$\gamma \neq 0$:

$$\ell(\theta) = n \log \sigma - \left(\frac{1}{\gamma}+1\right) \sum_{i=1}^{n} \log\left(1+\gamma \frac{x_i-\mu}{\sigma}\right) - \sum_{i=1}^{n} \left(1+\gamma \frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\gamma}},$$

if $1+\gamma \frac{x_i-\mu}{\sigma} > 0$ for all $i = 1, \dots, n$

and

$\ell(\theta) = -\infty$ else

$\gamma = 0$:

$$\ell(\theta) = -n \log \sigma - \sum_{i=1}^{n} \frac{x_i-\mu}{\sigma} - \sum_{i=1}^{n} e^{-\frac{x_i-\mu}{\sigma}}$$

→ maximization of $\ell(\theta)$ w.r.t. θ (in general numerically) gives the maximum likelihood estimator

$$\hat{\theta} = (\hat{\gamma}, \hat{\mu}, \hat{\sigma}) = \operatorname{argmax} \ell(\gamma, \mu, \sigma)$$

→ the latter estimation enables us to find the dike height z_p as a $(1-p)$-quantile of the annual maxima, that is as the (smallest) solution $\hat{z}_p$ to the equation

$$G_{\hat{\gamma},\hat{\mu},\hat{\sigma}}(\hat{z}_p) = 1-p$$

(4.2), (4.3) →

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}\left(1-(-\log(1-p))^{-\hat{\gamma}}\right), & \hat{\gamma} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log(-\log(1-p)), & \hat{\gamma} = 0 \end{cases} \tag{4.7}$$

→ Problem: goodness of fit of $\hat{G} = G_{\hat{\gamma},\hat{\mu},\hat{\sigma}}$ to the real distribution

→ possible methods:

1. PP-plot: Let $x_{(1)} \leq \dots \leq x_{(n)}$ be the order statistics of $x_1, \dots, x_n$

Def. 4.1 (Probability-Plot)

The probability-plot (PP-plot) is given by the graph

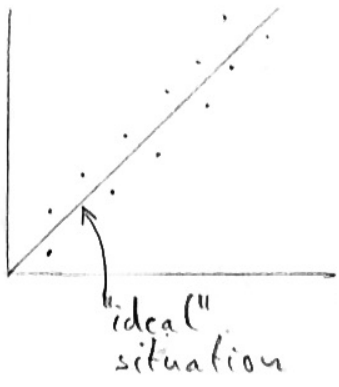
$$\left\{\left(\hat{G}(x_{(i)}), \frac{i}{n+1}\right), i = 1, \dots, n\right\}$$

Since the empirical distr. F_{n+1} given by

$$F_{n+1}(x) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{1}_{(-\infty, x]}(x_i)$$

can be regarded as a "good" approximation of the real distr., a "good" fit of $\hat{G}$ to the real distr. should imply that

$$\hat{G}(x_{(i)}) \approx F_{n+1}(x_{(i)}) \overset{\text{Def}}{=} \frac{i}{n+1}, \quad i = 1, \dots, n$$



→ disadvantage:

If $i \approx n$ then $\hat{G}(x_{(i)}) \approx 1$ and $\frac{i}{n+1} \approx 1$, regardless how well (or badly) $\hat{G}$ describes the data. However, especially the large values are of great interest here.

In order to overcome this deficiency one may resort to (see also the exercises)

2. QQ-plot:

Def 4.2 (Quantile-plot)

The quantile plot (QQ-plot) is given by the graph

$$\left\{ \left(\hat{G}^{\leftarrow}\left(\frac{i}{n+1}\right), x_{(i)} \right), i = 1, \dots, n \right\}$$

→ a "good" fit of $\hat{G}$ to the real distr. would imply that

$$\hat{G}^{\leftarrow}\left(\frac{i}{n+1}\right) \approx F_{n+1}^{\leftarrow}\left(\frac{i}{n+1}\right) = x_{(i)}, \quad i = 1, \dots, n$$

2. non-stationary model:

Assumption:

(i) $X_1, \dots, X_n$ indep. r.v.'s (but not necessarily identically distr.)

(ii) $X_i = \max_{j=1}^{365} X_{ij}$ for i.i.d. r.v.'s $X_{ij}, j = 1, \dots, 365$ for all i

The same reasoning as in the stationary case

also gives a justification for the additional assumption

(iii) $X_i \sim G_{\gamma(i),\mu(i),\delta(i)}$, $i=1,\dots,n$

$\longrightarrow$ the parameter vector $(\gamma(i),\mu(i),\delta(i))$ is a function of time i

$\longrightarrow$ this model is <u>useful</u>, if the data $x_1,\dots,x_n$ <u>exhibit a trend</u> (e.g. annual maxima increase over time)

$\longrightarrow$ <u>common choice</u> of $(\gamma(i),\mu(i),\delta(i))$:

$$\gamma(i)=\gamma,\ \mu(i)=\beta_0+\beta_1\cdot i,\ \delta(i)=\delta \qquad (4.8)$$

