1 Introduction to Non-Life Insurance Mathematics

Lit.: Straub, ch. 1-3, EKM, 1.1-1.2

1.1 Basic models and concepts

initial reserves of company:	u	
reserves at time t:	R_t	
accumulated premium up to t:	P_t	Ass: $P_t = \pi \cdot t,\ \pi$ annual premium

claims at times $T_1, T_2, \ldots$ with claim amounts $X_1, X_2, \ldots Y_1 = T_1,\ Y_j = T_j - T_{j-1}$, $j = 2, 3, \ldots$ interoccurence times between claims

risk process: $R_t = u + \pi t - S(t)$
$S(t) = \sum_{j=1}^{N(t)} X_j$ total claim amount up to time t
$N(t) = \max\{k;\ T_k \leq t\}$ no. of claims up to time t

Renewal model: a) $Y_1, Y_2, \ldots, X_1, X_2, \ldots$ independent positive random variables
b) $X_1, X_2, \ldots$ i.i.d. with $\mathcal{L}(X_j) = F$, finite mean $\mathcal{E}X_j = \mu$ and $\sigma^2 = \text{var}X_j \leq \infty$
c) $Y_1, Y_2, \ldots$ i.i.d. with finite mean $\mathcal{E}\mathcal{Y}_| = \frac{\infty}{\lambda}$

Special cases: Cramér-Lundberg (CL) model if $\mathcal{L}(\mathcal{Y}_|) = \mathcal{E}\S_{\surd}(\lambda)$
Erlang's model if $\mathcal{L}(\mathcal{Y}_|) = \mathcal{E}\S_{\surd}(\lambda)$ and $\mathcal{L}(\mathcal{X}_|) = \mathcal{E}\S_{\surd}\left(\frac{\infty}{\mu}\right)$

In the CL-model, $\{T_1, T_2, \ldots\}$ is homogeneous Poisson process with intensity λ and $\mathcal{L}(N(t)) = \mathcal{P}(\lambda \cdot t)$.

Lemma 1.1 *If $Y_1, Y_2, \ldots$ i.i.d. Exp(λ), then $T_n = \sum_{j=1}^{n} Y_j$ is Gamma-distributed ($\mathcal{L}(\mathcal{T}_\backslash) = -_\backslash(\lambda)$), i.e. its density is*

$$p_n(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x},\ x \geq 0.$$

1.2 Moment-generating functions and the total claim process

Z real random variable. If they exist:

$\mathcal{E}(\mathcal{Z}^\backslash)$	n-th moment of $Z,\ n \geq 1$
$\mathcal{E}(\mathcal{Z} - \mathcal{E}\mathcal{Z})^\backslash$	n-th central moment of $Z,\ n \geq 2$

Definition: a) moment-generating function of Z : $\mu_Z(\tau) = \mathcal{E}(\rceil^{\tau \mathcal{Z}})$
b) log-moment-generating function of Z : $\lambda_Z(\tau) = \log \mu Z(\tau)$

Lemma 1.2 *Let $\mu_Z(\tau)$ exist in a neighbourhood of $\tau = 0$. Then,*
a) $\mu_Z^{(n)}(0) = \mathcal{E}\mathcal{Z}^{\setminus},\ \setminus \geq \prime$
b) $\lambda'_Z(0) = \mathcal{E}\mathcal{Z},\ \lambda''_{\mathcal{Z}}(\prime) = var\mathcal{Z}\ ,\ \lambda'''_{\mathcal{Z}}(\prime) = \mathcal{E}(\mathcal{Z} - \mathcal{E}\mathcal{Z})^{\ni}$

Remark: If Y, Z independent, then $\mu_{Y+Z}(\tau) = \mu_Z(\tau) \cdot \mu_Z(\tau)$ and $\lambda_{Y+Z}(\tau) = \lambda_Y(\tau) + \lambda_Z(\tau)$
If $X_1, X_2, \ldots$ i.i.d., $S_n = X_1 + \ldots + X_n$, then $\lambda_{S_n}(\tau) = n\ \lambda_{X_1}(\tau)$

Lemma 1.3 *Let $N \geq 0$ random integer, independent of i.i.d. $X_1, X_2, \ldots, S_N = X_1 + \ldots + X_N$. Then, with $p_k = pr(N = k),\ k \geq 0$,*
a) $pr(S_N \leq x) = \sum_{k=0}^{\infty} p_k \underbrace{F * \ldots * F}_{k-times}(x)$
b) $\mu_{S_N}(\tau) = \mu_N(\lambda_{X_1}(\tau))$ *and* $\lambda_{S_N}(\tau) = \lambda_N(\lambda_{X_1}(\tau))$

For the total claim amount $S(t)$ of section 1.1:

$$\lambda_{S(t)} = \lambda_{N(t)}(\lambda_{X_1}(\tau))$$

Special case CL-model: $\lambda_{S(t)} = \lambda t(\mu_{X_1}(\tau) - 1)$

Corollary 1.1 *For the renewal model,* $\mathcal{E}\mathcal{R}_{\sqcup} = \sqcap + \pi \cdot \sqcup - \mu\mathcal{E}\mathcal{N}(\sqcup)$.
For the CL-model, $\mathcal{E}\mathcal{R}_{\sqcup} = \sqcap + (\pi - \mu\lambda) \cdot \sqcup$.

Definition $\rho = \frac{\pi}{\lambda\mu} - 1$ safety loading; net profit condition: $\rho > 0$ $\quad(\rightarrow \mathcal{E}\mathcal{R}_{\sqcup} > \prime$ for all t)

1.3 Poisson processes

Definition 1.1 *A stochastic process $N(t),\ t \geq 0$, in continuous time is a Poisson process if*

1) $N(0) = 0$ *a.s.*
2) *for all $n \geq 1,\ 0 = t_0 < t_1 < \ldots < t_n$, the increments $N(t_{i-1}, t_i] = N(t_i) - N(t_{i-1}),\ i = 1, \ldots, n$, are independent.*
3) *For some non-decreasing, right-continuous function $\nu : [0, \infty) \rightarrow [0, \infty)$ with $\nu(0) = 0$:*

$$\mathcal{L}(N(s,t]) = \mathcal{P}(\nu(s,t]) \quad \textit{for all } 0 \leq s < t,$$

where $N(s,t] = N(t) - N(s),\ \nu(s,t] = \nu(t) - \nu(s)$. $\nu(t)$ is called the mean function.

4) *With probability 1, the path $N(t),\ t \geq 0$, is right-continuous and has limits from the left everywhere (càdlàg).*

Remark: a) $\mathcal{L}(N) = \mathcal{P}(\lambda) \Longrightarrow \mathcal{E}N = \lambda = \text{var}N$
b) $N(t)$ Poisson process $\Longrightarrow \nu(t) = \mathcal{E}N(t)$ and

$$\text{pr}\ (N(t_1) = k_1,\ N(t_2) = k_1+k_2, \ldots, N(t_n) = k_1+\ldots+k_n) = e^{-\nu(t_n)} \prod_{j=1}^{n} \frac{(\nu(t_j) - \nu(t_{j-1}))^{k_j}}{k_j!}$$

for all $n \geq 1,\ 0 = t_0 < t_1 < \ldots t_n,\ k_1, \ldots, k_n \in \mathbb{N} \cup \{0\}$.

Definition 1.2 *a) A Poisson process is homogeneous with intensity $\lambda > 0$ if $\nu(t) = \lambda t$. If $\lambda = 1$, it is a standard homogeneous Poisson process (SHPP).*
b) If, for a general Poisson process, ν is absolutely continuous with density λ, i.e.

$$\nu(s, t] = \int_s^t \lambda(u) du\ ,\ 0 \leq s < t,$$

$\lambda(t)$ is called the intensity (function).

Definition 1.3 *A stochastic process $X(t),\ t \geq 0$, in continuous time is a Lévy process if*
1) $X(0) = 0$ a.s.
2) for all $n \geq 1,\ 0 = t_0 < t_1 < \ldots < t_n$, the increments $X(t_i) - X(t_{i-1}),\ i = 1, \ldots, n$, are independent
3) the increments are stationary, i.e. $\mathcal{L}(X(t) - X(s))$ depends only on $t - s$.
4) With probability 1, the sample path $X(t),\ t \geq 0$, is càdlàg.

Examples: a) homogeneous Poisson process
b) Brownian motion (Wiener process)
c) total claim amount $S(t)$ of Cramér-Lundberg model

Proposition 1.1 *Let $N(t)$ be a Poisson process with mean function ν, $N_0(t)$ a SHPP.*
a) $N_0(\nu(t)),\ t \geq 0$, is a Poisson process with mean function ν.
b) If ν is continuous, increasing and $\lim_{t\to\infty} \nu(t) = \infty$, then $N(\nu^{-1}(t)),\ t \geq 0$, is a SHPP.

Theorem 1.1 *a) Let $Y_1, Y_2, \ldots$, i.i.d. Exp(λ), $T_n = \sum_{j=1}^n Y_j$. Then, $N(t) = \#\{j;\ T_j \leq t\}$ is a homogeneous Poisson process with intensity λ.*
b) If $N(t)$ is a homogenous Poisson process with intensity λ, there are $Y_1, Y_2, \ldots$ i.i.d. Exp(λ) such that $N(t) = \#\{j,\ T_j \leq t\}$ with $T_n = \sum_{j=1}^n Y_j$.

Proposition 1.2 *Let $N(t)$ be a Poisson process with mean function $\nu(t)$ and continuous intensity function $\lambda(t) > 0$ a.e.. Then, for $Y_1, Y_2, \ldots, T_1, T_2, \ldots$ as in Theorem 1.1 b)*
a) $(T_1, \ldots, T_n)$ has the joint density

$$f(t_1, \ldots, t_n) = e^{-\nu(t_n)} \prod_{j=1}^{n} \lambda(t_j) \qquad \text{if } 0 < t_1 < \ldots < t_n,$$

and $f(t_1, \ldots, t_n) = 0$, else
b) $(Y_1, \ldots, Y_n)$ has the joint density

$$g(y_1, \ldots, y_n) = e^{-\nu(y_1+\ldots+y_n)} \prod_{j=1}^{n} \lambda(y_1 + \ldots y_j) \qquad \text{if } y_1, \ldots, y_n \geq 0,$$

and $g(y_1, \ldots, y_n) = 0$, else.

1.4 Cramér's inequality for ruin probabilities

ruin at time t if risk process $R_t < 0$ and $\lim_{s \uparrow t} R_s > 0$.
ruin occurs only at claim times $T_1, T_2, \ldots$ if at all.

ruin probability, finite horizon $t_0 < \infty$:	$\psi_{t_0}(u) = \mathrm{pr}(S(t) - \pi t > u \text{ for some } t \leq t_0)$
ruin probability, infinite horizon	$\psi(u) = \mathrm{pr}(S(t) - \pi t > u \text{ for some } t < \infty)$

Notation: $\delta(u) = 1 - \psi(u)$

Main renewal argument: If first claim at T_1 does not cause riun, then risk process starts anew with new initial time T_1 and new initial reserves $u_1 = u_1(X_1 T_1) = u + \pi T_1 - X_1 = R_{T_1}$.

Formally: $\mathcal{L}(\mathcal{R}_\sqcup - \mathcal{R}_\prime) = \mathcal{L}(\mathcal{R}_{\mathcal{T}_\infty + \sqcup} - \mathcal{R}_{\mathcal{T}_\infty} | \mathcal{T}_\infty, \mathcal{X}_\infty)$ for all $t > 0$.

Theorem 1.2 *Assume Cramér-Lundberg model and net profit condition $\rho > 0$. Then, $\delta(y)$ satisfies the integral equation*

$$\delta(y) - \delta(0) = \frac{\lambda}{\pi} \int_0^y \delta(y - x) \overline{F(x)} dx$$

with boundary condition

$$\psi(0) = 1 - \delta(0) = \frac{\lambda \mu}{\pi}$$

where $\overline{F}(x) = 1 - F(x)$ denotes the tail of $\mathcal{L}(\mathcal{X}_|)$.

Remark: By a limit argument, for $\rho \downarrow 0$, i.e. $\pi \downarrow \lambda \mu$, we get

$$\psi(u) = 1 \quad \text{for all} \quad u, \;\; \text{if } \rho \leq 0.$$

Corollary 1.2 *For the Erlang-model:*

$$\psi(u) = \begin{cases} \frac{\lambda\mu}{\pi} \exp(-\frac{1}{\mu}(1 - \frac{\lambda\mu}{\pi})u) & \text{if} \;\; \rho > 0 \\ 1 & \text{if} \;\; \rho \leq 0 \end{cases}$$

Definition 1.4 *Let $\overline{F}(x)$ denote the tail $1 - F(x)$ of $\mathcal{L}(\mathcal{X}_|)$. The risk process $\{R_t\}$ satisfies the Cramér-Lundberg condition if there is a $\nu > 0$ with*

$$\int_0^\infty e^{\nu x} \overline{F}(x) dx = \frac{\pi}{\lambda}.$$

ν is called the Lundberg exponent or adjustment coefficient.

