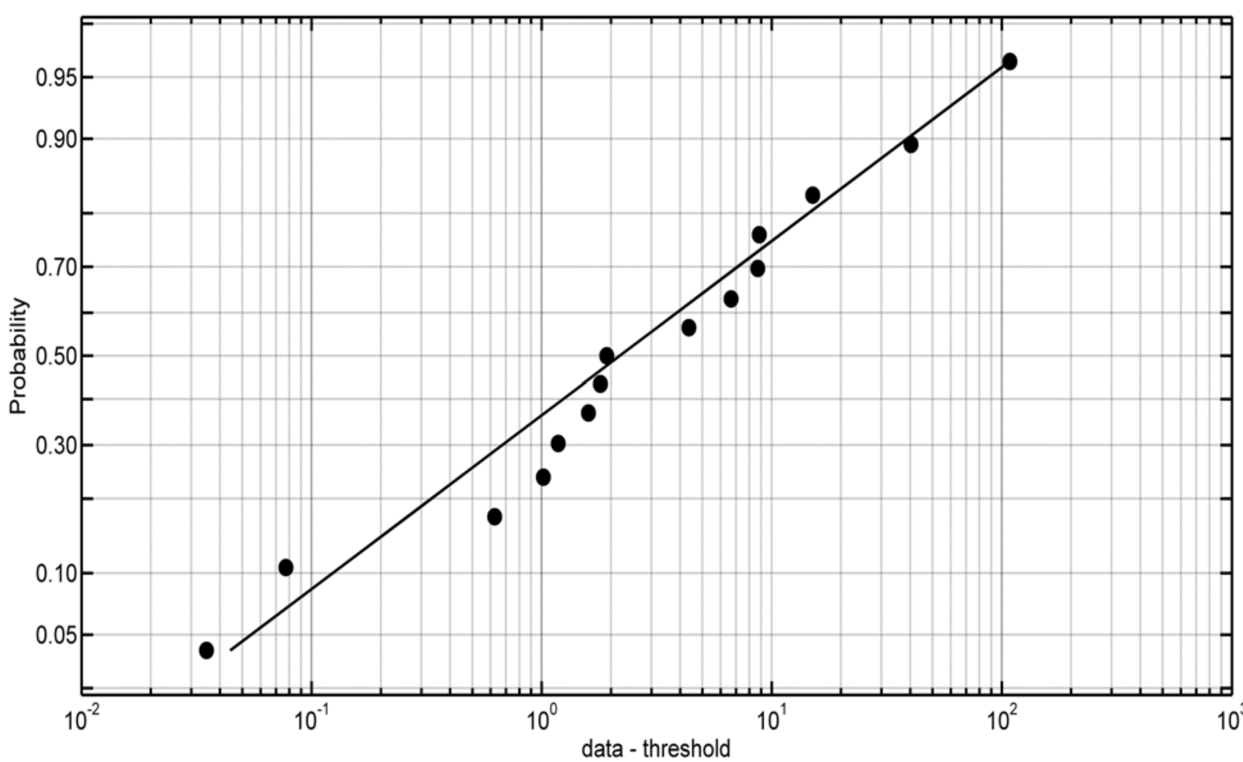
# Location–Scale Distributions

## Linear Estimation and Probability Plotting Using MATLAB

Horst Rinne



# Contents

# Preface

Statistical distributions can be grouped into families or systems. Such groupings are described in JOHNSON/KOTZ/KEMP (1992, Chapter 2), JOHNSON/KOTZ/BALAKRISHNAN (1994, Chapter 12) or PATEL(KAPADIA/OWEN (1976, Chapter 4). The most popular families are those of PEARSON, JOHNSON and BURR, the exponential, the stable and the infinitely divisible distributions or those with a monotone likelihood ratio or with a monotone failure rate. All these categories have attracted the attention of statisticians and they are fully discussed in the statistical literature. But there is one family, the location–scale family, which hitherto has not been discussed in greater detail. To my knowledge this book is the first comprehensive monograph on one–dimensional continuous location–scale distributions and it is organized as follows.

Chapter 1 goes into the details of location–scale distributions and gives their properties along with a short list of those distributions which are genuinely location–scale and which — after a suitable transformation of its variable — become member of this class. We will only consider the ln–transformation. Location–scale distributions easily lend themselves to an assessment by graphical methods. On a suitably chosen probability paper the cumulative distribution function of the universe gives a straight line and the cumulative distribution of a sample only deviates by chance from a straight line. Thus we can realize an informal goodness–of–fit test. When we fit the straight line free–hand or by eye we may read off the location and scale parameters as percentiles. Another and objective method is to find the straight line on probability paper by a least–squares technique. Then, the estimates of the location and scale parameters will be the parameters of that straight line.

Because probability plotting heavily relies on ordered observations Chapter 2 gives — as a prerequisite — a detailed representation of the theory of order statistics. Probability plotting is a graphical assessment of statistical distributions. To see how this kind of graphics fits into the framework of statistical graphics we have written Chapter 3.

A first core chapter is Chapter 4. It presents the theory and the methods of linear estimating the location and scale parameters. The methods to be implemented depend on the type of sample, i.e. grouped or non–grouped, censored or uncensored, the type of censoring and also whether the moments of the order statistics are easily calculable or are readily available in tabulated form or not. In the latter case we will give various approximations to the optimal method of general least–squares.

Applications of the exact or approximate linear estimation procedures to a great number of location–scale distributions will be presented in Chapter 5, which is central to this book. For each of 35 distributions we give a warrant of arrest enumerating the characteristics, the underlying stochastic model and the fields of application together with the pertinent probability paper and the estimators of the location parameter and the scale parameter. Distributions which have to be transformed to location–scale type sometimes have a third parameter which has to be pre–estimated before applying probability plotting and the linear estimation procedure. We will show how to estimate this third parameter.

The calculations and graphics of Chapter 5 have been done using MATLAB,[1] Version 7.4 (R2007a). The accompanying CD contains the MATLAB script M–file LEPP and all the function–files to be used by the reader when he wants to do inference on location–scale distributions. Hints how to handle the menu–driven program LEPP and how to organize the data input will be given in Chapter 6 as well as in the comments in the files on the CD.

Dr. HORST RINNE, Professor Emeritus of Statistics and Econometrics

Department of Economics and Management Science

Justus–Liebig–University, Giessen, Germany

[1] MATLAB® is a registered trade–mark of 'The MathWorks, Inc.'

# List of Figures

# List of Tables

# 1 The family of location–scale distributions

Statistical distributions can be grouped into families or systems such that all members of a family

- share the same properties and/or
- are constructed according to the same principles and/or
- possess the same structure.

For example, we have the PEARSON system, the JOHNSON system and the BURR system, the exponential family, the family of stable distributions and especially the location–scale family. The latter family is a good candidate for applying graphical procedures and linear estimation when the variate is **continuous**.

## 1.1 Properties of location–scale distributions

A random variable $X$ is said to belong to the **location–scale family** when its **cumulative distribution** (= CDF)

$$F_X(x \mid a, b) := \Pr(X \le x \mid a, b) \tag{1.1a}$$

is a function only of $(x-a)/b$:

$$F_X(x \mid a, b) = F\left(\frac{x-a}{b}\right);\quad a \in \mathbb{R},\ b > 0; \tag{1.1b}$$

where $F(\cdot)$ is a distribution having no other parameters. Different $F(\cdot)$'s correspond to different members of the family. The two–dimensional parameter $(a, b)$ is called a location–scale parameter, $a$ being the location parameter and $b$ being the scale parameter. For fixed $b = 1$ we have a subfamily which is a **location family** with parameter $a$, and for fixed $a = 0$ we have a **scale family** with parameter $b$. The variable

$$Y := \frac{X-a}{b} \tag{1.1c}$$

is called the **reduced variable**,[1] $y$ being a realization of $Y$. The reduced variable $Y$ has $a = 0$ and $b = 1$, and we will write the **reduced CDF** as

$$F_Y(y) := F\left(\frac{x-a}{b}\right). \tag{1.1d}$$

[1] Some authors call it the **standardized variable**. We will refrain from using this name because, conventionally, a standardized variable is defined as $Z = \big[X - \mathrm{E}(X)\big]\big/\sqrt{\mathrm{Var}(X)}$ and thus has mean $\mathrm{E}(Z) = 0$ and variance $\mathrm{Var}(Z) = 1$. The normal distribution, which is a member of the location–scale family, is the only distribution with $\mathrm{E}(X) = a$ and $\mathrm{Var}(X) = b^2$. So, in this case, reducing and standardizing are the same.

If the distribution of $X$ is absolutely continuous with the **density function** (= DF)

$$f_X(x\,|\,a,b) = \frac{\mathrm{d}F_X(x\,|\,a,b)}{\mathrm{d}\,x} \tag{1.2a}$$

then $(a, b)$ is a location scale–parameter for the distribution of $X$ if (and only if)

$$f_X(x\,|\,a,b) = \frac{1}{b}\,f_Y\left(\frac{x-a}{b}\right) \tag{1.2b}$$

for some density $f_Y(\cdot)$, called **reduced DF**.

The **location parameter** $a \in \mathbb{R}$ is responsible for the distribution's position on the abscissa. An enlargement (reduction) of $a$ causes a movement of the distribution to the right (left). The location parameter is either a **measure of central tendency** of a distribution, e.g.:

- $a$ is the **mean, median and mode** for a symmetric and unimodal distribution as for example with the **normal distribution** having DF

$$f_X(x\,|\,a,b) = \frac{1}{b\,\sqrt{2\,\pi}}\,\exp\left\{-\frac{(x-a)^2}{2\,b^2}\right\},\;\; x\in\mathbb{R},\; a\in\mathbb{R},\; b>0. \tag{1.3}$$

- $a$ is the **median and mode** for a symmetric distribution as for example with the **CAUCHY distribution** having DF

$$f_X(x\,|\,a,b) = \left\{\pi\,b\left[1+\left(\frac{x-a}{b}\right)^2\right]\right\}^{-1},\;\; x\in\mathbb{R},\; a\in\mathbb{R},\; b>0. \tag{1.4}$$

- $a$ is the **mode** for an asymmetric and unimodal distribution as for example with the **extreme value distribution of type I for the maximum** having DF

$$f_X(x\,|\,a,b) = \frac{1}{b}\,\exp\left\{-\frac{x-a}{b} - \exp\left(-\frac{x-a}{b}\right)\right\},\;\; x\in\mathbb{R},\; a\in\mathbb{R},\; b>0. \tag{1.5}$$

or it is the **threshold parameter** of a distribution, e.g.

- $a$ is the **lower threshold** of the **exponential distribution** having DF

$$f_X(x\,|\,a,b) = \frac{1}{b}\,\exp\left(-\frac{x-a}{b}\right),\;\; x\ge a,\; a\in\mathbb{R},\; b>0. \tag{1.6}$$

- $a$ is the **upper threshold** of the **reflected exponential distribution** having DF

$$f_X(x\,|\,a,b) = \frac{1}{b}\,\exp\left(-\frac{a-x}{b}\right),\;\; x\le a,\; a\in\mathbb{R},\; b>0. \tag{1.7}$$

The parameter $b,\ b>0$, is the **scale parameter**. It is responsible for the dispersion or variation of the variate $X$. Increasing (decreasing) $b$ results in an enlargement (reduction) of the spread and a corresponding reduction (enlargement) of the density. $b$ may be

- the **standard deviation** of $X$ as for example with the normal distribution,
- the **length of the support**[2] of $X$ as for example with the **uniform distribution** having DF
$$f_X(x\,|\,a,b) = \frac{1}{b},\;\; a \le x \le a+b,\; a\in\mathbb{R},\; b>0, \tag{1.8}$$
- the length of a **central $(1-\alpha)$–interval**[3] for $X$ as for example with the extreme value distribution of type I for the maximum, see (1.5),
$$b \approx x_{0.667169} - x_{0.332831},\;\; \alpha \approx 0.334338. \tag{1.9}$$

All distributions in a given family have the same shape, i.e. the same skewness and the same kurtosis. When the reduced variable $Y$ has mean $\mu_Y = \mathrm{E}(Y)$ and standard deviation $\sigma_Y = \sqrt{\mathrm{Var}(Y)}$ then, in the general case, the mean of $X$ is
$$\mathrm{E}(X) = a + b\,\mu_Y, \tag{1.10a}$$
and the standard deviation of $X$ is
$$\sqrt{\mathrm{Var}(X)} = b\,\sigma_Y. \tag{1.10b}$$
For $\mu_Y = 0$ and $\sigma_Y = 1$ we have the goal expectation $a$ and the goal standard deviation $b$, but this is not necessarily the case. It is possible, for example, that $\mu_Y$ and $\sigma_Y$ may not exist, as in the case of the CAUCHY distribution.

We have a lot of functions and parameters describing and measuring certain features of a variate. When we know such a function or parameter for the reduced variate $Y$, the corresponding function and parameter for the general variable $X = a + b\,Y$ follow in an easy way as is depicted in Tab. 1/1.

Table 1/1: Relations between functions and parameters of the reduced and the general variates of a continuous location–scale distribution

| Name | Definition for $Y$ | Relation for $X = a + b\,Y$ |
|---|---|---|
| density function (DF) | $f_Y(y) := \dfrac{\mathrm{d}F_Y(y)}{\mathrm{d}\,y}$ | $f_X(x) = \dfrac{1}{b}\,f_Y\!\left(\dfrac{x-a}{b}\right)$ |
| cumulative distribution function (CDF) | $F_Y(y) := \int\limits_{-\infty}^{y} f_Y(u)\,\mathrm{d}u$ | $F_X(x) = F_Y\!\left(\dfrac{x-a}{b}\right)$ |
| reliability function (CCDF) | $R_Y(y) := 1 - F_Y(y)$ | $R_X(x) = 1 - R_Y\!\left(\dfrac{x-a}{b}\right)$ |
| hazard function (HF) | $h_Y(y) := \dfrac{f_Y(y)}{R_Y(y)}$ | $h_X(x) = \dfrac{1}{b}\,h_Y\!\left(\dfrac{x-a}{b}\right)$ |

[2] For some symmetric location–scale distributions it is more convenient to have a parametrization which results in a length of $2\,b$ or a multiple of $2\,b$.

[3] Such an interval excludes the $\alpha/2$ smallest and the $\alpha/2$ largest realizations of $X$, where $\alpha < 1$.

| Name | Definition for $Y$ | Relation for $X = a + b\,Y$ |
|---|---|---|
| cumulative hazard function (CHF) | $H_Y(y) := \int\limits_{-\infty}^{y} h_Y(u)\,\mathrm{d}u$ | $H_X(x) = H_Y\left(\frac{x-a}{b}\right)$ |
| percentile, generally: $0 \le P \le 1$ | $y_P := F_Y^{-1}(P)$ | $x_P = a + b\,y_P$ |
| percentile distance, $0 \le P_1 < P_2 \le 1$ | $\mathrm{PD}_Y(P_2 - P_1) := y_{P_2} - y_{P_1}$ | $\mathrm{PD}_X(P_2 - P_1) = b\,\mathrm{PD}_Y(P_2 - P_1)$ |
| mode | $y_M$ such that $f_Y(y_M) = \max_y f_Y(y)$ | $x_M = a + b\,y_M$ |
| crude moments generating function | $M_Y(t) := \mathrm{E}\big(e^{t\,Y}\big)$ | $M_X(t) = \exp(a\,t)\,M_Y(b\,t)$ |
| crude moments, $r \in \mathbb{N}_0$ | $\mu'_r(Y) := \mathrm{E}\big(Y^r\big) = \left.\frac{\mathrm{d}^r M_Y(t)}{\mathrm{d}\,t^r}\right\vert_{t=0}$ | $\mu'_r(X) = \sum\limits_{j=0}^{r} \binom{r}{j} \mu'_{r-j}(Y)\,b^{r-j}\,a^j$ |
| mean | $\mu_Y := \mu'_1(Y) := \mathrm{E}(Y)$ | $\mu_X := \mu'_1(X) = a + b\,\mu_Y$ |
| central moments generating function | $Z_Y(t) := \mathrm{E}\Big[e^{t\,(Y-\mu_Y)}\Big]$ | $Z_X(t) = Z_Y(b\,t)$ |
| central moments | $\mu_r(Y) := \mathrm{E}\big[(Y-\mu_Y)^r\big] = \left.\frac{\mathrm{d}^r Z_Y(t)}{\mathrm{d}\,t^r}\right\vert_{t=0}$ | $\mu_r(X) = b^r\,\mu_r(Y)$ |
| variance | $\sigma_Y^2 := \mathrm{Var}(Y) := \mathrm{E}\big[(Y - \mu_Y)^2\big]$ | $\sigma_X^2 := \mathrm{Var}(X) = b^2\,\sigma_Y^2$ |
| standard deviation | $\sigma_Y := \sqrt{\mathrm{Var}(Y)}$ | $\sigma_X = b\,\sigma_Y$ |
| index of skewness | $\alpha_3(Y) := \frac{\mu_3(Y)}{\big[\mu_2(Y)\big]^{3/2}}$ | $\alpha_3(X) = \alpha_3(Y)$ |
| index of kurtosis | $\alpha_4(Y) := \frac{\mu_4(Y)}{\big[\mu_2(Y)\big]^2}$ | $\alpha_4(X) = \alpha_4(Y)$ |
| cumulants generating function | $K_Y(t) := \ln M_Y(t)$ | $K_X(t) = a\,t + K_Y(b\,t)$ |
| cumulants | $\kappa_r(Y) := \left.\frac{\mathrm{d}K_Y(t)}{\mathrm{d}t^r}\right\vert_{t=0}$ | $\kappa_1(X) = a + b\,\kappa_1(Y)$<br>$\kappa_r(X) = b^r\,\kappa_r(Y),\ r = 2, 3, \ldots$ |
| characteristic function | $C_Y(t) := \mathrm{E}\big(e^{i\,t\,Y}\big),\ \ i := \sqrt{-1}$ | $C_X(t) = \exp\,(i\,t\,a)\,C_Y(b\,t)$ |
| entropy[†] | $I(Y) := -\mathrm{E}\big\{\mathrm{ld}\big[f_Y(y)\big]\big\}$ | $I(X) = \mathrm{ld}(b) + I(Y)$ |
| LAPLACE transform | $L_Y(t) := \mathrm{E}\big(e^{-t\,Y}\big)$ | $L_X(t) = \exp\,(-a\,t)\,L_Y(b\,t)$ |

[†] ld(.) is the binary logarithm (logarithm with base 2). Some authors give the entropy in terms of the natural logarithm. Binary and natural logarithms are related as $\mathrm{ld}(x) = \ln(x)\big/\ln 2 \approx 1.4427\,\ln(x)$.

## 1.2 Genuine location–scale distributions — A short listing[4]

In this section we list — in alphabetic order — those continuous distributions which are directly of location–scale type. There also exist distributions which — after suitable transformation — are of location–scale type. They will be presented in Sect. 1.3. The following listing only gives the DF of the non–reduced variate $X$. A complete description of each distribution including all the parameters and functions of Tab. 1/1 will be given in Chapter 5 where we present the accompanying probability paper together with some auxiliary functions which are useful for the linear estimation procedure of the location–scale parameter.

**Arc–sine distribution**

$$\begin{aligned} f(x|a,b) &= \frac{1}{\pi\sqrt{b^2-(x-a)^2}} \\ &= \frac{1}{b\,\pi\sqrt{1-\left(\dfrac{x-a}{b}\right)^2}},\quad a-b \le x \le a+b,\ a\in\mathbb{R},\ b>0. \end{aligned} \tag{1.11a}$$

The name of this distribution is given by the fact that its CDF can be expressed — among others — by the arc–sine function $\arcsin(x) = \sin^{-1}(x)$:

$$F(x\,|\,a,b) = \frac{1}{2} + \frac{\arcsin\left(\dfrac{x-a}{b}\right)}{\pi}. \tag{1.11b}$$

**Beta distribution** (Special cases)

The beta distribution in its general form has DF

$$\left.\begin{aligned} f(x\,|\,a,b,c,d) &= \frac{\left(\dfrac{x-a}{b}\right)^{c-1}\left(1-\dfrac{x-a}{b}\right)^{d-1}}{b\,B(c,d)} \\ &= \frac{(x-a)^{c-1}(a+b-x)^{d-1}}{b^{c+d-1}\,B(c,d)} \end{aligned}\right\}\left\{\begin{aligned} &a\le x\le a+b, \\ &a\in\mathbb{R},\ b>0, \\ &c>0,\ d>0, \end{aligned}\right\} \tag{1.12a}$$

with the **complete beta function**

$$B(c,d) := \int_0^1 u^{c-1}\,(1-u)^{d-1}\,\mathrm{d}u = \frac{\Gamma(c)\,\Gamma(d)}{\Gamma(c+d)}\,, \tag{1.12b}$$

[4] Suggested reading for this and the following section: JOHNSON/KOTZ/BALAKRISHNAN (1994, 1995), PATEL/KAPADIA/OWEN (1976).

is not of location–scale type because it depends on two extra parameters $c$ and $d$ which are responsible for the shape of the DF. When these parameters are given special values we will arrive at a location–scale distribution, e.g.

- $c = d = 1$ gives the **uniform distribution** or **rectangular distribution**, see (1.33).
- $c = 1,\ d = 2$ gives the **positively skew right–angled triangular distribution**, see (1.32c).
- $c = 2,\ d = 1$ gives the **negatively skew right–angled triangular distribution**, see (1.32d).
- $c = d = 0.5$ gives an **U–shaped distribution**, see (5.52a).
- $c = 1$ and $d = 3, 4, \ldots$ gives a **power–function distribution of order $d - 1$**.
- $d = 1$ and $c = 3, 4, \ldots$ gives a **power–function distribution of order $c - 1$**, see (1.27a).

**CAUCHY distribution**

The CAUCHY distribution has DF

$$f(x \,|\, a, b) = \left\{ \pi\, b \left[ 1 + \left( \frac{x-a}{b} \right)^2 \right] \right\}^{-1}, \quad x \in \mathbb{R},\ a \in \mathbb{R},\ b > 0. \tag{1.13}$$

**Chi–distribution** (Special cases)

The $\chi$–distribution with DF

$$f(x|a,b,\nu) = \frac{1}{b\, 2^{(\nu/2)-1}\Gamma(\nu/2)} \left( \frac{x-a}{b} \right)^{\nu-1} \exp\left\{ -\frac{1}{2} \left( \frac{x-a}{b} \right)^2 \right\} \left\{ \begin{matrix} x \in \mathbb{R}, a \in \mathbb{R}, \\ b > 0, \nu \in \mathbb{N}, \end{matrix} \right\} \tag{1.14}$$

is not of location–scale type, but for given $\nu$ it is. We get

- the **half–normal distribution** for $\nu = 1$, see (1.20),
- the **RAYLEIGH distribution** for $\nu = 2$, see (1.28),
- the **MAXWELL–BOLTZMANN distribution** for $\nu = 3$, see (1.24).

**Cosine distribution**

The cosine distribution has DF

$$f(x \,|\, a, b) = \frac{1}{2\, b} \cos\left( \frac{x-a}{b} \right), \quad a - b\, \frac{\pi}{2} \leq x \leq a + b\, \frac{\pi}{2},\ a \in \mathbb{R},\ b > 0. \tag{1.15a}$$

Because $\cos u = \sin\left(1 + \dfrac{\pi}{2}\right)$ we can write the cosine distribution as **sine distribution**:

$$f(x\,|\,a,b) = \frac{1}{2\,b}\,\sin\left(\frac{\pi}{2} + \frac{x-a}{b}\right),\;\; a - b\,\frac{\pi}{2} \le x \le a + b\,\frac{\pi}{2}. \tag{1.15b}$$

**Exponential distribution**

The DF of this very popular distribution is given by

$$f(x\,|\,a,b) = \frac{1}{b}\,\exp\left(-\frac{x-a}{b}\right),\;\; x \ge a,\; a \in \mathbb{R},\; b > 0. \tag{1.16}$$

**Extreme value distributions**

Extreme value distributions are the limiting distributions of either the largest or the smallest value in a sample of size $n$ for $n \to \infty$. We have three types for each of the two cases (largest or smallest observation). Only the type–I distributions, which are of **GUMBEL–type** (EMIL JULIUS GUMBEL, 1891 – 1966) are of location–scale type:

- **Extreme value distribution of type I for the maximum**

$$f(x\,|\,a,b) = \frac{1}{b}\,\exp\left\{-\frac{x-a}{b} - \exp\left(-\frac{x-a}{b}\right)\right\},\; x \in \mathbb{R},\; a \in \mathbb{R},\; b > 0. \tag{1.17a}$$

  This is often called *the* extreme value distribution by some authors.

- **Extreme value distribution of type I for the minimum**

$$f(x\,|\,a,b) = \frac{1}{b}\,\exp\left\{\frac{x-a}{b} - \exp\left(\frac{x-a}{b}\right)\right\},\;\; x \in \mathbb{R},\; a \in \mathbb{R},\; b > 0. \tag{1.17b}$$

The type–II and the type–III extreme value distributions can be transformed to type–I distributions, see Sect. 1.3.

**Half–CAUCHY distribution**

This distribution results when the CAUCHY distribution (1.13) is folded around its location parameter $a$ so that the left–hand part for $x < a$ is added to the right–hand part $(x > a)$.

$$f(x\,|\,a,b) = 2\left\{\pi\,b\left[1 + \left(\frac{x-a}{b}\right)\right]\right\}^{-1},\;\; x \ge a,\; a \in \mathbb{R},\; b > 0. \tag{1.18}$$

**Half–logistic distribution**

The half–logistic distribution results from the logistic distribution (1.23) in the same way as the half–CAUCHY distribution is derived from the CAUCHY distribution:

$$f(x\,|\,a,b) = \frac{2\,\exp\left(-\dfrac{x-a}{b}\right)}{b\left[1 + \exp\left(-\dfrac{x-a}{b}\right)\right]^2},\; x \ge a,\; a \in \mathbb{R},\; b > 0. \tag{1.19}$$

### Half–normal distribution

The half–normal distribution results from normal distribution (1.25) like the half–CAUCHY distribution from the CAUCHY distribution as

$$f(x\,|\,a,b) = \frac{1}{b}\sqrt{\frac{2}{\pi}}\,\exp\left\{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right\},\;\; x \geq a,\; a \in \mathbb{R},\; b > 0. \tag{1.20}$$

### Hyperbolic secant distribution

The DF is

$$f(x\,|\,a,b) = \frac{1}{b\,\pi}\,\mathrm{sech}\left(\frac{x-a}{b}\right),\;\; x \in \mathbb{R},\; a \in \mathbb{R},\; b > 0. \tag{1.21}$$

### LAPLACE distribution

This distribution is a bilateral or two–tailed exponential distribution with DF

$$f(x\,|\,a,b) = \frac{1}{2\,b}\,\exp\left\{-\left|\frac{x-a}{b}\right|\right\},\;\; x \in \mathbb{R},\; a \in \mathbb{R},\; b > 0. \tag{1.22}$$

### Logistic distribution

This distribution can be written in different ways, see Sect. 5.2.9, one being

$$f(x\,|\,a,b) = \frac{\exp\left(\dfrac{x-a}{b}\right)}{b\left[1+\exp\left(\dfrac{x-a}{b}\right)\right]^2},\;\; x \in \mathbb{R},\; a \in \mathbb{R},\; b > 0. \tag{1.23}$$

### MAXWELL–BOLTZMANN distribution

This distribution is a special case of the $\chi$–distribution with $\nu = 3$:

$$f(x\,|\,a,b) = \frac{1}{b}\,\sqrt{\frac{2}{\pi}}\left(\frac{x-a}{b}\right)^2\exp\left\{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right\},\;\; x \geq a,\; a \in \mathbb{R},\; b > 0. \tag{1.24}$$

### Normal distribution

This well–known distribution has DF[5]

$$f(x\,|\,a,b) = \frac{1}{b\sqrt{2\,\pi}}\,\exp\left\{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right\},\;\; x \in \mathbb{R},\; a \in \mathbb{R},\; b > 0. \tag{1.25}$$

[5] We remark the $a = \mu(X) = \mathrm{E}(X)$ and $b^2 = \sigma_X^2 = \mathrm{Var}(X)$.

**Parabolic distributions of order 2**

We have two types, one has an **U–form** with DF

$$f(x\,|\,a,b) = \frac{3}{2\,b}\left(\frac{x-a}{b}\right)^2, \;\; a-b \le x \le a+b, \; a \in \mathbb{R}, \; b > 0, \tag{1.26a}$$

the other one has an **inverted U–form** with DF

$$f(x\,|\,a,b) = \frac{3}{4\,b}\left[1-\left(\frac{x-a}{b}\right)^2\right], \;\; a-b \le x \le a+b, \; a \in \mathbb{R}, \; b > 0. \tag{1.26b}$$

**Power–function distribution** (Special cases)

The power–function distribution with DF

$$f(x\,|\,a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1}, \;\; a \le x \le a+b, \; a \in \mathbb{R}, \; b > 0, \; c > 0, \tag{1.27a}$$

and CDF

$$F(x\,|\,a,b,c) = \left(\frac{x-a}{b}\right)^c \tag{1.27b}$$

is not of location–scale type. We can transform (1.27a,b) to a location–scale DF by considering

$$\widetilde{X} = \ln(X-a)$$

to arrive at the **reflected exponential distribution**, see (1.7) and (1.29). We also have a location–scale distribution when $c$ is known, e.g.:

- $c = 1$ gives an **uniform distribution** or **rectangular distribution**.
- $c = 2$ gives a **right–angled negatively skew triangular distribution**.
- $c = 3$ gives a **parabolic distribution of order 2**, but with a support of length $b$.

**RAYLEIGH distribution**

This distribution is a special case of the WEIBULL distribution, see (1.45a), with shape parameter $c = 2$:

$$f(x\,|\,a,b) = \frac{1}{b}\left(\frac{x-a}{b}\right)\exp\left\{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right\} \;\; x \ge a, \; a \in \mathbb{R}, \; b > 0. \tag{1.28}$$

**Reflected exponential distribution**

When the exponential distribution is reflected around $x = a$ we get the DF

$$f(x\,|\,a,b) = \frac{1}{b}\left(-\frac{a-x}{b}\right), \;\; x \le a, \; a \in \mathbb{R}, \; b > 0. \tag{1.29}$$

**Semi–elliptical distribution**

The graph of the DF

$$f(x\,|\,a,b) = \frac{1}{b\,\pi}\sqrt{1-\left(\frac{x-a}{b}\right)^2},\;\; a-b \le x \le a+b,\; a\in\mathbb{R},\; b>0 \tag{1.30}$$

is a semi–ellipse centered at $(a,0)$. For $b=\sqrt{2/\pi}\approx 0.7979$ we will have a semi–circle.

**TEISSIER distribution**

The DF of this distribution, named after the French biologist G. TEISSIER and published in 1934, is given by

$$f(x|a,b)=\frac{1}{b}\left\{\exp\left(\frac{x-a}{b}\right)-1\right\}\exp\left\{1+\frac{x-a}{b}-\exp\left(\frac{x-a}{b}\right)\right\},\; x\ge a, a\in\mathbb{R}, b>0 \tag{1.31}$$

**Triangular Distributions**

We have several types of triangular distributions:

- a **symmetric** version with DF

$$f(x\,|\,a,b) = \left\{\begin{array}{ll} \dfrac{x-a+b}{b^2}, & a-b\le x\le a, \\ \dfrac{a+b-x}{b^2}, & a\le x\le a+b, \end{array}\right\},\;\; a\in\mathbb{R},\; b>0, \tag{1.32a}$$

  or equivalently written as

$$f(x\,|\,a,b) = \frac{b-|x-a|}{b^2},\;\; a-b\le x\le a+b, \tag{1.32b}$$

- a **right–angled and positively skew** version with DF

$$f(x\,|\,a,b) = \frac{2}{b}\left(\frac{a+b-x}{b}\right),\; a\le x\le a+b,\; a\in\mathbb{R},\; b>0, \tag{1.32c}$$

- a **right–angled and negatively skew** version with DF

$$f(x\,|\,a,b) = \frac{2}{b}\left(\frac{x-a+b}{b}\right),\;\; a-b\le x\le a,\; a\in\mathbb{R},\; b>0, \tag{1.32d}$$

- **asymmetric** versions which have — besides $a$ and $b$ — a third parameter indicating the mode, thus they are not of location–scale type.

**Uniform** or **rectangular distribution**

This rather simple, but very important distribution — see Sect. 2.2.2 — has DF

$$f(x\,|\,a,b) = \frac{1}{b},\;\; a \le x \le a+b,\; a \in \mathbb{R},\; b > 0. \tag{1.33}$$

**U–shaped and inverted U–shaped distributions**

All parabolic distributions of order $k = 2\,m,\; m \in \mathbb{N}$, have an **U–shape** and are of location–scale type when $m$ is known. The DF is

$$f(x\,|\,a,b,m) = \frac{2\,m+1}{2\,b}\left(\frac{x-a}{b}\right)^{2\,m},\;\; a-b \le x \le a+b,\; a \in \mathbb{R},\; b > 0. \tag{1.34}$$

We arrive at a distribution whose graph has an **inverted U–shape** with DF

$$f(x\,|\,a,b) = \frac{2\,m+1}{4\,m\,b}\left[1-\left(\frac{x-a}{b}\right)^{2\,m}\right],\;\; a-b \le x \le a+b,\; a \in \mathbb{R},\; b > 0. \tag{1.35}$$

We will only study the case $m = 1$ which gives the parabolic distributions of order 2, see (1.26a,b).

**V–shaped distribution**

A symmetric V–shaped DF is given by

$$f(x\,|\,a,b) = \left\{\begin{array}{ll} \dfrac{a-x}{b^2}, & a-b \le x \le a, \\[2ex] \dfrac{x-a}{b^2}, & a \le x \le a+b \end{array}\right\},\;\; a \in \mathbb{R},\; b > 0, \tag{1.36a}$$

or equivalently written as

$$f(x\,|\,a,b) = \frac{|x-a|}{b^2},\;\; a-b \le x \le a+b. \tag{1.36b}$$

## 1.3 Distributions transformable to location–scale type

We start by giving the rules governing the transformation of a variate $X$ to another random variable $\widetilde{X}$.

Theorem: Let $X$ be continuously distributed with DF $f_X(x)$, CDF $F_X(x)$, mean $\mu_X = \mathrm{E}(X)$, variance $\sigma_X^2 = \mathrm{Var}(X)$ and percentiles $x_P,\; 0 \le P \le 1$, and let $\widetilde{x} = g(x)$ be a transforming function which is a one–to–one mapping over the entire range of $X$ and thus is monotonic so that $x = g^{-1}(\widetilde{x})$ exists. Furthermore $g(x)$ has to be differentiable twice. Then

$$\widetilde{X} = g(X)$$

has the DF

$$f_{\widetilde{X}}(\widetilde{x}) = f_X\left[g^{-1}(\widetilde{x})\right]\left|\frac{\mathrm{d}g^{-1}(\widetilde{x})}{\mathrm{d}\widetilde{x}}\right|, \tag{1.37a}$$

the CDF

$$F_{\widetilde{X}}(\widetilde{x}) = \left\{\begin{array}{ll} F_X\left[g^{-1}(\widetilde{x})\right] & \text{for} \quad g'(x) > 0 \\ 1 - F_X\left[g^{-1}(\widetilde{x})\right] & \text{for} \quad g'(x) < 0 \end{array}\right\}, \tag{1.37b}$$

the percentiles

$$\widetilde{x}_P = \left\{\begin{array}{ll} g(x_P) & \text{for} \quad g'(x) > 0 \\ g(x_{1-P}) & \text{for} \quad g'(x) < 0 \end{array}\right\}, \tag{1.37c}$$

the approximative mean

$$\mu_{\widetilde{X}} = \mathrm{E}(\widetilde{X}) \approx g(\mu_X) + \frac{\sigma_X^2}{2}\, g''(\mu_X), \tag{1.37d}$$

and the approximative variance

$$\sigma_{\widetilde{X}}^2 = \mathrm{Var}(\widetilde{X}) \approx \sigma_X^2\left[g'(\mu_X)\right]^2. \tag{1.37e}$$

(1.37e) should be used only when $\sigma_X \big/ \mu_X \ll 1$. ∎

We will concentrate on the most popular transformation, the **ln–transformation**, for rendering a non–location–scale distribution to a location–scale distribution. In this case, when the original variable $X$ has a location parameter $a \neq 0$, we either have to know its value, what rarely is the case, or we have to estimate it before forming $\ln(x-a)$ or $\ln(a-x)$. In Sect. 5.3.1 we will give several estimators of $a$.

**Extreme value distributions of type II and type III**

- The extreme value distribution of **type II for the maximum**, sometimes referred to as **FRÉCHET–type distribution** has DF

$$f(x\,|\,a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{-c-1}\exp\left\{-\left(\frac{x-a}{b}\right)^{-c}\right\}\left\{\begin{array}{l} x \geq a,\; a \in \mathbb{R}, \\ b > 0,\; c > 0, \end{array}\right\} \tag{1.38a}$$

  and CDF

$$F(x\,|\,a,b,c) = \left\{\begin{array}{ll} 0 & \text{for} \quad x < a, \\ \exp\left\{-\left(\dfrac{x-a}{b}\right)^{-c}\right\} & \text{for} \quad x \geq a. \end{array}\right\} \tag{1.38b}$$

  Forming

$$\widetilde{x} = g(x) = \ln(x-a) \tag{1.39a}$$

we first have

$$x = g^{-1}(\widetilde{x},) = \exp(\widetilde{x}) + a \tag{1.39b}$$

and

$$\frac{\mathrm{d}g^{-1}(\widetilde{x})}{\mathrm{d}\widetilde{x}} = \exp(\widetilde{x}). \tag{1.39c}$$

The CDF of $\widetilde{X}$ easily follows from (1.37b) with (1.38b) as

$$F(\widetilde{x}\,|\,b,c) = \exp\left\{-\left(\frac{\exp(\widetilde{x})}{b}\right)^{-c}\right\},\ \ \widetilde{x} = \ln(x-a) \in \mathbb{R}. \tag{1.40a}$$

Using the identity

$$b \equiv \exp(\ln b),$$

(1.40a) first can be written as

$$\begin{aligned} F(\widetilde{x}\,|\,\widetilde{b},\widetilde{c}) &= \exp\left\{-\left[\frac{\exp(\widetilde{x})}{\exp(\ln b)}\right]^{-c}\right\} \\ &= \exp\Big\{-\Big[\exp(\widetilde{x}-\ln b)\Big]^{-c}\Big\} \\ &= \exp\big\{-\exp[-c\,(\widetilde{x}-\ln b)]\big\}. \end{aligned} \tag{1.40b}$$

Introducing the **transformed location–scale parameter** $(\widetilde{a},\widetilde{b})$, where

$$\widetilde{a} = \ln b, \tag{1.40c}$$

$$\widetilde{b} = \frac{1}{c}, \tag{1.40d}$$

(1.40b) results in

$$F(\widetilde{x}\,|\,\widetilde{a},\widetilde{b}) = \exp\left\{-\exp\left(-\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right)\right\},\ \ \widetilde{x}\in\mathbb{R},\ \widetilde{a}\in\mathbb{R},\ \widetilde{b}>0, \tag{1.40e}$$

which is recognized as the CDF of the extreme value distribution of **type I for the maximum**. The DF belonging to (1.40e) can be found either by differentiating (1.40e) with respect to $\widetilde{x}$ or by applying (1.37a) together with (1.39b,c). The result is

$$f(\widetilde{x}\,|\,\widetilde{a},\widetilde{b}) = \frac{1}{\widetilde{b}}\,\exp\left\{-\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}} - \exp\left(-\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right)\right\}. \tag{1.40f}$$

Thus, when $X$ is of maximum type II, $\ln(X-a)$ is of maximum type I with location–scale parameter $(\widetilde{a},\widetilde{b}) = (\ln b, 1/c)$. The transformations of the other extreme value distributions follow along the same line.

- The extreme value distribution of **type III for the maximum**, sometimes referred to as **WEIBULL–type distribution**, has DF

$$f(x\,|\,a,b,c)=\frac{c}{b}\left(\frac{a-x}{b}\right)^{c-1}\exp\left\{-\left(\frac{a-x}{b}\right)^{c}\right\},\;\; x\le a,\; a\in\mathbb{R},\; b>0,\; c>0 \quad (1.41a)$$

and CDF

$$F(x\,|\,a,b,c)=\left\{\begin{array}{cl}\exp\left\{-\left(\dfrac{a-x}{b}\right)^{c}\right\} & \text{for} \quad x\le a,\\ 1 & \text{for} \quad x>a.\end{array}\right\} \quad (1.41b)$$

The transformed variable

$$\widetilde{X}=-\ln(a-X) \quad (1.42a)$$

has the **type I distribution for the maximum** (1.40e,f) with scale parameter $\widetilde{b}=1/c$, whereas the new location parameter is

$$\widetilde{a}=-\ln b. \quad (1.42b)$$

- The extreme value distribution of **type II for the minimum**, the **FRÉCHET distribution**, has DF

$$f(x\,|\,a,b,c)=\frac{c}{b}\left(\frac{a-x}{b}\right)^{-c-1}\exp\left\{-\left(\frac{a-x}{b}\right)^{-c}\right\},\; x\le a, a\in\mathbb{R}, b>0, c>0 \quad (1.43a)$$

and CDF

$$F(x\,|\,a,b,c)=\left\{\begin{array}{cl}1-\exp\left\{-\left(\dfrac{a-x}{b}\right)^{-c}\right\} & \text{for} \quad x\le a,\\ 1 & \text{for} \quad x>a.\end{array}\right\} \quad (1.43b)$$

Introducing once more

$$\widetilde{X}=-\ln(a-X) \quad (1.44a)$$

transforms this distribution to the extreme value distribution of **type I for the minimum**, see (1.17b), with DF

$$f(\widetilde{x}\,|\,\widetilde{a},\widetilde{b})=\frac{1}{\widetilde{b}}\exp\left\{\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}-\exp\left(\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right)\right\},\;\; \widetilde{x}\in\mathbb{R},\; \widetilde{a}\in\mathbb{R},\; \widetilde{b}>0, \quad (1.44b)$$

and CDF

$$F(\widetilde{x}\,|\,\widetilde{a},\widetilde{b})=1-\exp\left\{-\exp\left(\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right)\right\}, \quad (1.44c)$$

where $\widetilde{b}=1/c$ and $\widetilde{a}=-\ln b$.

- The extreme value distribution of **type III for the minimum**, the **WEIBULL distribution**,[6] has DF

$$f(x\,|\,a,b,c)=\frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1}\exp\left\{-\left(\frac{x-a}{b}\right)^{c}\right\},\; x\ge a, a\in\mathbb{R}, b>0, c>0 \quad (1.45a)$$

and CDF

$$F(x\,|\,a,b,c)=\left\{\begin{array}{cl} 0 & \text{for} \quad x<a, \\ 1-\exp\left\{-\left(\dfrac{x-a}{b}\right)^{c}\right\} & \text{for} \quad x\ge a. \end{array}\right\} \quad (1.45b)$$

With

$$\widetilde{X}=\ln(X-a)$$

we find the extreme value distribution of **type I for the minimum**, see (1.44b,c), where $\widetilde{b}=1/c$ but $\widetilde{a}=\ln b$. The extreme value of type I for the minimum is often referred to as **Log–WEIBULL distribution**. Comparing (1.45a) with (1.38a) we may also call the extreme value distribution of type II for the maximum an **inverse WEIBULL distribution**. Comparing (1.45a) with (1.41a) we see that the extreme value distribution of type III for the maximum may be called **reflected WEIBULL distribution**.

Tab. 1/2 summarizes the transformation procedures just described in a concise manner.

Table 1/2: Transformation of extreme value distributions

| Original distribution | Transformed variable | Transformed distribution | Transformed parameters |
|---|---|---|---|
| maximum of type II (1.38a,b) | $\widetilde{x}=\ln(x-a)$ | maximum of type I (1.40e,f) | $\widetilde{a}=\ln b$<br>$\widetilde{b}=1/c$ |
| maximum of type III (1.41a,b) | $\widetilde{x}=-\ln(a-x)$ | maximum of type I (1.40e,f) | $\widetilde{a}=-\ln b$<br>$\widetilde{b}=1/c$ |
| minimum of type II (1.43a,b) | $\widetilde{x}=-\ln(a-x)$ | minimum of type I (1.44b,c) | $\widetilde{a}=-\ln b$<br>$\widetilde{b}=1/c$ |
| minimum of Type III (1.45a,b) | $\widetilde{x}=\ln(x-a)$ | minimum of type I (1.44b,c) | $\widetilde{a}=\ln b$<br>$\widetilde{a}=1/c$ |

### Lognormal distribution

If there is a real number $a$ such that

$$\widetilde{X}=\ln(X-a) \quad (1.46a)$$

[6] For more details of this distribution see RINNE (2009).

is normally distributed, the distribution of $X$ is said to be **lognormal**. The DF of $X$ is

$$f(x\,|\,a,\widetilde{a},\widetilde{b}) = \frac{1}{(x-a)\,\widetilde{b}\sqrt{2\,\pi}}\,\exp\left\{-\frac{1}{2}\left(\frac{\ln(x-a)-\widetilde{a}}{\widetilde{b}}\right)^2\right\}\left\{\begin{array}{l} x \geq a,\ a \in \mathbb{R}, \\ \widetilde{a} \in \mathbb{R},\ \widetilde{b} > 0. \end{array}\right\} \quad (1.46b)$$

The graph of (1.46b) is positively skew. The distribution of $\widetilde{X}$ is normal with DF

$$f(\widetilde{x}\,|\,\widetilde{a},\widetilde{b}) = \frac{1}{\widetilde{b}\sqrt{2\,\pi}}\,\exp\left\{-\frac{1}{2}\left(\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right)^2\right\}\left\{\begin{array}{l} \widetilde{x} \in \mathbb{R}, \\ \widetilde{a} \in \mathbb{R},\ \widetilde{b} > 0. \end{array}\right\} \quad (1.46c)$$

The meaning of the three parameters in (1.46b) is as follows:

- $a$ is the location parameter of $X$ and gives a **lower threshold**.[7]
- $\widetilde{a} = \mathrm{E}[\ln(X-a)]$ is the location parameter of the transformed variate.
- $\widetilde{b} = \sqrt{\mathrm{Var}[\ln(X-a)]}$ is the scale parameter of $\widetilde{X}$.

In Sections 5.3.3.1 and 5.3.3.2 we will show that the variance of $X$ is dependent on $\widetilde{a}$ and $\widetilde{b}$ and that the shape, i.e. the skewness and the kurtosis, of $X$ is only dependent on $\widetilde{b}$.

### PARETO distribution

The PARETO distribution (of the first kind) has a negatively skew DF:

$$f(x\,|\,a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{-c-1},\ \ x \geq a+b,\ a \in \mathbb{R},\ b > 0,\ c > 0, \quad (1.47a)$$

and CDF

$$F(x\,|\,a,b,c) = 1 - \left(\frac{x-a}{b}\right)^{-c}. \quad (1.47b)$$

Introducing

$$\widetilde{X} = \ln(X-a)$$

leads to

$$f(\widetilde{x}\,|\,\widetilde{a},\widetilde{b}) = \frac{1}{\widetilde{b}}\exp\left(-\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right),\ \ \widetilde{x} \geq \widetilde{a},\ \widetilde{a} \in \mathbb{R},\ \widetilde{b} > 0, \quad (1.48a)$$

$$F(\widetilde{x}\,|\,\widetilde{a},\widetilde{b}) = 1 - \exp\left(-\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right) \quad (1.48b)$$

with

$$\widetilde{a} = \ln b,\ \ \widetilde{b} = 1/c. \quad (1.48c)$$

So $\widetilde{X}$ has an **exponential distribution**.

[7] The term 'lognormal' can also be applied to the distribution of $X$ if $\ln(a-X)$ is normally distributed. In this case $a$ is an **upper threshold** and $X$ has zero probability of exceeding $a$.

**Power–function distribution**

The power–function distribution with DF

$$f(x\,|\,a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1},\;\; a \le x \le a+b,\; a \in \mathbb{R},\; b>0,\; c>0, \tag{1.49a}$$

and CDF

$$F(x\,|\,a,b,c) = \left(\frac{x-a}{b}\right)^{c} \tag{1.49b}$$

by introducing

$$\widetilde{X} = \ln(X-a)$$

transforms to

$$f(\widetilde{x}\,|\,\widetilde{a},\widetilde{b}) = \frac{1}{\widetilde{b}}\exp\left(\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right),\;\; \widetilde{x} \le \widetilde{a},\; \widetilde{a} \in \mathbb{R},\; \widetilde{b} > 0, \tag{1.50a}$$

$$F(\widetilde{x}\,|\,\widetilde{a},\widetilde{b}) = \exp\left(\frac{\widetilde{x}-\widetilde{a}}{\widetilde{b}}\right) \tag{1.50b}$$

with parameters

$$\widetilde{a} = \ln b,\;\; \widetilde{b} = 1/c. \tag{1.50c}$$

(1.50a) is recognized as the **reflected exponential distribution**. Because the PARETO and the power–function distributions are related by a reciprocal transformation of their variables, the logarithms of these variables differ by sign and their distributions are related by a reflection.

# 2 Order statistics[1]

Order statistics and their functions play an important role in probability plotting and linear estimation of location–scale distributions. Plotting positions, see Sect. 3.3.2, and regressors as well as the elements of the variance–covariance matrix in the general least–squares (GLS) approach, see Sect. 4.2.1, are moments of the reduced order statistics. In most sampling situations the observations have to be ordered after the sampling, but in life–testing, when failed items are not replaced, order statistics will arise in a natural way.

## 2.1 Distributional concepts

Let $X_1, X_2, \ldots, X_n$ be **i**ndependently **i**dentically **d**istributed (iid) with CDF $F(x)$. The variables $X_i$ being arranged in ascending order and written as

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$$

are called **order statistics**. The CDF of $X_{r:n}$ is given by

$$\begin{aligned} F_{r:n}(x) &= \Pr\big(X_{r:n} \leq x\big) \\ &= \Pr(\text{at least } r \text{ of the } X_i \text{ are less than or equal to } x) \\ &= \sum_{i=r}^{n} \binom{n}{i} \big[F(x)\big]^i \big[1 - F(x)\big]^{n-i}, \end{aligned} \tag{2.1a}$$

since the term in the summand is the binomial probability that exactly $i$ of $X_1, \ldots, X_n$ are less than or equal to $x$. $F_{r:n}(x)$ can be written as the **incomplete beta function ratio** or **beta distribution function**:

$$F_{r:n}(x) = \frac{\int\limits_0^{F(x)} u^{r-1}\,(1-u)^{n-r}\,\mathrm{d}u}{B(r, n-r+1)}. \tag{2.1b}$$

Because the parameters $r$ and $n$ of the complete beta function are integers we have

$$B(r, n-r+1) = \frac{(r-1)!\,(n-r)!}{n!}\,. \tag{2.1c}$$

(2.1a,b) hold whether $X$ is discrete or continuous. In the following text we will always assume that $X$ is absolutely continuous with DF $f(x)$. Then differentiation of (2.1b) and

[1] Suggested reading for this chapter: ARNOLD/BALAKRISHNAN/NAGARAJA (1992), BALAKRISHNAN/COHEN (1991), BALAKRISHNAN/RAO (1998a,b), DAVID (1981), GALAMBOS (1978), SARHAN/GREENBERG (1962).

regarding (2.1c) gives the DF of $X_{r:n}$:

$$f_{r:n}(x) = \frac{n!}{(r-1)!\,(n-r)!}\,\big[F(x)\big]^{r-1}\,\big[1-F(x)\big]^{n-r}\,f(x). \tag{2.1d}$$

Two values of $r$ are of special interest:

- $r = 1$ – $X_{1:n}$ is the **sample minimum** with DF and CDF

$$f_{1:n}(x) = n\,f(x)\,\big[1-F(x)\big]^{n-1}, \tag{2.2a}$$

$$F_{1:n}(x) = \sum_{i=1}^{n}\binom{n}{i}\big[F(x)\big]^{i}\,\big[1-F(x)\big]^{n-i} = 1-\big[1-F(x)\big]^{n}. \tag{2.2b}$$

- $r = n$ – $X_{n:n}$ is the **sample maximum** with DF and CDF

$$f_{n:n}(x) = n\,f(x)\,\big[F(x)\big]^{n-1}, \tag{2.3a}$$

$$F_{n:n}(x) = \big[F(x)\big]^{n}. \tag{2.3b}$$

There are only a few variates whose order statistics' DF can be given in a simple and handy form. The most important example is the **reduced uniform variate**, denoted as $U$ in the following text, with DF and CDF

$$f(u) = 1,\;\; 0 \le u \le 1, \tag{2.4a}$$

$$F(u) = u. \tag{2.4b}$$

Upon inserting (2.4a,b) into (2.1a) and (2.1d) we find for $U_{r:n},\; 1 \le r \le n$:

$$\begin{aligned} F_{r:n}(u) &= \sum_{i=r}^{n}\binom{n}{i}u^{i}\,(1-u)^{n-i} \\ &= \frac{n!}{(r-1)!\,(n-r)!}\int\limits_{0}^{u} t^{r-1}\,(1-t)^{n-r}\,\mathrm{d}t,\;\; 0 \le u \le 1, \end{aligned} \tag{2.4c}$$

and

$$f_{r:n}(u) = \frac{n!}{(r-1)!\,(n-r)!}\,u^{r-1}\,(1-u)^{n-r}. \tag{2.4d}$$

Thus, $U_{r:n}$ has a reduced beta distribution.

Another example is the **reduced power–function** distribution with DF and CDF

$$f(y) = c\,y^{c-1},\;\; 0 \le y \le 1, \tag{2.5a}$$

$$F(y) = y^{c}. \tag{2.5b}$$

With (2.5a,b) the CDF and DF of $Y_{r:n},\ 1 \le r \le n$, follow from (2.1a,d) as

$$
\begin{aligned}
F_{r:n}(y) &= \sum_{i=r}^{n} \binom{n}{i} \big(y^c\big)^i \big(1-y^c\big)^{n-i} \\
&= \frac{n!}{(r-1)!\,(n-r)!} \int\limits_0^{y^c} t^{r-1}\,(1-t)^{n-r}\,\mathrm{d}t,\ \ 0 \le y \le 1, \qquad (2.5c)\\
f_{n:r}(y) &= \frac{n!}{(r-1)!\,(n-r)!}\, c\, y^{r\,c-1} \big(1-y^c\big)^{n-r}. \qquad (2.5d)
\end{aligned}
$$

The **joint DF** of $X_{r:n}$ and $X_{s:n},\ 1 \le r < s \le n$, is

$$
\begin{aligned}
f_{r,s:n}(x,y) &= \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \big[F(x)\big]^{r-1} \big[F(y)-F(x)\big]^{s-r-1} \times \\
&\quad \big[1-F(y)\big]^{n-s} f(x)\, f(y),\ \ x < y. \qquad (2.6a)
\end{aligned}
$$

Even if $X_1, \ldots, X_n$ are independent, their order statistics are not independent random variables. The joint CDF of $X_{r:n}$ and $X_{s:n}$ may be obtained by integration of (2.6a) as well as by a direct argument valid also in the discrete case.

