Analysis methods of heavy-tailed data

Natalia Markovich

Institute of Control Sciences
Russian Academy of Sciences, Moscow, Russia

June 27, 2022

Special References:

1. **Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004)** *Statistics of Extremes: Theory and Applications*. Wiley, Chichester, West Sussex.
2. **Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997)** *Modeling Extremal Events*. Springer, Berlin.
3. **Markovich, N.M. (2007)** *Nonparametric Analysis of Univariate Heavy-Tailed data: Research and Practice*. Wiley, Chichester, West Sussex.
4. **Resnick, S.I. (2006)** *Heavy-Tail Phenomena. Probabilistic and Statistical Modeling*. Springer, New York.
5. **Any basic course of probability theory and statistics**.

Required Knowledge:

Mathematical analysis, probability theory and statistics.

The course contains

- Practical exercises
- Theoretical exercises
- Control questions

Modul 1

The Modul 1 contains the introduction

with necessary definitions, basic properties and examples of heavy-tailed data. The tail index indicates the shape of the tail and therefore it is the basic characteristic of heavy-tailed data. Methods of tail index estimation are presented.
Finally, several rough tools for the detection of heavy-tails, the number of finite moments and dependence are considered.

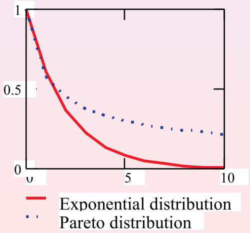
Definition of heavy-tailed distributions

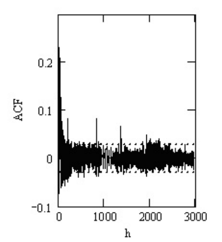
Let $X_1, \ldots, X_n$ be a sample of independent, identically distributed (i.i.d.) r.v.s X_i governed by the distribution function (DF) $F(x) = \mathbb{P}\{X \leq x\}$ with probability density function (PDF) $f(x) = dF(x)/dx$.

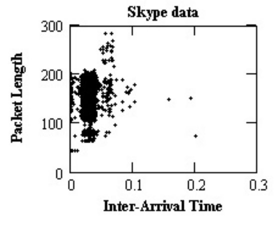
Definition

A DF $F(x)$ (or the r.v. X) is called heavy-tailed if its tail $\overline{F}(x) = 1 - F(x) > 0, \qquad x \geq 0$ satisfies for all $y > 0$

$$\lim_{x\to\infty} \mathbb{P}\{X > x + y | X > x\} = \lim_{x\to\infty} \overline{F}(x+y)/\overline{F}(x) = 1.$$







Examples of heavy- and light-tailed distributions

Heavy-tailed distributions:	**Subexponential:** Pareto, Lognormal, Weibull with shape parameter less than 1. **With regularly varying tails:** Pareto, Cauchy, Burr, Frechét, Zipf-Mandelbrot law. **Super heavy-tailed:** log-Cauchy
Light-tailed distributions	exponential, gamma, Weibull with shape parameter more than 1, normal, finite distributions.

Example of non-heavy tailed distribution: Exponential distribution $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$ satisfies

$$\bar{F}(x+y)/\bar{F}(x) = e^{-\lambda(x+y)}/e^{-\lambda x} = e^{-\lambda y} \nrightarrow 1$$

as $x \to \infty$, $x \geq 0$, $y > 0$

Exercise 1:

Prove that normal distribution is not heavy-tailed.

Heavy-tailed distributions have been accepted as realistic models for various phenomena:

- WWW-session characteristics
  - sizes and durations of sub-sessions; sizes of responses
  - inter-response time intervals
- on/off-periods of packet traffic
- file sizes
- service-time in queueing model
- flood levels of rivers
- major insurance claims
- extreme levels of ozon concentrations
- high wind-speed values
- wave heights during a storm
- low and high temperatures

Basic definitions and results: Max-stable law

Let $X^n = \{X_1, \ldots, X_n\}$ be a sample of i.i.d. r.v. distributed with the DF $F(x)$ and $M_n = \max(X_1, X_2, \ldots, X_n)$.

Gnedenko (1943): Extreme value theory assumes that for a suitable choice of normalizing constants $a_n > 0$, b_n real it holds

$$\mathbb{P}\{(M_n - b_n)/a_n \le x\} \quad = \quad F^n(b_n + a_n x) \to_{n\to\infty} H_\gamma(x), x \in R,$$

and an Extreme Value DF $H_\gamma(x)$ is of the following type:

$$H_\gamma(x) \quad = \quad \begin{cases} \exp(-x^{-1/\gamma}), & x > 0, \gamma > 0 \quad \text{'Fréchet'} \quad \Phi_\alpha(x), \\ \exp(-(-x)^{-1/\gamma}), & x < 0, \gamma < 0 \quad \text{'Weibull'} \quad \Psi_\alpha(x), \\ \exp(-e^{-x}), & \gamma = 0, x \in R \quad \text{'Gumbel'} \quad \Lambda(x). \end{cases}$$

Definition

The parameter γ is called the extreme value index (EVI) and defines the shape of the tail of the r.v. X.

The parameter $\alpha = 1/\gamma$ is called tail index.

Basic definitions and results: Max-stable law

The distribution $H_\gamma(x)$ can also be rewritten as

$$H_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}), & 1+\gamma x > 0, \quad \text{if } \gamma \neq 0 \\ \exp(-e^{-x}), & x \in R \quad \text{if } \gamma = 0. \end{cases}$$

Transformation of max-stable random variables

If $X > 0$ has a 'Fréchet' Φ_α distribution

- $\log X^\alpha$ has a 'Gumbel' distribution Λ
- $-X^{-1}$ has a 'Weibull' distribution Ψ_α

Exercise 2:

Prove these transformations

Basic definitions and results: Min-stable law

Since $\min_{1\le i\le n} X_i = -\max_{1\le i\le n}(-X_i)$ the min-stable law is determined by

$$G^*_\theta(x) \;=\; \begin{cases} 1-\exp(-(1-\theta x)^{-1/\theta}, & 1-\theta x>0, \text{ if } \quad \theta\ne 0 \\ 1-\exp(-e^x), & x\in R \quad \text{ if } \quad \theta=0. \end{cases}$$

θ is EVI for minima, measures the heaviness of the left-tail function $F(x)$, as $x\to-\infty$.