$\longrightarrow$ the location parameters $\mu(i)$, $i=1,\dots,n$ capture the trend in the data

<u>Just as in the stationary case</u> we can use <u>MLE</u> to <u>estimate the parameters</u> $(\gamma,\beta_0,\beta_1,\delta)$

<u>Further</u>, the <u>goodness of fit</u> of $G_{\hat\gamma(i),\hat\mu(i),\hat\delta(i)}$, $i=1,\dots,n$ to real distributions can be verified by using <u>PP- or QQ-plots</u> applied to the <u>i.i.d realizations</u>

$$x_i' := x_i-\hat\beta_0-\hat\beta_1\cdot i,\ i=1,\dots,n$$

(since PP- or QQ-plots are only applicable to i.i.d data)

<u>alternative choice</u> of $(\gamma(i),\mu(i),\delta(i))$:

$$\gamma(i)=\gamma,\ \mu(i)=e^{\beta_0+\beta_1 i},\ \delta(i)=\delta, \qquad (4.9)$$

$i=1,\dots,n$.

2. Peaks over threshold method: GP-distribution

The other estimation method we want to discuss is based on the generalized Pareto distribution (GPD)

In order to motivate this approach let us consider the following examples:

Ex. 4.3 ($F \sim Exp(\lambda) \in MDA(\Lambda)$)
Let X be exponentially distributed, that is $\overline{F}(x) = e^{-\lambda x}$, $x > 0$ for some $\lambda > 0$.
Then

$$P(X-u > x \mid X > u) \overset{Def.}{=} \frac{P(X > u+x,\ X > u)}{P(X > u)}$$

$$= \frac{P(X > u+x)}{P(X > u)} = \frac{e^{-\lambda(u+x)}}{e^{-\lambda u}} = e^{-\lambda x},$$

that is the exponential distr. is memoryless (see Exerc. 1)

$$\Rightarrow \lim_{u \to \infty} P(X-u \le x \mid X > u) = 1 - e^{-\lambda x}$$

for all $x > 0$

Ex. 4.4 ($F \in MDA(\Phi_\alpha)$)

$$F \in MDA(\Phi_\alpha) \overset{Th.\,3.2.5}{\Longleftrightarrow} \overline{F} \in RV_{-\alpha} \text{ and } x^* \overset{def}{=} \sup\{y \in \mathbb{R} : F(y) < 1\} = \infty$$

$$\Rightarrow P\left(\frac{X-u}{u} > x \mid X > u\right) \overset{Def.}{=} \frac{P(X > u + ux)}{P(X > u)}$$

$$= \frac{\overline{F}(u(x+1))}{\overline{F}(u)} \underset{u \to \infty}{\overset{\overline{F} \in RV_{-\alpha}}{\longrightarrow}} (1+x)^{-\alpha} \quad \text{for all } x > 0$$

$$\Rightarrow P\left(\frac{X-u}{u} \le x \mid X > u\right) \xrightarrow[u \to \infty]{} \underbrace{1 - (1+x)^{-\alpha}}_{=: G(x) \leftarrow \text{distr.}} \text{ for all } x > 0$$

$\overset{\text{Exerc. 9, Prob. 3}}{\Longrightarrow}$

$$\lim_{u \to \infty} \sup_y \left| P\left(\frac{X-u}{u} \le y \mid X > u\right) - G(y) \right| = 0$$

$$\Rightarrow \lim_{u \to \infty} \sup_x \left| P(X-u \le x \mid X > u) - G(\tilde{\beta}(u) x) \right| = 0$$

for $\tilde{\beta}(u) := \frac{1}{u}$

Ex. 4.3, 4.4 ⟶ observation:

The function $F_u(x) := P(X-u \le x \mid X > u)$, $u < x^*$, called excess function of X over the threshold u converges to a distribution function which is a special case of a generalized Pareto distribution

Def. 4.5 (Generalized Pareto distribution (GPD))

The generalized Pareto distribution with index $\gamma \in \mathbb{R}$ and scale parameter $\beta > 0$ is defined by the following distr.

$$P_{\gamma,\beta}(x) = \begin{cases} 1 - (1 + \frac{\gamma x}{\beta})^{-\frac{1}{\gamma}}, & x > 0 \quad \text{if } \gamma > 0 \\ 1 - (1 + \frac{\gamma x}{\beta})^{-\frac{1}{\gamma}}, & 0 \le x \le -\frac{\beta}{\gamma} \quad \text{if } \gamma < 0 \\ 1 - e^{-x/\beta}, & x > 0 \quad \text{if } \gamma = 0 \end{cases}$$

(exponential distr.)

The above observation gives rise to the following question:

Can the excess function F_u of X, whose distr. belongs to a maximum domain of attraction, be approximated by a generalized Pareto distr.?