- For $x < y$ we have:

$$
\begin{aligned}
F_{r,s:n}(x,y) &= \Pr(\text{at least } r\ X_i \le x \text{ and at least } s\ X_i \le y) \\
&= \sum_{j=s}^{n} \sum_{k=r}^{j} \Pr(\text{exactly } k\ X_i \le x \text{ and exactly } j\ X_i \le y) \\
&= \sum_{j=s}^{n} \sum_{k=r}^{j} \frac{n!}{k!\,(j-k)!\,(n-j)!} \big[F(x)\big]^{k} \big[F(y)-F(x)\big]^{j-k} \times \\
&\quad \big[1-F(y)\big]^{n-j},\ \ x < y. \qquad (2.6b)
\end{aligned}
$$

- For $x \ge y$ the inequality $X_{s:n} \le y$ implies $X_{r:n} \le x$, so that

$$
F_{r,s:n}(x,y) = F_{s:n}(y),\ \ x \ge y. \qquad (2.6c)
$$

By using the identity

$$
\begin{aligned}
&\sum_{j=s}^{n} \sum_{k=r}^{j} \frac{n!}{k!\,(j-k)!\,(n-j)!}\, p_1^k \big(p_2-p_1\big)^{j-k} \big(1-p_2\big)^{n-j} \\
&= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \int\limits_0^{p_1}\!\int\limits_{t_1}^{p_2} t_1^{r-1} \big(t_2-t_1\big)^{s-r-1} \big(1-t_2\big)^{n-s}\, \mathrm{d}t_2\, \mathrm{d}t_1,\ 0 < p_1 < p_2 < 1,
\end{aligned}
$$

we can write the joint CDF of $X_{r:n}$ and $X_{s:n}$ in (2.6b ) equivalently as

$$F_{r,s:n}(x,y) = \frac{n!}{(r-1)!\,(s-r-1)!(n-s)!} \int\limits_0^{F(x)} \int\limits_{t_1}^{F(y)} t_1^{r-1}\big(t_2-t_1\big)^{s-r-1} \times \big(1-t_2\big)^{n-s} \,\mathrm{d}t_2\,\mathrm{d}t_1,\ 0<x<y<\infty, \tag{2.7}$$

which is the CDF of the **reduced bivariate beta distribution**. The joint DF of $X_{r:n}$ and $X_{s:n}$ may be derived from (2.7) by differentiating with respect to both $x$ and $y$.

Let $U_1,\ U_2,\ \ldots,\ U_n$ be a sample of the reduced uniform distribution and $X_1,\ X_2,\ \ldots,\ X_n$ be a random sample from a population with CDF $F(x)$. Furthermore, let $U_{1:n} \le U_{2:n} \le \ldots \le\ U_{n:n}$ and $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ be the corresponding order statistics.

Specifically, when $F(x)$ is continuous the **probability integral transformation** $U = F(X)$ produces a reduced uniform distribution. Thus, when $F(x)$ is continuous we have

$$F(X_{r:n}) \stackrel{d}{=} U_{r:n},\ \ r=1,2,\ldots,n. \tag{2.8a}$$

where $\stackrel{d}{=}$ reads as "has the same distribution as". Furthermore, with the inverse CDF $F^{-1}(\cdot)$, it is easy to verify that

$$F^{-1}(U_r) \stackrel{d}{=} X_r,\ \ r=1,2,\ldots,n \tag{2.8b}$$

for an arbitrary $F(\cdot)$. Since $F^{-1}(\cdot)$ is also order preserving, it immediately follows that

$$F^{-1}(U_{r:n}) \stackrel{d}{=} X_{r:n},\ r=1,2,\ldots,n. \tag{2.8c}$$

We will apply the distributional relation (2.8c) in Sect. 2.2.2 in order to develop some series approximations for the moments of reduced order statistics $Y_{r:n}$ in terms of moments of the uniform order statistics $U_{r:n}$.

## 2.2 Moments of order statistics

For probability plotting and linear estimation we will need the first and second moments of the reduced order statistics.

### 2.2.1 Definitions and basic formulas

We will denote the **crude single moments** of the reduced order statistics, $\mathrm{E}(Y_{r:n}^k)$, by $\alpha_{r:n}^{(k)},\ 1\le r\le n$. They follow with

$$f(y) := f_Y(y) \ \text{ and } \ F(y) := F_Y(y)$$

as

$$\begin{aligned}\alpha_{r:n}^{(k)} \ := \ \mathrm{E}(Y_{r:n}^k) \ &= \int\limits_{-\infty}^{\infty} y^k\, f_{r:n}(y)\,\mathrm{d}y \\ &= \frac{n!}{(r-1)!\,(n-r)!} \int\limits_{-\infty}^{\infty} y^k\,[F(y)]^{r-1}\,[1-F(y)]^{n-r}\,f(y)\,\mathrm{d}y. \end{aligned} \tag{2.9a}$$

The most import case of (2.9a) is the mean, shortly denoted by $\alpha_{r:n}$:

$$\alpha_{r:n} := \mathrm{E}(Y_{r:n}) = n\binom{n-1}{r-1} \int\limits_{-\infty}^{\infty} y\,[F(y)]^{r-1}\,[1-F(y)]^{n-r}\,\mathrm{d}F(y). \tag{2.9b}$$

Since $0 \leq F(y) \leq 1$, it follows that

$$|\alpha_{r:n}| \leq n\binom{n-1}{r-1} \int\limits_{-\infty}^{\infty} |y|\,\mathrm{d}F(y), \tag{2.9c}$$

showing that $\alpha_{r:n}$ exists provided $\mathrm{E}(Y)$ exists, although the converse is not necessarily true.

We mention that for the general variate $X = a + b\,Y$ the crude single moments of $X_{r:n}$ are given by

$$\mu_{r:n}^{(k)} := \mathrm{E}(X_{r:n}^k) = \sum_{j=0}^{k}\binom{k}{j}\alpha_{r:n}^{(k-j)}\,b^{k-j}\,a^j,$$

especially

$$\mu_{r:n} := \mathrm{E}(X_{r:n}) = a + b\,\alpha_{r:n}.$$

An alternative formula for $\alpha_{r:n}$ may be obtained by integration by parts in

$$\alpha_{r:n} = \int\limits_{-\infty}^{\infty} y\,\mathrm{d}F_{r:n}(y).$$

To this end, note that for any CDF $F(y)$ the existence of $\mathrm{E}(Y)$ implies

$$\lim_{y\to-\infty} y\,F(y) = 0 \;\text{ and }\; \lim_{x\to\infty} y\,[1-F(y)] = 0,$$

so that we have

$$\begin{aligned}
\mathrm{E}(Y) &= \int\limits_{-\infty}^{\infty} y\,\mathrm{d}F(y) \\
&= \int\limits_{-\infty}^{0} y\,\mathrm{d}F(y) - \int\limits_{0}^{\infty} y\,\mathrm{d}[1-F(y)] \\
&= \int\limits_{0}^{\infty} [1-F(y)]\,\mathrm{d}y - \int\limits_{-\infty}^{0} F(y)\,\mathrm{d}y.
\end{aligned} \tag{2.10a}$$

This general formula gives $\alpha_{r:n} = \mathrm{E}(Y_{r:n})$ if $F(y)$ is replaced by $F_{r:n}(y)$ :

$$\alpha_{r:n} = \int\limits_{0}^{\infty} \big[1-F_{r:n}(y)\big]\,\mathrm{d}y - \int\limits_{-\infty}^{0} F_{r:n}(y)\,\mathrm{d}y. \tag{2.10b}$$

We may also write

$$\alpha_{r:n} = \int\limits_0^\infty \bigl[1 - F_{r:n}(y) - F_{r:n}(-y)\bigr]\,\mathrm{d}y, \tag{2.10c}$$

and when $f(y)$ is symmetric about $y = 0$ we have

$$\alpha_{r:n} = \int\limits_0^\infty \bigl[F_{n-r+1}(y) - F_{r:n}(y)\bigr]\,\mathrm{d}y. \tag{2.10d}$$

**Crude product moments** of reduced order statistics may be defined similarly:

$$\begin{aligned}\alpha_{r,s:n}^{(k,\ell)} &:= \mathrm{E}\Bigl(Y_{r:n}^k\,Y_{s:n}^\ell\Bigr)\\ &= \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \iint\limits_{-\infty<t<v<\infty} t^k\,v^\ell\,\bigl[F(t)\bigr]^{r-1}\bigl[F(v)-F(t)\bigr]^{s-r-1}\times\\ &\quad \bigl[1-F(v)\bigr]^{n-s} f(t)\,f(v)\,\mathrm{d}t\,\mathrm{d}v. \end{aligned}\tag{2.11a}$$

The most important case of(2.11a) has $k = \ell = 1$ and leads to the **covariance of $\boldsymbol{Y_{r:n}}$ and $\boldsymbol{Y_{s:n}}$**:

$$\begin{aligned}\mathrm{Cov}\bigl(Y_{r:n}, Y_{s:n}\bigr) &= \mathrm{E}\bigl(Y_{r:n}\,Y_{s:n}\bigr) - \mathrm{E}\bigl(Y_{r:n}\bigr)\,\mathrm{E}\bigl(Y_{s:n}\bigr)\\ &= \alpha_{r,s:n} - \alpha_{r:n}\,\alpha_{s:n}.\end{aligned}\tag{2.11b}$$

We introduce the notation

$$\beta_{r,s:n} := \mathrm{Cov}(Y_{r:n}, Y_{s:n}). \tag{2.12a}$$

A special case of (2.12a) for $r = s$ is **the variance of $\boldsymbol{Y_{r:n}}$**:

$$\begin{aligned}\beta_{r,r:n} &= \mathrm{Var}(Y_{r:n})\\ &= \alpha_{r:n}^{(2)} - \bigl(\alpha_{r:n}\bigr)^2.\end{aligned}\tag{2.12b}$$

We collect all the variances and covariances for a given sample size in the so–called **variance–covariance matrix $\boldsymbol{B}$**:

$$\boldsymbol{B} := \bigl(\beta_{r,s:n}\bigr);\ \ r, s = 1, \ldots, n. \tag{2.12c}$$

which is a symmetric matrix:

$$\boldsymbol{B} = \boldsymbol{B}' \ \text{ or } \ \beta_{r,s:n} = \beta_{s,r:n}\ \ \forall\, r, s. \tag{2.12d}$$

For the general order statistics we have

$$\begin{aligned}\sigma_{r,s:n} &= b^2\,\beta_{r,s:n}\\ \boldsymbol{\Sigma} &= \bigl(\sigma_{r,s:n}\bigr) = b^2\,\boldsymbol{B}.\end{aligned}$$

Generally, the moments of order statistics cannot be given in a handy and closed form. They have either to be evaluated by numeric integration of (2.9a) and (2.11a) in combination with some recurrence relation or to be approximated by expanding $Y_{r:n} = F^{-1}(U_{r:n})$ in a TAYLOR series around the point $\mathrm{E}(U_{r:n})$. Furthermore, when the **distribution of $\boldsymbol{Y}$** is **symmetric around zero** we may save time in computing the moments of order statistics making use of the following relations which are based on the distributional equivalences

$$\begin{aligned} Y_{r:n} &\overset{d}{=} (-Y_{n-r+1:n}),\ \ 1 \le r \le n, \\ (Y_{r:n}, Y_{s:n}) &\overset{d}{=} \big((-Y_{n-s+1:n}), (-Y_{n-r+1:n})\big),\ \ 1 \le r \le s \le n, \end{aligned}$$

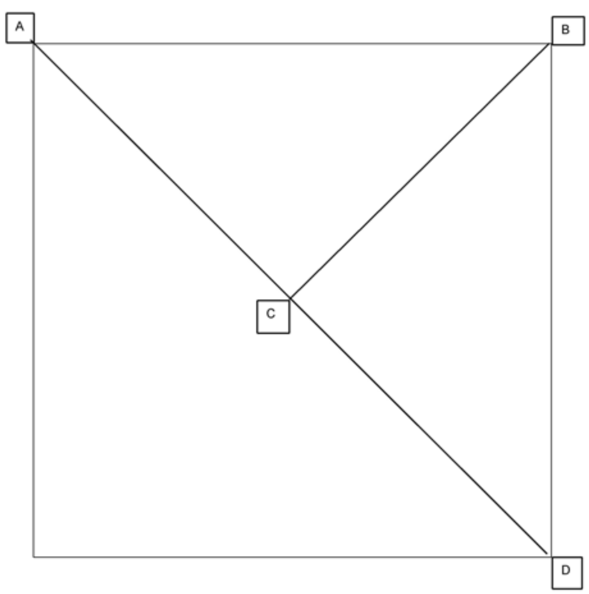
resulting in:

$$\alpha_{n-r+1:n}^{(k)} = (-1)^k\, \alpha_{r:n}^{(k)},\ \ 1 \le r \le n,\ k \ge 1, \tag{2.13a}$$

$$\alpha_{n-s+1,n-r+1:n} = \alpha_{r,s:n},\ \ 1 \le r \le s \le n, \tag{2.13b}$$

$$\begin{aligned} \beta_{n-s+1,n-r+1:n:n} &= \alpha_{n-s+1,n-r+1:n} - \alpha_{n-s+1:n}\,\alpha_{n-r+1:n} \\ &= \alpha_{r,s:n} - \alpha_{r:n}\,\alpha_{s:n} \\ &= \beta_{r,s:n},\ \ 1 \le r \le s \le n. \end{aligned} \tag{2.13c}$$

Figure 2/1: Structure of the variance–covariance matrix of order statistics from a distribution symmetric around zero




A special case of (2.13a) occurs when $n$ is odd, $n = 2\,\ell + 1,\ \ell \in \mathbb{N}$:

$$\alpha_{\ell+1:2\,\ell+1} = -\alpha_{\ell+1:2\,\ell+1} = 0,$$

i.e. the mean of the sample median is equal to zero. (2.13a) means that we have to evaluate $[\frac{n+1}{2}]$ means instead of $n$, where $[z]$ is the integer part of $z$. Looking at (2.13c) we have — besides $\beta_{r,s:n} = \beta_{s,r:n}$ — a second symmetry in the variance–covariance matrix of order statistics, see Fig. 2/1. The greater upper triangle ABD results from a reflection of the smaller triangle ABC at the line BC. We only have to evaluate the elements in the upper triangle ABC. Thus, instead of evaluating $(n+1)\big/2$ elements $\beta_{r,s:n}$ $(r = 1,\dots,n;\ s \ge r)$ we only have to evaluate $n\,(n+2)\big/4$ elements for $n = 2\,\ell$ and $(n+1)^2\big/4$ elements for $n = 2\,\ell + 1$.

Closed form expressions for moments of order statistics exist — among others — for the following distributions:

1. Reduced uniform distribution: $f(u) = 1,\ \ 0 \le u \le 1$

$$\mathrm{E}\big(U_{r:n}^k\big) = \frac{n!}{(n+k)!}\,\frac{(r+k-1)!}{(r-1)!} \tag{2.14a}$$

$$\alpha_{r:n} = \frac{r}{n+1} =: p_r \tag{2.14b}$$

$$\beta_{r,r:n} = \frac{r\,(n-r+1)}{(n+1)^2\,(n+2)} = \frac{p_r\,q_r}{n+2}\,,\ q_r := 1 - p_r \tag{2.14c}$$

$$\mathrm{E}\big(U_{r:n}\,U_{s:n}\big) = \frac{r\,(s+1)}{(n+1)\,(n+2)}\,,\ 1 \le 1 < r < s \le n \tag{2.14d}$$

$$\beta_{r,s:n} = \frac{r\,(n-s+1)}{(n+1)^2\,(n+2)} = \frac{p_r\,q_s}{n+2}\,,\ q_s := 1 - p_s \tag{2.14e}$$

2. Reduced power–function distribution: $f(y) = c\,y^{c-1},\ \ 0 \le y \le 1$

$$\mathrm{E}\big(Y_{r:n}^k\big) = \frac{\Gamma(n+1)}{\Gamma(n+1+k/c)}\,\frac{\Gamma(r+k/c)}{\Gamma(r)} = \frac{n!}{(r-1)!}\,\frac{\Gamma(r+k/c)}{\Gamma(n+1+k/c)} \tag{2.15a}$$

$$\alpha_{r:n} = \frac{n!}{(r-1)!}\,\frac{\Gamma(r+1/c)}{\Gamma(n+1+1/c)} \tag{2.15b}$$

$$\beta_{r,r:n} = \frac{n!}{(r-1)!}\left\{\frac{\Gamma(r+2/c)}{\Gamma(n+1+2/c)} - \frac{n!}{(r-1)!}\,\frac{\Gamma^2(r+1+1/c)}{\Gamma^2(n+1+1/c)}\right\} \tag{2.15c}$$

$$\mathrm{E}(Y_{r:n}\,Y_{s:n}) = \frac{n!}{(r-1)!}\,\frac{\Gamma(r+1/c)\,\Gamma(s+2/c)}{\Gamma(s+1/c)\,\Gamma(n+1+2/c)},\ 1 \le r < s \le n \tag{2.15d}$$

$$\begin{aligned}\beta_{r,s:n} = \frac{n!}{(r-1)!}&\left\{\frac{\Gamma(s+2/c)\,\Gamma(r+1/c)}{\Gamma(s+1/c)\,\Gamma(n+1+2/c)} - \right.\\ &\left.\frac{n!}{(s-1)!}\,\frac{\Gamma(s+1/c)\,\Gamma(r+1/c)}{\Gamma^2(n+1+1/c)}\right\}\end{aligned} \tag{2.15e}$$

3. Reduced exponential distribution: $f(y) = e^{-y}, \;\; y \geq 0$

$$\mathrm{E}\big(Y_{r:n}^k\big) = \sum_{i=1}^{r} \frac{k!}{n-i+1} \tag{2.16a}$$

$$\alpha_{r:n} = \sum_{i=1}^{r} \frac{1}{n-i+1} = \sum_{i=n-r+1}^{n} \frac{1}{i} \tag{2.16b}$$

$$\beta_{r,r:n} = \beta_{r,s:n} = \sum_{i=1}^{r} \frac{1}{(n-i+1)^2} = \sum_{i=n-r+1}^{n} \frac{1}{i^2} \tag{2.16c}$$

### 2.2.2 Identities, recurrence relations and approximations

For the computation of moments of order statistics and for checking the results we need some identities and recurrence relations. By using the basic identity

$$\left(\sum_{i=1}^{n} X_{i:n}^k\right)^{\ell} \equiv \left(\sum_{i=1}^{n} X_i^k\right)^{\ell} \tag{2.17a}$$

several **identities** for single and product moments of order statistics can be established which primarily serve the **purpose of checking**. By choosing $\ell = 1$ and taking expectations on both sides, we get the identity

$$\sum_{i=1}^{n} \mu_{i:n}^{(k)} = n\,\mathrm{E}\big(X^k\big) = n\,\mu_{1:1}^{(k)}. \tag{2.17b}$$

Similarly, by taking $k = 1$ and $\ell = 2$ we obtain

$$\sum_{i=1}^{n} X_{i:n}^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_{i:n}\,X_{j:n} = \sum_{i=1}^{n} X_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_i\,X_j. \tag{2.17c}$$

Now taking expectations on both sides leads to

$$\sum_{i=1}^{n} \mu_{i:n}^{(2)} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mu_{i,j:n} = n\,\mathrm{E}\big(X^2\big) + n\,(n-1)\big[\mathrm{E}(X)\big]^2 \tag{2.17d}$$

which, when used together with (2.17b), yields an identity for product moments of order statistics

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mu_{i,j:n} = \binom{n}{2}\big[\mathrm{E}(X)\big]^2 = \binom{n}{2}\mu_{1:1}^2 = \binom{n}{2}\mu_{1,2:2}. \tag{2.17e}$$

Furthermore, from

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mu_{i,j:n} = n\,\mu_{1:1}^{(2)} + n\,(n-1)\,\mu_{1:1}^2 \tag{2.18a}$$

and (2.17b) we get

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{i,j:n} &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mu_{i,j:n}-\left(\sum_{i=1}^{n}\mu_{i:n}\right)\left(\sum_{j=1}^{n}\mu_{j:n}\right)\\ &= n\Big\{\mu_{1:1}^{(2)}-\mu_{1:1}^{2}\Big\}\\ &= n\,\mathrm{Var}(X)\\ &= n\,\sigma_{1,1:n}. \end{aligned} \tag{2.18b}$$

Starting from (2.17a), one can establish the **triangle rule** leading to a **recurrence relation for single moments** of order statistics:[2]

$$r\,\mu_{r+1:n}^{(k)}+(n-r)\,\mu_{r:n}^{(k)}=n\,\mu_{r:n-1}^{(k)}. \tag{2.19a}$$

This triangle rule shows that it is enough to evaluate the $k$-th moment of a single order statistic in a sample of size $n$, if these moments in samples of size less than $n$ are already available. The $k$-th moment of the remaining $n-1$ moments can then be determined by repeated use of of (2.19a). For this purpose we could, for example, start with either $\mu_{1:n}^{(k)}$ or $\mu_{n:n}^{(k)}$. It is, therefore, desirable to reformulate (2.19a) so that $\mu_{r:n}^{(k)}$ is purely expressed in terms of the $k$-th moment of the smallest or the largest order statistics from samples of size up to $n$. The resulting recurrence relations are:

$$\mu_{r:n}^{(k)} = \sum_{i=n-r+1}^{n}(-1)^{i-n+r-1}\binom{n}{i}\binom{i-1}{n-r}\mu_{1:i}^{(k)},\ \ 2\le r\le n;\ k=1,2,\ldots \tag{2.19b}$$

$$\mu_{r:n}^{(k)} = \sum_{i=r}^{n}(-1)^{i-r}\binom{n}{i}\binom{i-1}{r-1}\mu_{i:i}^{(k)},\ \ 1\le r\le n-1,\ k=1,2,\ldots \tag{2.19c}$$

A warning has to be issued here. If any of the recurrence formulas (2.19a-c) is used in the computation of the moments of order statistics then the identity (2.17b) should not be employed to check these computations because this identity will be automatically satisfied if any of these recurrence relations is used.

A **recurrence relation for the product moments** of order statistics similar to (2.19a) is given by

$$(r-1)\,\mu_{r,s:n}^{(k,\ell)}+(s-r)\,\mu_{r-1,s:n}^{(k,\ell)}+(n-s+1)\,\mu_{r-1,s-1:n}^{(k,\ell)}=n\,\mu_{r-1,s-1:n}^{(k,\ell)},\ 2\le r<s\le n. \tag{2.20a}$$

(2.20a) shows that it is enough to evaluate the $(k,\ell)$-th moment of $n-1$ suitably chosen pairs of order statistics, if these moments in samples of size less than $n$ are already available. For example, the knowledge of $\{\mu_{1,s:n}^{(k,\ell)}$ for $2\le s\le n\}$ or $\{\mu_{r,n:n}^{(k,\ell)}$ for $1\le r\le n-1\}$

[2] More recurrence relations may be found in ARNOLD/BALAKRISHNAN (1989).

or $\{\mu_{r,r+1:n}^{k,\ell}$ for $1 \leq r \leq n-1\}$ will suffice. The $(k,\ell)$-th moment of the remaining $\binom{n-1}{2}$ pairs of order statistics can then be determined by repeated use of (2.20a). Hence, it is desirable to express the product moment of pairs of order statistics just given from samples of size up to $n$. One of the resulting recurrence relations is

$$\mu_{r,s:n}^{(k,\ell)} = \sum_{i=s-r}^{s-1} \sum_{j=n-s+i+1}^{n} (-1)^{j+n-r-s+1} \binom{i-1}{r-1} \binom{j-i-1}{n-s} \binom{n}{j} \mu_{i,i+1:j}^{(k,\ell)}, \quad 1 \leq r < s \leq n \text{ and } k,\ell = 1,2,\ldots \tag{2.20b}$$

The following recurrence relation for covariances of order statistics is similar to (2.19a):

$$\begin{aligned} &(r-1)\,\sigma_{r,s:n} + (s-r)\,\sigma_{r-1,s:n} + (n-s+1)\,\sigma_{r-1,s-1:n} \;= \\ &\quad n\left\{\sigma_{r-1,s-1:n-1} + \left(\mu_{r-1:n-1} - \mu_{r-1:n}\right)\left(\mu_{s-1:n-1} - \mu_{s:n}\right)\right\}, \;\; 2 \leq r < s \leq n. \end{aligned} \tag{2.21}$$

ARNOLD/BALAKRISHNAN (1989) have given the maximum number of single and product moments to be evaluated for the computation of all means, variances, and covariances of order statistics in a sample of size $n$ when these quantities are available in samples of sizes up to $n-1$ and when systematic usage of the recurrence relations is done:

> We have to evaluate at most two single moments and $(n-2)/2$ product moments if $n$ is even, and at most two single moments and $(n-1)/2$ product moments if $n$ is odd.

These maximum numbers can be decreased further for samples from a symmetric distribution.

It is possible to give **approximations for moments of reduced order statistics** based on the following two distributional equivalences:

$$Y_{r:n} \stackrel{d}{=} F^{-1}(U_{r:n}) \;\text{ and }\; (Y_{r:n}, Y_{s:n}) \stackrel{d}{=} \big(F^{-1}(U_{r:n}), F^{-1}(U_{s:n})\big),$$

where $U_{r:n}$ is an order statistic from a reduced uniform distribution, see (2.14a-e). Upon expanding $F^{-1}(U_{r:n})$ in a TAYLOR series around the point $\mathrm{E}(U_{r:n}) = r/(n+1) =: p_r$ we get a series expansion for $Y_{r:n}$ as

$$\left.\begin{aligned} Y_{r:n} \;=\;& F^{-1}(p_r) + F^{-1(1)}(p_r)\,(U_{r:n} - p_r) + \frac{1}{2}\,F^{-1(2)}(p_r)\,(U_{r:n} - p_r)^2 + \\ & \frac{1}{6}\,F^{-1(3)}(p_r)\,(U_{r:n} - p_r)^3 + \frac{1}{24}\,F^{-1(4)}(p_r)\,(U_{r:n} - p_r)^4 + \ldots, \end{aligned}\right\} \tag{2.22}$$

where $F^{-1(1)}(p_r),\; F^{-1(2)}(p_r), \ldots$ are the successive derivatives of $F^{-1}(u)$ evaluated at $p_r$. Now by taking expectations on both sides of (2.22) and using the expressions of the moments of reduced uniform order statistics given in (2.14a-e) [however, written in inverse powers of $n+2$ by DAVID (1981, p. 80) for computational ease and algebraic simplicity], we find the **approximate formula for the mean of $Y_{r:n}$**:

$$\left.\begin{aligned}
\alpha_{r:n} \approx\; & F^{-1}(p_r) + \frac{p_r\,q_r}{2\,(n+2)}\,F^{-1(2)}(p_r) + \\
& \frac{p_r\,q_r}{(n+2)^2}\left[\frac{1}{3}\,(q_r-p_r)\,F^{-1(3)}(p_r) + \frac{1}{8}\,p_r\,q_r\,F^{-1(4)}(p_r)\right] + \\
& \frac{p_r\,q_r}{(n+2)^3}\left[-\frac{1}{3}\,(q_r-p_r)\,F^{-1(3)}(p_r) + \frac{1}{4}\{(q_r-p_r)^2-p_r\,q_r\}\,F^{-1(4)}(p_r) + \right.\\
& \left.\frac{1}{6}\,p_r\,q_r\,(q_r-p_r)\,F^{-1(5)}(p_r) + \frac{1}{48}\,p_r^2\,q_r^2\,F^{-1(6)}(p_r)\right],
\end{aligned}\right\} \quad (2.23)$$

where $q_r = 1 - p_r = (n-r+1)\big/(n+1)$.

An **approximate formula for the variance of $Y_{r:n}$** is derived by taking expectation on both sides of the series expansion for $Y_{r:n}^2$, obtained from (2.22) and then subtracting from it the approximation of $\alpha_{r:n}^2$ obtained from (2.23).

$$\left.\begin{aligned}
\beta_{r,r:n} \approx\; & \frac{p_r\,q_r}{(n+2)}\left[F^{-1(1)}(p_r)\right]^2 + \\
& \frac{p_r\,q_r}{(n+2)^2}\left\langle 2\,(q_r-p_r)\,F^{-1(1)}(p_r)\,F^{-1(2)}(p_r) + \right.\\
& \left. p_r\,q_r\left\{F^{-1(1)}(p_r)\,F^{-1(3)}(p_r) + \frac{1}{2}\left[F^{-1(2)}(p_r)\right]^2\right\}\right\rangle + \\
& \frac{p_r\,q_r}{(n+2)^3}\left\langle -\,2\,(q_r-p_r)\,F^{-1(1)}(p_r)\,F^{-1(2)}(p_r) + \right.\\
& \left[(q_r-p_r)^2 - p_r\,q_r\right]\left\{2\,F^{-1(1)}(p_r)\,F^{-1(3)}(p_r) + \frac{3}{2}\,\left[F^{-1(2)}(p_r)\right]^2\right\} + \\
& p_r\,q_r\,(q_r-p_r)\left\{\frac{5}{3}\,F^{-1(1)}(p_r)\,F^{-1(4)}(p_r) + 3\,F^{-1(2)}(p_r)\,F^{-1(3)}(p_r)\right\} + \\
& \left.\frac{1}{4}\,p_r^2\,q_r^2\left\{F^{-1(1)}(p_r)F^{-1(5)}(p_r) + 2F^{-1(2)}(p_r)F^{-1(4)}(p_r) + \frac{5}{3}\,\left[F^{-1(3)}(p_r)\right]^2\right\}\right\rangle
\end{aligned}\right\} \quad (2.24)$$

By taking expectation on both sides of the series expansion for $(Y_{r:n}\,Y_{s:n})$ obtained from (2.22) and than subtracting from it the approximations for $\alpha_{r:n}$ and $\alpha_{s:n}$ obtained from

(2.23) we find an **approximate formula for the covariance of $Y_{r:n}$ and $Y_{s:n}$.**

$$\left.\begin{aligned}
&\beta_{r,s:n} \approx \frac{p_r\, q_s}{n+2}\, F^{-1(1)}(p_r)\, F^{-1(1)}(p_s) + \\
&\frac{p_r q_s}{(n+2)^2}\Big\langle (q_r - p_r)\, F^{-1(2)}(p_r) F^{-1(1)}(p_s) + (q_s - p_s)\, F^{-1(1)}(p_r) F^{-1(2)}(p_s) + \\
&\frac{1}{2}\Big\{ p_r q_r F^{-1(3)}(p_r) F^{-1(1)}(p_s) + p_s q_s F^{-1(1)}(p_r) F^{-1(3)}(p_s) + p_r q_s F^{-1(2)}(p_r) F^{-1(2)}(p_s) \Big\}\Big\rangle + \\
&\frac{p_r q_s}{(n+2)^3}\Big\langle -(q_r - p_r) F^{-1(2)}(p_r) F^{-1(1)}(p_s) - (q_s - p_s) F^{-1(1)}(p_r) F^{-1(2)}(p_s) + \\
&\Big[(q_r - p_r)^2 - p_r q_r\Big] F^{-1(3)}(p_r) F^{-1(1)}(p_s) + \Big[(q_s - p_s)^2 - p_s q_s\Big] F^{-1(1)}(p_r) F^{-1(3)}(p_s) + \\
&1.5\Big[(q_r - p_r)(q_s - p_s) + 0.5\, p_s q_r - 2\, p_r q_s\Big] F^{-1(2)}(p_r) F^{-1(2)}(p_s) + \\
&\frac{5}{6}\Big[p_r q_r (q_r - p_r) F^{-1(4)}(p_r) F^{-1(1)}(p_s) + p_s q_s (q_s - p_s) F^{-1(1)}(p_r) F^{-1(4)}(p_s)\Big] + \\
&\Big[p_r q_s (q_r - p_r) + \frac{1}{2} p_r q_r (q_s - p_s)\Big]\, F^{-1(3)}(p_r) F^{-1(2)}(p_s) + \\
&\Big[p_r q_s (q_s - p_s) + \frac{1}{2} p_s q_s (q_r - p_r)\Big]\, F^{-1(2)}(p_r) F^{-1(3)}(p_s) + \\
&\frac{1}{8}\Big[p_r^2 q_r^2 F^{-1(5)}(p_r) F^{-1(1)}(p_s) + p_s^2 q_s^2 F^{-1(1)}(p_r) F^{-1(5)}(p_s)\Big] + \\
&\frac{1}{4}\Big[p_r^2 q_r q_s F^{-1(4)}(p_r) F^{-1(2)}(p_s) + p_r p_s q_s^2 F^{-1(2)}(p_r) F^{-1(4)}(p_s)\Big] + \\
&\frac{1}{12}\Big[2 p_r^2 q_s^2 + 3 p_r p_s q_r q_s\Big] F^{-1(3)}(p_r) F^{-1(3)}(p_s) \Big\rangle
\end{aligned}\right\} \tag{2.25}$$

It is possible to **shorten the formulas** above by discontinuing after the term where the denominator of the multiplier reads $(n+2)^2$, but at the price of a smaller precision. As DAVID (1981, p. 81) notes, the convergence of the series expansions may be slow or even non–existent if $r/n$ is too close to 0 or to 1.

The evaluation of the derivatives of $F^{-1}(u)$ is rather straightforward when the inverse CDF is available explicitly. For example, for the **arc–sine distribution** with

$$F(y) = \frac{1}{2} + \frac{\arcsin(y)}{\pi}\,, \;\; -1 < y < 1, \;\text{ and }\; F^{-1}(u) = \sin\big[\pi\,(u - 0.5)\big], \;\; 0 < u < 1,$$

we obtain:

$$\begin{aligned}
F^{-1(1)}(u) &= \pi\,\cos\bigl[\pi\,(u-0.5)\bigr],\\
F^{-1(2)}(u) &= -\pi^2\,\sin\bigl[\pi\,(u-0.5)\bigr],\\
F^{-1(3)}(u) &= -\pi^3\,\cos\bigl[\pi\,(u-0.5)\bigr],\\
F^{-1(4)}(u) &= \pi^4\,\sin\bigl[\pi\,(u-0.5)\bigr],\\
F^{-1(5)}(u) &= \pi^5\,\cos\bigl[\pi\,(u-0.5)\bigr],\\
F^{-1(6)}(u) &= -\pi^6\,\sin\bigl[\pi\,(u-0.5)\bigr].
\end{aligned}$$

When the $F^{-1}(u)$ does not exist in explicit form we can base the forming of its successive derivatives on the fact that

$$F^{-1(1)}(u) = \frac{\mathrm{d}F^{-1}(u)}{\mathrm{d}u} = \frac{\mathrm{d}y}{\mathrm{d}u} = \frac{1}{\mathrm{d}u/\mathrm{d}y} = \frac{1}{f(y)} = \frac{1}{f\bigl[F^{-1}(u)\bigr]}, \tag{2.26}$$

which is nothing but the reciprocal of the DF of the population evaluated at $F^{-1}(u)$. Higher–order derivatives of $F^{-1}(u)$ may be obtained by successive differentiating the expression of $F^{-1(1)}(u)$ in (2.26).

The most popular distribution having no explicit formula for either its CDF nor its inverse CDF is the **normal distribution**. In this case we first have

$$\frac{\mathrm{d}f(y)}{\mathrm{d}y} = -y\,f(y)$$

and finally obtain

$$\begin{aligned}
F^{-1(1)}(u) &= \frac{1}{f\bigl[F^{-1}(u)\bigr]},\\
F^{-1(2)}(u) &= \frac{F^{-1}(u)}{\Bigl\{f\bigl[F^{-1}(u)\bigr]\Bigr\}^2},\\
F^{-1(3)}(u) &= \frac{1+2\,\bigl\{F^{-1}(u)\bigr\}^2}{\Bigl\{f\bigl[F^{-1}(u)\bigr]\Bigr\}^3},\\
F^{-1(4)}(u) &= \frac{F^{-1}(u)\Bigl\{7+6\,\bigl[F^{-1}(u)\bigr]^2\Bigr\}}{\Bigl\{f\bigl[F^{-1}(u)\bigr]\Bigr\}^4},\\
F^{-1(5)}(u) &= \frac{7+46\,\bigl[F^{-1}(u)\bigr]^2+24\,\bigl[F^{-1}(u)\bigr]^4}{\Bigl\{f\bigl[F^{-1}(u)\bigr]\Bigr\}^5},\\
F^{-1(6)}(u) &= \frac{F^{-1}(u)\Bigl\{127+326\,\bigl[F^{-1}(u)\bigr]^2+120\,\bigl[F^{-1}(u)\bigr]^4\Bigr\}}{\Bigl\{f\bigl[F^{-1}(u)\bigr]\Bigr\}^6}
\end{aligned}$$

**Example 2/1: Exact and approximate moments of reduced exponential order statistics**

Assuming an exponential population, the following Tab. 2/1 shows for a sample of size $n = 6$ the exact means, variances and covariances according to (2.16a–c) and their approximations computed with the long and the short version of the series expansions (2.23) through (2.25). From the CDF of the reduced exponential distribution, $F(x) = 1 - \exp(x)$ we first have the inverse CDF $F^{-1}(u) = -\ln(1-u),\ 0 \le u < 1$. The first six derivatives of $F^{-1}(u)$, needed for the approximations, are:

$$\begin{aligned} F^{-1(1)} &= (1-u)^{-1}, & F^{-1(2)} &= (1-u)^{-2}, \\ F^{-1(3)} &= 2\,(1-u)^{-3}, & F^{-1(4)} &= 6\,(1-u)^{-4}, \\ F^{-1(5)} &= 24\,(1-u)^{-5}, & F^{-1(6)} &= 120\,(1-u)^{-6}. \end{aligned}$$

Table 2/1: Exact and approximate means, variances and covariances of reduced exponential order statistics when $n = 6$

| Parameter | Exact value | Short approximation | Long approximation |
|---|---|---|---|
| $\alpha_{1:6}$ | $0.1\overline{6}$ | 0.166340 | 0.166620 |
| $\alpha_{2:6}$ | $0.3\overline{6}$ | 0.365847 | 0.366554 |
| $\alpha_{3:6}$ | $0.61\overline{6}$ | 0.615035 | 0.616440 |
| $\alpha_{4:6}$ | 0.95 | 0.946835 | 0.949584 |
| $\alpha_{5:6}$ | 1.45 | 1.443193 | 1.449296 |
| $\alpha_{6:6}$ | 2.45 | 2.430285 | 2.451770 |
| $\beta_{1,1:6},\ \dots\ \beta_{1,6:6}$ | $0.02\overline{7}$ | 0.026259 | 0.027481 |
| $\beta_{2,2:6},\ \dots,\ \beta_{2,6:6}$ | $0.0\overline{6}$ | 0.063750 | 0.066974 |
| $\beta_{3,3:6},\ \dots,\ \beta_{3,6:6}$ | $0.1302\overline{7}$ | 0.121582 | 0.128494 |
| $\beta_{4,4:6},\ \dots,\ \beta_{4,6:6}$ | $0.2413\overline{8}$ | $0.\overline{2}$ | 0.237317 |
| $\beta_{5,5:6},\ \beta_{5,6:6}$ | $0.4913\overline{8}$ | 0.439453 | 0.479940 |
| $\beta_{6,6:6}$ | $1.4913\overline{8}$ | 1.218750 | 1.435547 |

As to expected the approximations based on the longer formulas are closer to the exact values than those based on the shorter formulas.

## 2.3 Functions of order statistics

We will study some linear functions of order statistics and start with two special sums of two order statistics. The **median in a sample of even size** $n = 2\,m$ is defined as

$$\widetilde{\widetilde{X}} := \frac{X_{m:2\,m} + X_{m+1:2\,m}}{2}\,. \tag{2.27a}$$

Its DF $f_{\tilde{\tilde{X}}}(\cdot)$ may be derived from the joint DF of two order statistics (2.6a) by setting $n = 2\,m$, $r = m$ and $s = m + 1$ and by using standard transformation methods. The sum $T := X_{m:2\,m} + X_{m+1:2\,m}$ has DF

$$f_T(t) = \frac{(2\,m)!}{2\,(m-1)!} \int\limits_{-\infty}^{+\infty} \bigl\{F(v)\,\bigl[1 - F(t-v)\bigr]\bigr\}^{m-1} f(v)\,f(t-v)\,\mathrm{d}v, \qquad (2.27b)$$

so $\tilde{\tilde{X}} = T/2$ has DF

$$f_{\tilde{\tilde{X}}}(x) = 2\,f_T(2\,x) = \frac{(2\,m)!}{(m-1)!} \int\limits_{-\infty}^{\infty} \bigl\{F(v)\,\bigl[1 - F(2\,x - v)\bigr]\bigr\}^{m-1} f(v)\,f(2\,x - v)\,\mathrm{d}v. \qquad (2.27c)$$

The CDF of $\tilde{\tilde{X}}$ is given by

$$F_{\tilde{\tilde{X}}}(x) = \frac{2}{B(m,m)} \int\limits_{-\infty}^{x} \bigl[F(v)\bigr]^{m-1} \Bigl\{\bigl[1 - F(v)\bigr]^m - \bigl[1 - F(2\,x - v)\bigr]^m\Bigr\}\, f(v)\,\mathrm{d}v. \qquad (2.27d)$$

Another measure of central tendency in a sample — besides the median — is the **mid–range**:

$$M := \frac{X_{1:n} + X_{n:n}}{2} \qquad (2.28a)$$

with DF

$$f_M(w) = 2\,n\,(n-1) \int\limits_{-\infty}^{y} \bigl[F(2\,w - v) - F(v)\bigr]^{n-2} f(v)\,f(2\,w - v)\,\mathrm{d}v \qquad (2.28b)$$

and CDF

$$F_M(w) = n \int\limits_{-\infty}^{y} \bigl[F(2\,w - v) - F(v)\bigr]^{n-1} f(v)\,\mathrm{d}v. \qquad (2.28c)$$

The **difference of two arbitrary order statistics**

$$W_{rs} := X_{s:n} - X_{r:n}, \;\; 1 \le r < s \le n, \qquad (2.29a)$$

has DF

$$\begin{aligned} f_{W_{rs}}(w) \;\; = \;\; & \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \int\limits_{-\infty}^{\infty} \bigl[F(v)\bigr]^{r-1} \bigl[F(w+v) - F(v)\bigr]^{s-r-1} \times \\ & \bigl[1 - F(w+v)\bigr]^{n-s} f(v)\,f(w+v)\,\mathrm{d}v. \end{aligned} \qquad (2.29b)$$

A special case of (2.29a) is the **range** $W$:

$$W := X_{n:n} - X_{1:n} \tag{2.30a}$$

with DF and CDF

$$\begin{aligned} f_W(w) &= n\,(n-1)\int\limits_{-\infty}^{\infty}\big[F(w+v)-F(v)\big]^{n-2} f(v)\,f(w+v)\,\mathrm{d}v, & (2.30b)\\ F_W(w) &= n\int\limits_{-\infty}^{\infty}\big[F(w+v)-F(v)\big]^{n-1} f(v)\,\mathrm{d}v & (2.30c)\end{aligned}$$

and mean and variance

$$\begin{aligned} \mathrm{E}(W) &= \mathrm{E}\big(X_{n:n}\big) - \mathrm{E}\big(X_{1:n}\big) & (2.30d)\\ \mathrm{Var}(W) &= \mathrm{Var}\big(X_{n:n}\big) - 2\,\mathrm{Cov}\big(X_{1:n}, X_{n:n}\big) + \mathrm{Var}\big(X_{1:n}\big). & (2.30e)\end{aligned}$$

We now proceed to more **general linear functions of order statistics**:

$$L_n = \sum_{i=1}^{n} a_{in}\,X_{i:n}.$$

A major use of such functions arises in the estimation of location and scale parameters $a$ and $b$ for a location–scale distribution with a DF of the form

$$f_X(x\,|\,a,b) = \frac{1}{b}\,f_Y\left(\frac{x-a}{b}\right),\ a\in\mathbb{R},\ b>0,$$

where $f_Y(\cdot)$ is parameter–free. Denoting the reduced variate by

$$Y := \frac{X-a}{b}$$

and the moments of its order statistics by

$$\begin{aligned} \alpha_{r:n} &:= \mathrm{E}\big(Y_{r:n}\big) & (2.31a)\\ \beta_{r,s:n} &:= \mathrm{Cov}\big(Y_{r:n}, Y_{s:n}\big);\ \ r,s = 1,2,\ldots,n; & (2.31b)\end{aligned}$$

it follows that

$$Y_{r:n} = \frac{X_{r:n}-a}{b}$$

so that

$$\begin{aligned} \mathrm{E}\big(X_{r:n}\big) &= a + b\,\alpha_{r:n}, & (2.31c)\\ \mathrm{Cov}\big(X_{r:n}, X_{s:n}\big) &= b^2\,\beta_{r,s:n}. & (2.31d)\end{aligned}$$

Thus $\mathrm{E}\big(X_{r:n}\big)$ is linear in the parameters $a$ and $b$ with known coefficients, and $\mathrm{Cov}\big(X_{r:n}, X_{s:n}\big)$ is known apart from $b^2$. Therefore, the GAUSS–MARKOV least–squares theorem may be applied, in a slightly generalized form since the variance–covariance matrix is not diagonal. This gives the best linear unbiased estimators (BLUEs)

$$\widehat{a} = \sum_{i=1}^{n} a_{in}\, X_{i:n}, \tag{2.31e}$$

$$\widehat{b} = \sum_{i=1}^{n} b_{in}\, X_{i:n}, \tag{2.31f}$$

where the coefficients $a_{in}$ and $b_{in}$, which are functions of the $\alpha_{r:n}$ and $\beta_{r,s:n}$, can be evaluated once and for all, depending on the DF $f_Y(\centerdot)$ of the reduced variate. This technique will be applied in Chapter 4.

# 3 Statistical graphics[1]

An old Chinese truth says that a picture tells you more than thousand words. Statisticians, too, have made out the meaning of the slightly modified wisdom that a figure is able to convey more than a lot of numbers. One arrives at this conclusion when skimming statistical journals and monographs. Furthermore, computer science has supported the graphical approach in statistics by offering specialized and relatively inexpensive hardware and flexible software capable to even graphically display large and high-dimensional datasets. Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education and the mass media. Graphical methods, nowadays, play an important role in all aspects of statistical investigation, from the beginning exploratory plots, through various stages of analysis, to the final communication and display of results. Many persons consider graphical displays as the single most effective, robust statistical tool. Not only are graphical procedures helpful, but in many cases essential. TUKEY (1977) claims that "the greatest value of a picture is when it forces us to notice what we never expected to see".

## 3.1 Some historical remarks

Despite the venerable tradition of quantitative graphics for data analysis as sketched in this section and the revival of interest in statistical graphics in the last decades, a comprehensive and up–to–date description of this subject is still missing. There remains only a single monographic history of the subject: FUNKHOUSER (1938), now over 70 years old. A newer, but only brief history of quantitative graphics in statistics is given by BENIGER/ROBYN (1978). With the exception of a modest revival of interest in quantitative graphics among historians of early statistics, e.g. ROYSTON (1956) and TILLING (1975), visual forms have passed virtually unnoticed by historians and sociologists of knowledge and science.

Quantitative graphics have been central to the development of science, and statistical graphics, beginning with simple tables and plots, date from the earliest attempts to analyze empirical data. Many of the most familiar forms and techniques were well–established at least 200 years ago. At the turn of the 19th century, to use a convenient benchmark, a statistician might have resorted to the following graphical tools:

- **bivariate plots** of data points, used since the 17th century as witnessed by a paper of EDMUND HALLEY (1686) giving a graphical analysis of barometric pressure as a function of altitude. HALLEY's presentation rests upon the rectangular coordinate system which had been introduced in mathematics by RENÉ DESCARTES (1637)

[1] Suggested reading for this chapter: CLEVELAND (1993, 1994), TUFTE (1992, 2001).

after having been in use by old Egyptian surveyors three to four thousand years before;

- **line graphs of time series data** and **curve fitting** together with **interpolation** of empirical data points by J.H. LAMBERT (1760);

- the notion of **measurement error** as deviation from a rectangular graphed line, see J.H. LAMBERT (1765);

- **graphical analysis of periodic variation**, see J.H. LAMBERT (1779);

- **statistical mapping** by A.W.F. CROME (1782), who also introduced superimposed squares to compare areas and populations of European states;[2]

- **bar charts, pie charts circle graphs** by W. PLAYFAIR (1788, 1801).[3]

From the nineteenth century we mention the following inventions or new ideas in statistical graphics:

- the **ogive** or **cumulative frequency curve**, suggested by J.B.J. FOURIER (1821) to present the 1817 population of Paris by age grouping;

- the **histogram** by the French statistician A.M. GUERRY (1833) showing crime by age groupings;

- **logarithmic grid** by L. LALANNE (1843);[4]

- the **semi–logarithmic grid** by W.S. JEVONS (1863) to show the percentage changes in the prices of commodities;

- the **stereogram** by G. ZEUNER (1869) to depict demographic trends as surfaces in a three–coordinates system;[5]

- the **age pyramid** as a bilateral histogram or frequency polygon by F.A. WALKER (1874) to show results of the 1870 U.S. Census;

---

[2] A.W.F. CROME (1753 – 1833) held a chair in the Faculty of Economics at the University of Giessen, the first faculty of this kind, at least in Germany. Another famous statistican at the author's University of Giessen is E.L.E. LASPEYRES (1834 – 1913) who — in 1871 – suggested a formula for the price index which later on was given his name and which is used world–wide to measure inflation.

[3] There was a dispute between CROME and PLAYFAIR about the priority of comparing the size of statistical data by the area of geometric objects (squares, circles).

[4] This French engineer studied in detail the idea of linearizing plots by non–linear transformation of scales, what he called "anamorphisme", and so in particular introduced the **log–log plot**. He also suggested **polar coordinates**.

[5] This was a theoretical surface using axonometric projection to represent the history of various cohorts. About ten years later L. PEROZZO (1880) produced a colored relief drawing of ZEUNER's theoretical surface, based on actual data from the Swedish censuses of 1750 – 1875.

- **probability paper** suggested by F. GALTON (1899) and ruled so that the normal probability curve appeared as a straight line.[6] GALTON may be regarded as the one who laid the foundation–stone for one of this book's topic, the other one being the least–squares principle, tracing back to GAUSS and LEGENDRE.

The most recent history of graphical methods in the twentieth century is difficult to comment upon. We only mention the following highlights:

- the **LORENZ curve** of M.O. LORENZ (1905) to facilitate the study of concentration;
- the **graphical innovations for exploratory data analysis**, mainly due to J.W. TUKEY (1977) as, for example, the stem–and–leaf display, graphical lists, box–and–whisker display, hanging and suspended rootograms;
- the different **approaches to** graphically present **multivariate data**[7] like the circular glyphs with rays, see E. ANDERSON (1957), the ANDREWS' waves, see D.F. ANDREWS (1972), to generate plots of multivariate date by a form of FOURIER series or the CHERNOFF faces, see H. CHERNOFF (1973), to represent multivariate data by cartoons of human faces.

The new approaches and innovations exploit modern technology, and most have little practical value except when executed by computer. In the future, innovations in statistical graphics are likely to follow developments in computer graphics' hardware and software, and might be expected to include solutions to problems generated or made tractable by the computer and associated technologies. BENIGER/ROBYN (1978) listed seven problems which in the intervening time have been solved.

## 3.2 The role of graphical methods in statistics

The first part of this section discusses the relative merits of graphical versus numerical techniques. Then we turn to the dangers associated with the use of graphs and their ability to manipulate the observer of statistical graphs. In this context we will also comment upon graphical perception, i.e. the visual decoding of information encoded in graphs. Finally, in Sect. 3.2.3, we give an overview and a taxonomy of statistical graphs based on their uses. There we also show how the topic of this book — probability plotting and linear estimation — fits in the statistical graphics' building.

### 3.2.1 Graphical versus numerical techniques

Plots and graphs of a wide variety are used in the statistic literature for descriptive and inferential purposes. Visual representations have impact and they provide insight and understanding not readily found by other means. Formal numerical techniques are too often

[6] His so–called log–square paper was to be improved by MARTIN/LEAVENS (1931). Subsequently various special graph papers for probability plotting, including TUKEY's extremely ingenious **binomial paper**, became readily available commercially, see MOSTELLER/TUKEY (1949).

[7] The monograph of WANG (1978) is devoted to the graphical presentation of multivariate data.

designed to yield specific answers to rigidly defined questions. Graphical techniques are less formal and confining. They aid in understanding the numerous relationships reflected in the data. They help reveal departures from the assumed models and statistical distributions as is best demonstrated in ANSCOMBE (1973).[8] They help reveal the existence of peculiar looking observations or subsets of the data. Graphical data displays often uncover features of the data that were totally unanticipated prior to the analysis. It is difficult to obtain similar information from numerical procedures.

There is need in statistical data analysis for both numerical *and* graphical techniques. This text is a step in this direction by first assessing the validity of a chosen distribution by displaying the sample data in a graph paper special to that distribution and computing the distribution parameters in case of a linear fit to the data points. The numerical techniques serve as a yardstick against which one can evaluate the degree of evidence contained in the graphical display. Without such yardsticks it is quite easy to arrive at spurious conclusions.

Put another way — as TUKEY (1977) states — the relationship between graphical and numerical techniques is analogous to the relationship between a police detective and a judge. The graphical techniques are the counterpart of the detective looking for any clues to help uncover the mysteries of the data. The formal, numerical inference tools of classical statistics are the counterpart of the judge who weighs the degree of evidence contained in the clues to determine how much credence to put into them.

Once insights to a feature of the data have first been obtained from graphical procedures, one can then use a standard numerical procedure for a check on the conclusions obtained from the graphical analysis, for their increased precision and for their 'objectivity'. The graphical analyses suggest which numerical procedures to use and which of their assumptions are and are not satisfied. The results of these procedures in turn suggest new displays to prepare. Thus, graphical and numerical techniques are not contrary, but they complement each other and should be used iteratively as will be done in Chapter 5.

### 3.2.2 Manipulation with graphs and graphical perception

Graphical perception as defined by CLEVELAND/MCGILL (1984) is the visual decoding of information encoded on graphs. These authors — in a first attempt – identify ten elementary perceptual tasks that are carried out when people extract quantitative information from graphs:

- position on a common scale,
- position on non–aligned scales,
- length,

[8] ANSCOMBE constructed four data sets, each consisting of eleven $(x, y)$–pairs. Each of the four datasets yields the same standard output from a typical OLS regression program for an assumed linear relationship between $x$ and $y$. But when the datasets are graphed they show very different shapes: a linear relationship, a curved relationship, a strict linear relationship with one outlier and varying $y$–values at one and the same $x$–value together with one differing $(x, y)$–point.

- direction,
- angle,
- area,
- volume,
- curvature,
- shading,
- color saturation.

These items are not completely distinct tasks. For example, judging angle and direction are clearly related. A circle has an area associated with it, but it also has a length, and a person shown circles might well judge diameters or circumferences rather than areas.

Knowing what features of a graph are most attractive for the observer the designer may manipulate the graph. So, e.g. in designing ANDREWS' waves one can assign important characteristics to the high–frequency components of the trigonometric polynomial thus playing them down. The same thing is possible with CHERNOFF faces where one may hide important characteristics by assigning them to inconspicuous features of a face like the position of the pupil or the length and height of the eyebrow. Another very popular trick to manipulate graphs of time series consists in stretching the ordinate and compressing the abscissa (= time axis) to arrive at a steep course or doing the opposite to give the impression of a slow and smooth development. These actions are often combined with cutting of the ordinate without indication. The literature is full of examples showing graphs that intentionally or unintentionally manipulate the observer or that are badly designed, see for example KRÄMER (1994), LUKIESCH (1965), MONMONIER (1991), TOLANSKY (1964), TUFTE (1992, 2001) or the classical book on statistical lies of HUFF (1954).

On the other side there are a lot of papers and books containing good advice for the design of quantitative graphs like RIEDWYL (1987) or SCHMID/SCHMID (1979). Still, graph design for data analysis and presentation is largely unscientific, and there is a major need for a theory of graphical methods, although it is not at all clear what form such a theory should take. Of course, theory is not to be taken as meaning mathematical theory! What are required are ideas that will bring together in a coherent way things that previously appeared unrelated and which also provide a basis for dealing systematically with new situations.

Good graphs should be simple, self–explanatory and not deceiving. COX (1978) offers the following **guidelines** for the clarity of graphs:

1. The axes should be clearly labeled with the names of the variables and the units of measurement.
2. Scale breaks should be used for false origins.