Theorem 1.3 *(Cramér's inequality): Let $\{R_t\}$ be a Cramér-Lundberg process satisfying the Cramér-Lundberg condition. Then:*

$$\psi(u) \leq e^{-\nu u}.$$

Idea of proof: Ruin occurs at most at times $T_1, T_2, \ldots$
With $V_j := X_j - \pi Y_j$ loss in time interval $(T_{j-1}, T_j]$
$R_{T_k} = u - \sum_{j=1}^k V_j =: u - W_k$
$R_{T_k} < 0$, i.e. ruin at T_k iff $W_k > u$
Apply following theorem.

Theorem 1.4 *Let $V_1, V_2, \dots$ i.i.d. real random variables, $W_k = \sum_{j=1}^{k} V_j$ the corresponding random walk. Let there exist a $\nu > 0$ such that*

$$\mathcal{E}\rceil^{\nu\mathcal{V}_|} \equiv \psi_{\mathcal{V}_|}(\nu) = \infty.$$

Then, $\Psi(u) = pr(W_k > u$ for some $k \geq 1) \leq e^{-\nu u}$.

Corollary 1.3 *Under the assumption of Th. 1.2:*

$$\overline{F}(x) = 1 - F(x) \leq e^{-\nu x}(1 + \frac{\pi\nu}{\lambda})$$

Remark: The tail $\overline{F}(x)$ of the claim amount distribution $\mathcal{L}(\mathcal{X}_|)$ decreases exponentially with x, i.e. large values of X_j are quite unlikely. Therefore, the Cramér-Lundberg condition is a "small claim condition", i.e. the probability for extremely large individual claims is very small. This is not always a realistic assumption.

1.5 Premiums

For risk process R_t as above consider annual premium π in relation to annual risk (i.e. total claim amount in year $t+1$) $S(t+1) - S(t)$ which, for any t, is distributed as $S(1) \equiv Z$.

Premium calculation principle H: Given nonnegative random variable Z, representing a risk, with distribution $\mathcal{L}(\mathcal{Z})$. $\mathcal{H}$ is a mapping of $\mathcal{L}(\mathcal{Z})$ onto the premium π required for insuring the risk:

$$H : \mathcal{L}(\mathcal{Z}) \mapsto \pi \quad \text{(Notation: } \mathcal{H}(\mathcal{Z}) = \pi)$$

Examples:

a) pure risk premium: $\pi = \mathcal{E}\mathcal{Z}$ (certain ruin, even for large initial reserves)

b) loading proportional to $\mathcal{E}\mathcal{Z} : \pi = (\infty + \delta)\mathcal{E}\mathcal{Z}$ for some $\delta > 0$

c) σ-loading: $\pi = \mathcal{E}\mathcal{Z} + \beta\,\sigma(\mathcal{Z})$ for some $\beta > 0,\ \sigma^2(Z) = \text{var}Z$

d) σ^2-loading: $\pi = \mathcal{E}\mathcal{Z} + \beta\,\text{var}\mathcal{Z}$ for some $\beta > 0$

e) covariance loading: $\pi_1 = \mathcal{E}\mathcal{Z}_\infty + \beta\,\text{cov}(\mathcal{Z}_\infty, \mathcal{Z} + \mathcal{Z}_\infty)$ as a premium for including new risk Z_1 into existing portfolio of risks with total value Z.

These are heuristic, but practically important principles. In the rest of this section, we discuss two theoretically based principles.

A utility $U(x),\ x > 0$, measures the value attributed to an amount x of money. Two utilities U_1, U_2 are equivalent if $U_1(x) = U_2(x)\cdot a + b,\ a > 0$. Therefore, as standardization, we require $U(0) = 0,\ U'(0) = 1$.
Intuitively, $U'(x) \geq 0$, i.e. $U(x)$ increases. We require usually a bit more:

$$U \in C^2,\ U'(x) > 0,\ U''(x) \leq 0 \quad \text{for all} \quad x > 0.$$

Zero-utility principle: expected utility after 1 year $\geq$ zero utility, and $\pi = H(Z)$ defined by requiring equality:

$$\mathcal{E}\mathcal{U}(\pi - \mathcal{Z}) = \mathcal{U}(\prime) \quad (= \prime \quad \text{by standardization}).$$

Alternatively, taking initial reserves into account,

$$\mathcal{E}\mathcal{U}(\sqcap + \pi - \mathcal{Z}) = \mathcal{U}(\sqcap).$$

Special cases:
i) $U(x) = x \rightsquigarrow$ zero-utility principle $\equiv$ pure risk premium
ii) $U(x) = \frac{1}{a}(1 - e^{-ax})$ for some $a > 0$: exponential utility with risk aversion parameter a. Then,

$$\text{zero-utility principle} \rightsquigarrow \pi = \frac{1}{a} \log \mathcal{E}\rceil^{\dashv\mathcal{Z}} = \frac{\infty}{\dashv}\lambda_{\mathcal{Z}}(\dashv).$$

In particular, for $\mathcal{L}(\mathcal{Z}) = \mathcal{N}(\mu, \sigma^{\in})$:

$$\pi = \mu + \frac{a}{2}\sigma^2 \quad , \text{ i.e. } \sigma^2 - \text{loading}$$

Definition: Premium calculation principle H is additive if for independent risks Z_1, Z_2 : $H(Z_1 + Z_2) = H(Z_1) + H(Z_2)$.

Proposition 1.3 *Zero-utility principle additive iff* $U(x) = x$ *or*

$$U(x) = \frac{1}{a}(1 + e^{-ax}) \quad \textit{for some} \quad a > 0.$$

Expected value principle: Given $f : [0, \infty) \to [0, \infty)$ continuous, strictly increasing. $\pi = H(z)$ is defined by

$$f(\pi) = \mathcal{E}\{(\mathcal{Z}) \quad \text{or} \quad \pi = \{^{-\infty}(\mathcal{E}\{(\mathcal{Z}))$$

Special case: $f(x) = e^{ax} \rightsquigarrow$ zero utility principle with exponential utility

Definition: Premium calculation principle H is iterative if for any random variable Y, representing information on risk Z, and any risk Z

$$H(Z) = H(H(Z|Y))$$

(analogy to: $\mathcal{E}\mathcal{Z} = \mathcal{E}(\mathcal{E}\{\mathcal{Z}|\mathcal{Y}\})$).

Proposition 1.4 *H iterative iff H is an expected value principle for some suitable f.*

Remark: For $f(x) = e^{ax}$, H is iterative and additive.

Let $\nu > 0$ be given by $\mathcal{E}e^{\nu(z-\pi)} = 1$. For initial reserves u, ruin occurs if $Z - \pi > u$. As in proving Theorem 1.2 (but much simpler), a Cramér-type inequality follows:
$\text{pr}(Z - \pi > u) \leq e^{-\nu u}$ for all $u \geq 0$. If π is calculated by expected value principle with $f(x) = e^{ax}$($\equiv$ zero utility principle with exponential utility) $\rightsquigarrow \nu = a$.

1.6 Experience rating and credibility theory

Portfolio of risks partitioned into N tariffs or risk classes.
Available data from T years:

$$
\begin{array}{lll}
V_{tj} & & \text{insured volume in tariff } j \text{ in year } t\ ;\ j=1,\dots,N,\ \ t=1,\dots,T \\
X_{tj} & & \text{total claim amount in tariff } j \text{ in year } t\ ;\ j=1,\dots,N,\ \ t=1,\dots,T \\
\xi_{tj} & = & \dfrac{X_{tj}}{V_{tj}} \text{ relative costs in tariff } j \text{ in year } t\ ;\ j=1,\dots,N,\ \ t=1,\dots,T
\end{array}
$$

Notation: $V.j = \sum_{t=1}^{T} V_{tj}$, $V_{t\cdot} = \sum_{j=1}^{N} V_{tj}$, $V.. = \sum_{t,j} V_{tj}$, $X_{\cdot j}, X_{t\cdot}, X..$ analogously.
$\xi_{\cdot j} = X_{\cdot j}/V_{\cdot j}$, $\xi_{t\cdot} = X_{t\cdot}/V_{t\cdot}$, $\xi.. = X../V..$

The ξ's are also called **loss ratios**.

Credibility problem: Estimate expected loss ratio in each tariff:

$$\mu_j = \mathcal{E}(\xi_{\cdot j}) = ?$$

Naive solutions: $\hat{\mu_j} = \xi_{\cdot j}$ (ignoring large variability in small classes)
$\hat{\mu_j} = \xi..$ (ignoring differences between classes)

Alternative approach $\rightsquigarrow \hat{\mu_j} = \gamma_j\, \xi_{\cdot j} + (1-\gamma_j)\, \xi_0$

$$
\begin{aligned}
\text{with } \gamma_j &= \frac{w\ V_{\cdot j}}{v + w\ V_{\cdot j}}, \quad \begin{array}{lll} w & = & \text{variance of } X_{\cdot j} \text{ within portfolio} \\ v & = & \text{variance of } X_{tj} \text{ in time} \end{array} \\
\xi_0 &= \sum_{i=1}^{N} \frac{\gamma_i}{\gamma} X_{\cdot i}\ ,\ \gamma = \sum_{i=1}^{N} \gamma_i
\end{aligned}
$$

In general: $\xi_0 \neq \xi..$

Assumption: Each risk class is characterized by risk parameter ν_j, $j = 1, \dots, N$, such that:

$$
\begin{aligned}
\mathcal{E}\{\xi_{tj} | \nu_j = \nu\} &= \mu(\nu) \\
\text{var}\{\xi_{tj} | \nu_j = \nu\} &= \frac{1}{V_{tj}}\, \sigma^2(\nu)
\end{aligned}
$$

Example: Assuming a Cramér-Lundberg model for each risk with identical distribution F of individual claim amounts for all risk classes and Poisson-intensity λ_j of claim occurence times in risk class $j = 1, \dots, N$, we choose $\nu_j = \lambda_j$, i.e. risk parameter $\equiv$ intensity of claims. As individual claims have the same distribution and are insured in the same manner:

volume $V_{tj} \equiv$ number of risks insured in class j and year t.

Then, using notation of section 1.1

$$X_{tj} \underset{d}{=} \sum_{k=1}^{V_{tj}} S_k(1) \qquad (\underset{d}{=} \text{ stands for "is distributed as"})$$

where $S_1(1), S_2(1), \ldots$ are independent annual claim amounts distributed like:

$$S_k(1) \underset{d}{=} \sum_{l=1}^{N(1)} X_l$$

with $N(1),\ X_1, X_2, \ldots$ independent, $\mathcal{L}(N(1)) = \mathcal{P}(\lambda_j), \mathcal{L}(X_l) = F$. By Corollary 1.1 and using a similar argument for the variance:

$$\mathcal{E}S_k(1) = \mu\ \mathcal{E}N(1) = \mu\lambda_j$$

$$\mathrm{var}S_k(1) = \mu^2 \mathrm{var}N(1) + \sigma^2 \mathcal{E}N(1) = (\mu^2 + \sigma^2)\lambda_j$$

with $\mu = \mathcal{E}X_l,\ \sigma^2 = \mathrm{var}X_l$. We get $\mathcal{E}\{X_{tj}|\lambda_j\} = V_{tj}\mu\lambda_j$ and $\mathrm{var}\{X_{tj}|\lambda_j\} = V_{tj}(\mu^2+\sigma^2)\lambda_j$. Therefore,

$$\begin{aligned}\mathcal{E}\{\xi_{tj}|\lambda_j\} &= \frac{1}{V_{tj}}\,\mathcal{E}\{X_{tj}|\lambda_j\} = \mu\lambda_j \equiv \mu(\lambda_j)\\ \mathrm{var}\{\xi_{tj}|\lambda_j\} &= \frac{1}{V_{tj}^2}\,\mathrm{var}\{X_{tj}|\lambda_j\} = \frac{1}{V_{tj}}(\mu^2+\sigma^2)\ \lambda_j \equiv \frac{1}{V_{tj}}\sigma^2(\lambda_j).\end{aligned}$$

Bayesian approach: The unknown risk parameter $\nu_1, \ldots, \nu_N$ are treated as i.i.d. random variables with a-priori-distribution G and distribution function $G(u) = \mathrm{pr}(\nu_j \le u)$, called the **structure function** of the portfolio. $\mathrm{var}(\nu_j)$ is a measure for the heterogeneity of risks.

Assumption: The data $V_{tj}, \xi_{tj},\ t = 1, \ldots, T,\ j = 1, \ldots, N$ are given, where $\mathcal{E}\{\xi_{tj}|\nu_j\} = \mu(\nu_j),\ \mathrm{var}\{\xi_{tj}|\nu_j\} = \frac{1}{V_{tj}}\sigma^2(\nu_j)$ for all t, j and ξ_{tj}, ξ_{si} are conditionally independent given $\underset{\sim}{\nu} = (\nu_1, \ldots, \nu_N)$ if $t \ne s$ or $j \ne i$.