Basic results: Pickands's theorem

The limit distribution of the excess distribution of the i.i.d. X_i's is necessarily of the Generalized Pareto form

$$\lim_{u \uparrow x_F, u+x<x_F} \mathbb{P}(X_1 - u > x | X_1 > u) = \Psi_{\sigma,\gamma}(x), \qquad x \in R,$$

where

$$x_F = \sup\{x \in R : F(x) < 1\}$$

is the right endpoint of the distribution F and the shape parameter $\gamma \in R$.

Therefore, the Generalized Pareto distribution (GPD) with DF

$$\Psi_{\sigma,\gamma}(x) = \begin{cases} 1 - (1+\gamma x/\sigma)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - \exp(-x/\sigma), & \gamma = 0, \end{cases}$$

where $\sigma \geq 0$, $x \geq 0$ for $\gamma \geq 0$; $0 \leq x \leq -\sigma/\gamma$ for $\gamma < 0$, is often used as a **model of the tail** of the distribution.

Classification of distribution tails

1. $H_\gamma(x)$, $\gamma < 0 \hookrightarrow$ short tails with finite right endpoint (Beta, Uniform)

2. $H_\gamma(x)$, $\gamma = 0 \hookrightarrow$ exponentially decaying tails, light-tailed (Normal, Gamma) or moderate heavy-tailed (Lognormal)

$$\overline{F}(x) \sim \exp(-x), \qquad x \to +\infty$$

3. $H_\gamma(x)$, $\gamma > 0 \hookrightarrow$ polynomially decaying tails, heavy-tailed with infinite right endpoint
(Pareto, Cauchy, Student, Fréchet)

$$\overline{F}(x) \sim x^{-1/\gamma} = x^{-\alpha}, \qquad x \to +\infty$$

The classes of heavy-tailed distributions

- **distributions with regularly varying tails (*RVT*)**

($X \in \mathcal{R}_{-1/\gamma}$ or $X \in \mathcal{RV}_{-\alpha}$)

$$\mathbb{P}\{X > x\} = x^{-1/\gamma}\ell(x), \forall x > 0, \gamma > 0,$$

where $\ell(x)$ is a slowly varying function, i.e.

$$\lim_{x\to\infty} \ell(tx)/\ell(x) = 1, \qquad \forall t > 0.$$

Examples:

Pareto, Burr, Fréchet distributions belong to *RVT*.

Examples of $\ell(x)$ give $c > 0$, $c \ln x$, $c \ln(\ln x)$, $min(x, i)$ for $i \geq 1$

and all functions converging to positive constants.

Exercise 3:

Prove that $\ell(x) = \ln(\ln x)$ is slowly varying

Examples of not regularly varying functions

The following functions are not regularly varying

$$2 + \sin x, \qquad e^{[\ln(1+x)]}$$

By inequality

$$x^{-\alpha}\ell_1(x) \leq f(x) \leq x^{-\alpha}\ell_2(x)$$

it does not follow that $f(x)$ is regularly varying function.

Example: $\ell_1(x) = 1$, $\ell_2(x) = 3$, $f(x) = x^{-\alpha}(2 + \sin x)$.

The classes of heavy-tailed distributions

- **subexponential distributions (S) ($X \in \mathcal{S}$)**

  Let $X, X_1, ..., X_n$ be i.i.d. non-negative RV r.v.s.

  $$\mathbb{P}\{S_n > x\} \sim n\mathbb{P}\{X > x\} \sim \mathbb{P}\{M_n > x\} \quad \text{as} \quad x \to \infty,$$

  where $S_n = X_1 + ... + X_n$, $n \geq 2$, $M_n = \max_{i=1,...,n}\{X_i\}$.

Example:

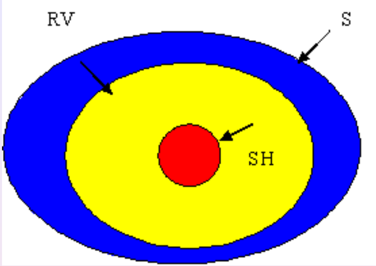
Weibull with shape parameter τ less than 1 belongs to S:

$$\overline{F}(x) = e^{-cx^{\tau}},$$

where $c > 0$, $0 < \tau < 1$.

Intuitively, subexponentiality means

that the only way the sum of i.i.d. r.v.s $X_1, ..., X_n$ can be large is by one of the summands getting large (in contrast, in the light-tailed case all summands are large if the sum is so).



Classes of heavy-tailed distributions:

subexponential (S), regularly varying (RV) and superheavy-tailed (SH)

Existence of distribution moments

- $\overline{F} \in RV_{-\alpha}$: heavy-tailed distributions have only finite moments of order $< \alpha$, $\alpha > 0$,
  $\lim_{t\to\infty} \overline{F}(tx)/\overline{F}(t) = x^{-\alpha}$ for any $x > 0$
- $\overline{F} \in RV_0$: super-heavy-tailed distributions have no finite moments of any order, $\alpha = 0$,
  $\lim_{t\to\infty} \overline{F}(tx)/\overline{F}(t) = 1$ for any $x > 0$, slowly varying tail

Examples of super-heavy tailed distributions

$$\overline{F}(x) = 1 - F(x) = x^{-1/\sqrt{\log x}}, \qquad x > 1 \tag{1}$$

$$\overline{F}(x) = (\log x)^{-\beta}, \qquad x \geq e, \ \beta > 0 \tag{2}$$

Cauchy distribution:

$$F(x) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}, \ \ x_0 \in R, \ \ \gamma > 0.$$

Exercise 4:

Prove that (1) and (2) are super-heavy-tailed, i.e. $\lim_{t\to\infty} \overline{F}(tx)/\overline{F}(t) = 1$ for any $x > 0$ is fulfilled.

Exercise 5:

Find other examples of super-heavy tailed distributions.

Property of super-heavy tailed distributions

Super-heavy tailed distribution does not need to belong to

the domain of attraction of any extreme value distribution $H_\gamma(x)$.