3. Comparison of related diagrams should be made easy, for example by using identical scales of measurement and placing diagrams side by side.
4. Scales should be arranged so that approximately linear relations are plotted at roughly $45°$ to the $x$–axis.
5. Legends should make diagrams as nearly self–explanatory, i.e. independent of the text, as is feasible.
6. Interpretation should not be prejudiced by the technique of presentation.

### 3.2.3 Graphical displays in statistics

Graphical approaches in statistics mainly serve four purposes.

1. **Data exploration** in the sense of finding a suitable model
   1.1. For **data condensation** we have, e.g.
      - the histogram,
      - the dot–array diagram,
      - the stem–and–leaf diagram,
      - the frequency polygon,
      - the ogive or
      - the box–and–whiskers plot.

   1.2. For showing **relationships**
      - among two variables we may use
        - the scatter plot,
        - the sequence plot,
        - the autocorrelation plot or
        - the cross–correlation plot and
      - among three or more variables we have
        - the labeled scatter plot,
        - the draftsman plot,
        - the casement plot,
        - glyphs and metroglyphs,
        - the weather–vane plot,
        - profile plots,
        - the biplot,
        - face plots,
        - ANDREWS' waves,
        - cluster trees,
        - similarity and preference maps,
        - multidimensional scaling displays.

2. **Data analysis** in the sense of estimation and validation of a model
   2.1. **Distribution assessment** by means of
      - the QQ–plot,
      - the PP–plot,
      - the probability plot on probability paper,
      - the hazard plot,
      - the TTT–plot,
      - the hanging histogram,
      - the rootogram or
      - the POISSONNESS plot, see D.A. HOAGLIN (1980).

   2.2 Model adequacy and **assumption verification** by
      - average versus standard deviation plot,
      - residual plots,
      - partial–residual plots,
      - component–plus–residual plot.

   2.3. **Decision making** based on
      - control charts,
      - cusum charts,
      - the YOUDEN plot
      - the half–normal plot,
      - the ridge trace.
3. **Communication** and **data presentation** as an appealing alternative to tables
   3.1. **Quantitative graphics** like
      - bar charts,
      - pie charts,
      - pictograms,
      - the contour plot,
      - the stereogram or
      - the color map.

   3.2. **Summary of statistical analysis** like
      - means plots,
      - the notched box plot,
      - the interaction plot,
      - the confidence plot.
4. **Tools** in statistical work like
   - power curves and OC–curves,

– sample–size curves,
– confidence limits,
– nomographs to quickly read the values of complicated mathematical–statistical functions,
– the binomial paper of MOSTELLER/TUKEY.

We remark that there are graphs that serve more than one of the four purposes mentioned above.

## 3.3 Distribution assessment by graphs

In this section we present graphs which play a dominant role in the inference of *univariate and continuous* distributions. We will distinguish five types:

- the quantile plot or QQ–plot,
- the percent plot or PP–plot,
- the probability plot on specially designed probability paper,
- the hazard plot and
- the TTT–plot.

The first four technique have a common basis, i.e. they all start with the CDF of a random variable. This CDF has to be a parametric function, more precisely, it should be a member of the location–scale family, see Chapter 1. The TTT–plot mainly is a technique in non–parametric statistics giving less information than hazard plotting and probability plotting.

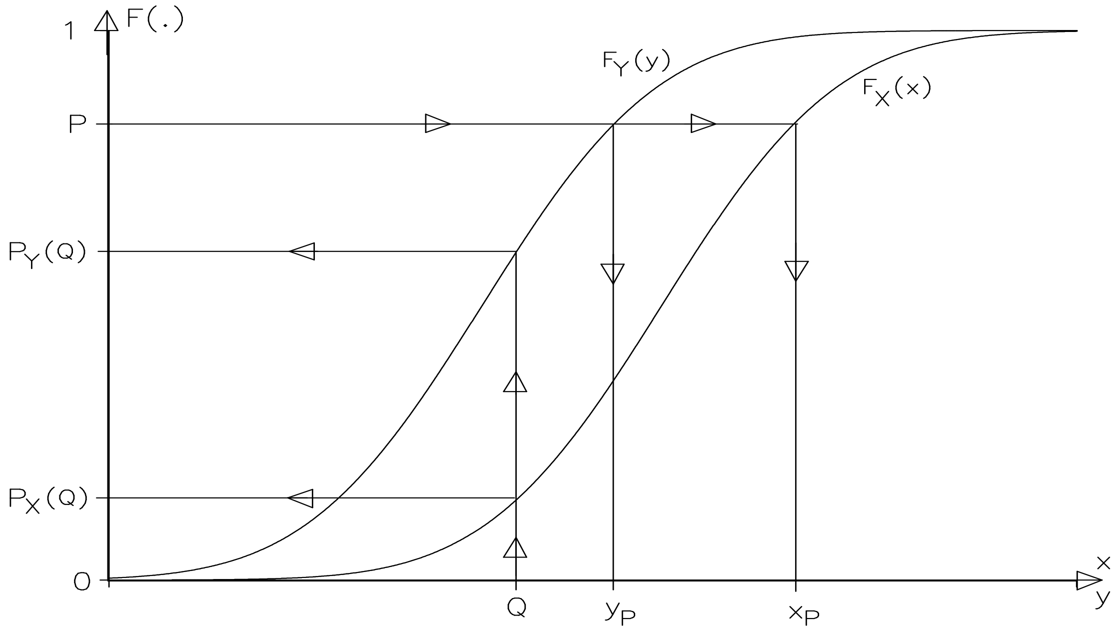
### 3.3.1 PP–plots and QQ–plots

Both types of plots are apt to compare

- two theoretical CDFs,
- two empirical CDFs and
- an empirical CDF to a theoretical CDF.

They do not help in estimating the parameters of the distribution, but they are the basis for probability plotting and hazard plotting which lead to estimates of the distribution parameters. Let $F_X(x)$ and $F_Y(y)$ be the CDFs of variates $X$ and $Y$, respectively, see Fig. 3/1. From this display we may deduce two types of graphs, the QQ–plot or quantile plot and the PP–plot or percent plot.

Figure 3/1: Explanation of the QQ–plot and the PP–plot




For each value $P$ on the ordinate axis displaying the CDF there are at most two values on the abscissa axis displaying the realizations of the variates, called quantiles:

$$x_P := Q_X(P) \;\text{ and }\; y_P := Q_Y(P).$$

Conversely, for each value $Q$ on the abscissa axis there are at most two values on the ordinate axis indicating the probability of each variate to show a realization up to $Q$:

$$\text{Pr}(X \le Q) = F_X(Q) = P_X(Q) \;\text{ and }\; \text{Pr}(Y \le Q) = F_Y(Q) = P_Y(Q).$$

A **PP–plot**, as is shown in Fig. 3/2, is a display where $F_Y(Q) = P_Y(Q)$ is plotted against $F_X(Q) = P_X(Q)$ for $Q$ varying, $Q \in \mathbb{R}$. The PP–plot is less important than the QQ–plot, see below. If $X$ and $Y$ were identically distributed, their CDFs will coincide in Fig. 3/1 and the resulting PP–plot will be a $45°$–line running from $\big(P_X(Q), P_Y(Q)\big)] = (0,0)$ to $\big(P_X(Q), P_Y(Q)\big) = (1,1)$. Variations from this line would indicate that the two distributions are not identical. Contrary to the QQ–plot the PP–plot will not be linear if one of the two variates is a linear transform of the other one, see Fig. 3/2 where $X$ and $Y$ are both normally distributed, $X \sim NO(0,1)$ and $Y \sim NO(-1,2)$.[9] $Y$ is — compared to $X$ — shifted to the left and thus will have a high probability when that of $X$ is still high and this effect is not compensated for by the smaller variance of $X$.

Despite the missing clear sensitivity of a PP–plot against a linear transformation of the variates it is of some importance. The PP–plot possesses a high discriminatory power in the region of high density because in that region the CDF, i.e. the value of $P$, is a more

[9] $X \sim NO(\mu,\sigma)$ reads: '$X$ is normally distributed with parameters $a = \mu$ and $b = \sigma$'.

rapidly changing function of $Q$ than in the region of low density, see Fig. 3/3, showing the normal and CAUCHY densities in the left–hand part and the corresponding PP–plot in the right–hand part. Both distributions are centered at $x = 0$ and the scale parameter $b$ of the CAUCHY distribution has been chosen so that both distributions have the same variation as measured by the length of a two–sided central $68.29\%$–interval. Furthermore, the idea of the PP–plot is — contrary to that of a QQ–plot — transferable to a multivariate distribution.

Figure 3/2: PP–plot comparing two normal distributions: $X \sim NO(0,1), Y \sim NO(-1,2)$

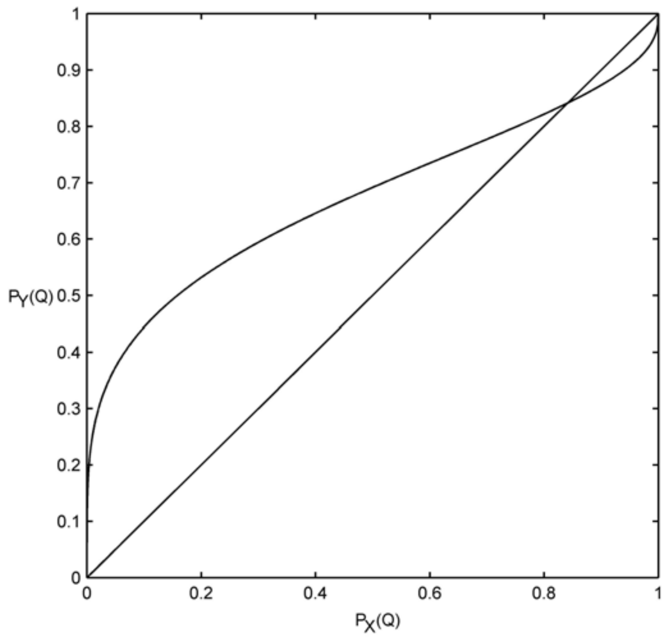
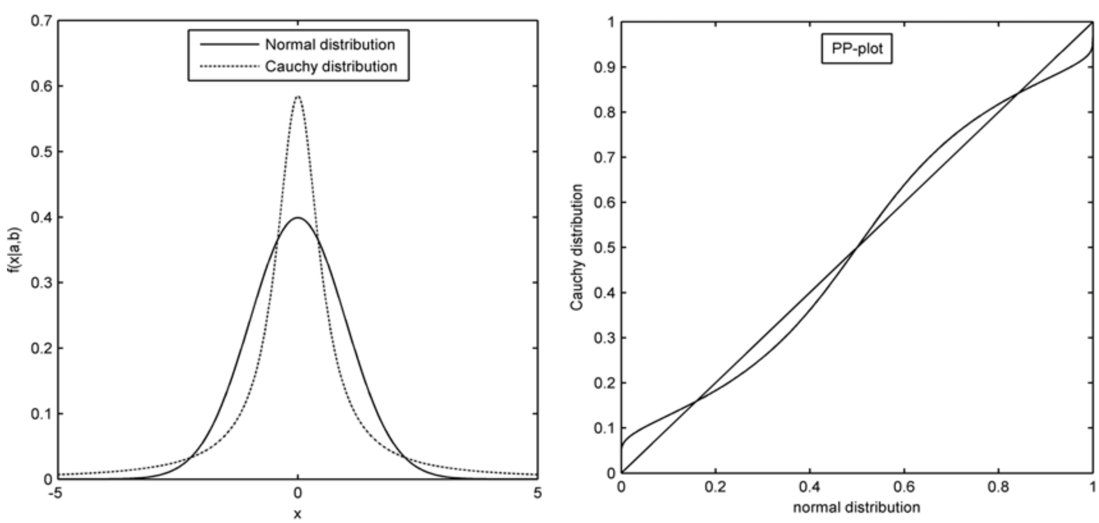



Figure 3/3: Normal and CAUCHY densities and the corresponding PP–plot



A **QQ–plot** is a display where $y_P$ is plotted against $x_P$ with $P$ varying, generally: $0 \leq P \leq 1$. For identically distributed variates $X$ and $Y$ the CDFs in Fig. 3/1 would coincide and we will get $x_P = y_P \;\forall\; P \in [0,1]$ and the QQ–plot is a $45°$–line running in the direction of the origin of the coordinate axes. If $x$ is a positive linear transform of $Y$, i.e. $X = a + b\,Y,\; b > 0$, then the QQ–plot will be straight line, which easily shown:

$$F_X(x) = F_Y\left(\frac{x-a}{b}\right),\; b > 0, \;\Longrightarrow\; y_P = \frac{x_P - a}{b} \text{ or } x_P = a + b\,y_P.$$

This property of linear invariance renders the QQ–plot especially useful in statistical analysis, because linearity is a form of course which is easily perceived by the human eye, as well as deviations from this course. There is a special case where QQ–plot and PP–plot are identical: $X$ and $Y$ both are uniformly distributed in $[0,1]$. This will happen when the two variates are probability integral transforms.

If one of the two distributions to be compared by a QQ–plot possesses very long tails with rather small DF–values, then the QQ–plot will emphasize this distributional difference whereas the difference in the "middle" region of the two distributions, where the density is relatively high, will be blurred somewhat. The reason for this kind of sensitivity of a QQ–plot is that a quantile changes rapidly with $P$ where the density is low, and only changes slowly with $P$ where the density is high. A greater concentration in the tails of one of the distributions as compared to the other one will cause the curve in the QQ–plot to deviate considerably from a straight line in the regions corresponding to $P < 0.05$ and $P > 0.95$, say, see the graphs on the right–hand side in Fig. 3/4.

It is also possible to use a **QQ–plot for** the comparison of **two empirical CDFs**, see the two graphs at the bottom of Fig. 3/4. When both samples are of equal size $n$, the empirical QQ–plot simply consists of plotting $y_{i:n}$ over $x_{i:n}$ for $i = 1, 2, \ldots, n$. For samples of different sizes the procedure is as follows.

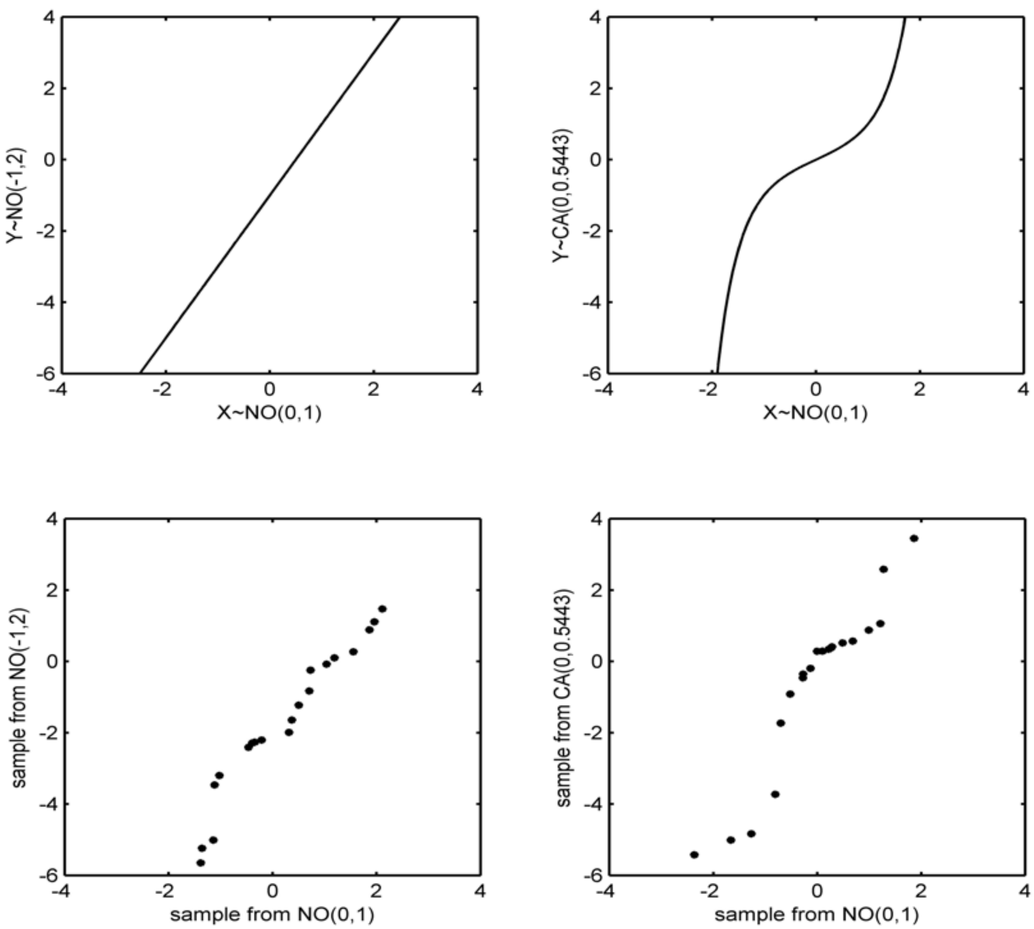
1. Let $n_1$ be the size of the smaller sample with observations $x_\nu,\; \nu = 1, 2, \ldots, n_1$ and $n_2$ be the size of the greater sample with observations $y_\kappa,\; \kappa = 1, 2, \ldots, n_2$.
2. The order of the empirical quantiles is chosen in such a way that the ordered $x$-observations are equal to the natural quantiles, i.e.
$$x_{p_\nu} = x_{\nu:n_1};\;\; p_\nu = \nu/n_1;\; \nu = 1, 2, \ldots, n_1. \tag{3.1a}$$
3. The $y$–quantile to be plotted over $x_{\nu:n_1}$ is an interpolated value:
$$y_{p_\nu} = y_{\kappa:n_2} + (n_2\,p_\nu - \kappa)\,(y_{\kappa+1:n_2} - y_{\kappa:n_2}) \tag{3.1b}$$
with
$$\kappa < n\,p_\nu \leq \kappa + 1. \tag{3.1c}$$

To compare an **empirical CDF**

$$F_n(x) = \begin{cases} 0 & \text{for} \quad x < x_{1:n} \\ i/n & \text{for} \quad x_{i:n} \leq x < x_{i+1:n},\; 1 \leq i \leq n-1, \\ 1 & \text{for} \quad x \geq x_{n:n} \end{cases}$$

on the abscissa axis to a **theoretical CDF** $F_Y(y)$ on the ordinate axis, we have to plot the $Y$–quantile of order $P = i/n$ against $x_{i:n}$. The resulting graph is a scatter–plot consisting of $n$ pairs $\left(x_{i:n}, y_{i/n}\right)$ or $n-1$ pairs when $y_{n/n} = \infty$ and the support of $X$ is unlimited to the right.

Figure 3/4: Theoretical and empirical QQ–plots




### 3.3.2 Probability paper and plotting positions

The two axes of a QQ–plot have a natural (= linear) scale. The most frequently used **probability papers** or **probability grids** result into a QQ–plot too, but

- on probability paper we compare empirical quantiles of a sample to the theoretical quantiles of a universe,
- the axes of a probability paper, especially the probability–labeled axis, are generally non–linear and distorted in such a way that the data points will scatter around a straight line when the sample has been drawn from the given universe.

Probability paper is predominant in the location–scale family of distributions where it is easily implemented and has a lot advantages. We will first give the theoretical background and show how to construct a probability paper. A special problem in application is the choice of the plotting position, i.e. the ordinate value, belonging to an ordered observation $x_{i:n}$.

A probability paper for a location–scale distribution is constructed by taking the **vertical axis** (ordinate) of a rectangular system of coordinates to lay off the quantiles of the reduced variable $Y$,[10]

$$y_P = F_Y^{-1}(P), \quad \text{generally: } 0 \leq P \leq 1, \tag{3.2}$$

but the labeling of this axis is according to the corresponding probability $P = F_Y^{-1}(y_P)$ or $100\,P\%$. This procedure gives a scaling with respect to $P$ which — in general — is non–linear, an exception being the uniform distribution over $[0, 1]$. Despite this probability labeling, which is chosen for reasons of an easier application and a better interpretation and understanding, the basis of this axis is a theoretical quantile function. For the exponential probability paper in Fig. 3/5 this quantile function is given by $y_P = -\ln(1 - P)$. Sometimes a second vertical axis is given with the quantile labeling, see Fig. 3/5, which will help to read off estimates of the parameters $a$ and $b$. There are cases where the inverse $F_Y^{-1}(P)$ cannot be given in closed, analytical form but has to be determined numerically, the normal distribution being the most prominent example of this case.

The second, **horizontal axis** of the system of coordinates is for the display of $X$, either in linear scaling or non–linear according to $\widetilde{X} = g(X)$ when a transformation to location–scale type has been made. The quantiles of the distribution of $X$ or of $g(X)$ will lie on a straight line

$$x_P = a + b\,y_P \quad \text{or} \quad \widetilde{x}_P = \widetilde{a} + \widetilde{b}\,\widetilde{y}_P. \tag{3.3}$$

The upper part Fig. 3/5 shows an exponential probability paper and — in the lower part — a PARETO probability paper. Both papers have the same scaling of the ordinate axis whereas their abscissa axes are different, because when $X$ is PARETO distributed, then $\ln(X - a)$ is exponentially distributed, see (1.47a) – (1.48c).

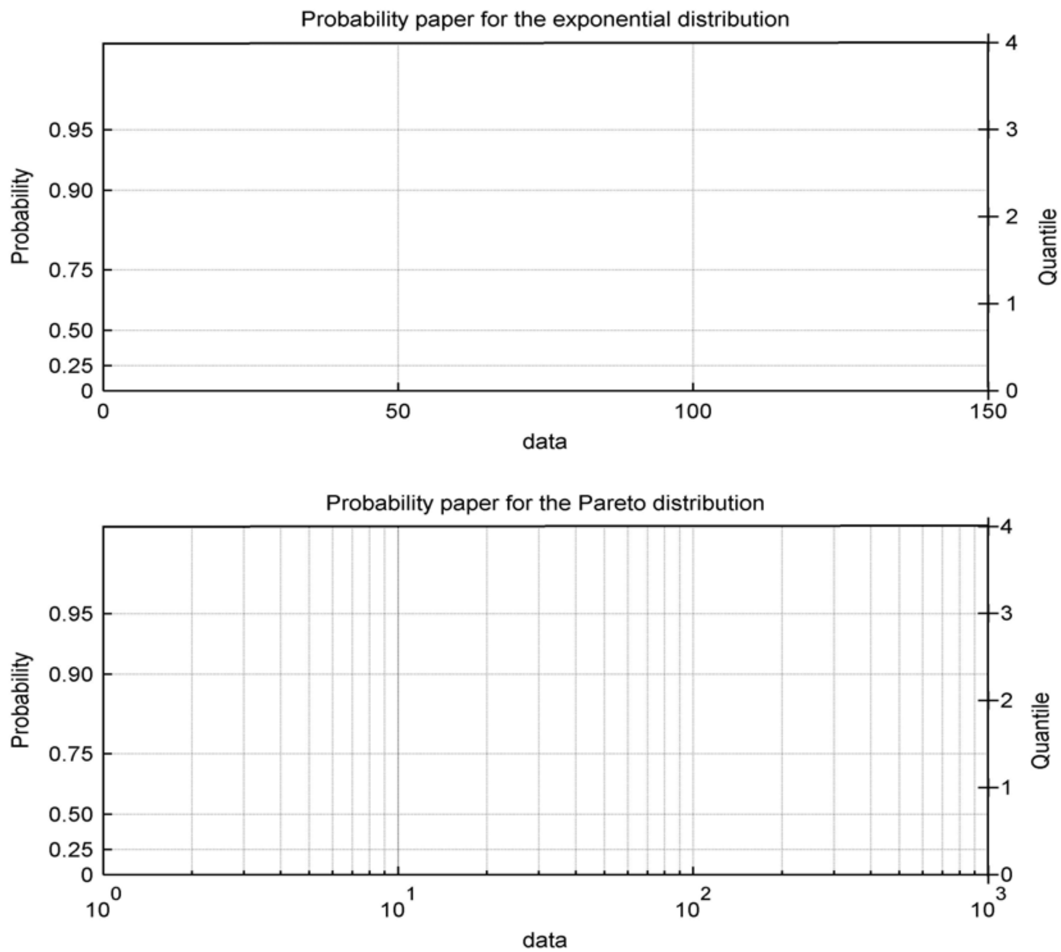
In application to sample data the ordered sample values or their transforms— both regarded as empirical quantiles — are laid off on the horizontal axis and their corresponding ordinate values, called **plotting positions**, on the vertical axis. Please observe the following items when looking at a probability paper:

- The estimation of the parameters $a$ and $b$ or $\widetilde{a}$ and $\widetilde{b}$ is done by minimizing the sum of squared **horizontal** distances, see Chapter 4, i.e. the plotting position is the explaining variable (regressor) and the empirical quantile is the variable to be explained (regressand). Thus, the estimated slope of the straight line is the tangent of the angle between the straight line and the ordinate axis and the estimated parameter $a$ or $\widetilde{a}$ is given by the intersection of the straight line and a horizontal line running through $y_P = 0$ or $\widetilde{y}_P = 0$.

[10] The quantiles of some reduced variables bear special names: **probit** for the normal distribution, **logit** for the logistic distribution and **rankit** for the uniform distribution.

- Log–transformed data are conventionally displayed with the log–axis scaled according to the common or decimal logarithm, see the lower graph in Fig. 3/5, whereas the transformation of the variate is done using the natural logarithm. Thus, when reading off estimates from the estimated straight line with a $\log_{10}$–scaled abscissa one has to pay regard to the modulus $M_{10} = 1\big/\ln 10 \approx 0.43429448$, i.e. one ln–unit is equivalent to $0.43429448\ \log_{10}$–units.

Figure 3/5: Exponential and PARETO probability papers




Let $x_{i:n}$; $i = 1, \ldots, n$; be the ordered observations of a sample from a location–scale distribution. The corresponding reduced observations would be

$$y_{i:n} = (x_{i:n} - a)/b, \tag{3.4a}$$

provided $a$ and $b$ are known. In the latter case we could even compute the corresponding $P$–value on the probability–labeled ordinate

$$P_i = F_Y\left(\frac{x_{i:n} - a}{b}\right) \tag{3.4b}$$

to be plotted over $x_{i:n}$. All these points would lie on a straight line. As $a$ and $b$ are unknown we have to ask how to estimate $y_{i:n}$ or equivalently $P_i$. These estimates are called plotting positions. We have to bear in mind two aims:[11]

1. achieving linearity when the distribution of $X$ has been chosen correctly,
2. efficient estimation of the parameters $a$ and $b$.

As two equivalent quantities can be laid off on the ordinate of a probability paper the search for a plotting position can either start at $y_P$ or at $P = F_Y(y_P)$. The various plotting conventions are based wholly on the sample size $n$ and on the nature of $F_Y(\cdot)$. The numerical values $x_{i:n}$ of the observations will not play a part. The conventions will indicate appropriate values $\widehat{P}_i$ on the $P$–scale or values $\widehat{y}_{i:n}$ on the axis for the reduced variate corresponding to the $\widehat{P}_i$. As in formulas (3.4a,b) the $\widehat{P}_i$ may then be expressed in terms of the $\widehat{y}_{i:n}$:

$$\widehat{P}_i = F_Y(\widehat{y}_{i:n}) \tag{3.5a}$$

or conversely

$$\widehat{y}_{i:n} = F_Y^{-1}(\widehat{P}_i). \tag{3.5b}$$

We first present the so–called "direct" method, see KIMBALL (1960, p. 549), since the rationale involved is based directly on the order number $i$ of $x_{i:n}$.

1. A **naïve estimator** — but simultaneously the maximum likelihood estimator (MLE) — of $F_X(x)=F_Y\big([x-a]/b\big) = F_Y(y)$ is the stair–case function

   $$\widehat{P} = \widehat{F}_X(x) = \begin{cases} 0 & \text{for} \quad x < x_{1:n}, \\ \dfrac{i}{n} & \text{for} \quad x_{i:n} \le x < x_{i+1:n},\ i = 1, 2, \ldots, n-1, \\ 1 & \text{for} \quad x \ge x_{n:n}, \end{cases}$$

   leading to the plotting position

   $$\widehat{P}_i = \frac{i}{n}\,. \tag{3.6a}$$

   A drawback of this proposal is that for all distributions with unlimited range to the right $P = 1$ is not found on the probability scale so that the largest sample value $x_{n:n}$ cannot be displayed.

2. For this reason WEIBULL (1939) has proposed

   $$\widehat{P}_i = \frac{i}{n+1}\,. \tag{3.6b}$$

   Another rationale for the choice of this **WEIBULL position** will be given below.

[11] Plotting positions are — among others — discussed by: BARNETT (1975, 1976), BLOM (1958), HARTER (1984), KIMBALL (1960) and LOONEY/GULLEDGE (1985).

3. The **midpoint position** is
$$\widehat{P}_i = \frac{i - 0.5}{n}\,, \tag{3.6c}$$
motivated by the fact that at $x_{i:n}$ the stair–case moves upwards from $(i-1)/n$ to $i/n$. Thus one believes that $x_{i:n}$ is a quantile of order $P_i$ somewhere between $(i-1)/n$ and $i/n$, and the average of these two is the estimator (3.6c).

4. BLOM (1958) has proposed
$$\widehat{P}_i = \frac{i - 0.375}{n + 0.25}\,. \tag{3.6d}$$
The **BLOM position** guarantees optimality of the linear fit on normal probability paper. Sometimes this plotting position is used for other than normal distributions.

There are plotting positions which rest on the theory of order statistics, see Chapter 2. A first approach along this line departs from the random portion $\Pi_i$ of sample values less than $X_{i:n}$ and tries to estimate $P_i$. A second approach — discussed further down — tries to estimate $y_{i:n}$ and departs from the distribution of $Y_{i:n}$. The random portion $\Pi_i$ is defined as
$$\Pi_i = \Pr(X \le X_{i:n}) = F_Y\left(\frac{X_{i:n} - a}{b}\right) \tag{3.7a}$$
and has the following CDF:
$$\begin{aligned} F_{\Pi_i}(p) &= \Pr\big(\Pi_i \le p\big) \\ &= \sum_{j=i}^{n} \binom{n}{j} p^j\,(1-p)^{n-j}. \end{aligned} \tag{3.7b}$$
The binomial formula (3.7b) results from the fact that we have $n$ independent observations $X_i$ each of them having a probability $p$ to fall underneath the quantile $x_p = F_X^{-1}(p)$. Then $X_{i:n}$ will be smaller than $F_X^{-1}(p)$, if $i$ or more sample values will turn out to be smaller than $F_X^{-1}(p)$. (3.7b) is identical to the CDF of the beta distribution with parameters $i$ and $n-i+1$, the DF being
$$f_{\Pi_i}(p) = \frac{n!}{(i-1)!\,(n-i)!}\,p^{i-1}\,(1-p)^{n-i},\;\; 0 \le p \le 1. \tag{3.7c}$$
Taking the mean, the median or the mode of $\Pi_i$ gives the following three plotting positions:

- $$\widehat{P}_i = \mathrm{E}(\Pi_i) = \frac{i}{n+1} \;-\; \textbf{mean plotting position}, \tag{3.7d}$$

which is equal to (3.6b).

- $$\widehat{P}_i \;\text{ such that }\; F_{\Pi_i}\big(\widehat{P}_i\big) = 0.5$$

This **median plotting position** cannot be given in closed form, but JOHNSON (1964) suggested the following approximation
$$\widehat{P}_i \approx \frac{i - 0.3}{n + 0.4}\,. \tag{3.7e}$$

- $$\widehat{P}_i = \frac{i-1}{n-1} - \quad \textbf{mode plotting position}. \tag{3.7f}$$

(3.7f) turns into $\widehat{P}_1 = 0$ for $i = 1$ and into $\widehat{P}_n = 1$ for $i = n$, and because most of the probability papers do not include the ordinate values $P = 0$ and $P = 1$, the mode plotting position is rarely used.

All plotting positions presented above are estimates of $P_i = F_X(X \leq x_{i:n})$ and all of them do not depend on the sampled distribution. Plotting positions on the scale of the reduced variable $Y = (X - a)/b$ depend on the distribution of the ordered variates $Y_{i:n}$, which on their turn depend on the sampled distribution, see (2.1a-d). The **plotting position** $\boldsymbol{\widehat{y}_{i:n}}$ is chosen as one of the functional parameters of $Y_{i:n}$, either the mean, see (2.9a)

$$\widehat{y}_{i:n} = \mathrm{E}(Y_{i:n}) =: \alpha_{i:n} \tag{3.8a}$$

or the median

$$\widehat{y}_{i:n} = \widetilde{\widetilde{y}}_{i:n} \tag{3.8b}$$

or the mode

$$\widehat{y}_{i:n} = y^*_{i:n}. \tag{3.8c}$$

Nearly all of these plotting positions cannot be given in closed form and have to be computed numerically, see Chapter 5.

Tab. 3/1 summarizes all the plotting positions discussed above. With respect to the choice we can finally state that in most applications it does not matter much how $P_i$ or $y_{i:n}$ are estimated. One will only notice marked differences when the sample size is small. But even these differences are blurred when the straight line is fitted to the data points free–hand.

Table 3/1: Plotting positions

| Name | Quantile axis' value $\widehat{y}_{i:n}$ | Probability axis' value $\widehat{P}_i$ |
|---|---|---|
| Naïve estimator | $\widehat{y}_{i:n} = F_Y^{-1}(i/n)$ | $\widehat{P}_i = i/n$ |
| Midpoint position | $\widehat{y}_{i:n} = F_Y^{-1}[(i-0.5)/n]$ | $\widehat{P}_i = (i-0.5)/n$ |
| BLOM position | $\widehat{y}_{i:n} = F_Y^{-1}[(i-0.375)/(n+0.25)]$ | $\widehat{P}_i = (i-0.375)/(n+0.25)$ |
| Mean position | | |
| – with respect to $\Pi_i$ | $\widehat{y}_{i:n} = F_Y^{-1}[i/(n+1)]$ | $\widehat{P}_i = i/(n+1)$ |
| – with respect to $Y_{i:n}$ | $\widehat{y}_{i:n} = \alpha_{i:n}$ | $\widehat{P}_i = F_Y(\alpha_{i:n})$ |
| Median position | | |
| – with respect to $\Pi_i$ | $\widehat{y}_{i:n} = F_Y^{-1}[(i-0.3)/(n+0.4)]$ | $\widehat{P}_i = (i-0.3)/(n+0.4)$ |
| – with respect to $U_{i:n}$ | $\widehat{y}_{i:n} = \widetilde{\widetilde{y}}_{i:n}$ | $\widehat{P}_i = F_Y(\widetilde{\widetilde{y}}_{i:n})$ |
| Mode position | | |
| – with respect to $\Pi_i$ | $\widehat{y}_{i:n} = F_Y^{-1}[(i-1)/(n-1)]$ | $\widehat{P}_i = (i-1)/(n-1)$ |
| – with respect to $Y_{i:n}$ | $\widehat{y}_{i:n} = y^*_{i:n}$ | $\widehat{P}_i = F_Y(y^*_{i:n})$ |

Probability papers[12] are often combined with a numerical analysis as in Chapter 5 because plots serve many purposes, which no single numerical method can. We conclude this section on probability plotting by first listing its **advantages** and then its limitations.

1. It is fast and simple to use. In contrast, numerical methods may be tedious to compute and may require analytic know–how or an expensive statistical consultant. Moreover, the added accuracy of numerical methods over plots often does not warrant the effort.

2. It presents data in an easily understandable form. This helps one to draw conclusions from data and also to present data to others. The method is easily understood, even by laymen.

3. It provides simple estimates for a distribution: its parameters, the so–called **percentile estimates**, its percentiles, its percentages below or above a given realization. When the paper is supplemented by auxiliary scales one can even read the hazard function, the cumulative hazard function, the mean and the standard deviation.

4. It helps to assess how well a given theoretical distribution fits the data. Sometimes it is even possible to identify and to estimate a mixture of two or at most three distributions.

5. It applies to both complete and censored data. Graphical extrapolation into the censored region is easily done.

6. It helps to spot unusual data. The peculiar appearance of a data plot or certain plotted points may reveal bad data or yield important insight when the cause is determined.

7. It lets one assess the assumptions of analytic methods which will be applied to the data in a later stage.

Some **limitations** of a data plot in comparison to analytic methods are the following:

1. It is not objective. Two people using the same plot may obtain somewhat different estimates. But they usually come to the same conclusion.

2. It does not provide confidence intervals or a statistical hypothesis test with given error–probabilities. However, a plot is often conclusive and leaves little need for such analytic results.

Usually a thorough statistical analysis combines graphical and analytical methods.

[12] For a short remark on the history of probability plotting see BARNETT (1975).

### 3.3.3 Hazard plot

The plotting of multiply and randomly censored data[13] on probability paper causes some problems, and the plotting positions are not easy to compute, see Sect. 4.3.2. Plotting positions in these cases are comfortably determined as shown by NELSON (1972) when **hazard paper** is used. We will first demonstrate how to construct a hazard paper with emphasis on the maximum extreme value distribution of type I and then shortly comment on the choice of the plotting position for this kind of paper.

We still analyze location–scale distributions where $F_X(x) = F_Y(y)$ for $y = (x - a)/b$. The **cumulative hazard function** (= CHF), see Table 1/1, is given by

$$H_X(x) = -\ln[1 - F_X(x)] = -\ln[1 - F_Y(y)] = H_Y(y), \;\; y = \frac{x-a}{b}\,. \tag{3.9a}$$

Let $\Lambda,\ \Lambda > 0$, be a value of the CHF, then

$$y_\Lambda = H_Y^{-1}(\Lambda) \tag{3.9b}$$

and consequently

$$x_\Lambda = a + b\,y_\Lambda. \tag{3.9c}$$

$y_\Lambda$ and $x_\Lambda$ may be called **hazard quantile**, **h–quantile** for short. A hazard paper for a location–scale distribution is constructed by taking the vertical axis of a rectangular system of coordinates to lay off $y_\Lambda$, but the labeling of this axis is according to the corresponding CHF–value $\Lambda$. This procedure gives a scaling with respect to $\Lambda$ which — in general — is non–linear, an exception being the exponential distribution.

The probability grid and the hazard grid for one and the same distribution are related to one another because

$$\Lambda = -\ln(1 - P) \tag{3.9d}$$

or

$$P = 1 - \exp(-\Lambda), \tag{3.9e}$$

where $P$ is a given value of the CDF. Thus, a probability grid may be used for hazard plotting when the $P$–scaling of the ordinate is supplemented by a $\Lambda$–scaling. Conversely, a hazard paper may be used for probability plotting.

The **reduced extreme value distribution of type I for the maximum** has

$$F_Y(y) = \exp\{-\exp(-y)\},$$

so that the CHF–function is

$$\begin{aligned} H_Y(y) &= -\ln[1 - F_Y(y)] \\ &= -\ln[1 - \exp\{-\exp(-y)\}] \end{aligned} \tag{3.10a}$$

[13] See Section 4.1 for an explication of these types of sampling.

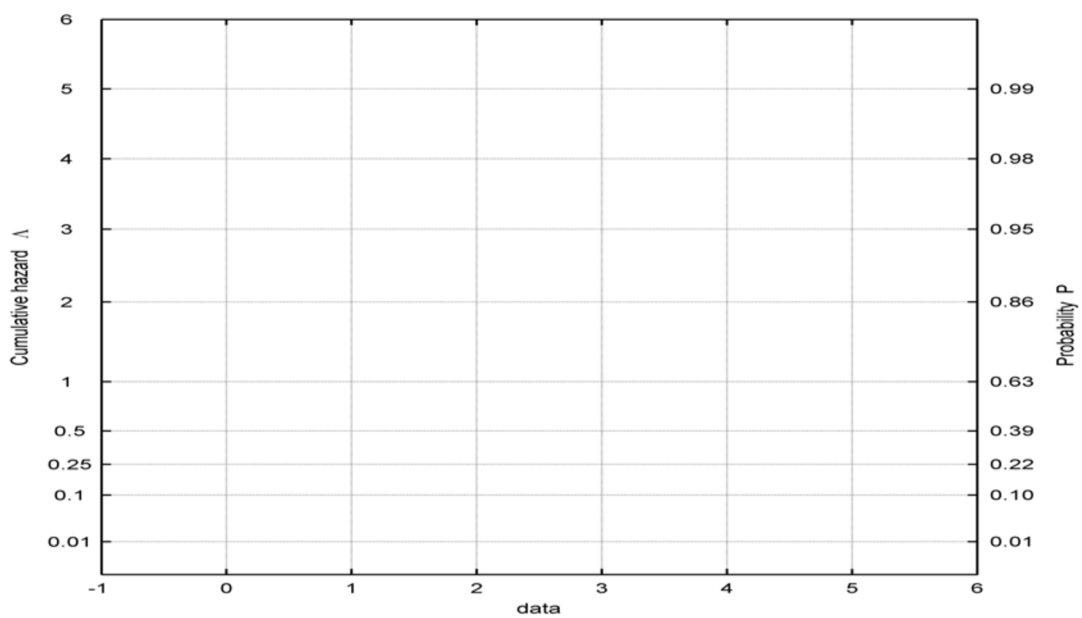
and

$$y_\Lambda = -\ln\{-\ln[1-\exp(-\Lambda)]\},\ \Lambda > 0, \tag{3.10b}$$

and finally

$$x_\Lambda = a - b\,\ln\{-\ln[1-\exp(-\Lambda)]\}. \tag{3.10c}$$

Figure 3/6: Hazard paper for the maximum extreme value distribution of type I




The cumulative hazard value $\Lambda_i$ for the $i$–th ordered observation $x_{i:n}$ has to be estimated. For each uncensored observation the hazard function $h(x\,|\,a,b)$ is estimated by the hazard value

$$\widehat{h}(x_{i:n}\,|\,a,b) = \frac{1}{n-i+1} \tag{3.11a}$$

where $n-i+1$ is the number of sampled items that have not been censored up to $x_{i:n}$. $n-i+1$ is nothing but the **reverse rank**

$$r_i := n-i+1 \tag{3.11b}$$

which results when all observations — censored as well as uncensored — would be ordered in descending order. The **hazard plotting position** is estimated by

$$\widehat{\Lambda}_i = \widehat{H}_X(x_{i:n}) = \sum_{j=1}^{i}\frac{1}{r_j} \tag{3.11c}$$

and the summation is only over those reverse ranks belonging to uncensored observations. NELSON (1972) proved the unbiasedness of (3.11c) when the data are type–II multiply censored.

### 3.3.4 TTT–plot

The TTT–plot is a graph which mainly serves to discriminate between different types of aging, i.e. between constant, decreasing or increasing hazard functions. Thus, this kind of plot is applicable to life–time distributions where the variate only takes non–negative realizations. We will first present TTT–plots for uncensored life tests, make some remarks on the censored–data case and close by listing the advantages and limitations.

Let $0 = X_{0:n} \leq X_{1:n} \leq \ldots \leq X_{n:n}$ denote an ordered sample from a life–time distribution $F(x)$ with survival function $R(x) = 1 - F(x)$. The **total time on test (TTT) statistics**

$$TTT_i = \sum_{j=1}^{i} (n - j + 1)\,(X_{j:n} - X_{j-1:n});\ i = 1, 2, \ldots, n; \tag{3.12a}$$

have been introduced by EPSTEIN/SOBEL (1953) in connection with the inference of the exponential distribution. For a graphical illustration of $TTT_i$ see Fig. 3/7. The sample mean may be expressed as

$$\overline{X} = \frac{1}{n}\,TTT_n. \tag{3.12b}$$

The normalized quantity

$$TTT_i^* = \frac{TTT_i}{TTT_n} = \frac{TTT_i}{n\,\overline{x}},\ \ 0 \leq TTT_i^* \leq 1, \tag{3.12c}$$

is called **scaled total time on test**. By plotting and connecting the points $(i/n, TTT_i^*)$; $i = 0, 1, \ldots, n;$ where $TTT_0 = 0$, by straight line segments we obtain a curve called the **TTT–plot**, see Fig. 3/7. This plotting technique was first suggested by BARLOW/CAMPO (1975) and shows what portion of the total time on test has been accumulated by the portion $i/n$ of items failing first. The TTT–plot has some resemblance to the LORENZ–curve, the difference being that the latter is always strictly convex and lies beneath the 45°–line.

To see what is revealed by a TTT–plot we look at the exponential distribution with $F(x) = 1 - \exp(-x/b),\ x \geq 0$. The theoretical counterpart of (3.12a) for this distribution is

$$\begin{aligned} G_F^{-1}(P) &:= \int\limits_0^{F^{-1}(P)} \exp(-x/b)\,\mathrm{d}x \\ &= \int\limits_0^{-b\,\ln(1-P)} \exp(-x/b)\,\mathrm{d}x \\ &= b\,P. \end{aligned} \tag{3.13a}$$

This is called the **TTT–transform** of $F_X(x)$. The scale invariant transform being the theoretical counterpart of (3.12c) is

$$\frac{G_F^{-1}(P)}{\mu} = \frac{b\,P}{b} = P,\ \ 0 \leq P \leq 1, \tag{3.13b}$$

and the TTT–plot will be the 45°–line starting at the origin. BARLOW/CAMPO (1975) have shown that the theoretical TTT–plot will be

- concave and lying above the 45°–line when $F(x)$ has an increasing hazard function (= IHF),
- convex and lying below the 45°–line when $F(x)$ has a decreasing hazard function (= DHF).

An **empirical TTT–plot** which

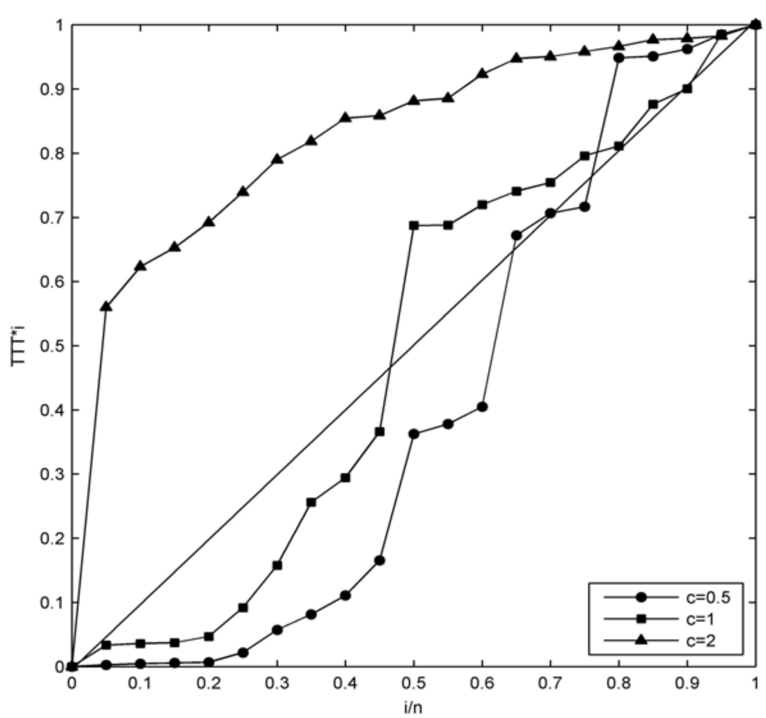
- takes its course randomly around the 45°–line indicates a sample from an exponential distribution,
- is nearly concave (convex) and is mainly above (below) the 45°–line indicates a sample from an IHF (DHF) distribution.

BARLOW/CAMPO (1975) formulated a test of $H_0$ : "$F(x)$ is an exponential distribution" against $H_1$: "$F(x)$ is IHF (DHF)". If the TTT–plot is completely above (below) the 45°–line $H_0$ is rejected in favor of IHF (DHF), the level of significance being $\alpha = 1/n$. Fig. 3/7 shows three empirical TTT–plots of samples from different WEIBULL distributions, where the hazard functon is given by

$$h(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1}, \;\; x \geq a.$$

$h(x)$ is increasing for $c > 1$, decreasing for $c < 1$ and constant for $c = 1$.

Figure 3/7: TTT–plots for several WEIBULL samples of size $n = 20$



With respect to censored life tests the scaled total time on test is defined as

- $$TTT_i^* = \frac{TTT_i}{TTT(T)} \tag{3.14a}$$

  for type–I singly censoring at $T$ and plotted against $i/k;\ i = 1, 2, \ldots, k$; and $k$ failures within $(0, T]$,

- $$TTT_i^* = \frac{TTT_i}{TTT(x_{r:n})} \tag{3.14b}$$

  for type–II singly censoring at the $r$–th failure and plotted against $i/r;\ i = 1, 2, \ldots, r$.

The plots generated in this way will generally lie above those of an uncensored sample of equal size. For the TTT–plot when the sample is multiply censored see BERGMAN/KLEFSJÖ (1984) and WESTBERG/KLEFSJÖ (1994).

Compared to plots on probability paper or on hazard paper the TTT–plot has several advantages:

1. The TTT–plot is well motivated in theory as well as in practice.
2. The TTT–plot is scale invariant.
3. The TTT–plot does not need a special system of coordinates, it is simply displayed in the linearly scaled unit square.
4. Several distributions — even from different families — can be compared.
5. Its interpretation is plain.

The limitations are:

1. It is only possible to give a rough classification into IHF, DHF or exponential or neither of them.
2. Parameter estimation is impossible as is the reading of percentages or of quantiles.

# 4 Linear estimation — Theory and methods

When we have great confidence that a sample comes from a special genuine location–scale distribution or a special transformed to location–scale distribution and this confidence is further supported by a nearly linearly ordered set of points in the matching probability paper we want to estimate the location–scale parameter $(a, b)$. Statistical theory has developed a number of approaches to parameter estimation, e.g. the maximum likelihood method, the method of moments or the method of percentiles. The approach that fits best to the use of probability paper and to the modeling of order statistics from location–scale populations is the **method of least–squares** that — in principle — linearly combines the data to parameter estimates. The special version of the least–squares method which will or has to be applied for estimating $(a, b)$ depends

- on the type of sampling data at hand and
- on what is known about the moments of order statistics needed in the estimation procedure.

We will first comment upon the commonly used types of sampling (Sect. 4.1). The availability of either exact or approximate moments of order statistics has influenced the structure of Sect. 4.2. The methods presented in Sect. 4.3 do not use exact or approximate moments of order statistics as regressors and plotting positions but use regressors which are deduced from the estimated CDF. This will be the case for great sample sizes so that either the observations are grouped and not known individually by their values or because the evaluation of the order statistics' moments is tedious and difficult. Another reason for relying on the estimated CDF is a special sampling plan, e.g. multiple and random censoring, resulting in a sequence of data that consists in a mixture of complete and incomplete observations. In Sect. 4.4 we comment on the goodness–of–fit.

## 4.1 Types of sampling data

In course of time both, statistical theory and practice have developed a great variety of sampling plans each implying a data–type of its own. The most common types of sampling data are depicted in Fig. 4/1. These types are admitted as input to the MATLAB program LEPP performing **l**inear **e**stimation and **p**robability **p**lottting for a great number of location–scale distributions, presented in Chapter 5. In Chapter 6, where LEPP is sketched, we will describe how the data–input has to be arranged. In this section we will comment upon the data–types of Fig. 4/1.

**Grouped data** will arise when the sample size $n$ is large enough to set up classes without loosing too much information. There are several rules of thumb telling how the number $k$

of classes should be linked to the sample size $n$:

$$\begin{aligned} k &\approx \sqrt{n} \text{ for } 30 < n \le 400, \\ k &\approx 5 \log n \text{ for } 30 < n \le 400, \\ k &\approx 1 + 3.3 \log n \text{ for } 30 < n \le 400, \\ k &= 20 \text{ for } n > 400. \end{aligned}$$

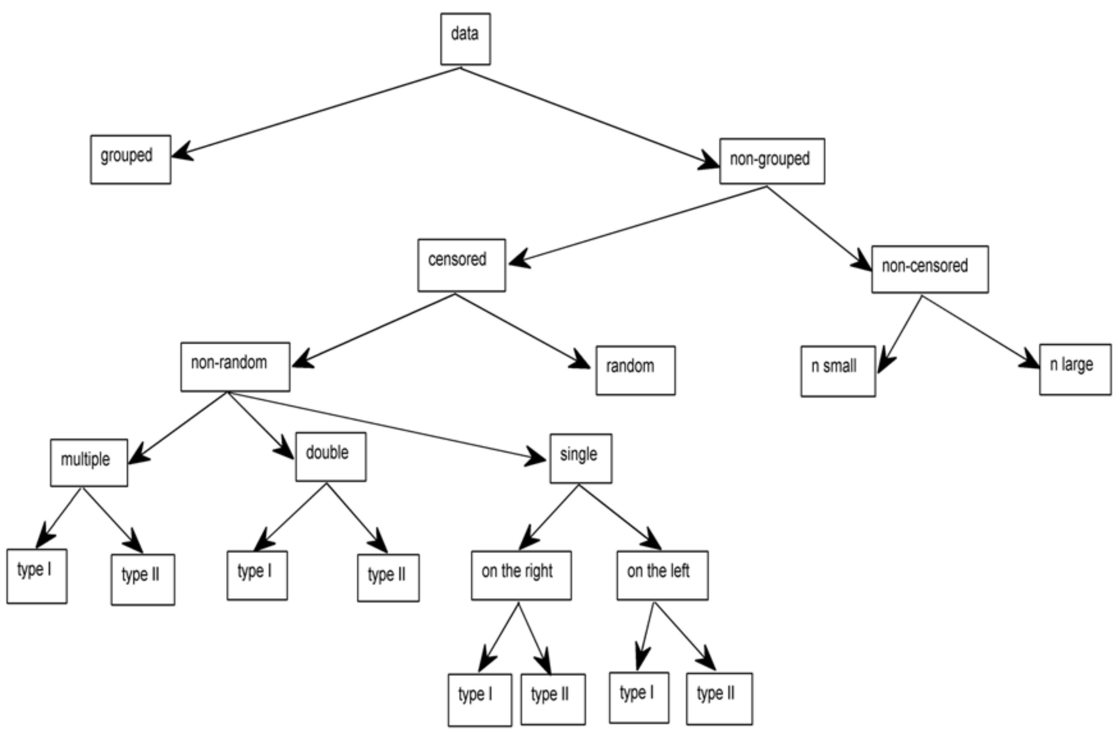
Let $j$ $(j = 1, \dots, k)$ be the class number. Then we define:

$$\begin{aligned} x_j^u &- \text{the upper class limit} \\ x_j^\ell &- \text{the lower class limit,} \end{aligned}$$

and have

$$\begin{aligned} x_j^\ell &< x_j^u \;\forall\, j, \\ x_j^u &< x_{j+1}^u,\; j = 1, \dots, k-1, \\ x_j^u &= x_{j+1}^\ell,\; j = 1, \dots, k-1. \end{aligned}$$

Figure 4/1: Types of sampling data




According to the right–hand continuity of the CDF the classes are open to the left, i.e. observation $x_i$ $(i = 1, \dots, n)$ falls into class $j$ when $x_j^\ell < x_i \le x_j^u$. The class frequency is denoted by $n_j$ with

$$n = \sum_{j=1}^{k} n_j.$$

The class widths $x_j^u - x_j^\ell$ may be equal (= equidistant grouping) or not, depending on the concentration of the observations along the $x$–axis. Often, the first class has $x_1^\ell = -\infty$ and the last class has $x_k^u = \infty$. In the latter case the point $x_k^u$ together with its plotting position cannot be depicted on probability paper. Generally, plotting positions for grouped data are given by cumulated relative frequencies and the regressors by the corresponding percentiles of a reduced distribution.

**Non–grouped data** may be censored or not. A sample is said to be censored when, either by chance or by design, the value of the variate under investigation is unobserved for some of the items. The data–input to LEPP for the censored as well as for the non–censored case has to consists of the ordered observations together with their corresponding ranks, see Sect. 6.2.