Notation: $m = \mathcal{E}\,\mu(\nu_j) = \int\ \mu(u)dG(u),\ v = \mathcal{E}\sigma^2(\nu_j),\ w = \mathrm{var}(\mu(\nu_j))$

A simple formulation of the **credibility problem**: For any given risk class k, estimate the average loss ratio $\mathcal{E}\{\xi_{tk}|\nu_k\} = \mu(\nu_k)$ by a linear, unbiased estimate $\hat{\mu}_k$:

$$\hat{\mu}_k = \sum_{t,j} \alpha_{tj}\xi_{tj},\ \mathcal{E}\hat{\mu}_k = \mathcal{E}\ \mu(\nu_k)\ \text{(recall: } \nu_k \text{ random!)}$$

such that

$$\mathcal{E}(\hat{\mu}_k - \mu(\nu_k))^2 = \min_{\alpha_{tj}}!$$

α_{tj} depend on k !

$$\begin{aligned}\text{Using }\ \mathcal{E}(\xi_{tj}\xi_{si})| &= \mathcal{E}(\mathcal{E}\{\xi_{tj}\xi_{si}|\underset{\sim}{\nu}\}) = m^2 + \delta_{ij}w + \delta_{ij}\delta_{st}\,\frac{v}{V_{tj}}\\ \mathcal{E}(\xi_{tj}\mu(\nu_k)) &= m^2 + \delta_{ik}w\\ \mathcal{E}\mu^2(\nu_k) &= m^2 + w\end{aligned}$$

this reduces to the following constrained minimization problem:

$$\min_{\alpha_{tj}} \sum_{t,j,s,i} \alpha_{tj}\alpha_{si}\left[m^2 + \delta_{ij}w + \delta_{ij}\delta_{st}\frac{v}{V_{tj}}\right] - 2\sum_{t,j}\alpha_{tj}[m^2 + \delta_{jk}w] + [m^2 + w]$$

under the constraint $\sum_{t,j} \alpha_{tj} = 1$

with solution: $\alpha_{tj} = \frac{V_{tj}}{V_{\cdot j}} \left(\frac{\gamma_j}{\gamma}(1-\gamma_k) + \delta_{jk}\gamma_k \right)$

and $\hat{\mu}_k = \sum_{t,j} \alpha_{ij}\xi_{tj} = \underbrace{\gamma_k \, \xi_{\cdot k}}_{\text{individual experience}} + (1-\gamma_k) \underbrace{\sum_{j=1}^{N} \frac{\gamma_j}{\gamma}\xi_{\cdot j}}_{\substack{=\xi_0 \\ \text{overall experience}}}$.

The **credibility factor** $\gamma_j = w\, V_{\cdot j}/(w\, V_{\cdot j} + v)$ depends on unknowns w, v.
Unbiased estimates are: $\hat{v} = V_{\cdot\cdot} \cdot s_t^2$, $\hat{w} = \frac{1}{w} = \frac{1}{Q}\,(s_g^2 - s_t^2)$

$$\begin{aligned} s_t^2 &= \frac{1}{N}\sum_{j=1}^{N}\frac{1}{T-1}\sum_{t=1}^{T}\frac{V_{tj}}{V_{\cdot\cdot}}(\xi_{tj}-\xi_{\cdot j})^2 \\ s_g^2 &= \frac{1}{TN-1}\sum_{t,j}\frac{V_{tj}}{V_{\cdot\cdot}}(\xi_{tj}-\xi_{\cdot\cdot})^2 \\ Q &= \frac{1}{TN-1}\sum_{j=1}^{N}\frac{V_{\cdot j}}{V_{\cdot\cdot}}(1-\frac{V_{\cdot j}}{V_{\cdot\cdot}}) \end{aligned}$$

1.7 Cramér-Lundberg Theory for Large Claims

Notation: For claim sizes X_j with tail distribution $\overline{F}(x) = \text{pr}(X_j > x)$ and $\mu = \mathcal{E}X_j$, the integrated tail distribution is given by

$$F_I(x) = \frac{1}{\mu}\int_0^x \overline{F}(y)dy.$$

F_I itself is a distribution function on $(0,\infty)$. In terms of F_I, the C-L-condition is: there is a $\nu > 0$ such that

$$\int_0^\infty e^{\nu x} dF_I(x) = \frac{1}{\mu}\int_0^\infty e^{\nu x}\overline{F}(x)dx \overset{!}{=} \frac{\pi}{\mu\lambda} = 1+\rho$$

with net profit condition $\rho \equiv \frac{\pi}{\mu\lambda} - 1 > 0$.

Definition: A distribution G on $(0,\infty)$ belongs to class $\mathcal{K}$ if, for all $\varepsilon > 0$,

$$\int_0^\infty e^{\varepsilon x} dG(x) = \infty.$$

If $F_I \in \mathcal{K}$, the corresponding claim size distribution cannot satisfy the C-L-condition.

Definition: a) A positive, measurable function L on $(0,\infty)$ is **slowly varying**, $(L \in \mathcal{R}_0)$, if $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$ for all $t > 0$.

b) A positive, measurable function H is **regularly varying of index** $\alpha \in \mathbb{R}$ ($H \in \mathcal{R}_\alpha$), if $\lim_{x\to\infty} \frac{H(tx)}{H(x)} = t^\alpha$ for all $t > 0$.

Remark: If $H \in \mathcal{R}_\alpha$, then $L(x) := \frac{H(x)}{x^\alpha} \in \mathcal{R}_0$, i.e. $H(x) = x^\alpha L(x)$. If $\overline{F}(x) \in \mathcal{R}_{-\alpha-1}$ for some $\alpha > 0$, then $F_I \in \mathcal{K}$.

Lemma 1.4 *Let X, Y be independent r.v. with tail distributions $\overline{F}(x) = \frac{1}{x^\alpha} L_1(x)$, $\overline{F}_2(x) = \frac{1}{x^\alpha} L_2(x) \in \mathcal{R}_{-\alpha}$, for some $\alpha > 0$. Then, the tail distribution $\overline{G}(z) = pr(X + Y > z)$ of $X + Y$ is in $\mathcal{R}_{-\alpha}$, too. More precisely:*

$$\overline{G}(x) \sim \frac{1}{x^\alpha}(L_1(x) + L_2(x)) \quad \textit{for } x \to \infty.$$

Corollary 1.4 *Let $X_1, X_2, \ldots$ i.i.d. with tail distribution $\overline{F}(x) = \frac{1}{x^\alpha} L_1(x) \in \mathcal{R}_{-\alpha}$ for some $\alpha > 0$. Then, the tail distribution of $S_n = X_1 + \ldots X_n$ satisfies*

$$pr(S_n > x) = \overline{F^{n*}}(x) \sim n\overline{F}(x) \textit{ for } x \to \infty \ , \quad \textit{for all } n \geq 1.$$

($F^{n*} = F * \ldots * F = \mathcal{L}(S_n)$, n-fold convolution)

Let $M_n = \max(X_1, \ldots, X_n)$. Then, $\mathrm{pr}(M_n \leq x) = \mathrm{pr}(X_1 \leq x, \ldots, X_n \leq x) = F^n(x)$, and the tail distribution of M_n satisfies:

$$\overline{F^n}(x) = \mathrm{pr}(M_n > x) = 1 - F^n(x) = \overline{F}(x) \sum_{k=0}^{n-1} F^k(x) \sim \overline{F}(x) \cdot n$$

for $x \to \infty$ and all $n \geq 1$.

Corollary 1.5 *Let $X_1, X_2, \ldots$ i.i.d. with tail distribution $\overline{F}(x) \in \mathcal{R}_{-\alpha}(x)$ for some $\alpha > 0$. Then, $pr(S_n > x) \sim pr(M_n > x)$ for $x \to \infty$, for all $n \geq 1$.*

Let $\psi(u)$ be the ruin probability, $\delta(u) = 1 - \psi(u)$ as in section 1.3. Using the integrated tail distribution, Theorem 1.1 can be reformulated as

a) $\delta(y) - \delta(0) = \frac{1}{1+\rho} \int_0^y \delta(y - x) dF_I(x)$

b) $\delta(0) = \frac{\rho}{1+\rho}$

Corollary 1.6 *Under the assumptions of Theorem 1.1:*

$$\delta(y) = \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} \frac{1}{(1+\rho)^n} F_I^{n*}(y), \quad y \geq 0.$$

Interpretation: If $Z_1, Z_2, \ldots$ i.i.d. with distribution F_I, independent of random N which is geometrically distributed with parameter $q = \frac{1}{1+\rho}$, then

$$\delta(y) = \text{pr}\left(\sum_{j=1}^{N} Z_j \leq y\right).$$

Corollary 1.6 is, therefore, called a geometric representation of the non-ruin probability.

Recall: N geometrically distributed if $\text{pr}(N = n) = (1 - q)q^n,\ \ n = 0, 1, 2, \ldots$

Analogously, we can derive a geometric representation of the ruin probability

$$\psi(y) = \text{pr}\left(\sum_{j=1}^{N} Z_j > y\right) = \frac{\rho}{1+\rho}\sum_{n=0}^{\infty}\frac{1}{(1+\rho)^n}\overline{F_I^{n*}}(y)$$

Therefore, if $\overline{F}_I \in \mathcal{R}_{-\alpha}$ for some $\alpha > 0$, we have from Corollary 1.4:

$$\frac{\psi(u)}{\overline{F}_I(u)} \underset{u\to\infty}{\sim} \frac{\rho}{1+\rho}\sum_{n=0}^{\infty}\frac{1}{(1+\rho)^n}\cdot n = \frac{1}{\rho}$$

provided that $\sum_{n=0}^{\infty}$ and $\lim_u \to \infty$ may be interchanged (which is the case - compare Theorem 1.4). As a consequence, we get that, for large initial reserves u, the ruin probability is essentially determined by the tail of the claim size distribution:

$$\psi(u) \underset{u\to\infty}{\sim} \frac{1}{\rho}\overline{F}_I(u) = \frac{1}{\rho\mu}\int_u^{\infty}\overline{F}(y)dy$$

Definition: A distribution G with support $(0, \infty)$ is subexponential if $\lim_{x\to\infty}\frac{\overline{G^{n*}}(x)}{\overline{G}(x)} = n$ for all $n \geq 1$. Notation: $G \in \mathcal{S}$.

Remark: If $\overline{G} \in \mathcal{R}_{-\alpha}$, $\alpha > 0$, then $G \in \mathcal{S}$ (by Corollary 1.4). Corollary 1.5 holds for $F \in \mathcal{S}$, too.

Lemma 1.5 *$G \in \mathcal{S}$ iff* $\limsup_{n\to\infty}\frac{\overline{G^{2*}}(x)}{\overline{G}(x)} \leq 2$.

Lemma 1.6 *a) If $G \in \mathcal{S}$, then $\lim_{x\to\infty}\frac{\overline{G}(x-y)}{\overline{G}(x)} = 1$ uniformly in $y \in C, C$ compact*

b) If assertion of a) holds, then $\lim_{x\to\infty} e^{\varepsilon x}\overline{G}(x) = \infty$ for all $\varepsilon > 0$
(explanation of term "subexponential")

c) If $G \in \mathcal{S}$, then for any $\varepsilon > 0$ there is a $K > 0$ such that

$$\frac{\overline{G^{n*}}(x)}{\overline{G}(x)} \leq K(1+\varepsilon)^n \quad \textit{for all } n \geq 1,\ x \geq 0.$$

Theorem 1.5 *(Cramér-Lundberg Theorem for large claims): Let $\{\mathcal{R}_t\}$ be a Cramér-Lundberg process with net profit condition $\rho = \frac{\pi}{\lambda\mu} - 1 > 0$ and $F_I \in \mathcal{S}$. Then, the ruin probability satisfies*

$$\psi(u) \sim \frac{1}{\rho}\overline{F}_I(u) = \frac{1}{\mu\rho}\int_u^\infty \overline{F}(y)dy \quad \textit{for } u \to \infty.$$

Consider now a general renewal model for which $N(t) = \#\{j;\ T_j \leq t\}$, the number of claims up to time t, has a general distribution given by

$$\text{pr}(N(t) = n) = p_t(n),\ n \geq 0.$$

Let $G_t(x) = \text{pr}(S(t) \leq x)$ be the distribution function of the total claim amount $S(t) = \sum_{j=1}^{N(t)} X_j$ up to time t, which, using independence, is

$$G_t(x) = \sum_{n=0}^{\infty} \text{pr}(N(t) = n,\ S_n \leq x) = \sum_{n=0}^{\infty} p_t(n)F^{n*}(x)$$

Let R_t be a renewal process with $F = \mathcal{L}(X_j) \in \mathcal{S}$, and, for fixed $t > 0:$
$\sum_{n=0}^{\infty} p_t(n)(1+\varepsilon)^n < \infty$ for some $\varepsilon > 0$.
Then, $G_t \in \mathcal{S}$, and $\overline{G}_t(x) = \text{pr}(S(t) > x) \sim \overline{F}(x) \cdot \mathcal{E}\mathcal{N}(\sqcup)$ for $x \to \infty$.

In the Cramér-Lundberg case: $\mathcal{L}(N(t)) = \mathcal{P}(\lambda t)$ and $\overline{G}_t(x) \sim \lambda t\ \overline{F}(x)$.

