson process with intensity λ are i.i.d. random variables with exponential distribution with mean $1/\lambda$.*

Proof. For r.v. τ_1 we obtain the exponential distribution from

$$\mathrm{P}(\tau_1 > t) = \mathrm{P}(\sigma_1 > t) = \mathrm{P}(N_t = 0) = \mathrm{e}^{-\lambda t}.$$

We prove the statement just for the first two interoccurrence times $\tau_1, \ \tau_2$ by deriving the joint probability

$$\mathrm{P}(\tau_1 \leq t, \ \tau_2 \leq s) \ \text{ for } t \geq 0, \ s \geq 0. \tag{1.5}$$

Let us denote $h = \frac{t}{n}$ for n sufficiently large, so that $h < s$.

The interval $(0, t]$ is divided into n equally sized intervals, the k-th interval denoted by $((k-1)\,h, k\,h]$.

Using the independence and Poisson distribution of the increments we construct a lower bound for (1.5) as

$$\begin{aligned}
&\sum_{k=1}^{n} \mathrm{P}\left(N_{(k-1)\,h} = 0\right) \, \mathrm{P}\left(N_{k\,h} - N_{(k-1)\,h} = 1\right) \, \mathrm{P}\left(N_{(k-1)\,h+s} - N_{k\,h} > 0\right) \\
&= \sum_{k=1}^{n} \mathrm{e}^{-\lambda\,(k-1)\,h} \, \lambda\,h \, \mathrm{e}^{-\lambda\,h} \, \left(1 - \mathrm{e}^{-\lambda\,(s-h)}\right). \qquad (1.6)
\end{aligned}$$

Similarly, an upper bound for (1.5) is

$$\begin{aligned}
&\sum_{k=1}^{n} \mathrm{P}\left(N_{(k-1)\,h} = 0\right) \, \Big(\mathrm{P}\left(N_{k\,h} - N_{(k-1)\,h} = 1\right) \, \mathrm{P}\left(N_{k\,h+s} - N_{k\,h} > 0\right) \\
&+ \mathrm{P}\left(N_{k\,h} - N_{(k-1)\,h} > 1\right)\Big) \\
&= \sum_{k=1}^{n} \mathrm{e}^{-\lambda\,(k-1)\,h} \, \left(\lambda\,h \, \mathrm{e}^{-\lambda\,h} \, \left(1 - \mathrm{e}^{-\lambda\,s}\right) + o(h)\right), \ h \to 0+. \qquad (1.7)
\end{aligned}$$

Letting $n \to \infty$ (i.e. $h \to 0+$) yields the same limit for (1.6) and (1.7). The joint d.f. (1.5) is

$$\mathrm{P}(\tau_1 \leq t, \tau_2 \leq s) = \left(1 - \mathrm{e}^{-\lambda\,t}\right) \, \left(1 - \mathrm{e}^{-\lambda\,s}\right).$$

□

Remark. $\sigma_k = \sum_{i=1}^{k} \tau_i$ is a sum of i.i.d. exponentially distributed random variables with the d.f. $1 - \mathrm{e}^{-\lambda\,t}, \ t \geq 0$. From this it follows that σ_k has gamma distribution with the density

$$f_k(t) = \frac{\lambda^k}{(k-1)!} \, t^{n-1} \, \mathrm{e}^{-\lambda\,t}, \ n = 1, 2, \ldots$$

From the relationship between $\{N_t,\ t \geq 0\}$ and $\{\sigma_k,\ k = 1, 2, \ldots\}$,

$$\mathrm{P}\left(N_t < k\right) = \mathrm{P}\left(\sigma_k > t\right),$$

we obtain the equality

$$\sum_{n=0}^{k-1} \frac{(\lambda\, t)^n}{n!}\ \mathrm{e}^{-\lambda\, t} = \int_t^\infty \frac{\lambda^k\, x^{k-1}}{(k-1)!}\ \mathrm{e}^{-\lambda\, x}\ \mathrm{d}x. \tag{1.8}$$

(1.8) could be proved without the probabilistic interpretation using integration by parts.

Remark. Memoryless property of exponential distribution. Assume that the history of a Poisson process is known up to time t. There were n events in $(0, t]$ and the last event occurred in time $\sigma_n < t$. We can easily derive the distribution of r.v. τ representing the waiting time until the occurrence of the next event: We denote $s = t - \sigma_n$. For $x \geq 0$ we have

$$\mathrm{P}(\tau > x) = \mathrm{P}\left(\tau_{n+1} > x + s \mid \tau_{n+1} > s\right) = \frac{\mathrm{e}^{-\lambda\,(x+s)}}{\mathrm{e}^{-\lambda\, s}} = \mathrm{e}^{-\lambda\, x}.$$

Waiting time τ has again exponential distribution with the d.f. $1 - \mathrm{e}^{-\lambda\, x}$.

From the property described in the above remark we conclude that we can choose any time t or any time point σ_k as a new origin of the observation and the resulting process is again a Poisson process with the same intensity.

1.2 Nonhomogeneous Poisson Process

We consider a generalization of Poisson process that lies in allowing the intensity to vary with time. In the definition 1.2 we change the condition *c)* to

c') for all $t \geq 0$ *it holds* $\mathrm{P}\left(N_{t+h} - N_t = 1\right) = \lambda(t)\, h + o(h),\ h \to 0+$, *where* $\lambda(t) > 0$ *is a piecewise continuous function.*

From the definition of nonhomogeneous Poisson process we derive the following system of differential equations for the distribution of increments:

$$\frac{\partial}{\partial t} p_0(s, t) = -p_0(s, t)\, \lambda(t), \tag{1.9}$$

$$\frac{\partial}{\partial t} p_n(s, t) = \lambda(t)\, p_{n-1}(s, t) - \lambda(t)\, p_n(s, t),\ n = 1, 2, \ldots \tag{1.10}$$

with the initial conditions

$$p_0(s,s) = 1,\ p_n(s,s) = 0 \text{ for } n = 1, 2, \ldots \tag{1.11}$$

The solution of (1.9) and (1.10) satisfying (1.11) is

$$p_n(s,t) = \mathrm{e}^{-(a(t)-a(s))}\ \frac{(a(t)-a(s))^n}{n!},\ n = 0, 1, \ldots$$

where

$$a(t) = \int_0^t \lambda(s)\ \mathrm{d}s.$$

So the nonhomogeneous Poisson process has again Poisson distributed increments, but they generally are not stationary.

1.3 Birth Processes

Definition 1.3. *A counting process $\{N_t,\ t \geq 0\}$ is called a **birth process** if it satisfies the following conditions:*

a) $N_0 = 0$ *a.s.*

b) The process $\{N_t,\ t \geq 0\}$ *is a Markov process.*

c) for all $t \geq 0$ *it holds* $\mathrm{P}\left(N_{t+h} - N_t = 1 \mid N_t = m\right) = \lambda_m(t)\, h + o(h),\ h \to 0+$, *m=0,1,...*

d) for all $t \geq 0$ *it holds* $\mathrm{P}\left(N_{t+h} - N_t > 1 \mid N_t = m\right) = o(h),\ h \to 0+$, *m=0,1,...*

The distribution of the increment on the time interval $(s,t]$ now has to be considered as conditional on the value of N_s. Denote for $0 \leq s \leq t$

$$p_{m,n}(s,t) = \mathrm{P}\left(N_t - N_s = n \mid N_s = m\right). \tag{1.12}$$

We can derive differential equations for the transition probabilities (1.12) using similar argument as in the case of Poisson process.

Let $h > 0$. From definition 1.3 we have for $m \geq 0$

$$\begin{aligned} p_{m,0}(s,t+h) &= \mathrm{P}\left(N_{t+h} = m \mid N_s = m\right) \\ &= \mathrm{P}\left(N_{t+h} = m \mid N_t = m,\ N_s = m\right) \mathrm{P}\left(N_t = m \mid N_s = m\right) \\ &= \mathrm{P}\left(N_{t+h} = m \mid N_t = m\right) \mathrm{P}\left(N_t = m \mid N_s = m\right) \\ &= p_{m,0}(s,t)\ \left[1 - \lambda_m(t)\, h + o(h)\right],\ h \to 0+. \end{aligned}$$

If we subtract $p_{m,0}(s,t)$ from both sides, divide by h and let $h \to 0+$, we obtain

$$\frac{\partial}{\partial t} p_{m,0}(s,t) = -\lambda_m(t)\, p_{m,0}(s,t). \tag{1.13}$$

Similarly, for $n \geq 1$ and $h > 0$ we have

$$\begin{aligned} p_{m,n}(s,t+h) &= \mathrm{P}\left(N_{t+h} = m+n \mid N_s = m\right) \\ &= \sum_{j=0}^{n} \mathrm{P}\left(N_{t+h} = m+n \mid N_t = m+j,\ N_s = m\right) \mathrm{P}\left(N_t = m+j \mid N_s = m\right) \\ &= \sum_{j=0}^{n} \mathrm{P}\left(N_{t+h} = m+n \mid N_t = m+j\right) \mathrm{P}\left(N_t = m+j \mid N_s = m\right) \\ &= p_{m,n}(s,t)\ [1 - \lambda_{m+n}(t)\, h] + p_{m,n-1}(s,t)\, \lambda_{m+n-1}(t)\, h + o(h),\ h \to 0+. \end{aligned}$$

From here we obtain

$$\frac{\partial}{\partial t} p_{m,n}(s,t) = \lambda_{m+n-1}(t)\, p_{m,n-1}(s,t) - \lambda_{m+n}(t)\, p_{m,n}(s,t),\ n = 1,2,\ldots \tag{1.14}$$

The general solution of (1.13) and (1.14) with the initial conditions

$$p_{m,0}(s,s) = 1,\ p_{m,n}(s,s) = 0 \text{ for } n = 1,2,,\ldots \tag{1.15}$$

satisfies recursive relations

$$\begin{aligned} p_{m,0}(s,t) &= \exp\left\{-\int_s^t \lambda_m(u)\ \mathrm{d}\,u\right\}, \\ p_{m,n}(s,t) &= \int_s^t p_{m,n-1}(s,u)\, \lambda_{m+n-1}(u)\, p_{m+n,0}(u,t)\ \mathrm{d}\,u,\ n = 1,2,\ldots \end{aligned}$$

Remark. For $\lambda_n(t) = \lambda(t)$ independently on n, we obtain a nonhomogeneous Poisson process with the intensity $\lambda(t)$. It follows from the fact that in such a case the system of equations (1.13), (1.14) coincides with the system (1.9) and (1.10), when we replace $p_{m,n}(s,t)$ with $p_n(s,t)$.

We now consider an important special case of a birth process.

1.3.1 Pólya Process

We specify the intensity function of **Pólya process** as

$$\lambda_n(t) = \frac{\lambda + b\, n}{1 + b\, t},\ n = 0,1,\ldots,\ t \geq 0,$$

where $\lambda > 0$, $b > 0$ are parameters of the intensity.

It can be shown from (1.13) and (1.14) that in this particular case the transition probabilities $p_{m,n}(s,t)$ correspond to a negative binomial distribution (see [1, p. 79]).

We restrict ourselves to unconditional probabilities

$$p_n(t) = \mathrm{P}\left(N_t = n\right), \ t \geq 0, \ n = 0, 1, \ldots$$

Obviously, $p_n(t) = p_{0,n}(0,t)$ in the previous notation, so the distribution of N_t solves the system of differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t} p_0(t) = -\lambda_0(t)\, p_0(t) \tag{1.16}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} p_n(t) = \lambda_{n-1}(t)\, p_{n-1}(t) - \lambda_n(t)\, p_n(t), \ n = 1, 2, \ldots \tag{1.17}$$

with the initial conditions

$$p_0(0) = 1, \ p_n(0) = 0 \text{ for } n = 1, 2, \ldots \tag{1.18}$$

We show that the solution of (1.16), (1.17) satisfying (1.18) is the probability function of negative binomial distribution of the form

$$p_n(t) = \binom{k+n-1}{n} \left(\frac{k}{\lambda\, t + k}\right)^k \left(\frac{\lambda\, t}{\lambda\, t + k}\right)^n, \ n = 0, 1, \ldots \tag{1.19}$$

where $k = \lambda / b$.

To prove this, we write (1.19) as

$$p_n(t) = \frac{k\,(k+1)\cdots(k+n-1)}{n!}\, b^n t^n \,(1+bt)^{-(n+k)}$$

and we take the derivative

$$\frac{\mathrm{d}}{\mathrm{d}t} p_n(t) = p_{n-1}(t)\, \frac{(k+n-1)\, b}{1 + b\, t} - p_n(t)\, \frac{n+k}{1 + b\, t}\, b.$$

1.4 Renewal Processes

Definition 1.4. *A point process $\{\sigma_k, \ k = 1, 2, \ldots\}$ (and the corresponding counting process $\{N_t, \ t \geq 0\}$) is called a **renewal process**, if $\sigma_k = \sum_{i=1}^k \tau_i$, where $\{\tau_i, \ i = 1, 2, \ldots\}$ is a sequence of mutually independent positive random variables such that $\tau_2, \ \tau_3, \ldots$ are identically distributed.*

Denote by F the d.f. of the common probability distribution of $\tau_2,\ \tau_3,\dots$ and by F_1 the d.f. of τ_1.

Remark. If $F = F_1$, we call the process **ordinary renewal process**.

Example. For Poisson process with intensity λ we have

$$F(x) = F_1(x) = 1 - \mathrm{e}^{-\lambda\, x},\ x \ge 0.$$

We assume for the sake of simplicity that the d.f. F and F_1 are absolutely continuous, i.e. there exist densities f and f_1 such that

$$F(x) = \int_0^x f(t)\,\mathrm{d}t,\ F_1(x) = \int_0^x f_1(t)\,\mathrm{d}t.$$

We denote by m and s^2 the mean and the variance of $\tau_2,\ \tau_3,\dots$ and we assume that those characteristics are finite:

$$m = \int_0^\infty t\, f(t)\,\mathrm{d}t < +\infty,\ s^2 = \int_0^\infty (t-m)^2\, f(t)\,\mathrm{d}t < +\infty.$$

Theorem 1.2 (Asymptotic normality of N_t)**.** *For $t \to \infty$ the distribution of a renewal process N_t is asymptotically normal, with mean t/m and variance $s^2\, t/m^3$, i.e.*

$$\lim_{t\to\infty} \mathrm{P}\left(\frac{N_t - t/m}{s\,\sqrt{t/m^3}} \le y\right) = \Phi(y),\ y \in \mathbb{R}, \tag{1.20}$$

where Φ denotes the d.f. of standard normal distribution.

Proof. We present here only the main ideas of the proof.

For given $y \in \mathbb{R}$ we denote

$$n_t = \frac{t}{m} + y\, s\,\sqrt{\frac{t}{m^3}}.$$

Then we have

$$\mathrm{P}\left(\frac{N_t - t/m}{s\,\sqrt{t/m^3}} < y\right) = \mathrm{P}\left(N_t < n_t\right) = \mathrm{P}\left(\sigma_{n_t^*} > t\right),$$

where n_t^* is the nearest integer greater than or equal to n_t.

We note that σ_n as a sum of independent and (with the exception of the first summand) identically distributed random variables satisfies the central limit theorem:

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{\sigma_n - n\, m}{\sqrt{n}\, s} \le x\right) = \Phi(x),\ x \in \mathbb{R}.$$

Similarly, it can be shown that

$$\frac{\sigma_{n_t^*} - n_t\, m}{\sqrt{n_t}\, s}$$

converges to $N(0,1)$ as $t \to \infty$.

We obtain

$$\lim_{t\to\infty} \mathrm{P}\left(\sigma_{n_t^*} > t\right) = \lim_{t\to\infty} \mathrm{P}\left(\frac{\sigma_{n_t^*} - n_t\, m}{\sqrt{n_t}\, s} > \frac{-y\, s\, \sqrt{t/m}}{s\, \sqrt{t/m + y\, s\, \sqrt{t/m^3}}} \right)$$
$$= 1 - \Phi(-y) = \Phi(y).$$

□

In what follows, we will study the properties of the **renewal function** defined by

$$h(t) = \mathrm{E}\, N_t,\ \ t \geq 0.$$

We derive an expression of $h(t)$ based on the distribution of σ_k. We denote by f_k the density of σ_k. According to the Definition 1.4, f_k is a convolution

$$f_k = f_1 * f^{(k-1)*},$$

where f^{n*} denotes an n-fold convolution of the density f.

We have

$$h(t) = \sum_{k=1}^{\infty} k\ \mathrm{P}\left(N_t = k\right) = \sum_{k=1}^{\infty} k\ \left[\mathrm{P}\left(N_t \geq k\right) - \mathrm{P}\left(N_t \geq k+1\right)\right]$$
$$= \sum_{k=1}^{\infty} \mathrm{P}\left(N_t \geq k\right) = \sum_{k=1}^{\infty} \mathrm{P}\left(\sigma_k \leq t\right) = \sum_{k=1}^{\infty} \int_0^t f_k(u)\, \mathrm{d}u. \tag{1.21}$$

For further work we introduce a technical tool called **Laplace transform**.

Laplace transform

The **Laplace integral** of a function $f(t),\ t \geq 0$, is

$$f^*(s) = \int_0^{\infty} e^{-s\, t}\, f(t)\, \mathrm{d}t, \tag{1.22}$$

where $s \in \mathbb{C}$. We assume that f is locally integrable on $[0, \infty)$.

If the integral (1.22) converges at $s = s_0$, then it converges for all s with $Re(s) > Re(s_0)$. Therefore, the set of values for which (1.22) converges (the region of convergence) is a half-plane of the form $Re(s) > a$, possibly including some points of the boundary line $Re(s) = a$ (a is known as the **abscissa of convergence**).

Laplace transform of f is defined by (1.22) for s from the region of convergence of the Laplace integral. The unambiguity of the definition of Laplace transform follows from the theorem proved in 1903 by Matyáš Lerch:

Theorem 1.3 (Lerch's theorem)**.** *Let f_1^* and f_2^* be the Laplace itegrals of the functions f_1 and f_2. If f_1^* and f_2^* converge at s_0 and if for some $l > 0$*

$$f_1^*(s_0 + n\,l) = f_2^*(s_0 + n\,l) \text{ for all } n = 0, 1, \dots,$$

then $f_1(t) = f_2(t)$ a.e.

We list some useful properties of Laplace transform.

1) Linearity: for $g(t) = a\,f_1(t) + b\,f_2(t)$ it holds

$$g^*(s) = a\,f_1^*(s) + b\,f_2^*(s).$$

2) Let a be the abscissa of convergence of $f^*(s)$. Let $g(t) = \int_0^t f(u)\,\mathrm{d}u$. Then

$$g^*(s) = \frac{1}{s}\,f^*(s) \text{ for } Re(s) > \max(0, a).$$

3) For f absolutely continuous and $g(t) = f'(t)$ we have

$$g^*(s) = s\,f^*(s) - f(0).$$

4) Laplace transform of a convolution $g(t) = \int_0^t f_1(u)\,f_2(t-u)\,\mathrm{d}u$ is

$$g^*(s) = f_1^*(s)\,f_2^*(s)$$

(provided that both $f_1^*(s)$ and $f_2^*(s)$ converge absolutely).

5) Laplace transform of $f(t) = t^a\,\mathrm{e}^{q\,t},\ a > -1,\ q \in \mathbb{C}$, is

$$f^*(s) = \frac{\Gamma(a+1)}{(s-q)^{a+1}} \text{ for } Re(s) > Re(q).$$

The above result is used when inverting a Laplace transform which has the form

$$f^*(s) = \frac{A(s)}{B(s)},$$

where A, B are polynomials such that the degree of A is less than the degree of B. We assume

$$B(s) = B_0\,(s - s_1)^{n_1} \ldots (s - s_k)^{n_k}.$$

We apply the method of partial fractions expansion to $f^*(s)$ and we obtain

$$\begin{aligned} f^*(s) =& \frac{A_{11}}{(s-s_1)^{n_1}} + \frac{A_{12}}{(s-s_1)^{n_1-1}} + \cdots + \frac{A_{1n_1}}{s-s_1} \\ &+ \frac{A_{21}}{(s-s_2)^{n_2}} + \cdots + \frac{A_{kn_k}}{s-s_k}. \end{aligned}$$

We can find inverse Laplace transform to each of the terms on the right-hand side:

$$f(t) = A_{11}\, \mathrm{e}^{s_1 t}\, \frac{t^{n_1-1}}{(n_1-1)!} + A_{12}\, \mathrm{e}^{s_1 t}\, \frac{t^{n_1-2}}{(n_1-2)!} + \cdots + A_{kn_k}\, \mathrm{e}^{s_k t}\,.$$

6) Derivative of Laplace transform:

$$\frac{\mathrm{d}^{(n)} f^*(s)}{\mathrm{d}s^n} = \int_0^\infty (-1)^n\, t^n\, \mathrm{e}^{-s\,t}\, f(t)\, \mathrm{d}t = g^*(s),$$

where $g(t) = (-1)^n\, t^n\, f(t)$.