**Censoring** is most popular with life testing[1] or for those situations where the variate under consideration is some duration like a mission time or the time of sojourn in a given state. The purpose of censoring in these cases is to come to an early end of the sampling process. Therefore, the following argumentation is mostly done in terms of failure and failure time. But censoring may occur for other types of variates, too, e.g. when the resolution of a measuring device is not high enough to record smaller values or when the exact value of items greater than a given limit are of no interest and only their frequency is what matters. Thus, censoring generally means that observations beyond a given threshold are only known by their numbers but not by their values.

**Censoring** may be at **random** or not. For example, in a medical trial patients may enter the study in a more or less random fashion, according to their time of diagnosis. Some of the patients are lost to follow–up because they move or they die of an illness other than that under investigation or the study is terminated at some prearranged date. Censoring times, that is the length of a patients's time under study, are random. A very simple random censoring process that is often realistic is one in which each individual is assumed to have a lifetime $X$ and a **censoring time** $C$, with $X$ and $C$ being independent continuous variates. $C$ may be the time of a competing risk. Let $(X_i, C_i);\ i = 1, 2, \ldots, n$; be independent and define

$$T_i = \min(X_i, C_i) \ \text{ and } \ \delta_i = \begin{cases} 0 & \text{if} \quad X_i \le C_i, \\ 1 & \text{if} \quad X_i > C_i. \end{cases}$$

The data from observations on $n$ individuals consist of the pairs $(t_i, \delta_i)$, where $\delta_i$ is an indicator telling whether the observation $t_i$ is censored $(\delta_i = 1)$ or not $(\delta_i = 0)$. For randomly censored data the plotting positions are derived from a CDF which is estimated by the KAPLAN–MEIER approach, see Sect. 4.3.2.

[1] See RINNE (2009) for more details on how to do life testing and on the parameters which constitute a life testing plan.

**Non–random censoring** means that the censoring time is planned in one way or the other. In **type–I censoring** or **time censoring**, testing is suspended when a pre–established and fixed testing time has been reached, i.e. the censoring time is directly specified. In this case the numbers of failing items inside and outside the censored interval are random variables and the ranks of the uncensored items are random numbers. With **type–II censoring** or **failure censoring** we indirectly set the censoring time by specifying a certain failure number where either the censoring starts and/or ends. In this case the ranks of the uncensored items are non–random whereas the the censoring time is given by a random order statistic.

**Single censoring** means that we only have one censoring limit either on the side of small values, called **censoring from below** (= **on the left**), or on the side of high values, called **censoring from above** (= **on the right**). With censoring on the left (on the right) the smallest (highest) values are only known by their frequencies. **Double censoring** (= **censoring on both sides**) is characterized by an upper and lower threshold so that only items within these limits are known by their values and those outside are only known by their frequencies. For single and double censoring we can arrange the uncensored items in ascending order and number them by an uninterrupted sequence of natural numbers. Thus, for single and double censoring the plotting positions are identical to the regressors and will be the exact or approximate means of order statistics.

In many practical situations, the initial censoring results only in withdrawal of a portion of the survivors. Those which remain on test continue under observation until ultimate failure or until a subsequent stage of censoring is performed. For sufficiently large samples, censoring may be progressive through several stages, thus leading to a **multiply** (**progressively**) **censored test** or (**hypercensored test**). There are several reasons for multiple censoring:

1. Certain specimens must be withdrawn from a life test prior to failure for use as test objects in related experimentation or to be inspected more thoroughly.
2. In other instances progressively censored samples result from a compromise between the need for more rapid testing and the desire to include at least some extreme life spans in the sample data.
3. When test facilities are limited and when prolonged tests are expensive, the early censoring of a substantial number of sample specimens frees facilities for other tests while specimens, which are allowed to continue on test until subsequent failure, permit observation of extreme sample values.

Multiple censoring results in an ordered series of data which is a mixture of uncensored and censored observations. Thus, we cannot dispose of an uninterrupted sequence of ranks for the uncensored items. We have to estimate a CDF in the same way as is the case of randomly censored data, i.e. by the KAPLAN–MEIER approach.

Resuming what has been said above we have three types of input data to the program LEPP described in Section 6.2:

1. ordered observations coming from an uncensored sample or from either a singly or doubly censored sample, each observation accompanied by its rank,

2. observations coming from either a randomly or a multiply censored sample, each observation bearing an indicator telling whether the observation is censored or not,
3. grouped data consisting of the pairs $(x_j^u, n_j);\ j = 1, \ldots, k$; i.e. the upper class limit and the class frequency.

## 4.2 Estimators based on moments of order statistics

The order statistics of the general and the reduced versions of any location–scale distribution are linked by

$$X_{r:n} = a + b\,Y_{r:n},\ r \in \{1, \ldots, n\}. \tag{4.1a}$$

The moments of $Y_{r:n}$ only depend on $n$ and $r$ and on the form of the reduced density $f_Y(y)$, but not on $a$ and $b$. Let

$$\text{E}(Y_{r:n}) =: \alpha_{r:n}, \quad \text{Cov}(Y_{r:n}, Y_{s:n}) =: \beta_{r,s:n};\ \ r, s \in \{1, \ldots, n\}; \tag{4.1b}$$

then

$$\mu_{r:n} \ = \ \text{E}(X_{r:n}) \ = \ a + b\,\alpha_{r:n} \tag{4.1c}$$

$$\sigma_{r,s:n} \ = \ \text{Cov}(X_{r:n}, X_{s:n}) \ = \ b^2\,\beta_{r,s:n}, \tag{4.1d}$$

where $\alpha_{r:n}$ and $\beta_{r,s:n}$ can be evaluated once and for all, see the hints to published tables for some location–scale distributions in Sect. 5.2. $\text{E}(X_{r:n})$ is linear in the parameters $a$ and $b$ with known coefficients $\alpha_{r:n}$ and $\text{Cov}(X_{r:n}, X_{s:n})$ is known apart from $b^2$. Reverting to (4.1a), giving an equation for the variates, we can establish the following **regression model**:

$$X_{r:n} = a + b\,\alpha_{r:n} + \varepsilon_r, \tag{4.2a}$$

where $\varepsilon_r$ is a variate expressing the difference between $X_{r:n}$ and its mean $\mu_{r:n} = \text{E}(X_{r:n})$. Thus,

$$\text{E}(\varepsilon_r) \ = \ 0,\ \ r \in \{1, \ldots, n\}, \tag{4.2b}$$

$$\text{Cov}(\varepsilon_r, \varepsilon_s) \ = \ \text{Cov}(X_{r:n}, X_{s:n});\ \ r, s \in \{1, \ldots, n\}. \tag{4.2c}$$

We collect the variances and covariances of the order statistics in a matrix, called **variance–covariance matrix**, which — for the vector $\boldsymbol{y}' = (Y_{1:n}, \ldots, Y_{n:n})$ of the reduced order statistics — reads

$$\text{Var}(\boldsymbol{y}) := \boldsymbol{B} := \big(\beta_{r,s:n}\big) \tag{4.3a}$$

and

$$\text{Var}(\boldsymbol{x}) := b^2\,\boldsymbol{B} \tag{4.3b}$$

for the vector $\boldsymbol{x}' = (X_{1:n}, \ldots, X_{n:n})$. The matrix $\boldsymbol{B}$ is square of size $n \times n$, symmetric and **positive semidefinite**, because the variance of a linear combination of $\boldsymbol{x}$, say

$$w = \boldsymbol{x}'\,\boldsymbol{q}, \tag{4.4a}$$

$\boldsymbol{q} \neq \boldsymbol{o}$ a real vector of weights, has variance

$$\text{Var}(\boldsymbol{x}'\,\boldsymbol{q}) = b^2\,\boldsymbol{q}'\,\boldsymbol{B}\,\boldsymbol{q} \geq 0. \tag{4.4b}$$

### 4.2.1 GLS estimators

The parameters $a$ and $b$ may be estimated by the method of least–squares, but the regression model (4.2a) does not fulfill all those conditions that are necessary for assuring the simple OLS estimation approach[2] to give BLU estimators,[3] i.e. estimators with minimum variance within the class of unbiased linear estimators. For OLS to be optimal in the sense of BLU the variance–covariance matrix of the regressands should be — up to a certain scalar — an identity matrix, meaning that the regressands have to be of equal variance (= **homoscedastic**) and **uncorrelated**. This will not be the case with the variance–covariance matrices of order statistics as is seen in the variance–covariance matrix $\boldsymbol{B}$ and correlation matrix $\boldsymbol{R}$ of reduced order statistics in a sample of size $n = 6$ from three distribution having different shapes: the uniform distribution with equal density, the normal distribution (symmetric and thin tails) and the exponential distribution (positively skew). We have only reproduced the upper triangle of these symmetric matrices and also omitted the leading zeros.

Uniform distribution, $n = 6$

$$\boldsymbol{B} = \begin{pmatrix} .0153 & .0128 & .0102 & .0077 & .0051 & .0026 \\ & .0255 & .0204 & .0153 & .0102 & .0021 \\ & & .0306 & .0230 & .0153 & .0027 \\ & & & .0306 & .0204 & .0102 \\ & & & & .0255 & .0128 \\ & & & & & .0153 \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} 1 & .6455 & .4714 & .3536 & .2582 & .1667 \\ & 1 & .7303 & .5477 & .4000 & .2582 \\ & & 1 & .7500 & .5477 & .3536 \\ & & & 1 & .7303 & .4714 \\ & & & & 1 & .6455 \\ & & & & & 1 \end{pmatrix}$$

Normal distribution, $n = 6$

$$\boldsymbol{B} = \begin{pmatrix} .4159 & .2085 & .1394 & .1024 & .0774 & .0563 \\ & .2796 & .1890 & .1397 & .1059 & .0774 \\ & & .2462 & .1833 & .1397 & .1024 \\ & & & .2462 & .1890 & .1394 \\ & & & & .2796 & .2085 \\ & & & & & .4159 \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} 1 & .6114 & .4357 & .3203 & .2269 & .1355 \\ & 1 & .7203 & .5323 & .3789 & .2269 \\ & & 1 & .7444 & .5323 & .3203 \\ & & & 1 & .7203 & .4357 \\ & & & & 1 & .6114 \\ & & & & & 1 \end{pmatrix}$$

Exponential distribution, $n = 6$

$$\boldsymbol{B} = \begin{pmatrix} .0278 & .0278 & .0278 & .0278 & .0278 & .0278 \\ & .0667 & .0667 & .0667 & .0667 & .0667 \\ & & .1303 & .1303 & .1303 & .1303 \\ & & & .2424 & .2414 & .2414 \\ & & & & .4914 & .4914 \\ & & & & & 1.4914 \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} 1 & .6402 & .4618 & .3392 & .2379 & .1365 \\ & 1 & .7213 & .5299 & .3714 & .2132 \\ & & 1 & .7364 & .5148 & .2965 \\ & & & 1 & .7009 & .4023 \\ & & & & 1 & .5704 \\ & & & & & 1 \end{pmatrix}$$

[2] OLS – **o**rdinary **l**east–**s**quares

[3] BLU – **b**est **l**inear **u**nbiased

When we take a closer look at the variance–covariance matrix of the order statistics we see two characteristics:

- **heteroscedasticity**, i.e. differing variances on the diagonal and
- **autocorrelation**, i.e. non–zero off–diagonal elements.

The autocorrelation pattern, which will be visible in the **correlation matrix**[4]

$$\boldsymbol{R} = \mathrm{diag}(\boldsymbol{B})^{-1/2}\, \boldsymbol{B}\, \mathrm{diag}(\boldsymbol{B})^{-1/2}$$

shows

- **positive correlation coefficients** meaning that when an order statistic $X_{r:n}$ has a realization above (below) its mean $\mu_{r:n}$ the realization of the following order statistic $X_{r+1:n}$ will deviate from its mean $\mu_{r+1:n}$ in the same direction with high probability, thus leading to a wave–like pattern of the scatter plot around a straight line on the probability paper,
- a **strength of correlation** that **declines** with growing distance between the orders $r$ and $s$.

To assure the validity of the **GAUSS–MARKOV theorem**[5] AITKEN (1935) proposed a modification of OLS method when the variance–covariance matrix of the regressands — up to a scalar — is not the identity matrix. This GLS method (**g**eneral **l**east–**s**quares) was first applied to the estimation of the location–scale parameter $(a, b)$ by LLOYD (1952) leading to the so–called **LLOYD's estimator**. We will present this estimator for an arbitrary location–scale distribution in Sect. 4.2.1.1 before turning to a symmetric parent distribution (Sect. 4.2.1.2), in both cases assuming an uncensored sample. Finally (Sect. 4.2.1.3), we comment upon how to proceed with censored samples.

#### 4.2.1.1 GLS for a general location–scale distribution

For reasons of compact notation we introduce the following vectors and matrices:

$$\boldsymbol{x} := \begin{pmatrix} X_{1:n} \\ X_{2:n} \\ \vdots \\ X_{n:n} \end{pmatrix},\ \boldsymbol{1} := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},\ \boldsymbol{\alpha} := \begin{pmatrix} \alpha_{1:n} \\ \alpha_{2:n} \\ \vdots \\ \alpha_{n:n} \end{pmatrix},\ \boldsymbol{\varepsilon} := \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},\ \boldsymbol{\theta} := \begin{pmatrix} a \\ b \end{pmatrix},\ \boldsymbol{A} := \begin{pmatrix} \boldsymbol{1} & \boldsymbol{\alpha} \end{pmatrix}. \quad (4.5a)$$

4 $\mathrm{diag}(\boldsymbol{B})$ is the diagonal matrix with the diagonal elements of $\boldsymbol{B}$ and zeros otherwise.

5 This theorem states that least–squares estimators are unbiased and best (= having minimum variance) in the class of linear estimators. However, there is no assurance that estimators in other classes, such as maximum likelihood estimators, may not be more efficient. We will not discuss questions of efficiency in this text but refer the reader to other textbooks.

The regression model (4.2a), pertaining to the sample, now reads

$$\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{4.5b}$$

with variance–covariance matrix

$$\text{Var}(\boldsymbol{x}) = b^2\,\boldsymbol{B}. \tag{4.5c}$$

Under GLS we have to minimize the following generalized variance with respect to $\boldsymbol{\theta}$:

$$Q = (\boldsymbol{x} - \boldsymbol{A}\,\boldsymbol{\theta})'\,\boldsymbol{\Omega}\,(\boldsymbol{x} - \boldsymbol{A}\,\boldsymbol{\theta}) \ \text{ where } \ \boldsymbol{\Omega} = \boldsymbol{B}^{-1}, \tag{4.6a}$$

yielding the **GLS estimator**[6] of $\boldsymbol{\theta}$

$$\widehat{\boldsymbol{\theta}} = \big(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{x}. \tag{4.6b}$$

$\widehat{\boldsymbol{\theta}}$ is a linear combination of $\boldsymbol{x}$ and according to (4.4b) the **variance–covariance matrix of $\widehat{\boldsymbol{\theta}}$** is given by

$$\begin{aligned}\text{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) &= \big(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}\,\big(b^2\,\boldsymbol{B}\big)\,\boldsymbol{\Omega}\,\boldsymbol{A}\,\big(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}\big)^{-1}\\ &= b^2\,\big(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}\big)^{-1}.\end{aligned} \tag{4.6c}$$

Because $\boldsymbol{B}$ is symmetric its inverse $\boldsymbol{\Omega}$ is symmetric, too. From the definitions in (4.5a) the matrix $\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}$ follows as

$$\boldsymbol{D} := \boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A} = \begin{pmatrix}\boldsymbol{1}'\\ \boldsymbol{\alpha}'\end{pmatrix}\boldsymbol{\Omega}\,\big(\boldsymbol{1}\ \ \boldsymbol{\alpha}\big) = \begin{pmatrix}\boldsymbol{1}'\,\boldsymbol{\Omega}\,\boldsymbol{1} & \boldsymbol{1}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}\\ \boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{1} & \boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}\end{pmatrix}, \tag{4.7}$$

all the elements of this $(2 \times 2)$–matrix being scalar. The two elements $\widehat{a}$ and $\widehat{b}$ of $\widehat{\boldsymbol{\theta}}$ in (4.6b) are explicitly given when reverting to the familiar representation of the inverse of the $(2 \times 2)$–matrix $\boldsymbol{D}$ using its determinant

$$\Delta = \det \boldsymbol{D}$$

by

$$\begin{aligned}\widehat{\boldsymbol{\theta}} = \begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix} &= \frac{1}{\Delta}\begin{pmatrix}\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha} & -\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{1}\\ -\boldsymbol{1}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha} & \boldsymbol{1}'\,\boldsymbol{\Omega}\,\boldsymbol{1}\end{pmatrix}\begin{pmatrix}\boldsymbol{1}'\,\boldsymbol{\Omega}\\ \boldsymbol{\alpha}'\,\boldsymbol{\Omega}\end{pmatrix}\boldsymbol{x}\\ &= \frac{1}{\Delta}\begin{pmatrix}\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}\,\boldsymbol{1}'\,\boldsymbol{\Omega} - \boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{1}\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\\ -\boldsymbol{1}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}\,\boldsymbol{1}'\,\boldsymbol{\Omega} + \boldsymbol{1}'\,\boldsymbol{\Omega}\,\boldsymbol{1}\,\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\end{pmatrix}\boldsymbol{x}\end{aligned} \tag{4.8a}$$

[6] When $\boldsymbol{B}$ is the identity matrix $\boldsymbol{I}$, we have $\boldsymbol{\Omega} = \boldsymbol{B}^{-1} = \boldsymbol{I}$ and (4.6b) turns into the familiar OLS estimator.

or

$$\widehat{a} = -\boldsymbol{\alpha}' \boldsymbol{H} \boldsymbol{x} \tag{4.8b}$$

$$\widehat{b} = \mathbf{1}' \boldsymbol{H} \boldsymbol{x}, \tag{4.8c}$$

where $\boldsymbol{H}$ is the skew–symmetric matrix[7]

$$\boldsymbol{H} = \frac{\boldsymbol{\Omega}\left(\mathbf{1}\,\boldsymbol{\alpha}' - \boldsymbol{\alpha}\,\mathbf{1}'\right)\boldsymbol{\Omega}}{\Delta}. \tag{4.8d}$$

The row vector

$$\boldsymbol{a}' := -\boldsymbol{\alpha}' \boldsymbol{H} \tag{4.8e}$$

contains the weights which linearly combine the order statistics of $\boldsymbol{x}$ into the estimator $\widehat{a} = \boldsymbol{a}' \boldsymbol{x}$. Likewise

$$\boldsymbol{b}' := \mathbf{1}' \boldsymbol{H} \tag{4.8f}$$

is another row vector of weights combining the elements of $\boldsymbol{x}$ into the estimator $\widehat{b} = \boldsymbol{b}' \boldsymbol{x}$. The weights in $\boldsymbol{a}$ and $\boldsymbol{b}$ only depend on the known moments $\alpha_{r:n}$ and $\beta_{r,s:n}$ of the reduced order statistics and they can be evaluated and tabulated once and for all what has been done by several authors for selected distributions, see Sect. 5.2. The sums of the weights are:

$$-\boldsymbol{\alpha}' \boldsymbol{H}\,\mathbf{1} = 1 \quad \text{and} \quad \mathbf{1}' \boldsymbol{H}\,\mathbf{1} = 0.$$

We further notice that

$$\begin{pmatrix} \mathbf{1}' \\ \boldsymbol{\alpha}' \end{pmatrix} \left(-\boldsymbol{H}'\boldsymbol{\alpha},\ \boldsymbol{H}'\,\mathbf{1}\right) = \begin{pmatrix} -\mathbf{1}'\boldsymbol{H}\,\boldsymbol{\alpha} & \mathbf{1}'\boldsymbol{H}'\,\mathbf{1} \\ -\boldsymbol{\alpha}'\boldsymbol{H}'\boldsymbol{\alpha} & \boldsymbol{\alpha}'\boldsymbol{H}'\,\mathbf{1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} -\boldsymbol{\alpha}'\boldsymbol{H} \\ \mathbf{1}'\boldsymbol{H} \end{pmatrix} \left(\mathbf{1},\ \boldsymbol{\alpha}\right) = \begin{pmatrix} -\boldsymbol{\alpha}'\boldsymbol{H}\,\mathbf{1} & -\boldsymbol{\alpha}'\boldsymbol{H}\,\boldsymbol{\alpha} \\ \mathbf{1}'\boldsymbol{H}\,\mathbf{1} & \mathbf{1}'\boldsymbol{H}\,\boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The elements of the variance–covariance matrix in (4.6c) can also be written in more detail as

$$\text{Var}\big(\widehat{a}\big) = b^2\,\frac{\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}}{\Delta}, \tag{4.9a}$$

$$\text{Var}\big(\widehat{b}\big) = b^2\,\frac{\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1}}{\Delta} \tag{4.9b}$$

$$\text{Cov}\big(\widehat{a},\widehat{b}\big) = -b^2\,\frac{\mathbf{1}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}}{\Delta}. \tag{4.9c}$$

[7] A square matrix $\boldsymbol{C}$ is said to be skew–symmetric if $\boldsymbol{C}' = -\boldsymbol{C}$.

As is the rule with regression estimating a straight line the estimators of the two parameters will be negatively correlated. Applying the estimator $\widehat{b}$ to (4.9a–c) gives the estimated variances and covariance.

**Excursus: Proving the BLU property of GLS estimators**

The **linearity** of the estimators $\widehat{a}$ and $\widehat{b}$ is evident by looking at (4.8b,c) and (4.8e,f). To prove the **unbiasedness** of $\widehat{\boldsymbol{\theta}}$ given by (4.6b) we substitute $\boldsymbol{x}$ by $\boldsymbol{A}\,\boldsymbol{\theta}+\boldsymbol{\varepsilon}$, see (4.5b), and get

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= (\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{a})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}\,(\boldsymbol{A}\,\boldsymbol{\theta}+\boldsymbol{\varepsilon})\\ &= \underbrace{(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})}_{=\boldsymbol{I}}\,\boldsymbol{\theta}+(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{\varepsilon}\\ &= \boldsymbol{\theta}+\underbrace{(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}}_{=\boldsymbol{K}}\,\boldsymbol{\varepsilon}.\end{aligned}$$

Now, upon forming expectation we find

$$\mathrm{E}\big(\widehat{\boldsymbol{\theta}}\big)=\boldsymbol{\theta}+\boldsymbol{K}\,\mathrm{E}(\boldsymbol{\varepsilon}),$$

and as $\boldsymbol{\varepsilon}$ is the vector of deviations between the $X_{r:n}$ and their means $\mu_{r:n}$ with $\mathrm{E}(X_{r:n}-\mu_{r:n})=0\ \forall\, r$, we have $\mathrm{E}(\boldsymbol{\varepsilon})=\mathrm{E}(\boldsymbol{x}-\boldsymbol{\mu})=\boldsymbol{o},\ \ \boldsymbol{\mu}':=(\mu_{1:n},\ldots,\mu_{n:n})$. Thus, the unbiasedness of $\widehat{\boldsymbol{\theta}}$ has been proven.

To prove the **minimum variance property** we assume that

$$\widetilde{\boldsymbol{\theta}}=\boldsymbol{C}\boldsymbol{x} \tag{4.10a}$$

is some other linear estimator of $\boldsymbol{\theta}$, $\boldsymbol{C}$ being a $(2\times n)$–matrix of non–random weights. We have to show that $\mathrm{Var}\big(\widetilde{\boldsymbol{\theta}}\big)-\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\big)$ is a positive semidefinite matrix, meaning among others that $\mathrm{Var}\big(\widetilde{a}\big)\geq\mathrm{Var}\big(\widehat{a}\big)$ and $\mathrm{Var}\big(\widetilde{b}\big)\geq\mathrm{Var}\big(\widehat{b}\big)$. Upon taking expectation in (4.10a) we have

$$\begin{aligned}\mathrm{E}\big(\widetilde{\boldsymbol{\theta}}\big) &= \mathrm{E}\big[\boldsymbol{C}(\boldsymbol{A}\,\boldsymbol{\theta}+\boldsymbol{\varepsilon})\big]\\ &= \boldsymbol{C}\boldsymbol{A}\,\boldsymbol{\theta}+\boldsymbol{C}\,\mathrm{E}(\boldsymbol{\varepsilon}),\end{aligned} \tag{4.10b}$$

so $\widetilde{\boldsymbol{\theta}}$ will be unbiased iff

$$\boldsymbol{C}\boldsymbol{A}=\boldsymbol{I}\quad\text{and}\quad\mathrm{E}(\boldsymbol{\varepsilon})=\boldsymbol{o}. \tag{4.10c}$$

The variance–covariance matrix of $\widetilde{\boldsymbol{\theta}}$ can be found by replacing the GLS weighing vector $(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}$ in (4.6c) with $\boldsymbol{C}$. The result is

$$\mathrm{Var}\big(\widetilde{\boldsymbol{\theta}}\big)=b^2\,\boldsymbol{C}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{C}', \tag{4.11a}$$

remember $\boldsymbol{\Omega}^{-1}=\boldsymbol{B}$. Now let

$$\boldsymbol{L}=\boldsymbol{C}-(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega} \tag{4.11b}$$

so

$$\boldsymbol{L}\,\boldsymbol{x}=\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}. \tag{4.11c}$$

Then,

$$\operatorname{Var}\big(\widetilde{\boldsymbol{\theta}}\,\big) = b^2\Big\{\big[\boldsymbol{L} + (\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}\big]\boldsymbol{B}\big[\boldsymbol{L} + (\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}\big]'\Big\}. \tag{4.11d}$$

We know that

$$\boldsymbol{C}\boldsymbol{A} = \boldsymbol{I} = \boldsymbol{L}\,\boldsymbol{A} + (\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A},$$

so $\boldsymbol{L}\,\boldsymbol{A}$ must be equal to $\boldsymbol{O}$. Therefore, after some matrix manipulation, we find:

$$\begin{aligned}\operatorname{Var}\big(\widetilde{\boldsymbol{\theta}}\,\big) &= b^2\left(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}\right)^{-1} + b^2\,\boldsymbol{L}\,\boldsymbol{B}\,\boldsymbol{L}' \\ &= \operatorname{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) + b^2\,\boldsymbol{L}\,\boldsymbol{B}\,\boldsymbol{L}'. \end{aligned} \tag{4.11e}$$

A quadratic form of $\boldsymbol{L}\,\boldsymbol{B}\,\boldsymbol{L}'$ is $\boldsymbol{q}'\,\boldsymbol{L}\,\boldsymbol{B}\,\boldsymbol{L}'\,\boldsymbol{q} = \boldsymbol{z}'\,\boldsymbol{B}\,\boldsymbol{z}$, $\boldsymbol{q}$ being a real vector with two elements. Because $\boldsymbol{B}$ is positive semidefinite we have

$$\boldsymbol{z}'\,\boldsymbol{B}\,\boldsymbol{z} \geq 0.$$

Thus, $\operatorname{Var}\big(\widetilde{\boldsymbol{\theta}}\,\big)$ is the variance–covariance matrix of the GLS estimator plus a non–negative definite matrix. Therefore, every quadratic form in $\operatorname{Var}\big(\widetilde{\boldsymbol{\theta}}\,\big)$ is larger or equal than the corresponding quadratic form in $\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}\,\big)$. Especially for $\boldsymbol{z}' = (1,\ 0)$ we find $\operatorname{Var}(\,\widetilde{a}\,) \geq \operatorname{Var}(\,\widehat{a}\,)$ and for $\boldsymbol{z}' = (0,\ 1)$ we have $\operatorname{Var}(\,\widetilde{b}\,) \geq \operatorname{Var}(\,\widehat{b}\,)$.

---

---

**Example 4/1: GLS estimating $(a, b)$ for an exponential distribution ( uncensored sample)**

The following $n = 10$ observations resulted from a Monte Carlo simulation of $X \sim EX(10, 100)$:

$$\boldsymbol{x}' = (25,\ 26,\ 40,\ 40,\ 81,\ 89,\ 130,\ 149,\ 232,\ 330)$$

For $n = 10$ we have the vector of reduced means:

$$\boldsymbol{\alpha}' = (0.1000,\ 0.2111,\ 0.3361,\ 0.4790,\ 0.6456,\ 0.8456,\ 1.0956,\ 1.4290,\ 1.9290,\ 2.9290),$$

and the elements of the upper triangle of $\boldsymbol{B}$ are:

$$\begin{aligned}
\beta_{1,1:10} = \beta_{1,2:10} = \ldots = \beta_{1,10:10} &= 0.0100,\\
\beta_{2,2:10} = \beta_{2,3:10} = \ldots = \beta_{2,10:10} &= 0.0223,\\
\beta_{3,3:10} = \beta_{3,4:10} = \ldots = \beta_{3,10:10} &= 0.0380,\\
\beta_{4,4:10} = \beta_{4,5:10} = \ldots = \beta_{4,10:10} &= 0.0584,\\
\beta_{5,5:10} = \beta_{5,6:10} = \ldots = \beta_{5,10:10} &= 0.0862,\\
\beta_{6,6:10} = \beta_{6,7:10} = \ldots = \beta_{6,10:10} &= 0.1262,\\
\beta_{7,7:10} = \beta_{7,8:10} = \ldots = \beta_{7,10:10} &= 0.1887,\\
\beta_{8,8:10} = \beta_{8,9:10} = \beta_{8,10:10} &= 0.2998,\\
\beta_{9,9:10} = \beta_{9,10:10} &= 0.5498,\\
\beta_{10,10:10} &= 1.5498.
\end{aligned}$$

$\boldsymbol{\Omega}$, the inverse of $\boldsymbol{B}$, is a banded matrix:

$$\boldsymbol{\Omega} = \begin{pmatrix} 181 & -81 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -81 & 145 & -64 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -64 & 113 & -49 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -49 & 85 & -36 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -36 & 61 & -25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -25 & 41 & -16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -16 & 25 & -9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -9 & 13 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & 5 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

The estimated parameters according to (4.6b) are

$$\widehat{a} = 15.0889 \quad \text{and} \quad \widehat{b} = 99.1111,$$

and the estimated variance–covariance according to (4.6c) is

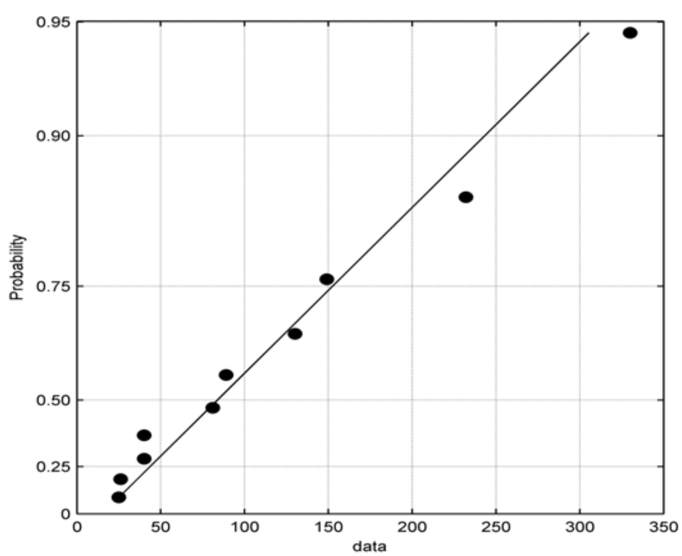
$$\widehat{\text{Var}}\big(\widehat{a}, \widehat{b}\,\big) = 1000 \begin{pmatrix} 0.1091 & -0.1091 \\ -0.1091 & 1.0914 \end{pmatrix}.$$

The weighting vectors according to (4.8a,b) read:

$$\begin{aligned} \boldsymbol{a}' &= (1.1,\ -0.0\overline{1},\ -0.0\overline{1},\ -0.0\overline{1},\ -0.0\overline{1},\ -0.0\overline{1},\ -0.0\overline{1},\ -0.0\overline{1},\ -0.0\overline{1},\ -0.0\overline{1}\,), \\ \boldsymbol{b}' &= (-1,\ 0.\overline{1},\ 0.\overline{1},\ 0.\overline{1},\ 0.\overline{1},\ 0.\overline{1},\ 0.\overline{1},\ 0.\overline{1},\ 0.\overline{1},\ 0.\overline{1}\,). \end{aligned}$$

The following graph shows the data together with the estimated regression line on exponential probability paper.

Figure 4/2: Probability plot and estimated regression line (uncensored sample of size $n = 10$ from an exponential distribution)



#### 4.2.1.2 GLS for a symmetric location–scale distribution

A great number of location–scale distributions are symmetric. In these cases some simplification of the GLS formulas (4.6) through (4.9) is possible. First, the distribution of $\boldsymbol{x'} = (X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ is the same as that of $(-X_{n:n}, -X_{n-1:n}, \ldots, -X_{1:n})$. Let

$$\begin{pmatrix} -X_{n:n} \\ -X_{n-1:n} \\ \vdots \\ -X_{1:n} \end{pmatrix} = -\boldsymbol{J} \begin{pmatrix} X_{1:n} \\ X_{2:n} \\ \vdots \\ X_{n:n} \end{pmatrix} = -\boldsymbol{J}\,\boldsymbol{x} \tag{4.12a}$$

where

$$\boldsymbol{J} = \begin{pmatrix} 0 & \cdot & \cdot & \cdot & 1 \\ \cdot & & & \cdot & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot \\ 1 & \cdot & \cdot & \cdot & 0 \end{pmatrix} \tag{4.12b}$$

is a symmetric **permutation matrix** with the following properties:

$$\boldsymbol{J} = \boldsymbol{J'} = \boldsymbol{J}^{-1}, \;\; \boldsymbol{J'}\,\boldsymbol{1} = \boldsymbol{1}, \tag{4.12c}$$

$$\boldsymbol{J}^{2\,k} = \boldsymbol{I}, \;\; \boldsymbol{J}^{2\,k+1} = \boldsymbol{J};\; k = 1, 2, \ldots; \tag{4.12d}$$

$\boldsymbol{I}$ being an identity matrix. Since

$$\boldsymbol{x} \stackrel{d}{=} -\boldsymbol{J}\,\boldsymbol{x}, \tag{4.13a}$$

i.e. both vectors are equivalent in distribution, we have

$$\mathrm{E}(\boldsymbol{x}) = -\boldsymbol{J}\,\mathrm{E}(\boldsymbol{x}) \tag{4.13b}$$

$$\mathrm{Var}(\boldsymbol{x}) = \mathrm{Var}(\boldsymbol{J}\,\boldsymbol{x}) = \boldsymbol{J}\,\mathrm{Var}(\boldsymbol{x})\,\boldsymbol{J}. \tag{4.13c}$$

For the reduced order statistics $\boldsymbol{y}$ with $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\alpha}$ and $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{B}$ the formulas (4.13b,c) turn into

$$\boldsymbol{\alpha} = -\boldsymbol{J}\,\boldsymbol{\alpha} \tag{4.13d}$$

$$\boldsymbol{B} = \boldsymbol{J}\,\boldsymbol{B}\,\boldsymbol{J}. \tag{4.13e}$$

For $\boldsymbol{\Omega} = \boldsymbol{B}^{-1}$ we notice that

$$\boldsymbol{\Omega} = \boldsymbol{J}^{-1}\boldsymbol{B}^{-1}\boldsymbol{J}^{-1} = \boldsymbol{J}\,\boldsymbol{\Omega}\,\boldsymbol{J} \tag{4.13f}$$

because of (4.12c). It follows that

$$\begin{aligned} \boldsymbol{1'}\,\boldsymbol{\Omega}\,\boldsymbol{\alpha} &= \boldsymbol{1'}\,(\boldsymbol{J}\,\boldsymbol{\Omega}\,\boldsymbol{J})\,(-\boldsymbol{J}\,\boldsymbol{\alpha}) \\ &= -(\boldsymbol{1'}\,\boldsymbol{J})\,\boldsymbol{\Omega}\,\boldsymbol{J}^2\,\boldsymbol{\alpha} \\ &= -\boldsymbol{1'}\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}. \end{aligned} \tag{4.14a}$$

Thus, the scalars $\mathbf{1}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}$ and $-\mathbf{1}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}$ must be zero as is also the case with their transposes $\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\mathbf{1}$ and $-\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\mathbf{1}$. Looking at (4.9c) we recognize that $\widehat{a}$ and $\widehat{b}$ are uncorrelated. Furthermore, (4.7) simplifies to a diagonal matrix:

$$\boldsymbol{D} = \begin{pmatrix} \mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1} & 0 \\ 0 & \boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha} \end{pmatrix} \tag{4.14b}$$

with determinant

$$\Delta = \det \boldsymbol{D} = (\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1})\,(\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}), \tag{4.14c}$$

and the GLS estimators resulting from (4.8a) are

$$\widehat{a} = \frac{(\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha})\,(\mathbf{1}'\,\boldsymbol{\Omega})\,\boldsymbol{x}}{(\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1})\,(\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha})} = \frac{\mathbf{1}'\,\boldsymbol{\Omega}\,\boldsymbol{x}}{\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1}}, \tag{4.15a}$$

$$\widehat{b} = \frac{(\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1})\,(\boldsymbol{\alpha}'\,\boldsymbol{\Omega})\,\boldsymbol{x}}{(\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1})\,(\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha})} = \frac{\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{x}}{\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}} \tag{4.15b}$$

with variances

$$\text{Var}\big(\widehat{a}\big) = b^2\,\frac{\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}}{(\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1})\,(\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha})} = \frac{b^2}{\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1}}, \tag{4.15c}$$

$$\text{Var}\big(\widehat{b}\big) = b^2\,\frac{\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1}}{(\mathbf{1}'\,\boldsymbol{\Omega}\,\mathbf{1})\,(\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha})} = \frac{b^2}{\boldsymbol{\alpha}'\,\boldsymbol{\Omega}\,\boldsymbol{\alpha}}. \tag{4.15d}$$

We note that $\widehat{a}$ reduces to the sample mean $\overline{X}_n$ if

$$\mathbf{1}'\,\boldsymbol{\Omega} = \mathbf{1}',$$

or equivalently $\boldsymbol{B}\,\mathbf{1} = \mathbf{1}$, i.e. if all the rows (or columns) of the variance–covariance matrix add to unity, which holds for the normal distribution.

**Example 4/2: GLS estimating $(a, b)$ for a normal distribution (uncensored sample)**

The following $n = 6$ observations resulted from a Monte Carlo simulation of $X \sim NO(10, 50)$ :

$$\boldsymbol{x}' = (-57,\ -32,\ 5,\ 25,\ 46,\ 91).$$

$\boldsymbol{B}$ can be taken from Sect. 4.1 and the vector $\boldsymbol{\alpha}$ is

$$\boldsymbol{\alpha}' = (-1.2672,\ -0.6418,\ -0.2016,\ 0.2016,\ 0.6418,\ 1.2672).$$

We first have, only giving the upper triangle of the symmetric matrix:

$$\boldsymbol{\Omega} = \boldsymbol{B}^{-1} = \begin{pmatrix} 3.8401 & -2.8967 & 0.0332 & 0.0146 & 0.0065 & 0.0023 \\ & 9.6197 & -5.7823 & 0.0355 & 0.0173 & 0.0065 \\ & & 13.5676 & -6.8686 & 0.0355 & 0.0146 \\ & & & 13.5676 & -5.7823 & 0.0332 \\ & & & & 9.6197 & -2.8967 \\ & & & & & 3.8401 \end{pmatrix}.$$

The estimated parameters result as

$$\widehat{a} = \widehat{\mu} = 13.0000 \quad \text{and} \quad \widehat{b} = \widehat{\sigma} = 58.6656$$

with estimated variance-covariance matrix

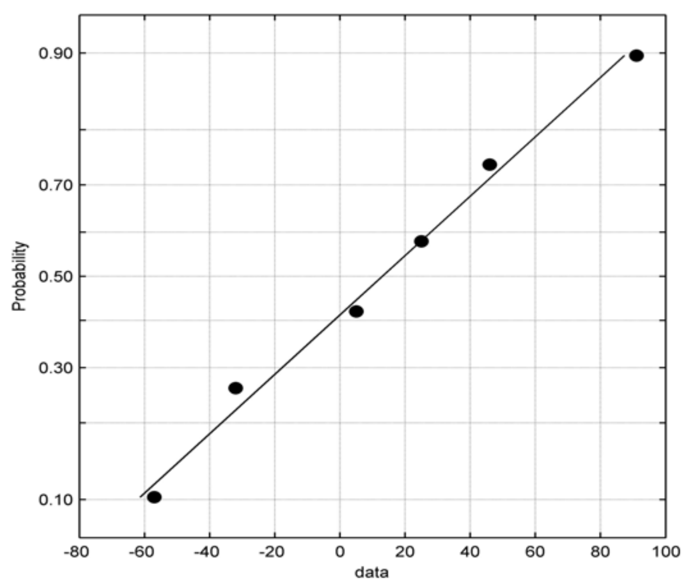
$$\widehat{\mathrm{Var}}\big(\widehat{a}, \widehat{b}\,\big) = \begin{pmatrix} 573.6084 & 0 \\ 0 & 363.7904 \end{pmatrix}.$$

The weighting vectors read

$$\begin{aligned} \boldsymbol{a'} &= (1/6,\ 1/6,\ 1/6,\ 1/6,\ 1/6,\ 1/6), \\ \boldsymbol{b'} &= (-0.3175,\ -0.1386,\ -0.0432,\ 0.0432,\ 0.1386,\ 0.3175). \end{aligned}$$

The following graph depicts the data together with the estimated regression line on normal probability paper.

Figure 4/3: Probability plot and estimated regression line (uncensored sample of size $n = 6$ from a normal distribution)



#### 4.2.1.3 GLS and censored samples

The discussion so far has concentrated on data consisting of the full set of order statistics. If the data to be used consist of a **fixed subset** of order statistics the general formulas (4.6) through (4.9) continue to hold. We have to cancel those $\alpha_{r:n}$ in the vector $\boldsymbol{\alpha}$ as regressors for which the ordered observations do not exist. Likewise, we have to skip the rows and columns of $\boldsymbol{B}$ belonging to these observations. The BLU property is preserved, but the efficiency of the estimators will decrease, see Example 4/3. Such a fixed subset of order statistics may arise in different ways.

1. We have a type–II censored sample. In this case the numbers of the ordered observations run from
   - $\ell+1$ through $n$ for censoring from below (= on the left) and the first $\ell$ ($\ell \geq 1$) observations are not known by their values,
   - 1 through $n-u$ for censoring from above (= on the right) and the last $u$ ($u \geq 1$) observations are not known by their values,
   - $\ell+1$ through $n-u$ for double censoring (= censoring on both sides) and $\ell+u$ observation are not known by their values.
2. The estimation procedure is based on selected order statistics. The selection may be done either arbitrarily or consciously and optimally.[8]

The formulas for the symmetric case in Sect. 4.2.1.2 hold whenever (4.13a) is satisfied. This occurs, for example, when we have a type–II censored sample from a symmetric population where censoring is symmetric.

**Example 4/3: Variance–covariance matrix of exponential parameter estimators under different modes of type–II censoring ($n = 10$)**

We want to demonstrate how different modes of type–II censoring affect the variance–covariance matrix of $\widehat{a}$ and $\widehat{b}$ for a sample of size $n = 10$ from an exponential distribution. We have chosen the unscaled variance–covariance matrix, i.e. we have dropped the factor $b^2$ in (4.6c). So, Tab. 4/1 displays $\text{Var}(\,\widehat{a}\,)/b^2$, $\text{Var}(\,\widehat{b}\,)/b^2$ and $\text{Cov}(\,\widehat{a},\widehat{b}\,)/b^2$. $\ell = 0$ ($u = 0$) means no censoring from below (from above). With single censoring we skip at most 50% of the observations, and with double censoring at most 60% ($\ell = u = 3$).

[8] We will not pursue this approach her. The earliest work in statistical inference using selected order statistics was done by MOSTELLER (1946). The approach was developed as a compromise between lack of efficiency and quickness and ease of computation. But it was found that in many situations the $n$–th observation $X_{n:n}$ must be used. Thus, the experimental time itself — the time until $X_{n:n}$ is available to be measured — will not necessarily be shortened. The reader who is interested in this topic is referred to the survey article by ALI/UMBACH (1998).

Table 4/1: Elements of the unscaled variance–covariance matrix $\left(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}\right)^{-1}$ for different modes of type–II censoring (exponential distribution, $n = 10$)

| censoring | | Elements of $\left(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}\right)^{-1}$ | | |
|---|---|---|---|---|
| $\ell$ | $u$ | Var$(\widehat{a})/b^2$ | Var$(\widehat{b})/b^2$ | Cov$(\widehat{a},\widehat{b})/b^2$ |
| 0 | 0 | 0.0111 | 0.1111 | −0.0111 |
| 0 | 1 | 0.0113 | 0.1250 | −0.0125 |
| 0 | 2 | 0.0114 | 0.1429 | −0.0143 |
| 0 | 3 | 0.0117 | 0.1667 | −0.0167 |
| 0 | 4 | 0.0120 | 0.2000 | −0.0200 |
| 0 | 5 | 0.0125 | 0.2500 | −0.0250 |
| 1 | 0 | 0.0279 | 0.1250 | −0.0264 |
| 2 | 0 | 0.0541 | 0.1429 | −0.0480 |
| 3 | 0 | 0.0966 | 0.1667 | −0.0798 |
| 4 | 0 | 0.1695 | 0.2000 | −0.1291 |
| 5 | 0 | 0.3049 | 0.2500 | −0.2114 |
| 1 | 1 | 0.0287 | 0.1429 | −0.0302 |
| 2 | 2 | 0.0606 | 0.2000 | −0.0672 |
| 3 | 3 | 0.1348 | 0.3333 | −0.1597 |

For the exponential distribution we notice that an increasing censoring level $(\ell + u)$ leads to higher variances of both parameter estimators. The effect of left and right censoring on Var$(\widehat{b})/b^2$ is the same, but left censoring has a much greater effect on Var$(\widehat{a})/b^2$ than right censoring. A censoring amount of $50\%$ leads to a variance of $\widehat{b}$ which is $225\%$ of the variance in the uncensored case $(\ell = u = 0)$.

If the data to be used do **not** consist of a **fixed subset** of order statistics the general formulas of Sections 4.2.1.1 and 4.2.1.2 do not hold anymore, i.e. the estimator are neither unbiased nor best. We do not have a fixed subset when type–I censoring is practised. With this censoring mode the numbers of censored items are random. Let $x_\ell$ be the left–hand censoring time, then the number of failing items before and up to $x_\ell$ is a binomially distributed variable $L$:

$$\Pr(L = \ell) = \binom{n}{\ell} P_\ell^{\ell}\,(1 - P_\ell)^{n-\ell};\ \ell = 0, 1, \ldots, n; \tag{4.16a}$$

where

$$P_\ell = \Pr(X \leq x_\ell). \tag{4.16b}$$

Likewise, the number $U$ of items failing beyond the right–hand censoring time $x_u$ is binomially distributed:

$$\Pr(U = u) = \binom{n}{u} P_u^{u}\,(1 - P_u)^{n-u};\ u = 0, 1, \ldots, n; \tag{4.17a}$$

where

$$P_u = \Pr(X > x_u). \tag{4.17b}$$

With double censoring the numbers $L$ and $U$ failing outside the interval $(x_\ell, x_u]$ have a multinomial distribution:

$$\Pr(L = \ell,\ U = u,\ R = r) = \frac{n!}{\ell!\,u!\,r!}\,P_\ell^\ell\,P_u^u\,P_r^r \tag{4.18}$$

with

$$\left.\begin{aligned} R &= n - L - U \\ r &= n - \ell - u \end{aligned}\right\} \quad - \quad \text{number of uncensored items,}$$

$$P_r = 1 - P_\ell - P_u \quad - \quad \text{probability of realizing an uncensored item,}$$

$$0 \le \ell, u, r \le n,\ \text{but}\ \ell + u + r = n.$$

## 4.2.2 Approximations to GLS

GLS requires the evaluation of the mean vector $\boldsymbol{\alpha}$ and the variance–covariance matrix $\boldsymbol{B}$ of the reduced order statistics. This will be a tedious task for most location–scale distributions. So, various approaches to reduce this burden have been suggested. In this section we will adhere to $\boldsymbol{\alpha}$, but look at possibilities to simplify $\boldsymbol{B}$. The resulting estimators will still be unbiased,[9] but their variances will increase, i.e. the efficiency will lessen.

### 4.2.2.1 $B$ approximated by a diagonal matrix

The evaluation of the covariances $\beta_{r,s:n},\ r > s$, of the reduced order statistics is cumbersome, either by evaluating the double integrals or by applying a recurrence formula or even by using the approximation in (2.25). The variances $\beta_{r,r:n}$ are much easier to evaluate and, therefore, a first approximation to the original variance–covariance matrix $\boldsymbol{B}$ is to do without the covariances, i.e. the covariances will be set equal to zero, and to use the diagonal matrix

$$\boldsymbol{B}_d = \begin{pmatrix} \beta_{1,1:n} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_{n,n:n} \end{pmatrix}. \tag{4.19a}$$

Its inverse is rather simple:

$$\boldsymbol{\Omega}_d = \boldsymbol{B}_d^{-1} = \begin{pmatrix} \beta_{1,1:n}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_{n,n:n}^{-1} \end{pmatrix}. \tag{4.19b}$$

[9] Unbiasedness is preserved as long as we apply the regressor matrix $\boldsymbol{A}$ of (4.5a).

Then, the approximate GLS estimators follow from (4.6b) by changing $\boldsymbol{\Omega}_d$ against $\boldsymbol{\Omega}$:

$$\widehat{\boldsymbol{\theta}}_d = \begin{pmatrix} \widehat{a}_d \\ \widehat{b}_d \end{pmatrix} = \big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{x}. \tag{4.20a}$$

The differences between these estimators and the proper GLS estimators of (4.6b) are given by, see (4.11c):

$$\widehat{\boldsymbol{\theta}}_d - \widehat{\boldsymbol{\theta}} = \boldsymbol{L}_d\,\boldsymbol{x} \tag{4.20b}$$

with

$$\boldsymbol{L}_d = \big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}_d - \big(\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A}\big)^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}. \tag{4.20c}$$

The amounts and the signs of the differences depend on the parent distribution.

The special and simple structure of $\boldsymbol{\Omega}_d$ allows an explicit formulation of the weights combining the data $\boldsymbol{x}$ into the estimators. Using the abbreviations

$$\beta_i^\circ := \beta_{i,i:n}^{-1}, \quad \alpha_i := \alpha_{i:n},$$

$\boldsymbol{D}$ of (4.7) turns into[10]

$$\boldsymbol{D}_d = \big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big) = \begin{pmatrix} \sum \beta_i^\circ & \sum \alpha_i\,\beta_i^\circ \\ \sum \alpha_i\,\beta_i^\circ & \sum \alpha_i^2\,\beta_i^\circ \end{pmatrix} \tag{4.21a}$$

with determinant

$$\Delta_d = \det \boldsymbol{D}_d = \sum \beta_i^\circ \sum \alpha_i^2\,\beta_i^\circ - \Big(\sum \alpha_i\,\beta_i^\circ\Big)^2. \tag{4.21b}$$

The skew–symmetric matrix $\boldsymbol{H}$ of (4.8d) now reads[11]

$$\begin{aligned} \boldsymbol{H}_d &= \frac{\boldsymbol{\Omega}_d\,\big(\boldsymbol{1}\,\boldsymbol{\alpha} - \boldsymbol{\alpha}\,\boldsymbol{1}'\big)\boldsymbol{\Omega}_d}{\Delta_d} \\ &= \frac{1}{\Delta_d}\begin{pmatrix} 0 & \beta_2^\circ\beta_1^\circ(\alpha_2-\alpha_1) & \beta_3^\circ\beta_1^\circ(\alpha_3-\alpha_1) & \cdots & \beta_n^\circ\beta_1^\circ(\alpha_n-\alpha_1) \\ \beta_1^\circ\beta_2^\circ(\alpha_1-\alpha_2) & 0 & \beta_3^\circ\beta_2^\circ(\alpha_3-\alpha_2) & \cdots & \beta_1^\circ\beta_2^\circ(\alpha_n-\alpha_2) \\ \beta_1^\circ\beta_3^\circ(\alpha_1-\alpha_3) & \beta_2^\circ\beta_3(\alpha_2-\alpha_3) & 0 & \cdots & \beta_n^\circ\beta_3^\circ(\alpha_n-\alpha_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_1^\circ\beta_n^\circ(\alpha_1-\alpha_n) & \beta_2^\circ\beta_n^\circ(\alpha_2-\alpha_n) & \beta_3^\circ\beta_n^\circ(\alpha_3-\alpha_n) & \cdots & 0 \end{pmatrix}. \end{aligned} \tag{4.21c}$$

[10] Summation is over those indices which belong to uncensored observations.

[11] Rows and columns belonging to censored observations have to be omitted.

The weighting vectors in (4.8e,f) follow as[12]

$$\boldsymbol{a}_d' = -\boldsymbol{\alpha}'\,\boldsymbol{H}_d$$
$$= -\frac{1}{\Delta_d}\Big(\beta_1^\circ \sum \alpha_i\beta_i^\circ(\alpha_1-\alpha_i),\ \beta_2^\circ \sum \alpha_i\beta_i^\circ(\alpha_2-\alpha_i), \ldots,\ \beta_n^\circ \sum \alpha_i\beta_i^\circ(\alpha_n-\alpha_i)\Big), \quad (4.21\text{d})$$
$$\boldsymbol{b}_d' = \boldsymbol{1}'\,\boldsymbol{H}_d$$
$$= \frac{1}{\Delta_d}\Big(\beta_1^\circ \sum \beta_i^\circ(\alpha_1-\alpha_i),\ \beta_2^\circ \sum \beta_i^\circ(\alpha_2-\alpha_i), \ldots,\ \beta_n^\circ \sum \beta_i^\circ(\alpha_n-\alpha_i)\Big). \quad (4.21\text{e})$$

We now turn to the variance–covariance matrix of $\widehat{\boldsymbol{\theta}}_d$. Starting with (4.6c) and substituting $\boldsymbol{\Omega}$ by $\boldsymbol{\Omega}_d$ the resulting variance–covariance matrix, called **approximate variance–covariance matrix**,

$$\text{Var}_d\big(\widehat{\boldsymbol{\theta}}_d\big) = b^2\big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1} \quad (4.22)$$

will not be the true and correct variance–covariance matrix but nothing but a crude approximation, based on the $\boldsymbol{B}$–substitute. The **true variance–covariance matrix** of $\widehat{\boldsymbol{\theta}}_d$ can be found on different ways.

1. From (4.11e) we have

   $$\text{Var}\big(\widehat{\boldsymbol{\theta}}_d\big) = \text{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) + b^2\,\boldsymbol{L}_d\,\boldsymbol{B}\,\boldsymbol{L}_d', \quad (4.23)$$

   with $\boldsymbol{L}_d$ given in (4.20c).

2. Starting from (4.20a) and replacing $\boldsymbol{x}$ by $\boldsymbol{A}'\,\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ we first have

   $$\begin{aligned}\widehat{\boldsymbol{\theta}}_d &= \big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,(\boldsymbol{A}\,\boldsymbol{\theta} + \boldsymbol{\varepsilon})\\ &= \underbrace{\big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}}_{=\boldsymbol{I}}\,\boldsymbol{\theta} + \big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{\varepsilon}\\ &= \boldsymbol{\theta} + \big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{\varepsilon}. \end{aligned} \quad (4.24\text{a})$$

   The variance–covariance matrix of $\widehat{\boldsymbol{\theta}}_d$ is

   $$\text{Var}\big(\widehat{\boldsymbol{\theta}}_d\big) = \text{E}\big[\big(\widehat{\boldsymbol{\theta}}_d - \boldsymbol{\theta}\big)\big(\widehat{\boldsymbol{\theta}}_d - \boldsymbol{\theta}\big)'\,\big]. \quad (4.24\text{b})$$

   Inserting (4.24a) into (4.24b) gives

   $$\begin{aligned}\text{Var}\big(\widehat{\boldsymbol{\theta}}_d\big) &= \text{E}\big[\big\{\big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{\varepsilon}\big\}\,\big\{\boldsymbol{\varepsilon}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\,\big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{\varepsilon}\big)^{-1}\big\}\big]\\ &= \big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\underbrace{\text{E}(\boldsymbol{\varepsilon}'\,\boldsymbol{\varepsilon})}_{=b^2\,\boldsymbol{B}}\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\,(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A})^{-1}\\ &= b^2\,\big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1}\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{B}\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\,(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A})^{-1}. \end{aligned} \quad (4.24\text{c})$$

[12] Weights belonging to censored observations have to be omitted.