Example 1.1 $p_t(n) = \binom{\gamma+n-1}{n}\left(\frac{\beta}{\beta+t}\right)^\gamma\left(\frac{t}{\beta+t}\right)^n,\ n \geq 0$
i.e. $N(t)$ a negative binomial process. The above condition on $p_t(n)$ is satisfied and $\mathcal{E}N(t) = \frac{\gamma^t}{\beta}$ such that $\overline{G}_t(x) \sim \frac{\gamma^t}{\beta}\overline{F}(x)$. Apart from the Cramér-Lundberg assumption of a homogeneous Poisson process, this is the most frequent model for claim number distribution in insurance practice. It is appropriate in the case of overdispersion, i.e. where $varN(t) > \mathcal{E}N(t)$. In the negative binomial case: $varN(t) = \mathcal{E}\mathcal{N}(\sqcup) \cdot (\infty + \frac{\sqcup}{\beta})$. For the Poisson process we have in contrast $\mathcal{E}\mathcal{N}(\sqcup) = var\mathcal{N}(\sqcup)$. One cause for overdispersion is intersubject variability where the intensity of claims varies between subjects, i.e. one observes a mixed Poisson process:

Definition 1.5 *Let Λ be a positive random variable, independent of a homogeneous Poisson process $N_1(t)$ with intensity 1. The process $N(t) = N_1(\Lambda \cdot t)$ is a mixed Poisson process.*

If Λ is Gamma-distributed, i.e. its density is $f_\Lambda(y) = \frac{\beta^\gamma}{\Gamma(\gamma)}y^{\gamma-1}e^{-\beta y},\ y > 0$, then $N(t)$ is a negative binomial process.
If $\text{pr}(\Lambda = \lambda) = 1,\ N(t)$ is a homogeneous Poisson process with intensity λ.

$F \in \mathcal{S}$ does in general not imply $F_I \in \mathcal{S}$ and vice versa. Therefore, conditions for $F_I \in \mathcal{S}$ are of interest. Some of them use the hazard rate $q(x) = f(x)/\overline{F}(x)$, known from survival analysis. If τ is a survival time with distribution F and density f, then

$$q(x)dx = \frac{f(x)dx}{\overline{F}(x)} \approx \frac{1}{\overline{F}(x)}\int_x^{x+dx} f(t)dt = \text{pr}(\tau \leq x + dx/\tau > x)$$

Lemma 1.7 *a) If* $\limsup_{x\to\infty} xq(x) < \infty$, *then* $F_I \in \mathcal{S}$.

b) $F_I \in \mathcal{S}$ *if* $\lim_{x\to\infty} xq(x) = \infty,\ \lim_{x\to\infty} q(x) = 0$ *and one of the following conditions holds:*
(i) $\limsup_{x\to\infty} \frac{xq(x)}{-\log \overline{F}(x)} < 1$

(ii) $q \in \mathcal{R}_{-\delta}$ *for some* $0 < \delta \leq 1$

As a consequence, for the following list of heavy-tailed distributions we have $F \in \mathcal{S}$ and $F_I \in \mathcal{S}$, provided the first moment exists:
Pareto, Weibull ($\tau < 1$), lognormal, Benktander type I and II, Burr, loggamma.

2 Fluctuations of sums and maxima

Throughout this chapter: $X_1, X_2, \ldots, X$ i.i.d. with distribution $\mathcal{L}(X_j) = F$.

$$S_n = X_1 + \ldots + X_n,\ \overline{X}_n = \frac{1}{n} S_n,\ M_n = \max(X_1, \ldots, X_n)$$

2.1 Limit behaviour of sums

Classical central limit theorem (CLT):

$\mathrm{var} X_j = \sigma^2 < \infty,\ \mathcal{E}\mathcal{X}_| = \mu \Longrightarrow$

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(\prime, \infty)$$

Definition: A distribution G is stable if for i.i.d. Z_0, Z_1, Z_2 with law G and for all c_1, c_2, there are $b > 0$ and a such that

$$\mathcal{L}(\rfloor_\infty \mathcal{Z}_\infty + \rfloor_\in \mathcal{Z}_\in) = \mathcal{L}(\dashv \mathcal{Z}_\prime + \lfloor).$$

As a consequence, if F is stable there are normalizing sequences $b_n > 0, a_n$, such that

$$\mathcal{L}\left(\frac{\mathcal{S}_\backslash - \dashv_\backslash}{\lfloor_\backslash}\right) = \mathcal{L}(\mathcal{X}_\infty),\ \ \backslash \geq \infty.$$

Theorem 2.1 *If for some* $b_n > 0, a_n$:

$$\frac{S_n - a_n}{b_n} \underset{\mathcal{L}}{\longrightarrow} Z$$

with nondegenerate $\mathcal{L}(\mathcal{Z}) = \mathcal{G}$, *then* G *is stable.*

Theorem 2.2 *(Characterization of stable laws:)* $G = \mathcal{L}(\mathcal{Z})$ *is stable if there are* $\gamma \in \mathbb{R},\ c > 0, \alpha \in (0, 2],\ \beta \in [-1, +1]$ *such that the characteristic function of* Z *is*

$$\chi_Z(u) = \mathcal{E}\rceil^{\rangle \sqcap \mathcal{Z}} = \exp\{\rangle \gamma \sqcap - \rfloor |\sqcap|^\alpha (\infty - \rangle \beta\, sgn(\sqcap) \xi(\sqcap, \alpha))\}$$

with $\xi(u, \alpha) = \tan \frac{\pi\alpha}{2}$ *for* $\alpha \neq 1$ *and* $= -\frac{2}{\pi} \log |u|$ *for* $\alpha = 1$.

If the location parameter γ and the skewness parameter β are both 0,then G is symmetric around 0, and we talk of a symmetric α-stable (sαs) distribution. In this case: $\chi_Z(u) = e^{-c|u|^\alpha}$.

If $\alpha = 2, \gamma = 0$, then $G \equiv \mathcal{N}(\prime, \sigma^{\in})$ with $\sigma^2 = 2c$. In general, c is a dispersion parameter. α is called the characteristic exponent and determines the tail behaviour. For $\alpha = 1$, the sαs-distribution is the Cauchy distribution.

Remark: If G is α-stable, $\mathcal{L}(\mathcal{Z}) = \mathcal{G}$, then, for $\alpha < 2$,

$$\mathcal{E}|\mathcal{Z}|^\delta < \infty \quad \text{iff } \delta < \alpha.$$

Definition: The random variable X with $\mathcal{L}(\mathcal{X}) = \mathcal{F}$ belongs to the domain of attraction DA(G_α) if for suitable $b_n > 0, a_n$:

$$\frac{S_n - a_n}{b_n} \underset{\mathcal{L}}{\longrightarrow} G_\alpha.$$

$X \in$ DA(α) if $X \in$ DA(G_α) for some α-stable law G_α.

Theorem 2.3 *a) $X \in DA(2)$, i.e. the domain of attraction of the normal law iff $V(x) = \int_{-x}^{x} y^2 dF(x) = \mathcal{E}\mathcal{X}^{\in} \cdot \infty_{[-\S,\S]}(\mathcal{X})$ is a slowly varying function.*

b) $X \in DA(\alpha)$ for some $\alpha < 2$ iff there are a slowly varying function $L(x)$ and constants $c_1, c_2 \geq 0, \; c_1 + c_2 > 0$ such that

$$F(x) \sim \frac{c_1 L(x)}{|x|^\alpha} \; \textit{for}\, x \to -\infty \;\; \textit{and} \;\; \overline{F}(x) \sim \frac{c_2 L(x)}{|x|^\alpha} \; \textit{for } x \to +\infty.$$

If $\mathcal{E}\mathcal{X}^{\in} < \infty$, then $V(x) \longrightarrow \mathcal{E}\mathcal{X}^{\in}$ and, therefore, $X \in DA(2)$. Otherwise, $V(x)$ slowly varying (i.e. $\in \mathcal{R}_0$) iff the following tail condition is satisfied:

$$\text{pr}(|X| > z) = O(\frac{1}{z^2} V(z)).$$

Analogously, if $X \in DA(\alpha)$ for some $\alpha < 2$:

$$\text{pr}(|X| > z) = \frac{L(z)}{z^\alpha} \quad \text{for some} \quad L \in \mathcal{R}_0 \; \text{ and}$$

$$\frac{z^2 \text{pr}(|X| > z)}{V(z)} \longrightarrow \frac{2 - \alpha}{\alpha} \text{ for } z \to \infty.$$

Corollary 2.1 *If $X \in DA(\alpha)$, then:*
$\mathcal{E}|\mathcal{X}|^\delta < \infty$ for $\delta < \alpha$
$\mathcal{E}|\mathcal{X}|^\delta = \infty$ for $\delta > \alpha, \; \alpha < 2$.

Proposition 2.1 *a) Let* $Q(z) = pr(|X| > z) + \frac{1}{z^2}V(z)$.
Then, b_n *can be chosen as solution of* $Q(b_n) = \frac{1}{n}$, $n \geq 1$. *In particular,* $b_n \sim \sigma\sqrt{n}$ *for* $\sigma^2 = varX < \infty$ *and* $\mathcal{E}\mathcal{X} = \prime$, *and* $b_n = n^{1/\alpha}L(n)$ *for some* $L(n) \in \mathcal{R}_0$ *if* $\alpha < 2$.

b) a_n *can be chosen as*

$$a_n = \begin{cases} \mu \cdot n & \text{if } 1 < \alpha \leq 2,\ \mu = \mathcal{E}\mathcal{X} \\ 0 & \text{if } 0 < \alpha < 1 \\ 0 & \text{if } \alpha = 1 \text{ and } F \text{ symmetric} \end{cases}$$

Theorem 2.4 *(General CLT for i.i.d. random variables): Let* $X \in DA(\alpha)$ *for some* $0 < \alpha \leq 2$. *Then,*

a) If $\mathcal{E}\mathcal{X}^{\in} < \infty$,

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(\prime, \infty)$$

a) If $\mathcal{E}\mathcal{X}^{\in} = \infty$, *then*

$$\frac{S_n - a_n}{n^{1/\alpha}L(n)} \underset{\mathcal{L}}{\longrightarrow} G_\alpha$$

for an α*-stable law* G_α *where* $L \in \mathcal{R}_0$ *and* a_n *as in Proposition 2.1.*

Corollary 2.2 *Theorem 2.4 b) holds for* $L(n) \equiv c$ *iff* $\alpha < 2$ *and the condition of Theorem 2.3 holds with* $L(x) \equiv c$.

Let $0 \leq T_1 \leq T_2 \leq \ldots$ be an increasing sequence of random variables,

$$N(t) = \max\{k;\ T_k \leq t\}, \quad S(t) = \sum_{j=1}^{N(t)} X_j.$$

Let $T_1, T_2, \ldots$ be independent of $X_1, X_2, \ldots$

Theorem 2.5 *(Anscombe's CLT): Suppose* $X \in DA(\alpha)$ *for some* $0 < \alpha \leq 2$ *and*

$$\frac{N(t)}{t} \underset{p}{\longrightarrow} \lambda > 0.$$

Then, Theorem 2.4 holds with $N(t)$ *replacing* n *and* $S(t)$ *replacing* S_n. *Moreover,*

$$\frac{S(t) - a_{N(t)}}{(\lambda t)^{1/\alpha}L(t)} \underset{\mathcal{L}}{\longrightarrow} G_\alpha$$

i.e. $N(t)$ *may be replaced by its asymptotic approximation in the denominator.*

2.2 Limit behaviour of maxima

The distribution function of $M_n = \max(X_1, \dots, X_n)$ is given by

$$\mathrm{pr}(M_n \le x) = \mathrm{pr}(X_1 \le x, \dots, X_n \le x) = F^n(x).$$

Let $x_F = \sup\{x \in \mathbb{R};\ F(x) < 1\} \le \infty$. Then, for $n \to \infty$,

$$F^n(x) \longrightarrow 0 \text{ for } x < x_F,\ \ F^n(x) \to 1 \quad \text{for } x > x_F, \text{ i.e.}$$

$M_n \underset{p}{\longrightarrow} x_F$ (also: $M_n \underset{a.s.}{\longrightarrow} x_F$ due to monotonicity). To get an interesting limit behaviour we have to standardize M_n.

Definition 2.1 *The random variable X belongs to the maximum domain of attraction MDA(H) of a nondegenerate law H if for suitable $c_n > 0, d_n$:*

$$\frac{M_n - d_n}{c_n} \underset{\mathcal{L}}{\longrightarrow} H$$

i.e. $F^n(c_n x + d_n) \longrightarrow H(x)$ for all points x of continuity of $H(x)$.

Definition 2.2 *Extreme value distributions with distribution functions*

a) Fréchet: $\Phi_\alpha(x) = \exp\{-1/x^\alpha\},\ x \ge 0$, *for some* $\alpha > 0$
b) Gumbel: $\Lambda(x) = \exp\{-e^{-x}\},\ x \in \mathbb{R}$
c) Weibull: $\Psi_\alpha(x) = \exp\{-|x|^\alpha\},\ x \le 0$, *for some* $\alpha > 0$
The Fréchet distributions are supported on $[0, \infty)$, *the Weibull distributions on* $(-\infty, 0]$.

Definition 2.3 *The generalized extreme value distribution (GEV) with shape parameter $\xi \in \mathbb{R}$ has the distribution function:*

$H_\xi(x) = \exp\{-(1 + \xi x)^{-1/\xi}\},\ 1 + \xi x > 0$ *for* $\xi \ne 0$
$H_0(x) = \Lambda(x)$
This notion just combines the three different distributions of the previous definition:

$$H_\xi\Big(\frac{x-1}{\xi}\Big) = \Phi_{1/\xi}(x) \text{ for } \xi > 0, \qquad H_\xi\Big(-\frac{x+1}{\xi}\Big) = \Psi_{-1/\xi}(x) \text{ for } \xi < 0.$$

The definition describes the standard forms. In general, we may apply shifts and scale transformations to get other GEV-laws: $H(x) = H_\xi(\frac{x-\mu}{\sigma})$ for some $\xi, \mu \in \mathbb{R},\ \sigma > 0$. In the asymptotic theory this does not matter as the standardizing sequences can always be chosen such that the limit is in standard form ($\mu = 0, \sigma = 1$).