7) The relationship between the limit behavior of $f(t)$ for $t \to +\infty$ and the limit behavior of $f^*(s)$ for $s \to 0+$. (We consider a real variable s here.)

Assume $\lim_{t\to\infty} t^{-a}\, f(t) = K$ for some $a > -1$. Then

$$\lim_{s\to 0+} s^{a+1}\, f^*(s) = K\, \Gamma(a+1).$$

Particularly, when $\lim_{t\to\infty} f(t)$ exists, then

$$\lim_{s\to 0+} s\, f^*(s) = \lim_{t\to\infty} f(t). \tag{1.23}$$

We apply properties 1), 2) and 4) to (1.21) and we arrive at the Laplace transform of renewal function $h(t)$:

$$h^*(s) = \sum_{k=1}^{\infty} \frac{1}{s}\, f_1^*(s)\, \left[f^*(s)\right]^{k-1} = \frac{1}{s}\, \frac{f_1^*(s)}{1 - f^*(s)}. \tag{1.24}$$

Remark. (1.24) can be rewritten as

$$h^*(s) = \frac{1}{s}\, f_1^*(s) + h^*(s)\, f^*(s),$$

so the original function h fulfills the following integral equation

$$h(t) = F_1(t) + \int_0^t h(t-y)\, f(y)\, \mathrm{d}y. \tag{1.25}$$

An equation of the form (1.25), where F_1 and f are known functions, is called **renewal equation**. It is met in various applications of mathematics and it is often solved by means of Laplace transform.

From (1.24) an analytic form of $h(t)$ can be obtained in case we are able to find the inverse Laplace transform to the right-hand side of (1.24).

Example. In Poisson process with intensity λ we have $f_1(t) = f(t) = \lambda\, \mathrm{e}^{-\lambda\, t}$ and $f_1^*(s) = f^*(s) = \frac{\lambda}{\lambda+s}$.

Inserting this into (1.24) we obtain

$$h^*(s) = \frac{1}{s}\left(\frac{\lambda}{\lambda+s}\right) / \left(\frac{s}{\lambda+s}\right) = \frac{\lambda}{s^2},$$

from where it is seen that $h(t) = \lambda\, t$.

Similar result is available for a broader class of renewal processes. let us assume that

$$f_1(t) = \frac{1-F(t)}{m},\ t \geq 0, \tag{1.26}$$

where $m = \int_0^\infty t\, f(t)\, \mathrm{d}t < \infty$. Then we have

$$f_1^*(t) = \frac{1}{s}\,\frac{1-f^*(s)}{m}$$

and from (1.24)

$$h^*(s) = \frac{1}{s^2\, m},$$

which means

$$h(t) = \frac{t}{m}. \tag{1.27}$$

Renewal process satisfying (1.26) is called **stationary renewal process**. (Note that in Poisson process $f_1(t) = \lambda\, \mathrm{e}^{-\lambda\, t}$, $F(t) = 1 - \mathrm{e}^{-\lambda\, t}$, $m = 1/\lambda$.)

For a general renewal process, (1.24) can be used to study the asymptotic behavior of the renewal function for $t \to +\infty$.

We start with examining the expected value of the number of events in an interval $(t-a, t]$ for some $a > 0$, and its limit when $t \to +\infty$.

Theorem 1.4 (Blackwell's theorem)**.** *For the renewal function $h(t)$ of a renewal process with $m = \int_0^\infty t\, f(t)\, \mathrm{d}t < \infty$ it holds*

$$\lim_{t \to +\infty} [h(t) - h(t-a)] = \frac{a}{m}. \tag{1.28}$$

Proof. (partial) To prove Blackwell's theorem, we should consider the existence of the limit on the left-hand side of (1.28). We only show its evaluation by means of Laplace transform and (1.24).

Denote $g(t) = h(t) - h(t-a)$ for $t > a$. It is easily derived that

$$g^*(s) = h^*(s) - \mathrm{e}^{-a\,s}\, h^*(s).$$

From (1.24) we have

$$g^*(s) = \frac{1 - \mathrm{e}^{-a\,s}}{s}\, f_1^*(s)\, \frac{1}{1 - f^*(s)}.$$

Given the existence of $\lim_{t\to+\infty} g(t)$, (1.23) yields

$$\lim_{t\to+\infty} g(t) = \lim_{s\to 0+} s\, g^*(s).$$

From here we obtain (1.28) thanks to the fact that

$$\lim_{s\to 0+} \frac{1 - f^*(s)}{s} = -\frac{\mathrm{d}}{\mathrm{d}s} f^*(s)\,|_{s=0} = \int_0^\infty t\, f(t)\, \mathrm{d}t = m$$

and

$$f_1^*(0) = \int_0^\infty f_1(t)\, \mathrm{d}t = 1.$$

□

The renewal function $h(t)$ is continuous (see (1.21)) and for its derivative

$$h'(t) = \lim_{a\to 0} \frac{h(t) - h(t-a)}{a}$$

we conclude from Blackwell's theorem

$$\lim_{t\to\infty} h'(t) = \frac{1}{m}.$$

For the asymptotic behavior of $h(t)$ we thus have the result

$$h(t) \approx \frac{t}{m},\ t \to +\infty. \tag{1.29}$$

Forward recurrence time

Consider a fixed time $t > 0$. We define the **forward recurrence time** as the time until the next event strictly after time t:

$$\tau_t = \sigma_{N_t+1} - t.$$

We use Blackwell's theorem to derive the limit distribution of τ_t when $t \to +\infty$:

Theorem 1.5. *The forward recurrence time of a renewal process converges in distribution as $t \to +\infty$ to the distribution with the density $\frac{1-F(y)}{m}$.*

Proof. We write for a given $x > 0$

$$\begin{aligned}
\mathrm{P}(\tau_t > x) = \mathrm{P}(\sigma_1 > t+x) + \sum_{k=1}^{\infty} \mathrm{P}(\sigma_k \le t,\ \sigma_{k+1} > t+x) \\
&= \mathrm{P}(\sigma_1 > t+x) + \sum_{k=1}^{\infty} \mathrm{P}(\sigma_k \le t,\ \tau_{k+1} > t+x-\sigma_k) \\
&= \int_{t+x}^{\infty} f_1(u)\,\mathrm{d}u + \sum_{k=1}^{\infty} \int_0^t f_k(u) \int_{t-u+x}^{\infty} f(y)\,\mathrm{d}y\,\mathrm{d}u \\
&= \int_{t+x}^{\infty} f_1(u)\,\mathrm{d}u + \sum_{k=1}^{\infty} \int_x^{\infty} f(y) \int_{\max(0,t-y+x)}^{t} f_k(u)\,\mathrm{d}u\,\mathrm{d}y \\
&= \int_{t+x}^{\infty} f_1(u)\,\mathrm{d}u + \int_x^{t+x} f(y)\ (h(t) - h(t+x-y))\,\mathrm{d}y \\
&\quad + h(t) \int_{t+x}^{\infty} f(y)\,\mathrm{d}y. \qquad (1.30)
\end{aligned}$$

We use the fact that the joint density of (σ_k, τ_{k+1}) is $f_k(u)\,f(y)$. The last equality in (1.30) follows from (1.21). The last term at the right-hand side of (1.30) vanishes as $t \to +\infty$, since

$$t \int_{t+x}^{\infty} f(y)\,\mathrm{d}y \le \int_{t+x}^{\infty} y\,f(y)\,\mathrm{d}y \to 0 \text{ for } t \to +\infty$$

(we assume $m < \infty$). We can see that

$$\lim_{t\to+\infty} \mathrm{P}(\tau_t > x) = \lim_{t\to+\infty} \int_x^{t+x} f(y)\ (h(t) - h(t+x-y))\,\mathrm{d}y.$$

We have an integrable majorant for the integral at the right-hand side in the form of *const* $f(y)\,\frac{y-x}{m}$, so we obtain

$$\lim_{t\to+\infty} \mathrm{P}\left(\tau_t > x\right) = \int_x^\infty f(y)\,\frac{y-x}{m}\,\mathrm{d}y. \tag{1.31}$$

We rewrite (1.31) using the integration by parts: For $T > x$ we write

$$\begin{aligned}\int_x^T f(y)\,\frac{y-x}{m}\,\mathrm{d}y &= F(y)\,\frac{y-x}{m}\,|_{y=x}^{T} - \frac{1}{m}\int_x^T F(y)\,\mathrm{d}y \\ &= \frac{1}{m}\int_x^T (1-F(y))\,\mathrm{d}y - (1-F(T))\,\frac{T-x}{m}.\end{aligned}$$

Taking the limit for $T \to +\infty$ yields

$$\lim_{t\to+\infty} \mathrm{P}\left(\tau_t > x\right) = \int_x^\infty \frac{1-F(y)}{m}\,\mathrm{d}y.$$

□

The limit distribution in theorem 1.5 is the distribution of τ_1 in a stationary renewal process. Using similar argument as in the proof of theorem 1.5 we obtain the following result:

Theorem 1.6. *The forward recurrence time of a stationary renewal process has the density* $f_1(y) = \frac{1-F(y)}{m}$.

Proof. Inserting (1.27) into (1.30) we obtain

$$\begin{aligned}\mathrm{P}\left(\tau_t > x\right) &= \int_{t+x}^\infty \frac{1-F(y)}{m} + \int_x^{t+x} f(y)\,\frac{y-x}{m}\,\mathrm{d}y + \frac{t}{m}\int_{t+x}^\infty f(y)\,\mathrm{d}y \\ &= \int_{t+x}^\infty \frac{1-F(y)}{m}\,\mathrm{d}y + F(y)\,\frac{y-x}{m}|_{y=x}^{t+x} - \frac{1}{m}\int_x^{t+x} F(y)\,\mathrm{d}y \\ &\quad + \frac{t}{m}\,(1-F(t+x)) \\ &= \int_{t+x}^\infty \frac{1-F(y)}{m}\,\mathrm{d}y + \frac{1}{m}\int_x^{t+x} (1-F(y))\,\mathrm{d}y = \int_x^\infty \frac{1-F(y)}{m}\,\mathrm{d}y.\end{aligned}$$

□

If we assume that random variables $\tau_2,\ \tau_3,\ldots$ have also a finite variance $s^2 = \int_0^\infty (y-m)^2\,f(y)\,\mathrm{d}y < \infty$, we can derive an expression for the expected value of the distribution with density (1.26):

$$\begin{aligned}\int_0^\infty y\,\frac{1-F(y)}{m}\,\mathrm{d}y &= \lim_{T\to+\infty}\int_0^T y\,\frac{1-F(y)}{m}\,\mathrm{d}y \\ &= \lim_{T\to+\infty}\left\{\frac{y^2}{2m}\,(1-F(y))\,|_{y=0}^T + \int_0^T \frac{y^2}{2m}\,f(y)\,\mathrm{d}y\right\} \\ &= \int_0^\infty \frac{y^2}{2m}\,f(y)\,\mathrm{d}y = \frac{s^2+m^2}{2m}.\end{aligned}$$

(Note that in Poisson process $s^2 = \frac{1}{\lambda^2}$, $m = \frac{1}{\lambda}$.)

Chapter 2

Collective Risk Model with Continuous Time

We present basic results of the traditional ruin theory originating in works of F. Lundberg and further developed by H. Cramér.

We consider development in time of an insurer's **surplus** given by the excess of an initial capital plus premiums collected over claims paid.

We concentrate on the concept of **ruin probability** and its expression or approximation looked at as a tool for assessing the riskiness of the insurer's business (e.g. assessing the level of the safety loading contained in the premium).

2.1 Ruin Probability

For $t \geq 0$, let $U(t)$ denote the surplus of the insurer at time t. We assume that premiums are paid continuously at a constant rate $c > 0$.

We denote by $S(t)$ the aggregate claims up to time t and by u the initial capital,

$$U(0) = u \ a.s.$$

We have a simple model

$$U(t) = u + c\,t - S(t), \ t \geq 0. \tag{2.1}$$

We write

$$S(t) = \sum_{k=1}^{N_t} X_k,$$

where X_i denotes the amount of the ith claim and $\{N_t, \ t \geq 0\}$ is the counting process for the number of claims. We denote by $\{\sigma_k, \ k = 1, 2, \ldots\}$ the point process of claim occurrence times corresponding to $\{N_t, \ t \geq 0\}$.

For results derived in this chapter the following assumptions concerning the aggregate claim process $S(t)$ are crucial:

1. $\{X_k,\ k = 1, 2, \ldots\}$ is a sequence of i.i.d. positive random variables. We denote the d.f. of X_k by $P(x),\ x \geq 0$.

2. $\{N_t,\ t \geq 0\}$ is a Poisson process with intensity $\lambda > 0$.

3. $\{X_k,\ k = 1, 2, \ldots\}$ and $\{\sigma_k,\ k = 1, 2, \ldots\}$ are mutually independent sequences of random variables.

From assumptions 1. and 2. it follows that $S(t)$ has a compound Poisson distribution. From the independence and stationarity of the increments of Poisson process and from assumptions 1.-3. we can see that the compound Poisson process $\{S(t),\ t \geq 0\}$ has also stationary and independent increments.

Namely, the increment

$$S(t+s) - S(t) = \sum_{k=N_t+1}^{N_{t+s}} X_k = \sum_{k=1}^{N_{t+s}-N_t} X'_k,$$

where $\{X'_k,\ k = 1, 2, \ldots\}$ are i.i.d. random variables with d.f. $P(x)$, has a compound Poisson distribution with the expected value

$$\mathrm{E}\,[S(t+s) - S(t)] = \lambda\, s\, p_1,$$

where we denote by $p_1 = \int_0^\infty x\,\mathrm{d}P(x)$ the expected value of one claim size.

In (2.1) we take into account only the risk premium with a certain safety loading. Since c is the amount of premium paid in a unit time interval, we further assume

$$c = (1+\theta)\,\lambda\, p_1, \tag{2.2}$$

where $\theta > 0$ is the **safety loading** and $\lambda\, p_1$ is the expected value of aggregate claims occurred in any time interval of unit length.

We assume that the process (2.1) continues in an infinite horizon under steady conditions (as expressed by constant parameters c, θ, λ, $P(x)$).

We define the random variable

$$T = \inf\,\{t :\ U(t) < 0\} \tag{2.3}$$

as the time when **ruin** occurs (with possible value $T = +\infty$ in case $U(t) \geq 0$ for all $t \geq 0$).

Further, we denote by

$$\psi(u) = \mathrm{P}(T < \infty) \tag{2.4}$$

the **probability of ruin** considered as a function of the initial capital u.

2.2 The Adjustment Coefficient

I some cases the probability (2.4) can be quantified by means of **adjustment coefficient**. To define it we consider a sequence $\{Y_k,\ k = 1, 2, \ldots\}$ of random variables,

$$Y_k = X_k - c\,\tau_k, \tag{2.5}$$

where $\tau_k = \sigma_k - \sigma_{k-1}$ as before. (2.5) is a loss (difference between premiums and claims) incurred in the time interval between the $(k-1)$st and the kth claim. According to assumptions 1.-3., $\{Y_k,\ k = 1, 2, \ldots\}$ are i.i.d. random variables. Let us assume that the distribution of $\{X_k,\ k = 1, 2, \ldots\}$ is such that the moment generating function $M_X(r)$ exists for $r \in (-\infty,\ \gamma)$ for some $0 < \gamma \le +\infty$ and that $\lim_{r\to\gamma-} M_X(r) = +\infty$. We than define the **adjustment coefficient** as the positive solution of

$$M_Y(R) = 1. \tag{2.6}$$

we know from assumptions 2.-3. that τ_k is an exponentially distributed random variable, independent on X_k. Then

$$M_Y(r) = \mathrm{E}\,\mathrm{e}^{r\,(X-c\tau)} = M_X(r)\,\frac{\lambda}{\lambda + c\,r},$$

where $r > -\frac{\lambda}{c}$ and such that $M_X(r)$ exists.

(2.6) becomes

$$\lambda\,M_X(R) = \lambda + c\,R. \tag{2.7}$$

Inserting (2.2) into (2.7) we obtain an equation

$$M_X(R) = 1 + (1+\theta)\,p_1\,R. \tag{2.8}$$

It is

$$M_X'(0) = p_1 < (1+\theta)\,p_1. \tag{2.9}$$

Also

$$M_X''(r) = \mathrm{E}\,X^2\ \mathrm{e}^{rX} > 0 \text{ for } r < \gamma.$$

So, $M_X(r)$ is a convex function tending to $+\infty$ as $r \to \gamma-$. From here and from (2.9) we conclude that there exists one positive solution of (2.8).

The following theorem shows how the probability of ruin is expressed by means of the adjustment coefficient.

Theorem 2.1. *For $u \ge 0$,*

$$\psi(u) = \frac{\mathrm{e}^{-R\,u}}{\mathrm{E}\,[\mathrm{e}^{-R\,U(T)}\mid T < \infty]}. \tag{2.10}$$

Proof. For $t > 0$ we write

$$\begin{aligned}\mathrm{E}\,\mathrm{e}^{-RU(t)} = &\,\mathrm{E}\left[\mathrm{e}^{-RU(t)} \mid T \le t\right]\, \mathrm{P}(T \le t)\\ &+ \mathrm{E}\left[\mathrm{e}^{-RU(t)} \mid T > t\right]\, \mathrm{P}(T > t).\end{aligned} \tag{2.11}$$

Now we rewrite the left-hand side of (2.11) using (2.1)

$$\mathrm{E}\,\mathrm{e}^{-RU(t)} = \mathrm{E}\,\mathrm{e}^{-R(u+ct-S(t))} = \mathrm{e}^{-R\,u} \cdot \mathrm{e}^{-R\,c\,t}\, M_{S(t)}(R).$$

Since $S(t)$ has compound Poisson distribution with m.g.f.

$$M_{S(t)}(r) = \mathrm{e}^{\lambda\, t\, (M_X(r)-1)},$$

we have

$$\mathrm{E}\,\mathrm{e}^{-RU(t)} = \mathrm{e}^{-R\,u} \cdot \mathrm{e}^{-\lambda\, t\, ((1+\theta)\, p_1\, R - M_X(R)+1)} = \mathrm{e}^{-R\,u}, \tag{2.12}$$

thanks to (2.8).

In the first term on the right-hand side of (2.11) we write

$$\begin{aligned}U(t) &= U(T) + [U(t) - U(T)]\\ &= U(T) + c\,(t - T) - [S(t) - S(T)]\,.\end{aligned}$$

For a given T, $S(t) - S(T)$ is independent of $U(T)$ and has a compound Poisson distribution with m.g.f.

$$\mathrm{e}^{\lambda\,(t-T)\,(M_X(r)-1)}\,.$$

Thus

$$\mathrm{E}\left[\mathrm{e}^{-RU(t)} \mid T \le t\right] = \mathrm{E}\left[\mathrm{e}^{-RU(T)} \mid T \le t\right]. \tag{2.13}$$

We insert (2.12) and (2.13) into (2.11). Finally,

$$\begin{aligned}\mathrm{e}^{-Ru} = &\,\mathrm{E}\left[\mathrm{e}^{-RU(T)} \mid T \le t\right]\, \mathrm{P}(T \le t)\\ &+ \mathrm{E}\left[\mathrm{e}^{-RU(t)} \mid T > t\right]\, \mathrm{P}(T > t).\end{aligned} \tag{2.14}$$

It holds

$$\lim_{t\to+\infty} \mathrm{E}\left[\mathrm{e}^{-RU(T)} \mid T \le t\right]\, \mathrm{P}(T \le t) = \mathrm{E}\left[\mathrm{e}^{-RU(T)} \mid T < +\infty\right]\, \psi(u).$$

To finish the proof, we need to show that the second term on the right-hand side of (2.14) vanishes as $t \to +\infty$.