3. The elements of $\operatorname{Var}(\boldsymbol{\theta}_d)$ can be given explicitly by using the weights $a_{d,i}$ and $b_{d,i}$ in the vectors $\boldsymbol{a}'_d$ and $\boldsymbol{b}'_d$ of (4.21d,e):[13]

$$\begin{aligned}
\operatorname{Var}\big(\widehat{a}_d\big) &= b^2 \sum_i \sum_j a_{d,i}\, a_{d,j}\, \beta_{i,j:n}, && \text{(4.25a)}\\
\operatorname{Var}\big(\widehat{b}_d\big) &= b^2 \sum_i \sum_j b_{d,i}\, b_{d,j}\, \beta_{i,j:n}, && \text{(4.25b)}\\
\operatorname{Cov}\big(\widehat{a}_d, \widehat{b}_d\big) &= b^2 \sum_i \sum_j a_{d,i}\, b_{d,j}\, \beta_{i,j:n}. && \text{(4.25c)}
\end{aligned}$$

The difference matrix between the true variance–covariance matrix $\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}_d\big)$ and its approximation (4.22) is

$$\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}_d\big) - \operatorname{Var}_d\big(\widehat{\boldsymbol{\theta}}_d\big) = b^2 \left(\boldsymbol{A}'\,\boldsymbol{\Omega}_d \boldsymbol{A}\right)^{-1} \left\{\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{B}\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big(\boldsymbol{A}'\,\boldsymbol{\Omega}_d\,\boldsymbol{A}\big)^{-1} - \boldsymbol{I}\right\} \qquad (4.26)$$

and depends on the parent distribution via $\boldsymbol{\alpha}$ and $\boldsymbol{B}$, thus no general statement can be made. We see in the following example that we will commit severe errors in applying $\operatorname{Var}_d\big(\widehat{\boldsymbol{\theta}}_d\big)$ to estimate the variance of the simplified estimators.

**Example 4/4: Comparisons of GLS and GLS with diagonal variance–covariance matrix (exponential, uniform and normal distributions, $n = 10$)**

We first look at the **exponential distribution** and take the data $\boldsymbol{x}$ of Example 4/1. The estimates result as

$$\widehat{a}_d = 9.7580, \;\; \widehat{b}_d = 100.9885,$$

whereas the proper GLS estimates read

$$\widehat{a} = 15.0889, \;\; \widehat{b} = 99.1111.$$

The approximate variance–covariance matrix (4.22) is

$$\operatorname{Var}_d\big(\widehat{\boldsymbol{\theta}}_d\big) = b^2 \begin{pmatrix} 0.0082 & -0.0119 \\ -0.0119 & 0.0399 \end{pmatrix},$$

whereas the correct variance–covariance matrix, using (4.23), reads

$$\begin{aligned}
\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}_d\big) &= b^2 \begin{pmatrix} 0.0111 & -0.0111 \\ -0.0111 & 0.1111 \end{pmatrix} + b^2 \begin{pmatrix} 0.0030 & -0.0009 \\ -0.0009 & 0.0047 \end{pmatrix}\\
&= b^2 \begin{pmatrix} 0.0141 & -0.0120 \\ -0.0120 & 0.1158 \end{pmatrix}.
\end{aligned}$$

[13] Summation is over the indices of uncensored observations.

The relative efficiencies are

$$\frac{\mathrm{Var}\big(\widehat{a}\,\big)}{\mathrm{Var}\big(\widehat{a}_d\big)} = \frac{0.0111}{0.0141} = 0.7896, \quad \frac{\mathrm{Var}\big(\widehat{b}\,\big)}{\mathrm{Var}\big(\widehat{b}_d\big)} = \frac{0.1111}{0.1158} = 0.9596.$$

The loss of efficiency for $b$ is small ($\approx 4\%$) compared with the loss for $a$ ($\approx 21\%$).

For the **uniform distribution** we find

- the GLS variance–covariance matrix

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) = b^2 \begin{pmatrix} 0.0084 & -0.0093 \\ -0.0093 & 0.0185 \end{pmatrix},$$

- the approximate variance–covariance matrix according to (4.22)

$$\mathrm{Var}_d\big(\widehat{\boldsymbol{\theta}}_d\big) = b^2 \begin{pmatrix} 0.0047 & -0.0068 \\ -0.0068 & 0.0136 \end{pmatrix},$$

- the correct variance–covariance matrix according to (4.23)

$$\begin{aligned} \mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_d\big) &= b^2 \begin{pmatrix} 0.0084 & -0.0093 \\ -0.0093 & 0.0185 \end{pmatrix} + b^2 \begin{pmatrix} 0.0035 & -0.0020 \\ -0.0020 & 0.0041 \end{pmatrix} \\ &= b^2 \begin{pmatrix} 0.0119 & -0.0113 \\ -0.0113 & 0.0226 \end{pmatrix}, \end{aligned}$$

- the relative efficiencies

$$\frac{\mathrm{Var}\big(\widehat{a}\,\big)}{\mathrm{Var}\big(\widehat{a}_d\big)} = \frac{0.0084}{0.0119} = 0.7081, \quad \frac{\mathrm{Var}\big(\widehat{b}\,\big)}{\mathrm{Var}\big(\widehat{b}_d\big)} = \frac{0.0185}{0.0226} = 0.8188.$$

The loss of efficiency is of moderate size.

For a sample of size $n = 10$ from a **normal distribution** using the tabulated values[14]

$$\alpha_{6:10} = 0.122668,\ \alpha_{7:10} = 0.375265,\ \alpha_{8:10} = 0.656059,\ \alpha_{9:10} = 1.001357,\ \alpha_{10:10} = 1.538753,$$

[14] Missing values may be found by the symmetry relations:

$$\begin{aligned} \alpha_{r:n} &= -\alpha_{n-r:n} \\ \beta_{r,s:n} &= \beta_{n-s+1,n-r+1:n}. \end{aligned}$$

Source for $\alpha_{r:10}$ : TEICHROEW (1956)
Source for $\beta_{r,s:10}$: SARHAN/GREENBERG (1956).

$$
\begin{array}{lll}
\beta_{1,1:10} = 0.344344, & \beta_{2,2:10} = 0.214524, & \beta_{3,5:10} = 0.107745, \\
\beta_{1,2:10} = 0.171263, & \beta_{2,3:10} = 0.146623, & \beta_{3,6:10} = 0.089225, \\
\beta_{1,3:10} = 0.116259, & \beta_{2,4:10} = 0.111702, & \beta_{3,7:10} = 0.074918, \\
\beta_{1,4:10} = 0.088249, & \beta_{2,5:10} = 0.089743, & \beta_{3,8:10} = 0.063033, \\
\beta_{1,5:10} = 0.070741, & \beta_{2,6:10} = 0.074200, & \beta_{4,4:10} = 0.157939, \\
\beta_{1,6:10} = 0.058399, & \beta_{2,7:10} = 0.062228, & \beta_{4,5:10} = 0.127509, \\
\beta_{1,7:10} = 0.048921, & \beta_{2,8:10} = 0.052307, & \beta_{4,6:10} = 0.105786, \\
\beta_{1,8:10} = 0.041084, & \beta_{2,9:10} = 0.043371, & \beta_{4,7:10} = 0.088946, \\
\beta_{1,9:10} = 0.034041, & \beta_{3,3:10} = 0.175003, & \beta_{5,5:10} = 0.151054, \\
\beta_{1,10:10} = 0.026699, & \beta_{3,4:10} = 0.133802, & \beta_{5,6:10} = 0.125599.
\end{array}
$$

We find

- the GLS variance–covariance matrix

$$
\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\big) = b^2 \begin{pmatrix} 0.1000 & 0.0000 \\ 0.0000 & 0.0576 \end{pmatrix},
$$

- the approximate variance–covariance matrix according to (4.22)

$$
\mathrm{Var}_d\big(\widehat{\boldsymbol{\theta}}_d\big) = b^2 \begin{pmatrix} 0.0191 & 0.0000 \\ 0.0000 & 0.0333 \end{pmatrix},
$$

- the correct variance–covariance matrix according to (4.23)

$$
\begin{aligned}
\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_d\big) &= b^2 \begin{pmatrix} 0.1000 & 0.0000 \\ 0.0000 & 0.0576 \end{pmatrix} + b^2 \begin{pmatrix} 0.0016 & 0.0000 \\ 0.0000 & 0.0023 \end{pmatrix} \\
&= b^2 \begin{pmatrix} 0.1016 & 0.0000 \\ 0.0000 & 0.0599 \end{pmatrix},
\end{aligned}
$$

- the relative efficiencies

$$
\frac{\mathrm{Var}\big(\widehat{a}\big)}{\mathrm{Var}\big(\widehat{a}_d\big)} = \frac{0.1000}{0.1016} = 0.9840, \quad \frac{\mathrm{Var}\big(\widehat{b}\big)}{\mathrm{Var}\big(\widehat{b}_d\big)} = \frac{0.0576}{0.0599} = 0.9609.
$$

Both losses of efficiency are rather small.

#### 4.2.2.2 $B$ approximated by an identity matrix

A more radical approximation than that of the preceding section consists of substituting $\boldsymbol{B}$ by an identity or unity matrix $\boldsymbol{I}$ of proper dimension, so we might expect a greater loss of efficiency:

$$\boldsymbol{B} \approx \boldsymbol{I}. \tag{4.27a}$$

$\boldsymbol{I}$ is of order $n$ for an uncensored sample and of order $n-\ell-u$ for a censored sample. This approach goes back to GUPTA (1952) and is known as **GUPTA's simplified linear estimator**. Instead of the GLS method we now use the **OLS method** and the estimators and their approximate variance–covariance matrix are given by

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}} = \begin{pmatrix}\widehat{a}_{\boldsymbol{I}}\\ \widehat{b}_{\boldsymbol{I}}\end{pmatrix} = \left(\boldsymbol{A}'\,\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\,\boldsymbol{x} \tag{4.27b}$$

$$\mathrm{Var}_{\boldsymbol{I}}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2\left(\boldsymbol{A}'\,\boldsymbol{A}\right)^{-1}. \tag{4.27c}$$

The formulas of the preceding section hold here with $\boldsymbol{\Omega}_d$ replaced by $\boldsymbol{I}$ and $\beta_i^\circ$ by $1$. Because $\boldsymbol{I}$ and $1$ are the neutral elements of multiplication we can simply omit $\boldsymbol{\Omega}_d$ and $\beta_i^\circ$ in the formulas of Sect. 4.2.2.1 to arrive at the formulas valid for the OLS method. The elements of $\boldsymbol{D}_{\boldsymbol{I}} = \boldsymbol{A}'\,\boldsymbol{A}$ are:[15]

$$\boldsymbol{1}'\,\boldsymbol{I}\,\boldsymbol{1} = n^* \quad \text{with} \quad \left\{\begin{array}{l} n^* = n \ \text{ for an uncensored sample,}\\ n^* = n-\ell-u \ \text{ for a censored sample,}\end{array}\right\} \tag{4.28a}$$

$$\boldsymbol{1}'\,\boldsymbol{I}\,\boldsymbol{\alpha} = \boldsymbol{\alpha}'\,\boldsymbol{I}\,\boldsymbol{1} = \sum \alpha_i \tag{4.28b}$$

$$\boldsymbol{\alpha}'\,\boldsymbol{I}\boldsymbol{\alpha} = \sum \alpha_i^2. \tag{4.28c}$$

The determinant of $\boldsymbol{D}_{\boldsymbol{I}}$ is[15]

$$\Delta_{\boldsymbol{I}} = \det \boldsymbol{D}_{\boldsymbol{I}} = n^* \sum \alpha_i^2 - \big(\sum \alpha_i\big)^2 \tag{4.29a}$$

and the skew–symmetric matrix $\boldsymbol{H}$ in (4.8d) turns into

$$\boldsymbol{H}_{\boldsymbol{I}} = \frac{\boldsymbol{1}\,\boldsymbol{\alpha}' - \boldsymbol{\alpha}\,\boldsymbol{1}'}{\Delta_{\boldsymbol{I}}} = \frac{1}{\Delta_{\boldsymbol{I}}}\begin{pmatrix} 0 & \alpha_2-\alpha_1 & \alpha_3-\alpha_1 & \cdots & \alpha_n-\alpha_1\\ \alpha_1-\alpha_2 & 0 & \alpha_3-\alpha_2 & \cdots & \alpha_n-\alpha_2\\ \alpha_1-\alpha_3 & \alpha_2-\alpha_3 & 0 & \cdots & \alpha_n-\alpha_3\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \alpha_1-\alpha_n & \alpha_2-\alpha_n & \alpha_3-\alpha_n & \cdots & 0\end{pmatrix}. \tag{4.29b}$$

[15] We only have to sum those $\alpha_i$ belonging to uncensored observations.

The weighting vectors in (4.8e,f) follow as[16]

$$\boldsymbol{a}_{\boldsymbol{I}}' = -\boldsymbol{\alpha}'\,\boldsymbol{H}_{\boldsymbol{I}} = -\frac{1}{\Delta_{\boldsymbol{I}}}\Big(\sum \alpha_i(\alpha_1-\alpha_i), \sum \alpha_i(\alpha_2-\alpha_i), \ldots, \sum \alpha_i(\alpha_n-\alpha_i)\Big), \quad (4.29c)$$

$$\boldsymbol{b}_{\boldsymbol{I}}' = \boldsymbol{1}'\,\boldsymbol{H}_{\boldsymbol{I}} = \frac{1}{\Delta_{\boldsymbol{I}}}\Big(\sum(\alpha_1-\alpha_i),\ \sum(\alpha_2-\alpha_i), \ldots, \sum(\alpha_n-\alpha_i)\Big). \quad (4.29d)$$

The difference between the GLS estimators (4.6b) and the simplified estimators (4.27b) is

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}} - \widehat{\boldsymbol{\theta}} = \boldsymbol{L}_{\boldsymbol{I}}\,\boldsymbol{x} \quad (4.30a)$$

with

$$\boldsymbol{L}_{\boldsymbol{I}} = (\boldsymbol{A}'\,\boldsymbol{A})^{-1}\,\boldsymbol{A}' - (\boldsymbol{A}'\,\boldsymbol{\Omega}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{\Omega}. \quad (4.30b)$$

The correct variance–covariance matrix of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}$ is

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = \mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\big) + b^2\,\boldsymbol{L}_{\boldsymbol{I}}\,\boldsymbol{B}\,\boldsymbol{L}_{\boldsymbol{I}}' \quad (4.31a)$$

or from (4.24c) with $\boldsymbol{\Omega}_d$ substituted by $\boldsymbol{I}$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2(\boldsymbol{A}'\,\boldsymbol{A})^{-1}\,\boldsymbol{A}'\,\boldsymbol{B}\,\boldsymbol{A}\,(\boldsymbol{A}'\,\boldsymbol{A})^{-1}, \quad (4.31b)$$

and its elements are explicitly given by[16]

$$\mathrm{Var}(\widehat{a}_{\boldsymbol{I}}) = b^2\sum_i\sum_j a_{\boldsymbol{I},i}\,a_{\boldsymbol{I},j}\,\beta_{i,j:n}, \quad (4.32a)$$

$$\mathrm{Var}(\widehat{b}_{\boldsymbol{I}}) = b^2\sum_i\sum_j b_{\boldsymbol{I},i}\,b_{\boldsymbol{I},j}\,\beta_{i,j:n}, \quad (4.32b)$$

$$\mathrm{Cov}(\widehat{a}_{\boldsymbol{I}},\widehat{b}_{\boldsymbol{I}}) = b^2\sum_i\sum_j a_{\boldsymbol{I},i}\,b_{\boldsymbol{I},j}\,\beta_{i,j:n}, \quad (4.32c)$$

the weights $a_{\boldsymbol{I},i},\ b_{\boldsymbol{I},i}$ coming from (4.29c,d). The difference between the correct variance–covariance matrix (4.31a) and the approximate variance–covariance matrix (4.27c) is

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) - \mathrm{Var}_{\boldsymbol{I}}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2(\boldsymbol{A}'\,\boldsymbol{A})^{-1}\big\{\boldsymbol{A}'\,\boldsymbol{B}\,\boldsymbol{A}\,(\boldsymbol{A}'\,\boldsymbol{A})^{-1} - \boldsymbol{I}\big\}. \quad (4.33)$$

The difference depends on the parent distribution and no general statement is possible.

**Example 4/5: Comparison of GLS and OLS estimators for exponential, normal and uniform distributions** ($\boldsymbol{n = 10}$)

We first look at the **exponential distribution** and — with the data $\boldsymbol{x}$ of Example 4/1 — find the following OLS estimates

$$\widehat{a}_{\boldsymbol{I}} = 1.6876,\ \ \widehat{b}_{\boldsymbol{I}} = 112.5124,$$

whereas the proper GLS estimates read

$$\widehat{a} = 15.0889,\ \ \widehat{b} = 99.1111.$$

We further have

[16] Summation is over the indices of uncensored observations.

- the approximate variance–covariance matrix of (4.27c)

$$\operatorname{Var}_{\boldsymbol{I}}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2 \begin{pmatrix} 0.2414 & -0.1414 \\ -0.01414 & 0.1414 \end{pmatrix},$$

- the correct variance–covariance matrix (4.31a)

$$\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2 \begin{pmatrix} 0.0698 & -0.0698 \\ -0.0698 & 0.1698 \end{pmatrix},$$

- the GLS variance–covariance matrix

$$\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}\big) = b^2 \begin{pmatrix} 0.0111 & -0.0111 \\ -0.0111 & 0.1111 \end{pmatrix},$$

- the relative efficiencies

$$\frac{\operatorname{Var}(\widehat{a})}{\operatorname{Var}(\widehat{a}_{\boldsymbol{I}})} = \frac{0.0111}{0.0698} = 0.1591, \quad \frac{\operatorname{Var}(\widehat{b})}{\operatorname{Var}(\widehat{b}_{\boldsymbol{I}})} = \frac{0.1111}{0.1698} = 0.6542.$$

  We observe heavy losses of efficiency.

For the **uniform distribution** we find

- the GLS variance–covariance matrix

$$\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}\big) = b^2 \begin{pmatrix} 0.0084 & -0.0093 \\ -0.0093 & 0.0185 \end{pmatrix},$$

- the approximate variance–covariance matrix according to (4.27c)

$$\operatorname{Var}_{\boldsymbol{I}}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2 \begin{pmatrix} 0.4667 & -0.7333 \\ -0.7333 & 1.4667 \end{pmatrix},$$

- the correct variance–covariance matrix according to (4.31a)

$$\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2 \begin{pmatrix} 0.0156 & -0.0144 \\ -0.0144 & 0.0289 \end{pmatrix},$$

- the relative efficiencies

$$\frac{\operatorname{Var}(\widehat{a})}{\operatorname{Var}(\widehat{a}_{\boldsymbol{I}})} = \frac{0.0084}{0.0156} = 0.5411, \quad \frac{\operatorname{Var}(\widehat{b})}{\operatorname{Var}(\widehat{b}_{\boldsymbol{I}})} = \frac{0.0185}{0.0289} = 0.6410.$$

The efficiencies decrease considerably.

For the **normal distribution** we have

- the GLS variance–covariance matrix

$$\text{Var}\big(\widehat{\boldsymbol{\theta}}\big) = b^2 \begin{pmatrix} 0.1000 & 0.0000 \\ 0.0000 & 0.0576 \end{pmatrix},$$

- the approximate variance–covariance matrix according to (4.27c)

$$\text{Var}_{\boldsymbol{I}}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2 \begin{pmatrix} 0.1000 & 0.0000 \\ 0.0000 & 0.1264 \end{pmatrix},$$

- the correct variance–covariance matrix according to (4.31a)

$$\text{Var}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}\big) = b^2 \begin{pmatrix} 0.100008 & 0.0000 \\ 0.0000 & 0.057648 \end{pmatrix},$$

- the relative efficiencies

$$\frac{\text{Var}\big(\widehat{a}\big)}{\text{Var}\big(\widehat{a}_{\boldsymbol{I}}\big)} = \frac{0.1000}{0.100008} \approx 1, \quad \frac{\text{Var}\big(\widehat{b}\big)}{\text{Var}\big(\widehat{b}_{\boldsymbol{I}}\big)} = \frac{0.0576}{0.057648} = 0.9989.$$

  The efficiencies are approximately the same for GLS and OLS. This result holds for other sample sizes, too.

#### 4.2.2.3 BLOM's estimator

Another approach of simplifying the variance–covariance matrix $\boldsymbol{B}$ of the reduced order statistics $Y_{r:n}$ consists in replacing $\boldsymbol{B}$ by the asymptotic variance–covariance matrix, as has been proposed by BLOM (1958, 1962). The resulting estimator is called **unbiased nearly best linear**. If exact unbiasedness is given up, one may also approximate asymptotically to the expectations $\alpha_{r:n}$ and obtain **nearly unbiased, nearly best linear estimator**, see Sect. 4.2.3.

The elements of $\boldsymbol{B}$ are replaced by the first term in the series approximations (2.24) and (2.25). Let

$$y_{p_r} = F^{-1}(p_r) \tag{4.34a}$$

denote the reduced percentile of order

$$p_r = \frac{r}{n+1}\,. \tag{4.34b}$$

Then, the first derivative of this percentile

$$y'_{p_r} = \frac{\mathrm{d}y_{p_r}}{\mathrm{d}p_r} = F^{-1(1)}(p_r), \tag{4.34c}$$

which is part of the first term in (2.24) and (2.25), can be substituted by the reciprocal of $f(y_{p_r})$, because

$$y'_{p_r} = \frac{1}{\mathrm{d}p_r/\mathrm{d}y_{p_r}} = \frac{1}{f(y_{p_r})}, \tag{4.34d}$$

$f(y_{p_r})$ being the reduced density evaluated at $y_{p_r}$. Thus, for large $n$ we have the following elements of the upper triangle of the **asymptotic variance–covariance matrix $\boldsymbol{B}_B$**:

$$\beta_{rs} = \frac{p_r\, q_s}{(n+2)\, f(y_{p_r})\, f(y_{p_s})}, \quad r \le s, \tag{4.35}$$

where $q_s = 1 - p_s$. $\boldsymbol{B}_B$ is a symmetric matrix. **BLOM's estimator** follows as

$$\widehat{\boldsymbol{\theta}}_B = \left(\boldsymbol{A}'\,\boldsymbol{B}_B^{-1}\,\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\,\boldsymbol{B}_B^{-1}\,\boldsymbol{x} \tag{4.36a}$$

with the approximate variance–covariance matrix

$$\mathrm{Var}_B\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2\left(\boldsymbol{A}'\,\boldsymbol{B}_B^{-1}\,\boldsymbol{A}\right)^{-1}. \tag{4.36b}$$

The true and correct variance–covariance matrix of BLOM's estimator is

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_B\big) = \mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) + \boldsymbol{L}_B\,\boldsymbol{B}\,\boldsymbol{L}'_B \tag{4.36c}$$

where

$$\boldsymbol{L}_B = \left(\boldsymbol{A}'\,\boldsymbol{B}_B^{-1}\,\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\,\boldsymbol{B}_B^{-1} - \left(\boldsymbol{A}'\,\boldsymbol{B}^{-1}\,\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\,\boldsymbol{B}^{-1}. \tag{4.36d}$$

We will see in Example 4/6 that with $n \to \infty$ $\mathrm{Var}_B\big(\widehat{\boldsymbol{\theta}}_B\big)$ approaches $\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_B\big)$ which in turn approaches the GLS variance–covariance matrix $\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\,\big)$. For a censored sample the rows and columns of $\boldsymbol{B}_B$ and the elements of $\boldsymbol{\alpha}$ in $\boldsymbol{A}$ corresponding to the censored observations have to be omitted.

The matrix $\boldsymbol{B}_B$ is a $(n \times n)$ non–singular symmetric matrix with a special pattern. Its elements are of the form

$$\beta_{rs} = \frac{1}{n+2}\, c_r\, d_s, \quad r \le s, \tag{4.37a}$$

where

$$c_r = \frac{p_r}{f(y_{p_r})}, \quad d_s = \frac{q_s}{f(y_{p_s})}\,. \tag{4.37b}$$

The inverse of $\boldsymbol{B}_B$ is a **banded matrix** with the $(r,s)$–th element $(r \le s)$ given by

$$\beta^{rs} = \begin{pmatrix} c_2\,\{c_1\,(c_2\,d_1 - c_1\,d_2)\}^{-1} & \text{for } r = s = 1, \\ \dfrac{c_{r+1}\,d_{r-1} - c_{r-1}\,d_{r+1}}{(c_r\,d_{r-1} - c_{r-1}\,d_r)\,(c_{r+1}\,d_r - c_r\,d_{r+1})} & \text{for } r = s = 2 \text{ to } n-1, \\ d_{n-1}\,\{d_n\,(c_n\,(c_n\,d_{n-1} - c_{n-1}\,d_n)\}^{-1} & \text{for } r = s = n \\ -(c_{r+1}\,d_r - c_r\,d_{r+1})^{-1} & \text{for } s = r+1 \text{ and } r = 1 \text{ to } n-1, \\ 0 & \text{for } s > r+1. \end{pmatrix} \tag{4.37c}$$

BLOM (1958, 1962) has exploited this feature of $\boldsymbol{B}_B$ to obtain solutions for $\widehat{\boldsymbol{\theta}}_B$ and $\text{Var}_B\big(\widehat{\boldsymbol{\theta}}_B\big)$, equivalent to (4.36a,b), that avoid the inversion of the $(n \times n)$ matrix $\boldsymbol{B}_B$ and only requires the inversion of a $(2 \times 2)$ matrix. Nowadays, inversion of large–order matrices is no problem. For those readers interested in the BLOM procedure we give the results without any proofs.[17]

**Excursus: BLOM's unbiased nearly best linear estimator in explicit form**

We introduce the following variables:

$$f_r := f(y_{p_r});\ r = 1,\dots,n \text{ and } f_0 = f_{n+1} = 0; \tag{4.38a}$$

$$\alpha_r := \alpha_{r:n};\ r = 1,\dots,n \text{ and } \alpha_0 = \alpha_{n+1} = 0; \tag{4.38b}$$

$$C_{1r} := f_r - f_{r-1};\ r = 0,\dots,n; \tag{4.38c}$$

$$C_{2r} := \alpha_r\,f_r - \alpha_{r+1}\,f_{r+1};\ r = 0,\dots,n; \tag{4.38d}$$

$$\boldsymbol{D} := (d_{ij});\ i,j = 1,2; \tag{4.38e}$$

$$d_{ij} := \sum_{\ell=0}^{n} C_{i\ell}\,C_{j\ell};\ i,j = 1,2; \tag{4.38f}$$

$$\boldsymbol{D}^{-1} := (d^{ij});\ i,j = 1,2; \tag{4.38g}$$

$$w_{ir} := f_r\big[d^{i1}\,(C_{1r} - C_{1,r-1}) + d^{i2}\,(C_{2r} - C_{2,r-1})\big];\ i = 1,2;\ r = 1,\dots,n; \tag{4.38h}$$

$$\widehat{a}_B := \sum_{r=1}^{n} w_{1r}\,X_{r:n}; \tag{4.38i}$$

[17] Details of this procedure and the derivation of the formulas can be found in DAVID (1981, p. 133–135) or BALAKRISHNAN/COHEN (1991, p. 100–104).

$$\widehat{b}_B \;:=\; \sum_{r=1}^{n} w_{2r}\, X_{r:n}; \tag{4.38j}$$

$$\mathrm{Var}_B\big(\widehat{\boldsymbol{\theta}}_B\big) \;:=\; \frac{b^2}{(n+1)\,(n+2)} \begin{pmatrix} d^{11} & d^{12} \\ d^{12} & d^{22} \end{pmatrix}. \tag{4.38k}$$

For censored samples the variables $C_{1r}$ and $C_{2r}$ have special values for the indices belonging to censored observations, $\ell$ being the left–hand censoring number and $n-u$ being the right–hand censoring number:

$$C_{1r} \;=\; \left\{ \begin{array}{ll} -\dfrac{1}{\ell+1}\, f_{\ell+1}\,, & 0 \le r \le \ell, \\ f_r - f_{r+1}\,, & \ell+1 \le r \le n-u-1, \\ \dfrac{1}{u+1}\, f_{n-u}\,, & n-u \le r \le n; \end{array} \right\} \tag{4.38l}$$

$$C_{2r} \;=\; \left\{ \begin{array}{ll} -\dfrac{1}{\ell+1}\, \alpha_{\ell+1}\, f_{\ell+1}, & 0 \le r \le \ell, \\ \alpha_r\, f_r - \alpha_{r+1}\, f_{r+1}, & \ell+1 \le r \le n-u-1, \\ \dfrac{1}{u+1}\, \alpha_{n-u}\, f_{n-u}, & n-u \le r \le n. \end{array} \right\} \tag{4.38m}$$

**Example 4/6: BLOM's unbiased nearly best linear estimator for the exponential distribution $(n = 10,\ 20,\ 50,\ 100)$**

For a sample of size $\underline{n=10}$ we use the data $\boldsymbol{x}$ of Example 4/1 and find the BLOM's estimates

$$\widehat{a}_B = 14.8808, \quad \widehat{b}_B = 101.1293,$$

whereas the GLS estimates are

$$\widehat{a} = 15.0889, \quad \widehat{b} = 99.1111.$$

The variance matrices are:

$$\mathrm{Var}_B\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2 \begin{pmatrix} 0.009108 & -0.007749 \\ -0.007749 & 0.077493 \end{pmatrix},$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2 \begin{pmatrix} 0.011158 & -0.011576 \\ -0.011576 & 0.115760 \end{pmatrix},$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) = b^2 \begin{pmatrix} 0.0\overline{1} & -0.0\overline{1} \\ -0.0\overline{1} & 0.\overline{1} \end{pmatrix}.$$

For samples of sizes $n = 20,\ 50,\ 100$ we only give the variance matrices to demonstrate the process of convergence.

$\underline{n = 20}$

$$\mathrm{Var}_B\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2 \begin{pmatrix} 0.002379 & -0.002117 \\ -0.002117 & 0.042334 \end{pmatrix}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2 \begin{pmatrix} 0.002636 & -0.002723 \\ -0.002723 & 0.054463 \end{pmatrix}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) = b^2 \begin{pmatrix} 0.002632 & -0.002632 \\ -0.002632 & 0.052632 \end{pmatrix}$$

$\underline{n = 50}$

$$\mathrm{Var}_B\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2 \begin{pmatrix} 0.000392 & -0.000367 \\ -0.000367 & 0.018339 \end{pmatrix}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2 \begin{pmatrix} 0.000408 & -0.000417 \\ -0.000417 & 0.020408 \end{pmatrix}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) = b^2 \begin{pmatrix} 0.000408 & -0.000408 \\ -0.000408 & 0.020408 \end{pmatrix}$$

$\underline{n = 100}$

$$\mathrm{Var}_B\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2 \begin{pmatrix} 0.000099 & -0.000095 \\ -0.000095 & 0.009505 \end{pmatrix}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_B\big) = b^2 \begin{pmatrix} 0.000101 & -0.000102 \\ -0.000102 & 0.010227 \end{pmatrix}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\,\big) = b^2 \begin{pmatrix} 0.000\overline{10} & -0.000\overline{10} \\ -0.000\overline{10} & 0.\overline{01} \end{pmatrix}$$

We observe that

- the estimated variances $\mathrm{Var}_B\big(\widehat{a}_B\big)$ and $\mathrm{Var}_B\big(\widehat{b}_B\big)$ of BLOM's estimators are smaller than the GLS variances $\mathrm{Var}\big(\widehat{a}\,\big)$ and $\mathrm{Var}\big(\widehat{b}\,\big)$, but the differences diminish as $n$ gets larger,
- the true variances $\mathrm{Var}\big(\widehat{a}_B\big)$ and $\mathrm{Var}\big(\widehat{b}_B\big)$ are very similar to the GLS variances and for $n \geq 50$ are nearly equal to them.

#### 4.2.2.4 DOWNTON's estimator

The following approach, which is due to DOWNTON (1966a,b), is only applicable when the sample is not censored. For the latter reason we only give a short introduction and present the main idea of this method.[18] DOWNTON and several other authors had observed that — especially for the Log–WEIBULL distribution — the efficiencies of linear estimators for the location–scale parameter are not particularly sensitive to changes in the values of the weights combining the ordered observations. Thus, he suggested that efficient estimators might be found, where the coefficients or weights are chosen for convenience and mathematical tractability, rather than because they conform to some optimizing process. DOWNTON (1966a) demonstrated that these estimators are quite highly efficient for the normal and the extreme value distributions. DOWNTON proposed so–called **linear estimators with polynomial coefficients** of the form

$$\widehat{a}_{D,p} = \sum_{k=0}^{p} (k+1)\,\vartheta_k \sum_{i=1}^{n} \frac{(i-1)^{(k)}}{n^{(k+1)}}\, X_{i:n}, \tag{4.39a}$$

$$\widehat{b}_{D,p} = \sum_{k=0}^{p} (k+1)\,\lambda_k \sum_{i=1}^{n} \frac{(i-1)^{(k)}}{n^{(k+1)}}\, X_{i:n}, \tag{4.39b}$$

where $\ell^{(r)}$ $(r,\ \ell$ being integers) denotes the $r$–th factorial power of $\ell$, i.e.

$$\ell^{(r)} = \frac{\ell!}{(\ell-r)!} = \ell\,(\ell-1)\ldots(\ell-r+1). \tag{4.40a}$$

We further have, see DOWNTON (1966a), the identities

$$\sum_{i=1}^{n} (i-1)^{(k)} = \frac{n^{(k+1)}}{k+1}, \tag{4.40b}$$

and with $\mathrm{E}\big[(X_{i:n}-a)\big/b\big] = \alpha_{i:n}$,

$$\sum_{i=1}^{n} (i-1)^{(k)}\,\alpha_{i:n} = \frac{n^{(k+1)}}{k+1}\,\alpha_{k+1:k+1}. \tag{4.40c}$$

Thus, taking expectation of (4.39a,b) we get

$$\mathrm{E}\big(\widehat{a}_{D,p}\big) = a \sum_{k=0}^{p} \vartheta_k + b \sum_{i=1}^{n} \vartheta_k\, \alpha_{k+1:k+1}, \tag{4.41a}$$

$$\mathrm{E}\big(\widehat{b}_{D,p}\big) = a \sum_{k=0}^{p} \lambda_k + b \sum_{i=1}^{n} \lambda_k\, \alpha_{k+1:k+1}. \tag{4.41b}$$

[18] DOWNTON's method has not been implemented in the program LEPP of Chapter 6.

The polynomial coefficients $\vartheta_k$ and $\lambda_k$ $(k = 0, \ldots, p)$ have to be determined such that the estimators are unbiased and best (in least–squares sense). Introducing the vectors

$$\begin{aligned} \boldsymbol{\vartheta} &= (\vartheta_0, \vartheta_1, \ldots, \vartheta_p)', && \text{(4.42a)} \\ \boldsymbol{\lambda} &= (\lambda_0, \lambda_1, \ldots, \lambda_p)', && \text{(4.42b)} \\ \boldsymbol{\alpha}^\diamond &= (\alpha_{1:1}, \alpha_{2:2}, \ldots, \alpha_{p+1:p+1})', && \text{(4.42c)} \end{aligned}$$

and $\mathbf{1}$ for the column vector of $(p+1)$ ones, unbiasedness is given, see Sect. 4.2.1.1, when

$$\begin{pmatrix} \boldsymbol{\vartheta} \\ \boldsymbol{\lambda} \end{pmatrix} \big(\mathbf{1} \;\; \boldsymbol{\alpha}^\diamond\big) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.43a}$$

and when

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_{D,p}\big) = b^2 \begin{pmatrix} \boldsymbol{\vartheta} \\ \boldsymbol{\lambda} \end{pmatrix} \boldsymbol{\Omega}^\diamond \big(\boldsymbol{\vartheta} \;\; \boldsymbol{\lambda}\big), \quad \widehat{\boldsymbol{\theta}}_{D,p} = \big(\,\widehat{a}_{D,p} \;\; \widehat{b}_{D,p}\big)', \tag{4.43b}$$

is a minimum. $\boldsymbol{\Omega}^\diamond$ is the variance–covariance matrix of the random variables

$$W_k = (k+1) \sum_{i=1}^{n} \frac{(i-1)^{(k)}}{n^{(k+1)}}\, Y_{i:n}, \quad Y_{i:n} = \frac{X_{i:n} - a}{b}\,. \tag{4.43c}$$

This minimum variance–covariance matrix reads

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}_{D,p}\big) = b^2 \left[\begin{pmatrix} \mathbf{1}' \\ \boldsymbol{\alpha}^{\diamond\prime} \end{pmatrix} \big(\boldsymbol{\Omega}^\diamond\big)^{-1} \big(\mathbf{1} \;\; \boldsymbol{\alpha}^\diamond\big)\right]^{-1}. \tag{4.43d}$$

The procedure is rather simple for $p = 1$ (polynomial of degree 1), which gives an estimator consisting simply of two terms having constant and linear coefficients, respectively, for the two parameters $a,\ b$. In this case minimization is not necessary, as only two coefficients are required for each estimator and these are determined by the two conditions due to unbiasedness, see (4.43a). The solutions are

$$\begin{aligned} \vartheta_0 &= -\frac{\alpha_{2:2}}{\alpha_{1:1} - \alpha_{2:2}}, \qquad \vartheta_1 = \frac{\alpha_{1:1}}{\alpha_{1:1} - \alpha_{2:2}}, && \text{(4.44a)} \\ \lambda_0 &= -\frac{1}{\alpha_{2:2} - \alpha_{1:1}}, \qquad \lambda_1 = \frac{1}{\alpha_{2:2} - \alpha_{1:1}}\,. && \text{(4.44b)} \end{aligned}$$

### 4.2.3 Approximations to GLS with approximated moments of order statistics

In (2.23) through (2.25) we have presented the series approximation of $\alpha_{r:n},\ \beta_{r,r:n}$ and $\beta_{r,s:n}, r < s$. These approximated moments may be used as substitutes in the estimation approaches of Sections 4.2.1 and 4.2.2 when the exact moments are difficult to evaluate or tabulated values are not available. We will denote the **approximated moments** as

$$\alpha^*_{r:n}, \quad \beta^*_{r,s:n}, \; r \le s,$$

and the corresponding vector, matrix and inverse matrix as

$$\boldsymbol{\alpha}^*, \quad \boldsymbol{B}^*, \quad \boldsymbol{\Omega}^* := (\boldsymbol{B}^*)^{-1}.$$

$\alpha^*_{r:n}$ and $\beta^*_{r,s:n}$ will be different from the correct values $\alpha_{r:n} = \mathrm{E}(Y_{r:n})$ and $\beta_{r,s:n} = \mathrm{E}\big[(Y_{r:n} - \alpha_{r:n})\,(Y_{s:n} - \alpha_{s:n})\big]$. Thus, we will commit a systematic (= non–random) error with respect to GLS when using the approximated moments. These differences or biases will carry over to the estimator of $\boldsymbol{\theta} = (a,\ b)'$ and its total error has to be evaluated by the MSE matrix incorporating the systematic error and the random error of sampling.

Applying $\alpha^*_{r:n}$ and $\beta^*_{r,s:n}$ to the GLS method of Sect. 4.2.1.1 we have

$$\widehat{\boldsymbol{\theta}}^* = \big(\boldsymbol{A}^{*\prime}\,\boldsymbol{\Omega}^*\,\boldsymbol{A}^*\big)^{-1}\,\boldsymbol{A}^{*\prime}\,\boldsymbol{\Omega}^*\,\boldsymbol{x} \tag{4.45}$$

with

$$\boldsymbol{A}^* = \big(\boldsymbol{1} \quad \boldsymbol{\alpha}^*\big).$$

In order to find the bias of $\widehat{\boldsymbol{\theta}}^*$ we first substitute $\boldsymbol{x}$ by its definition (4.5a) and then take expectation. According to (4.10b) with

$$\boldsymbol{C} = \big(\boldsymbol{A}^{*\prime}\,\boldsymbol{\Omega}^*\,\boldsymbol{A}^*\big)^{-1}\,\boldsymbol{A}^{*\prime}\,\boldsymbol{\Omega}^* \tag{4.46a}$$

and observing that $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{o}$ we arrive at

$$\mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big) = \boldsymbol{C}\,\boldsymbol{A}\,\boldsymbol{\theta}. \tag{4.46b}$$

Thus, the **bias of** $\widehat{\boldsymbol{\theta}}^*$ is

$$\boldsymbol{\delta} = \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big) - \boldsymbol{\theta} = \big(\boldsymbol{C}\,\boldsymbol{A} - \boldsymbol{I}\big)\,\boldsymbol{\theta}. \tag{4.46c}$$

The difference between $\widehat{\boldsymbol{\theta}}^*$ and its mean $\mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big)$ is

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^* - \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big) &= \boldsymbol{C}\,\boldsymbol{x} - \boldsymbol{C}\,\boldsymbol{A}\,\boldsymbol{\theta} \\
&= \boldsymbol{C}\big(\boldsymbol{A}\,\boldsymbol{\theta} + \boldsymbol{\varepsilon} - \boldsymbol{A}\,\boldsymbol{\theta}\big) \\
&= \boldsymbol{C}\,\boldsymbol{\varepsilon}.
\end{aligned} \tag{4.47a}$$

So, the **variance–covariance matrix of** $\widehat{\boldsymbol{\theta}}^*$ is

$$\begin{aligned}
\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}^*\big) &= \mathrm{E}\Big\{\big[\widehat{\boldsymbol{\theta}}^* - \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big)\big]\,\big[\widehat{\boldsymbol{\theta}}^* - \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big)\big]'\Big\} \\
&= \mathrm{E}\big\{(\boldsymbol{C}\,\boldsymbol{\varepsilon})\,(\boldsymbol{C}\,\boldsymbol{\varepsilon})'\big\} \\
&= \mathrm{E}\big\{\boldsymbol{C}\,\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'\,\boldsymbol{C}'\big\} \\
&= \boldsymbol{C}\,\underbrace{\mathrm{E}(\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}')}_{=\,b^2\,\boldsymbol{B}}\,\boldsymbol{C}' \\
&= b^2\,\boldsymbol{C}\boldsymbol{B}\,\boldsymbol{C}'.
\end{aligned} \tag{4.47b}$$

$\operatorname{Var}\big(\widehat{\boldsymbol{\theta}}^*\big)$ only incorporates the random fluctuation of $\widehat{\boldsymbol{\theta}}^*$ around its mean $\mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big)$ but not the systematic errors due to the approximations. The total squared–error matrix of $\widehat{\boldsymbol{\theta}}^*$, the **MSE–matrix of $\widehat{\boldsymbol{\theta}}^*$**, is

$$\begin{aligned}
\operatorname{MSE}\big(\widehat{\boldsymbol{\theta}}^*\big) &= \mathrm{E}\Big\{\big(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}\big)\big(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}\big)'\Big\} \\
&= \mathrm{E}\Big\{\Big(\big[\widehat{\boldsymbol{\theta}}^* - \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big)\big] + \big[\underbrace{\mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big) - \boldsymbol{\theta}}_{=\boldsymbol{\delta}}\big]\Big)\Big(\big[\widehat{\boldsymbol{\theta}}^* - \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big)\big] + \big[\underbrace{\mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big) - \boldsymbol{\theta}}_{=\boldsymbol{\delta}}\big]\Big)'\Big\} \\
&= \underbrace{\mathrm{E}\Big\{\big[\widehat{\boldsymbol{\theta}}^* - \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big)\big]\big[\widehat{\boldsymbol{\theta}}^* - \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*\big)\big]'\Big\}}_{=\operatorname{Var}\left(\widehat{\boldsymbol{\theta}}^*\right)} + \boldsymbol{\delta}\,\boldsymbol{\delta}' \\
&= b^2\,\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}' + \boldsymbol{\delta}\,\boldsymbol{\delta}'. \qquad (4.47c)
\end{aligned}$$

**Excursus: Best linear invariant estimators**

In the context of biased estimation we shortly mention the **BLIE** approach (best linear invariant estimator) of MANN (1969). The resulting estimators are biased, but invariant under location and scale transformations and have minimum mean–squared error among the class of linear estimators. This minimum MSE is sometimes smaller than the variance of a BLUE. The BLIEs are obtained from the BLUEs $\widehat{a}$, $\widehat{b}$ as

$$\widehat{a}_{BLIE} = \widehat{a} - \frac{C}{1+B}\,\widehat{b}, \quad \widehat{b}_{BLIE} = \frac{\widehat{b}}{1+B}$$

with the following MSEs:

$$\operatorname{MSE}\big(\widehat{a}_{BLIE}\big) = b^2\left(A - \frac{C^2}{1+B}\right), \quad \operatorname{MSE}\big(\widehat{b}_{BLIE}\big) = b^2\,\frac{B}{1+B},$$

and expected cross–product of the estimation errors

$$\mathrm{E}\big[\big(\widehat{a}_{BLIE} - a\big)\big(\widehat{b}_{BLIE} - b\big)\big] = b^2\,\frac{C}{1+B},$$

where

$$A = b^2\operatorname{Var}\big(\widehat{a}\big),\; B = b^2\operatorname{Var}\big(\widehat{b}\big),\; C = b^2\operatorname{Cov}\big(\widehat{a},\widehat{b}\big).$$

**Example 4/7: Bias and MSE of $\widehat{\theta^*}$ in a sample from an exponential and a normal distribution ($n = 10$)**

The series approximations (2.23) through (2.25) for the moments of the reduced **exponential distribution** are

$$\boldsymbol{\alpha}^* = (0.0100,\; 0.2111,\; 0.3361,\; 0.4789,\; 0.6456,\; 0.8456,\; 1.0955,\; 1.4288,\; 1.9289,\; 2.9332)'$$

and — only giving the upper triangle of the variance–covariance matrix:

$$\boldsymbol{B}^* = \begin{pmatrix} 0.0100 & 0.0100 & 0.0100 & 0.0100 & 0.0100 & 0.0100 & 0.0100 & 0.0100 & 0.0100 & 0.0100 \\ & 0.0223 & 0.0223 & 0.0223 & 0.0223 & 0.0223 & 0.0233 & 0.0233 & 0.0233 & 0.0233 \\ & & 0.0378 & 0.0378 & 0.0378 & 0.0378 & 0.0378 & 0.0378 & 0.0378 & 0.0378 \\ & & & 0.0581 & 0.0581 & 0.0581 & 0.0581 & 0.0581 & 0.0581 & 0.0581 \\ & & & & 0.0858 & 0.0858 & 0.0858 & 0.0858 & 0.0858 & 0.0858 \\ & & & & & 0.1255 & 0.1255 & 0.1255 & 0.1255 & 0.1255 \\ & & & & & & 0.1875 & 0.1875 & 0.1875 & 0.1875 \\ & & & & & & & 0.2974 & 0.2974 & 0.2974 \\ & & & & & & & & 0.5433 & 0.5433 \\ & & & & & & & & & 1.5230 \end{pmatrix}.$$

The true values are in Example 4/1. With the data $\boldsymbol{x}$ of Example 4/1 we find the approximated GLS estimates according to (4.45) as

$$\widehat{\boldsymbol{\theta}^*} = \begin{pmatrix} \widehat{a^*} \\ \widehat{b^*} \end{pmatrix} = \begin{pmatrix} 15.0844 \\ 99.1610 \end{pmatrix}.$$

These estimates are very close to the true GLS estimates

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \begin{pmatrix} 15.0889 \\ 99.1111 \end{pmatrix}.$$

We have generated the observations with $\boldsymbol{\theta} = (10,\ 100)'$, so we can evaluate the bias according to (4.46c):

$$\boldsymbol{\delta} = \begin{pmatrix} 0.0048 \\ -0.0430 \end{pmatrix}.$$

The variance–covariance matrix (4.47b) is

$$\text{Var}\big(\widehat{\boldsymbol{\theta}^*}\big) = b^2 \begin{pmatrix} 0.0111 & -0.0111 \\ -0.0111 & 0.1110 \end{pmatrix}$$

which — because $\boldsymbol{\delta}$ is very small — turns out to be practically equal to the MSE–matrix (4.47c). The variance–covariance matrix of the true GLS estimates is

$$\text{Var}\big(\widehat{\boldsymbol{\theta}}\big) = b^2 \begin{pmatrix} 0.0111 & -0.0111 \\ -0.0111 & 0.1111 \end{pmatrix}.$$

For a sample of size $n = 10$ from a **normal distribution**[19] and using the short versions of the

[19] Missing values may be found by the symmetry relations:

$$\begin{aligned} \alpha_{r:n} &= -\alpha_{n-r:n} \\ \beta_{r,s:n} &= \beta_{n-s+1,n-r+1:n}. \end{aligned}$$

approximating formulas, which turned out to be more accurate than the long versions,[20] we find:

$$\alpha^*_{6:10} = 0.1226,\ \alpha^*_{7:10} = 0.3755,\ \alpha^*_{8:10} = 0.6556,\ \alpha^*_{9:10} = 1.0006,\ \alpha*_{10:10} = 1.5393,$$

$$\begin{aligned}
&\beta^*_{1,1:10} = 0.3322, \quad &&\beta^*_{2,2:10} = 0.2097, \quad &&\beta^*_{3,5:10} = 0.1063,\\
&\beta^*_{1,2:10} = 0.1680, &&\beta^*_{2,3:10} = 0.1442, &&\beta^*_{3,6:10} = 0.0881,\\
&\beta^*_{1,3:10} = 0.1148, &&\beta^*_{2,4:10} = 0.1102, &&\beta^*_{3,7:10} = 0.0740,\\
&\beta^*_{1,4:10} = 0.0874, &&\beta^*_{2,5:10} = 0.0886, &&\beta^*_{3,8:10} = 0.0624,\\
&\beta^*_{1,5:10} = 0.0704, &&\beta^*_{2,6:10} = 0.0734, &&\beta^*_{4,4:10} = 0.1555,\\
&\beta^*_{1,6:10} = 0.0580, &&\beta^*_{2,7:10} = 0.0616, &&\beta^*_{4,5:10} = 0.1257,\\
&\beta^*_{1,7:10} = 0.0486, &&\beta^*_{2,8:10} = 0.0518, &&\beta^*_{4,6:10} = 0.1044,\\
&\beta^*_{1,8:10} = 0.0409, &&\beta^*_{2,9:10} = 0.0431, &&\beta^*_{4,7:10} = 0.0878,\\
&\beta^*_{1,9:10} = 0.0339, &&\beta^*_{3,3:10} = 0.1719, &&\beta^*_{5,5:10} = 0.1488,\\
&\beta^*_{1,10:10} = 0.0267, &&\beta^*_{3,4:10} = 0.1318, &&\beta^*_{5,6:10} = 0.1238.
\end{aligned}$$

Let the true parameters be $a = 0,\ b = 1$. Then the vector of biases is

$$\boldsymbol{\delta} = 10^{-4}\begin{pmatrix}0.0000\\0.5671\end{pmatrix},$$

i.e. the $a$–estimator will be unbiased and the $b$–estimator is positively biased by only $0.005671\%$. The variance–covariance matrix of $\widehat{\boldsymbol{\theta}^*}$ is

$$\text{Var}\big(\widehat{\boldsymbol{\theta}^*}\big) = \begin{pmatrix} 0.1000 & 0.0000 \\ 0.0000 & 0.0576 \end{pmatrix}$$

and the MSE–matrix is practically the same and nearly identical to the GLS variance–covariance matrix $\text{Var}\big(\widehat{\boldsymbol{\theta}}\big)$.

We may also apply the series approximated moments to the GLS approaches where $\boldsymbol{B}$ has been substituted by

1. the diagonal matrix $\boldsymbol{B}_d$,
2. the identity matrix $\boldsymbol{I}$,
3. the asymptotic variance–covariance matrix $\boldsymbol{B}_B$.

[20] For the definition of the short version see text following formula (2.25).

In the first case ($\boldsymbol{B}_d \implies \boldsymbol{B}$) we have — besides the means — to substitute the variances $\beta_{r,r:n}$ in $\boldsymbol{B}_d$ by their series approximations $\beta^*_{r,r:n}$. The approximated matrix is denoted as $\boldsymbol{B}^*_d$ and its inverse as $\boldsymbol{\Omega}^*_d$. The results are:

$$\widehat{\boldsymbol{\theta}}^*_d = \boldsymbol{C}_d\,\boldsymbol{x}, \tag{4.48a}$$

$$\boldsymbol{C}_d = \big(\boldsymbol{A}^{*\prime}\,\boldsymbol{\Omega}^*_d\,\boldsymbol{A}^*\big)^{-1}\,\boldsymbol{A}^{*\prime}\,\boldsymbol{\Omega}^*_d, \tag{4.48b}$$

$$\mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*_d\big) = \boldsymbol{C}_d\,\boldsymbol{A}\,\boldsymbol{\theta}, \tag{4.48c}$$

$$\boldsymbol{\delta}_d = \mathrm{E}\big(\widehat{\boldsymbol{\theta}}^*_d\big) - \boldsymbol{\theta} = \big(\boldsymbol{C}_d\,\boldsymbol{A} - \boldsymbol{I}\big)\,\boldsymbol{\theta}, \tag{4.48d}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}^*_d\big) = b^2\,\boldsymbol{C}_d\,\boldsymbol{B}\,\boldsymbol{C}'_d, \tag{4.48e}$$

$$\mathrm{MSE}\big(\widehat{\boldsymbol{\theta}}^*_d\big) = \mathrm{Var}\big(\widehat{\boldsymbol{\theta}}^*_d\big) + \boldsymbol{\delta}_d\,\boldsymbol{\delta}'_d. \tag{4.48f}$$

In the second case ($\boldsymbol{I} \implies \boldsymbol{B}$) we only have to substitute the means by their series approximates. The results are:

$$\widehat{\boldsymbol{\theta}^*_{\boldsymbol{I}}} = \boldsymbol{C}_{\boldsymbol{I}}\,\boldsymbol{x}, \tag{4.49a}$$

$$\boldsymbol{C}_{\boldsymbol{I}} = \big(\boldsymbol{A}^{*\prime}\,\boldsymbol{A}^*\big)^{-1}\,\boldsymbol{A}^{*\prime}, \tag{4.49b}$$

$$\mathrm{E}\big(\widehat{\boldsymbol{\theta}^*_{\boldsymbol{I}}}\big) = \boldsymbol{C}_{\boldsymbol{I}}\,\boldsymbol{A}\,\boldsymbol{\theta}, \tag{4.49c}$$

$$\boldsymbol{\delta}_{\boldsymbol{I}} = \mathrm{E}\big(\widehat{\boldsymbol{\theta}^*_{\boldsymbol{I}}}\big) - \boldsymbol{\theta} = \big(\boldsymbol{C}_{\boldsymbol{I}}\boldsymbol{A} - \boldsymbol{I}\big)\,\boldsymbol{\theta}, \tag{4.49d}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}^*_{\boldsymbol{I}}}\big) = b^2\,\boldsymbol{C}_{\boldsymbol{I}}\,\boldsymbol{B}\,\boldsymbol{C}'_{\boldsymbol{I}}, \tag{4.49e}$$

$$\mathrm{MSE}\big(\widehat{\boldsymbol{\theta}^*_{\boldsymbol{I}}}\big) = \mathrm{Var}\big(\widehat{\boldsymbol{\theta}^*_{\boldsymbol{I}}}\big) + \boldsymbol{\delta}_{\boldsymbol{I}}\,\boldsymbol{\delta}'_{\boldsymbol{I}}. \tag{4.49f}$$

In third case ($\boldsymbol{B}_B \implies \boldsymbol{B}$) we have BLOM's **nearly unbiased, nearly best linear estimator**. The results are:

$$\widehat{\boldsymbol{\theta}^*_B} = \boldsymbol{C}_B\,\boldsymbol{x}, \tag{4.50a}$$

$$\boldsymbol{C}_B = \big(\boldsymbol{A}^{*\prime}\,\boldsymbol{B}_B^{-1}\,\boldsymbol{A}^*\big)^{-1}\,\boldsymbol{A}^{*\prime}\,\boldsymbol{B}_B^{-1}, \tag{4.50b}$$

$$\mathrm{E}\big(\widehat{\boldsymbol{\theta}^*_B}\big) = \boldsymbol{C}_B\,\boldsymbol{A}\,\boldsymbol{\theta}, \tag{4.50c}$$

$$\boldsymbol{\delta}_B = \mathrm{E}\big(\widehat{\boldsymbol{\theta}^*_B}\big) - \boldsymbol{\theta} = \big(\boldsymbol{C}_B\,\boldsymbol{A} - \boldsymbol{I}\big)\,\boldsymbol{\theta}, \tag{4.50d}$$

$$\mathrm{Var}\big(\widehat{\boldsymbol{\theta}^*_B}\big) = b^2\,\boldsymbol{C}_B\,\boldsymbol{B}\,\boldsymbol{C}'_B, \tag{4.50e}$$

$$\mathrm{MSE}\big(\widehat{\boldsymbol{\theta}^*_B}\big) = \mathrm{Var}\big(\widehat{\boldsymbol{\theta}^*_B}\big) + \boldsymbol{\delta}_B\,\boldsymbol{\delta}'_B. \tag{4.50f}$$

## 4.3 Estimators based on empirical percentiles

Sometimes there are situations where the data are such that we have to base the plotting positions on an estimate of the CDF and take regressors deduced from this CDF. These regressors are the **empirical percentiles**. In Sect. 4.3.1 we show how to proceed with grouped data and in Sect. 4.3.2 we describe how to treat data sets consisting of a mixture of complete and censored data arising from progressively censored samples or randomly censored samples.