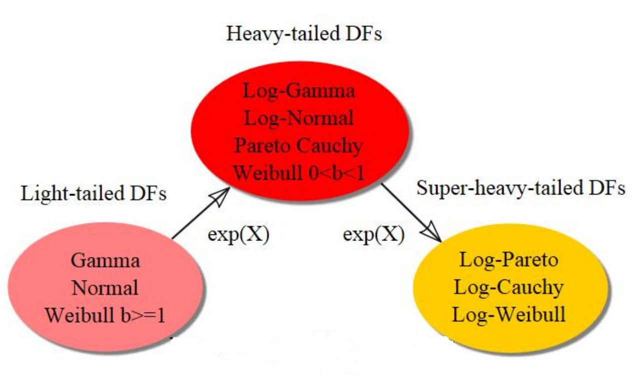
Let us consider this on example (1).
Since for any $c > 0$ we have $x^{-c}/(1 - F(x)) \to 0$ as $x \to \infty$, there are no normalizing sequences a_n and b_n such that

$$\mathbb{P}\{\max_{1 \le j \le n} X_j \le a_n x + b_n\} = F^n(a_n x + b_n) \to_{n\to\infty} H_\gamma(x), \qquad x \in R.$$

Gnedenko's theorem

is not fulfilled for super-heavy tailed distributions.

Transformations Between Classes of Distributions



Exercise 6:

If X is Weibull distributed with shape parameter $\tau \geq 1$. Will $Y = \exp X$ be heavy- or super-heavy-tailed distributed?

Basic properties of heavy-tailed distributions

Heavy-tailed distributions	Not all moments of the distribution exist or no one moment exists; Tail index $\alpha \geq 0$ The distribution function $F(x) < 1$ for any x; Infinite end-point $x_F = \sup\{x \in R : F(x) < 1\}$; Hazard rate $h(x) = f(x)/\overline{F(x)}$ tends to 0 as $x \to \infty$ $\overline{F(x)} \gg f(x) \gg f'(x) \gg ...$, where $\overline{F(x)} = 1 - F(x)$ is the tail function, $f(x) = F'(x)$ is the probability density function
Light-tailed distributions	All moments of the distribution exist; Tail index $\alpha < 0$ The distribution function $F(x) \leq 1$ for any x; For Weibull class end-point x_F is finite; For Gumbel class x_F is (in)finite Hazard rate $h(x)$ tends to ∞ as $x \to \infty$ or is constar $\overline{F(x)}$, $f(x)$, $f'(x)$,... have the same magnitude.

Basic properties of regularly varying distributions:

Lemma

Let $X \in \mathcal{R}_{-\alpha}$. *Then,*

(i) $X \in S$.

(ii) $\mathbb{E}\{X^{\beta}\} < \infty$ *if* $\beta < \alpha$, $\mathbb{E}\{X^{\beta}\} = \infty$ *if* $\beta \geq \alpha$.

(iii) *If* $\alpha > 1$, *then* $X^r \in \mathcal{R}_{1-\alpha}$ *and*

$$\mathbb{P}\{X^r > x\} \sim \ell(x)x^{1-\alpha}/((\alpha-1)\mathbb{E}\{X\}) \qquad \text{as} \qquad x \to \infty.$$

(iv) *If* Y *is non-negative and independent of* X *such that* $\mathbb{P}\{Y > x\} = \ell_2(x)x^{-\alpha_2}$, *then* $X + Y \in \mathcal{R}_{-\min(\alpha,\alpha_2)}$ *and* $\mathbb{P}\{X+Y > x\} \sim \mathbb{P}\{X > x\} + \mathbb{P}\{Y > x\}$ *as* $x \to \infty$.

(v) *If* Y *is non-negative and independent of* X *such that* $\mathbb{E}\{Y^{\alpha+\varepsilon}\} < \infty$ *for some* $\varepsilon > 0$ *then* $XY \in \mathcal{R}_{-\alpha}$ *and*

$$\mathbb{P}\{XY > x\} \sim \mathbb{E}\{Y^{\alpha}\}\mathbb{P}\{X > x\} \qquad \text{as} \qquad x \to \infty.$$

Important property for the rough detection of heavy tails and the number of finite moments:

Let $X \in \mathcal{R}_{-\alpha}$.

Then $\mathbb{E}\{X^\beta\} < \infty$, if $\beta < 1/\gamma$; $\mathbb{E}\{X^\beta\} = \infty$, if $\beta > 1/\gamma$.

Examples:

- If $\alpha = 2$, $\gamma = 0.5$, then
  $\mathbb{E}X_1 < \infty$ (the first moment is finite),
  $\mathbb{E}X_1^2 = \infty$ (the second moment is infinite, i.e. it does not exist).
- If $\alpha = 0.5$, $\gamma = 2$, then
  $\mathbb{E}X_1 = \infty$, $\mathbb{E}X_1^2 = \infty$... (all moments are infinite, i.e. they do not exist).

Estimators of tail index

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$$

are order statistics of the sample $X^n = \{X_1, X_2, ..., X_n\}$

For $\gamma > 0$:

- **Hill's estimator**

$$\hat{\gamma}^H(n,k) = \frac{1}{k}\sum_{i=1}^{k} \ln X_{(n-i+1)} - \ln X_{(n-k)}$$

- **Ratio estimator** Goldie, Smith, (1987)

$$a_n = a_n(x_n) = \sum_{i=1}^{n} \ln(X_i/x_n) I\{X_i > x_n\} / \sum_{i=1}^{n} I\{X_i > x_n\},$$

x_n is am arbitrary threshold level, e.g., $x_n = X_{(n-k)}$

Bias reduced modification of the Hill's estimator

The Hill's estimator is biased, i.e. $E\widehat{\gamma}^H(n,k) - \gamma \neq 0$. A bias reduced modification is **the generalized Jackknife estimator**

$$\widehat{\gamma}_k^{GJ} = 2\widehat{\gamma}_k^V - \widehat{\gamma}^H(n,k),$$

where $\widehat{\gamma}^H(n,k)$ is the Hill's estimator of the extreme value index $\gamma = 1/\alpha$,

$$\widehat{\gamma}_k^V = \frac{M_{n,k}}{2\widehat{\gamma}^H(n,k)}, \; M_{n,k} = \frac{1}{k}\sum_{i=1}^{k} Y_{(i,k)}^2, \; Y_{(i,k)} = \log(\frac{X_{(n-i+1)}}{X_{(n-k)}}).$$

is proposed in Gomes et al. (2000)[a]

[a] Gomes, I., Martins, J., Neves, M. Alternatives to a Semi-Parametric Estimator of Parameters of Rare Events - The Jackknife Methodology. Extremes (2000) 3, 207-229