We assume $p_2 = \int_0^\infty x^2 \, \mathrm{d}P(x) < \infty$. We write

$$\begin{aligned} \mathrm{E}\, U(t) &= \mathrm{E}\, [u + c\, t - S(t)] = u + \alpha\, t, \\ \mathrm{Var}\, U(t) &= \mathrm{Var}\, S(t) = \beta^2\, t, \end{aligned}$$

where $\alpha = c - \lambda\, p_1$, $\beta^2 = \lambda\, p_2$.

We consider $u + \alpha\, t - \beta\, t^{2/3}$, which is positive for t sufficiently large, we rewrite the second term on the right-hand side of (2.14) and we find an upper bound from the Chebychev's inequality:

$$\begin{aligned} &\mathrm{E}\left[\mathrm{e}^{-R\,U(t)} \mid T > t,\ 0 \le U(t) \le u + \alpha\, t - \beta\, t^{2/3}\right] \\ &\quad \times \mathrm{P}\left(T > t,\ 0 \le U(t) \le u + \alpha\, t - \beta\, t^{2/3}\right) \\ &+ \mathrm{E}\left[\mathrm{e}^{-R\,U(t)} \mid T > t,\ U(t) > u + \alpha\, t - \beta\, t^{2/3}\right] \\ &\quad \times \mathrm{P}\left(T > t,\ U(t) > u + \alpha\, t + \beta\, t^{2/3}\right) \\ &\le \mathrm{P}\left(U(t) \le u + \alpha\, t - \beta\, t^{2/3}\right) + \exp\left(-R\left(u + \alpha\, t - \beta\, t^{2/3}\right)\right) \\ &\le t^{-1/3} + \exp\left(-R\left(u + \alpha\, t - \beta\, t^{2/3}\right)\right). \end{aligned}$$

We have an upper bound for the second term on the right-hand side of (2.14) that vanishes for $t \to +\infty$.

□

Remark. As the surplus at time of ruin T, $U(T)$, is negative, it follows from Theorem 2.1 immediately that

$$\psi(u) < \mathrm{e}^{-R\,u},$$

an upper bound for the probability of ruin known as **Lundberg's inequality**.

Remark. From Theorem 2.1 it is also seen that $\psi(u) \to 1$ as $R \to 0$, which is the case when $\theta \to 0$ (see the discussion of the solution of (2.8) above).

Example. We illustrate the application of Theorem 2.1 in the simple case when individual claims $X_1,\ X_2, \ldots$ are exponentially distributed with d.f.

$$P(x) = 1 - \mathrm{e}^{-\beta\, x},\ x \ge 0,\ \beta > 0. \tag{2.15}$$

First, we find the adjustment coefficient. We have

$$M_X(r) = \frac{\beta}{\beta - r},\ r < \beta,\ \text{and } p_1 = \frac{1}{\beta}.$$

(2.8) becomes

$$1 + \frac{(1+\theta)\, R}{\beta} = \frac{\beta}{\beta - R},$$

which is a quadratic equation

$$(1+\theta)\, R^2 - \theta\, \beta\, R = 0.$$

The positive solution is

$$R = \frac{\theta\, \beta}{1+\theta}. \tag{2.16}$$

For application of Theorem 2.1 we need the conditional distribution of $U(T)$, given that ruin occurs. Let $\hat{u}$ be the amount of surplus just prior to the time of ruin T. Let us denote by X the size of the claim causing ruin. Then

$$\begin{aligned} \mathrm{P}\,(-U(T) > y) &= \mathrm{P}\,(X > \hat{u} + y \mid X > \hat{u}) \\ &= \frac{\mathrm{e}^{-\beta\,(\hat{u}+y)}}{\mathrm{e}^{-\beta\,\hat{u}}} = \mathrm{e}^{-\beta\, y}\,. \end{aligned}$$

So $-U(T)$, given $T < \infty$, has also exponential distribution with d.f. (2.15) and we have

$$\mathrm{E}\left[\mathrm{e}^{-R\,U(T)} \mid T < \infty\right] = \frac{\beta}{\beta - R}.$$

Combining this with (2.16) gives

$$\psi(u) = \frac{\beta - R}{\beta}\ \mathrm{e}^{-R\,u} = \frac{1}{1+\theta}\ \exp\left\{-\frac{\theta}{1+\theta}\,\frac{u}{p_1}\right\}.$$

In general, an explicit evaluation of the denominator in (2.10) is not possible.

2.3 The First Surplus Below The Initial Level

In this section we present some preparatory results that will be used in the next section for further expressing the probability of ruin.

We consider now the amount of the surplus at the time it first falls below the initial level u (given this ever happens). The main result is summarized in the following theorem:

Theorem 2.2. *The probability that $U(t)$ will ever fall below its initial level u, and will be between $u - y$ and $u - y - \mathrm{d}y$ when it happens for the first time, is*

$$\frac{\lambda}{c}\ (1 - P(y))\ \mathrm{d}y = \frac{1 - P(y)}{(1+\theta)\, p_1}\ \mathrm{d}y,\ \ y > 0.$$

Proof. Let $w(x),\ x<0$, be a bounded function with $w(x)\geq 0$. We define

$$\psi(u;w)=\mathrm{E}\left[w\left(U(T)\right)\mid T<\infty\right]\,\psi(u). \tag{2.17}$$

We may interpret $w(x)$ as a penalty if the surplus at the time of ruin is x. In this case $\psi(u;w)$ is the expected value of the penalty.

We rewrite (2.17) using conditioning on the time to the first claim occurrence, τ_1, and on the size of the first claim, X_1.

$$\psi(u;w)=\mathrm{E}\left[\mathrm{E}\left[w\left(U(T)\right)\mid T<\infty,\ \tau_1,\ X_1\right]\mathrm{P}\left(T<\infty\mid \tau_1,\ X_1\right)\right]. \tag{2.18}$$

We distinguish between two cases: whether the first claim causes ruin or not. From the stationarity and independence of the compound Poisson process $\{S(t),\ t\geq 0\}$ it follows that for $u+c\,\tau_1-X_1\geq 0$ we have

$$\mathrm{E}\left[w\left(U(T)\right)\mid T<\infty,\ \tau_1,\ X_1\right]\mathrm{P}\left(T<\infty\mid \tau_1,\ X_1\right)=\psi\left(u+c\,\tau_1-X_1;w\right),$$

and for $u+c\,\tau_1-X_1<0$ we have $\mathrm{P}\left(T<\infty\mid \tau_1,\ X_1\right)=1$ and

$$\mathrm{E}\left[w\left(U(T)\right)\mid T<\infty,\ \tau_1,\ X_1\right]=w\left(u+c\,\tau_1-X_1\right).$$

The expected value in (2.18) is evaluated with the joint density of $(\tau_1,\ X_1)$ given by $\lambda\ \mathrm{e}^{-\lambda t}\ p(y),\ t\geq 0,\ y\geq 0$.

$$\begin{aligned}\psi(u;w)=&\int_0^\infty\int_0^{u+c\,t}\psi(u+c\,t-y;w)\,\lambda\ \mathrm{e}^{-\lambda t}\ p(y)\,\mathrm{d}y\,\mathrm{d}t\\&+\int_0^\infty\int_{u+c\,t}^\infty w(u+c\,t-y)\,\lambda\ \mathrm{e}^{-\lambda t}\ p(y)\,\mathrm{d}y\,\mathrm{d}t,\end{aligned}$$

which is rewritten with the substitution $x=u+c\,t$ in the form

$$\begin{aligned}\psi(u;w)=&\frac{\lambda}{c}\int_u^\infty \mathrm{e}^{-\frac{\lambda}{c}(x-u)}\left(\int_0^x\psi(x-y;w)\,p(y)\,\mathrm{d}y\right.\\&\left.+\int_x^\infty w(x-y)\,p(y)\,\mathrm{d}y\right)\mathrm{d}x.\end{aligned}$$

We can see that $\psi(u;w)$ is a continuous function of $u>0$ and it has the derivative

$$\begin{aligned}\psi'(u;w)=&\frac{\lambda}{c}\,\psi(u;w)-\frac{\lambda}{c}\int_0^u\psi(u-y;w)\,p(y)\,\mathrm{d}y\\&-\frac{\lambda}{c}\int_u^\infty w(u-y)\,p(y)\,\mathrm{d}y.\end{aligned} \tag{2.19}$$

We will integrate (2.19) over u from 0 to z. The double integrals can be reduced to simple integrals by a change in variables. Then

$$\begin{aligned}\int_0^z \int_0^u \psi(u-y;w)\,p(y)\,\mathrm{d}y\,\mathrm{d}u &= \int_0^z \int_0^u \psi(x;w)\,p(u-x)\,\mathrm{d}x\,\mathrm{d}u \\ = \int_0^z \psi(x;w) \int_x^z p(u-x)\,\mathrm{d}u\,\mathrm{d}x &= \int_0^z \psi(x;w)\,P(z-x)\,\mathrm{d}x,\end{aligned}$$

and

$$\begin{aligned}\int_0^z \int_u^\infty w(u-y)\,p(y)\,\mathrm{d}y\,\mathrm{d}u &= \int_0^z \int_0^\infty w(-x)\,p(x+u)\,\mathrm{d}x\,\mathrm{d}u \\ = \int_0^\infty w(-x) \int_0^z p(x+u)\,\mathrm{d}u\,\mathrm{d}x &= \int_0^\infty w(-x)\,\left(P(x+z)-P(x)\right)\mathrm{d}x.\end{aligned}$$

Thus the integration of (2.19) gives

$$\begin{aligned}\psi(z;w)-\psi(0;w) =& \frac{\lambda}{c}\int_0^z \psi(x;w)\,\left(1-P(z-x)\right)\mathrm{d}x \\ &- \frac{\lambda}{c}\int_0^\infty w(-x)\,\left(P(x+z)-P(x)\right)\mathrm{d}z.\end{aligned}$$

For $z \to +\infty$ the first terms on both sides vanish ($\psi(z) < \mathrm{e}^{-Rz}$ and $w(x)$ is assumed to be bounded), so

$$\psi(0;w) = \frac{\lambda}{c}\int_0^\infty w(-y)\,\left(1-P(y)\right)\mathrm{d}y. \tag{2.20}$$

We use (2.20) with a particular choice of $w(x)$. For $s > 0$ we set

$$\begin{aligned}w_s(x) &= 1,\ x < -s, \\ &= 0,\ -s \le x \le 0.\end{aligned}$$

We obtain

$$\mathrm{P}\left(U(T) < -s \mid T < \infty\right)\,\psi(0) = \psi(0;w_s) = \frac{\lambda}{c}\int_s^\infty \left(1-P(y)\right)\mathrm{d}y.$$

It means that for $u = 0$, the probability that the surplus ever falls below 0 and will be between $-y$ and $-y-\mathrm{d}y$ is $\frac{\lambda}{c}\,(1-P(y))\,\mathrm{d}y$.

For $u > 0$, an event with equal probability is that the surplus will ever fall below u and will be between $u-y$ and $u-y-\mathrm{d}y$.

□

As an application of Theorem 2.2, we note that the probability that $U(t)$ will ever fall below the initial level u is

$$\int_0^\infty \frac{1-P(y)}{(1+\theta)\,p_1}\,\mathrm{d}y = \frac{1}{1+\theta}, \tag{2.21}$$

since

$$p_1 = \int_0^\infty (1-P(y))\,\mathrm{d}y.$$

As the fall of the surplus below the initial level is equivalent to the ruin when $u = 0$, we conclude from (2.21)

$$\psi(0) = \frac{1}{1+\theta}. \tag{2.22}$$

Observe that $\psi(0)$ depends on the safety loading θ and not on the specific form of the claim size distribution.

Let L_1 be a random variable denoting the amount by which the surplus falls below the initial level for the first time, given that this ever happens. We can derive the density of L_1 from Theorem 2.2 and from (2.21):

$$p_1(y) = \left(\frac{1-P(y)}{(1+\theta)\,p_1}\right) / \left(\frac{1}{1+\theta}\right) = \frac{1-P(y)}{p_1},\ y > 0. \tag{2.23}$$

We will also need the m.g.f. of L_1. It is evaluated (at point r for which $M_X(r)$ exists) by integration by parts:

$$\begin{aligned} M_{L_1}(r) &= \frac{1}{p_1}\int_0^\infty \mathrm{e}^{r\,y}\,(1-P(y))\,\mathrm{d}y \\ &= \frac{1}{p_1}\left\{\frac{\mathrm{e}^{r\,y}}{r}\,(1-P(y))\,\Big|_{y=0}^{\infty} + \frac{1}{r}\int_0^\infty \mathrm{e}^{r\,y}\,p(y)\,\mathrm{d}y\right\} \\ &= \frac{1}{p_1\,r}\,(M_X(r)-1)\,. \end{aligned} \tag{2.24}$$

(Note that $\lim_{y\to+\infty} \mathrm{e}^{r\,y}\,(1-P(y)) = 0$ since $M_X(r) < \infty$.)

2.3.1 Maximal Aggregate Loss

The **maximal aggregate loss** is defined as

$$L = \max_{t\geq 0}\left\{S(t) - c\,t\right\},$$

that is, as the maximal excess of aggregate claims over premiums received.

From $S(0) = 0$ a.s. it follows $L \geq 0$ a.s.

There is an obvious connection between the distribution of the maximal aggregate loss and the probability of ruin as a function of the initial capital:

$$\mathrm{P}(L \leq u) = \mathrm{P}\left(S(t) - c\,t \leq u,\ t \geq 0\right) = 1 - \psi(u). \tag{2.25}$$

In particular, we have

$$\mathrm{P}(L = 0) = 1 - \psi(0) = \frac{\theta}{1+\theta}, \tag{2.26}$$

so the distribution of L is of mixed type. There is a point mass of $1 - \psi(0)$ at the origin with the remaining probability distributed continuously over positive values.

In what follows, we use the m.g.f. of L to obtain information about the probability of ruin $\psi(u)$.

An explicit formula for the m.g.f. of L is given in Theorem 2.3.

Theorem 2.3. *The m.g.f. of* L *is*

$$M_L(r) = \frac{\theta\, p_1\, r}{1 + (1+\theta)\, p_1\, r - M_X(r)} \tag{2.27}$$

(provided that $r < \log(1+\theta)$*,* $M_X(r)$ *exists and* $r \neq R$*,* $r \neq 0$*).*

Proof. The proof is based on the fact that the compound Poisson process $\{S(t),\ t \geq 0\}$ has stationary and independent increments.

We consider the times when the loss process, $S(t) - c\,t$, assumes new record highs. We conclude that when a new record high is established, the amount of the increase in the maximum of $S(t) - c\,t$ is distributed as L_1, introduced in the previous section. We can also see that after each record high there is a probability of $1 - \psi(0)$ that this record will not be broken and $\psi(0)$ that it will be broken. To sum up, the maximal aggregate loss has a representation of the form

$$L = \sum_{i=1}^{N} L_i, \tag{2.28}$$

where $L_1,\ L_2, \ldots$ are i.i.d. random variables with the density (2.23) and N has a geometric distribution with the probability function

$$\mathrm{P}(N = n) = (1 - \psi(0))\ \psi(0)^n = \theta \left(\frac{1}{1+\theta}\right)^{n+1},\ n = 0, 1, \ldots$$

corresponding to the event that a new record high is attained by the loss process n times in the infinite horizon.

The p.g.f. of N is

$$\Pi_N(r) = \mathrm{E}\, r^N = \frac{\theta}{1+\theta} \sum_{n=0}^{\infty} \left(\frac{r}{1+\theta}\right)^n$$
$$= \frac{\theta}{1+\theta-r} \quad \text{for } |r| < 1+\theta. \tag{2.29}$$

Combining (2.24) and (2.29) we obtain

$$M_L(r) = \Pi_N\left(M_{L_1}(r)\right) = \frac{\theta}{1+\theta-\frac{M_X(r)-1}{p_1\, r}}$$
$$= \frac{\theta\, p_1\, r}{1+(1+\theta)\, p_1\, r - M_X(r)},$$

which was to be shown.

□

Remark. From (2.28) it follows the representation of the probability of ruin in the form (Beekman's formula)

$$\psi(u) = \mathrm{P}(L > u) = \sum_{n=0}^{\infty} \theta \left(\frac{1}{1+\theta}\right)^{n+1} \int_u^{\infty} \left[\frac{1-P(y)}{p_1}\right]^{n*} \mathrm{d}y. \tag{2.30}$$

An equivalent formula for $M_L(r)$ is

$$M_L(r) = \frac{\theta}{1+\theta} \frac{(1+\theta)\, p_1\, r}{1+(1+\theta)\, p_1\, r - M_X(r)}$$
$$= \frac{\theta}{1+\theta} + \frac{\theta}{1+\theta} \frac{M_X(r)-1}{1+(1+\theta)\, p_1\, r - M_X(r)}. \tag{2.31}$$

It reflects the contribution to the m.g.f. of the point mass (2.26) at the origin.

As L has a positive density $-\psi'(u)$ for positive values of u, it follows from (2.31) that

$$\int_0^{\infty} \mathrm{e}^{r\,u}\, (-\psi'(u))\, \mathrm{d}u = \frac{\theta}{1+\theta} \frac{M_X(r)-1}{1+(1+\theta)\, p_1\, r - M_X(r)}. \tag{2.32}$$

(2.32) can be inverted to find an explicit expression for $\psi(u)$ for certain families of claim size distributions. We show the application of (2.32) in the following example.

Example. Let the density of individual claim sizes be a mixture of exponential densities, i.e.

$$p(x) = \sum_{i=1}^{n} A_i\, \beta_i\, \mathrm{e}^{-\beta_i\, x},$$

where $\beta_i > 0$, $A_i > 0$, $\sum_{i=1}^{n} A_i = 1$.

In this case,

$$M_X(r) = \sum_{i=1}^{n} A_i\, \frac{\beta_i}{\beta_i - r} \tag{2.33}$$

and $M_X(r)$ is defined for $r < \min(\beta_1, \ldots, \beta_n)$.

We substitute (2.33) into (2.32) and we see that the right-hand side is a rational function of r. We apply the method of partial fractions and obtain

$$\begin{aligned} \int_0^\infty \mathrm{e}^{r\, u}\, (-\psi'(u))\, \mathrm{d}u &= \sum_{i=1}^{n} \frac{C_i\, r_i}{r_i - r} \\ &= \int_0^\infty \mathrm{e}^{r\, u} \left(\sum_{i=1}^{n} C_i\, r_i\, \mathrm{e}^{-r_i\, u} \right) \mathrm{d}u. \end{aligned} \tag{2.34}$$

Note that (2.34) means the equality of Laplace integrals at all points $-r$, $r < \min(\beta_1, \ldots, \beta_n)$. Hence we have

$$-\psi'(u) = \sum_{i=1}^{n} C_i\, r_i\, \mathrm{e}^{-r_i\, u} \quad a.e.$$

due to Lerch's theorem.

From $\lim_{u \to +\infty} \psi(u) = 0$ we obtain the unique solution as

$$\psi(u) = \sum_{i=1}^{n} C_i\, \mathrm{e}^{-r_i\, u}\, .$$

In the case when an explicit formula for $\psi(u)$ cannot be found, Theorem 2.3 is used to derive various approximations.