### 4.3.1 Grouped data

When the sample size is great the data naturally come in grouped form. As has been described in Sect. 4.1 we now dispose of **upper class limits** $x_j^u$ $(j = 1, \ldots, k)$ and **cumulated frequencies**

$$n_j^u = \sum_{i=1}^{j} n_i; \;\; j = 1, \ldots, k; \tag{4.51}$$

$n_i$ being the class frequency. For plotting the data on probability paper we have to resort to plotting positions that use an estimator $\widehat{P}_j$ of the CDF $P_j = F(x_j^u)$. From Tab. 3/1 we may choose among the following estimators to be plotted on the probability–labeled ordinate against $x_j^u$ on the abscissa:

- the **nïve estimator**

$$\widehat{P}_j = \frac{n_j^u}{n}\,, \tag{4.52a}$$

- the **midpoint position**

$$\widehat{P}_j = \frac{n_j^u - 0.5}{n}\,, \tag{4.52b}$$

- the **BLOM position**[21]

$$\widehat{P}_j = \frac{n_j^u - 0.375}{n + 0.25}\,, \tag{4.52c}$$

- the**WEIBULL position** or **mean plotting position**

$$\widehat{P}_j = \frac{n_j^u}{n + 1}\,, \tag{4.52d}$$

[21] BLOM (1958) suggested that a good approximation to the mean $\alpha_{i:n}$ of the reduced order statistic $Y_{i:n}$ is

$$\widehat{\alpha}_{i:n} = F_Y^{-1}(\pi_i)$$

where

$$\pi_i = \frac{i - \alpha_i}{n - \alpha_i - \beta_i + 1}\,.$$

BLOM used $3/8$ for both these constants and arrived at (4.52c).

- the **median plotting position**

$$\widehat{P}_j = \frac{n_j^u - 0.3}{n + 0.4}\,, \tag{4.52e}$$

- the **mode plotting position**

$$\widehat{P}_j = \frac{n_j^u - 1}{n - 1}\,. \tag{4.52f}$$

The regressor to be used in the least–squares estimator's formula will be the estimated reduced percentile of order $\widehat{P}_j$ of the relevant location–scale distribution:

$$\widehat{\alpha}_j = F_Y^{-1}\big(\,\widehat{P}_j\big) \tag{4.53a}$$

and the regression equation now reads

$$x_j^u = a + b\,\widehat{\alpha}_j + \eta_j;\ j = 1, \ldots, k; \tag{4.53b}$$

where $\eta_j$ is an error variable incorporating sampling errors as well as non–random errors due to approximating the means of the order statistics. The asymptotic variance–covariance matrix of the model (4.53b), according to MOSTELLER (1946), is

$$\text{Var}\big(\boldsymbol{x}^u\,\big) = b^2\,\widehat{\boldsymbol{B}};\ \ \boldsymbol{x}^u = (x_1^u, \ldots, x_k^u)'; \tag{4.53c}$$

where the elements $\widehat{\beta}_{j\ell}$ of $\widehat{\boldsymbol{B}}$ read

$$\widehat{\beta}_{j\ell} = \frac{\widehat{P}_j\,\big(1 - \widehat{P}_\ell\big)}{n\,f_Y\big(\,\widehat{\alpha}_j\big)\,f_Y\big(\,\widehat{\alpha}_\ell\big)}\,;\ \ 1 \le j \le \ell \le k. \tag{4.53d}$$

$f_Y\big(\,\widehat{\alpha}_j\big)$ is the reduced DF evaluated at $\widehat{\alpha}_j$. When the last class is open to the right, i.e. $x_k^u = \infty$, we have to do without this class.[22]

The approximate GLS estimator of $\boldsymbol{\theta} = (a,\ b)'$, based on the empirical percentiles resulting from the estimated CDF, is

$$\widehat{\widehat{\boldsymbol{\theta}}} = \widehat{\boldsymbol{C}}\,\boldsymbol{x}^u \tag{4.54a}$$

with

$$\widehat{\boldsymbol{C}} \;=\; \big(\widehat{\boldsymbol{A}}'\,\widehat{\boldsymbol{B}}^{-1}\,\widehat{\boldsymbol{A}}\big)^{-1}\,\widehat{\boldsymbol{A}}'\,\widehat{\boldsymbol{B}}^{-1}, \tag{4.54b}$$

$$\widehat{\boldsymbol{A}} \;=\; (\boldsymbol{1}\ \ \widehat{\boldsymbol{\alpha}}),\ \ \widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_k)', \tag{4.54c}$$

$$\boldsymbol{x}^u \;=\; (x_1^u, x_2^u, \ldots, x_k^u). \tag{4.54d}$$

[22] The reader should compare this asymptotic variance–covariance matrix of the empirical quantiles with the asymptotic variance–covariance matrix $\boldsymbol{B}_B$ of BLOM's estimator and notice the similarity as well as the difference.

The estimator $\widehat{\widehat{\boldsymbol{\theta}}}$ will be biased and not best. We may estimate its variance–covariance matrix based on the asymptotic variance–covariance matrix of (4.53c) as

$$\widehat{\text{Var}\big(\widehat{\widehat{\boldsymbol{\theta}}}\big)} = \widehat{b}^2 \left(\widehat{\boldsymbol{A}}' \, \widehat{\boldsymbol{B}}^{-1} \, \widehat{\boldsymbol{A}}\right)^{-1}. \tag{4.54e}$$

We have to decide what estimator $\widehat{P}_j$ to be used. The plotting positions in (4.52a) and (4.52f) have to be excluded because for most distributions the last observation point $(x_k^u,\ 1)$ cannot be depicted on most probability papers, even if $x_k^u$ is finite. CHERNOFF/LIEBERMAN (1954) have shown in the normal distribution case that the choice of the frequently recommended WEIBULL position leads to much poorer estimates of $\sigma = b$, as measured by mean squared error, than those obtained with the midpoint position. The program LEPP asks the reader to make his choice among the positions (4.52b) through (4.52e).

**Excursus: The idea behind the procedure for grouped data**

The procedure proposed for treating grouped data rests upon the asymptotic theory for constructing estimators of $\boldsymbol{\theta}$ based on a few selected order statistics. This approach is as follows: A $k$–tuple $(P_1,\ P_2,\ \ldots,\ P_k)$ with $0 < P_1 < P_2 < \ldots < P_k < 1$ is specified.[23] In a random sample of size $n$ the order statistics $X_{n_j:n}$ are chosen, where $n_j = [n\,P_j] + 1$ ($[m]$ indicating the greatest integer contained in $m$). The regressor values in vector $\boldsymbol{\alpha}$ are $\alpha_j = F_Y^{-1}(P_j)$. Asymptotically, the random vector $\boldsymbol{x} = (X_{n_1:n},\ \ldots,\ X_{n_k:n})'$ has a $k$–variate normal distribution with mean vector $\boldsymbol{\mu} = a\,\mathbf{1} + b\,\boldsymbol{\alpha}$ and variance–covariance matrix $\text{Var}(\boldsymbol{x}) = b^2\boldsymbol{B}$ where the elements of $\boldsymbol{B}$ are $\beta_{j\ell} = P_j\,(1-P_\ell)\big/\big[n\,f_Y(\alpha_j)\,f_Y(\alpha_\ell)\big]$.

The procedure proposed to deal with grouped data differs from the theory above in so far as the upper class limits $x_j^u$ — taken as substitutes for $X_{n_j:n}$ — are fix and the $k$–tuple $(\widehat{P}_1,\ \ldots,\ \widehat{P}_k)$ — taken as a substitute for $(P_1,\ \ldots,\ P_k)$ — is random. Thus, the procedure used is nothing but a crude approximation.

**Example 4/8: Estimating with grouped data from an exponential distribution**

$n = 50$ exponentially distributed random numbers have been generated with $a = 10$ and $b = 100$ and grouped into $k = 7$ classes with the following result:

| $x_j^u$ | 30 | 70 | 120 | 200 | 300 | 500 | $\infty$ |
|---|---|---|---|---|---|---|---|
| $n_j^u$ | 13 | 25 | 34 | 42 | 46 | 49 | 50 |

[23] The problem of optimally choosing the $k$–tuple is treated in the fundamental work of OGAWA (1951, 1962). An up–to–date presentation of this topic is to be found in BALAKRISHNAN/COHEN (1991, Chapter 7).

The intermediate results of estimation — omitting the last class with $x_k^u = \infty$ — are:

| $j$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\widehat{P}_j$ | 0.2549 | 0.4902 | 0.6667 | 0.8235 | 0.9020 | 0.9608 |
| $\widehat{\alpha}_j$ | 0.2942 | 0.6737 | 1.0986 | 1.7346 | 2.3224 | 3.2387 |
| $f_Y(\widehat{\alpha}_j)$ | 0.7451 | 0.5098 | 0.3333 | 0.1756 | 0.0980 | 0.0392 |

$\widehat{\boldsymbol{B}}$ — only showing the upper triangle of this symmetric matrix – reads

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} 0.0068 & 0.0068 & 0.0068 & 0.0068 & 0.0068 & 0.0068 \\ & 0.0192 & 0.0192 & 0.0192 & 0.0192 & 0.0192 \\ & & 0.0400 & 0.0400 & 0.0400 & 0.0400 \\ & & & 0.0933 & 0.0933 & 0.0933 \\ & & & & 0.1840 & 0.1840 \\ & & & & & 0.4900 \end{pmatrix}$$

The estimated parameters are

$$\begin{pmatrix} \widehat{\widehat{a}} \\ \widehat{\widehat{b}} \end{pmatrix} = \begin{pmatrix} -7.9964 \\ 129.1343 \end{pmatrix}$$

with estimated variance–covariance matrix

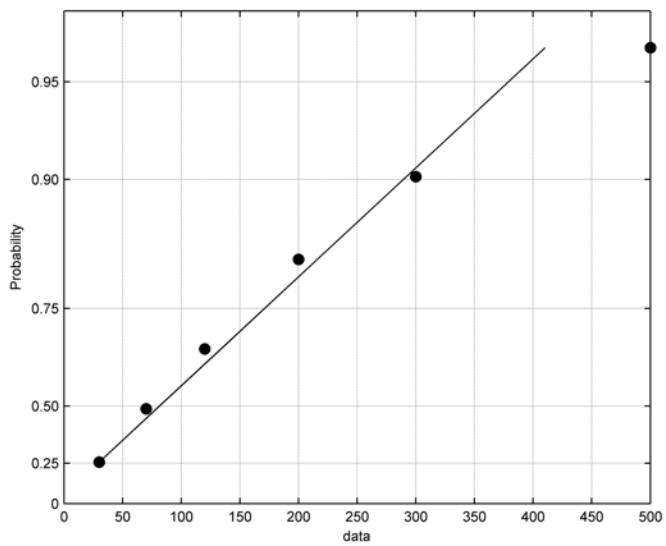
$$\widehat{\mathrm{Var}\big(\widehat{\widehat{\boldsymbol{\theta}}}\big)} = \begin{pmatrix} 155.9987 & -142.4077 \\ -142.4077 & 483.9857 \end{pmatrix}.$$

Fig. 4/4 shows the data together with the estimated regression line. For an explanation of the apparently bad fit of the regression line see Example 5/1. The fit in the upper $x$–region does not seem to be good due to the great variance of the last ranking observation, but when we refer to the asymptotic normal distribution of the estimators and set up two–sided 95%–confidence intervals for both parameters we have

$$\begin{aligned} -7.9964 - 1.96\sqrt{155.9987} &\le a \le -7.9964 + 1.96\sqrt{155.9987} \\ -32.4767 &\le a \le 16.4839 \\ 129.1343 - 1.96\sqrt{483.9857} &\le b \le 129.1343 + 1.96\sqrt{483.9857} \\ 86.0149 &\le b \le 172.2537. \end{aligned}$$

The true parameter values $a = 10$ and $b = 100$ of the sample generating distribution are within these intervals.

Figure 4/4: Probability plot and estimated regression line — $n = 50$, grouped, from an exponential distribution




### 4.3.2 Randomly and multiply censored data

When the data come from a sample which has been randomly or multiply censored the sorted sequence of observations (sorted in ascending order) will be a series where censored and uncensored observations follow one another in a non–systematic pattern. Each item in such a series will bear two entries:

1. a real number showing a measured value $t_{i:n}$ $(i = 1, \ldots, n)$ that may be either the value of some censoring variable $C$ or the value of the variable $X$ under study,

2. a binary indicator $\vartheta_i$ telling whether $t_{i:n}$ is censored or not:

$$\vartheta_i = \begin{cases} 0, & \text{if } t_{i:n} \text{ is uncensored,} \\ 1, & \text{if } t_{i:n} \text{ is censored.} \end{cases}$$

The estimation process and the probability plotting explicitly make use of the uncensored observations only, but the censored observations, which bear some information, will be allowed for implicitly. The non–parametric approach, which is applied here to estimate $R_X(x)$, the complementary CDF (= reliability or survival function) of $X$, is the **product–limit estimator** of KAPLAN/MEIER (1958). Once we have found the estimate of $R_X(x)$ probability plotting and linear estimation are executed along the lines of the preceding section.

In the following description of the KAPLAN–MEIER approach we will call $X$ a lifetime. Corresponding to each $x_i$, i.e. an uncensored observation $t_{i:n}$, is $n_i$ the **number at risk** just prior to time $x_i$ and $d_i$ the number of failures at $x_i$. The KAPLAN–MEIER estimator is the non–parametric maximum–likelihood estimator of $R_X(x)$. It is a product of the form

$$\widehat{R}_X(x) = \prod_{x_i \le x} \frac{n_i - d_i}{n_i}\,. \tag{4.55a}$$

When there is no censoring, $n_i$ is the number of survivors just prior to time $x_i$. With censoring, $n_i$ is the number of survivors less the number of losses (censored cases). It is only these surviving cases that are still being observed (have not been censored) and that are at risk of an (observed) failure. An alternate definition is

$$\widehat{\widehat{R}}_X(x_i) = \prod_{x_i < x} \frac{n_i - d_i}{n_i}\,. \tag{4.55b}$$

The two definitions differ only with respect to the observed event times. The definition in (4.55b) is left–continuous whereas that in (4.55a) is right–continuous. Note that

$$R_X(x) = \Pr(X > x) = 1 - \Pr(X \le x) = 1 - F_X(x).$$

Consequently, the right–continuous definition $\widehat{R}_X(x)$ in (4.55a) may be preferred in order to make the estimator compatible with the right–continuous estimator of $F_X(x)$. The KAPLAN–MEIER approach rests upon the intuitive idea called **redistribution–to–the–right**. The algorithm starts with an empirical distribution that puts mass $1/n$ at each observation $t_{i:n}$ and then moves the mass of each censored observation by distributing it equally to all observations to the right of it.

On probability paper we will plot

$$\widehat{P}_i = \widehat{F}_X(x_i) = 1 - \widehat{R}_X(x_i) \tag{4.56a}$$

against $x_i$. For the least–squares estimator we then take the regressor

$$\widehat{\alpha}_i = F^{-1}\big(\widehat{P}_i\big), \tag{4.56b}$$

which is the estimated reduced percentile of order $\widehat{P}_i$ of the relevant location–scale distribution and then proceed as has been described in Sect. 4.3.1 using formulas (4.53a) through (4.54e) where the vector $\boldsymbol{x}^u$ now consists of the uncensored observations $x_i$.

**Example 4/9: Randomly censored data from an exponential distribution**

The following $n = 20$ observations which have been randomly censored come from an exponential distribution with $a = 10$ and $b = 100$. The KAPLAN–MEIER estimates $\widehat{P}_i$ in the following table have been found with the command `ecdf` of MATLAB. The estimates are

$$\widehat{\widehat{a}} = 8.8158 \quad \text{and} \quad \widehat{\widehat{b}} = 83.6105$$

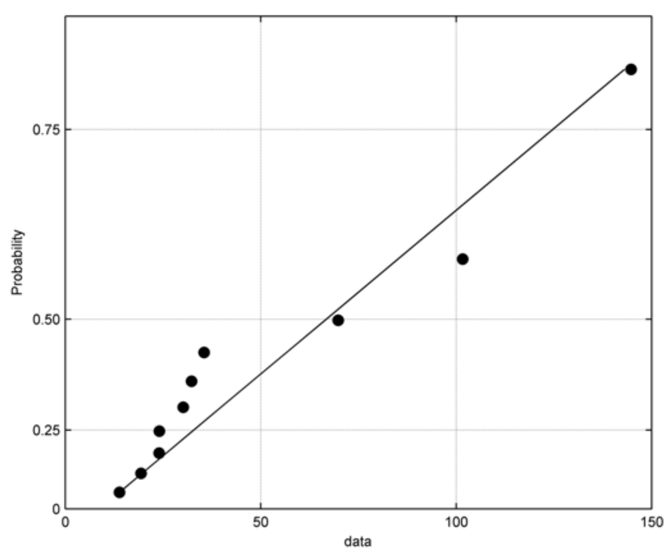
and are in good accordance with the true parameter values $a = 10$ and $b = 100$. The estimated variance of the estimates is

$$\widehat{\mathrm{Var}\big(\widehat{\widehat{\boldsymbol{\theta}}}\big)} = \begin{pmatrix} 23.6014 & -28.9558 \\ -28.9558 & 477.6240 \end{pmatrix}.$$

| $i$ | $t_{i:20}$ | $\delta_i$ | $x_i$ | $\widehat{P}_i$ | $\widehat{\alpha}_i$ |
|---|---|---|---|---|---|
| 1 | 5.6796 | 1 | – | – | – |
| 2 | 10.2347 | 1 | – | – | – |
| 3 | 10.5301 | 1 | – | – | – |
| 4 | 13.8847 | 0 | 13.8847 | 0.0588 | 0.0606 |
| 5 | 15.8959 | 1 | – | – | – |
| 6 | 19.3599 | 0 | 19.3599 | 0.1216 | 0.1296 |
| 7 | 24.0076 | 0 | 24.0076 | 0.1843 | 0.2037 |
| 8 | 24.0764 | 0 | 24.0764 | 0.2471 | 0.2838 |
| 9 | 30.1745 | 0 | 30.1745 | 0.3098 | 0.3708 |
| 10 | 32.3058 | 0 | 30.3058 | 0.3727 | 0.4661 |
| 11 | 35.5008 | 0 | 35.5008 | 0.4353 | 0.5715 |
| 12 | 69.8091 | 0 | 69.8091 | 0.4980 | 0.6892 |
| 13 | 71.1588 | 1 | – | – | – |
| 14 | 71.4876 | 1 | – | – | – |
| 15 | 87.4029 | 1 | – | – | – |
| 16 | 101.6834 | 0 | 101.6834 | 0.5984 | 0.09124 |
| 17 | 104.7105 | 1 | – | – | – |
| 18 | 142.7466 | 1 | – | – | – |
| 19 | 144.7572 | 0 | 144.7572 | 0.7992 | 1.6055 |
| 20 | 203.1333 | 0 | 203.1333 | 1.0000 | $\infty$ |

Fig. 4/5 displays the uncensored data $x_i$ of the sample together with the estimated regression line on exponential probability paper.

Figure 4/5: Probability plot and estimated regression line — $n = 20$ randomly censored from an exponential distribution



## 4.4 Goodness–of–fit

After the data have been plotted on probability paper a decision must be made whether these values tend to fall on a straight line and the chosen distribution could be the correct model. Since the order statistics are not independent there could be runs of points above and below a hypothetical straight line when the model is appropriate. Thus, tests of runs cannot be used in the assessment. In addition the variance of the order statistics in the unrestricted tails of the distribution is higher than in the central region or restricted tails. Thus, greater deviations from linearity must be allowed for in these regions. Systematic departures from linearity are the indicators of lack of fit and hence model rejection. Also the smaller the sample size the greater the tolerance that must be allowed for departures from a straight line.

Often an analytical test procedure — see below — is based in conjunction with the plot so that an objective assessment can be made. In these cases the plot serves as a graphical guide as to what the analytical procedure is telling and can indicate when the model is rejected if the rejection was due to one or a few outlying observations and give some idea of where the discrepancy in the data and the model occurred.

As LOCKHART/STEPHENS (1998) carry out there can be identified three main approaches to measuring the fit. The first is simply to measure the **correlation coefficient** $r(X_{i:n}, \alpha_i)$, $r(X_{i:n}, \alpha_i^*)$ or $r(X_{i:n}, \widehat{\alpha}_i)$ between the paired sets $X_{i:n}$ and the chosen regressor. A second method is to estimate the straight line using GLS to take into account the autocorrelation and heteroscedasticity of order statistics, and then to base the test of fit on the **sum of squares of residuals**. A closely related procedure is to fit a higher–order polynomial regression, and then to test the hypothesis that the coefficients of the higher–order terms are zero. A third technique is to estimate $b$ from the regression using GLS, and to **compare** this estimate **with** the estimate of scale given by the **sample standard deviation** for those location–scale distributions where $b$ is the standard deviation. The latter test statistic will be a modification of the well–known $W$ statistic of SHAPIRO/WILK (1965). All three approaches mentioned above have been applied to the WEIBULL distribution by RINNE (2009, pp. 667–674).

Testing with the correlation coefficient is no easy task in so far as the conventional theory for this statistic assumes that the correlation coefficient applies to two random variables, but here one of the variables — $\alpha_i,\ \alpha_i^*$ or $\widehat{\alpha}_i$, respectively — are constants which are functions of the order number $i$. For this reason LOCKHART/STEPHENS (1998) have developed the distribution theory for such a type of correlation coefficient. The program LEPP will give the correlation coefficient for the data in the probability plot, but no formal test of significance of $\mathrm{H}_0:\ \rho(\boldsymbol{x}, \boldsymbol{\alpha}) = 0$ is executed because the critical points of the test statistic are not easy to find and depend on the special location–scale distribution. Thus, the correlation coefficient shown by LEPP has to be interpreted in the descriptive sense: the nearer to one the better.

# 5 Probability plotting and linear estimation — Applications

After enumerating in Sect. 5.1 what will be shown for each location–scale distribution we present in Sect. 5.2 a great number of original distributions of this type and in Sect. 5.3 several distributions which after a log–transformation turn into this type.

## 5.1 What will be given for each distribution?

- The description of each distribution starts with its genesis, gives fields of application and enumerates related distributions together with bibliographical notes. For more details the reader is referred to the two volumes of JOHNSON/KOTZ/BALAKRISHNAN (1994, 1995) on continuous univariate distributions.

- Then we will specify the functions and parameters listed in Tab. 1/1 for the non–reduced version of the distribution.[1] Functions and parameters of the reduced version easily follow by setting $a$ and $b$ to zero and to one, respectively. We will also say by which percentiles and how to estimate the location and scale parameters which may be read off using a line fitted to the data on the probability paper. When this line has been estimated by one of the linear procedures of this book the read–off parameter values will be identical to those of the linear estimation procedure up to minor errors of reading off.

- We will give a chart displaying the DF $f_Y(y)$, the CDF $F_Y(y)$, the hazard function $h_Y(y)$ and the cumulated hazard function $H_Y(y)$ for the reduced variate $Y = (X - a)/b$.

- When there are closed form expressions for the DF, CDF and moments of the order statistics, they will follow. We will also give hints to tables and special recurrence relations and identities pertaining to these moments.

- We will give the percentile function $F_Y^{-1}(P)$ and — when LEPP uses an approximation for the moments of order statistics — its first six derivatives.

- For the processing of grouped data as well as of multiply or progressively censored data we will give the reduced DF at $\widehat{\alpha}_j$, i.e. $f_Y(\widehat{\alpha}_j)$ where $\widehat{\alpha}_j = F_Y^{-1}(\widehat{P}_j)$.

- We we will display the probability paper pertaining to that distribution together with a Monte Carlo generated data set and the estimated regression line.

[1] We only give the characteristic function and/or the crude moment generating function, because the remaining generating functions simply depend on those two functions.

- For grouped data and randomly or multiply censored data the linear estimation procedure implemented in LEPP is always based on (4.54a,e) and (4.56a,b). But when the data input is a set of order statistics the linear estimation process in LEPP uses different approaches depending on the respective distribution. We therefore indicate which method of Sect. 4.2 has been used for that distribution.

## 5.2 The case of genuine location–scale distributions

This section describes 27 distributions which originally come in location–scale form. We hope to not have omitted any important member of this family.

### 5.2.1 Arc–sine distribution — $X \sim AS(a, b)$

The **arc–sine distribution**[2] is a special case of the **beta distribution** (1.12a) having shape parameters $c = d = 0.5$. A beta distribution with $c + d = 1$, but $c \neq 0.5$, is sometimes called a generalized arc–sine distribution. The name is derived from the fact that the CDF (5.1b) is written in terms of the arc–sine function, the inverse of the sine function: $\arcsin(x) = \sin^{-1}(x)$. The arc–sine distribution describes the location, velocity and related attributes at random time of a particle in simple harmonic motion. The arc–sine distribution with parameter $a = 0$ and $b > 0$ having support $[-b,\ b]$ gives the position at random time[3] of a particle engaged in simple harmonic motion with amplitude $b > 0$.

A random variable $X$ is arc–sine distributed with parameters $a \in \mathbb{R}$ and $b > 0$, denoted $X \sim AS(a, b)$, when its DF is given by

$$f(x|a,b) = \frac{1}{b\,\pi\sqrt{1-\left(\dfrac{x-a}{b}\right)^2}}\,,\quad a-b<x<a+b. \tag{5.1a}$$

$f(x|a,b)$ is symmetric around $a$ and U–shaped. There is an antimode $\big($= minimum of $f(x|a,b)\big)$ at $x = a$ and $f(x|a,b) \to \infty$ for $x \to a \pm b$.

$$F(x|a,b) = \frac{\pi + 2\,\arcsin\left(\dfrac{x-a}{b}\right)}{2\,\pi} = \frac{1}{2} + \frac{\arcsin\left(\dfrac{x-a}{b}\right)}{\pi} \tag{5.1b}$$

$$R(x|a,b) = \frac{\pi - 2\,\arcsin\left(\dfrac{x-a}{b}\right)}{2\,\pi} = \frac{1}{2} - \frac{\arcsin\left(\dfrac{x-a}{b}\right)}{\pi} \tag{5.1c}$$

$$h(x|a,b) = \frac{2}{b\,\pi\sqrt{1-\left(\dfrac{x-a}{b}\right)^2}\left[\pi - 2\,\arcsin\left(\dfrac{x-a}{b}\right)\right]} \tag{5.1d}$$

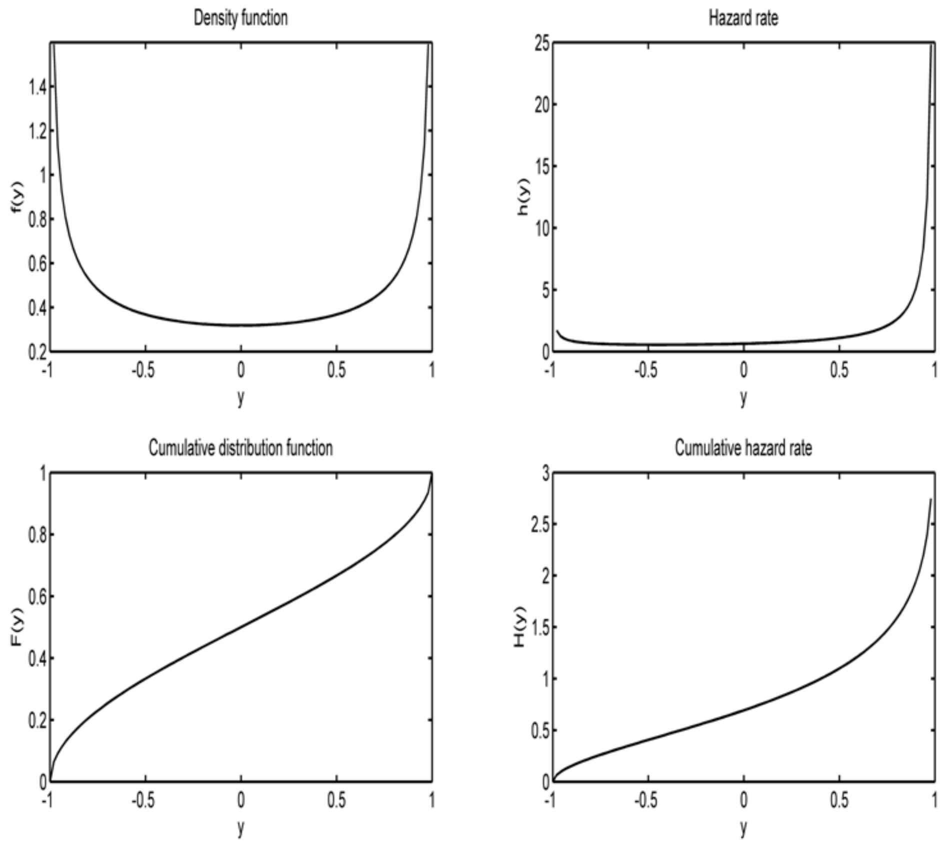
[2] Suggested reading for this section: ARNOLD/GROENEVELD (1980), NORTON (1975, 1978, 1982), SHANTARAM (1978, 1981), KEMPERMAN/SKIBINSKY (1982).

[3] "Random time" means that the time of observation is independent of the initial angle, $0 \leq \theta_0 \leq 2\,\pi$.

The hazard function $h(x)$ has a bathtub shape with a minimum at $x \approx a - 0.4421\, b$.

$$H(x|a,b) \;=\; \ln(2\,\pi) - \ln\left[\pi - 2\,\arcsin\left(\frac{x-a}{b}\right)\right] \tag{5.1e}$$

Figure 5/1: Several functions for the reduced arc–sine distribution




$$F_X^{-1}(P) \;=\; x_P \;=\; a + b\,\sin[\pi\,(P-0.5)],\;\; 0 \le P \le 1 \tag{5.1f}$$

$$a \;=\; x_{0.5} \tag{5.1g}$$

$$b \;=\; x_{0.\overline{6}} - x_{0.\overline{3}} \tag{5.1h}$$

$$C_X(t) \;=\; \exp(i\,t\,a)\sum_{k=0}^{\infty}(-1)^k\left(\frac{b\,t}{2}\right)2\,k \Big/ (k!)^2 \tag{5.1i}$$

$$\mu'_r(Y) \;=\; \mathrm{E}\big(Y^r\big) \;=\; \left\{\begin{array}{cl} \binom{r}{r/2}\Big/2^r & \text{for } r = 2\,k;\; k = 1,2,\ldots \\ 0 & \text{for } r = 2\,k+1;\; k = 0,1,\ldots \end{array}\right\} \tag{5.1j}$$

$$\mu'_r(X) \;=\; \mathrm{E}\big[(a+b\,Y)^r\big] \;=\; \sum_{j=0}^{r}\binom{r}{j} b^{r-j}\,a^j\,\mu'_{r-j}(Y) \tag{5.1k}$$

$$\mu_1'(X) = \mathrm{E}(X) = a \tag{5.1l}$$

$$\mu_2'(X) = \mathrm{E}(X^2) = \frac{b^2}{2} + a^2 \tag{5.1m}$$

$$\mathrm{Var}(X) = \frac{b^2}{2} \tag{5.1n}$$

$$\alpha_3 = 0 \tag{5.1o}$$

$$\alpha_4 = 1.5 \tag{5.1p}$$

$$I(X) = \mathrm{ld}\, b + \mathrm{ld}(\pi/2) \approx \mathrm{ld}\, b + 0.6515 \tag{5.1q}$$

$$F_Y^{-1}(P) = y_P = \sin[\pi\,(P-0.5)],\ \ 0 \le P \le 1 \tag{5.1r}$$

$$f_Y(y_P) = \frac{1}{\pi\,\sqrt{1-\{\sin[\pi\,(P-0.5)]\}^2}} \tag{5.1s}$$

In LEPP linear estimation using order statistics is realized with BLOM's unbiased, nearly best linear estimator. The means of the reduced order statistics are evaluated by numerical integration.

**Example 5/1: OLS and GLS with a sample of size $n = 10$ from an arc–sine distribution**

The following ten observations are order statistics in an uncensored sample of size $n = 10$ from an arc–sine distribution having $a = 1$ and $b = 2$:

| $r$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_{r:10}$ | −0.9400 | −0.9311 | −0.7712 | −0.5045 | 0.6418 | 0.6905 | 1.2403 | 2.7080 | 2.9263 | 2.9999 |

The correlation coefficient of the data and the plotting positions (here: means of the reduced order statistics) is $r = 0.9762$, indicating a good fit. Fig. 5/2 shows on the left–hand side the data together with the OLS-estimated regression line on arc–sine probability paper. The estimates of the parameters and their estimated variance matrix are

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \begin{pmatrix} 0.8056 \\ 2.3476 \end{pmatrix} \qquad \widehat{\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\big)} = \begin{pmatrix} 0.5611 & 0 \\ 0 & 1.3598 \end{pmatrix}.$$

The right–hand side of Fig. 5/2 shows the same data, but with the unbiased nearly best linear BLOM–estimated regression line. The estimates of the parameters and their estimated variance matrix are

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \begin{pmatrix} 1.0442 \\ 2.0738 \end{pmatrix} \qquad \widehat{\mathrm{Var}\big(\widehat{\boldsymbol{\theta}}\big)} = \begin{pmatrix} 0.0102 & 0 \\ 0 & 0.0108 \end{pmatrix}.$$

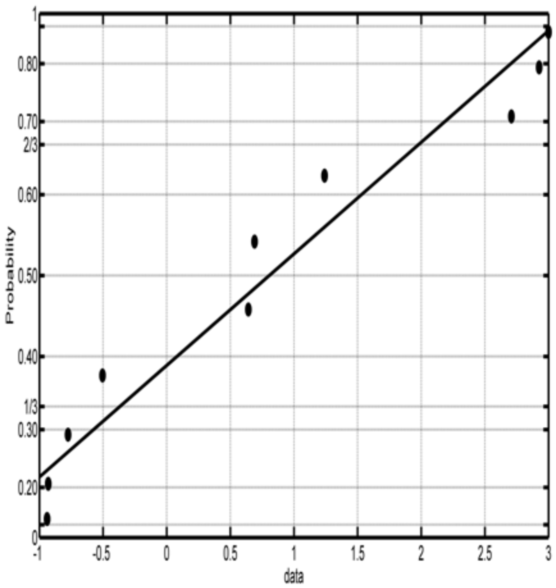
The BLOM-estimates, being GLS-like, are closer to the true parameter values than the OLS–estimates. Furthermore, OLS severely overestimates the variances. Looking at both pictures in Fig. 5/2 the OLS-line gives the better eye–fit to the data.

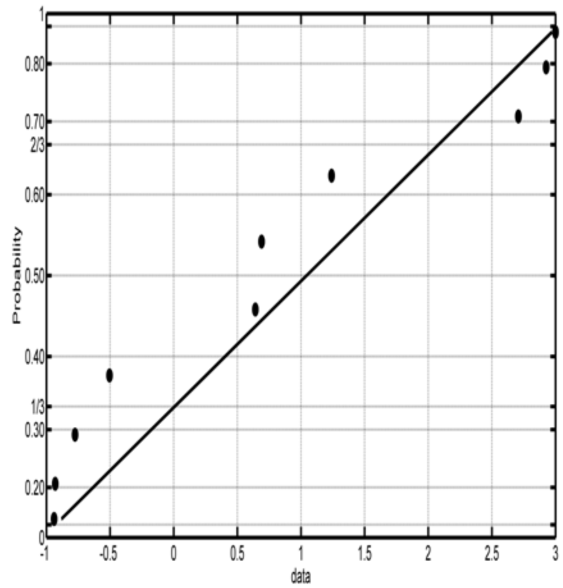
If we have had to draw an eye–fitted line to the data we surely would have come out with a line similar to the OLS–line, because we intuitively give equal weights to the deviations. GLS takes into account the fact of variance–heterogeneity of order statistics and weighs down those deviations which belong to an order statistic with high variance. The GLS–approach consists in minimizing the generalized variance

$$(\boldsymbol{x} - \widehat{a}\,\mathbf{1} - \widehat{b}\,\boldsymbol{\alpha})'\boldsymbol{B}^{-1}(\boldsymbol{x} - \widehat{a}\,\mathbf{1} - \widehat{b}\,\boldsymbol{\alpha}),$$

i.e. the deviations $(\boldsymbol{x} - \widehat{a}\,\mathbf{1} - \widehat{b}\,\boldsymbol{\alpha})$ are weighted with the inverse of the variance matrix. Thus, a GLS–line normally is no good eye–fitted line.

Figure 5/2: Arc–sine probability paper with data and OLS–line (left) and GLS–line (right)




## 5.2.2 CAUCHY distribution — $X \sim CA(a, b)$

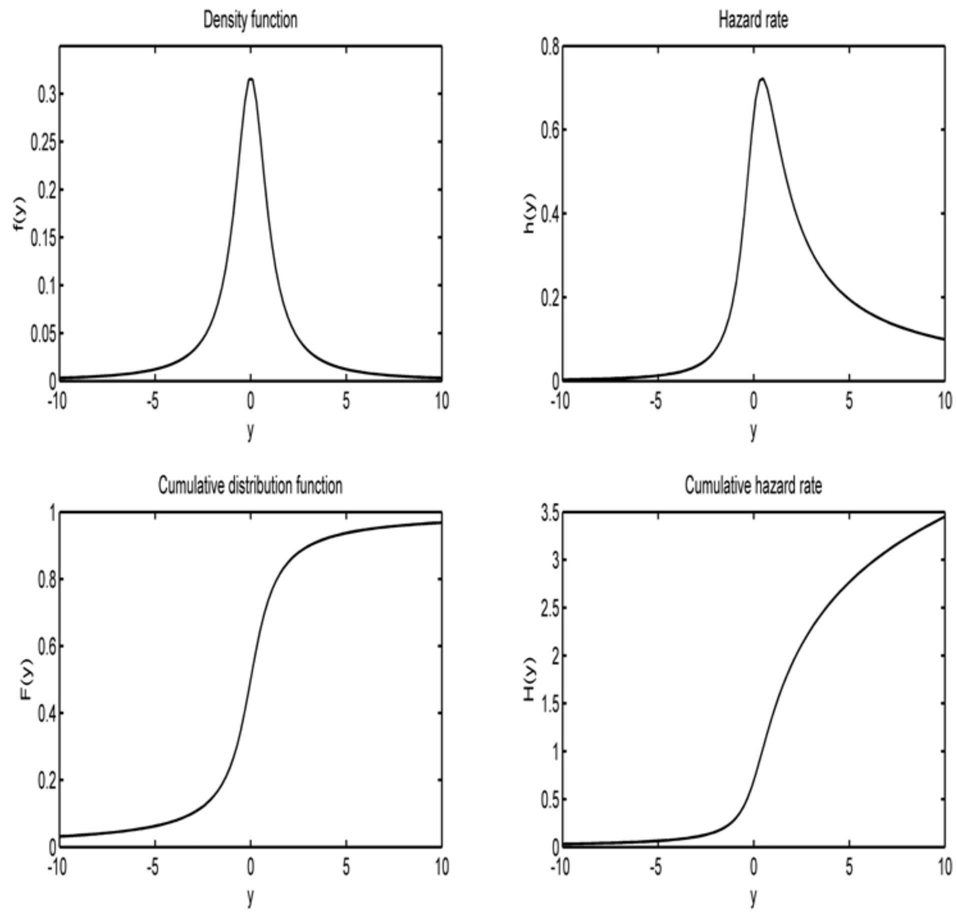
The **CAUCHY distribution**, named after the French mathematician AUGUSTIN CAUCHY (1789 –1857), is the distribution of the ratio of two independent normal variates, more precisely: if $Y_1 \sim NO(0,1)$ and $Y_2 \sim NO(0,1)$, both independent, then $Y = Y_1\big/Y_2 \sim CA(0,1)$. Of course, the reciprocal $Y_2\big/Y_1$ is CAUCHY distributed, too. However, it should be noted that the common distribution of $Y_1$ and $Y_2$ need not be normal. Among physicists this distribution is known as LORENTZ distribution (HENDRIK ANTOON LORENTZ, 1853 – 1928, Nobel Prize winner for physics in 1902) or BREIT–WIGNER function. In physics it is of importance due to being the solution of the differential equation describing forced resonance. It also describes the line–shape of spectral lines which are subject to homogeneous broadening in which all atoms interact in the same way with the frequency range contained in the line–shape. For more details on the genesis and applications of the CAUCHY distribution the interested reader is referred to STIGLER (1974).

The reduced CAUCHY distribution $CA(0,1)$ is identical to the **central *t*–distribution** with one degree of freedom. The **generalized CAUCHY distribution** with DF

$$\left.\begin{aligned} f(x|a,b,d,m) = \; & \frac{m\,\Gamma(d)}{2\,b\,\Gamma\big(m^{-1}\big)\,\Gamma\big(d-m^{-1}\big)}\left[1+\left|\frac{x-a}{b}\right|^m\right]^{-d}; \\ & x \in \mathbb{R};\; b,d,m>0;\; d \geq m^{-1}; \end{aligned}\right\} \tag{5.2}$$

results into the simple CAUCHY distribution with DF (5.3a) for $m = 2$ and $d = 1$. For $m = 2,\ a = 0$ we have the DF of $\big[b\,(2\,d-1)^{-0.5}\big]$ times a central $t$–variable with $2\,(d-1)$ degrees of freedom.

Figure 5/3: Several functions for the reduced CAUCHY distribution




A variate $X$ is CAUCHY distributed with parameters $a \in \mathbb{R}$ and $b > 0$, denoted $CA(a,b)$, when its DF is given by

$$f(x|a,b) = \left\{\pi\,b\left[1+\left(\frac{x-a}{b}\right)^2\right]\right\}^{-1},\ x \in \mathbb{R}. \tag{5.3a}$$

$f(x|a,b)$ is symmetric around $a$ and bell–shaped like the normal distribution, but it has thicker tails than a normal distribution. That is why the CAUCHY distribution has no finite moments of order $r > 1$, because $\int\limits_{-\infty}^{\infty} x^r\, f(x|a,b)\,\mathrm{d}x$ does not converge on $\mathbb{R}$.

$$F(x|a,b) = \frac{1}{2} + \frac{1}{\pi}\,\arctan\left(\frac{x-a}{b}\right) \tag{5.3b}$$

$$R(x|a,b) = \frac{1}{2} - \frac{1}{\pi}\,\arctan\left(\frac{x-a}{b}\right) \tag{5.3c}$$

$$h(x|a,b) = \left\{ b\left[1 + \left(\frac{x-a}{b}\right)^2\right]\left[\frac{\pi}{2} - \arctan\left(\frac{x-a}{b}\right)\right]\right\}^{-1} \tag{5.3d}$$

The hazard rate has an inverted bath–tub shape with a maximum at $x \approx a + 0.4290\,b$.

$$H(x|a,b) = \ln(2\,\pi) - \ln\left[\pi - 2\,\arctan\left(\frac{x-a}{b}\right)\right] \tag{5.3e}$$

$$F_X^{-1}(P) = x_P \;=\; a + b\,\tan\big[\pi\,(P-0.5)\big] \tag{5.3f}$$

$$a = x_{0.5} \tag{5.3g}$$

$$b = \big(x_{0.75} - x_{0.25}\big)/2 \tag{5.3h}$$

$$x_M = x_{0.5} \;=\; a \tag{5.3i}$$

$$C_X(t) = \exp\big(i\,t\,a - |t|\,b\big), \quad i = \sqrt{-1} \tag{5.3j}$$

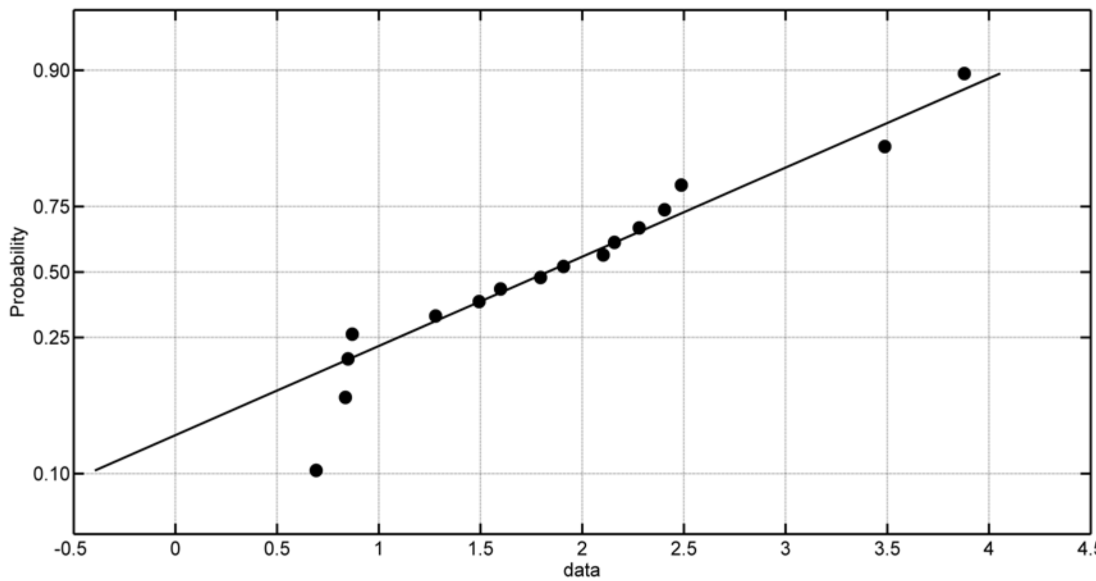
$$I(X) = \mathrm{ld}\,b + \mathrm{ld}(4\,\pi) \;\approx\; \mathrm{ld}\,b + 3.6515 \tag{5.3k}$$

$$F_Y^{-1}(P) = y_P \;=\; \tan[\pi\,(P-0.5)], \quad 0 < P < 1 \tag{5.3l}$$

$$f_Y(y_P) = \Big\{\pi\big(1 + \tan^2\big[\pi\,(P-0.5)\big]\big)\Big\}^{-1} \tag{5.3m}$$

$$\left.\begin{aligned}
F_Y^{-1(1)}(P) &= \pi\,\sec^2[\pi\,(P-0.5)]\\
F_Y^{-1(2)}(P) &= 2\,\pi^2\,\sec^2[\pi\,(P-0.5)]\,\tan[\pi\,(P-0.5)]\\
F_Y^{-1(3)}(P) &= -2\,\pi^3\,\sec^4[\pi\,(P-0.5)]\,\big\{\cos[2\,\pi\,(P-0.5)] - 2\big\}\\
F_Y^{-1(4)}(P) &= -2\,\pi^4\,\sec^5[\pi\,(P-0.5)]\,\big\{\sin[3\,\pi\,(P-0.5)]\\
&\quad - 11\,\sin[\pi\,(P-0.5)]\big\}\\
F_Y^{-1(5)}(P) &= 2\,\pi^5\,\sec^6[\pi\,(P-0.5)]\big\{33 - 26\,\cos[2\,\pi\,(P-0.5)]\\
&\quad + \cos[4\,\pi\,(P-0.5)]\big\}\\
F_Y^{-1(6)}(P) &= 2\,\pi^6\,\sec^7[\pi\,(P-0.5)]\big\{302\,\sin[\pi\,(P-0.5)] + \sin[5\,\pi\,(P-0.5)]\\
&\quad + 57\,\sin[3\,\pi\,(P-0.5)]\big\}
\end{aligned}\right\} \tag{5.3n}$$

Figure 5/4: CAUCHY probability paper with data and regression line




Since the reduced CAUCHY variate does not possess finite moments of order $r > 1$ the expected values of $Y_{1:n}$ and $Y_{n:n}$ and the variances of $Y_{1:n},\ Y_{2:n},\ Y_{n-1:n}$ and $Y_{n:n}$ are infinite. Explicit expressions involving infinite series have been given by VAUGHAN (1994). BARNETT (1966) has computed the means, variances and covariances of reduced order statistics for sample sizes $n = 5(1)16(2)20$ based on the following formulas which result from the general formulas (2.9a) and (2.11a) after a transformation of the variable:

$$\alpha_{r:n} = \frac{n!}{\pi^n\,(r-1)!\,(n-r)!}\int\limits_{-\pi/2}^{\pi/2}\Big(\frac{\pi}{2}+v\Big)^{r-1}\Big(\frac{\pi}{2}-v\Big)^{n-r}\tan v\,\mathrm{d}v \tag{5.3o}$$

$$\alpha_{r,r:n} = \frac{n!}{\pi^n\,(r-1)!\,(n-r)!}\int\limits_{-\pi/2}^{\pi/2}\Big(\frac{\pi}{2}+v\Big)^{r-1}\Big(\frac{\pi}{2}-v\Big)^{n-r}\tan^2 v\,\mathrm{d}v \tag{5.3p}$$

$$\alpha_{r,s:n} = \frac{n!}{\pi^n\,(r-1)!\,(s-r-1)!\,(n-r)!}\int\limits_{-\pi/2}^{\pi/2}\int\limits_{-\pi/2}^{w}\tan v\,\tan w\ \times$$

$$\Big(\frac{\pi}{2}+v\Big)^{r-1}(w-v)^{s-r-1}\Big(\frac{\pi}{2}-w\Big)^{n-s}\mathrm{d}v\,\mathrm{d}w,\ r<s. \tag{5.3q}$$

(5.3o) can be transformed to

$$\alpha_{r,r:n} = \frac{n}{\pi}\big(\alpha_{r:n-1}-\alpha_{r-1:n-1}\big)-1$$

so that the variance is

$$\beta_{r,r:n} = \frac{n}{\pi}\big(\alpha_{r:n-1}-\alpha_{r-1:n-1}\big)-1-\alpha^2_{r:n}. \tag{5.3r}$$

Computation of the means by (5.3o) causes no numerical difficulties as well as the computation of the variances which are obtained from the means according to (5.3r). But there are considerable difficulties in evaluating (5.3q), see BARNETT (1966) for resolving them. In LEPP linear estimation with order statistics is realized as follows:

- for $7 \leq n \leq 20$ LLOYD's estimator with means computed according to (5.3o) and tabulated variances and covariances taken from BARNETT (1966),
- for $n > 20$ LLOYD's estimator with means computed according to (5.3o) and approximated variance–covariance matrix.

### 5.2.3 Cosine distributions

The cosine function $y = \cos x$ describes a wave with constant length of $2\,\pi$ and a height of $2$, measured from the bottom to the top. When we pick out that piece where $y \geq 0$, i.e. the interval $-\pi/2 \leq x \leq \pi/2$ — marked bold in Fig. 5/5 —, normalize the enclosed area to unity and rename the variables, we arrive at the rather simple looking DF

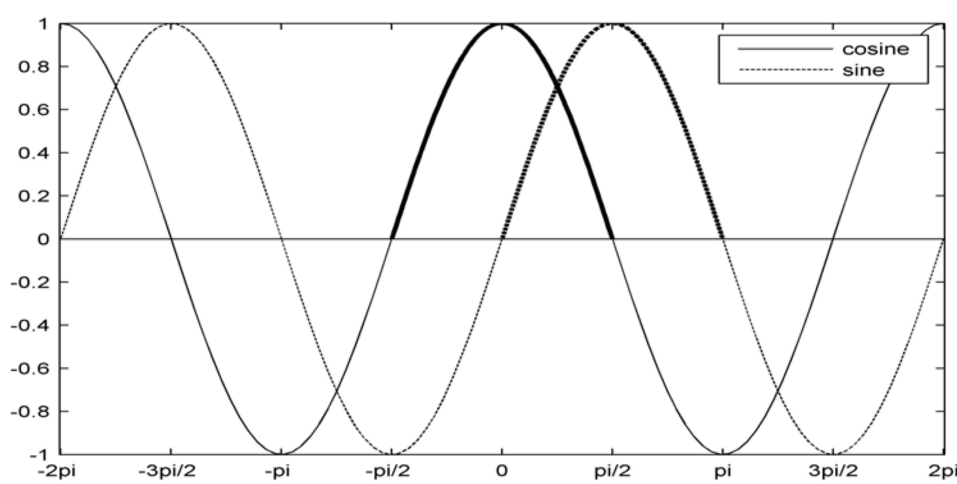
$$f(y) = \frac{1}{2}\cos y, \;\; -\pi/1 \leq y \leq \pi/2. \tag{5.4}$$

This is the **reduced form of the simple** or **ordinary cosine distribution**. In class–room teaching it often serves as an object of demonstrating the handling of continuous distributions. The ordinary cosine distribution is convex over $[-\pi/2, \pi/2]$ and thus is no good approximation to the bell–shaped normal distribution. We find a bell–shaped DF of cosine–type when we modify (5.4) to the so–called **raised cosine distribution**. Its reduced form reads

$$f(y) = \frac{1}{2}\big[1 + \cos(\pi\,y)\big], \;\; -1 \leq y \leq 1. \tag{5.5}$$

The sine function and the cosine function differ by a horizontal shift of $\pi/2$ — see Fig. 5/5 — so that we can equally express (5.4) and (5.5) in terms of the sine function.