Estimators of tail index

For $\gamma \in R$:

- **"Moment estimator",** Dekkers, Einmahl, de Haan, (1989):

$$\hat{\gamma}^M(n,k) = \hat{\gamma}^H(n,k) + 1 - 0.5\left(1 - (\hat{\gamma}^H(n,k))^2/S_{n,k}\right)^{-1},$$

$$S_{n,k} = (1/k)\sum_{i=1}^{k}\left(\ln X_{(n-i+1)} - \ln X_{(n-k)}\right)^2$$

- **"UH estimator",** Berlinet, (1998)

$$\hat{\gamma}^{UH}(n,k) = (1/k)\sum_{i=1}^{k}\ln UH_i - \ln UH_{k+1},\ UH_i = X_{(n-i)}\hat{\gamma}^H(n,i)$$

- **Pickands's estimator**

$$\hat{\gamma}^P(n,k) = \frac{1}{\ln 2}\ln\frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{(n-4k+1)}},\ k \le n/4$$

Group estimator, Davydov, Paulauskas, Račkauskas, (2000):

The sample X^n is divided into l groups $V_1, ..., V_l$, each group containing m r.v.s, i.e. $n = l \cdot m$.

Estimator of the function of the tail index:

$$z_l = (1/l) \sum_{i=1}^{l} k_{li} = \frac{\hat{\alpha}}{\hat{\alpha} + 1} = \frac{1}{1 + \hat{\gamma}},$$

where

$$k_{li} = M_{li}^{(2)} / M_{li}^{(1)}, \qquad M_{li}^{(1)} = \max\{X_j : X_j \in V_i\}$$

and $M_{li}^{(2)}$ is the second largest element in the same group V_i.

Group estimator, Davydov, Paulauskas, Račkauskas, (2000).Theoretical background.

1. For distributions with regularly varying tails

$$1 - F(x) = x^{-\alpha}\ell(x),$$

and $l = m = [\sqrt{n}]$, it is proved

$$z_l \to^{a.s.} \frac{\alpha}{\alpha+1} = \frac{1}{1+\gamma}.$$

2. For distributions

$$1 - F(x) = C_1 x^{-\alpha} + C_2 x^{-\beta} + o(x^{-\beta}),$$

with $\beta = 2\alpha$ it holds

$$l(l^{-1}\sum_{i=1}^{l} k_{l,i} - \alpha(1+\alpha)^{-1}) \left(\sum_{i=1}^{l} k_{l,i}^2 - l^{-1}\left(\sum_{i=1}^{l} k_{l,i}\right)^2\right)^{-1/2} \to^d N(0,1) \quad (3)$$

Confidence interval of group estimate

Exercise 7:

Using (3) obtain the confidence interval of α.

Recursiveness of the estimate z_l. On-line estimation.

Having the next group of observations V_{l+1} it follows

$$\begin{aligned}\gamma_{l+1} &= \left(\frac{1}{l+1}\sum_{i=1}^{l+1} k_{l+1,i}\right)^{-1} - 1 \\ &= \left(\frac{l}{l+1}\cdot\frac{1}{1+\gamma_l} + \frac{k_{l+1,l+1}}{l+1}\right)^{-1} - 1\end{aligned}$$

After getting i additional groups with m elements each $V_{l+1}, ..., V_{l+i}$

$$\gamma_{l+i} = (l+i)\left(\frac{l}{1+\gamma_l} + k_{l+1,l+1} + ... + k_{l+i,l+i}\right)^{-1} - 1, \qquad (4)$$

i.e.

$\boxed{\gamma_{l+i} \text{ is obtained using } \gamma_l \text{ by } O(1) \text{ calculations.}}$

Since

$$z_{l+i} = 1/(1+\gamma_{l+i})$$

it holds from (4) that

$$z_{l+i} = \left(l z_l + \sum_{j=1}^{i} k_{l+j,l+j} \right) /(l+i),$$

The bias of z_{l+i} is the same as for z_l assuming the process is weak-sense stationary ($Ek_{ij} = const$, $\forall i, j$), but the variance is less if $\{k_{li}\}$ are uncorrelated

$$bias(z_{l+i}) = bias(z_l), \qquad var(z_{l+i}) = var(z_l)l/(l+i), \tag{5}$$

$$var(z_{l+i}) < var(z_l) \qquad \text{for} \qquad \forall i > 0$$

Exercise 8:

Prove (5).

What estimator of the tail index is the best?

It is difficult to compare the estimators of γ.

One can only look at the asymptotic variances and biases of estimates for known distributions.

Example

For Pareto tail the moment estimator is unbiased for any γ, but the variance of this estimate is larger than the variance of the Hill's estimate. Besides, it is known that

$$\sqrt{k}\left(\hat{\gamma}^M(n,k)-\gamma\right)\to^d$$

$$\begin{cases} N(0,1+\gamma^2), & \gamma\ge 0,\\ N(0,(1-\gamma)^2(1-2\gamma)\left(4-8\frac{1-2\gamma}{1-3\gamma}+\frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right)), & \gamma<0. \end{cases}$$

What estimator of the tail index is the best?

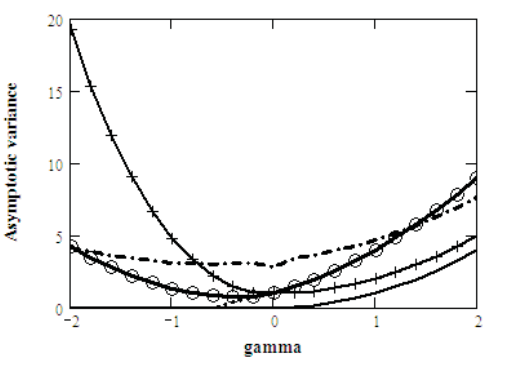


Figure1.4. Asymptotic variances of $\sqrt{k}(\hat{\gamma} - \gamma)$ for the Hill's, moment, *UH*, Pickands' and the POT ML estimators: *VarHill* (solid line), *VarM* (solid line marked by $+$), *VarUH* (solid line marked by $\circ$), *VarP* $(- \cdot - \cdot -)$ and *VarMLP* $(\cdot \cdot \cdot \cdot \cdot)$.

Modul 1: Lesson 4

Methods for the selection of the number of the largest order statistics in Hill estimator.