One such approximation has the form

$$\psi(u) \doteq \psi(0)\, \mathrm{e}^{-K\, u}$$

for some positive constant K. We obtain the value of K from the requirement

$$\mathrm{E}\, L = \int_0^\infty \psi(u)\, \mathrm{d}u = \psi(0) \int_0^\infty \mathrm{e}^{-K\, u}\, \mathrm{d}u = \frac{1}{(1+\theta)\, K}. \tag{2.35}$$

Since L has a compound distribution, we have from (2.28)

$$\mathrm{E}\, L = \mathrm{E}\, N \;\mathrm{E}\, L_1,$$

where

$$\mathrm{E}\, N = \frac{\theta}{(1+\theta)^2} \sum_{n=0}^{\infty} n \left(\frac{1}{1+\theta}\right)^{n-1} = \frac{1}{\theta}$$

and

$$\begin{aligned} \mathrm{E}\, L_1 &= \int_0^\infty y\, \frac{1-P(y)}{p_1}\, \mathrm{d}y \\ &= \frac{1}{p_1} \left\{ \frac{y^2}{2}\, (1-P(y)) \,\Big|_{y=0}^{\infty} + \frac{1}{2} \int_0^\infty y^2\, p(y)\, \mathrm{d}y \right\} = \frac{p_2}{2\, p_1}, \end{aligned}$$

if we assume $p_2 = \int_0^\infty y^2\, p(y)\, \mathrm{d}y < \infty$.

From (2.35),

$$K = \frac{2\,\theta\, p_1}{(1+\theta)\, p_2},$$

and the resulting approximation is

$$\psi(u) \doteq \frac{1}{1+\theta}\, \exp\left\{ -\frac{2\,\theta\, p_1\, u}{(1+\theta)\, p_2} \right\}, \;\; u > 0.$$

Another well known approximation of $\psi(u)$ for large values of u follows from the asymptotic formula

$$\psi(u) \approx C\; \mathrm{e}^{-Ru} \;\text{ for } u \to +\infty.$$

More precisely,

Theorem 2.4 (Cramér's asymptotic formula)**.** *When the adjustment coefficient R exists, it holds*

$$\psi(u) = \frac{\theta\, p_1}{M_X'(R) - (1+\theta)\, p_1}\; \mathrm{e}^{-R\,u} + O\left(\mathrm{e}^{-(R+\epsilon)\,u}\right), \;\; u \to +\infty,$$

where $\epsilon > 0$ is such that $\int_0^\infty \mathrm{e}^{(R+\epsilon)\,x}\, \mathrm{d}P(x) < \infty$.

Proof. (part)

We only prove

$$\lim_{u \to +\infty} \psi(u)\, e^{R\,u} = \frac{\theta\, p_1}{M_X'(R) - (1+\theta)\, p_1}, \tag{2.36}$$

without proving the existence of the limit in (2.36).

By considering two possibilities for the ruin with the initial capital u - that the ruin occurs when the surplus falls below the initial level for the first time (i.e. the amount of the drop is greater than u) and that the first fall below the initial level does not cause the ruin. From the properties of the compound Poisson process $S(t)$ we have

$$\psi(u) = \frac{1}{1+\theta}\int_u^\infty \frac{1-P(y)}{p_1}\,\mathrm{d}y + \frac{1}{1+\theta}\int_0^u \psi(u-y)\,\frac{1-P(y)}{p_1}\,\mathrm{d}y \tag{2.37}$$

We rewrite (2.37) as

$$\begin{aligned}\psi(u)\ \mathrm{e}^{Ru} =& \frac{\mathrm{e}^{Ru}}{(1+\theta)\,p_1}\int_u^\infty (1-P(y))\,\mathrm{d}y \\ &+ \int_0^u \psi(u-y)\ \mathrm{e}^{-R(u-y)}\ \frac{\mathrm{e}^{Ry}\ (1-P(y))}{(1+\theta)\,p_1}\,\mathrm{d}y.\end{aligned} \tag{2.38}$$

It is a special case of a renewal equation (see (1.25))

$$h(u) = g(u) + \int_0^u h(u-y)\,f(y)\,\mathrm{d}y \tag{2.39}$$

for $h(u) = \psi(u)\ \mathrm{e}^{Ru}$ with

$$g(u) = \frac{\mathrm{e}^{Ru}}{(1+\theta)\,p_1}\int_u^\infty (1-P(y))\,\mathrm{d}y$$

and

$$f(y) = \frac{\mathrm{e}^{Ry}\ (1-P(y))}{(1+\theta)\,p_1}.$$

From (2.8) we have

$$(1+\theta)\,p_1 = \int_0^\infty \frac{\mathrm{e}^{Ry}-1}{R}\,p(y)\,\mathrm{d}y = \int_0^\infty \mathrm{e}^{Ry}\ (1-P(y))\,\mathrm{d}y,$$

where the second equality is derived by integration by parts. Hence,

$$\int_0^\infty f(y)\,\mathrm{d}y = 1$$

and f is a probability density function with the expectation

$$\begin{aligned}\mu &= \int_0^\infty y\,f(y)\,\mathrm{d}y = \int_0^\infty y\ \mathrm{e}^{Ry}\ \frac{1-P(y)}{(1+\theta)\,p_1}\,\mathrm{d}y \\ &= \frac{1}{(1+\theta)\,p_1\,R}\left\{-\int_0^\infty \mathrm{e}^{Ry}\ (1-P(y))\,\mathrm{d}y + \int_0^\infty y\ \mathrm{e}^{Ry}\ p(y)\ \mathrm{d}y\right\} \\ &= \frac{M_X'(R)-(1+\theta)\,p_1}{(1+\theta)\,p_1\,R}.\end{aligned} \tag{2.40}$$

From (2.39) Laplace transform of $h(u)$ is derived as

$$h^*(s) = \frac{g^*(s)}{1 - f^*(s)}.$$

If $\lim_{u\to+\infty} h(u)$ exists, from (1.23) we deduce

$$\begin{aligned} \lim_{u\to+\infty} h(u) &= \lim_{s\to 0+} s\, h^*(s) = \lim_{s\to 0+} \frac{g^*(s)}{\frac{1-f^*(s)}{s}} \\ &= \frac{\int_0^\infty g(y)\,\mathrm{d}y}{\mu}, \end{aligned} \tag{2.41}$$

provided that $\int_0^\infty |g(y)|\,\mathrm{d}y < \infty$.

In our case

$$\begin{aligned} \int_0^\infty g(u)\,\mathrm{d}u &= \frac{1}{(1+\theta)\,p_1} \int_0^\infty \mathrm{e}^{R\,u} \int_u^\infty (1 - P(y))\,\mathrm{d}y\,\mathrm{d}u \\ &= \frac{1}{(1+\theta)\,p_1} \frac{1}{R} \int_0^\infty \left(\mathrm{e}^{R\,y} - 1\right) (1 - P(y))\,\mathrm{d}y \\ &= \frac{1}{(1+\theta)\,p_1\,R} \left((1+\theta)\,p_1 - p_1\right) = \frac{\theta\,p_1}{(1+\theta)\,p_1\,R}. \end{aligned} \tag{2.42}$$

Inserting (2.40) and (2.42) into (2.41) (when assuming the existence of the limit) we obtain

$$\lim_{u\to+\infty} \psi(u)\,\mathrm{e}^{R\,u} = \frac{\theta\,p_1}{M_X'(R) - (1+\theta)\,p_1}.$$

□

2.4 Subexponential Claim Sizes

The results for the probability of ruin $\psi(u)$ presented in the previous sections use the adjustment coefficient defined in (2.6). For the adjustment coefficient to be defined we need the existence of $M_X(r)$ for some positive values of r.

One of the most popular models for insurance claims is Pareto distribution which is particularly useful when claims data indicate heavy tails of the corresponding distribution. We recall that Pareto distribution has d.f.

$$P(x) = 1 - \left(\frac{x}{a}\right)^{-\alpha},\ x \geq a,\ \alpha > 0,\ a > 0, \tag{2.43}$$

and the density

$$p(x) = \alpha\, a^\alpha\, x^{-\alpha-1},\ x \geq a.$$

Obviously,

$$M_X(r) = \int_0^{\infty} \mathrm{e}^{r\,x}\, p(x)\,\mathrm{d}x = +\infty \text{ for all } r > 0,$$

hence the methods of the preceding sections are not applicable to assess the probability of ruin.

We will show an asymptotic formula for $\psi(u)$ suitable for Pareto distributed claim sizes.

First, we note that Pareto distribution can be classified as **subexponential distribution**.

Definition 2.1. *The distribution of a nonnegative random variable is **subexponential**, if its d.f. $P(x)$ satisfies*

$$\lim_{x\to+\infty} \frac{1 - P^{n*}(x)}{1 - P(x)} = n, \ \ n = 2, 3, \ldots \tag{2.44}$$

To give an interpretation to Definition 2.1, we consider i.i.d. random variables $X_1, \ldots, X_n$ with the d.f. $P(x)$ satisfying (2.44). Then

$$\lim_{x\to+\infty} \frac{\mathrm{P}\left(X_1 + \cdots + X_n > x\right)}{\mathrm{P}\left(\max\left(X_1, \ldots, X_n\right) > x\right)} = \lim_{x\to+\infty} \frac{1 - P^{n*}(x)}{1 - P^{n}(x)}$$
$$= \lim_{x\to+\infty} \frac{1 - P^{n*}(x)}{\left(1 - P(x)\right)\,\left(1 + \cdots + P^{n-1}(x)\right)} = 1.$$

It means that for x large, the probability that the aggregate loss exceeds the threshold x is close to the probability that at least one claim is greater than x (large aggregate loss is caused by one large claim).

Remark. Since for nonnegative i.i.d. random variables it always holds

$$\frac{\mathrm{P}\left(X_1 + \cdots + X_n > x\right)}{\mathrm{P}\left(\max\left(X_1, \ldots, X_n\right) > x\right)} \geq 1,$$

the inequality

$$\liminf_{x\to+\infty} \frac{1 - P^{n*}(x)}{1 - P(x)} \geq n, \ \ n = 2, 3, \ldots$$

is always satisfied. To prove subexponentiality, it suffices to verify

$$\limsup_{x\to+\infty} \frac{1 - P^{n*}(x)}{1 - P(x)} \leq n, \ \ n = 2, 3, \ldots \tag{2.45}$$

Evaluating convolutions of d.f. $P(x)$ is not practically convenient, we present a sufficient condition for subexponentiality based only on the d.f. $P(x)$ itself.

Theorem 2.5. *Assume that for all t, $0 < t \leq 1$, there exists the limit*

$$\gamma(t) = \lim_{x \to +\infty} \frac{1 - P(t\,x)}{1 - P(x)} \tag{2.46}$$

continuous for $t = 1$. Then $P(x)$ is subexponential.

Proof. We will show that (2.45) holds for $n = 2$. (For $n = 3, 4, \ldots$ we could proceed using induction.)

We take $x > 0$ and we divide the interval $[0, x]$ into n pieces of length $\frac{x}{n}$. We denote $u_k = k\,\frac{x}{n}$. Since $P(x - u)$ is non-increasing in u, we obtain a lower bound for the convolution $P^{2*}(x)$:

$$\begin{aligned} P^{2*}(x) &= \int_0^x P(x-u)\;\mathrm{d}P(u) \\ &\geq P\left(u_1\right)\;P\left(x - u_1\right) + \sum_{k=1}^{n-2} P\left(x - u_{k+1}\right)\;\left(P\left(u_{k+1}\right) - P\left(u_k\right)\right). \end{aligned}$$

Hence,

$$\begin{aligned} 1 - P^{2*}(x) \leq & 1 - P\left(x - u_1\right) + \left(1 - P\left(u_1\right)\right)\;P\left(x - u_1\right) \\ & - \sum_{k=1}^{n-2} P\left(x - u_{k+1}\right)\;\left[1 - P\left(u_k\right) - \left(1 - P\left(u_{k+1}\right)\right)\right]. \end{aligned} \tag{2.47}$$

Applying (2.46) to the right-hand side of (2.47) gives

$$\begin{aligned} \limsup_{x \to +\infty} \frac{1 - P^{2*}(x)}{1 - P(x)} &\leq \gamma\left(\frac{n-1}{n}\right) + \gamma\left(\frac{1}{n}\right) - \sum_{k=1}^{n-2}\left(\gamma\left(\frac{k}{n}\right) - \gamma\left(\frac{k+1}{n}\right)\right) \\ &= 2\,\gamma\left(\frac{n-1}{n}\right). \end{aligned}$$

from here (2.45) follows when $n \to +\infty$, thanks to the continuity of $\gamma(t)$ at $t = 1$. □

Example. The subexponetiality of Pareto distribution with d.f. (2.43) is verified by (2.46), since

$$\gamma(t) = \lim_{x \to +\infty} \frac{\left(\frac{t\,x}{a}\right)^{-\alpha}}{\left(\frac{x}{a}\right)^{-\alpha}} = t^{-\alpha}$$

is continuous at $t = 1$.

For assessing the probability of ruin $\psi(u)$, we will use Beekman's formula (2.30). If we denote by P_1 the d.f. of random variable L_1 with the density given in (2.23), we can rewrite (2.30) as

$$\psi(u) = \sum_{n=0}^{\infty} \theta \left(\frac{1}{1+\theta} \right)^{n+1} (1 - P_1^{n*}(u)) . \tag{2.48}$$

Now suppose that the claim size d.f. $P(x)$ satisfies (2.46). Then

$$\begin{aligned} \lim_{x \to +\infty} \frac{1 - P_1(t\,x)}{1 - P_1(x)} &= \lim_{x \to +\infty} \frac{\int_{tx}^{\infty} \frac{1-P(y)}{p_1} \,\mathrm{d}y}{\int_{x}^{\infty} \frac{1-P(y)}{p_1} \,\mathrm{d}!y} = \lim_{x \to +\infty} \frac{\int_{x}^{\infty} \frac{1-P(t\,u)}{p_1} \,t \,\mathrm{d}u}{\int_{x}^{\infty} \frac{1-P(u)}{p_1} \,\mathrm{d}u} \\ &= t\,\gamma(t), \end{aligned}$$

by L'Hospital's rule.

Hence, the d.f. P_1 is subexponential. We use (2.44) and (2.48) to derive

$$\lim_{n \to +\infty} \frac{\psi(u)}{1 - P_1(u)} = \sum_{n=1}^{\infty} \theta\, n \left(\frac{1}{1+\theta} \right)^{n+1} = \frac{1}{\theta}.$$

Thus the assymptotic formula for the probability of ruin is

$$\psi(u) \approx \frac{1}{\theta} \,(1 - P_1(u)) = \frac{1}{\theta\, p_1} \int_u^{\infty} (1 - P(y)) \,\mathrm{d}y, \;\; u \to +\infty. \tag{2.49}$$

Example. For Pareto distribution with the d.f. (2.43) we have

$$p_1 = a \,\frac{\alpha}{\alpha - 1} \;\text{ for } \alpha > 1.$$

In this case we can use

$$\begin{aligned} 1 - P_1(u) &= \frac{a^{\alpha}}{a\, \frac{\alpha}{\alpha-1}} \int_u^{\infty} x^{-\alpha} \,\mathrm{d}x = \frac{a^{\alpha}}{a\, \frac{\alpha}{\alpha-1}} \left(- \frac{u^{-\alpha+1}}{-\alpha+1} \right) \\ &= \frac{1}{\alpha} \left(\frac{u}{a} \right)^{1-\alpha}, \;\; u \to +\infty. \end{aligned}$$

From (2.49) we have

$$\psi(u) \approx \frac{1}{\alpha\,\theta} \left(\frac{u}{a} \right)^{1-\alpha}, \;\; u \to +\infty.$$

Chapter 3

Extreme Value Theory

Extreme value theory (EVT) deals with modeling and analysing events that occur with low frequency, but have a significant (usually negative) impact.

Examples from the area of finance and insurance are large insurance claims caused by catastrophic events or losses observed in series of returns of financial assets. Historically, an area important for the development of mathematical extreme value models was hydrology, namely flood frequency analysis based on observation of water levels. There are many other areas of application of EVT, e.g. analysis of strength of materials, concentration of ecological quantities, seismic risk analysis etc.

EVT is based on parametric models - parameters are estimated from available data and are used for inference about events that are rarely observed. We distinguish between two types of models according to the type of data used for the estimation.

The most traditional models are **block maxima** models. In this case observed data $X_1, X_2, \ldots$ are grouped into m blocks of size n. The models work with block maxima $M_{n1}, \ldots, M_{nm}$. Examples are annual maxima of daily observed water levels, annual maximal losses in series of daily stock returns.

Models for **threshold exceedances** explore from data $X_1, X_2, \ldots$ the values that exceed some high level. These models are considered to be more useful for applications, due to their more efficient use of data.

3.1 Extreme Value Distributions

Parametric models used in block maxima analysis are called **extreme value distributions** (EVD).

They are characterized as limit distributions for sequences of appropriately normalized maxima M_n of n i.i.d. random variables (for $n \to +\infty$). (Note that EVD play a role analogous to that of the normal distribution for sums of random variables.)

There are three basic types of distributional families belonging to EVD. We start with presenting their distribution functions in so called α-**parametrization**.

The d.f. of **Gumbel distribution** is defined for $x \in \mathbb{R}$ by

$$G_0(x) = \exp\left(-\,\mathrm{e}^{-x}\right). \tag{3.1}$$

The standard form of the d.f. of **Fréchet distribution** with a parameter $\alpha > 0$ is defined for $x > 0$ by

$$G_{1,\alpha}(x) = \exp\left(-x^{-\alpha}\right). \tag{3.2}$$

The standard form of the d.f. of **Weibull distribution** with a parameter $\alpha < 0$ is defined for $x < 0$ by

$$G_{2,\alpha}(x) = \exp\left(-(-x)^{-\alpha}\right). \tag{3.3}$$

(3.1)-(3.3) are standard versions of the three types of EVD distribution functions. The complete distributional families are obtained by introducing a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$:

$$G_{0,\mu,\sigma} = \exp\left(-e^{-\frac{x-\mu}{\sigma}}\right), \; x \in \mathbb{R}, \tag{3.4}$$

$$G_{1,\alpha,\mu,\sigma}(x) = \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}\right), \; x > \mu, \; \alpha > 0, \tag{3.5}$$

$$G_{2,\alpha,\mu,\sigma}(x) = \exp\left(-\left(-\frac{x-\mu}{\sigma}\right)^{-\alpha}\right), \; x < \mu, \; \alpha < 0. \tag{3.6}$$

Generalized extreme value distribution (GEVD). The generalized form of EVD encompasses all three above mentioned distributional families. It is obtained by reparametrization $\gamma = \frac{1}{\alpha}$ and by a specific choice of location and scale parameters.

For Fréchet distribution we have $\gamma > 0$ and we choose $\mu = -\frac{1}{\gamma}$, $\sigma = \frac{1}{\gamma}$. We obtain

$$G_{1,\frac{1}{\gamma},-\frac{1}{\gamma},\frac{1}{\gamma}}(x) = \exp\left(-(1+\gamma\, x)^{-1/\gamma}\right), \; x > -\frac{1}{\gamma}. \tag{3.7}$$

For Weibull distribution we have $\gamma < 0$ and we choose $\mu = -\frac{1}{\gamma}$, $\sigma = -\frac{1}{\gamma}$. It leads to

$$G_{2,\frac{1}{\gamma}-\frac{1}{\gamma},-\frac{1}{\gamma}}(x) = \exp\left(-(1+\gamma\, x)^{-1/\gamma}\right),\ x < -\frac{1}{\gamma}. \tag{3.8}$$

Both distribution functions have the same form, they differ in the sign of parameter γ and they also have different supports. We observe that the d.f. of Gumbel distribution is obtained as a limit for $\gamma \to 0$:

$$G_0(x) = \lim_{\gamma\to 0} \exp\left(-(1+\gamma\, x)^{-1/\gamma}\right).$$

We summarize the definition of GEVD d.f. as

$$G_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma\, x)^{-1/\gamma}\right), & \gamma \neq 0, \\ \exp\left(-e^{-x}\right), & \gamma = 0, \end{cases} \tag{3.9}$$

where the support is defined by the condition $1+\gamma\, x > 0$ (which gives $x \in \mathbb{R}$ in case $\gamma = 0$).

The value of parameter γ defines the type of the distribution. We can consider also location and scale parameters $\mu \in \mathbb{R}$, $\sigma > 0$, and work with the three-parameter family

$$G_{\gamma,\mu,\sigma}(x) = \begin{cases} \exp\left(-\left(1+\gamma\,\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right), & \gamma \neq 0, \\ \exp\left(-e^{-\frac{x-\mu}{\sigma}}\right), & \gamma = 0, \end{cases} \tag{3.10}$$

where $1+\gamma\,\frac{x-\mu}{\sigma} > 0$.