Figure 5/5: Cosine and sine functions



#### 5.2.3.1 Ordinary cosine distribution — $X \sim COO(a, b)$

We list the following results for an ordinary cosine distribution.

$$f(x|a,b) = \frac{1}{2\,b}\cos\left(\frac{x-a}{b}\right), \quad a-b\,\frac{\pi}{2} \le x \le a+b\,\frac{\pi}{2},\ a \in \mathbb{R},\ b>0 \tag{5.6a}$$

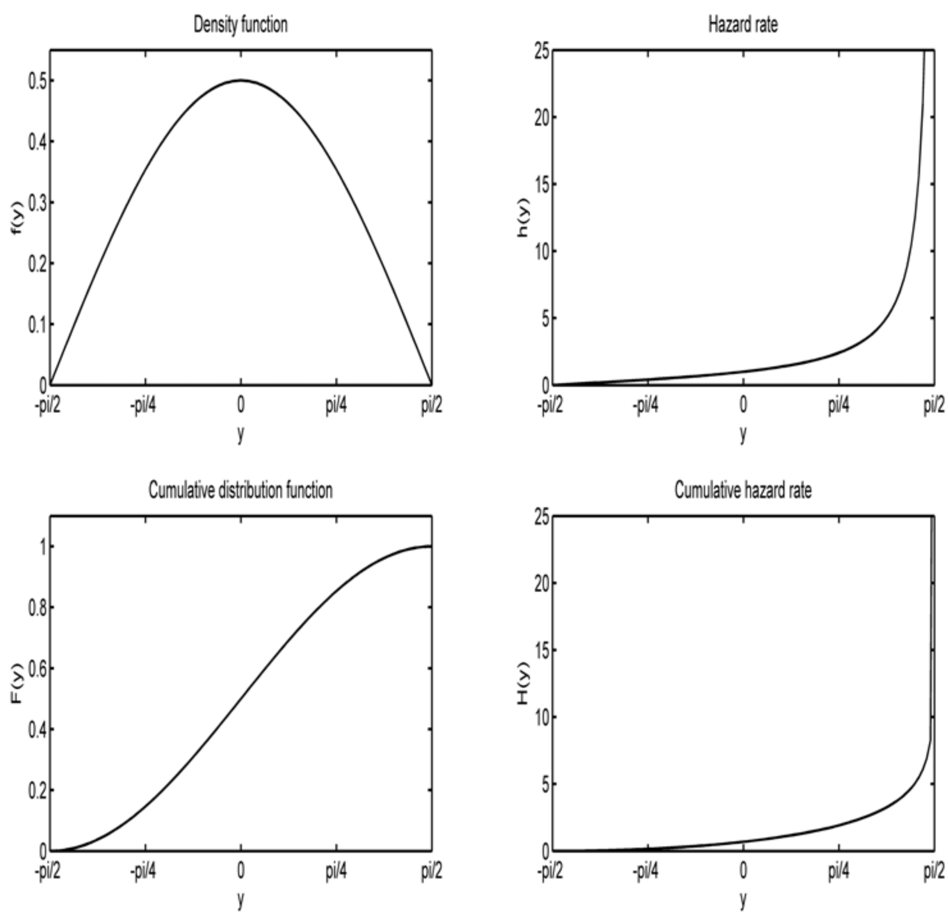
$$F(x|a,b) = 0.5\left[1+\sin\left(\frac{x-a}{b}\right)\right] \tag{5.6b}$$

$$R(x|a,b) = 0.5\left[1-\sin\left(\frac{x-a}{b}\right)\right] \tag{5.6c}$$

$$h(x|a,b) = \frac{1}{b\left[\sec\left(\frac{x-a}{b}\right)-\tan\left(\frac{x-a}{b}\right)\right]} \tag{5.6d}$$

$$H(x|a,b) = \ln 2 - 2\,\ln\left[\cos\left(\frac{x-a}{b}\right)-\sin\left(\frac{x-a}{b}\right)\right] \tag{5.6e}$$

Figure 5/6: Several functions for the reduced ordinary cosine distribution

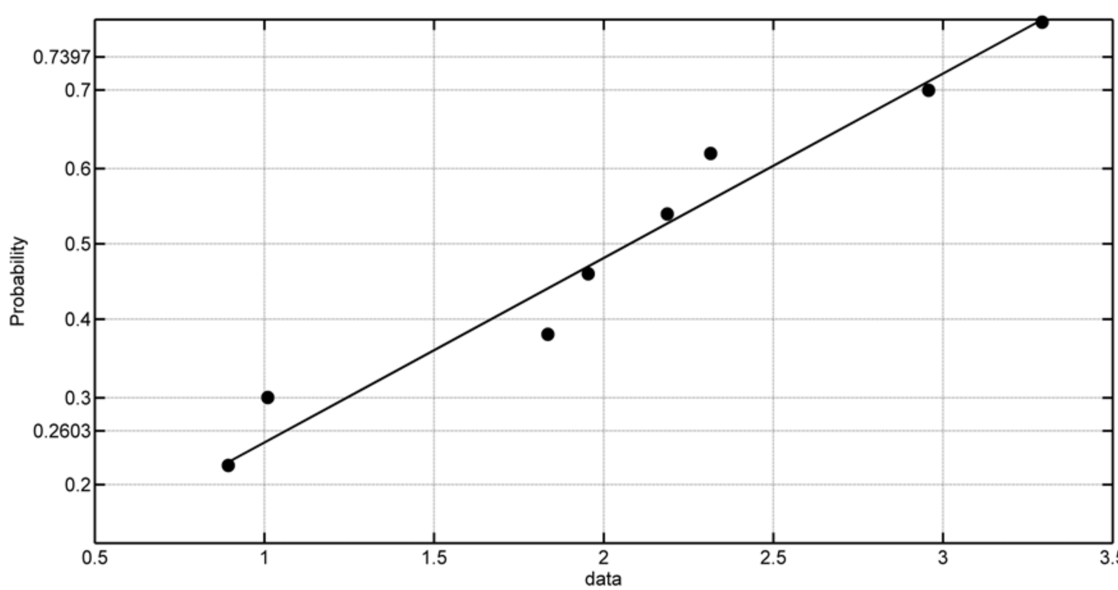


$$\begin{aligned}
F_X^{-1}(P) &= x_P \;=\; a + b\,\arcsin\bigl[2\,(P-0.5)\bigr],\;\; 0 \le P \le 1 && \text{(5.6f)}\\
a &= x_{0.5} && \text{(5.6g)}\\
b &\approx x_{0.7397} - x_{0.2603} && \text{(5.6h)}\\
x_M &= x_{0.5} \;=\; a && \text{(5.6i)}\\
M_X(t) &= \frac{\exp\bigl(a\,t - \pi\,b\,t/2\bigr)\bigl[1+\exp\bigl(\pi\,b\,t\bigr)\bigr]}{2\bigl(1+b^2\,t^2\bigr)} && \text{(5.6j)}\\
C_X(t) &= \frac{\exp\bigl[i\,\bigl(a\,t - \pi\,b\,t/2\bigr)\bigr]\bigl[1+\exp\bigl(i\,\pi\,b\,t\bigr)\bigr]}{2\bigl(1-b^2\,t^2\bigr)},\quad i=\sqrt{-1} && \text{(5.6k)}\\
\mu_r'(Y) &= \frac{\pi^{r+1}}{1+r}\Bigl\{2^{-r-2}\bigl[1+(-1)^r\bigr]\;{}_1F_3(\alpha;\boldsymbol{\beta};z)\Bigr\} && \text{(5.6l)}
\end{aligned}$$

${}_1F_3(\alpha;\boldsymbol{\beta};z)$ is a **generalized hypergeometric function** with $\alpha = 0.5\,(1+r)$, $\boldsymbol{\beta} = (0.5, 1.5, 0.5\,r)$, $z = -\pi^2\big/16$, see ABRAMOWITZ/STEGUN (1965).

$$\begin{aligned}
\mu_1'(X) &= \mathrm{E}(X) \;=\; a && \text{(5.6m)}\\
\mu_2'(X) &= a^2 + \frac{b^2}{4}\,(\pi^2-8) && \text{(5.6n)}\\
\mu_3'(X) &= a^3 + 0.75\,a\,b^2\,(\pi^2-8) && \text{(5.6o)}\\
\mu_4'(X) &= a^4 + 1.5\,a^2\,b^2(\pi^2-8) + \frac{b^4}{16}\,(384 - 48\,\pi^2 + \pi^4) && \text{(5.6p)}
\end{aligned}$$

Figure 5/7: Ordinary cosine probability paper with data and regression line



$$\operatorname{Var}(X) = \frac{b^2\,\pi^2}{4} - 2\,b^2 \approx 0.4674\,b^2 \tag{5.6q}$$

$$\alpha_3 = 0 \tag{5.6r}$$

$$\alpha_4 \approx 2.1938 \tag{5.6s}$$

$$I(X) = \frac{1+\ln b}{\ln 2} \approx 1.4427\,(1+\ln b) \tag{5.6t}$$

$$F_Y^{-1}(P) = y_P = \arcsin\big[2\,(P-0.5)\big],\;\; 0 \le P \le 1 \tag{5.6u}$$

$$f_Y(y_P) = \sqrt{P\,(1-P)} \tag{5.6v}$$

In LEPP linear estimation for order statistics input is realized with LLOYD's estimator using computed means and variance–covariance matrix.

#### 5.2.3.2 Raised cosine distribution — $X \sim COR(a,b)$

We give the following description and features for a raised cosine distribution.

$$f(x|a,b) = \frac{1}{2\,b}\left[1+\cos\left(\pi\,\frac{x-a}{b}\right)\right],\quad a-b \le x \le a+b,\; a\in\mathbb{R},\; b>0 \tag{5.7a}$$

$$F(x|a,b) = \frac{1}{2}\left[1+\frac{x-a}{b}+\frac{1}{\pi}\,\sin\left(\pi\,\frac{x-a}{b}\right)\right] \tag{5.7b}$$

$$R(x|a,b) = \frac{1}{2}\left[1-\frac{x-a}{b}-\frac{1}{\pi}\,\sin\left(\pi\,\frac{x-a}{b}\right)\right] \tag{5.7c}$$

$$h(x|a,b) = \frac{1+\cos\left(\pi\,\dfrac{x-a}{b}\right)}{b\left[1-\dfrac{x-a}{b}-\dfrac{1}{\pi}\,\sin\left(\pi\,\dfrac{x-a}{b}\right)\right]} \tag{5.7d}$$

$$H(x|a,b) = \ln(2\,\pi) - \ln\left[\pi\,\left(\frac{x-a}{b}-1\right)+\sin\left(\pi\,\frac{x-a}{b}\right)\right] \tag{5.7e}$$

The percentiles cannot be given in explicit form. For the reduced variate the percentile $y_P$ is the solution of

$$\pi\big(2\,P-1\big) = \pi\,y_P + \sin(\pi\,y_P).$$

We have OLS–fitted the percentile function for the reduced distribution by a polynomial of degree two, where the regressor is

$$t = \sqrt{-2\,\ln P},$$

and found

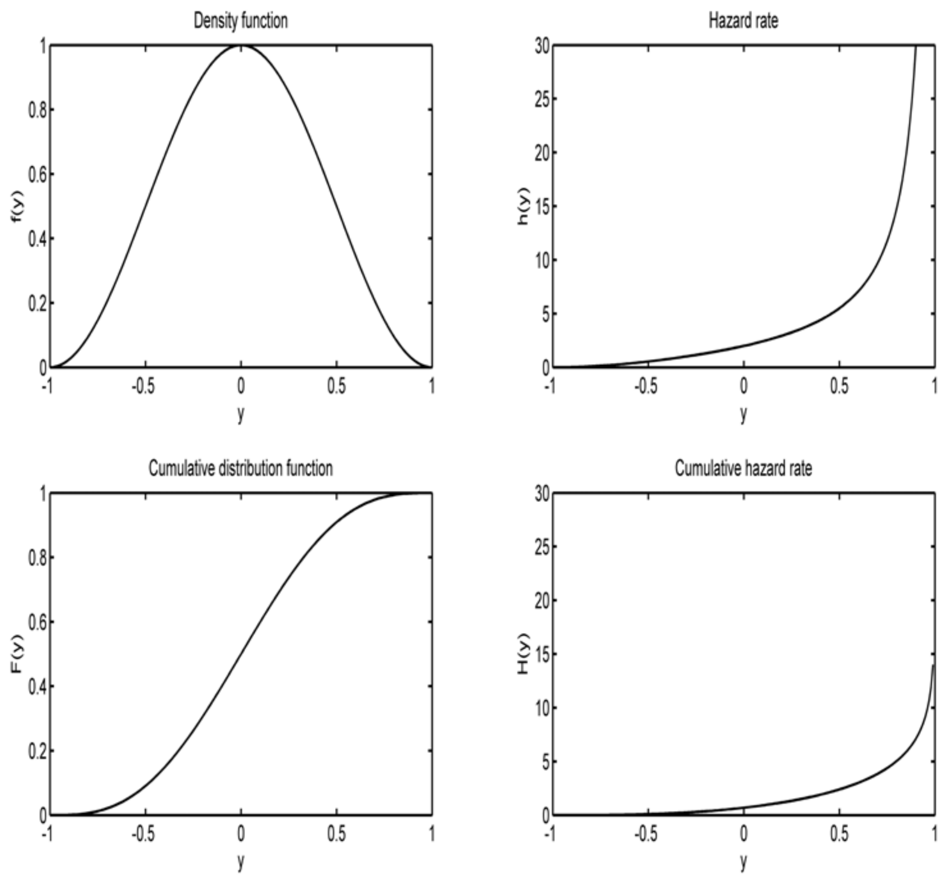
$$y_P \approx 0.820251 - 0.806770\,t + 0.093247\,t^2,\quad 0 < P \le 0.5. \tag{5.7f}$$

The absolute error of this approximation is less than $3 \cdot 10^{-3}$. For $P > 0.5$ the symmetry relation

$$y_{1-P} = -y_P, \quad 0 < P \leq 0.5,$$

holds.

Figure 5/8: Several functions for the reduced raised cosine distribution




$$\begin{aligned}
F_X^{-1}(P) &= x_P = a + b\,y_P, \quad 0 \leq P \leq 1 && (5.7g)\\
a &= x_{0.5} && (5.7h)\\
b &\approx x_{0.9092} - x_{0.0908} && (5.7i)\\
x_M &= x_{0.5} = a && (5.7j)\\
M_X(t) &= \exp(a\,t)\,\frac{\pi^2 \sinh(b\,t)}{b\,t\,(\pi^2 + b^2\,t^2)} && (5.7k)\\
C_X(t) &= \exp(i\,a\,t)\,\frac{\pi^2 \sinh(b\,t)}{b\,t\,(\pi^2 - b^2\,t^2)}, \quad i = \sqrt{-1} && (5.7l)\\
\mu_r'(Y) &= \frac{1}{r+1} + \frac{1}{1+2\,r}\,{}_1F_2\left(r + 0.5; 0.5, r + 1.5; -\frac{\pi^2}{4}\right) && (5.7m)
\end{aligned}$$

${}_1F_2(\cdot)$ is a generalized hypergeometric function.

$$\mu_1'(X) = \mathrm{E}(X) = a \tag{5.7n}$$

$$\mu_2'(X) = a^2 + b^2\,\frac{\pi^2-6}{3\,\pi^2} \tag{5.7o}$$

$$\mu_3'(X) = a^3 + a\,b^3\left(1-\frac{6}{\pi^2}\right) \tag{5.7p}$$

$$\mu_4'(X) = a^4 + \frac{2\,a^2\,b^2\,(\pi^2-6)}{\pi^2} + \frac{b^4\,(120-20\,\pi^2+\pi^4)}{5\,\pi^4} \tag{5.7q}$$

$$\mathrm{Var}(X) = b^2\left(\frac{1}{3}-\frac{6}{\pi^2}\right) \tag{5.7r}$$

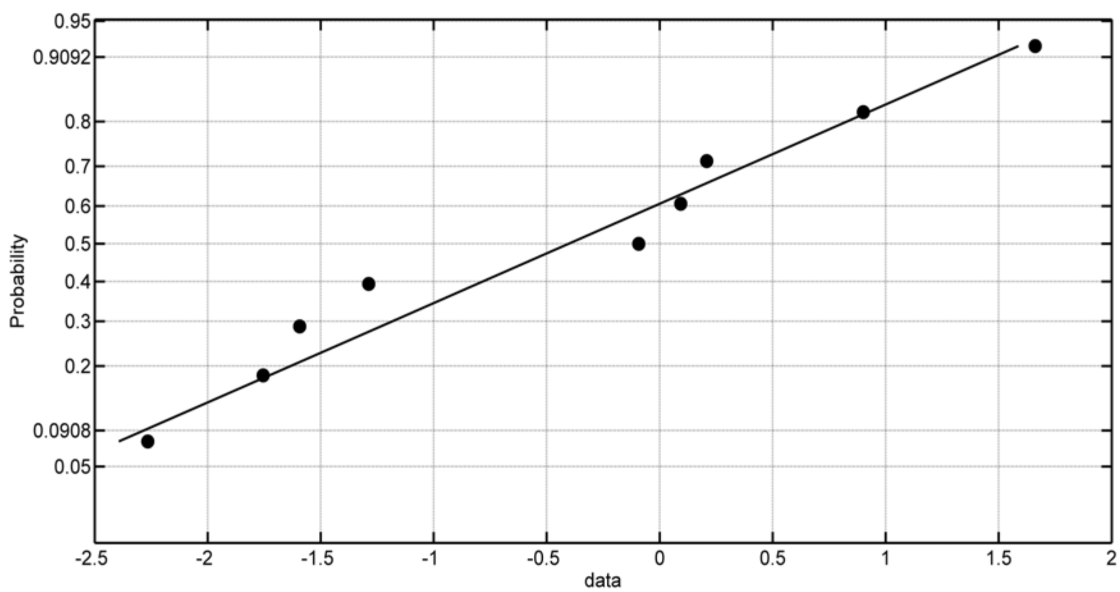
$$\alpha_3 = 0 \tag{5.7s}$$

$$\alpha_4 \approx 2.4062 \tag{5.7t}$$

$$I(X) \approx 0.5573 + 1.4427\,\ln b \tag{5.7u}$$

In LEPP linear estimation for order statistics input is realized with LLOYD's estimator using computed means and variance–covariance matrix.

Figure 5/9: Raised cosine probability paper with data and regression line




## 5.2.4 Exponential distribution — $X \sim EX(a,b)$

The **exponential distribution**[4] gives the probability for the distance between successive events in a POISSON process. It is widely used as a lifetime distribution and as such it is

[4] Suggested reading for this section: JOHNSON/KOTZ/BALAKRISHNAN (1994, Chapter 19), BALAKRISHNAN/RAO (1998b, Chapters 1 and 2).

characterized by the **property of no–aging**, i.e. the future lifetime of an individual aged $x > 0$ has the same distribution of a new–born individual. It is further characterized by a constant hazard function, see (5.8d). The exponential distribution is related to a great number of other distributions.

- It is a special case of the **WEIBULL distribution**, see Sect. 5.3.2.4, having shape parameter $c = 1$.
- If $Y = \exp(-X)$ has an exponential distribution with a value of zero for the location parameter then $X$ has a **extreme value distribution of type I for the maximum**, see Sect. 5.2.5.1.
- The **double** or **bilateral exponential distribution**, also known as **LAPLACE distribution**, see Sect. 5.2.8, is a combination of the exponential and reflected exponential distributions.
- The **central $\chi^2$–distribution** $\chi^2(\nu)$ with $\nu = 2$ degrees of freedom is identical to $EX(0, 2)$.
- When $U$ has a reduced **uniform distribution**, i.e. $U \sim UN(0, 1)$, then $X = -\ln U \sim \chi^2(2) = EX(0, 2)$.
- The exponential distribution with $a = 0$ is a special case of the **gamma distribution** having shape parameter $c = 1$.
- The logarithm of a **PARETO variable** is exponentially distributed, see Sect. 5.3.4.
- When $X$ has **power–function distribution** then $V = -\ln X$ has an exponential distribution, see Sect. 5.3.5.
- When $X$ is exponentially distributed then $V = -X$ has a **reflected exponential distribution**, see Sect. 5.2.14.

$$f(x|a,b) = \frac{1}{b}\exp\left(-\frac{x-a}{b}\right), \quad x \ge a,\ a \in \mathbb{R},\ b > 0 \tag{5.8a}$$

$$F(x|a,b) = 1 - \exp\left(-\frac{x-a}{b}\right) \tag{5.8b}$$

$$R(x|a,b) = \exp\left(-\frac{x-a}{b}\right) \tag{5.8c}$$

$$h(x|a,b) = \frac{1}{b} \tag{5.8d}$$

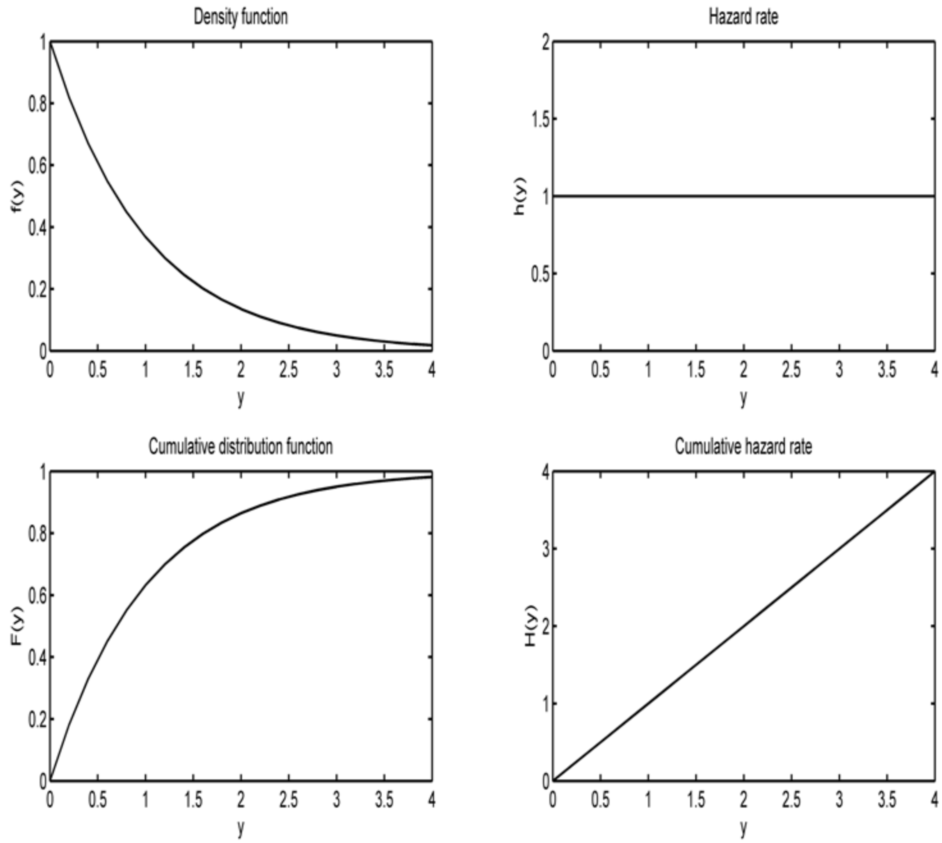
$$H(x|a,b) = \frac{x}{b} \tag{5.8e}$$

$$F_X^{-1}(P) = x_P = a - b\,\ln(1-P), \quad 0 \le P < 1 \tag{5.8f}$$

$$x_{0.5} = a + b\,\ln 2 \approx a + 0.6931\,b \tag{5.8g}$$

$$x_{0.6321} \approx a + b \ - \ \textbf{characteristic life} \tag{5.8h}$$

Figure 5/10: Several functions for the reduced exponential distribution




$$a = x_0 \tag{5.8i}$$

$$b \approx x_{0.6321} - x_0 \tag{5.8j}$$

$$x_M = a \text{ with } f(x_M|a,b) = 1/b \tag{5.8k}$$

$$M_X(t) = \frac{\exp(a\,t)}{1 - b\,t} \tag{5.8l}$$

$$C_X(t) = \frac{\exp(i\,a\,t)}{1 - i\,b\,t}, \quad i = \sqrt{-1} \tag{5.8m}$$

$$\mu'_r(X) = b^r \exp(a/b)\,\Gamma(1+r, a/b) \tag{5.8n}$$

$$\Gamma(1+r, a/b) = \int\limits_{a/b}^{\infty} u^r \exp(-u)\,\mathrm{d}u - \text{incomplete gamma function}$$

$$\mu'_1(X) = \mathrm{E}(X) = a + b \tag{5.8o}$$

$$\mu'_2(X) = (a+b)^2 + b^2 \tag{5.8p}$$

$$\mu'_3(X) = (a+b)^3 + b^2\,(3\,a + 5\,b) \tag{5.8q}$$

$$\mu'_4(X) = (a+b)^4 + b^2\,(6\,a^2 + 20\,a\,b + 23\,b^2) \tag{5.8r}$$

$$\mu'_r(X^*) = r!\,b^r = r\,b\,\mu'_{r-1}(X^*), \quad \text{where } X^* = X - a \tag{5.8s}$$

$$\mu_r(X) = b^r \sum_{j=0}^{r} \binom{r}{j} (-1)^j (r-j)! \tag{5.9a}$$

$$\mu_2(X) = \text{Var}(X) = b^2 \tag{5.9b}$$

$$\mu_3(X) = 2\,b^3 \tag{5.9c}$$

$$\mu_4(X) = 9\,b^4 \tag{5.9d}$$

$$\alpha_3 = 2 \tag{5.9e}$$

$$\alpha_4 = 9 \tag{5.9f}$$

$$\kappa_1(X) = a + b \tag{5.9g}$$

$$\kappa_r(X) = (r-1)!\,b^r; \quad r = 2, 3, \ldots \tag{5.9h}$$

$$I(X) = \text{ld}\,2 + \text{ld}\,b \approx 1.4427\,(1 + \ln b) \tag{5.9i}$$

$$F_Y^{-1}(P) = y_P = -\ln(1-P), \;\; 0 \le P < 1 \tag{5.9j}$$

$$f_Y(y_P) = 1 - P \tag{5.9k}$$

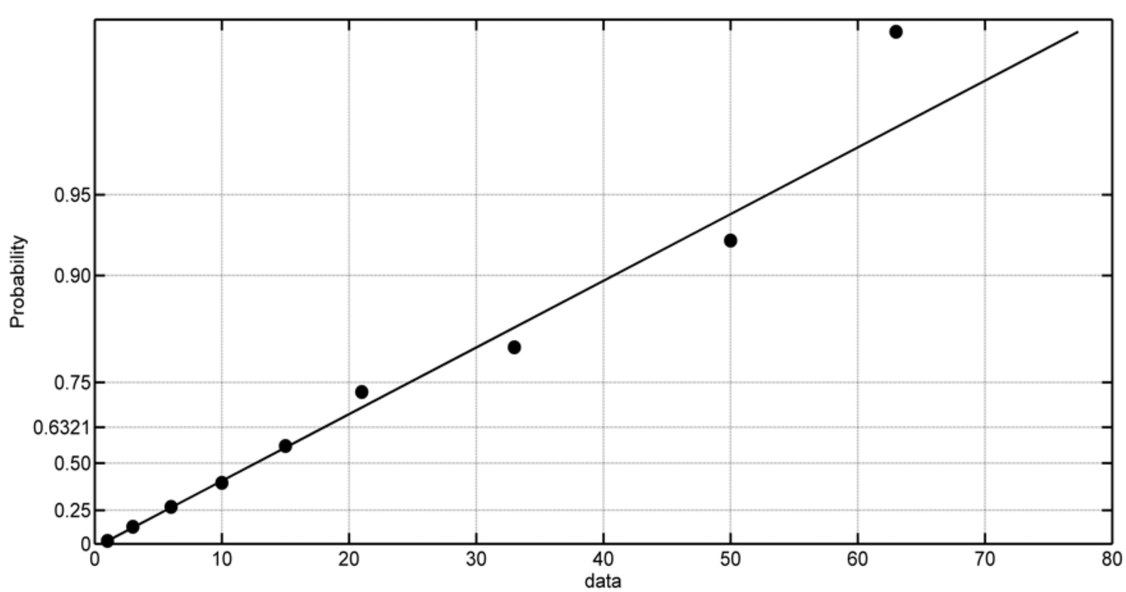
The exponential distribution is one of the few distributions whose moments of order statistics can be given in closed form, and in LEPP they are input to LLOYD's estimator:

$$\alpha_{r:n} = \text{E}(Y_{r:n}) = \sum_{j=1}^{r} \frac{1}{n-j+1} = \sum_{i=n-r+1}^{n} i^{-1}, \quad 1 \le r \le n \tag{5.9l}$$

$$\beta_{r,r:n} = \text{Var}(Y_{r:n}) = \beta_{r,s:n} = \text{Cov}(Y_{r:n}, Y_{s:n}) \tag{5.9m}$$

$$= \sum_{j=1}^{r} \frac{1}{(n-j+1)^2} = \sum_{i=n-r+1}^{n} i^{-2}, \quad 1 \le r < s \le n \tag{5.9n}$$

Figure 5/11: Exponential probability paper with data and regression line



### 5.2.5 Extreme value distributions of type I

An extreme value distribution[5] is the distribution of either the smallest or the largest observation in a sample for $n \to \infty$. Thus, it is a limiting distribution. The discovery of and the pioneering work on these types of distributions took place in the second and third decades of the last century. The main contributors are mathematicians and statisticians from England ( R.A. FISHER, L.H.C. TIPPETT), France (M. FRÉCHET), Germany (L. VON BORTKIEWICZ, R. VON MISES, E.J. GUMBEL) and Russia ( B.V. GNEDENKO). For more details of the genesis of extreme value distributions see JOHNSON/KOTZ/BALAKRISHNAN (1995, Chapter 22) and RINNE (2009, Chapter 1).

FISHER/TIPPETT (1928) were the first to prove that there only exist three types of limiting extreme value distributions, each type having a maximum and a minimum variant.[6] The type depends on the behavior of the sampled distribution on the relevant side, i.e. on the left–hand (right–hand) for the distribution of the minimum (maximum). The main results are:

- **Type I** will come up if the sampled distribution is unlimited towards the relevant side and is of **exponential type** on that side, meaning that the CDF of the sampled distribution, is increasing towards unity with $x \to \infty$ (decreasing towards zero with $x \to -\infty$) at least as quickly as an exponential function. Prototypes are the exponential, normal and the $\chi^2$–distributions.

- **Type II** will come up if the sampled distribution has a range which is unlimited from below (for the minimum type) or unlimited from above (for the maximum type), respectively, and if its CDF is of **CAUCHY–type**. This means, that for some positive $k$ and $A$:

  $$\lim_{x\to\infty} x^k \big[1 - F(x)\big] = A \;\text{ in case of a maximum or}$$

  $$\lim_{x\to-\infty} (-x)^k \, F(x) = A \;\text{ in case of a minimum.}$$

  The convergence of $F(x)$ or $\big[1 - F(x)\big]$, respectively, is slower than exponential. The prototype is the CAUCHY distribution itself.

- **Type III** will come up if the sampled distribution has a range which is **bounded** from above (for the maximum type) or bounded from below (for the minimum type), the bound being $x_0$. Besides the CDF must behave like

  $$\beta\,(x_0 - x)^\alpha \;\text{ for some }\; \alpha,\, \beta > 0 \;\text{ as }\; x \to x_0^- \;\text{ (case of a maximum) and like}$$

  $$\beta\,(x - x_0)^\alpha \;\text{ for some }\; \alpha,\, \beta > 0 \;\text{ as }\; x \to x_0^+ \;\text{ (case of a minimum).}$$

  A prototype is the uniform distribution over some interval $[A, B]$.

[5] Suggested reading for this section: GALAMBOS (1978, 1998), GUMBEL (1958), JOHNSON/KOTZ/BALAKRISHNAN (1995).

[6] The order of enumerating the extreme value distributions goes back to these authors.

We finally mention that type II distributions are said to be of **FRÉCHET type**. Type I distributions which have been studied extensively and applied frequently by E.J. GUMBEL, 1891 – 1966, see GUMBEL (1958), are said to be of **GUMBEL type**, whereas type III distributions are said to be of **WEIBULL type**.

The distributions of the asymptotically smallest and largest sample value are linked because of

$$\min_i X_i = -\big\{\max_i(-X_i)\big\} \tag{5.10a}$$

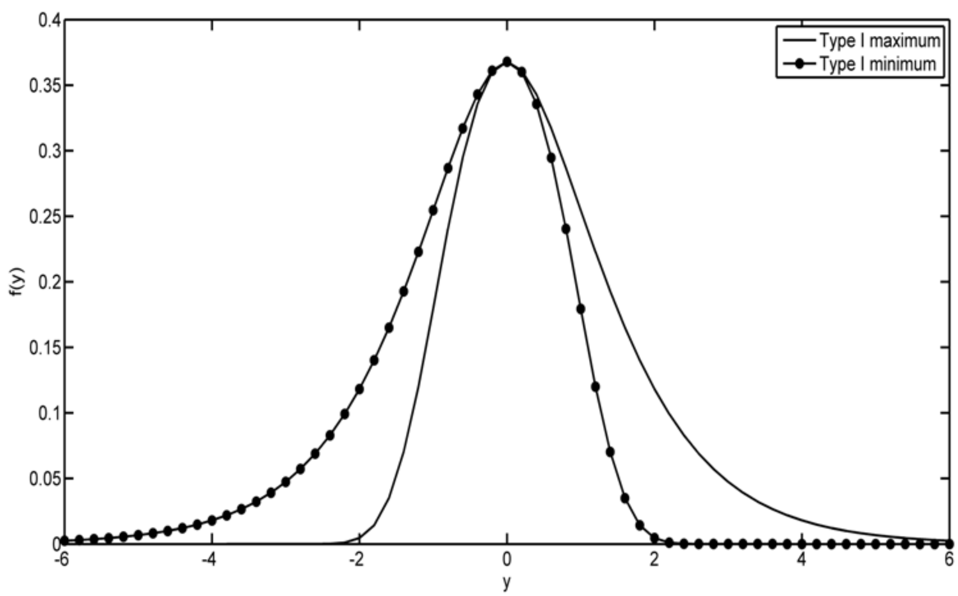
as follows, where $W$ is the maximum variable and $X$ the minimum variable, respectively, both being continuous:

$$\left.\begin{aligned} \Pr(X \leq t) = \Pr(W \geq -t) = 1 - \Pr(W < t), \\ F_X(t) = 1 - F_W(-t), \\ f_X(t) = f_W(-t). \end{aligned}\right\} \tag{5.10b}$$

Fig. 5/12 demonstrates this mirror–effect with the reduced versions of the type I distributions. Thus, it is sufficient to study only one of these versions. Furthermore, it has been shown in Sect. 1.3 and Table 1/2 that the type II and type III distributions, which are not of location–scale type but have a third parameter responsible for the shape, can be transformed to a type I distribution which is of location–scale type. In principle, we thus can trace back five of the distributions to a sixth one, e.g. to the type I maximum distribution.

Extreme value distributions play an important role in the natural sciences, especially in material research (strength of material) and in the life–sciences (lifetime distribution).

Figure 5/12: Densities of the reduced extreme value distributions of type I for the maximum and minimum



#### 5.2.5.1 Maximum distribution (GUMBEL distribution) — $X \sim EMX1(a,b)$

This type I extreme value distribution is also known as **GUMBEL distribution**, honoring the main contributor to its development, or **double exponential distribution** due to the mathematical structure of its DF and CDF, see (5.11a,b). Besides the relationships with the other extreme value distributions this extreme value distribution is related to several other distributions.

- When $X \sim EMX1(0,b)$ then $V = \exp(-X/b) \sim EX(0,b)$.
- When $X_i \overset{\text{iid}}{\sim} EMX1(a,b)$; $i = 1,\ldots,n$; then $W = \max(X_1,\ldots,X_n) \sim EMX1(a^*,b)$ where $a^* = a + \ln n$, i.e. the extreme value distribution of type I for the maximum is **reproductive with respect to** the formation of **sample maximum**.
- When $X_1,\ X_2 \overset{\text{iid}}{\sim} EMX1(a,b)$ than $W = X_1 - X_2$ has a **logistic distribution**.
- When $X \sim EMX1(0,b)$ then $W = b\big\{1 - \exp[-\exp(-X)]\big\}^{1/c}$ has a **PARETO distribution** with parameters $a = 0,\ b$ and $c$.
- When $X$ has a **power–function distribution** with parameters $a = 0,\ b$ and $c$ then $Y = -\ln[-c\,\ln X] \sim EMX1(0,1)$.
- When $Y \sim EMX1(0,1)$ then $X = \exp(-Y)$ has a **WEIBULL distribution**.

We have the following results for the type I maximum distribution:

$$f(x|a,b) = \frac{1}{b}\exp\left\{-\frac{x-a}{b} - \exp\left(-\frac{x-a}{b}\right)\right\},\quad a\in\mathbb{R},\ b>0,\ x\in\mathbb{R} \tag{5.11a}$$

$$F(x|a,b) = \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right] \tag{5.11b}$$

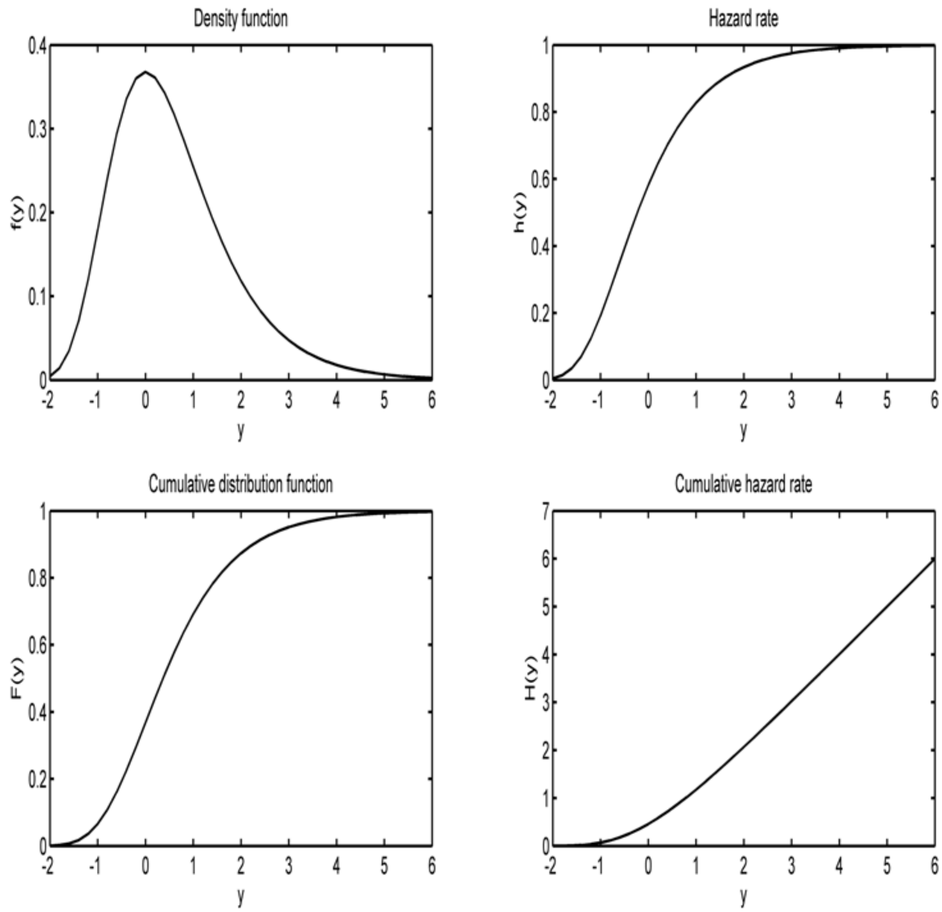
$$R(x|a,b) = 1 - \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right] \tag{5.11c}$$

$$h(x|a,b) = \frac{\exp\left(-\dfrac{x-a}{b}\right)}{b\left\{\exp\left[\exp\left(-\dfrac{x-a}{b}\right)\right] - 1\right\}} \tag{5.11d}$$

$$H(x|a,b) = -\ln\left\{1 - \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right]\right\} \tag{5.11e}$$

$$F_X^{-1}(P) = x_P = a - b\,\ln(-\ln P),\quad 0 < P < 1 \tag{5.11f}$$

Figure 5/13: Several functions of the reduced extreme value distribution of type I for the maximum




$$
\begin{aligned}
x_{0.5} &\approx a + 0.3665\, b && (5.11g)\\
a &\approx x_{0.3679} && (5.11h)\\
b &\approx x_{0.3679} - x_{0.0660} && (5.11i)\\
x_M &= a && (5.11j)\\
M_X(t) &= \exp(a\,t)\,\Gamma(1 - b\,t), \quad b\,|t| < 1 && (5.11k)\\
\mu_1'(X) &= \mathrm{E}(X) \;=\; a + b\,\gamma && (5.11l)\\
&\quad\; \gamma \approx 0.5772 - \textbf{Euler–Mascheroni's constant} &&\\
\mu_2(X) &= \mathrm{Var}(X) \;=\; b^2\,\frac{\pi^2}{6} \approx 1.6449\,b^2 && (5.11m)\\
\alpha_3 &\approx 1.1400 && (5.11n)\\
\alpha_4 &\approx 5.4 && (5.11o)\\
\kappa_r(X) &= b^r\,(r-1)!\sum_{k=0}^{\infty} k^{-r}, \quad r \geq 2 && (5.11p)
\end{aligned}
$$

$$\begin{aligned}
\kappa_2(X) &\approx 1.6449\,b^2 && (5.11q)\\
\kappa_3(X) &\approx 2.4041\,b^3 && (5.11r)\\
\kappa_4(X) &\approx 6.4939\,b^4 && (5.11s)\\
F_Y^{-1}(P) &= y_P = -\ln(-\ln P), \;\; 0 \le P \le 1 && (5.11t)\\
f_Y(y_P) &= -P\,\ln P && (5.11u)
\end{aligned}$$

With respect to the moments of reduced order statistics we have from LIEBLEIN (1953):[7]

$$\mathrm{E}\big(Y_{r:n}^k\big) = \frac{n!}{(r-1)!(n-r)!}\sum_{j=0}^{n-r}(-1)^j\binom{n-r}{j}\int\limits_{-\infty}^{\infty} u^k \exp\big[-u-(r+j)\exp(-u)\big]\,\mathrm{d}u. \quad (5.12a)$$

The integral in (5.12a)

$$g_k(c) := \int\limits_{-\infty}^{\infty} u^k\, \exp\big[-u - c\,\exp(-u)\big]\,\mathrm{d}u \quad (5.12b)$$

has the following solutions for $k = 1$ and $k = 2$:

$$\begin{aligned}
g_1(c) &= -\left[\frac{\Gamma'(c)}{c} - \frac{\Gamma(1)}{c}\,\ln c\right],\\
&= \frac{1}{c}\,\big(\gamma + \ln c\big), && (5.12c)\\
g_2(c) &= \frac{1}{c}\left[\frac{\pi^2}{6} + (\gamma + \ln c)^2\right]. && (5.12d)
\end{aligned}$$

As a special case of (5.12a) for $r = n$ (sample maximum) we have:

$$\begin{aligned}
\alpha_{n:n} &= \mathrm{E}(Y_{n:n}) = \gamma + \ln n \approx 0.57722 + \ln n, && (5.12e)\\
\alpha_{n:n}^{(2)} &= \mathrm{E}(Y_{n:n}^2) = \frac{\pi^2}{6} + (\gamma + \ln n)^2 \approx 1.6449 + (0.5772 + \ln n)^2, && (5.12f)\\
\beta_{n,n:n} &= \mathrm{Var}(Y_{n:n}) = \frac{\pi^2}{6} \approx 1.6449. && (5.12g)
\end{aligned}$$

It is to be mentioned that the variance of the last order statistic (5.12g) does not depend on the sample size $n$ and is always equal to $\pi^2/6$. Because of (5.13a) the variance of the first order statistic from a type I minimum distribution is independent of the sample size, too, and also equal to $\pi^2/6$.

[7] Another approach to evaluate the means and variances of $-Y_{r:n}$, i.e. of the minimum type I distribution, is given in WHITE (1969).

The product moment $\mathrm{E}(Y_{r:n}\,Y_{s:n})$ has the representation

$$\left.\begin{aligned}\mathrm{E}(Y_{r:n}\,Y_{s:n}) \;=\; & \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\sum_{k=0}^{s-r-1}\sum_{\ell=0}^{n-s}(-1)^{k+\ell}\binom{s-r-1}{k}\binom{n-s}{\ell}\\ & \times\,\phi(r+k,s-r-k+\ell),\;\; r<s\le n.\end{aligned}\right\}\tag{5.12h}$$

LIEBLEIN's $\phi$–function is the following double integral:

$$\phi(t,u)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{y} x\,y\exp\big[-x–t\exp(-x)\big]\exp\big[-y–u\exp(-y)\big]\,\mathrm{d}x\,\mathrm{d}y;\; t,u>0. \tag{5.12i}$$

LIEBLEIN shows how to evaluate the $\phi$–function based on SPENCE's integral

$$L(1+x)=\int\limits_{1}^{1+x}\frac{\ln w}{w-1}\,\mathrm{d}w=\sum_{m=1}^{\infty}\frac{(-1)^{m+1}}{m^2}.$$

There exist several tables providing means, variances and covariances of type I order statistics. The most comprehensive set of tables seems to be that of BALAKRISHNAN/CHAN (1992a) for $n=1(1)15(5)$ and from the same authors (1992b) a set of complete tables for all sample sizes up to 30. Another set of tables for $n=2(1)20$ has been compiled by WHITE (1964). All three tables refer to the type I minimum distribution, but because of (5.10a) they may be used for the maximum type I distribution as follows:

$$\alpha_{r:n} \;=\; -\widetilde{\alpha}_{n-r+1:n} \tag{5.13a}$$

$$\beta_{r,r:n} \;=\; \widetilde{\beta}_{n-r+1,n-r+1:n:n} \tag{5.13b}$$

$$\beta_{r,s:n} \;=\; \widetilde{\beta}_{n-s+1,n-r+1:n},\;\; r<s \tag{5.13c}$$

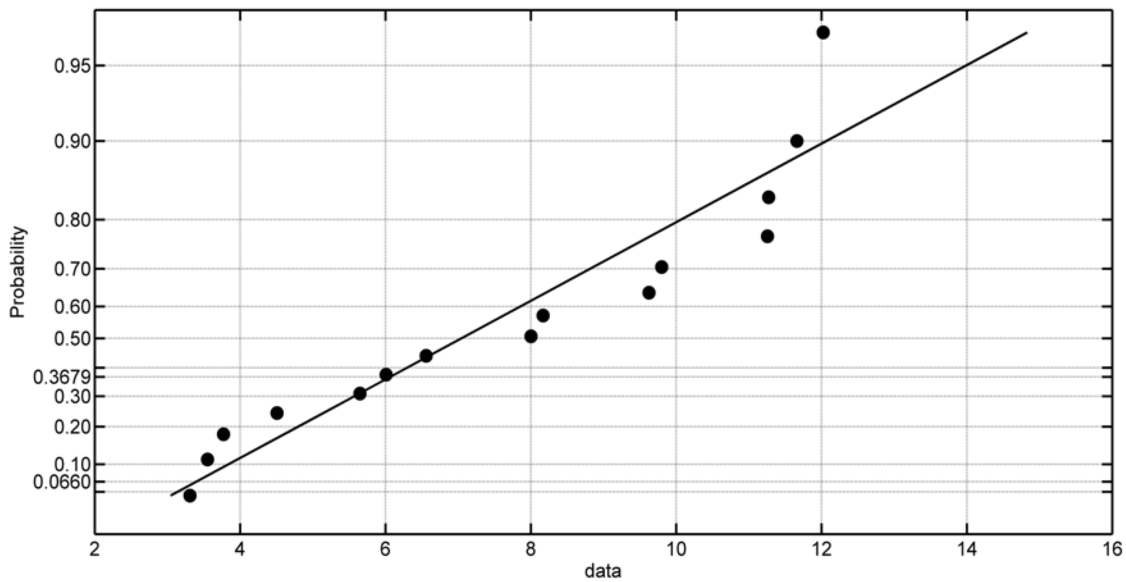
where $\widetilde{\alpha}_{\cdot:n}$ and $\widetilde{\beta}_{\cdot,\cdot:n}$ denote the moments of the minimum variable.

When order statistics are the data input to LEPP, linear estimation is done as follows:

- for $3\le n\le 10$ by LLOYD's estimator with computed means according to (5.12a) and tabulated variance–covariance matrix,

- for $n\ge 10$ by BLOM's unbiased, nearly best linear estimator with means according to (5.12a).

Figure 5/14: Type I maximum extreme value probability paper with data and regression line




### 5.2.5.2 Minimum distribution (Log–WEIBULL) — $X \sim EMN1(a,b)$

The results for this distribution easily follow from those of the maximum distribution in the preceding Section because for $X \sim EMX1(a,b)$ we have $-X \sim EMN1(a,b)$. The results are:

$$f(x|a,b) = \frac{1}{b}\left\{\exp\left[\frac{x-a}{b} - \exp\left(\frac{x-a}{b}\right)\right]\right\}, \quad a \in \mathbb{R},\ b>0,\ x \in \mathbb{R} \tag{5.14a}$$

$$F(x|a,b) = 1 - \exp\left[-\exp\left(\frac{x-a}{b}\right)\right] \tag{5.14b}$$

$$R(x|a,b) = \exp\left[-\exp\left(\frac{x-a}{b}\right)\right] \tag{5.14c}$$

$$h(x|a,b) = \frac{1}{b}\exp\left(\frac{x-a}{b}\right) \tag{5.14d}$$

$$H(x|a,b) = \exp\left(\frac{x-a}{b}\right) \tag{5.14e}$$

$$F_X^{-1}(P) = x_P = a + b\,\ln[-\ln(1-P)], \quad 0<P<1 \tag{5.14f}$$

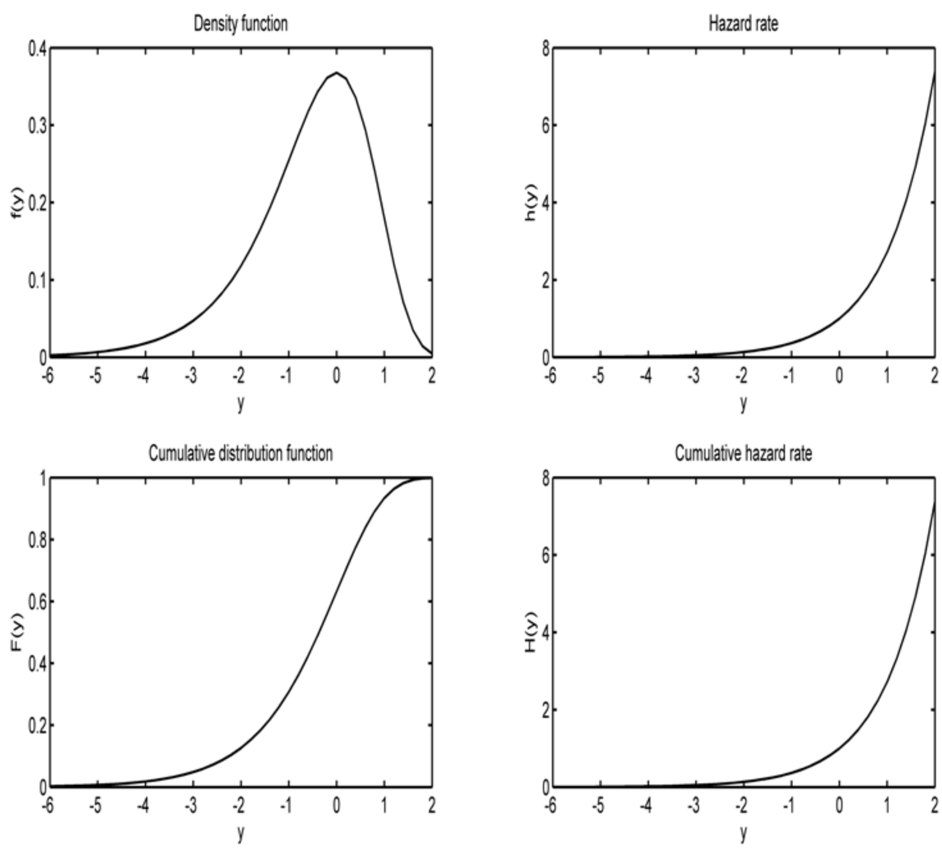
$$x_{0.5} = a - 0.3665\,b \tag{5.14g}$$

$$a = x_{0.6321} \tag{5.14h}$$

$$b = x_{0.9340} - x_{0.6321} \tag{5.14i}$$

$$
\begin{aligned}
x_M &= a && (5.14j)\\
M_X(t) &= \exp(a\,t)\,\Gamma(1+b\,t) && (5.14k)\\
\mu_1'(X) &= \mathrm{E}(X) \;=\; a - b\,\gamma \;\approx\; a - 0.5772\,b && (5.14l)\\
\mu_2(X) &= \mathrm{Var}(X) \;=\; b^2\,\pi^2/6 \;\approx\; 1.6449\,b^2 && (5.14m)\\
\alpha_3 &\approx -1.1396 && (5.14n)\\
\alpha_4 &= 5.4 && (5.14o)\\
\kappa_r(X) &= (-b)^r\,(r-1)!\sum_{k=0}^{r} k^{-r}; \quad r \ge 2 && (5.14p)\\
\kappa_2(X) &\approx 1.6449\,b^2 && (5.14q)\\
\kappa_3(X) &\approx -2.4041\,b^3 && (5.14r)\\
\kappa_4(X) &\approx 6.4939\,b^4 && (5.14s)\\
F_Y^{-1}(P) &= y_P \;=\; \ln[-\ln(1-P)],\;\; 0 < P < 1 && (5.14t)\\
f_Y(y_P) &= (P-1)\,\ln(1-P) && (5.14u)
\end{aligned}
$$

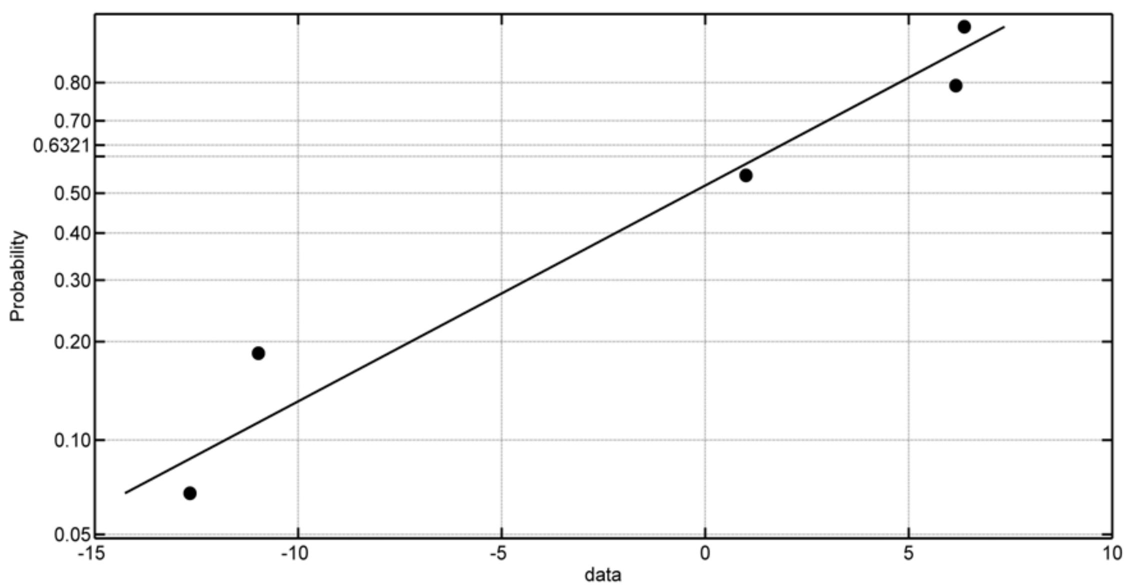
Figure 5/15: Several functions for the reduced extreme value distribution of type I for the minimum



The moments of the reduced order statistics of the minimum distribution easily from those of the maximum distribution, see (5.13a–c). Means, variances and covariances of reduced type I minimum order statistics are tabulated in BALAKRISHNAN/CHAN (1992a,b) and WHITE (1964). When order statistics are the data input to LEPP, linear estimation is done as follows:

- for $3 \leq n \leq 10$ by LLOYD's estimator with computed means according to (5.12a) and and (5.13a) tabulated variance–covariance matrix,
- for $n \geq 10$ by BLOM's unbiased, nearly best linear estimator with means according to (5.12a) and (5.13a).