The visual selection of the number of the largest order statistics in Hill estimator

- **A Hill plot** $\{(k,\hat{\gamma}^H(n,k)), 1 \le k \le n-1\})$:
  the estimate of $\hat{\gamma}^H(n,k)$ is chosen from an interval in which this function demonstrate stability.
- **Plot of the mean excess function** $\{(u, e(u)) : X_{(1)} < u < X_{(n)}\}$**,** where

$$e(u) = \mathbb{E}(X - u | X > u), \qquad 0 \le u < x_F \le \infty,$$

$$e_n(u) = \sum_{i=1}^{n} (X_i - u)\mathbf{1}\{X_i > u\} / \sum_{i=1}^{n} \mathbf{1}\{X_i > u\}$$

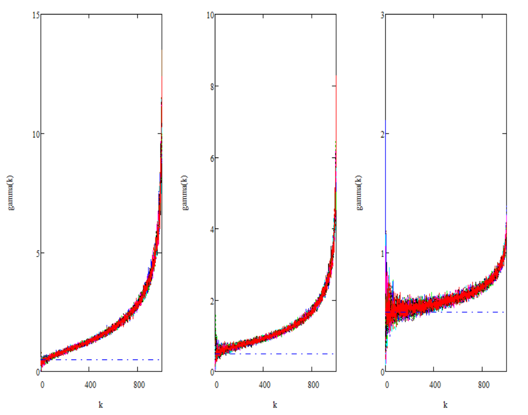
  is the **sample mean excess function** over the threshold u. If this plot follows a reasonably straight line above a certain value of u, then this indicates that excesses over u follow $e^P(u) = (1+\gamma u)/(1-\gamma)$ of generalized Pareto distribution with positive shape parameter. The number of the nearest order statistics to u is used as the estimate of k.

Mean excess function of Pareto distribution

Exercise 9:

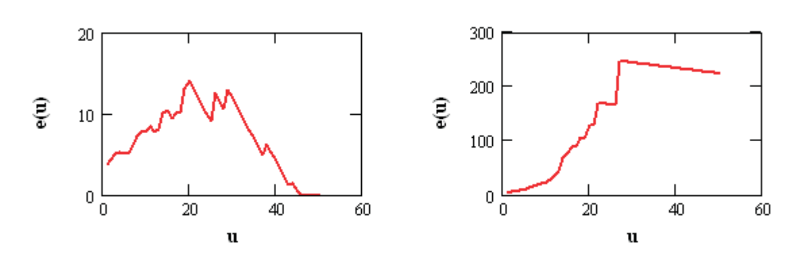
Prove $e^P(u)$.

The sensitivity of the Hill estimate to k.



The Hill's estimate against k for 15 samples of Weibull (left), Pareto (middle) and Frechét (right) distributions, all with shape parameter $\alpha = 1/\gamma = 0.5$. Sample size is $n = 1000$.

Plot of the mean excess function.



Left: Weibull distribution. Right: Pareto distribution. The shape parameter is 0.5**, sample size** 1000**.**

Bootstrap method for automatic selection of k.

- Minimizing of the empirical bootstrap estimate of the mean squared error of γ:

$$\boxed{MSE(\hat{\gamma}_k) = bias^2 + variance = \mathbb{E}\left(\hat{\gamma}_k - \gamma\right)^2 \to \min_k .}$$

$$bias = \mathbb{E}\hat{\gamma}_k - \gamma, \qquad variance = \mathbb{E}(\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k)^2$$

Since γ is unknown we have to substitute MSE by its bootstrap estimate.

We have to select such k that

real MSE and its bootstrap analog will be close

Exercise 10:

Prove $MSE(\hat{\gamma}) = bias^2 + variance$.

Bootstrap method for k selection.

- The bootstrap estimate is obtained by drawing B samples with replacement from the original data set X^n.
- Smaller re-samples $\{X_1^*, ..., X_{n_1}^*\}$ of the size

  $$n_1 = n^\beta, \qquad 0 < \beta < 1,$$

  are used.
- The corresponding smaller k_1 and an optimal k are related by:

  $$k = k_1(n/n_1)^\alpha, \qquad 0 < \alpha < 1,$$

  where $\beta = 1/2$ and $\alpha = 2/3$ for $\overline{F}(x) = C_0 x^{-1/\gamma} + C_1 x^{-2/\gamma} + o(x^{-2/\gamma})$ and the Hill's estimator, **P.Hall, (1990)**.

  Such k provides the minimum of $MSE(\hat{\gamma})$.

Empirical bootstrap estimate of the $MSE(\hat{\gamma})$ is

$$MSE^*(n_1,k_1) = \left(\hat{b}^*(n_1,k_1)\right)^2 + \widehat{var}^*(n_1,k_1) \to \min_{k_1},$$

where

$$\hat{b}^*(n_1,k_1) = \frac{1}{B}\sum_{b=1}^{B}\hat{\gamma}_b^*(n_1,k_1) - \hat{\gamma}(n,k),$$

$$\widehat{var}^*(n_1,k_1) = \frac{1}{B-1}\sum_{b_1=1}^{B}\left(\hat{\gamma}_{b_1}^*(n_1,k_1) - \frac{1}{B}\sum_{b_2=1}^{B}\hat{\gamma}_{b_2}^*(n_1,k_1)\right)^2$$

are the **empirical bootstrap estimates of the bias and the variance,**

$\hat{\gamma}_b^*$ is the Hill's estimate constructed by some re-sample of the size n_1 with the parameter k_1.

Double bootstrap method for k selection

Danielsson, de Haan, Peng and de Vries, (1997)

requires less parameters than bootstrap method, Hall, (1990): n_1 **and** B **are required,** α **is not required.**

Auxiliary statistic:

$$z_{n,k} = M_{n,k} - 2(\hat{\gamma}^H(n,k))^2,$$

where

$$M_{n,k} = \frac{1}{k}\sum_{j=1}^{k}\left(\log X_{(n-j+1)} - \log X_{(n-k)}\right)^2$$

$M_{n,k}/2\hat{\gamma}^H(n,k)$ and $\hat{\gamma}^H(n,k)$ are consistent estimates for γ,

$\Longrightarrow \qquad z_{n,k} \to 0 \qquad \text{as} \qquad n \to \infty$, since

$$\frac{z_{n,k}}{2\hat{\gamma}^H(n,k)} = \frac{M_{n,k}}{2\hat{\gamma}^H(n,k)} - \hat{\gamma}^H(n,k) \to \gamma - \gamma = 0$$

$\Longrightarrow \qquad AMSE(z_{n,k}) = \mathbb{E}(z_{n,k})^2 \to \min_k,$

The value $\widehat{k}_n^{opt}$ of k, which minimizes $AMSE(z_{n,k})$[1], has the same order in n as k_n^{opt} that minimizes $AMSE(\widehat{\gamma}^H_{n,k})$.