Thus, Fréchet d.f. ($\gamma > 0$) is given by (3.10) in the interval $\left(\mu - \frac{\sigma}{\gamma}, +\infty\right)$ and Weibul d.f. ($\gamma < 0$) is given by (3.10) in the interval $\left(-\infty, \mu + \frac{\sigma}{|\gamma|}\right)$.

3.2 Limiting Distributions of Maxima

The role of GEVD in the analysis of maxima is based on Fisher-Tippet theorem.

Asssume that $X_1,\ X_2,\ldots$ is a sequence of i.i.d. random variables with d.f. F and denote

$$M_n = \max\left(X_1,\ldots,X_n\right).$$

The d.f. of M_n is

$$\mathrm{P}\left(M_n \le x\right) = \mathrm{P}\left(X_1 \le x,\ldots,X_n \le x\right) = F^n(x).$$

Theorem 3.1 (Fisher-Tippet theorem)**.** *If there exist sequences of real constants* $\{c_n\}$ *and* $\{d_n\}$*, where* $c_n > 0$ *for all* n*, such that*

$$\lim_{n\to+\infty} \mathrm{P}\left(\frac{M_n - d_n}{c_n} \le x\right) = G(x) \tag{3.11}$$

for some non-degenerate d.f. G*, then* G *is d.f. of GEVD.*

Proof. (Sketch.)

We only mention here the result from the probability theory that is used to obtain the three types of limit distribution of maxima. We say that random variables X and Y **are of the same type** (and so are the corresponding distributions and distribution functions), if there exist constants $a \in \mathbb{R}$ and $b > 0$ such that $X \stackrel{d}{=} bY + a$.

Theorem 3.2 (Convergence to types theorem)**.** *Let* A, B, A_1, $A_2, \dots$ *be random variables and* $b_n > 0$, $\beta_n > 0$ *and* $a_n \in \mathbb{R}$, $\alpha_n \in \mathbb{R}$ *be constants. Suppose that for* $n \to \infty$

$$\frac{A_n - a_n}{b_n} \stackrel{d}{\to} A.$$

Then the relation

$$\frac{A_n - \alpha_n}{\beta_n} \stackrel{d}{\to} B \tag{3.12}$$

holds for $n \to +\infty$ *if and only if*

$$\lim_{n\to\infty} \frac{b_n}{\beta_n} = b \in [0, +\infty)$$

and

$$\lim_{n\to\infty} \frac{a_n - \alpha_n}{\beta_n} = a \in \mathbb{R}.$$

If (3.12) *holds, then* $B \stackrel{d}{=} bA + a$ *and* a, b *are the unique constants for which this holds.*

When (3.12) *holds,* A *is non-degenerate if and only if* $b > 0$ *and then* A *and* B *are of the same type.*

Convergence to types theorem tells us that the limit distribution for a sequence of random variables is uniquely determined up to changes in location and scale.

We sketch its application to the convergence of normalized maxima:
From (3.11) it follows for all $t > 0$

$$\lim_{n\to\infty} F^{[nt]}\left(c_{[nt]}\, x + d_{[nt]}\right) = G(x),$$

where [] denotes the integer part.

At the same time,

$$\lim_{n\to\infty} F^{[nt]}(c_n x + d_n) = \lim_{n\to\infty} \left(F^n(c_n x + d_n)\right)^{[nt]/n} = G^t(x),$$

so by the convergence to types theorem there exist functions $\gamma(t) > 0$, $\delta(t) \in \mathbb{R}$ satisfying

$$\lim_{n\to\infty} \frac{c_n}{c_{[nt]}} = \gamma(t),$$

$$\lim_{n\to\infty} \frac{d_n - d_{[nt]}}{c_{[nt]}} = \delta(t)$$

and

$$G^t(x) = G\left(\gamma(t)\, x + \delta(t)\right). \tag{3.13}$$

Similarly, for $s, t > 0$ we have

$$\begin{aligned} G^{ts}(x) &= G\left(\gamma(ts)\, x + \delta(ts)\right) \\ &= G^t\left(\gamma(s)\, x + \delta(s)\right) \\ &= G\left(\gamma(t)\left(\gamma(s)\, x + \delta(s)\right) + \delta(t)\right). \end{aligned}$$

From here we deduce that for $s, t > 0$ it holds

$$\gamma(ts) = \gamma(s)\,\gamma(t), \tag{3.14}$$

$$\delta(ts) = \gamma(t)\,\delta(s) + \delta(t). \tag{3.15}$$

Solving the functional equations (3.14)-(3.15) together with (3.13) leads to the three types of EVD.

□

Remark. It is seen from the convergence to types theorem that if convergence of normalized maxima takes place, the type of the limiting distribution (as specified by γ) is uniquely determined, while the location parameter μ and the scale parameter σ depend on the chosen normalizing sequences. It is always possible to choose these sequences such that the limit distribution has the standard d.f. $G_{\gamma,0,1} = G_\gamma$, given by (3.9).

Definition 3.1. *If the assumption of Fisher-Tippet theorem is fulfilled, we say that F belongs to the **maximum domain of attraction** of the d.f. G, $F \in \mathrm{MDA}(G)$.*

For checking the convergence of normalized maxima to certain type of EVD we can use the following theorem which expresses the convergence to EVD by means of the tail function $\bar{F}(x) = 1 - F(x)$.

Theorem 3.3 (Characterization of MDA(G)). *The d.f. F belongs to the maximum domain of attraction of EVD with d.f. G with normalizing constants $c_n > 0$, $d_n \in \mathbb{R}$, if and only if*

$$\lim_{n\to\infty} n\,\bar{F}\,(c_n\,x + d_n) = -\log G(x),\ \ x \in \mathbb{R}. \tag{3.16}$$

When $G(x) = 0$, the limit is interpreted as $+\infty$.

Proof. Consider first x such that $0 < G(x) \le 1$ and denote $\tau = -\log G(x)$. It is $0 \le \tau < +\infty$.

If (3.16) holds, then

$$\begin{aligned} \mathrm{P}\,(M_n \le c_n\,x + d_n) &= F^n\,(c_n\,x + d_n) = \big(1 - \bar{F}\,(c_n\,x + d_n)\big)^n \\ &= \left(1 - \frac{\tau}{n} + o\left(\frac{1}{n}\right)\right)^n,\ \ n \to +\infty, \end{aligned}$$

which implies

$$\lim_{n\to\infty} \mathrm{P}\,(M_n \le c_n\,x + d_n) = e^{-\tau} = G(x).$$

Conversely, if

$$\lim_{n\to\infty} \mathrm{P}\,(M_n \le c_n\,x + d_n) = \lim_{n\to\infty} F^n\,(c_n\,x + d_n) = G(x) \tag{3.17}$$

holds, then

$$\lim_{n\to+\infty} \bar{F}\,(c_n\,x + d_n) = 0.$$

(Otherwise, $\bar{F}\,(c_{n_k}\,x + d_{n_k})$ would be bounded away from 0 for some subsequence $\{n_k\}$. Then

$$\mathrm{P}\,(M_{n_k} \le c_{n_k}\,x + d_{n_k}) = \big(1 - \bar{F}\,(c_{n_k}\,x + d_{n_k})\big)^{n_k}$$

would imply

$$\lim_{k\to\infty} \mathrm{P}\,(M_{n_k} \le c_{n_k}\,x + d_{n_k}) = 0.)$$

Taking logarithms in (3.17) we have

$$\lim_{n\to\infty} \big[-n\,\log\big(1 - \bar{F}\,(c_n\,x + d_n)\big)\big] = -\log G(x). \tag{3.18}$$

Since $-\log(1 - x) \approx x$ for $x \to 0$, (3.18) implies

$$n\,\bar{F}\,(c_n\,x + d_n) = -\log G(x) + o(1),\ \ n \to \infty,$$

giving (3.16).

If $G(x) = 0$ and (3.16) holds, but

$$\lim_{n\to\infty} \mathrm{P}\left(M_n \le c_n\, x + d_n\right) \ne 0,$$

there must be a subsequence $\{n_k\}$ such that

$$\lim_{k\to\infty} \mathrm{P}\left(M_{n_k} \le c_{n_k}\, x + d_{n_k}\right) = \mathrm{e}^{-\tau'}$$

for some $\tau' < \infty$. But then

$$\lim_{k\to\infty} n_k\, \bar{F}\left(c_{n_k}\, x + d_{n_k}\right) = \tau' < \infty,$$

contradicting (3.16) with the limit $+\infty$.

Similarly,

$$\lim_{n\to\infty} \mathrm{P}\left(M_n \le c_n\, x + d_n\right) = 0$$

implies (3.16) with the limit equal to $+\infty$.

□

In financial and insurance applications we are especially interested in models describing the occurence of high losses. In the framework of modelling maxima, the most important model is Fréchet distribution and its maximum domain of attraction. The distributions belonging to the MDA of Fréchet distribution are characterized as distributions with heavy tails. We present a result that gives the description of the tail behavior of members of MDA $(G_{1,\alpha})$.

Theorem 3.4. *For $\alpha > 0$ and a distribution function F,*

$$F \in \mathrm{MDA}\left(G_{1,\alpha}\right)$$

if and only if

$$\bar{F}(x) = x^{-\alpha}\, L(x), \tag{3.19}$$

where L is a positive, Lebesgue-measurable function on $(0, +\infty)$ satisfying

$$\lim_{x\to+\infty} \frac{L(t\,x)}{L(x)} = 1,\;\; t > 0. \tag{3.20}$$

Remark. Function L satisfying (3.20) is referred to as **slowly varying** at ∞. The function on the right-hand side of (3.19) is called **regularly varying** at ∞ with index $-\alpha$. According to Theorem 3.4, the tails of members of MDA of Fréchet distribution are regularly varying functions with a negative index of variation.

3.3 Block Maxima Method

We present the application of EVD in the statistical analysis of extreme values. We assume that we have independent data from a distribution with d.f. $F \in \mathrm{MDA}(G_\gamma)$ for some GEVD d.f. G_γ. We divide the observed data into m blocks of size n. The maxima of the blocks are denoted by

$$M_{n1}, \ldots, M_{nm}. \tag{3.21}$$

By Fisher-Tippet theorem, for a sufficiently large size of one block n we can approximate the distribution of maxima (3.21) by a three-parametric d.f. $G_{\gamma,\mu,\sigma}$. The parameters are usually estimated by means of maximum likelihood method. The log-likelihood function is

$$\begin{aligned} & l(\gamma, \mu, \sigma; M_{n1}, \ldots, M_{nm}) = \sum_{i=1}^{m} \log g_{\gamma,\mu,\sigma}(M_{ni}) \\ & = -m \log \sigma \\ & \quad - (1 + 1/\gamma) \sum_{i=1}^{m} \log\left(1 + \gamma \frac{M_{ni} - \mu}{\sigma}\right) - \sum_{i=1}^{m} \left(1 + \gamma \frac{M_{ni} - \mu}{\sigma}\right)^{-1/\gamma}, \end{aligned} \tag{3.22}$$

where $g_{\gamma,\mu,\sigma}(x)$ is the density of (3.10). We maximize (3.22) under the conditions $\sigma > 0$, $1 + \gamma (M_{ni} - \mu)/\sigma > 0$ for all i. Note that the set of possible parameter values depends on the observations. We cannot use the usual regularity conditions to deduce properties of maximum likelihood estimates.

Remark. For a given amount of data we have to choose a compromise between a sufficiently large size of blocks n that is necessary for a satisfactory approximation of block maxima distribution by GEVD and on the other hand, a sufficiently large number of blocks that is necessary for a reasonable bias of maximum likelihood estimates.

The distribution fitted by maximization of (3.22) can be used to calculate characteristics traditionally applied to analysis of extremal events. Possible questions are

- What is the size of an extremal event that occurs with given average frequency?
- What is the average frequency of an extremal event of given size?

Let us denote by G the real d.f. of the maximum of a block of size n. We further introduce

$$r_{n,k} = G^{-1}(1 - 1/k), \tag{3.23}$$

the $(1-1/k)$-quantile of d.f. G. It holds

$$\mathrm{P}\left(M_n > r_{n,k}\right) = \frac{1}{k}.$$

The value of $r_{n,k}$ thus represents the level exceeded on average in one in k blocks. We call the characteristic $r_{n,k}$ **return level**. For n sufficiently large we replace d.f. G by the GEVD distribution with maximum likelihood estimates of parameters. For the estimate of the return level we obtain from (3.23)

$$\widehat{r}_{n,k} = \widehat{\mu} + \frac{\widehat{\sigma}}{\widehat{\gamma}}\left((-\log(1-1/k)^{-\widehat{\gamma}} - 1\right).$$

To solve the second question, we introduce for given u the characteristic

$$k_{n,u} = \frac{1}{\bar{G}(u)} = \frac{1}{1-G(u)}. \tag{3.24}$$

Obviously,

$$\mathrm{P}\left(M_n > u\right) = \frac{1}{k_{n,u}},$$

i.e. we expect that in $k_{n,u}$ blocks there is one block in which the level u is exceeded. The characteristic $k_{n,u}$ is called **return period**. We estimate its value using the fitted GEVD distribution by

$$\widehat{k}_{n,u} = \frac{1}{\bar{G}_{\widehat{\gamma},\widehat{\mu},\widehat{\sigma}}(u)}.$$

Remark. The analysis of extremal events based on the approximation of maxima by GEVD is applicable also in some cases of dependent observations. Suppose that the original data $\{X_i\}$ form a strictly stationary random sequence with the stationary distribution with d.f. F. Denote by $\left\{\tilde{X}_i\right\}$ a sequence od i.i.d. random variables with d.f. F. We further denote

$$M_n = \max\left(X_1, \ldots, X_n\right),$$
$$\tilde{M}_n = \max\left(\tilde{X}_1, \ldots, \tilde{X}_n\right).$$

For some processes $\{X_i\}$ it can be shown that there exists $\theta \in (0,1]$ such that

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{\tilde{M}_n - d_n}{c_n} \le x\right) = G(x) \tag{3.25}$$

holds for some non-degenerate d.f. G if and only if

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) = G^\theta(x). \tag{3.26}$$

The number θ is called **extremal index** of the process $\{X_i\}$.

The equivalence of (3.25) and (3.26) can be interpreted as follows: when $u = c_n\, x + d_n$, in the process with the extremal index θ we have for n sufficiently large

$$\mathrm{P}\left(M_n \leq u\right) \approx \mathrm{P}^\theta\left(\tilde{M}_n \leq u\right) = F^{n\theta}(u).$$

We can say that the distribution of the maximum of n observations is approximated by the distribution of the maximum of $n\,\theta$ independent observations with the same d.f. F.

3.4 Threshold Exceedances

The main distributional model for excesses over given high threshold is the **generalized Pareto distribution** (GPD).

First we introduce an α-parametrization corresponding to that of EVD. There are three standard versions of the d.f. for different choices of the parameter α.

The d.f. of **exponential distribution** is defined for $x \geq 0$ by

$$W_0(x) = 1 - \mathrm{e}^{-x}\,. \tag{3.27}$$

The d.f. of **Pareto distribution** with a parameter $\alpha > 0$ is given for $x \geq 1$ by

$$W_{1,\alpha}(x) = 1 - x^{-\alpha}. \tag{3.28}$$

The standard form of the d.f. of **beta distribution** with a parameter $\alpha < 0$ is given for $-1 \leq x \leq 0$ by

$$W_{2,\alpha}(x) = 1 - (-x)^{-\alpha}. \tag{3.29}$$

We can introduce a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$ as in (3.4)-(3.6).

Remark. By choosing $\mu = 0$, $\sigma = 1/\beta$ for $\beta > 0$, we obtain the usual form of the exponential distribution,

$$W_{0,0,1/\beta}(x) = 1 - \mathrm{e}^{-\beta\, x},\ x \geq 0.$$

Similarly, for $\mu = 0$, $\sigma = a$, we obtain Pareto distribution at the interval $(a, +\infty)$ with d.f.

$$W_{1,\alpha,0,a}(x) = 1 - \left(\frac{x}{a}\right)^{-\alpha}.$$

Beta distribution is usually introduced in statistical literature as the distribution defined on (0,1) by the density

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, x^{a-1}\,(1-x)^{b-1},\ a > 0,\ b > 0,\ 0 \le x \le 1. \tag{3.30}$$

If we set $\mu = 1$, $\sigma = 1$, in our parametric model, we obtain the d.f.

$$W_{2,\alpha,1,1}(x) = 1 - (1-x)^{-\alpha},\ 0 \le x \le 1,$$

with the density

$$w_{2,\alpha,1,1}(x) = -\alpha\,(1-x)^{-\alpha-1}.$$

This is a special case of (3.30) with $a = 1$, $b = -\alpha$.

Generalized Pareto distribution (GPD). Using the same reparametrization and choosing the same location and scale parameters as in (3.7) and (3.8), we obtain the standard form of GPD

$$W_\gamma(x) = 1 - (1 + \gamma\, x)^{-1/\gamma}, \tag{3.31}$$

from where Pareto distribution with the support $(0, +\infty)$ is obtained by choosing $\gamma > 0$, in the case $\gamma < 0$ we have beta distribution defined by (3.31) for $0 \le x \le 1/|\gamma|$. From (3.31) it is also seen the continuity in γ in the sense of the convergence to the d.f. of exponential distribution for $\gamma \to 0$.

The general form of the d.f. of GPD is derived by introducing a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$:

$$W_{\gamma,\mu,\sigma}(x) = 1 - \left(1 + \gamma\,\frac{x-\mu}{\sigma}\right)^{-1/\gamma}. \tag{3.32}$$

Note that μ is always the right endpoint of the support of (3.32).

We want to study the distribution of the values in the data that exceed a given threshold, or equivalently, the distribution of the excesses over the threshold. The following notation will be useful.

Let X be a r.v. with d.f. F. For the distribution of X conditional on the event $X > u$ for some $u > 0$ we introduce the notation

$$F^{[u]}(x) = \mathrm{P}\,(X \le x \mid X > u) = \frac{F(x) - F(u)}{1 - F(u)},\ x \ge u. \tag{3.33}$$

Similarly, for the distribution of the excess above u, given that $X > u$ we have

$$F^{(u)}(x) = \mathrm{P}\left(X - u \le x \,|\, x > u\right) = \frac{F(x+u) - F(u)}{1 - F(u)}, \; 0 \le x < x_F - u, \tag{3.34}$$

where x_F is the right endpoint of the support of d.f. F.

The importance of GPD in modelling excesses over high thresholds is explained by the following theorem which we state without a proof.

Theorem 3.5 (Pickands-Balkema-de Haan theorem)**.** *Let F be a d.f. with the right endpoint $x_F \le +\infty$ and let $F^{(u)}$ be the corresponding d.f. of excesses defined in* (3.34)*. We can find a positive measurable function $\sigma(u)$ such that*

$$\lim_{u \to x_F} \sup_{0 \le x < x_F - u} |\, F^{(u)}(x) - W_{\gamma,0,\sigma(u)}(x) \,| = 0, \tag{3.35}$$

if and only if $F \in \mathrm{MDA}\,(G_\gamma)$, $\gamma \in \mathbb{R}$.

Pickands-Balkema-de Haan theorem says that the set of distributions, for which normalized maxima converge to a d.f. of GEVD with a shape parameter γ, is the same as the set of distributions, for which the distribution of excesses converges with increasing value of the threshold to a GPD with the same value of the parameter γ. The theorem thus can serve also as a tool for characterization of the maximum domain of attraction for a given type of GEVD.

3.5 POT Method

We present an approach to analysis of data based on values exceeding given threshold. Such an approach is sometimes referred to as **POT method**, where POT stands for "peaks over threshold".

It follows from Pickands-Balkema-de Haan theorem that for the distribution of excesses over a high threshold in the data coming from a distribution with a d.f. $F \in \mathrm{MDA}\,(G_\gamma)$, the GPD distribution can be used as a suitable approximation.