Theorem 2.6 *(Fisher-Tippett): If there are $c_n > 0, d_n$ and a non-degenerate H such that*

$$\frac{M_n - d_n}{c_n} \underset{\mathcal{L}}{\longrightarrow} H,$$

then H is a GEV distribution.

Lemma 2.1 *(Convergence of types theorem): Let $U_1, U_2, \ldots, V, W$ be random variables, $b_n, \beta_n > 0,\ a_n, \alpha_n \in \mathbb{R}$. Suppose:*

$$\frac{U_n - a_n}{b_n} \xrightarrow[\mathcal{L}]{} V.$$

Then:

$$\frac{U_n - \alpha_n}{\beta_n} \xrightarrow[\mathcal{L}]{} W \qquad \textit{iff} \qquad \frac{b_n}{\beta_n} \longrightarrow b \geq 0,\ \frac{a_n - \alpha_n}{\beta_n} \longrightarrow a \in \mathbb{R}.$$

In this case: $\mathcal{L}(W) = \mathcal{L}(bV + a)$.

Sketch of proof of Theorem 2.6: $t > 0, [\,] =$ integer part, $\frac{M_n - d_n}{c_n} \xrightarrow[\mathcal{L}]{} H$. As $F^{[nt]}$ is distribution function of $M_{[nt]}$,

$$F^{[nt]}(c_{[nt]}x + d_{[nt]}) \longrightarrow H(x) \text{ for } [nt] \to \infty, \quad \text{i.e.} \quad n \to \infty.$$

On the other hand,

$$F^{[nt]}(c_n x + d_n) = (F^n(c_n x + d_n))^{\frac{[nt]}{n}} \longrightarrow H^t(x) \text{ for } n \to \infty.$$

Therefore:

$$\frac{M_{[nt]} - d_{[nt]}}{c_{[nt]}} \xrightarrow[\mathcal{L}]{} H\ ,\ \ \frac{M_{[nt]} - d_n}{c_n} \xrightarrow[\mathcal{L}]{} H^t.$$

By Lemma 2.1:

$$\frac{c_n}{c_{[nt]}} \longrightarrow \gamma(t) \geq 0,\ \frac{d_n - d_{[nt]}}{c_{[nt]}} \longrightarrow \delta(t),$$

and

$$H^t(x) = H(\gamma(t)x + \delta(t)),\ t > 0,\ x \in \mathbb{R}. \tag{A}$$

Applying that argument to t, s and $s \cdot t$ implies

$$\gamma(st) = \gamma(s)\ \gamma(t),\ \delta(st) = \gamma(t)\delta(s) + \delta(t). \tag{B}$$

Solving the functional equations $(A), (B)$ for $H(x), \gamma(t), \delta(t)$ implies $H \in \{\Lambda, \Phi_\alpha, \Psi_\alpha\}$, i.e. GEV.

Remark The GEV - laws coincide with the class of max-stable laws $F = \mathcal{L}(X)$ defined by the existence of $c_n > 0, d_n, n \geq 1$ such that

$$\mathcal{L}(M_n) = \mathcal{L}(c_n X + d_n),\ n \geq 1.$$

Definition 2.4 *F distribution function. The generalized inverse F^- of F:*

$$F^-(t) = \inf\{x \in \mathbb{R};\ F(x) \geq t\},\ \ 0 < t < 1$$

is called the quantile function. $F^-(t)$ is the t-quantile.

Theorem 2.7 *(Characterization of MDA of Fréchet and Weibull-laws):*

a) $F \in MDA(\Phi_\alpha)$ *for some* $\alpha > 0$ *iff* $\overline{F}(x) = \frac{1}{x^\alpha} L(x)$ *for some* $L \in \mathcal{R}_0$. *In this case,* $x_F = \infty$ *and*

$$\frac{M_n}{c_n} \underset{\mathcal{L}}{\longrightarrow} \Phi_\alpha$$

with $c_n = F^{-}(1 - \frac{1}{n})$.

b) $F \in MDA(\Psi_\alpha)$ *for some* $\alpha > 0$ *iff* $x_F < \infty$ *and* $\overline{F}(x_F - \frac{1}{x}) = \frac{1}{x^\alpha} L(x)$ *for some* $L \in R_0$. *In this case,*

$$\frac{M_n - x_F}{c_n} \underset{\mathcal{L}}{\longrightarrow} \Psi_\alpha$$

with $c_n = x_F - F^{-}(1 - \frac{1}{n})$.

Proposition 2.2 *Let* $0 \le \tau \le \infty,\ u_n \in \mathbb{R}$. *Then, for* $n \to \infty$,

$$n\overline{F}(u_n) \longrightarrow \tau \quad \textit{iff} \quad pr(M_n \le u_n) \longrightarrow e^{-\tau}.$$

Corollary 2.3 $F \in MDA(H)$ *with normalizing sequences* c_n, d_n *iff*

$$n\overline{F}(c_n x + d_n) \longrightarrow -\log H(x) \quad \textit{for all } x \in \mathbb{R},\ n \to \infty.$$

Remark Part b) of the Theorem follows rather immediately from part a) by exploiting the relation $\Psi_\alpha(-\frac{1}{x}) = \Phi_\alpha(x),\ x > 0$. Mark also that $\overline{\Phi}_\alpha(x) \sim \frac{1}{x^\alpha}, \alpha > 0$.

Theorem 2.8 *(MDA of the Gumbel law): Let* $x_F \le \infty$. $F \in MDA(\Lambda)$ *iff there exists* $z < x_F$, *measurable scaling functions* $c(x) \longrightarrow c > 0,\ g(x) \longrightarrow 1$ *for* $x \uparrow x_F$ *and an absolutely continuous function* $e(x) > 0$ *with* $e'(x) \longrightarrow 0$ *for* $x \uparrow x_F$ *such that*

$$\overline{F}(x) = c(x) \exp\{-\int_z^x \frac{g(y)}{e(y)} dy\},\ z < x < x_F.$$

In this case,

$$\frac{M_n - d_n}{c_n} \underset{\mathcal{L}}{\longrightarrow} \Lambda$$

with $d_n = F^{-}(1 - \frac{1}{n})$ *and* $c_n = e(d_n)$. *The function* $e(x)$ *may be chosen as the mean excess function*

$$e(x) = \int_x^{x_F} \frac{\overline{F}(y)}{\overline{F}(x)} dy,\ \ x < x_F.$$

Lemma 2.2 F *satisfies the conditions of Theorem 2.8 if* $F \in C^2$ *on* (z, x_F) *with density* $f > 0,\ f' < 0$ *on* (z, x_F), *i.e.* F *is strictly concave on* (z, x_F), *and* $\{\overline{F}(x) f'(x)\}/f^2(x) \to -1$ *for* $x \uparrow x_F$. *Then,* $e(x)$ *may be chosen simply as* $e(x) = \overline{F}(x)/f(x)$.

Examples:

a) The Pareto-, Burr- and α-stable laws ($\alpha < 2$) all have *Pareto-tails*, i.e.

$$\overline{F}(x) \sim \frac{K}{x^\alpha} \quad \text{for} \quad x \to \infty,$$

and, therefore, are in MDA(Φ_α). As $\overline{F}^{-1}(t) \sim (\frac{K}{t})^{1/\alpha},\ c_n \sim (Kn)^{1/\alpha}$ and:

$$\frac{M_n}{(Kn)^{1/\alpha}} \underset{\mathcal{L}}{\longrightarrow} \Phi_\alpha \quad \text{for } n \to \infty.$$

b) For the uniform distribution $U(0,1)$, we have $x_F = 1, \overline{F}(x) = 1 - x,\ 0 \le x \le 1$, and $\overline{F}(1 - \frac{1}{x}) = \frac{1}{x}$. Therefore, $U(0,1) \in MDA(\Psi_1), c_n = \frac{1}{n}$ and:

$$n(M_n - 1) \underset{\mathcal{L}}{\longrightarrow} \Psi_1 \quad \text{for} \quad n \to \infty.$$

c) For Exp(λ), we have $\overline{F}(x) = e^{-\lambda x}$, satisfying the assumption of Theorem 2.8 with $c(x) = 1,\ g(x) = 1,\ z = 0$ and $e(x) = \frac{1}{\lambda}$. Therefore, Exp(λ) $\in MDA(\Lambda), c_n = \frac{1}{\lambda}, d_n = \frac{1}{\lambda}\log n$ and:

$$\lambda(M_n - \frac{1}{\lambda}\log n) \underset{\mathcal{L}}{\longrightarrow} \Lambda \quad \text{for} \quad n \to \infty.$$

d) Let Φ, φ denote distribution function and density of $\mathcal{N}(0,1)$. We have $\overline{\Phi}(x) \sim \frac{1}{x}\varphi(x)$ for $x \to \infty$ (Mill's ratio), from which we get the conditions of Lemma 2.2. Therefore, $\mathcal{N}(0,1) \in MDA(\Lambda)$, and some asymptotic approximations show

$$\sqrt{2\log n}(M_n - d_n) \underset{\mathcal{L}}{\longrightarrow} \Lambda \quad \text{for} \quad n \to \infty$$

with $d_n = \sqrt{2\log n} - \{\log\log n + \log(4\pi)\}/\{2\sqrt{2\log n}\}$.

Definition 2.5 *Let $u < x_F$ be a given threshold.*

a) $F_u(x) = pr\{X - u \le x/X > u\} = (F(u + x) - F(u))/\overline{F}(u),\ 0 \le x < x_F - u$ is called the excess distribution function above the threshold u.

b) $e(u) = \mathcal{E}\{X - u/X > u\},\ u < x_F$, is the mean excess function.

Remarks

i) Integration by parts implies (compare Theorem 2.8):

$$e(u) = \int_u^{x_F} \frac{\overline{F}(y)}{\overline{F}(u)} dy.$$

ii) If Δ_u is a random variable with distribution F_u, then $\mathcal{E}\Delta_u = e(u)$.

Proposition 2.3 *Let X be a positive random variable with density f.*

a) The mean excess function characterizes F uniquely:

$$\overline{F}(x) = \frac{e(0)}{e(x)} \exp\{-\int_0^x \frac{1}{e(u)} du\}, \quad x > 0.$$

b) If $\overline{F}(x) = \frac{1}{x^\alpha} L(x)$ for some $L \in \mathcal{R}_0, \ \alpha > 1$, then $e(u) \sim \frac{u}{\alpha - 1}$ for $u \to \infty$.

Definition 2.6 *The generalized Pareto distribution (GPD) with parameters $\beta > 0, \ \xi$ has the distribution function*

$$G_{\xi,\beta}(x) = 1 - (1 + \frac{\xi x}{\beta})^{1/\xi} \quad for \quad \begin{cases} x \geq 0 & if \ \ \xi > 0 \\ 0 \leq x \leq \frac{-\beta}{\xi} & if \ \ \xi < 0, \end{cases}$$

and

$$G_{0,\beta}(x) = 1 - e^{-\frac{1}{\beta}x}, \ x \geq 0.$$

$G_\xi(x) \equiv G_{\xi,1}(x)$ are called <u>standard</u> GPDs.

For $\xi = 0$, the GPD is $\mathrm{Exp}(\frac{1}{\beta})$, for $\xi > 0$ it is a Pareto distribution reparametrized (in particular: $\xi = \frac{1}{\alpha}$). For $\xi < 0$, this type of law is called a Pareto distribution of type II.

Theorem 2.9 *$F \in MDA(H_\xi)$ for a GEV H_ξ with shape parameter ξ iff there exists a (measurable) function $\beta(x) > 0$ such that*

$$\sup_{0 \leq x \leq x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| \longrightarrow 0 \, for \ \ u \uparrow x_F,$$

where $G_{\xi,\beta}$ denotes a GPD.

Corollary 2.4 *If $F \in MDA(H_\xi)$, then for an appropriate scaling function $\beta(u)$:*

$$pr\{\frac{X - u}{\beta(u)} > x | X > u\} \longrightarrow \overline{G}_\xi(x) \quad for \quad u \uparrow x_F.$$

(Approximation of the law of scaled excesses).

Examples: a) By the memorylessness of $\mathrm{Exp}(\lambda)$, we have $F_u(x) = F(x)$ for all $u > 0$, and, therefore, $F_u \equiv G_{0,\beta}$ with $\beta(u) \equiv \beta = \frac{1}{\lambda}$.

b) Stability of GPDs under truncation: If $F = G_{\xi,\beta}$, then for all $u > 0$

$$F_u(x) \equiv G_{\xi,\beta+\xi u}(x) \quad \text{for} \quad \begin{cases} x \geq 0 & \text{if} \ \ \xi \geq 0 \\ 0 \leq x < -\frac{\beta}{\xi} - u & \text{if} \ \ \xi < 0, \end{cases}$$

i.e. $\beta(u) = \beta + \xi \, u$ in this case.