[1] AMSE is an asymptotic mean squared error

Double bootstrap procedure is

- Draw B bootstrap sub-samples of the size $n_1 \in (\sqrt{n}, n)$ (e.g., $n_1 \sim n^{3/4}$) from the original sample and determine the value $\hat{k}^\star_{n_1}$ that minimizes MSE of $z_{n_1,k}$.
- Repeat this for B sub-samples of the size $n_2 = [n_1^2/n]$ ($[x]$ is the integer part of the number) and determine the value $\hat{k}^\star_{n_2}$ that minimizes MSE of $z_{n_2,k}$.
- Calculate $\hat{k}_n^{opt}$ from the formula

$$\hat{k}_n^{opt} = [\frac{(\hat{k}^\star_{n_1})^2}{\hat{k}^\star_{n_2}}\left(1 - \frac{1}{\hat{\rho}_1}\right)^{\frac{2}{2\hat{\rho}_1 - 1}}], \qquad \hat{\rho}_1 = \frac{\log \hat{k}^\star_{n_1}}{2\log(\hat{k}^\star_{n_1}/n_1)},$$

and estimate γ by the Hill's estimate with $\hat{k}_n^{opt}$.

The method is robust with respect to the choice of n_1, Gomes, Oliveira, (2000).

Sequential procedure for k selection

is based on the theoretical result:

$$\sqrt{i}(\hat{\gamma}^H(n,i) - \gamma) \sim (\log\log n)^{1/2}, \qquad 2 \leq i \leq k$$

in probability, **Drees & Kaufmann, (1998).**[a]

[a] $f(n) \sim g(n)$ denotes $\mathbb{P}\{\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1\} = 1$

Sequential procedure for k selection. Algorithm.

- An initial estimate $\hat{\gamma}_0 = \hat{\gamma}^H(n, 2\sqrt{n})$ for the parameter γ is obtained by the Hill's estimate.
- For $r_n = 2.5\hat{\gamma}_0 n^{0.25}$ we compute

$$\hat{k}(r_n) = \min\{k \in 2, \ldots, n-1 | \max_{2 \le i \le k} \sqrt{i}(\hat{\gamma}^H(n,i) - \hat{\gamma}^H(n,k)) > r_n\}.$$

  If r_n is too large and $\max_{2 \le i \le k} \sqrt{i}(\hat{\gamma}^H(n,i) - \hat{\gamma}^H(n,k)) > r_n$ is not satisfied it is recommended repeatedly replace r_n by $0.9r_n$ until $\hat{k}(r_n)$ is well defined.
- Similarly, $\hat{k}(r_n^\varepsilon)$ for $\varepsilon = 0.7$ is computed.
- Optimal k

$$\hat{k}^{opt} = \frac{1}{3}\left(\frac{\hat{k}(r_n^\varepsilon)}{(\hat{k}(r_n))^\varepsilon}\right)^{1/(1-\varepsilon)} (2\hat{\gamma}_0)^{1/3}$$

  is calculated and γ is estimated by $\hat{\gamma}^H(n, \hat{k}^{opt})$.

The method is sensitive to the choice of r_n.

Eye-Ball method

By Eye-Ball method the first stability interval is found using a moving window. The number of the largest order statistics is

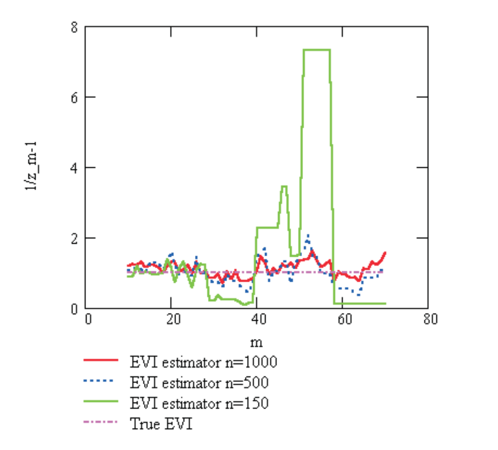
$$k^*_{eye} \quad = \quad \min\{k \in 2,...,n^+ - \omega | h < \frac{1}{\omega}\sum_{i=1}^{\omega} \mathbf{1}\{\hat{\alpha}_{n,k+i} < \hat{\alpha}_{n,k} \pm \varepsilon\}\}.$$

ω is the size of moving window, e.g. 1% of the full sample. n^+ is the number of positive observations in the data. Not less than h% of the estimates should be within the bounds $\hat{\alpha}_{n,k} \pm \varepsilon$ ($\hat{\alpha}_{n,k} = 1/\hat{\gamma}^H(n,k)$). One can take $h = 90\%$ and $\varepsilon = 0.3$.

The Eye-Ball principle can be applied to other estimators of the tail index not only the Hill's one.[2]

[2] Danielsson, J., Ergun, L.M., de Haan, L., De Vries, C. Tail Index Estimation: Quantile Driven Threshold Selection. SSRN Electronic Journal (2016)

Plot method for m selection in the group estimator.



The plot $\{(m, 1/z_m - 1)\}$ for Pareto distribution with $\gamma = 1$, the true γ is shown by dotted line. Sample sizes are $n = \{150, 500, 1000\}$.

Plot method for m selection in the group estimator.

Plot: $\{(m, z_m), m_0 < m < M_0\}$, $m_0 > 2$, $M_0 < n/2$

$$m = n/l,\ z_m = (m/n)\sum_{i=1}^{[n/m]} k_{(n/m)_i}$$

From consistency result $z_l \to^{a.s.} \frac{1}{1+\gamma}$ it follows that

there must be an interval $[m_-, m_+]$

such that $z_m \approx \alpha/(1+\alpha) = (1+\gamma)^{-1}$ for all $m \in [m_-, m_+]$.