Note that for the approximation of $F^{(u)}$, which is the d.f. with the left endpoint equal to zero, we use a two-parametric version of (3.32) with $\mu = 0$ and with the density

$$w_{\gamma,0,\sigma}(x) = \frac{1}{\sigma}\left(1 + \gamma\,\frac{x}{\sigma}\right)^{-1/\gamma - 1} \tag{3.36}$$

with the support $(0, +\infty)$ for $\gamma > 0$, and $(0, -1/\gamma)$ for $\gamma < 0$. Given the observed data

$$X_1, \ldots, X_n, \tag{3.37}$$

independent realizations of a r.v. with d.f. F, for a given high threshold u we denote by N_u the number of values in the data exceeding u. The exceedances over u in (3.37) are denoted by

$$\tilde{X}_1, \ldots, \tilde{X}_{N_u}. \tag{3.38}$$

From (3.38) we derive for each $j = 1, \ldots, N_u$ the excesses $Y_j = \tilde{X}_j - u$.

The maximum likelihood estimation is based on the assumption, that the distribution of excesses Y_j can be satisfactorily approximated by a GPD d.f. $W_{\gamma,0,\sigma}$. Parameters γ and σ are estimated by maximizing the log-likelihood

$$l\left(\gamma, \sigma; Y_1, \ldots, Y_{N_u}\right) = -N_u \log \sigma - (1 + 1/\gamma) \sum_{j=1}^{N_u} \log\left(1 + \gamma \frac{Y_j}{\sigma}\right). \tag{3.39}$$

(3.39) is maximized under conditions $\sigma > 0$, $1 + \gamma \frac{Y_j}{\sigma} > 0$ for all $j = 1, \ldots, N_u$. Similarly as in the case of fitting a GEVD to normalized maxima, we again have here the situation non-compliant with the regularity conditions.

A useful property of GPD is **POT-stability**. For the d.f. (3.32) we derive, using (3.33),

$$\begin{aligned} W^{[u]}_{\gamma,\mu,\sigma}(x) &= \frac{\left(1 + \gamma \frac{u-\mu}{\sigma}\right)^{-1/\gamma} - \left(1 + \gamma \frac{x-\mu}{\sigma}\right)^{-1/\gamma}}{\left(1 + \gamma \frac{u-\mu}{\sigma}\right)^{-1/\gamma}} \\ &= 1 - \left(1 + \frac{\gamma \frac{x-\mu}{\sigma}}{1 + \gamma \frac{u-\mu}{\sigma}}\right)^{-1/\gamma} \\ &= 1 - \left(1 + \gamma \frac{x - u}{\sigma + \gamma\,(u - \mu)}\right)^{-1/\gamma}. \end{aligned} \tag{3.40}$$

From (3.40) we see that

$$W^{[u]}_{\gamma,\mu,\sigma}(x) = W_{\gamma,u,\sigma+\gamma(u-\mu)}(x), \; x \geq u. \tag{3.41}$$

When we consider GPD with the location parameter equal to zero, (3.41) yields for the distribution of excesses

$$W^{(u)}_{\gamma,0,\sigma}(x) = W_{\gamma,0,\sigma+\gamma u}(x), \; x \geq 0. \tag{3.42}$$

If the approximation of the excesses over given threshold by GPD with d.f. $W_{\gamma,0,\sigma}$ is plausible, then the excesses over any higher threshold can be modelled again by GPD, with the same shape parameter γ.

The expected value of the excess distribution (3.34), denoted by

$$e(u) = \mathrm{E}\left[X - u \,|\, X > u\right], \tag{3.43}$$

is called **mean excess function**.

Assume that X in (3.43) has the density (3.36). The expected value $\mathrm{E}\,X$ exists for $\gamma < 1$. It holds

$$\mathrm{E}\,X = \int_0^{\infty} \frac{x}{\sigma}\left(1+\gamma\,\frac{x}{\sigma}\right)^{-1/\gamma-1} \mathrm{d}\,x = \sigma \int_0^{\infty} y\,(1+\gamma\,y)^{-1/\gamma-1}\,\mathrm{d}\,y \tag{3.44}$$

for $\gamma < 0$,

$$\mathrm{E}\,X = \int_0^{-1/\gamma} \frac{x}{\sigma}\left(1+\gamma\,\frac{x}{\sigma}\right)^{-1/\gamma-1} \mathrm{d}\,x = \sigma \int_0^{-1/\gamma} y\,(1+\gamma\,y)^{-1/\gamma-1}\,\mathrm{d}\,y \tag{3.45}$$

for $\gamma < 0$.

Using integration by parts, the same result is obtained from (3.44) and (3.45), i.e.

$$\mathrm{E}\,X = \frac{\sigma}{1-\gamma}. \tag{3.46}$$

(3.46) is applicable also in the case of exponential distribution with the d.f. $W_{0,0,\sigma}(x) = 1 - \mathrm{e}^{-\frac{x}{\sigma}}$. Thus (3.46) is the expression of the expected value of GPD for $\gamma < 1$.

Combining this result with (3.42) we see an important property of the mean excess function of GPD:

$$e(u) = \frac{\sigma + \gamma\,u}{1-\gamma}. \tag{3.47}$$

In case that the excesses over given threshold are modelled by GPD with d.f. $W_{\gamma,0,\sigma}$, the expected value of excesses over higher thresholds, taken as the function of the threshold u, is linear.

This feature is often used in practice to decide whether the chosen threshold is high enough for the GPD approximation based on Picands-Balkema-de Haan theorem.

For given u, mean excess function can be estimated from data (3.37) by computing

$$e_n(u) = \frac{1}{N_u}\sum_{i=1}^{N_u}\left(\tilde{X}_i - u\right).$$

It is recommended to create a graph consisting of points

$$\{(X_{in}, e_n(X_{in}))\,,\ i = 2, \ldots, n\}\,, \tag{3.48}$$

where $X_{nn} \le \cdots \le X_{1n}$ are ordered observations (3.37). Such a graph is called **mean excess plot**. We can consider the right choice of the threshold u as a value from a region where the shape of the graph shows approximately linear trend. On the other hand, the threshold should be chosen low enough, so that there is sufficient number of exceedances in the data, as needed for maximum likelihood estimation.

From (3.47) it is seen that the mean excess function is increasing for $\gamma > 0$, decreasing for $\gamma < 0$ and constant for $\gamma = 0$. The slope of the linear trend in the graph (3.48) thus indicates whether the appropriate model for excesses over high thresholds is Pareto, beta or exponential distribution.

We now explain how the estimated GPD distribution of excesses can be applied to modelling the right tail of the distribution of our data. We are interested in the estimation of the probability $\bar{F}(x) = \mathrm{P}(X > x)$ for such a high value of x that is exceeded very rarely or never in the observed data.

Assume that u is a threshold such that we have in our data sufficient number of exceedances over u, but high enough so that the excesses over u can be approximated by GPD. We estimate parameters γ and σ by maximum likelihood procedure using (3.39). We write for $x \ge u$

$$\bar{F}(x) = \bar{F}(u)\,\bar{F}^{(u)}(x-u) \doteq \bar{F}(u)\left(1 + \gamma\,\frac{x-u}{\sigma}\right)^{-1/\gamma}. \tag{3.49}$$

It is estimated from the data (3.37) by

$$\widehat{\bar{F}(x)} = \frac{N_u}{n}\left(1 + \widehat{\gamma}\,\frac{x-u}{\widehat{\sigma}}\right)^{-1/\widehat{\gamma}}, \tag{3.50}$$

where $\widehat{\gamma}$ and $\widehat{\sigma}$ are maximum likelihood estimates. Note that the probability $\bar{F}(u)$ is estimated by the ratio $\frac{N_u}{n}$. This is possible thanks to the fact that we have enough data exceeding the chosen threshold u.

By inverting (3.49) we get the approximation for the quantile function corresponding to the d.f. F

$$F^{-1}(\alpha) = u + \frac{\sigma}{\gamma}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\gamma} - 1\right), \tag{3.51}$$

valid for $\alpha \ge F(u)$. This enables us to estimate α-quantile of d.f. F using data (3.37) by

$$\widehat{F^{-1}(\alpha)} = u + \frac{\widehat{\sigma}}{\widehat{\gamma}}\left(\left(\frac{1-\alpha}{N_u/n}\right)^{-\widehat{\gamma}} - 1\right).$$

3.6 The Hill Method

We describe an approach suitable for estimation of the tails of distributions belonging to the maximum domain of attraction of Fréchet distribution. So we consider d.f. F satisfying (3.19) with some $\alpha > 0$. Since (3.19) describes the tail behavior of F, the parameter α is referred to as **tail index**.

The Hill method deals with estimation of the tail index α based on identically distributed data $X_1, \ldots, X_n$ with d.f. F. We present a sketch of the derivation of so called **Hill estimator** $\widehat{\alpha}^{(H)}$.

We denote by

$$e^*(v) = \mathrm{E}\left(\log X - v \mid \log X > v\right),$$

the mean excess function of $\log X$. Then for $u > 0$

$$\begin{aligned} e^*(\log u) &= \mathrm{E}\left(\log X - \log u \mid \log X > \log u\right) \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty (\log x - \log u) \,\mathrm{d}\, F(x). \end{aligned} \tag{3.52}$$

Integration by parts gives

$$\begin{aligned} &\int_u^M (\log x - \log u)\ \mathrm{d}\, F(x) = \int_u^M \log\left(\frac{x}{u}\right) \mathrm{d}\, F(x) \\ &= \log\left(\frac{M}{u}\right)\ F(M) - \int_u^M \frac{1}{x}\, F(x) \,\mathrm{d}\, x \\ &= \log\left(\frac{M}{u}\right)\ (F(M) - 1) + \int_u^M \frac{1}{x}\ (1 - F(x)) \,\mathrm{d}\, x. \end{aligned} \tag{3.53}$$

From

$$\log\left(\frac{M}{u}\right) \int_M^\infty \mathrm{d}\, F(x) \le \int_M^\infty \log\left(\frac{x}{u}\right) \mathrm{d}\, F(x)$$

it is seen that the first term on the right-hand side of (3.53) vanishes as $M \to +\infty$, provided that $\mathrm{E} \log X < +\infty$. In such a case we have from (3.52)

$$e^*(\log u) = \frac{1}{\bar{F}(u)} \int_u^\infty \frac{\bar{F}(x)}{x} \,\mathrm{d}\, x. \tag{3.54}$$

We rewrite (3.54) using (3.19),

$$e^*(\log u) = \frac{1}{\bar{F}(u)} \int_u^\infty L(x)\, x^{-(\alpha+1)} \,\mathrm{d}\, x, \tag{3.55}$$

where $L(x)$ is the corresponding slowly varying function.

We further use the fact that for u sufficiently large, the slowly varying function $L(x)$ for $x \ge u$ can be approximated by a constant which is taken out of the integral in (3.55). Formally, this is based on the following theorem:

Theorem 3.6 (Karamata's theorem)**.** *Let L be a slowly varying function which is locally bounded in $[x_0, +\infty)$ for some $x_0 \geq 0$. Then for $\kappa < -1$*

$$\int_u^\infty x^\kappa\, L(x)\, \mathrm{d}\, x \approx \frac{1}{\kappa + 1}\, u^{\kappa+1}\, L(u),\ \ u \to +\infty. \tag{3.56}$$

Applying Karamata's theorem to (3.55) we obtain

$$e^*(\log u) \approx \frac{L(u)\, u^{-\alpha}\, \alpha^{-1}}{\bar{F}(u)} = \alpha^{-1},\ \ u \to +\infty. \tag{3.57}$$

Hill estimator of the tail index α is the inverse of the estimate $e_n^*(\log X_{kn})$, where X_{kn} is the k-th largest value in (3.37),

$$\widehat{\alpha}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^{k} \log X_{jn} - \log X_{kn} \right)^{-1}. \tag{3.58}$$

Obviously, the estimator (3.58) depends on the choice of k. It is recommended to create **Hill plot** consisting of points

$$\left\{ \left(k, \widehat{\alpha}_{k,n}^{(H)} \right),\ k = 2, \ldots, n \right\},$$

from which we can find the area where the values of Hill estimator computed for different values of k are close to each other.

At the end we show an application of Hill estimator to modelling the tail of a distribution belonging to the maximum domain of attraction of Fréchet distribution. We write

$$\bar{F}(x) = C\, x^{-\alpha},\ \ x \geq u > 0, \tag{3.59}$$

for a high threshold u (we substitute the slowly varying function by a constant C). Then we can write

$$C = u^\alpha\, \bar{F}(u),$$

where we substitute the index α by the Hill estimator (3.58) and for the value u we insert the statistic X_{kn}. For this choice of u we have $\frac{k}{n}$ as the estimate of $\bar{F}(u)$. Using (3.59) we come to the estimate of the tail function at point $x > X_{kn}$ in the form

$$\widehat{\bar{F}(x)} = \frac{k}{n} \left(\frac{x}{X_{kn}} \right)^{-\widehat{\alpha}_{k,n}^{(H)}}.$$

Chapter 4

Modelling Dependence

4.1 Copulas

In this section we deal with multivariate distributions and modelling the dependence among components of a random vector using copulas.

Definition 4.1. *A d-**dimensional copula** is a distribution function of a d-dimensional random vector for which all univariate distributions are uniform on (0,1).*

Equivalently, copula C is a mapping of the form $C : [0,1]^d \to [0,1]$, satisfying the following conditions:

1. $C(u_1, \ldots, u_d)$ is increasing in each component u_i.
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i = 1, \ldots, d$, $u_i \in [0,1]$.
3. For all $\left(u_1^{(1)}, \ldots, u_d^{(1)}\right)$, $\left(u_1^{(2)}, \ldots, u_d^{(2)}\right)$ in $[0,1]^d$ such that $u_i^{(1)} \le u_i^{(2)}$ for all $i = 1, \ldots, d$, it holds
$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C\left(u_1^{(i_1)}, \ldots, u_d^{(i_d)}\right) \ge 0.$$

The first condition is fulfilled by any multivariate distribution function. The second condition is the requirement of uniform marginal distributions. The third property ensures that if the random vector $(U_1, \ldots, U_d)'$ has the d.f. C, then
$$\mathrm{P}\left(u_1^{(1)} \le U_1 \le u_1^{(2)}, \ldots, u_d^{(1)} \le U_2 \le u_d^{(2)}\right) \ge 0.$$

For example, for $d = 2$ we have

$$\begin{aligned}
&\mathrm{P}\left(u_1^{(1)} \le U_1 \le u_1^{(2)},\ u_2^{(1)} \le U_2 \le u_2^{(2)}\right) \\
&= \mathrm{P}\left(U_1 \le u_1^{(2)},\ U_2 \le u_2^{(2)}\right) - \mathrm{P}\left(U_1 \le u_1^{(1)},\ U_2 \le u_2^{(2)}\right) \\
&- \mathrm{P}\left(U_1 \le u_1^{(2)},\ U_2 \le u_2^{(1)}\right) + \mathrm{P}\left(U_1 \le u_1^{(1)},\ U_2 \le u_2^{(1)}\right).
\end{aligned}$$

The importance of copulas is in "connecting" multivariate distribution functions with their univariate marginals.

The essential result of the theory of copulas is the following theorem. It shows that all multivariate distributions contain copulas and also that a copula may be used to construct a multivariate distribution with given univariate marginals.

Theorem 4.1 (Sklar's theorem). *Let F be a joint d.f. with marginal distribution functions $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, \ldots, x_d \in [-\infty, +\infty]$,*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{4.1}$$

If the marginal distributions are continuous, then C is unique. Otherwise, C is uniquely determined on $RanF_1 \times \cdots \times RanF_d$, where $RanF_i$ denotes the range of F_i.

Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F defined in (4.1) *is a joint d.f. with margins $F_1, \ldots, F_d$.*

Proof. (Part.)

We prove the existence and uniqueness of a copula in the case when $F_1, \ldots, F_d$ are continuous.

If F is the d.f. of a random variable X, then

$$\mathrm{P}(F(X) \le F(x)) = \mathrm{P}(X \le x). \tag{4.2}$$

The equality (4.2) follows from the fact that F is nondecreasing function, so

$$\mathrm{P}(F(X) \le F(x)) = \mathrm{P}(X \le x) + \mathrm{P}(F(X) = F(x),\ X > x),$$

where the second probability on the right-hand side is zero, since it corresponds to the event, where X attains a value from the interval on which d.f. $F(x)$ is constant.

Using (4.2) we can write

$$F(x_1,\dots,x_d) = \mathrm{P}(X_1 \le x_1,\dots,X_d \le x_d)$$
$$= \mathrm{P}(F_1(X_1) \le F_1(x_1),\dots,F_d(X_d) \le F_d(x_d)).$$

We thus obtain the equality (4.1), where C is the joint d.f. of the random vector $(F_1(X_1),\dots,F_d(X_d))'$.

To see that C is a copula, we have to show that for $i = 1,\dots,d$ the r.v. $F_i(X_i)$ has uniform distribution on $(0,1)$.

For a d.f. F we define its quantile function by

$$F^{-1}(y) = \inf\{x,\ F(x) \ge y\}. \tag{4.3}$$

If F is continuous, then F^{-1} is increasing and also

$$F\left(F^{-1}(y)\right) = y,\ y \in (0,1). \tag{4.4}$$

We use this together with the equality

$$\mathrm{P}\left(F^{-1}(F(X)) = X\right) = 1,$$

holding true for any d.f. F, to derive for $y \in (0,1)$

$$\mathrm{P}(F(X) \le y) = \mathrm{P}\left(F^{-1}(F(X)) \le F^{-1}(y)\right) = \mathrm{P}\left(X \le F^{-1}(y)\right)$$
$$= F\left(F^{-1}(y)\right) = y. \tag{4.5}$$

From (4.5) it is seen that for a continuous d.f. F, random variable $F(X)$ has uniform distribution.

For $(u_1,\dots,u_d) \in [0,1]^d$ we insert

$$x_i = F_i^{-1}(u_i),\ i = 1,\dots,d,$$

into (4.1) and we obtain, using (4.4),

$$C(u_1,\dots,u_d) = F\left(F_1^{-1}(u_1),\dots,F_d^{-1}(u_d)\right). \tag{4.6}$$

In case of continuous margins $F_1,\dots,F_d$, copula C is by (4.6) determined uniquely.

For te converse implication we assume that C is a copula and $F_1,\dots,F_d$ are univariate d. functions. Let $U = (U_1,\dots,U_d)'$ be a random vector with the joint d.f. C. We set

$$X_i = F_i^{-1}(U_i),\ i = 1,\dots,d. \tag{4.7}$$

For a uniformly distributed r.v. U, a d.f. F and $x \in \mathbb{R}$ it holds

$$\mathrm{P}\left(F^{-1}(U) \leq x\right) = \mathrm{P}\left(U \leq F(x)\right) = F(x).$$

Here we use the equivalence

$$F(x) \geq y \;\Leftrightarrow\; F^{-1}(y) \leq x$$

which is valid for any right-continuous d.f. F.

Random variable X_i defined by (4.7) thus has the d.f. F. For the joint d.f. of the vector $(X_1, \ldots, X_d)'$ we derive

$$\begin{aligned}&\mathrm{P}\left(X_1 \leq x_1, \ldots, X_d \leq x_d\right) = \mathrm{P}\left(F_1^{-1}\left(U_1\right) \leq x_1, \ldots, F_d^{-1}\left(U_d\right) \leq x_d\right) \\ &= \mathrm{P}\left(U_1 \leq F_1\left(x_1\right), \ldots, U_d \leq F_d\left(x_d\right)\right) = C\left(F_1\left(x_1\right), \ldots, F_d\left(x_d\right)\right).\end{aligned}$$

□

Definition 4.2. *If the random vector* $\mathbf{X} = (X_1, \ldots, X_d)'$ *has joint d.f.* F *with continuous margins* $F_1, \ldots, F_d$*, then the* ***copula of*** F *(or* $\mathbf{X}$*) is the d.f.* C *of* $\left(F_1\left(X_1\right), \ldots, F_d\left(X_d\right)\right)$.

A useful property of a copula is its invariance under increasing transformations of the margins.