2.3 Exploiting Fisher-Tippett in practice

To estimate characteristics of the tail behaviour, one may use for large n that, by Theorem 2.6, $\mathcal{L}(M_n)$ is approximately GEV, i.e. there are a shape parameter ξ, a location parameter μ and a scale parameter σ (the latter two depending on n) such that

$$\text{pr}(M_n \leq y) \approx H_\xi(\frac{y-\mu}{\sigma}).$$

To estimate $\vartheta = (\xi, \mu, \sigma)^T$, we need a whole sample of maxima $M_n^{(1)}, \ldots, M_n^{(b)}$. For that purpose, a very large sample of size $N = b \cdot n$ is partitioned into b blocks of size n each, and $M_n^{(k)}$ is chosen as the maximum of the observations in the k-th block, $k = 1, \ldots, b$.

Assumption: $M_n^{(1)}, \ldots, M_n^{(b)}$ are i.i.d. with distribution function $H_\xi(\frac{y-\mu}{\sigma})$. Let $h_\vartheta(y)$ denote the density of $H_\xi(\frac{y-\mu}{\sigma})$ with support

$$D_\vartheta = \{y; 1 + \xi \, \frac{y-\mu}{\sigma} > 0\}.$$

We get the ML-estimates of $\vartheta : \hat{\vartheta} = (\hat{\xi}, \hat{\mu}, \hat{\sigma})^T$ by maximizing the log-likelihood (stressing the dependence of the support D_ϑ of h_ϑ on ϑ):

$$\ell(\vartheta / M_n^{(1)}, \ldots, M_n^{b)}) = \sum_{k=1}^{b} \log\{h_\vartheta(M_n^{(k)}) \cdot 1_{D_\vartheta}(M_n^{(k)})\} = \max!$$

Though D_ϑ depends of ϑ, contradicting one of the usual assumptions of asymptotics for ML-estimates in standard situations, it can be shown that, for $\xi > -\frac{1}{2}$, the same asymptotic results as usual hold, e.g. asymptotic normality of $\hat{\vartheta}$ with rate $\frac{1}{\sqrt{N}}$ and asymptotic efficiency (compare also Proposition 3.1).

As usual, one may calculate asymptotic confidence regions for ϑ, construct LR-tests for, e.g., the size of ξ and so on. The same applies for functions of ϑ like:

Definition: Let $M_n = \max\{X_1, \ldots, X_n\}$, $X_1, \ldots, X_n$ i.i.d. The ℓ n-block return level $R_{n,\ell}$ (e.g., for daily data $n = 365, \ell = 10$, the 10-year return level) is the $(1 - \frac{1}{\ell})$-quantile of M_n, i.e.

$$\text{pr}(M_n > R_{n,\ell}) = \frac{1}{\ell}.$$

We are talking of a 10-year (resp. a centennial) event if in year no. k we have $M_n^{(k)} > R_{365,10}$ (resp. $> R_{365,100}$).

Let $F = \mathcal{L}(X_j)$ and, therefore, $\mathcal{L}(M_n) = F^n$ and $F^n(R_{n,\ell}) = 1 - \frac{1}{\ell}$. Therefore, $R_{n,\ell}$ is the $(1 - \frac{1}{\ell})^{\frac{1}{n}}$-quantile of F. In applications, $(1 - \frac{1}{\ell})^{\frac{1}{n}} \approx 1$ such that the empirical quantile of the X_j is based on too few observations and highly variable. For estimating $R_{n,\ell}$, one uses again Fisher-Tippett from which we have, using the notation $H_\vartheta(y) = H_\xi(\frac{y-\mu}{\sigma})$:

$$R_{n,\ell} \approx H_\vartheta^{-1}(1 - \frac{1}{\ell}) \approx \mu + \frac{\sigma}{\xi}\{-\log(1 - \frac{1}{\ell})\}^{-\xi}.$$

Replacing ξ, μ, σ by their ML-estimates $\hat{\xi}, \hat{\mu}, \hat{\sigma}$, we get the ML-estimate $\hat{R}_{n,\ell}$ of $R_{n,\ell}$.

A critical point in using that type of extreme value statistics is the choice of block length n and number of blocks b given the total sample size N. If b is too small, the sample size for the ML-procedure is small, and the variance of $\hat{\vartheta}$ or $\hat{R}_{n,\ell}$ is large. If n is too small, the approximation of $\mathcal{L}(M_n)$ by a GEV is not a good one which causes a bias in estimates like $\hat{R}_{n,\ell}$ to appear. As $N = n \cdot b$, we have one of the common bias-variance dilemmas.

Example: The annual maxima of the water level of River Nidd (England) are given from 1936-1970, i.e. $n = 365,\ b = 35,\ N = 12\ 775$.

1) What is the probability that next year's maximum exceeds all previous annual maxima? We use the estimation procedure above and get

$$1 - H_{\hat{\vartheta}}(\max\{M_{365}^{(k)},\ k = 1, \ldots, 35\}) \approx 0.04$$

2) What is the 10-year return level $R_{365,10}$? From the data, we get $\hat{R}_{365,10} \approx 222$. The plot shows this level together with a confidence interval for $R_{365,10}$.

3 Statistics for Extremal Events

Literatur: EKM, chapter 6

Throughout this chapter, $X, X_1, \ldots, X_N$ are i.i.d. random variables with $\mathcal{L}(X_j) = F$ and $x_F = \infty$ (unbounded support on the right-hand side).

Notation: $X_{(1)} \leq \ldots \leq X_{(N)}$ and $X^{(1)} \geq \ldots \geq X^{(N)}$ denote the order statistics, i.e. the ordered data, in ascending and descending order. Of course, $X_{(1)} = X^{(N)},\ X_{(N)} = X^{(1)}$ etc.

Definition: Let $K_N(u) = \{j \leq N;\ X_j > u\}$ and $\hat{F}_N(x) = \frac{1}{N} \sum_{j=1}^N 1_{(-\infty,x]}(X_j)$ be the empirical distribution function, $N_u = \# K_N(u)$.

$$e_N(u) = \int_u^\infty \overline{\hat{F}_N}(y)dy / \overline{\hat{F}_N}(u) = \frac{1}{N_u} \sum_{j \in K_N(u)} (X_j - u) = \frac{1}{N_u} \sum_{j=1}^N (X_j - u)^+$$

is called the empirical mean excess function.

$e_N(u)$ approximates the mean excess function $e(u)$ of section 2.2.

For exploratory data analysis, one frequently considers the following plots:

$$\begin{aligned} &\text{probability plot} && (F(X_{(k)}),\ \frac{N-k+1}{N+1}), \quad k = 1, \ldots, N, \\ &\text{quantile plot} && (X_{(k)},\ F^{-}(\frac{N-k+1}{N+1})), \quad k = 1, \ldots, N, \\ &\text{mean excess plot} && (X_{(k)}, e_N(X_{(k)})), \quad k = 1, \ldots, N. \end{aligned}$$

By the Glivenko-Cantelli theorem, the first two plots should be approximately linear, if the assumption $\mathcal{L}(X_j) = F$ really holds.

3.1 The POT-method (peaks-over-threshold)

The goal of this and the following sections is the derivation of tail estimates, i.e. estimates of $\overline{F}(x) = 1 - F(x)$ for large x, and of corresponding quantities like quantiles $F^{-}(q)$ for $q \approx 1$.

Assumption: $F \in MDA(H_\xi)$ for some GEV $H_\xi,\ \xi \geq 0$.

Let $K_N(u)$ and N_u denote the set of indices and the number of indices for which X_j exceeds the given threshold u (as above).

Definition: The <u>excesses</u> above the threshold u are the random variables $Y_l, l = 1, \ldots, N_u$, with $\{Y_1, \ldots, Y_{N_u}\} = \{X_j - u;\ j \in K_N(u)\}$.

The POT (peaks-over-threshold) method is based on considering the $Y_l,\ l \leq N_u$, as the main information about the tail behaviour of the original data $X_j, j \leq N$.

Remark: By definition, given $N_u, Y_1, \ldots, Y_{N_u}$ are i.i.d. with $\mathcal{L}(Y_l) = F_u$, the excess distribution (compare section 2.2). Therefore, by Theorem 2.10, $F_u(y) \approx G_{\xi,\beta(u)}(y)$ for some GPD, provided <u>u is large enough.</u>

By definition, $\overline{F}_u(y) = \text{pr}(X - u > y/X > u) = \overline{F}(y+u)/\overline{F}(u)$, or:

$$\overline{F}(x) = \overline{F}(u) \cdot \overline{F_u}(x-u), \ \ u < x < \infty.$$

u is large, therefore F_u may be approximated by $G_{\xi,\beta}$ for appropriate ξ, β. $\overline{F}(u)$ is replaced by $\hat{F}_N(u)$, the empirical distribution function:

$$\hat{F}_N(u) = \frac{N - N_u}{N} = 1 - \frac{N_u}{N}.$$

For u itself, this works, but not for $x \gg u$. The estimate $1 - \hat{F}_N(x)$ for $\overline{F}(x)$ depends for extreme x only on very few observations and is too unreliable. Using the above identity for $\overline{F}(x)$ and replacing the two factors of the right-hand side by their approximations we get:

Definition: The <u>POT-tail estimate</u> $\overline{F}^{\wedge}(x)$ for $\overline{F}(x), x$ large, is given by

$$\overline{F}^{\wedge}(x) = \frac{N_u}{N}\overline{G}_{\hat{\xi},\hat{\beta}}(x-u) = \frac{N_u}{N}(1 + \frac{\hat{\xi}(x-u)}{\hat{\beta}})^{-1/\hat{\xi}}, \ u < x < \infty,$$

where $\hat{\xi}, \hat{\beta}$ are some appropriate estimates (e.g. ML-estimates) for ξ, β based on the excesses $Y_1, \ldots, Y_{N_u}$.

We consider the ML-estimation of ξ, β for a sample $Y_1, \ldots, Y_M$ of fixed, not random, size M which are assumed to be i.i.d. with $\mathcal{L}(Y_j) = G_{\xi,\beta},\ \xi > 0$. As the density of the Pareto distribution is

$$g(y) = \frac{1}{\beta}(1 + \frac{\xi y}{\beta})^{-\frac{1}{\xi}-1},\ \ x \geq 0,$$

the log-likelihood function is, denoting $Y = (Y_1, \ldots, Y_M)^T$:

$$l(\xi, \beta / Y) = -M \log \beta - (\frac{1}{\xi} + 1) \sum_{j=1}^{M} \log(1 + \frac{\xi}{\beta} Y_j).$$

Maximizing it, we get the ML-estimates $\hat{\xi}, \hat{\beta}$.

Proposition 3.1 *If $\xi > -\frac{1}{2}$, then for $M \to \infty$:*

$$\sqrt{M}(\hat{\xi} - \xi,\ \frac{\hat{\beta}}{\beta} - 1)^T \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}_{\in}(\prime, \mathcal{D}^{-\infty})$$

with $D = (1+\xi) \begin{pmatrix} 1+\xi & -1 \\ -1 & 2 \end{pmatrix}$.
Moreover, $\hat{\xi}, \hat{\beta}$ are asymptotically efficient.

In the POT-approach, $M = N_u$ is random. Then, $\hat{\xi}, \hat{\beta}$ are the conditional ML-estimates given N_u. The limit theory is known for that case; to avoid an asymptotic bias, $\overline{F}$ has to satisfy some second-order conditions.

Definition: The POT-quantile estimate $\hat{x}_q$ for the q-quantile $x_q = F^{-}(q)$ is given as the solution of $\overline{F}^{\wedge}(\hat{x}_q) = 1 - q$, i.e.

$$\hat{x}_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left\{ (\frac{N}{N_u}(1-q))^{-\hat{\xi}} - 1 \right\}.$$

To compare this estimate with the common empirical quantile, assume that u is chosen such that there are exactly k exceedances: $N_u = k > N(1-q)$, i.e. $u = X^{(k+1)}$. Then, depending on the choice of u resp. k, the POT-quantile estimate is:

$$\hat{x}_{q,k} = X^{(k+1)} + \frac{\hat{\beta}_k}{\hat{\xi}_k} \left\{ \left(\frac{N}{k}(1-q) \right)^{-\hat{\xi}_k} - 1 \right\},$$

stressing the dependence of the ML-estimates for ξ, β on k. The empirical quantile is

$$\hat{x}_q^e = X^{([N(1-q)]+1)}$$

which corresponds to $\hat{x}_{q,k}$ for the minimal choice $k = [N(1-q)] + 1$ roughly.

Simulation studies show that the optimal choice k_0 for k which minimizes $\text{mse}(\hat{x}_{q,k}) = \mathcal{E}(\hat{\S}_{\text{II},\|} - \S_{\text{II}})^{\in}$ is much larger than $[N(1-q)]+1$, i.e. the POT-estimate differs considerably from the very variable empirical quantile.

The quality of the POT-estimate essentially depends on the choice of the threshold u. Qualitatively, we have the following bias-variance dilemma:

- if u is too large, there are too few exceedances $Y_l,\ l \leq N_u$, and the variance of estimates increases,
- u is too small, the approximation of the excess distribution F_u by a GPD is not good, and a nonneglible bias occurs.