We suggest choosing the average value

$\overline{z} = mean\{1/z_m - 1 : m \in [m_-, m_+]\}$

and $m^* \in [m_-, m_+]$ as a point such that $z_{m^*} = \overline{z}$.

Bootstrap method for m automatical selection.

- Minimizing of the empirical bootstrap estimate of the mean squared error of $(1+\gamma_l)^{-1}$, $l = n/m$:

$$MSE(\gamma_l) = \mathbb{E}\left(\frac{1}{l}\sum_{i=1}^{l} k_{l_i} - \frac{1}{1+\gamma}\right)^2 \to \min_m .$$

- The bootstrap estimate is obtained by drawing B samples with replacement from the original data set X^n.
- Smaller re-samples $\{X_1^*, ..., X_{n_1}^*\}$ of the size
  $n_1 = n^d, \qquad 0 < d < 1,$
  are used.
- The re-sample is divided into l_1 subgroups.
- The size of subgroups m_1 and m are related by:
  $m = m_1(n/n_1)^c, \qquad 0 < c < 1,$
  where m_1 is the size of subgroups in re-samples.

Empirical bootstrap estimate of the MSE

$$MSE^*(l_1,m_1) = \left(\hat{b}^*(l_1,m_1)\right)^2 + \widehat{var}^*(l_1,m_1) \to \min_{m_1},$$

where

$$\hat{b}^*(l_1,m_1) = \frac{1}{B}\sum_{b=1}^{B} z_{l_1}^b - z_l,$$

$$\widehat{var}^*(l_1,m_1) = \frac{1}{B-1}\sum_{b_1=1}^{B}\left(z_{l_1}^{b_1} - \frac{1}{B}\sum_{b_2=1}^{B} z_{l_1}^{b_2}\right)^2$$

are the **empirical bootstrap estimates of the bias and the variance,**
$z_{l_1}^b = \frac{1}{l_1}\sum_{i=1}^{l_1} k_{l_{1i}}$ is the group estimator constructed by some re-sample.

How to select c and d?

Simulation study: the selection of c and d.

- Asymptotic theory (P.Hall, (1990)) recommends

$$d = 1/2 \qquad \text{and} \qquad c = 2/3$$

for the bootstrap estimation of the parameter k of the Hill's estimate of γ.

- We check $c = \{0.05, 0.1(0.1); 0.5\}$ for a fixed $d = 0.5$. Samples of the Pareto, Fréchet and Weibull distributions with known γ were generated.

Relative bias and the square root of the mean squared error:

$$Bias\gamma = \frac{1}{\gamma}\left(\frac{1}{N_R}\sum_{i=1}^{N_R}\hat{\gamma}_i - \gamma\right),$$

$$RMSE\gamma = \frac{1}{\gamma}\sqrt{\frac{1}{N_R}\sum_{i=1}^{N_R}(\hat{\gamma}_i - \gamma)^2}$$

Conclusions:

- the best values of c for the fixed $d = 0.5$ are $c = \{0.3 \div 0.5\}$;
- the bias of the group estimator is larger for Weibull distribution.

Further research:

- proof of theoretically best values of c and d for the group estimator using the bootstrap.

Derivation of the Hill's estimator, Hill (1975)

Assume:

$$\overline{F}(x) = \mathbb{P}\{X > x\} = x^{-\alpha}, \qquad x \geq 1$$

Then the r.v. $Y = \ln X$ has the tail function

$$\mathbb{P}\{Y > y\} = \mathbb{P}\{\ln X > y\} = \mathbb{P}\{X > e^{y}\} = e^{-\alpha y}, \qquad y \geq 0,$$

i.e. Y is exponentially distributed with DF $G(y) = 1 - e^{-\alpha y}$ and PDF $g(y) = \alpha e^{-\alpha y}$

Maximum likelihood estimator:

$$\begin{aligned}
\ln \mathcal{L}(\alpha | X_1, ..., X_n) &= \sum_{i=1}^{n} \ln g(Y_i|\alpha) = \sum_{i=1}^{n} (\ln \alpha - \alpha Y_i) \\
&= \sum_{i=1}^{n} (\ln \alpha - \alpha \ln X_i)
\end{aligned}$$

Maximum likelihood estimator:

$$\ln' \mathcal{L}(\alpha|X_1, ..., X_n) = \sum_{i=1}^{n} (1/\alpha - \ln X_i) = 0$$

$$\frac{n}{\alpha} = \sum_{i=1}^{n} \ln X_i \qquad \Rightarrow \qquad \widehat{\alpha} = \left(\frac{1}{n} \sum_{i=1}^{n} \ln X_{(i)} \right)^{-1}$$

$X_{(1)} \leq ... \leq X_{(n)}$

A trivial generalization concerns

$$\overline{F}(x) = Cx^{-\alpha}, \qquad x \geq u > 0,$$

with u known. If $C = u^{\alpha}$ then $\overline{G}(y) = u^{\alpha} e^{-\alpha y}$, $g(y) = \alpha u^{\alpha} e^{-\alpha y}$

Maximum likelihood estimator:

$$\ln \mathcal{L}(\alpha|X_1,...,X_n) = \sum_{i=1}^{n}(\ln\alpha + \alpha\ln u - \alpha\ln X_i)$$

$$\ln' \mathcal{L}(\alpha|X_1,...,X_n) = \sum_{i=1}^{n}(1/\alpha + \ln u - \ln X_i) = 0$$

$$1/\alpha = 1/n\sum_{i=1}^{n}\ln X_{(i)} - \ln u$$

Taking $u = X_{(n-k)}$ and since $x \geq u$ we obtain the Hill's estimator

$$\hat{\gamma}^H(n,k) = \frac{1}{k}\sum_{i=1}^{k}\ln X_{(n-i+1)} - \ln X_{(n-k)}$$

Theoretical properties of the Hill's estimator

Mason (1982): Hill's estimator is weakly consistent if

$k \to \infty$, $k/n \to 0$ as $n \to \infty$

Häusler & Teugels (1985): Hill's estimator is asymptotically normal with mean γ and variance γ^2/k,

$$\sqrt{k}\left(\widehat{\gamma}^H(n,k) - \gamma\right) \to^d N(0,\gamma^2)$$

Let $X_*^{n_1} = \{X_1^*, ..., X_{n_1}^*\}$ be bootstrap re-sample.