Theorem 4.2. *Let* $(X_1, \ldots, X_d)'$ *be a random vector with continuous margins and with the copula* C*. Let* $T_1, \ldots, T_d$ *be increasing functions. Then the random vector* $\left(T_1\left(X_1\right), \ldots, T_d\left(X_d\right)\right)'$ *also has the copula* C*.*

Proof. We denote by $\tilde{F}_i$ the d.f. of the transformed variable $T_i\left(X_i\right)$. We write

$$\tilde{F}_i(y) = \mathrm{P}\left(T_i\left(X_i\right) \leq y\right) = \mathrm{P}\left(X_i \leq T_i^{-1}(y)\right) = F_i\left(T_i^{-1}(y)\right). \tag{4.8}$$

In (4.8) we use the generalized inverse defined as in (4.3). The second equality in (4.8) holds thanks to the continuity of F, from which

$$\mathrm{P}\left(X_i = T_i^{-1}(y),\ T\left(X_i\right) > y\right) = 0.$$

From (4.8) it folows

$$\tilde{F}_i\left(T_i(x)\right) = F_i\left(T_i^{-1}\left(T_i(x)\right)\right) = F_i(x).$$

C is the copula of $(X_1, \ldots, X_d)'$, it is the joint d.f. of $\left(F_1\left(X_1\right), \ldots, F_d\left(X_d\right)\right)'$:

$$\begin{aligned}C\left(u_1, \ldots, u_d\right) &= \mathrm{P}\left(F_1\left(X_1\right) \leq u_1, \ldots, F_d\left(X_d\right) \leq u_d\right) \\ &= \mathrm{P}\left(\tilde{F}_1\left(T_1\left(X_1\right)\right) \leq u_1, \ldots, \tilde{F}_d\left(T_d\left(X_d\right)\right) \leq u_d\right).\end{aligned}$$

Thus C is also the copula of $\left(T_1\left(X_1\right), \ldots, T_d\left(X_d\right)\right)'$.

□

The following theorem concerns so called Fréchet bounds for copulas.

Theorem 4.3. *For every copula $C(u_1, \ldots, u_d)$ we have the bounds*

$$\left(\sum_{i=1}^{d} u_i + 1 - d\right)_+ \le C(u_1, \ldots, u_d) \le \min(u_1, \ldots, u_d). \tag{4.9}$$

Remark. The bounds in (4.9) are called **Fréchet upper bound** and **Fréchet lower bound**.

Proof. The second inequality in (4.9) follows from

$$\mathrm{P}(U_1 \le u_1, \ldots, U_d \le u_d) \le \mathrm{P}(U_i \le u_i),\ i = 1, \ldots, d,$$

where $(U_1, \ldots, U_d)'$ has d-dimensional uniform distribution with d.f. C.

For the first inequality we observe that

$$\begin{aligned}
&\mathrm{P}(U_1 \le u_1, \ldots, U_d \le u_d) = 1 - \mathrm{P}(\exists i:\ U_i > u_i) \\
&\ge 1 - \sum_{i=1}^{d} \mathrm{P}(U_i > u_i) = 1 - \sum_{i=1}^{d} (1 - u_i).
\end{aligned}$$

□

Remark. Using the same reasoning as in the proof of Theorem 4.3 we can derive bounds for any joint d.f. F in the form

$$\left(\sum_{i-1}^{d} F_i(x_i) + 1 - d\right)_+ \le F(x_1, \ldots, x_d) \le \min(F_1(x_1), \ldots, F_d(x_d)).$$

4.1.1 Fundamental Copulas

We give examples of copulas representing important special cases of dependence structure.

The **independence copula** is

$$\Pi(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i,\ u_i \in [0,1],\ i = 1, \ldots, d. \tag{4.10}$$

It is clear from Sklar's theorem that random variables $X_1, \ldots, X_d$ with continuous distribution functions are independent if and only if their joint d.f. is

$$F(x_1, \ldots, x_d) = \Pi(F_1(x_1), \ldots, F_d(x_d)).$$

The **comonotonicity copula** is the Fréchet upper bound

$$M(u_1, \dots, u_d) = \min(u_1, \dots, u_d), \ u_i \in [0,1], \ i = 1, \dots, d. \tag{4.11}$$

It is the joint d.f. of the random vector $(U, \dots, U)$, where U has uniform distribution on (0,1).

Definition 4.3. *Random variables $X_1, \dots, X_d$ are **comonotonic** if their joint d.f. has copula $M(u_1, \dots, u_d)$ given in* (4.11)*.*

An equivalent definition of comonotonicity is given in the following theorem.

Theorem 4.4. *Random variables $X_1, \dots, X_d$ are comonotonic if and only if the vector $(X_1, \dots, X_d)'$ has the same distribution as $(\nu_1(Z), \dots, \nu_d(Z))'$ for some r.v. Z and non-decreasing functions $\nu_1, \dots, \nu_d$.*

Proof. Let $X_1, \dots, X_d$ be comonotonic random variables with joint d.f. F and margins $F_1, \dots, F_d$. Let U be a r.v. uniformly distributed on (0,1). We have

$$\begin{aligned} F(x_1, \dots, x_d) &= \min(F_1(x_1), \dots, F_d(x_d)) \\ &= \mathrm{P}(U \le \min(F_1(x_1), \dots, F_d(x_d))) \\ &= \mathrm{P}(U \le F_1(x_1), \dots, U \le F_d(x_d)) \\ &= \mathrm{P}\left(F_1^{-1}(U) \le x_1, \dots, F_d^{-1}(U) \le x_d\right). \end{aligned}$$

We see that

$$(X_1, \dots, X_d) \stackrel{d}{=} \left(F_1^{-1}(U), \dots, F_d^{-1}(U)\right).$$

Conversely, assume

$$(X_1, \dots, X_d) \stackrel{d}{=} (\nu_1(Z), \dots, \nu_d(Z))$$

for a r.v. Z and non-decreasing functions $\nu_1, \dots, \nu_d$.

We have

$$\begin{aligned} F(x_1, \dots, x_d) &= \mathrm{P}(\nu_1(Z) \le x_1, \dots, \nu_d(Z) \le x_d) \\ &= \mathrm{P}(Z \in A_1, \dots, Z \in A_d), \end{aligned}$$

where each A_i is an interval of the form $(-\infty, k_i]$ or $(-\infty, k_i)$, so one interval A_i is a subset of all other intervals. Then

$$\begin{aligned} F(x_1, \dots, x_d) &= \min(\mathrm{P}(Z \in A_1), \dots, \mathrm{P}(Z \in A_d)) \\ &= \min(F_1(x_1), \dots, F_d(x_d)). \end{aligned}$$

□

Corollary. Random variables X_1, X_2 with continuous distribution functions are comonotonic if and only if $X_2 = T(X_1)$ almost surely for some non-decreasing transformation T.

Proof. From the proof of Theorem 4.4 it is seen that for $U = F_1(X_1)$ we obtain

$$(X_1, X_2) \overset{d}{=} \left(X_1, F_2^{-1}(F_1(X_1))\right).$$

From

$$(X_1, X_2) \overset{d}{=} (X_1, T(X_1))$$

we obtain $\mathrm{P}(X_2 = T(X_1)) = 1$.

□

The dependence structure contained in the comonotonicity copula is referred to as **perfect positive dependence**.

The **countermonotonicity copula** is the Fréchet lower bound in case $d = 2$:

$$W(u_1, u_2) = (u_1 + u_2 - 1)_+, \ u_i \in [0, 1], \ i = 1, 2. \tag{4.12}$$

It is the joint d.f. of random vector $(U, 1 - U)$, where U is uniformly distributed on (0,1).

Definition 4.4. *Random variables X_1 and X_2 are* ***countermonotonic*** *if their joint distribution has copula $W(u_1, u_2)$ given in* (4.12)*.*

An equivalent definition of countermonotonicity is given in the following theorem.

Theorem 4.5. *Random variables X_1 and X_2 are comonotonic if and only if the vector (X_1, X_2) has the same distribution as $(\nu_1(Z), \nu_2(Z))$ for some r.v. Z and some functions ν_1, ν_2 from which one is non-decreasing and the other is non-increasing.*

Corollary. Random variables X_1 and X_2 with continuous distribution functions are countermonotonic if and only if $X_2 = T(X_1)$ almost surely for some non-increasing function T.

The concept of countermonotonicity cannot be extended to higher dimensions. It is also referred to as **perfect negative dependence**.

4.1.2 Implicit Copulas

Implicit copulas are derived using Sklar's theorem as copulas of specific multivariate distributions with continuous marginal distribution functions.

An example of an implicit copula is **Gauss copula** defined as the copula of multivariate normal distribution. Consider vector $\mathbf{X} = (X_1, \ldots, X_d)'$ with the joint density

$$f(x_1, \ldots, x_d) = f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}, \tag{4.13}$$

for $\mathbf{x} \in \mathbb{R}^d$, $\boldsymbol{\mu} \in \mathbb{R}^d$ and a $d \times d$ positive definite matrix Σ. From Theorem 4.2 it follows that the same copula as vector $\mathbf{X}$ has vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$, where Y_i is an increasing transformation of X_i given by

$$Y_i = \frac{X_i - \mu_i}{\sqrt{\operatorname{Var} Y_i}}.$$

Vector $\mathbf{Y}$ has d-dimensional normal distribution $N_d(0, R)$, where R is the correlation matrix of $\mathbf{X}$, its marginal distributions are standard normal. Gauss copula is given by

$$\begin{aligned} C_R^{Ga}(u_1, \ldots, u_d) &= \mathrm{P}\left(\Phi(Y_1) \le u_1, \ldots, \Phi(Y_d) \le u_d\right) \\ &= \mathrm{P}\left(Y_1 \le \Phi^{-1}(u_1), \ldots, Y_d \le \Phi^{-1}(u_d)\right) \\ &= \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right), \end{aligned} \tag{4.14}$$

where Φ is d.f. of $N(0,1)$ and Φ_R is d.f. of $N_d(0, R)$.

In the special case of $d = 2$, Gauss copula is a copula of bivariate distribution determined by one parameter ρ, corresponding to the correlation coefficient of the bivariate normal vector from which the copula is derived. In case $|\rho| \neq 1$, (4.14) can be expresed by means of the integral from the density (4.13). For $\rho = 0$ the resulting copula is the independence copula. For $\rho = 1$ we obtain from (4.14) the comonotonicity copula and for $\rho = -1$ we obtain the countermonotonicity copula.

t-**copula** is a copula based on multivariate t-distribution with the joint density

$$f(x_1, \ldots, x_d) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\,(\nu\,\pi)^{d/2}\,|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]. \tag{4.15}$$

Vector $\mathbf{X}$ with the density (4.15) has the expected value $\mathrm{E}\,\mathbf{X} = \boldsymbol{\mu}$, provided that $\nu > 1$, and the covariance matrix $\operatorname{Cov}(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma$ in case $\nu > 2$.

Analogously as in the case of Gauss copula, the t-copula is written in the form

$$C^t_{\nu,R}\left(u_1,\ldots,u_d\right)=T_{\nu,R}\left(T^{-1}_\nu\left(u_1\right),\ldots,T^{-1}_\nu\left(u_d\right)\right). \tag{4.16}$$

T_ν in (4.16) stands for the d.f. of standard univariate t-distribution with ν degrees of freedom and $T_{\nu,R}$ is the joint d.f. of d-dimensional t-distribution with zero expected value and with the correlation matrix R.

4.1.3 Explicit Copulas

There are many copulas that can be written in a simple closed form as functions on $[0,1]^d$.

An example of bivariate explicit copula is **Gumbel copula** given by

$$C^{Gu}_\theta\left(u_1,u_2\right)=\exp\left\{-\left(\left(-\log u_1\right)^\theta+\left(-\log u_2\right)^\theta\right)^{1/\theta}\right\},\ 1\leq\theta<+\infty. \tag{4.17}$$

For $\theta=1$ copula (4.17) is equal to the independence copula (4.10). For $\theta\to+\infty$ we obtain bivariate comonotonicity copula (4.11) as a limit.

Clayton copula is bivariate copula of the form

$$C^{Cl}_\theta\left(u_1,u_2\right)=\left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1/\theta},\ 0<\theta<+\infty. \tag{4.18}$$

Similarly as for Gumbel copula, we have the independence copula and the comonotonicity copula as limit cases of (4.18) for $\theta\to 0+$ or $\theta\to+\infty$ respectively.

4.1.4 Archimedean Copulas

The Gumbel copula (4.17) and the Clayton copula (4.18) can be both represented in the general form

$$C\left(u_1,u_2\right)=\phi^{-1}\left(\phi\left(u_1\right)+\phi\left(u_2\right)\right), \tag{4.19}$$

with

$$\phi(x)=\left(-\log x\right)^\theta \tag{4.20}$$

in the case of the Gumbel copula and

$$\phi(x)=\frac{1}{\theta}\left(x^{-\theta}-1\right) \tag{4.21}$$

for the Clayton copula. Observe that in both cases, for the range of values of theta given in (4.17) or (4.18), $\phi(x)$ given by (4.20) or (4.21) is a convex decreasing function from $(0,1]$ to $[0,+\infty)$, satisfying $\phi(0+)=+\infty$ and $\phi(1)=0$.

We want to use a construction similar to (4.19), where we accept also functions $\phi(x)$ such that $\phi(0) < +\infty$. In such a case it is necessary to use a generalized form of inverse function.

Definition 4.5. *Suppose $\phi : [0,1] \to [0,+\infty]$ is continuous and decreasing with $\phi(1) = 0$ and $\phi(0) \leq +\infty$.* ***Pseudo-inverse*** *of ϕ is a function on $[0,+\infty]$ defined by*

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & 0 \leq t \leq \phi(0), \\ 0, & \phi(0) < t \leq +\infty. \end{cases} \tag{4.22}$$

Definition 4.6. ***Bivariate Archimedean copula*** *is defined by*

$$C\left(u_1, u_2\right) = \phi^{[-1]}\left(\phi\left(u_1\right) + \phi\left(u_2\right)\right), \tag{4.23}$$

where $\phi : [0,1] \to [0,+\infty]$ is a continuous, decreasing, convex function satisfying $\phi(1) = 0$.

Remark. It can be shown that the convexity of ϕ in (4.23) is a necessary and sufficient condition for C being a copula.

The function ϕ with the properties given in the definition 4.6 is called **Archimedean copula generator**. It is called **strict generator** if $\phi(0) = +\infty$.

Example. We generalize the definition of Clayton copula (4.18) by extending the set of possible values for the parameter θ to $-1 \leq \theta < +\infty$. For $0 < \theta < +\infty$ the generator (4.21) is strict. For $-1 \leq \theta < 0$, ϕ is decreasing from $\frac{1}{|\theta|}$ to 0 on [0,1]. In this case we use the pseudo-inverse of (4.21) in the form

$$\phi^{[-1]}(x) = \begin{cases} (\theta\, x + 1)^{-1/\theta}, & \text{for } 0 \leq x \leq \frac{1}{|\theta|} \\ 0, & \text{for } x > \frac{1}{|\theta|}. \end{cases} \tag{4.24}$$

Inserting (4.24) into (4.23) leads to a generalized form of Clayton copula

$$C_\theta^{Cl}\left(u_1, u_2\right) = \left[\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)_+\right]^{-1/\theta}, \ -1 \leq \theta < +\infty. \tag{4.25}$$

For $\theta = -1$ (4.25) becomes the countermonotonicity copula (4.12).

4.2 Dependence Measures

We present three approaches to measuring dependence between random variables. We briefly recall Pearson's correlation coefficient and its properties. Further, we concentrate on measures of rank correlation and on coefficients of tail dependence. In case of continuously distributed random variables, the last two concepts are determined solely by the unique copula of the corresponding bivariate distribution.

4.2.1 Linear Correlation

Pearson's correlation coefficient for random variables X_1 and X_2, where X_1, X_2 have finite positive variances, is defined by

$$\rho\,(X_1, X_2) = \frac{\mathrm{Cov}\,(X_1, X_2)}{\sqrt{\mathrm{Var}\,X_1}\,\sqrt{\mathrm{Var}\,X_2}}. \tag{4.26}$$

Correlation coefficient (4.26) takes values from [-1,1].

If X_1, X_2 are independent, it is $\rho\,(X_1, X_2) = 0$ (the converse implication does not hold).

We have $\rho\,(X_1, X_2) = 1$ if and only if $X_2 \overset{a.s.}{=} \alpha + \beta\, X_1,\ \beta > 0$. Similarly, $\rho\,(X_1, X_2) = -1$ if and only if $X_2 \overset{a.s.}{=} \alpha + \beta\, X_1,\ \beta < 0$. Thus, the value 1 (-1) correspond to the cases of **perfect positive (negative) linear dependence**.

Coefficient (4.26) is invariant under increasing linear transforms:

$$\rho\,(\alpha_1 + \beta_1\, X_1, \alpha_2 + \beta_2\, X_2) = \rho\,(X_1, X_2) \text{ for } \beta_1,\ \beta_2 > 0.$$

One shortcoming of (4.26) is in the fact, that the knowledge of marginal distributions together with the value of $\rho = \rho\,(X_1, X_2)$ does not determine the joint distribution of $(X_1, X_2)'$.

Also, for given marginal distribution functions F_1 and F_2 it is not always possible to construct a joint distribution with given value of $\rho \in [-1, 1]$. It can be shown that for given F_1 and F_2 the **attainable correlations** form a closed interval $[\rho_{min}, \rho_{max}]$ with $\rho_{min} < 0 < \rho_{max}$. It holds $\rho = \rho_{min}$ if and only if X_1 and X_2 are countermonotonic and $\rho = \rho_{max}$ if and only if X_1 and X_2 are comonotonic.

4.2.2 Rank Correlation

There are two characteristics used to measure rank correlation, Kendall's tau and Spearman's rho.

Definition 4.7. *For random variables X_1 and X_2* ***Kendall's tau*** *is given by*

$$\rho_\tau (X_1, X_2) = \mathrm{E}\left[\mathrm{sign}\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right)\right], \tag{4.27}$$

where $\left(\tilde{X}_1, \tilde{X}_2\right)'$ is an independent copy of $(X_1, X_2)'$.

Remark. Two points in $\mathbb{R}^2$, (x_1, x_2), $(\tilde{x}_1, \tilde{x}_2)$ are said to be **concordant** if $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0$ and to be **discordant** if $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0$.

For random vector $(X_1, X_2)'$ and its independent copy $\left(\tilde{X}_1, \tilde{X}_2\right)'$ we expect the probability of concordance be high relative to the probability of discordance, if X_2 tends to increase with X_1.

Definition 4.8. *For random variables X_1 and X_2 with distribution functions F_1 and F_2* ***Spearman's rho*** *is given by*

$$\rho_S (X_1, X_2) = \rho\left(F_1(X_1), F_2(X_2)\right). \tag{4.28}$$

Both coefficients (4.27) and (4.28), like the correlation coefficient (4.26), take values from [-1,1] and for independent random variables are equal to zero. Unlike (4.26), rank correlations do not depend on the marginal distributions of X_1 and X_2. It can be shown, that they take the value 1 when X_1 and X_2 are comonotonic and the value -1 when they are countermonotonic.

For random variables X_1 and X_2 with continuous distribution functions, (4.27) and (4.28) depend only on the unique copula of $(X_1, X_2)'$ and thanks to the Theorem 4.2 they are invariant under increasing transformations.

The connection between rank correlations and the copula is shown by the following theorem:

Theorem 4.6. *Suppose X_1 and X_2 have continuous marginal distributions and a copula C. Then*

$$\rho_\tau (X_1, X_2) = 4\int_0^1\int_0^1 C(u_1, u_2)\,\mathrm{d}C(u_1, u_2) - 1, \tag{4.29}$$

$$\rho_S (X_1, X_2) = 12\int_0^1\int_0^1 \left(C(u_1, u_2) - u_1 u_2\right)\mathrm{d}u_1\,\mathrm{d}u_2. \tag{4.30}$$

Proof. From the definition of Kendall's tau it follows

$$\begin{aligned} &\rho_\tau (X_1, X_2) \\ &= \mathrm{P}\left(\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right) > 0\right) - \mathrm{P}\left(\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right) < 0\right) \\ &= 2\,\mathrm{P}\left(\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right) > 0\right) - 1. \end{aligned}$$

From the interchangeability of (X_1, X_2) and $\left(\tilde{X}_1, \tilde{X}_2\right)$ we have

$$\begin{aligned}
\rho_\tau(X_1, X_2) &= 4\ \mathrm{P}\left(X_1 < \tilde{X}_1, X_2 < \tilde{X}_2\right) - 1 \\
&= 4\ \mathrm{E}\left[\mathrm{P}\left(X_1 < \tilde{X}_1, X_2 < \tilde{X}_2 \,|\, \tilde{X}_1, \tilde{X}_2\right)\right] - 1 \\
&= 4 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{P}(X_1 < x_1, X_2 < x_2)\,\mathrm{d}F(x_1, x_2) - 1,
\end{aligned}$$

where $F(x_1, x_2)$ is the joint d.f. of $(X_1, X_2)'$ (and also of $\left(\tilde{X}_1, \tilde{X}_2\right)'$).