Exploratory methods for choosing a useful threshold are based on results like:

Proposition 3.2 *If $\mathcal{L}(Z) = G_{\xi,\beta}$ is a GPD, then the mean excess function is linear:*

$$e(u) = \mathcal{E}\{\mathcal{Z} - \sqcap | \mathcal{Z} > \sqcap\} = \frac{\beta + \xi\sqcap}{\infty + \xi},\quad \sqcap \geq \prime,\quad \textit{for } \prime \leq \xi < \infty.$$

For the Pareto distribution in its usual representation, $\xi = \frac{1}{\alpha}$. The condition $\xi < 1$ is therefore equivalent to $\alpha > 1$, i.e. $\mathcal{E}|\mathcal{Z}| < \infty$.

Threshold selection rule: For POT-estimates, select the threshold u such that the empirical mean excess function $e_N(v)$ is approximately linear for $v \geq u$. For this purpose, the mean excess plot is considered where it is often advisable to omit the highly variable right-most points $(X_{(k)}, e_N(X_{(k)})),\ k \approx N$, which disturb the visual impression.

3.2 Measures of Risk

Definition: Let $0 < q < 1$, and $F = \mathcal{L}(X)$ be the distribution of claims or losses. Typically, $q = 0.95$ or $q = 0.99$.

a) The Value-at-Risk (VaR) is the q-quantile

$$\mathrm{VaR}_q(X) \equiv x_q = F^{-}(q).$$

b) The Expected Shortfall is

$$\mathrm{ES}_q(X) \equiv S_q = \mathcal{E}\{\mathcal{X}|\mathcal{X} > \S_{\amalg}\}.$$

Definition: (Artzner, Delbaen, Eber, Heath): A coherent risk measure is a function ρ on the space of real-valued random variables (corresponding to the losses) with

A1)	$X \geq Y$ a.s. $\Longrightarrow \rho(X) \geq \rho(Y)$	(monotonicity)
A2)	$\rho(X + Y) \leq \rho(X) + \rho(Y)$	(subadditivity)
A3)	$\rho(\lambda X) = \lambda\rho(X)$ for $\lambda \geq 0$	(positive homogeneity)
A4)	$\rho(X + a) = \rho(X) + a$	(translation equivariance)

ES is a coherent risk measure, VaR isn't. Consider, e.g., independent X, Y, assuming only the two values 0 and 100, with

$$\mathcal{L}(X) = \mathcal{L}(Y) = 0.9\,\delta_0 + 0.1\,\delta_{100}$$

and, therefore,

$$\mathcal{L}(X+Y) = 0.9^2\,\delta_0 + 0.1^2\,\delta_{200} + 2 \cdot 0.1 \cdot 0.9\,\delta_{100}.$$

For $q = 0.95$:

$$\mathrm{VaR}_q(X) = \mathrm{VaR}_q(Y) = 0 \;\text{ but }\; \mathrm{VaR}_q(X+Y) = 100.$$

The expected shortfall is closely related to the mean excess function at $u = x_q$:

$$S_q = e(x_q) + x_q.$$

Proposition 3.3 *a) If $F \in MDA(H_\xi),\ 0 < \xi < 1$ (Fréchet-case), then*

$$\lim_{u \to \infty} \frac{1}{u}\, e(u) = \frac{\xi}{1-\xi}.$$

b) If $F \in MDA(H_0)$ (Gumbel-case), then

$$\lim_{u \to \infty} \frac{1}{u}\, e(u) = 0.$$

Consider the expected shortfall-to-quantile ratio $\frac{S_q}{x_q} = \frac{e(x_q)}{x_q} + 1$ for

a) $F = \mathcal{N}(0,1) \in MDA(H_0) \Longrightarrow \lim_{q \to 1} \frac{S_q}{x_q} = 1.$

b) $F = t_\nu \in MDA(H_\xi)$ with $\xi = \frac{1}{\nu} \Longrightarrow \lim_{q \to 1} \frac{S_q}{x_q} = \frac{1}{1-\xi} = \frac{\nu}{\nu-1} > 1.$

q	0.95	0.99	0.995	$q \to 1$
$\mathcal{N}(0,1)$	1.25	1.15	1.12	1
t_4	1.50	1.39	1.37	1.33
t_2	2.11	2.02	2.01	2

Losses exceeding VaR_q exceed it by 15% on the average for $q = 0.99$, $\mathcal{N}(0,1)$, but by 102% on the average for $q = 0.99$, t_2.

The expected shortfall may be estimated by the POT-method. $F_u(x) \approx G_{\xi,\beta}(x)$ for large enough threshold u implies

$$e(\nu) \approx \frac{\beta + (\nu - u)\xi}{1-\xi} \;\text{ for }\; \nu > u.$$

Therefore, for $x_q > u$, we have

$$\frac{S_q}{x_q} = 1 + \frac{e(x_q)}{x_q} \approx \frac{1}{1-\xi} + \frac{\beta - \xi u}{x_q(1-\xi)}.$$

The POT-estimate for S_q is, then, with $\hat{x}_q$ denoting the POT-quantile estimate

$$\hat{S}_{q,u} = \frac{\hat{x}_q}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} u}{1-\hat{\xi}}.$$

3.3 The Hill estimator

Recall: For $\xi > 0$, the GEV H_ξ is a Fréchet distribution Φ_α with $\alpha = \frac{1}{\xi}$.

Assumption: $X_1, \ldots, X_n$ i.i.d. with $\mathcal{L}(X_j) = F \in MDA(\Phi_\alpha)$ for some $\alpha > 0$.

Definition: Let $X^{(1)} \geq X^{(2)} \geq \ldots \geq X^{(n)}$ be the order statistics in descending order. The Hill estimator $\hat{\alpha}_H$ of the tail index α is, for appropriate $k = k(n)$, given by

$$\hat{\alpha}_H = \{\frac{1}{k}\sum_{j=1}^{k} \log X^{(j)} - \log X^{(k)}\}^{-1}.$$

We motivate this form of an estimate by considering ML-estimates in a series of successively more complicated situations. Recall $F \in \text{MDA}(\Phi_\alpha)$ iff $\overline{F}(x) = \frac{L(x)}{x^\alpha}$ for some slowly varying function L.

1) Assume $\overline{F}(x) = \frac{1}{x^\alpha},\ x \geq 1$. Then, for $Y_j = \log X_j$ we have

$$\text{pr}(Y_j > y) = \text{pr}(X_j > e^y) = \overline{F}(e^y) = e^{-\alpha y},\ y \geq 0,$$

i.e. $Y_1, \ldots, Y_n$ are i.i.d. $\text{Exp}(\alpha)$. It is well-known that the ML-estimate of $\alpha = (\mathcal{E}Y)^{-1}$ is:

$$\hat{\alpha} = (\overline{Y}_N)^{-1} = \{\frac{1}{n}\sum_{j=1}^{n} \log X_j\}^{-1} = \{\frac{1}{n}\sum_{j=1}^{n} \log X^{(j)}\}^{-1}.$$

2) Assume $\overline{F}(x) = \frac{C}{x^\alpha},\ x \geq u > 0$, with $C = u^\alpha$. Dividing X_j by u, we are back in case 1). Therefore, the ML-estimate of α is now

$$\hat{\alpha} = \{\frac{1}{n}\sum_{j=1}^{n} \log \frac{X_j}{u}\}^{-1} = \{\frac{1}{n}\sum_{j=1}^{n} \log X^{(j)} - \log u\}^{-1}.$$

3) For general $F \in \text{MDA}(\Phi_\alpha)$, we have $\overline{F}(x) \approx \frac{C}{x^\alpha}$ for $x \geq u$ where u is an appropriately large threshold. Let again $N_u = \#\{j \leq N; X_j \geq u\}$. We condition on the event $N_u = k$, i.e. for $x \geq X^{(k)}$ we are approximately in case 2). We calculate the conditional maximum-likelihood (CML-)estimate for α given $N_u = k$. For that purpose we need:

Proposition 3.4 *Let $X_1, \ldots, X_n$ be i.i.d. with $\mathcal{L}(X_j) = F$ and density f.*
a) The joint density of the order statistics $X^{(1)} \geq \ldots \geq X^{(n)}$ is

$$f^{(n)}(x_1, \ldots, x_n) = \begin{cases} n!\Pi_{j=1}^{n} f(x_j) & \text{for } x_1 > \ldots > x_n \\ 0 & \text{else} \end{cases}$$

b) The joint density of $X^{(1)}, \ldots, X^{(k)},\ k \leq n$, is

$$f^{(k)}(x_1, \ldots, x_k) = \begin{cases} \frac{n!}{(n-k)!} F^{n-k}(x_k)\Pi_{j=1}^{k} f(x_j) & \text{for } x_1 > \ldots > x_k \\ 0 & \text{else} \end{cases}$$

Idea of proof: a) density of $(X_1, \ldots, X_n)^T$ is $\Pi_{j=1}^n f(x'_j),\ x'_1, \ldots, x'_n \in \mathbb{R}$. Every possible value $(x_1, \ldots, x_n)^T,\ x_1 > \ldots > x_n$, of the vector of order statistics is the result of ordering one out of $n!$ possible vectors $(x'_1, \ldots, x'_n)$.

b) Integrate $f^{(n)}$ with respect to $x_{k+1}, \ldots, x_n$ successively, remembering $x_n < x_{n-1},\ x_{n-1} < x_{n-2}$, etc.

If $\overline{F}(x) \approx \frac{C}{x^\alpha}$ for $x > u$, then the density of F is $f(x) \approx \frac{\alpha C}{x^{\alpha+1}}$ for $x > u$, and, moreover, we have $F^{n-k}(x) \approx (1 - \frac{C}{x^\alpha})^{n-k}$ for $x > u$. The conditional likelihood given $N_u = k$ is therefore

$$L_k(\alpha, C) \approx \frac{n!}{(n-k)!}(1 - \frac{C}{x_k^\alpha})^{n-k}(\alpha C)^k \Pi_{j=1}^k \frac{1}{x_j^{\alpha+1}},\ u < x_k < \ldots < x_1.$$

Replacing the variable $x_k, \ldots, x_1$ by the observed $X^{(k)}, \ldots, X^{(1)}$ and maximizing w.r.t. α, C, we get the CML-estimates:

$$\hat{\alpha} = \{\frac{1}{k}\sum_{j=1}^k \log X^{(j)} - \log X^{(k)}\}^{-1} = \hat{\alpha}_H$$
$$\hat{C} = \frac{k}{n}(X^{(k)})^{\hat{\alpha}}.$$

Another approach, leading also to the Hill estimate, uses $\overline{F}(x) = \frac{L(x)}{x^\alpha}$ for some $L \in \mathcal{R}_0$ directly. By definition of L as a slowly varying function:

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = \frac{1}{t^\alpha} \quad \text{for all } t > 0.$$

Using partial integration and Karamata's theorem on the tail behaviour of integrals of regularly varying functions, this implies:

$$\frac{1}{\overline{F}(x)}\int_x^\infty (\log t - \log x)dF(t) \longrightarrow \frac{1}{\alpha} \text{ for } x \to \infty.$$

We replace F by the empirical distribution $F_n(t) = \frac{1}{n}\#\{j \le n;\ X_j \le t\}$ and x by a large, data-dependent level, e.g. $x = X^{(k)}$ for $k = k(n)$. If, for $n \to \infty$, $k \to \infty$ and $\frac{k}{n} \to 0$, we still have $X^{(k)} \to \infty$.

$$\frac{1}{\tilde{\alpha}} = \frac{1}{\overline{F}_n(X^{(k)})}\int_{x^{(k)}}^\infty (\log t - \log X^{(k)})dF_n(t) = \frac{1}{k-1}\sum_{j=1}^{k-1} \log X^{(j)} - \log X^{(k)}$$

i.e. $\tilde{\alpha} \approx \hat{\alpha}_H$.

Theorem 3.1 *Let $X_1, X_2, \ldots,$ i.i.d. with $\mathcal{L}(X_j) = F \in MDA\ (\Phi_\alpha),\ \alpha > 0$.*

a) weak consistency: $\hat{\alpha}_H \xrightarrow[p]{} \alpha$ for $n, k \to \infty$ such that $\frac{k}{n} \to 0$.

b) strong consistency: $\hat{\alpha}_H \underset{a.s.}{\longrightarrow} \alpha$ *for* $n, k \to \infty$, $\frac{k}{n} \to 0$, $\frac{k}{\log\log n} \to \infty$.

c) asymptotic normality: $\sqrt{k}(\hat{\alpha}_H - \alpha) \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, \alpha^2)$ *under further assumptions on* k, F.

Choosing an appropriate value for k is again a bias-variance-dilemma. If $k \nearrow$, then $\mathrm{var}(\hat{\alpha}_H) \searrow$ but bias $\uparrow$ as in the POT-method. The following result describes a k which achieves some kind of balance between bias and variance. However, the second-order assumptions on F are not verifiable in practice.