Bootstrap estimator of the bias $\mathbb{E}\widehat{\gamma}^H - \gamma$:

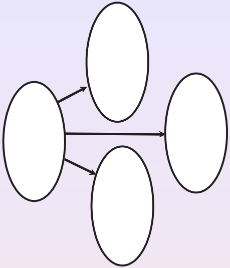
$$b^*(n_1, k_1) = \mathbb{E}\{\widehat{\gamma}^{*^H}(n_1, k_1)|X^n\} - \widehat{\gamma}^H(n, k)$$

Bootstrap estimator of the variance $var = \mathbb{E}\left(\widehat{\gamma}^H(n, k) - \mathbb{E}\widehat{\gamma}^H(n, k)\right)^2$:

$$var^*(n_1, k_1) = \mathbb{E}\{\left(\widehat{\gamma}^{*^H}(n_1, k_1) - \mathbb{E}\{\widehat{\gamma}^{*^H}(n_1, k_1)|X^n\}\right)^2 |X^n\}$$

X^n is fixed and the expectation is calculated among all theoretically possible re-samples $X_*^{n_1}$.
Bootstrap mean is not expectation in the usual sense since bootstrap "expectation" is random variable.

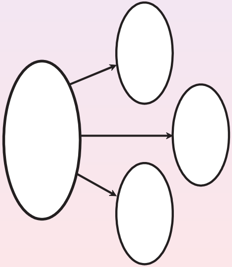
Types of Bootstrap



Classical bootstrap:

re-samples are of the same size as an original sample

It leads to underestimating of the bias



Non-classical bootstrap:

re-samples are of the smaller size than an original sample $n_1 < n$

It helps to avoid the bootstrap bias $= 0$ regardless the true bias $\neq 0$.

This is typical for linear estimates: linear regressions, kernel estimates of the density.

Examples of linear estimates

Kernel density estimator $1/(nh)\sum_{i=1}^{n} K((x-X_i)/h)$, linear regression

Let $\widehat{\theta}=\sum_{i=1}^{n}\varphi(X_i)$ be linear function built by sample $X_1,...,X_n$.
Let $\theta^* = \sum_{i=1}^{n}\varphi(X_i^*)$ be the same function built by re-sample $X_1^*,...,X_n^*$.
Then

$$\mathbb{E}(\theta^*|X^n) = n\mathbb{E}\{\varphi(X_i^*)|X^n\} = n\sum_{i=1}^{n}\frac{1}{n}\varphi(X_i) = \widehat{\theta},$$

since X_i^* may be selected by n ways from X^n.
$\Rightarrow$ bootstrap bias $\mathbb{E}(\theta^*|X^n)-\widehat{\theta}=0$, but the true bias $\mathbb{E}\widehat{\theta}-\theta\neq 0$

Next problem of classical bootstrap

Bickel and Sakov (2002): that the statistic

$$a_n(F_n)\left(\max(X_1^*, ..., X_n^*) - b_n(F_n)\right)$$

(a_n, b_n are the same normalized constants as in Gnedenko (1943) theorem) does not converge to $H_\gamma(x)$ for bootstrap with re-samples of the size n.
If the re-samples of smaller size $n_1 < n$ are used, $n_1 \to \infty$, $n_1/n \to 0$ and von Mises's condition

$$x\frac{f(x)}{1 - F(x)} \to_{x\to\infty} \frac{1}{\gamma}$$

is satisfied ($f(x)$ is probability density function), then

$$a_{n_1}(F_n)\left(\max(X_1^*, ..., X_{n_1}^*) - b_{n_1}(F_n)\right) \to H_\gamma(x).$$

Confidence intervals of bootstrap estimates

Assumptions

the bootstrap estimates $\gamma_1^*, ..., \gamma_B^*$ are normal distributed with the mean $Mean\gamma$ and standard deviation $StDev\gamma$, constructed by B bootstrap estimates,
B is the number of bootstrap re-samples.

Smirnov and Dunin-Barkovsky (1965): Tolerant bounds of confidence intervals

$$(u_1, u_2) = (Mean\gamma - \rho \cdot StDev\gamma; Mean\gamma + \rho \cdot StDev\gamma)$$

The interval is constructed in such a way that the $(1-p)$th part of the distribution falls into this interval with the probability P:

$$\rho = \rho_\infty \left(1 + \frac{t_p}{\sqrt{2B}} + \frac{5t_p^2 + 10}{12B}\right).$$

Confidence intervals of bootstrap estimates. Cont.

ρ_∞ is defined by the equation

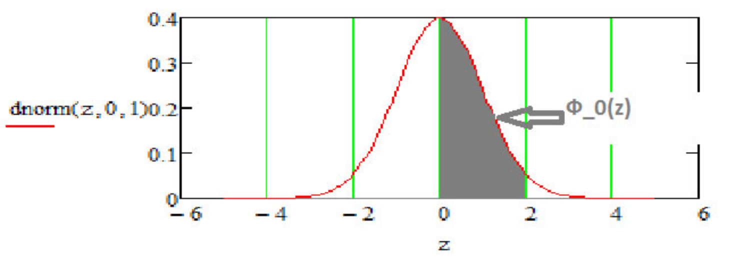
$$\frac{1}{\sqrt{2\pi}}\int_{-\rho_\infty}^{\rho_\infty} e^{-t^2/2}dt = 2\Phi_0(\rho_\infty) = 1 - p,$$

where $\Phi_0(z) = \frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}dt$ is Laplace's function.

Normal distribution function $N(z;0,1) = 0.5 + \Phi_0(z)$ for $z > 0$, $\Phi_0(-z) = -\Phi_0(z)$, $\Phi_0(0) = 0$, $\Phi_0(-\infty) = -0.5$, $\Phi_0(\infty) = 0.5$

t_p is calculated by the equation

$$\frac{1}{\sqrt{2\pi}}\int_{t_p}^{\infty} e^{-t^2/2}dt = 0.5 - \Phi_0(t_p) = 1 - P.$$



Example

For $P = 0.99$ we have $\Phi_0(t_p) = P - 0.5 = 0.49$
$t_p = 2.33$ is found from the table of Laplace's function.
$\rho_\infty \in \{2.245, 1.97, 1.645\}$ for $p \in \{0.025, 0.05, 0.1\}$, respectively.