Using Sklar's theorem we write

$$\rho_\tau(X_1, X_2) = 4 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C(F_1(x_1), F_2(x_2))\,\mathrm{d}C(F_1(x_1), F_2(x_2)) - 1,$$

from where we obtain (4.29) by substitution $u_1 = F_1(x_1)$, $u_2 = F_2(x_2)$.

The proof of (4.30) uses Höffding's formula

$$\mathrm{Cov}(X_1, X_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (F(x_1, x_2) - F_1(x_1)\ F_2(x_2))\,\mathrm{d}x_1\,\mathrm{d}x_2 \qquad (4.31)$$

that holds for any pair of random variables X_1, X_2, with joint d.f. F with finite covariance and marginal distribution functions F_1 and F_2.

Then (4.30) is obtained by (4.31) taking into account that the distribution of $(F_1(X_1), F_2(X_2))'$ is bivariate uniform on $(0,1)^2$ and that the variance of the uniform distribution is equal to $\frac{1}{12}$.

The proof of (4.31) uses an independent copy $\left(\tilde{X}_1, \tilde{X}_2\right)'$ of $(X_1, X_2)'$. We have

$$2\ \mathrm{Cov}(X_1, X_2) = \mathrm{E}\left(\left(X_1 - \tilde{X}_1\right)\ \left(X_2 - \tilde{X}_2\right)\right). \qquad (4.32)$$

For $a, b \in \mathbb{R}$ we can write

$$a - b = \int_{-\infty}^{+\infty} \left(I_{[b \le x]} - I_{[a \le x]}\right) \mathrm{d}x.$$

We rewrite similarly the product on the right-hand side of (4.32) and then we apply the expected value. We obtain

$$\begin{aligned}
&2\ \mathrm{Cov}(X_1, X_2) \\
&= \mathrm{E}\left[\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(I_{[\tilde{X}_1 \le x_1]} - I_{[X_1 \le x_1]}\right)\ \left(I_{[\tilde{X}_2 \le x_2]} - I_{[X_2 \le x_2]}\right) \mathrm{d}x_1\,\mathrm{d}x_2\right] \\
&= 2 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (\mathrm{P}(X_1 \le x_1, X_2 \le x_2) - \mathrm{P}(X_1 \le x_1)\ \mathrm{P}(X_2 \le x_2))\,\mathrm{d}x_1\,\mathrm{d}x_2.
\end{aligned}$$

□

4.2.3 Tail Dependence

Coefficients of tail dependence dealt with in this subsection measure the dependence in the tails of a bivariate distribution.

Definition 4.9. *The* ***coefficient of upper tail dependence*** *of random variables X_1 and X_2 with distribution functions F_1 and F_2 is*

$$\lambda_u = \lambda_u (X_1, X_2) = \lim_{q \to 1-} \mathrm{P}\left(X_2 > F_2^{-1}(q) \mid , X_1 > F_1^{-1}(q)\right) \tag{4.33}$$

provided that a limit $\lambda_u \in [0, 1]$ exists.

The ***coefficient of lower tail dependence*** *is*

$$\lambda_l = \lambda_l (X_1, X_2) = \lim_{q \to 0+} \mathrm{P}\left(X_2 \leq F_2^{-1}(q) \mid , X_1 \leq F_1^{-1}(q)\right) \tag{4.34}$$

provided that a limit $\lambda_l \in [0, 1]$ exists.

Remark. If $\lambda_u = 0$ ($\lambda_l = 0$), random variables are said to be **asymptotically independent** in the upper (lower) tail.

For continuous distribution functions F_1 and F_2 the coefficients of tail dependence can be expressed by means of the unique copula C of the bivariate distribution.

For (4.34) we write

$$\begin{aligned} \lambda_l &= \lim_{q \to 0+} \frac{\mathrm{P}\left(X_2 \leq F_2^{-1}(q), X_1 \leq F_1^{-1}(q)\right)}{\mathrm{P}\left(X_1 \leq F_1^{-1}(q)\right)} \\ &= \lim_{q \to 0+} \frac{C(q, q)}{q}. \end{aligned} \tag{4.35}$$

For (4.33) we have

$$\begin{aligned} \lambda_u &= \lim_{q \to 1-} \frac{\mathrm{P}\left(X_2 > F_2^{-1}(q), X_1 > F_1^{-1}(q)\right)}{\mathrm{P}\left(X_1 > F_1^{-1}(q)\right)} \\ &= \lim_{q \to 1-} \frac{\mathrm{P}\left(1 - F_2(X_2) \leq 1 - q, 1 - F_1(X_1) \leq 1 - q\right)}{1 - q}. \end{aligned} \tag{4.36}$$

For expression of λ_u it is useful to mention the concept of **survival copula**. For a d-dimensional d.f. $F(x_1, \ldots, x_d)$ we use the following generalization of univariate tail function:

$$\bar{F}(x_1, \ldots, x_d) = \mathrm{P}(X_1 > x_1, \ldots, X_d > x_d). \tag{4.37}$$

When $X_1, \dots, X_d$ have continuous distribution functions, we write

$$\begin{aligned}\bar{F}(x_1, \dots, x_d) &= \mathrm{P}\left(F_1(X_1) > F_1(x_1), \dots, F_d(X_d) > F_d(x_d)\right) \\ &= \mathrm{P}\big(1 - F_1(X_1) \le \bar{F}_1(x_1), \dots, 1 - F_d(X_d) \le \bar{F}_d(x_d)\big).\end{aligned}$$

We thus have for (4.37) the expression

$$\bar{F}(x_1, \dots, x_d) = \widehat{C}\left(\bar{F}_1(x_1), \dots, \bar{F}_d(x_d)\right), \tag{4.38}$$

where $\widehat{C}$ is the joint d.f. of $(1 - F_1(X_1), \dots, 1 - F_d(X_d))'$. Obviously, $\widehat{C}$ is a copula, it is called survival copula.

If copula C is the joint d.f. of random vector $\mathbf{U} = (U_1, \dots, U_d)'$ with d-dimensional uniform distribution on $(0,1)^d$, then the survival copula of copula C is the joint d.f. $\widehat{C}$ of $\mathbf{1} - \mathbf{U} = (1 - U_1, \dots, 1 - U_d)$.

Now we can rewrite (4.36) as

$$\lambda_u = \lim_{q \to 1-} \frac{\widehat{C}(1-q, 1-q)}{1-q} = \lim_{q \to 0+} \frac{\widehat{C}(q,q)}{q}, \tag{4.39}$$

where $\widehat{C}$ is the survival copula of the copula C.

In the case of a bivariate distribution it is easy to derive a relationship between a copula and its survival copula:

$$\begin{aligned}\widehat{C}(1 - u_1, 1 - u_2) &= \mathrm{P}(1 - U_1 < 1 - u_1, 1 - U_2 < 1 - u_2) \\ &= 1 - \mathrm{P}(U_1 < u_1) - \mathrm{P}(U_2 < u_2) \\ &\quad + \mathrm{P}(U_1 < u_1, U_2 < u_2) \\ &= 1 - u_1 - u_2 + C(u_1, u_2). \end{aligned} \tag{4.40}$$

Inserting (4.40) into (4.39) yields

$$\lambda_u = \lim_{q \to 1-} \frac{1 - 2q + C(q,q)}{1-q} = \lim_{q \to 1-} \left(2 - \frac{1 - C(q,q)}{1-q}\right)$$

and by L'Hospital's rule

$$\lambda_u = 2 - \lim_{q \to 1-} \frac{\mathrm{d}C(q,q)}{\mathrm{d}q}. \tag{4.41}$$

Example. Coefficient of upper tail dependence for Gumbel copula (4.17) is computed by means of (4.41) taking into account $C_\theta^{Gu}(u,u) = u^{2^{1/\theta}}$:

$$\lambda_u = 2 - 2^{1/\theta}.$$

For $\theta > 1$ Gumbel copula shows upper tail dependence that is getting stronger with higher values of θ (with $\lambda_u = 1$ as the limit value in case $\theta \to +\infty$, when we obtain comonotonicity as the limiting dependence structure).

In some cases the coefficients (4.33) and(4.34) have the same value. It occurs when the joint distribution of (X_1, X_2) is **radially symmetric**.

Definition 4.10. *A random vector* $\mathbf{X} = (X_1, \ldots, X_d)'$ *is radially symmetric about point* $\mathbf{a} = (a_1, \ldots, a_d)'$ *if* $\mathbf{X} - \mathbf{a} \overset{d}{=} \mathbf{a} - \mathbf{X}$.

For example, multivariate normal distribution with the density (4.14) is radially symmetric about $\boldsymbol{\mu}$.

For d-dimensional uniform distribution of vector $\mathbf{U}$, the only possible center of symmetry is the point $(0.5, \ldots, 0.5)$, vector U is radially symmetric if and only if $\mathbf{U} - \mathbf{0.5} \overset{d}{=} \mathbf{0.5} - \mathbf{U}$, which is equivalent to $\mathbf{U} \overset{d}{=} \mathbf{1} - \mathbf{U}$. Thus if C is a copula of a radially symmetric vector and $\widehat{C}$ is its survival copula, we have $\widehat{C} = C$. From (4.35) and (4.39) it is seen that if the copula of given bivariate distribution is radially symmetric, there is only one coefficient of tail dependence $\lambda = \lambda_u = \lambda_l$.

Example. Coefficient of tail dependence of Gauss copula. Consider the Gauss copula defined in (4.14) in bivariate case. We have

$$\begin{aligned} C_\rho^{Ga}(u_1, u_2) &= \mathrm{P}\left(\Phi(Y_1) \le u_1, \Phi(Y_2) \le u_2\right) \\ &= \Phi_\rho\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right), \end{aligned} \tag{4.42}$$

where $(Y_1, Y_2)'$ has bivariate normal distribution with standard marginal distribution functions and with the correlation coefficient ρ. From the symmetry of standard normal distribution about zero it follows that Gauss copula (4.14) is radially symmetric. We derive the coefficient of tail dependence using (4.35). By applying L'Hospital's rule to the limit in (4.35) we obtain

$$\lambda = \lim_{q \to 0+} \frac{\mathrm{d}C(q, q)}{\mathrm{d}q}. \tag{4.43}$$

An equality useful for expressing the derivative of copula concerns the conditional distribution of U_2 when $U_1 = u_1$ in the bivariate uniform distribution with the joint d.f. C. It holds

$$\begin{aligned} \mathrm{P}\left(U_2 \le u_2 \mid U_1 = u_1\right) &= \lim_{\delta \to 0} \frac{C(u_1 + \delta, u_2) - C(u_1, u_2)}{\delta} \\ &= \frac{\partial}{\partial u_1} C(u_1, u_2), \end{aligned}$$

which together with (4.43) yields

$$\lambda = \lim_{q \to 0+} \mathrm{P}\left(U_2 \le q \mid U_1 = q\right) + \lim_{q \to 0+} \mathrm{P}\left(U_1 \le q \mid U_2 = q\right). \tag{4.44}$$

From (4.42) it is seen that bivariate Gauss copula is the joint d.f. of

$$(U_1, U_2) = (\Phi(Y_1), \Phi(Y_2)),$$

which satisfies

$$(U_1, U_2) \stackrel{d}{=} (U_2, U_1). \tag{4.45}$$

Random vector with the property (4.45) is called **exchangeable**. Its joint d.f. is called **exchangeable copula**.

Thanks to the exchangeability of Gauss copula we rewrite (4.44) as

$$\begin{aligned} \lambda &= 2 \lim_{q \to 0+} \mathrm{P}(U_2 \le q \mid U_1 = q) \\ &= 2 \lim_{q \to 0+} \mathrm{P}\left(\Phi^{-1}(U_2) \le \Phi^{-1}(q) \mid \Phi^{-1}(U_1) = \Phi^{-1}(q)\right) \\ &= 2 \lim_{y \to -\infty} \mathrm{P}(Y_2 \le y \mid Y_1 = y). \end{aligned}$$

From the bivariate normal distribution of (Y_1, Y_2) we know that the conditional distribution of Y_2 when $Y_1 = y$ is normal, $N(\rho y, 1 - \rho^2)$. For the coefficient of tail dependence of the Gauss copula with parameter ρ we finally obtain

$$\lambda = 2 \lim_{y \to -\infty} \Phi\left(y\sqrt{1-\rho}/\sqrt{1+\rho}\right).$$

For $\rho < 1$ we have $\lambda = 0$. The Gauss copula is asymptotically independent in both tails.

4.3 Fitting Copulas to Data

In this section we briefly describe possible approaches to estimating the dependence structure from multivariate data.

We assume that we observe d-dimensional data vectors $\mathbf{X_1}, \ldots, \mathbf{X_n}$ with the joint d.f. F. We denote $\mathbf{X}_t = (X_{t,1}, \ldots, X_{t,d})$ for $t = 1, \ldots, n$. We assume that the d.f. F has continuous marginal distribution functions $F_1, \ldots, F_d$ and the unique copula C.

We start with the approach based on empirical values of rank correlation coefficients. Then we discuss fitting a copula to data by the method of maximum likelihood.

4.3.1 Fitting copulas using rank correlations

From the d-dimensional data we can estimate pairwise coefficients of rank correlation. The estimator of Kandall's tau (4.27) for the ith and jth element of the d-dimensional vector with d.f. F is

$$\rho_\tau \widehat{(X_i, X_j)} = \binom{n}{2}^{-1} \sum_{t<s} \text{sign}\left(\left(X_{t,i} - X_{s,i}\right)\left(X_{t,j} - X_{s,j}\right)\right). \tag{4.46}$$

The estimator of Spearman's rho (4.28) can be based on the rank data $(\text{rank}\,(X_{t,i}), \text{rank}\,(X_{t,j}))$, where $\text{rank}\,(X_{t,i})$ denotes the rank of $X_{t,i}$ in the data $X_{1,i},\ldots,X_{n,i}$ (its position in the ordered sample). We calculate the estimator of Spearman's rho as the empirical correlation coefficient for the pairs of rank data.

$$\begin{aligned} &\rho_S \widehat{(X_i, X_j)} \\ &= \frac{12}{n\left(n^2-1\right)} \sum_{t=1}^{n} \left(\text{rank}\,(X_{t,i}) - \frac{1}{2}(n+1)\right)\left(\text{rank}\,(X_{t,j}) - \frac{1}{2}(n+1)\right). \end{aligned} \tag{4.47}$$

The estimators (4.46) and(4.47) are applicable in the situation when we consider a specific copula for which we know the relationship between the parameters of the copula and a rank correlation coefficient. For example, assume that we have bivariate data with the joint d.f.

$$F(x_1, x_2) = C_\theta\left(F_1(x_1), F_2(x_2)\right),$$

where C_θ is a bivariate Archimedean copula with a parameter θ, generated by ϕ. For d.f. F with continuous marginal distributions it is possible to derive Kendall's tau in the form

$$\rho_\tau(X_1, X_2) = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}\,\mathrm{d}t.$$

In general we have

$$\rho_\tau(X_1, X_2) = f(\theta).$$

We can estimate the parameter θ by solving

$$\rho_\tau \widehat{(X_1, X_2)} = f\left(\widehat{\theta}\right).$$

We illustrate this approach on the example of Gumbel copula (4.17). For Gumbel copula Kendall's tau has the form

$$\rho_\tau = 1 - \frac{1}{\theta}$$

and the range of possible values of the parameter is $\theta \geq 1$. We thus obtain the estimate

$$\widehat{\theta} = (1 - \widehat{\rho_\tau})^{-1},$$

provided that $\widehat{\rho_\tau} \geq 0$.

For the bivariate Gauss copula (4.42) it holds

$$\rho_\tau = \frac{2}{\pi} \arcsin \rho, \tag{4.48}$$

$$\rho_S = \frac{6}{\pi} \arcsin \frac{1}{2} \rho. \tag{4.49}$$

The right-hand side of (4.49) is well approximated by the value of parameter ρ. For d-mensional data it is convenient to estimate the elements of the matrix R in (4.14) by the empirical pairwise Spearman's rank correlations.

4.3.2 Fitting copulas by maximum likelihood method

Let $C_{\boldsymbol{\theta}}$ be a copula determined by a vector of parameters $\boldsymbol{\theta}$. We could use the d-dimensional data directly for fitting the joint d.f.

$$F(x_1, \ldots, x_2) = C_{\boldsymbol{\theta}}(F_1(x_1), \ldots, F_d(x_d)).$$

It is convenient to work with an absolutely continuous copula, for which we have a density

$$c(u_1, \ldots, u_d) = \frac{\partial C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}.$$

It is recommended, instead of estimating from the data simultaneously parameters of marginal distributions and of the copula, to proceed in two steps.

In the first step we fit univariate marginal distributions F_i using the data $X_{1,i}, \ldots, X_{n,i}$. We fit an appropriate parametric distribution or alternatively, we can just derive the empirical distribution functions from the univariate data.

In the second step we estimate parameters of the copula by maximizing log-likelihood

$$\log L\left(\boldsymbol{\theta}; \widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n\right) = \sum_{t=1}^{n} \log c_{\boldsymbol{\theta}}\left(\widehat{\mathbf{U}}_t\right), \tag{4.50}$$

where the vectors $\widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n$ are pseudo-sample from the distribution with d.f. C formed by means of the fitted univariate distribution functions:

$$\widehat{\mathbf{U}}_t = \left(\widehat{F}_1(X_{t,1}), \ldots, \widehat{F}_d(X_{t,d})\right)'.$$

Example. The density of the Gauss copula (4.14) has the form

$$c(u_1, \ldots, u_d) = \frac{\phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)}{\phi\left(\Phi^{-1}(u_1)\right) \ldots \phi\left(\Phi^{-1}(u_d)\right)},$$

where ϕ_R is the density of multivariate normal distribution with standard marginal distributions and the correlation matrix R. The log-likelihood (4.50) becomes

$$\begin{aligned}&\log L\left(R, \widehat{U}_1, \ldots, \widehat{U}_n\right)\\ &= \sum_{t=1}^{n} \log \phi_R\left(\Phi^{-1}(U_{t,1}), \ldots, \Phi^{-1}(U_{t,d})\right) - \sum_{t=1}^{n}\sum_{i=1}^{d} \log \phi\left(\Phi^{-1}(U_{t,j})\right).\end{aligned}$$

For the maximum likelihood estimate of the matrix R it holds

$$\widehat{R} = \operatorname*{argmax}_{\Sigma} \sum_{t=1}^{n} \log \phi_\Sigma\left(\Phi^{-1}(U_{t,1}) \ldots, \Phi^{-1}(U_{t,d})\right),$$

where the maximum is taken over all possible correlation matrices.

Bibliography

[1] H.Panjer, G.E.Willmot: *Insurance Risk Models*, Society of Actuaries, Schaumburg, Illinois, 1992.

[2] N.L.Bowers, H.U.Gerber, J.C.Hickman, D.A.Jones, C.J.Nesbitt: *Actuarial Mathematics*, Society of Actuaries, Itasca, Illinois, 1986.

[3] A.J.McNeil, R.Frey, P.Embrechts: *Quantitative Risk Management*, Princeton University Press, 2005.

[4] P. Embrechts, C.Klüppelberg, T.Mikosch: *Modelling Extremal Events*, Springer-Verlag, Berlin Heidelberg, 1997.

[5] R.D.Reiss, M.Thomas: *Statistical Analysis of Extreme Values*, Birkhäuser, Basel, 1997.