Proposition 3.5 *Let* $F \in MDA(\Phi_\alpha), \alpha > 0$, *and, moreover,*

$$\lim_{x\to\infty} \frac{1}{a(x)} \left\{ \frac{\overline{F}(tx)}{\overline{F}(x)} - \frac{1}{t^\alpha} \right\} = \begin{cases} \frac{1}{t^\alpha} \frac{t^\rho - 1}{\rho} & , t > 0, \quad \rho < 0 \\ \frac{1}{t^\alpha} \log t & , t > 0, \quad \rho = 0 \end{cases}$$

for some function $a(x)$ *with* $|a(x)| \in \mathcal{R}_\rho$ *and sgn* $a(x) = const.$ *Let*

$$A(t) = \frac{1}{\alpha^2}\, a(F^{-}(1 - \frac{1}{t})), \ t > 0.$$

If $k \to \infty$, $\frac{k}{n} \to 0$ *such that* $\sqrt{k}A(\frac{n}{k}) = \lambda \in \mathbb{R}$ *for all* n, *then:*

$$\sqrt{k}(\hat{\alpha}_H - \alpha) \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(\frac{\alpha^{\ni}\lambda}{\rho - \alpha}, \alpha^{\in})$$

Based on the Hill estimate for the tail parameter α, we also get estimates for the tail of F or for quantiles of F:

If $F \in MDA(\Phi_\alpha)$ and, therefore, $\overline{F}(x) = \frac{L(x)}{x^\alpha}$ for some $L \in \mathcal{R}_0$, we have for $x \geq X^{(k)}$:

$$\frac{\overline{F}(x)}{\overline{F}(X^{(k)})} = \frac{L(x)}{L(X^{(k)})} \left(\frac{X^{(k)}}{x} \right)^\alpha \approx \left(\frac{X^{(k)}}{x} \right)^\alpha$$

as a slowly varying function is nearly constant in the tails. Using $\overline{F}_n(X^{(k)}) = \frac{k}{n} \approx \overline{F}(X^{(k)})$, where F_n denotes the empirical distribution, we get as an estimate of $\overline{F}(x)$:

$$\overline{F}^{\wedge}_H(x) = \frac{k}{n}(\frac{X^{(k)}}{x})^{\hat{\alpha}_H}$$

as the <u>Hill tail estimate</u>. Inverting this estimate, we get the <u>Hill quantile estimate</u> for $q \approx 1$:

$$\begin{aligned} \hat{x}_{q,H} &= X^{(k)} \left\{ \frac{n}{k}(1-q) \right\}^{-1/\hat{\alpha}_H} \\ &= X^{(k)} + X^{(k)} \left(\{\frac{n}{k}(1-q)\}^{-\hat{\xi}_H} - 1 \right) \end{aligned}$$

with $\hat{\xi}_H = 1/\hat{\alpha}_H$, where the latter form stresses the similarities and the differences to the POT-quantile estimate.

3.4 Extreme value theory for time series

Let $Z_j, \ -\infty < j < \infty$, be a strictly stationary time series with $\mathcal{L}(Z_j) = F$ i.e.

$$\mathrm{pr}(Z_{j_1} \le x_1, \ldots, Z_{j_k} \le x_k) = \mathrm{pr}(Z_{j_1+t} \le x_1, \ldots, Z_{j_k+t} \le x_k)$$

for all $k \ge 1, \ -\infty < j_1 < j_2 < \ldots < j_k < \infty, x_1, \ldots, x_k \in \mathbb{R}, \ t \in \mathbb{Z}$. Let $X_1, X_2, \ldots$ be i.i.d. with the same distribution $\mathcal{L}(X_j) = F$.

Let $M_n = \max\{Z_1, \ldots, Z_n\}, \ M_n^x = \max\{X_1, \ldots, X_n\}$. A simple fundamental relation for the previous chapters was, using independence of the X_j,

$$\mathrm{pr}(M_n^x \le y) = (\mathrm{pr}(X_j \le y))^n = F^n(y).$$

In the dependent situation of time series, this argument no longer applies, and $\mathcal{L}(M_n)$ is not known in terms of F. However, often there is at least a similar approximation:

$$\mathrm{pr}(M_n \le y) \approx F^{n\delta}(y) \ge F^n(y) \quad \text{for large } n,$$

where $\delta \in [0, 1]$ is the so-called extremal index. For a precise definition, recall Proposition 2.2 for the i.i.d. case:

$$n\overline{F}(u_n) \to \tau \quad \text{iff} \quad \mathrm{pr}(M_n^x \le u_n) \longrightarrow e^{-\tau}.$$

Definition: $\delta \in [0, 1]$ is called the extremal index of the time series $Z_j, \ -\infty < j < \infty$, if for some τ, u_n

$$n\overline{F}(u_n) \to \tau \quad \text{and} \quad \mathrm{pr}(M_n \le u_n) \to e^{-\delta\tau}.$$

(δ, if it exists, does not depend on the special choice of τ, u_n).

Remarks: a) This definition implies the approximation above as

$$\mathrm{pr}(M_n \le u_n) \approx e^{-\delta\tau} \approx e^{-\delta n\overline{F}(u_n)} = (e^{-\overline{F}(u_n)})^{n\delta} \approx (1 - \overline{F}(u_n))^{n\delta} = F^{n\delta}(u_n).$$

b) Not every stationary time series has an extremal index. Consider, e.g., $Z_j = A \cdot X_j$, where X_j are i.i.d., $A > 0$ is a random variable independent of the X_j, $\mathcal{L}(X_j) \in MDA(\Phi_\alpha)$ for some $\alpha > 0$ with normalizing sequence $c_n > 0$. By Theorem 2.7,

$$\mathrm{pr}(\frac{1}{c_n} M_n \le y) = \mathrm{pr}(\frac{1}{c_n} M_n^x \le \frac{y}{A}) = \mathcal{E}\{\mathrm{pr}(\frac{\infty}{\rfloor\backslash}\mathcal{M}^{\S}_{\backslash} \le \frac{\dagger}{\mathcal{A}})|\mathcal{A}\}$$

$$\longrightarrow \mathcal{E}\oplus_\alpha(\frac{\dagger}{\mathcal{A}}) = \mathcal{E}\exp(-\frac{\mathcal{A}^\alpha}{\dagger^\alpha}) \text{ for } \dagger > \prime,$$

whereas, again by Theorem 2.7, using $\overline{F}(x) = \frac{L(x)}{x^\alpha}$ and $\overline{F}(c_n) \approx \frac{1}{n}$

$$n\overline{F}(c_n y) = n\overline{F}(c_n) \cdot \frac{\overline{F}(c_n y)}{\overline{F}(c_n)} \approx 1 \cdot \frac{1}{y^\alpha}.$$

Therefore, with $\tau = \frac{1}{y^\alpha}, u_n = c_n y$, we have $n\overline{F}(u_n) \to \tau$, but $\mathrm{pr}(M_n \le u_n) \longrightarrow \mathcal{E}\rceil^{-\tau\mathcal{A}^\alpha}$.

c) For white noise, i.e. i.i.d. Z_j, we have trivially $\delta = 1$.

d) If $\{Z_j\}$ is a Gaussian ARMA(p, q)-process, e.g. for $p = q = 1$:

$$Z_{t+1} = aZ_t + b\,\varepsilon_t + \varepsilon_{t+1},\ \varepsilon_t \quad \text{i.i.d.} \quad \mathcal{N}(0, \sigma^2),$$

then $\delta = 1$.

e) If $\{Z_j\}$ is an ARCH(1)-process, i.e.

$$Z_t = \sigma_t\,\varepsilon_t\ ,\ \varepsilon_t \quad \text{i.i.d.} \quad \mathcal{N}(0, 1),\ \sigma_t^2 = \omega + a Z_{t-1}^2,$$

then $\delta = \delta(a) < 1$. δ can be calculated approximatively, e.g. for $a = \frac{1}{2}$, we have $\delta \approx 0.835$.

There are two conditions $D(u_n), D'(u_n)$ which guarantee $\delta = 1$ as in the i.i.d. case. Financial time series violate the second one as, due to the practically observed *stylized fact* of volatility clustering, extremal observations tend to be closer together than in the standard situation.

Definition For any sequence of thresholds u_n as in the definition of the extremal index:

a) $D(u_n)$-condition: For $n, l \geq 1$ let

$$\alpha_{n,l} = \sup |\text{pr}(\max_{i \in A \cup B} Z_i \leq u_n) - \text{pr}(\max_{i \in A} Z_i \leq u_n) \cdot \text{pr}(\max_{i \in B} Z_i \leq u_n)|$$

where the supremum is taken over all $1 \leq k \leq n - l,\ A \subset \{1, \ldots, k\},\ B \subset \{k + l, \ldots, n\}$. Assume that for $n \to \infty$ and some sequence $l = l(n) \to \infty$ with $\frac{l}{n} \to 0$, we have $\alpha_{n,l} \to 0$.

b) $D'(u_n)$-condition:

$$\lim_{k \to \infty} \limsup_{n \to \infty} n \sum_{j=2}^{[n/k]} \text{pr}(X_1 > u_n, X_j > u_n) = 0.$$

$D(u_n)$ states the asymptotic independence of maxima taken over index sets A, B which are far apart in time (at least by a lag $l \to \infty$). This is a rather weak form of the mixing conditions in time series analysis which guarantee that the presence is only weakly depending on the remote past. A consequence of the approximate independence of block maxima is:

$$\text{pr}(M_n \leq u_n) \approx (\text{pr}(M_{[n/k]} \leq u_n))^k \qquad (*)$$

for fixed (or slowly increasing) number k of blocks of length $[n/k]$ each.

$D'(u_n)$ is an anti-clustering condition which states that the occurence of two extrema (exceeding threshold u_n) close together (separated by a time lag of at most $[n/k]$) has a very small probability.

Theorem 3.2 *Let $\{Z_j\}$ be a stationary time series with extremal index $\delta > 0$, $X_1, X_2, \ldots$ i.i.d. with $\mathcal{L}(\mathcal{X}_|) = \mathcal{L}(\mathcal{Z}_|) = \mathcal{F}$. Then, for a GEV H_ξ,*

$$pr\left(\frac{M_n^x - d_n}{c_n} \le x\right) \longrightarrow H_\xi(x) \quad \textit{iff} \quad pr\left(\frac{M_n - d_n}{c_n} \le x\right) \longrightarrow H_\xi^\delta(x)$$

for all x in the support of H_ξ.

The maxima of the time series have asymptotically the same type of distribution as the i.i.d. data, as H_ξ^δ is itself a GEV with the same shape parameter as H_ξ, e.g. for $\xi > 0$:

$$H_\xi^\delta(x) = \exp\{-\delta(1+\xi x)^{-1/\xi}\} = H_\xi(\frac{x-\mu}{\sigma}), \;\; 1+\xi x > 0$$

with $\sigma = \delta^\xi$ and $\mu = -(1-\delta^\xi)$.

Based on this result, many techniques developped for extreme value statistics of i.i.d. data may be used for time series if appropriately adapted. One of the main problems is that the effective sample size is $n\delta$ instead of n, i.e. more data are needed. For the POT-method, e.g., the idea of approximating the excess distribution by a GPD still works. However, for estimating ξ, β by ML, we have to make particular model assumptions to be able to write down the likelihood function as, in particular due to clustering of extremes as consequence of the violation of $D'(u_n)$, the excesses $Y_1, Y_2, \ldots$ are no longer independent too. One approach tries instead to make them "more independent" by replacing clusters of exceedances by just one exceedance, e.g. by the maximum value in a cluster. Cluster size has to be chosen such that the number of exceedances is reduced by a factor δ approximately (a sample of n time series data corresponds to $n\delta$ independent observations with respect to its information about extremes). Then, the standard POT-method is applied to the reduced data set.

For applying such modifications of the methods of previous chapters one needs the extremal index δ. It may be estimated by various methods. We consider only one of them which may be easily explained without additional background information:

The blocks method: Partition $Z_1, \ldots, Z_N$ into b blocks of size n each ($N = bn$, b, n large). Let $M_n^{(k)}$ be the maximum in the k-th block:

$$M_n^{(k)} = \max(Z_{(k-1)n+1}, \ldots, Z_{kn}), \;\; k = 1, \ldots, b.$$

For a large threshold $u = u_N$:

$$N_u = \#\{j \le N;\; Z_j > u\}, \;\; B_u = \#\{k \le b;\; M_n^{(k)} > u\},$$

$$\hat{\delta} := \frac{1}{n} \frac{\log(1 - \frac{B_u}{b})}{\log(1 - \frac{N_u}{N})}.$$

This estimate of δ is justified by the following three arguments:

a) For large N, $\operatorname{pr}(M_N \leq u) \approx F^{\delta N}(u)$ if $u = u_N \to \infty$ such that $N\overline{F}(u_N) \to \tau$. Solving for δ we get:

$$\delta \approx \frac{\log \operatorname{pr}(M_N \leq u)}{N \log F(u)}.$$

b) Estimate F by empirical distribution $F_N : F(u) = 1 - \operatorname{pr}(Z_j > u) \approx 1 - \frac{N_u}{N}$.
c) Use (*) above, recalling $N = bn$:

$$\operatorname{pr}(M_N \leq u) \approx \{\operatorname{pr}(M_n \leq u)\}^b \approx \{\frac{1}{b}\sum_{k=1}^{b} 1_{(-\infty,u]}(M_n^{(k)})]^b = (1 - \frac{B_u}{b})^b.$$

Combining a)-c), we get:

$$\delta \approx \frac{b \log(1 - \frac{B_u}{b})}{N \log(1 - \frac{N_u}{N})} = \hat{\delta}.$$