⟶ Yes, that is in the sense of

Th. 4.6 (Pickands-Balkema-de Haan, 1974)

Let $X \sim F$. Assume that F is continuous in $x^* := \sup\{y \in \mathbb{R} : F(y) < 1\}$. (see (4.1))

Then

$$F \in MDA(G_{\gamma,0,1})$$

if and only if there ex. a positive measurable function β such that

$$\lim_{u \uparrow x^*} \sup_{x \in (0, x^*-u]} |F_u(x) - P_{\gamma, \beta(u)}(x)| = 0 .$$

Proof:

"$\Rightarrow$": (consider the case $G_{\gamma,0,1} = \Lambda$ ← Gumbel distr.

$\overset{\text{Th. 3.2.12}}{\Rightarrow}$ $\exists$ positive meas. funct. β such that

$$\lim_{u \uparrow x^*} \frac{\overline{F}(u + x\beta(u))}{\overline{F}(u)} = e^{-x}$$

for all x.

$$\Rightarrow F_u(\beta(u)x) \overset{\text{Def.}}{=} 1 - P(X > u + x\beta(u) \mid X > u)$$

$$= 1 - \frac{\overline{F}(u + x\beta(u))}{\overline{F}(u)} \xrightarrow[u \uparrow x^*]{} P_{0,1}(x)$$

for all $x \geq 0$

$\overset{\text{Excerc. 9, Prob. 3}}{\Rightarrow}$ $P_{0,1}$ continuous

$$\lim_{u \uparrow x^*} \sup_{0 \leq x < \infty} |F_u(\beta(u)x) - P_{0,1}(x)| = 0$$

$\overset{x = \frac{1}{\beta(u)} y}{\Rightarrow}$

$$\lim_{u \uparrow x^*} \sup_{0 \leq x < \infty} |F_u(x) - \underbrace{P_{0,1}\left(\tfrac{1}{\beta(u)} x\right)}_{= P_{0,\beta(u)}(x)}| = 0$$

$$\Rightarrow \lim_{u \uparrow x^*} \sup_{0 < x < x^* - u} |F_u(x) - P_{0,\beta(u)}(x)| = 0$$

The other cases for Ψ_α and Φ_α (see Ex. 4.4) are similar

"$\Leftarrow$": omitted.

Application of Th. 4.6 : Peaks over threshold method

Let $x_1, \dots, x_n$ be i.i.d. realizations of r.v.'s X_i w.r.t. water levels at a dike during n days

In view of Th. 4.6 we are only interested in large observations $x_{i_1}, \dots, x_{i_k}$ which exceed a threshold u

Then

$$y_1 = x_{i_1} - u, \dots, y_k = x_{i_k} - u$$

are the realizations of i.i.d r.v.'s $Y_1, \dots, Y_k$ given by

$$Y_i := X_{\tilde{\tau}_i} - u,$$

where $\tilde{\tau}_j$ are the exceedance times

$$\tilde{\tau}_1 := \inf\{\ell \geq 1 : X_\ell > u\}$$

$$\vdots$$

$$\tilde{\tau}_k := \inf\{\ell > \tilde{\tau}_{k-1} : X_\ell > u\}$$

See Exerc. 7, Prob. 1

In Exerc. 7, Prob. 1 it is shown that

$X_{\tilde{\tau}_i}, i \geq 1$ is i.i.d. with

$$P(X_{\tilde{\tau}_i} \leq x) = P(X_i \leq x \mid X_i > u), \quad x > u$$

So $P(Y_i \leq x) = F_u(x), \quad x > u$

Since the conditions on $F = F_{X_i}$ in Th. 4.6 are rather "mild" we may assume for large u that

$$F_u(x) = P_{\gamma, \beta(u)}(x) \qquad (4.10)$$

$\longrightarrow$ we can interprete $y_1, \dots, y_k$ as i.i.d realizations of a generalized Pareto distribution

$\longrightarrow$ using $Y_1, \dots, Y_K$ we can apply <u>MLE</u> to <u>estimate the parameters</u> $\gamma, \beta = \beta(u)$ of the distr. $P_{\gamma,\beta}$:

The prob. density of $P_{\gamma,\beta}$ is given by

$$f_{\gamma,\beta}(x) = \begin{cases} \frac{1}{\beta}\left(1+\frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}-1} & , x > 0 \text{ , if } \gamma > 0 \\ \frac{1}{\beta}\left(1+\frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}-1} & , 0 \le x \le -\frac{\beta}{\gamma} \text{ , if } \gamma < 0 \\ \beta^{-1} e^{-x/\beta} & , x \in \mathbb{R} \text{ , if } \gamma = 0 \end{cases}$$

$\longrightarrow$ log-likelihood function:

$$l(\gamma,\beta) = \sum_{i=1}^{K} \log\left(f_{\gamma,\beta}(Y_i)\right)$$

$$= -K\log\beta - \left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{K}\log\left(1+\frac{\gamma Y_i}{\beta}\right)$$

for $\gamma \neq 0$

$\longrightarrow$ ML estimator: $(\hat{\gamma},\hat{\beta}) = \underset{\gamma,\beta}{\operatorname{argmax}}\, l(\gamma,\beta)$