Figure 5/16: Type I minimum extreme value probability paper with data and regression line




### 5.2.6 Half–distributions

The starting point for any half–distribution is a unimodal symmetric distribution of a continuous variate, i.e. we have $a = \mathrm{E}(X) = x_M = x_{0.5}$. Then, a **half–distribution** is obtained by folding the density $f(x|a,b)$ about $x = a$ to the right–hand side. Thus, the density $f_X(a-\Delta|a,b)$, where $x$ is $\Delta$ units less than $a$, is added to the density $f_X(a+\Delta|a,b)$ where $x$ is $\Delta$ units greater then $a$, $\Delta \geq 0$, so that the density at $x = a + \Delta$ is doubled. Stated otherwise, we have found the distribution of the variate $a + b\,|Y|$, where $Y$ is the reduced variate. We may also call the half–distribution a **special truncated distribution**, singly truncated from below at $x = a$. A half–distribution differs from a **folded distribution**, where the point of folding is $x = 0$. Both types coincide for $x = a = 0$, i.e. when the parent distribution is non-shifted.

We will discuss three types of half–distributions, the CAUCHY, the logistic and the normal cases, but there are other half distributions known by another name. Looking at the

LAPLACE distribution (Sect. 5.2.8) which is unimodal and symmetric, we find its half–distribution as the exponential distribution (Sect. 5.2.4). Likewise, the symmetric triangular distribution of Sect. 5.2.17.1 has a half–distribution which is the positively skew and right–angled distribution of Sect. 5.2.17.3 and the parabolic U–shaped distribution of Sect. 5.1.12.1 has the power–function distribution with $c = 2$ as its half–distribution.

We give the following hint to construct the probability paper for a half–distribution. On a sheet of paper for the parent unimodal symmetric distribution delete the printed probability scale $P$ for the range of $P$ less than $0.50$. For the range of $P$ greater than $0.50$ replace the value of $P$ by the corresponding value $P^* = 2\,P - 1$. This description follows from the fact that

$$\begin{aligned} P^* \;=\; F_{X^*}(x|a,b) &= \int\limits_a^x f_{X^*}(v|a,b)\,\mathrm{d}v \\ &= 2\int\limits_a^x f_X(v|a,b)\,\mathrm{d}v \\ &= 2\big[\underbrace{F_X(x|a,b)}_{=\,P} - 0.5\big], \end{aligned}$$

where $X$ denotes the parent variable and $X^*$ the half–distributed variable.

### 5.2.6.1 Half–CAUCHY distribution — $X \sim HC(a,b)$

Starting with the DF (5.3a) of the common CAUCHY distribution we find for the half–CAUCHY distribution:

$$f(x|a,b) = 2\left\{\pi\,b\left[1 + \left(\frac{x-a}{b}\right)^2\right]\right\}^{-1}, \quad x \ge a,\; a \in \mathbb{R},\; b > 0 \tag{5.15a}$$

$$F(x|a,b) = \frac{2}{\pi}\,\mathrm{arctan}\left(\frac{x-a}{b}\right) \tag{5.15b}$$

$$R(x|a,b) = 1 - \frac{2}{\pi}\,\mathrm{arctan}\left(\frac{x-a}{b}\right) \tag{5.15c}$$

$$h(x|a,b) = \frac{2}{b\left[1 - \left(\frac{x-a}{b}\right)^2\right]\left[\pi - 2\,\mathrm{arctan}\left(\frac{x-a}{b}\right)\right]} \tag{5.15d}$$

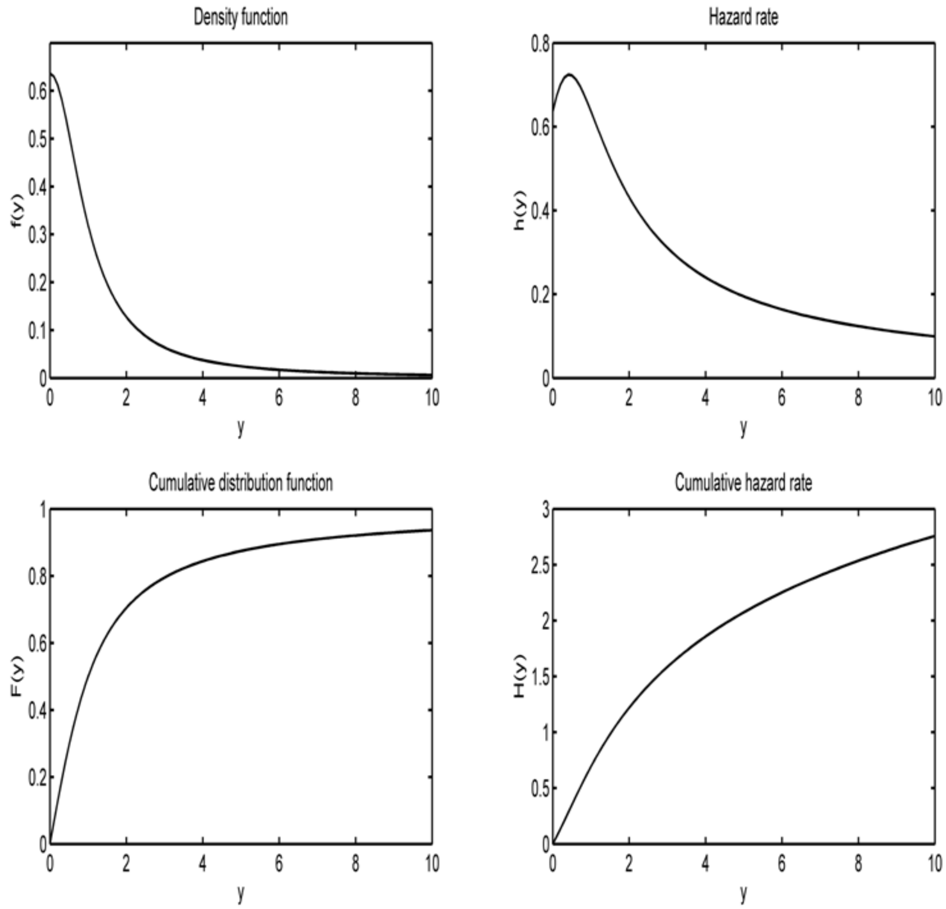
$$H(x|a,b) = -\ln\left[1 - \frac{2}{\pi}\,\mathrm{arctan}\left(\frac{x-a}{b}\right)\right] \tag{5.15e}$$

$$F_X^{-1}(P) = x_P = a + b\,\tan\left(\frac{\pi}{2}\,P\right), \quad 0 \le P < 1 \tag{5.15f}$$

$$x_{0.5} = a + b \tag{5.15g}$$

$$a = x_0 \tag{5.15h}$$

Figure 5/17: Several functions for the reduced half–CAUCHY distribution




$$b = x_{0.5} - x_0 \tag{5.15i}$$

$$x_M = a \tag{5.15j}$$

$$F_Y^{-1}(P) = y_P = \tan\left(\frac{\pi}{2}\,P\right), \quad 0 \le P < 1 \tag{5.15k}$$

$$f_Y(y_P) = 2\left\{\pi\left[1+\tan^2\left(\frac{\pi}{2}\,P\right)\right]\right\}^{-1} \tag{5.15l}$$

Moments of order $k > 1$ do not exist for the half–CAUCHY distribution because on $[0,\infty)$ the integral $\int\limits_a^\infty 2\,x^k \left\{\pi\,b\left[1+\left(\frac{x-a}{b}\right)^2\right]\right\}^{-1} \mathrm{d}x$ does not converge. The means of $Y_{1:n}$ and $Y_{n:n}$ and the variances of $Y_{1:n},\ Y_{2:n},\ Y_{n-1:n}$ and $X_{n:n}$ are infinite as is the case with the common CAUCHY distribution. The means of $Y_{r:n};\ r = 2,\ldots,n-1$; may be found by numerical integration. First the general formula (2.9b) gives

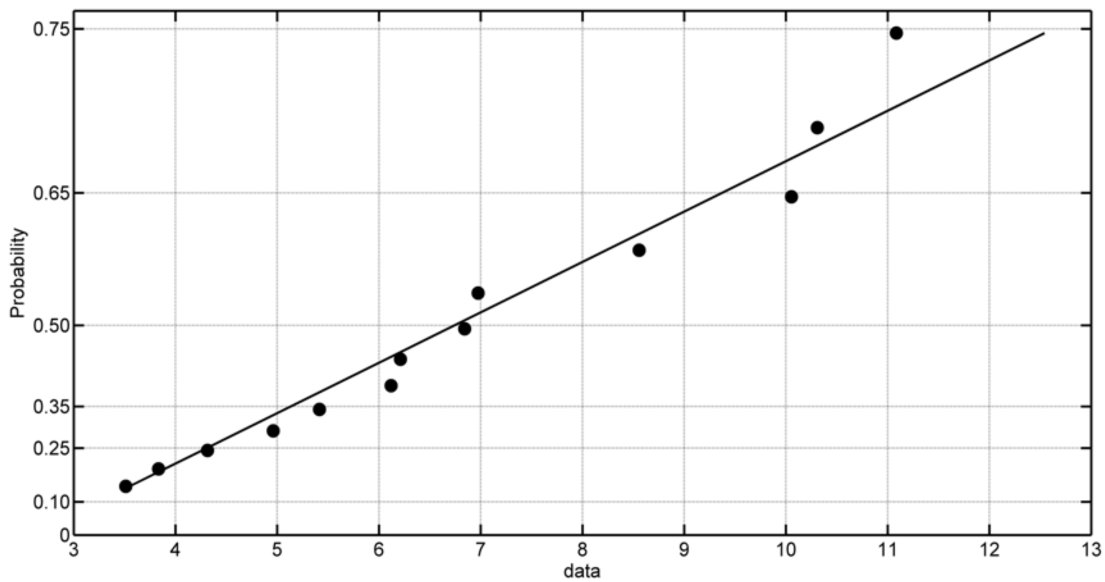
$$\alpha_{r:n} = \frac{n!}{(r-1)!(n-r)!}\int\limits_0^\infty \frac{2\,y}{\pi\,(1+y^2)}\left[\frac{2}{\pi}\,\arctan y\right]^{r-1}\left[1-\frac{2}{\pi}\,\arctan y\right]^{n-r} \mathrm{d}y \tag{5.15m}$$

and, finally, after the substitution $v = \arctan y$ with $\mathrm{d}y = (1 + y^2)\,\mathrm{d}v$ and some manipulations we have

$$\alpha_{r:n} = \frac{2^r\, n!}{\pi^n\,(r-1)!\,(n-r)!} \int\limits_0^{\pi/2} v^{r-1}(\pi - 2\,v)^{n-r}\,\tan v\,\mathrm{d}v. \tag{5.15n}$$

When order statistics are the data input to LEPP, linear estimation is done by BLOM's unbiased, nearly best linear estimator, where the means are evaluated by (5.15n).

Figure 5/18: Half–CAUCHY probability paper with data and regression line




### 5.2.6.2 Half–logistic distribution — $X \sim HL(a, b)$

We may write the DF, CDF and CCDF of the half–logistic distribution[8] in different ways as is the case with the common logistic distribution, see Sect. 5.2.9. The half–logistic distribution has been suggested as a possible life–time model, having an increasing hazard function, by BALAKRISHNAN (1985).

$$f(x|a,b) \;=\; \frac{2\,\exp\left(\dfrac{x-a}{b}\right)}{b\left[1+\exp\left(\dfrac{x-a}{b}\right)\right]^2}\,,\;\; x \ge a,\; a \in \mathbb{R},\; b > 0 \tag{5.16a}$$

[8] Suggested reading for this section: BALAKRISHNAN (1985), BALAKRISHNAN (1992), BALAKRISHNAN/PUTHENPURA (1986).

$$= \frac{2\,\exp\left(-\dfrac{x-a}{b}\right)}{b\left[1+\exp\left(-\dfrac{x-a}{b}\right)\right]^2} \tag{5.16b}$$

$$= \frac{1}{2\,b}\,\text{sech}^2\left(\frac{x-a}{2\,b}\right) \tag{5.16c}$$

$$F(x|a,b) = \frac{\exp\left(\dfrac{x-a}{b}\right)-1}{\exp\left(\dfrac{x-a}{b}+1\right)} \tag{5.16d}$$

$$= \frac{1-\exp\left(-\dfrac{x-a}{b}\right)}{1+\exp\left(-\dfrac{x-a}{b}\right)} \tag{5.16e}$$

$$= \tanh\left(\frac{x-a}{2\,b}\right) \tag{5.16f}$$

From (5.16a) and (5.16d) — or from (5.16b) and (5.16e) or from (5.16c) and (5.16f) — we immediately observe the relations

$$f(x|a,b) = F(x|a,b)\big[1-F(x|a,b)\big]+0.5\big[1-F(x|a,b)\big]^2, \tag{5.16g}$$

$$f(x|a,b) = \big[1-F(x|a,b)\big]-0.5\big[1-F(x|a,b)\big]^2, \tag{5.16h}$$

$$f(x|a,b) = 0.5\big[1-F^2(x|a,b)\big] \tag{5.16i}$$

These relations serve to establish numerous recurrence relations satisfied by the single and product moments of order statistics as given by (5.18a) through (5.19f).

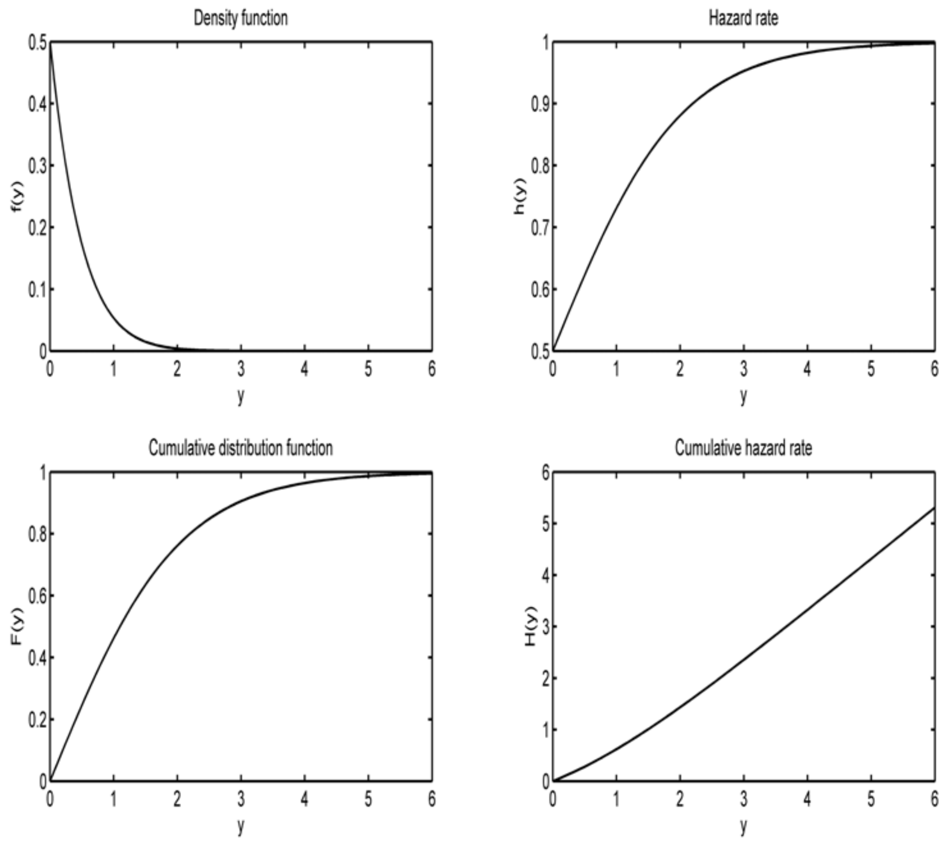
$$R(x|a,b) = \frac{2}{\exp\left(\dfrac{x-a}{b}+1\right)} \tag{5.16j}$$

$$= \frac{2\,\exp\left(-\dfrac{x-a}{b}\right)}{1+\exp\left(-\dfrac{x-a}{b}\right)} \tag{5.16k}$$

$$= 1-\tanh\left(\frac{x-a}{2\,b}\right) \tag{5.16l}$$

$$h(x|a,b) = \frac{1}{b\left[1+\exp\left(-\dfrac{x-a}{b}\right)\right]} = \frac{\exp\left(\dfrac{x-a}{b}\right)}{b\left[1+\exp\left(\dfrac{x-a}{b}\right)\right]} \tag{5.16m}$$

Figure 5/19: Several functions for the reduced half–logistic distribution




$$H(x|a,b) \;=\; -\ln\left[\frac{2}{1+\exp\left(\dfrac{x-a}{b}\right)}\right] \tag{5.16n}$$

$$F_X^{-1}(P) \;=\; x_P \;=\; a+b\,\ln\left(\frac{1+P}{1-P}\right) \tag{5.16o}$$

$$x_{0.5} \;=\; a+b\,\ln 3 \;\approx\; a+1.0986\,b \tag{5.16p}$$

$$a \;=\; x_0 \tag{5.16q}$$

$$b \;\approx\; x_{0.4621}-x_0 \tag{5.16r}$$

$$x_M \;=\; a \tag{5.16s}$$

$$M_X(t) \;=\; \exp(a\,t)\left\{b\,t\left[\psi\left(1-\frac{b\,t}{2}\right)-\psi\left(\frac{1-b\,t}{2}\right)\right]-1\right\}\;^{9} \tag{5.16t}$$

$$C_X(t) \;=\; \exp(i\,a\,t)\left\{i\,b\,t\left[\psi\left(1-\frac{i\,b\,t}{2}\right)-\psi\left(\frac{1-i\,b\,t}{2}\right)\right]+1\right\} \tag{5.16u}$$

[9] $\psi(z)$ is the **digamma function**: $\psi(z)=\mathrm{d}\ln\Gamma(z)\big/\mathrm{d}z=\Gamma'(z)\big/\Gamma(z)$.

$$\mu_1'(X) = \mathrm{E}(X) = a + b\,\ln 4 \approx a + 1.3863\,b \tag{5.17a}$$

$$\mu_2(X) = \mathrm{Var}(X) = b^2\left[\frac{\pi^2}{3} - \big(\ln 4\big)^2\right] \approx 1.3681\,b^2 \tag{5.17b}$$

$$\alpha_3 = \frac{9\,\zeta(3)}{\left[\dfrac{\pi^2}{3} - (\ln 4)^2\right]^{3/2}} \approx 6.7610 \;^{10} \tag{5.17c}$$

$$\alpha_4 = \frac{7\,\pi^4/15}{\left[\dfrac{\pi^2}{3} - (\ln 4)^2\right]^{2}} \approx 24.2884 \tag{5.17d}$$

$$F_Y^{-1}(P) = y_P = \ln\left(\frac{1+P}{1-P}\right), \;\; 0 \le P < 1 \tag{5.17e}$$

$$f_Y(y_P) = \frac{1-P^2}{2} \tag{5.17f}$$

$$\left.\begin{aligned}
F_Y^{-1(0)}(P) &= \ln\left(\frac{1+P}{1-P}\right)\\
F_Y^{-1(1)}(P) &= -\frac{2}{P^2-1}\\
F_Y^{-1(2)}(P) &= \frac{4\,P}{\big(P^2-1\big)^2}\\
F_Y^{-1(3)}(P) &= -\frac{4\big(1+3\,P^2\big)}{\big(P^2-1\big)^3}\\
F_Y^{-1(4)}(P) &= \frac{48\,P\big(1+P^2\big)}{\big(P^2-1\big)^4}\\
F_Y^{-1(5)}(P) &= -\frac{48\big(1+10\,P^2+5\,P^4\big)}{\big(P^2-1\big)^5}\\
F_Y^{-1(6)}(P) &= \frac{480\,P\big(3+10\,P^2+3\,P^4\big)}{\big(P^2-1\big)}
\end{aligned}\right\} \tag{5.17g}$$

Moments of order statistics for the reduced half–logistic distribution can be computed recursively. The following sets of recursion formulas are taken from ARNOLD/BALAKRISHNAN/NAGARAJA (1992, p. 105). **Single moments** are given by

[10] $\zeta(s) = \sum\limits_{k=1}^{\infty} k^{-s}$ is **RIEMANN's zeta function**.

$$\alpha_{1:n+1}^{(m)} = 2\left[\alpha_{1:n}^{(m)} - \frac{m}{n}\,\alpha_{1:n}^{(m-1)}\right], \;\; n \geq 1, \tag{5.18a}$$

$$\alpha_{2:n+1}^{(m)} = \frac{(n+1)\,m}{n}\,\alpha_{1:n}^{(m-1)} - \frac{n-1}{2}\,\alpha_{1:n+1}^{(m)}, \;\; n \geq 1, \tag{5.18b}$$

$$\alpha_{r:n+1}^{(m)} = \frac{1}{r}\left[\frac{(n+1)\,m}{n-r+1}\,\alpha_{r:n}^{(m-1)} + \frac{n+1}{2}\alpha_{r-1:n}^{(m)} - \frac{n-2r+1}{2}\alpha_{r:n+1}^{(m)}\right], \; 2 \leq r \leq n. \tag{5.18c}$$

The starting values are

$$\alpha_{1:1} = \alpha_{1:1}^{(1)} = \mathrm{E}(Y) = \ln 4, \tag{5.18d}$$

$$\alpha_{1:1}^{(2)} = \mathrm{E}\big(Y^2\big) = \pi^2/3. \tag{5.18e}$$

Furthermore, observe that

$$\alpha_{r:n}^{(0)} = 1 \;\text{ for }\; 1 \leq r \leq n. \tag{5.18f}$$

For computing the **product moments** in the triangle above the main diagonal $\alpha_{r,s:n+1} = \mathrm{E}\big(Y_{r:n+1}Y_{s:n+1}\big),\; 1 \leq r \leq n,\; r+1 \leq s \leq n+1$, for a sample of size $n+1$, we need the following formulas:

$$\alpha_{r,r+1:n+1} = \alpha_{r:n+1}^{(2)} + \frac{2\,(n+1)}{n-r+1}\left[\alpha_{r,r+1:n} - \alpha_{r:n}^{(2)} - \frac{1}{n-r}\,\alpha_{r:n}\right], \tag{5.19a}$$
$$1 \leq r \leq n-1, r \neq 2,$$

$$\alpha_{2,3:n+1} = \alpha_{3:n+1}^{(2)} + (n+1)\left[\alpha_{2:n} - \frac{n}{2}\,\alpha_{1:n-1}^{(2)}\right], \; n \geq 2, \tag{5.19b}$$

$$\alpha_{r+1,r+2:n+1} = \alpha_{r+2:n+1}^{(2)} + \frac{n+1}{r(r+1)}\left\{2\,\alpha_{r+1:n} + n\left[\alpha_{r-1,r:n-1} - \alpha_{r:n-1}^{(2)}\right]\right\}, \tag{5.19c}$$
$$2 \leq r \leq n-1,$$

$$\alpha_{r,s:n+1} = \alpha_{r,s-1:n+1} + \frac{2\,(n+1)}{n-s+2}\left[\alpha_{r,s:n} - \alpha_{r,s-1:n} - \frac{1}{n-s+1}\,\alpha_{r:n}\right], \tag{5.19d}$$
$$1 \leq r < s \leq n,\; s-r \geq 2,$$

$$\alpha_{2,s+1:n+1} = \alpha_{3,s+1:n+1} + (n+1)\left[\alpha_{s:n} - \frac{n}{2}\,\alpha_{1,s-1:n-1}\right], \; 3 \leq s \leq n, \tag{5.19e}$$

$$\alpha_{r+1,s+1:n+1} = \alpha_{r+2,s+1:n+1} + \frac{n+1}{r(r+1)}\left\{2\,\alpha_{s:n} - n\left[\alpha_{r,s-1:n-1} - \alpha_{r-1,s-1:n-1}\right]\right\}, \tag{5.19f}$$
$$2 \leq r < s \leq n,\; s-r \geq 2.$$

The starting value for the recursion is

$$\alpha_{1,2:2} = \alpha_{1,1}^2 = \big(\ln 4\big)^2. \tag{5.19g}$$

These recursion formulas have to be used in suitable order to find all elements in the upper triangle of the matrix. The element $\alpha_{1,n+1:n+1},\ n \geq 2$, is not given by any of the formulas (5.19a–f) but has to be derived from the identity

$$\sum_{r=1}^{n+1}\sum_{s=r+1}^{n+1}\alpha_{r,s:n+1} = \binom{n+1}{2}\alpha_{1:1}^2 = \frac{(n+1)\,n}{2}\big(\ln 4\big)^2$$

as

$$\alpha_{1,n+1:n+1} = \frac{(n+1)\,n}{2}\big(\ln 4\big)^2 - \left[\sum_{s=2}^{n}\alpha_{1,s:n+1} + \sum_{r=2}^{n+1}\sum_{s=r+1}^{n+1}\alpha_{r,s:n+1}\right]. \tag{5.19h}$$

**Example 5/3: Application of the recursions for $n = 5$**

We first have to evaluate (5.18a) for $m = 1$ leading to the following matrix of means $\alpha_{r:n} = \alpha_{r:n}^{(1)}$:

| | $n$ | | | | |
|---|---|---|---|---|---|
| $r$ | 1 | 2 | 3 | 4 | 5 |
| 1 | 1.3863 | 0.7726 | 0.5452 | 0.4237 | 0.3474 |
| 2 | | 2.0000 | 1.2274 | 0.9096 | 0.7289 |
| 3 | | | 2.3863 | 1.5452 | 1.1807 |
| 4 | | | | 2.6667 | 1.7887 |
| 5 | | | | | 2.8863 |

Then we execute (5.18a) for $m = 2$ using the preceding results and arrive at the following matrix of second moments $\alpha_{r:n}^{(2)}$:

| | $n$ | | | | |
|---|---|---|---|---|---|
| $r$ | 1 | 2 | 3 | 4 | 5 |
| 1 | 3.2899 | 1.0346 | 0.5239 | 0.3210 | 0.2183 |
| 2 | | 5.5452 | 2.0558 | 1.1328 | 0.7318 |
| 3 | | | 7.2899 | 2.9788 | 1.7343 |
| 4 | | | | 8.7269 | 3.8084 |
| 5 | | | | | 9.9565 |

The preceding two matrices in combination with (5.19a-f) lead to the following symmetric matrices of product moments $\alpha_{r,s:n}$, where we only reproduce the upper parts:

$$n = 2 \;\rightarrow\; \begin{pmatrix} 0 & 1.9218 \\ & 0 \end{pmatrix}, \quad n = 3 \;\rightarrow\; \begin{pmatrix} 0 & 0.8679 & 1.4772 \\ & 0 & 3.4203 \\ & & 0 \end{pmatrix},$$

$$n = 4 \;\rightarrow \begin{pmatrix} 0 & 0.5114 & 0.7679 & 1.2324 \\ & 0 & 1.6811 & 2.6762 \\ & & 0 & 4.66.19 \\ & & & 0 \end{pmatrix},$$

$$n = 5 \;\rightarrow \begin{pmatrix} 0 & 0.3413 & 0.4901 & 0.6943 & 1.0699 \\ & 0 & 1.0432 & 1.4707 & 2.2581 \\ & & 0 & 2.4241 & 3.6971 \\ & & & 0 & 5.7294 \\ & & & & 0 \end{pmatrix}.$$

The variance–covariance matrices follow in the usual way. They read, only giving the upper triangles:

$$n = 2 \;\rightarrow \begin{pmatrix} 0.4377 & 0.3766 \\ & 1.5452 \end{pmatrix}, \quad n = 3 \;\rightarrow \begin{pmatrix} 0.2267 & 0.1988 & 0.1763 \\ & 0.5493 & 0.1413 \\ & & 1.5955 \end{pmatrix},$$

$$n = 4 \;\rightarrow \begin{pmatrix} 0.1415 & 0.1260 & 0.1132 & 0.1026 \\ & 0.3054 & 0.2755 & 0.2505 \\ & & 0.5912 & 0.5414 \\ & & & 1.6158 \end{pmatrix},$$

$$n = 5 \;\rightarrow \begin{pmatrix} 0.0976 & 0.0880 & 0.0800 & 0.0731 & 0.0673 \\ & 0.2005 & 0.1825 & 0.1672 & 0.1541 \\ & & 0.3403 & 0.3128 & 0.2892 \\ & & & 0.6109 & 0.5683 \\ & & & & 1.6258 \end{pmatrix}.$$

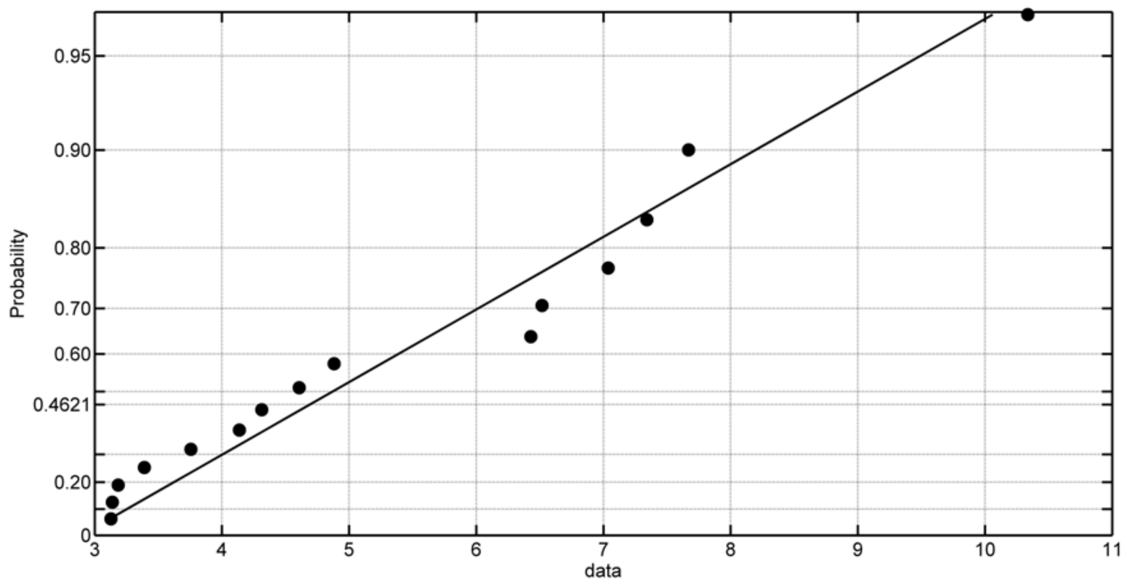
---

It is not to be recommended to apply the recursion formulas for greater sample sizes because the rounding errors cumulate and distort the results. In LEPP we have decided to proceed as follows:

- For $n > 35$ we compute all moments by the short versions of the approximating formulas (2.23) through (2.25) and use LLOYD's GLS estimator.
- For $20 \leq n \leq 35$ we use BLOM's unbiased, nearly best linear estimator, where the means are computed by the recursions (5.18a–c) which are not so sensitive to the cumulation of rounding errors.

- For $n < 20$ all the moments — means as well as the variance–covariance matrix — are computed using the recursion formulas and inserted into LLOYD's estimator.

Figure 5/20: Half–logistic probability paper with data and regression line




#### 5.2.6.3 Half–normal distribution — $X \sim HN(a, b)$

The half–normal distribution[11] is the distribution of $X = a + b\,|Z|$ where $Z$ has a standard normal distribution. The reduced form of the half–normal distribution is also known as the $\chi$–distribution with $\nu = 1$ degree of freedom.[12] We note the following functions and characteristics of a half–normal distribution:

$$f(x|a,b) = \frac{1}{b}\sqrt{\frac{2}{\pi}}\,\exp\left[-\frac{(x-a)^2}{2\,b^2}\right],\quad x \geq a,\ a \in \mathbb{R},\ b > 0 \qquad (5.20a)$$

$$F(x|a,b) = 2\,\Phi\left(\frac{x-a}{b}\right) - 1,^{13} \qquad (5.20b)$$

[11] Suggested reading for this section: BALAKRISHNAN/RAO (1998a, p. 172–175), BALAKRISHNAN/COHEN (1991, p. 660–62), JOHNSON/KOTZ/BALAKRISHNAN (1994, p. 421).

[12] The DF of a $\chi(\nu)$– distribution reads

$$f(x|\nu) = \frac{x^{\nu-1}\,\exp\big(-x^2/2\big)}{2^{\nu/2-1}\,\Gamma(\nu/2)}.$$

$$R(x|a,b) = 2\left[1-\Phi\left(\frac{x-a}{b}\right)\right] \tag{5.20c}$$

$$h(x|a,b) = \frac{1}{b\sqrt{2\,\pi}}\,\frac{\exp\left[-\dfrac{(x-a)^2}{2\,b^2}\right]}{1-\Phi\left(\dfrac{x-a}{b}\right)} \tag{5.20d}$$

$$H(x|a,b) = -\ln\left\{2\left[1-\Phi\left(\frac{x-a}{b}\right)\right]\right\} \tag{5.20e}$$

$$F^{-1}(P) = x_P = a + b\,\Phi^{-1}\left(\frac{1+P}{2}\right),\ 0<P<1, \quad [14] \tag{5.20f}$$

$$x_{0.5} \approx a + 0.6745\,b \tag{5.20g}$$

$$a = x_0 \tag{5.20h}$$

$$b \approx x_{0.6827} - a \tag{5.20i}$$

$$x_M = a \tag{5.20j}$$

$$C_X(t) = M\left(\frac{1}{2},\frac{1}{2},-\frac{t^2}{2}\right) + \frac{i\,t\sqrt{2}}{\Gamma\left(\dfrac{1}{2}\right)}\,M\left(1,\frac{3}{2},-\frac{t^2}{2}\right) \quad [15] \tag{5.20k}$$

$$\mu'_r(Y) = \mathrm{E}(Y^r) = 2^{r/2}\,\frac{\Gamma\left(\dfrac{1+r}{2}\right)}{\Gamma\left(\dfrac{1}{2}\right)},\ Y = \frac{X-a}{b}, \tag{5.20l}$$

$$\mu'_1(Y) = \mathrm{E}(Y) = \sqrt{\frac{2}{\pi}} \approx 0.7979 \tag{5.20m}$$

$$\mu'_2(Y) = 1 \tag{5.20n}$$

$$\mu'_3(Y) = 2\sqrt{\frac{2}{\pi}} \approx 1.5958 \tag{5.20o}$$

$$\mu'_4(Y) = 3 \tag{5.20p}$$

---

[13] $\Phi(z) = \int\limits_0^z \frac{1}{\sqrt{2\,\pi}}\exp(-v^2/2)\,\mathrm{d}v$ is the CDF of the standard normal variate. $\Phi(z)$ cannot be given in closed form.

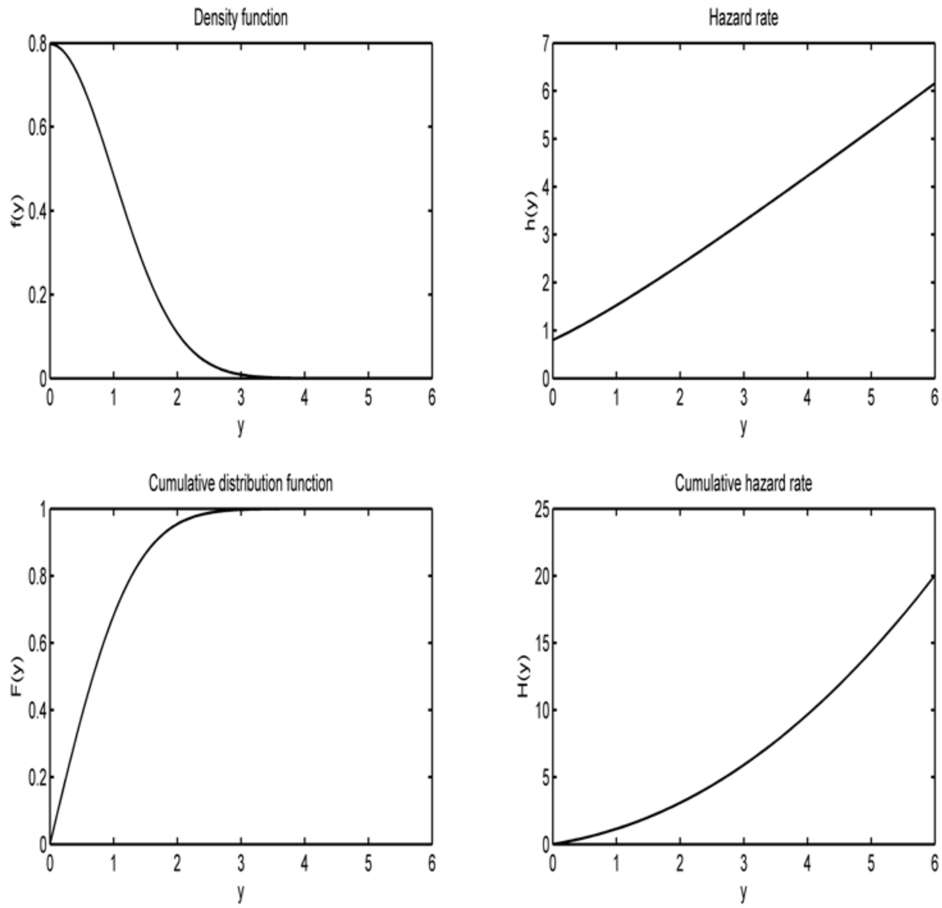
[14] $\Phi^{-1}(\bullet)$, which cannot be given in closed form, is the percentile function of the standard normal distribution.

[15] $M(a,b,z)$ is **KUMMER's function**:

$$M(a,b,z) = 1 + \frac{a\,z}{b} + \frac{(a)_2\,z^2}{(b)_2\,2!} + \ldots + \frac{(a)_n\,z^n}{(b)_n\,n!} + \ldots$$

where $(a)_n = a\,(a+1)\ldots(a+n-1)$ and $(a)_0 = 1$.

Figure 5/21: Several functions for the reduced half–normal distribution




$$\mu_1'(X) \; = \; \mathrm{E}(X) \; = \; a + b\sqrt{\frac{2}{\pi}} \; \approx \; a + 0.7979\,b \tag{5.20q}$$

$$\mu_2(X) \; = \; \mathrm{Var}(X) \; = \; b^2\left(1 - \frac{2}{\pi}\right) \; \approx \; 0.3634\,b^2 \tag{5.20r}$$

$$\alpha_3 \; = \; \frac{\sqrt{2}\,(4-\pi)}{(\pi-2)^{3/2}} \; \approx \; 0.9953 \tag{5.20s}$$

$$\alpha_4 \; = \; 3 + \frac{8\,(\pi-3)}{(\pi-2)^2} \; \approx \; 3.8692 \tag{5.20t}$$

$$F_Y^{-1}(P) \; = \; y_P \; = \; \Phi^{-1}\left(\frac{1+P}{2}\right), \;\; 0 < P < 1 \tag{5.20u}$$

$$f_Y(y_P) \; = \; \sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{2}\,y_P^2\right) \tag{5.20v}$$

For the half–normal distribution we have — as in the case of the common normal distribution —

$$\frac{\mathrm{d}f(y)}{\mathrm{d}y} = -y\,f(y). \tag{5.20w}$$

From (5.20w) we find the following derivatives of $y_P = F_P^{-1}(P)$:

$$\left.\begin{aligned}
F_Y^{-1(0)}(P) &= y_P \;=\; \Phi^{-1}\left(\frac{1+P}{2}\right),\\
F_Y^{-1(1)}(P) &= \frac{1}{f_Y(y_P)},\\
F_Y^{-1(2)}(P) &= \frac{y_P}{\big[f_Y(y_P)\big]^2},\\
F_Y^{-1(3)}(P) &= \frac{1-2\,y_P^2}{\big[f_Y(y_P)\big]^3},\\
F_Y^{-1(4)}(P) &= \frac{y_P\big[7+6\,y_P^2\big]^2}{\big[f_Y(y_P)\big]^4},\\
F_Y^{-1(5)}(P) &= \frac{7+46\,y_P^2+24\,y_P^4}{\big[f_Y(y_P)\big]^5},\\
F_Y^{-1(6)}(P) &= \frac{y_P\big(127+326\,y_P^2+120\,y_P^4\big)}{\big[f_Y(y_P)\big]^6}.
\end{aligned}\right\} \tag{5.20x}$$

GOVINDARAJULU/EISENTSTAT (1965) have tabulated all means and covariances of reduced order statistics for $n = 1(1)20(10)100$. BALAKRISHNAN/COHEN (1991, p. 61) give the following formulas for the product moments $\alpha_{r,s:n}$ and covariances $\beta_{r,s:n}$ of reduced order statistics from a half–normal distribution:

$$\sum_{s=1}^{n} \alpha_{r,s:n} = 1 + n\,\alpha_{1:1}\,\alpha_{r-1:n-1},\;\; 1 \le r \le n, \tag{5.21a}$$

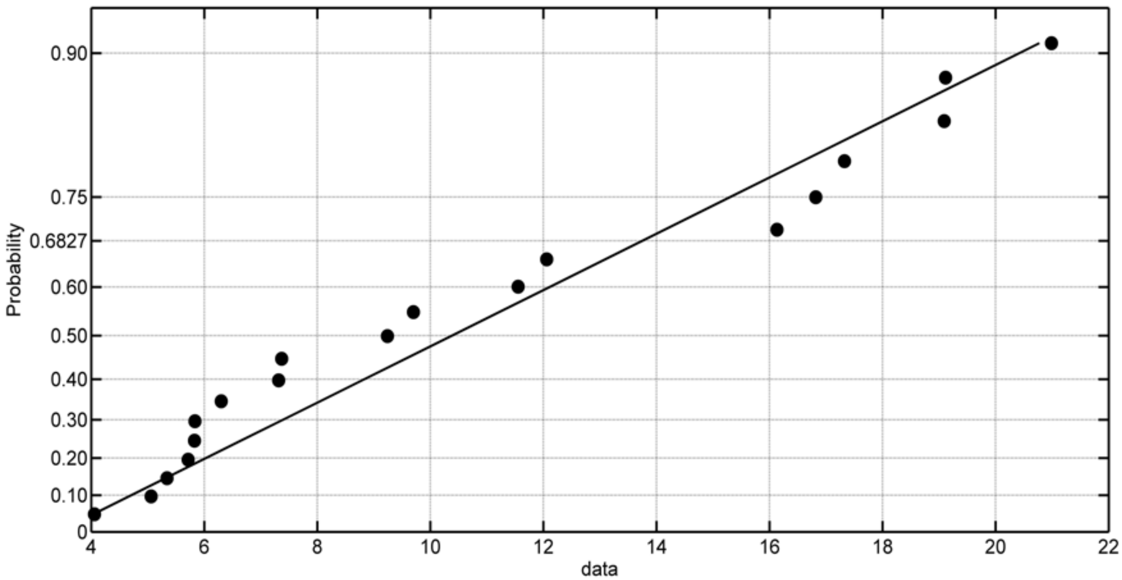
$$\sum_{s=1}^{n} \beta_{r,s:n} = 1 - (n-r+1)\,\alpha_{11:1}\,(\alpha_{r:n} - \alpha_{r-1:n}),\;\; 1 \le r \le n, \tag{5.21b}$$

where

$$\begin{aligned}
\alpha_{0:n} &= 0 \quad \text{for} \quad n \ge 1,\\
\alpha_{1:1} &= \mathrm{E}(Y) \;=\; \sqrt{2/\pi}.
\end{aligned}$$

In LEPP linear estimation for order statistic input is realized with BLOM's unbiased, nearly best linear estimator where the means are approximated by the short version of (2.23).

Figure 5/22: Half–normal probability paper with data and regression line




### 5.2.7 Hyperbolic secant distribution — $X \sim HS(a, b)$

When we look at a reduced variate with a symmetric DF of type

$$f(y) = \frac{c_1}{\exp y + \exp(-y)}, \quad y \in \mathbb{R}, \tag{5.22a}$$

the condition

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

requires $c_1 = 2/\pi$. Furthermore, because the **hyperbolic secant function** is given by

$$\operatorname{sech} y = \frac{2}{\exp y + \exp(-y)}$$

we may equally write (5.22a) as

$$f(y) = \frac{1}{\pi} \operatorname{sech} y \tag{5.22b}$$

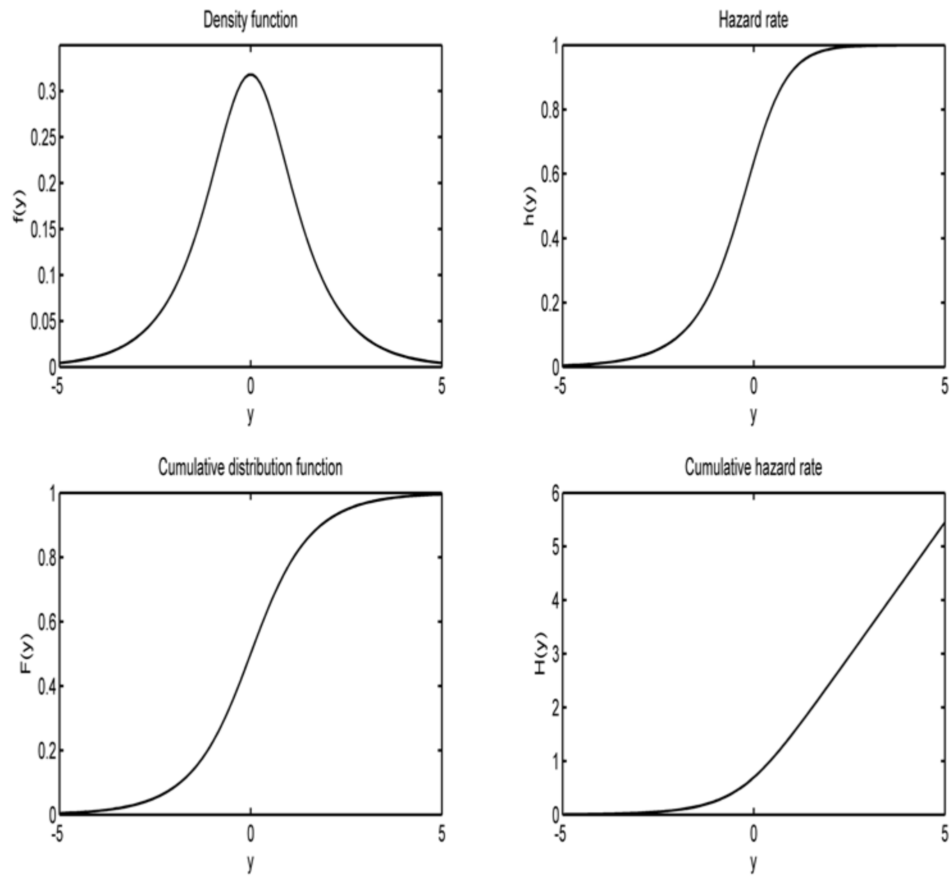
which is the reduced form DF of the hyperbolic secant distribution.[16] For the general hyperbolic secant distribution we have the following functions and characteristics:[17]

---

[16] Suggested reading for this section: JOHNSON/KOTZ/BALAKRISHNAN (1995, Chapter 5), TALACKO (1956).

[17] When $Y$ has the DF given by (5.22b) then $\exp(Y)$ has a half–CAUCHY distribution.

Figure 5/23: Several functions for the reduced hyperbolic secant distribution




$$f(x|a,b) = \frac{1}{b\,\pi}\,\text{sech}\left(\frac{x-a}{b}\right), \quad x \in \mathbb{R},\ a \in \mathbb{R},\ b > 0 \tag{5.23a}$$

$$F(x|a,b) = \frac{2}{\pi}\,\arctan\left[\exp\left(\frac{x-a}{b}\right)\right] = \frac{1}{2} + \frac{1}{\pi}\,\arctan\left[\sinh\left(\frac{x-a}{b}\right)\right] \tag{5.23b}$$

$$R(x|a,b) = 1 - \frac{2}{\pi}\,\arctan\left[\exp\left(\frac{x-a}{b}\right)\right] = \frac{1}{2} - \frac{1}{\pi}\,\arctan\left[\sinh\left(\frac{x-a}{b}\right)\right] \tag{5.23c}$$

$$h(x|a,b) = \frac{\text{sech}\left(\dfrac{x-a}{b}\right)}{b\left\{\pi - 2\,\arctan\left[\exp\left(\dfrac{x-a}{b}\right)\right]\right\}}\ {}^{18} \tag{5.23d}$$

$$= \frac{\text{sech}\left(\dfrac{x-a}{b}\right)}{b\left\{\dfrac{\pi}{2} - \arctan\left[\sinh\left(\dfrac{x-a}{b}\right)\right]\right\}} \tag{5.23e}$$

[18] We notice that $h(x|0,1)$ looks like a CDF, see Fig. 5/23.

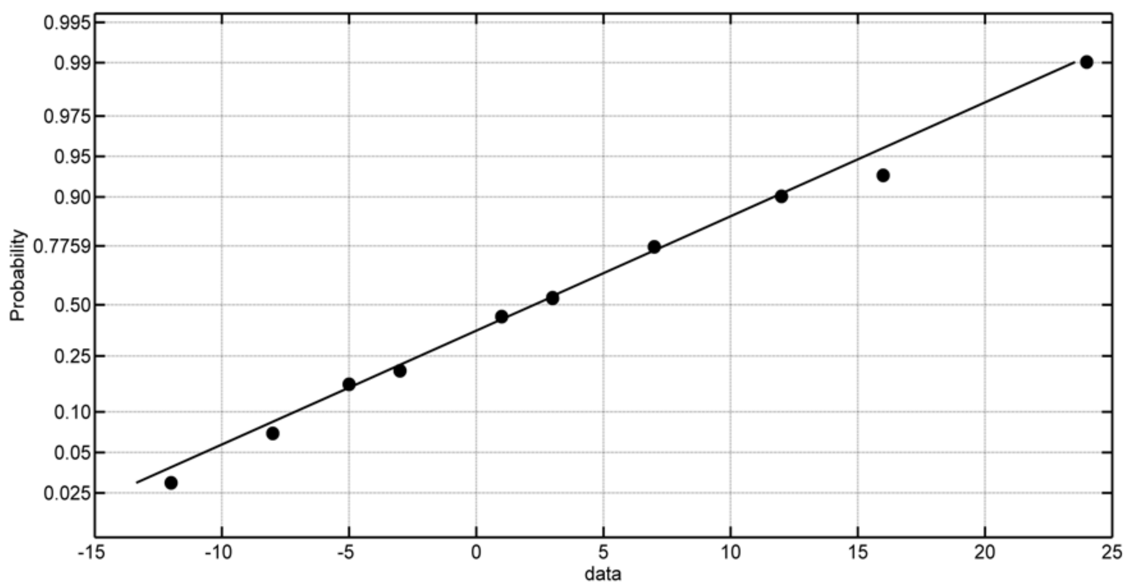
$$
\begin{aligned}
H(x|a,b) &= -\ln\left\{1-\frac{2}{\pi}\,\arctan\left[\exp\left(\frac{x-a}{b}\right)\right]\right\} &\text{(5.23f)}\\
&= -\ln\left\{\frac{1}{2}-\frac{1}{\pi}\,\arctan\left[\sinh\left(\frac{x-a}{b}\right)\right]\right\} &\text{(5.23g)}\\
F_X^{-1}(P) &= x_P \;=\; a+b\,\ln\left[\tan\left(\frac{\pi}{2}\,P\right)\right],\;\; 0<P<1 &\text{(5.23h)}\\
a &= x_{0.5} &\text{(5.23i)}\\
b &\approx x_{0.7759}-a &\text{(5.23j)}\\
x_M &= a &\text{(5.23k)}\\
C_X(t) &= \exp(i\,t\,a)\,\text{sech}\left(\frac{\pi}{2}\,b\,t\right) &\text{(5.23l)}\\
M_X(t) &= \exp(t\,a)\,\sec\left(\frac{\pi}{2}\,b\,t\right) &\text{(5.23m)}\\
\mu'_{2k+1}(Y) &= 0;\;\; k=0,1,\ldots;\;\; Y=(X-a)/b &\text{(5.23n)}\\
\mu'_2(Y) &= \frac{\pi^2}{4}\;\approx\;2.4674 &\text{(5.23o)}\\
\mu'_4(Y) &= 5 &\text{(5.23p)}\\
\mu'_1(X) &= \text{E}(X)\;=\;a &\text{(5.23q)}\\
\mu_2(X) &= \text{Var}(X)\;=\;\frac{\pi^2}{4}\;\approx\;2.4674 &\text{(5.23r)}\\
\alpha_3 &= 0 &\text{(5.23s)}\\
\alpha_4 &= 5 &\text{(5.23t)}\\
F_Y^{-1}(P) &= y_P \;=\; \ln\left[\tan\left(\frac{\pi}{2}\,P\right)\right],\;\; 0<P<1 &\text{(5.23u)}\\
f_Y(y_P) &= \frac{1}{\pi}\,\sin(\pi\,P) &\text{(5.23v)}
\end{aligned}
$$

For $a = 0$ and $b = 2/\pi$ the hyperbolic secant distribution shares many properties with the standard normal distribution. It is symmetric with mean, median and mode all equal to zero, with unit variance and a DF which is proportional to its characteristic function. However, the hyperbolic secant distribution is leptokurtic, i.e. it has a more acute peak near its mean and heavier tails compared with the standard normal distribution.

$$\left.\begin{aligned}
F^{-1(0)}(P) &= \ln\Big[\tan\Big(\frac{\pi}{2}\,P\Big)\Big] \\
F^{-1(1)}(P) &= \pi\,\csc(\pi\,P) \\
F^{-1(2)}(P) &= -\pi^2\,\cot(\pi\,P)\,\csc(\pi\,P) \\
F^{-1(3)}(P) &= \frac{\pi^3}{2}\big[3+\cos(2\,\pi\,P)\big]\,\csc^3(\pi\,P) \\
F^{-1(4)}(P) &= -\frac{\pi^4}{4}\big[23\,\cos(\pi\,P)+\cos(3\,\pi\,P)\big]\,\csc^4(\pi\,P) \\
F^{-1(5)}(P) &= \frac{\pi^5}{8}\big[115+76\,\cos(2\,\pi\,P)+\cos(4\,\pi\,P)\big]\,\csc^5(\pi\,P) \\
F^{-1(6)}(P) &= -\frac{\pi^6}{16}\big[1682\,\cos(\pi\,P)+237\,\cos(3\,\pi\,P)+\cos(5\,\pi\,P)\big]\,\csc^6(\pi P)
\end{aligned}\right\} \quad (5.23w)$$

Linear estimation in LEPP for order statistics input is done by LLOYD's estimator with approximated means and variance–covariance matrix, see (2.23) – (2.25).

Figure 5/24: Hyperbolic secant probability paper with data and regression line




### 5.2.8 LAPLACE distribution — $X \sim LA(a, b)$

This distribution[19] is named after the French mathematician PIERRE SIMON DE LAPLACE (1749 – 1827), who discovered it in 1774 as the distribution form for which the likelihood function is maximized by setting the location parameter equal to the sampling median

[19] Suggested reading for this section: JOHNSON/KOTZ/BALAKRISHNAN (1995, Chapter 24).

when the sample size is an odd number. The LAPLACE distribution arises as the distribution of the difference of two independent and identically distributed exponential variates, too. The distribution is also known as the **first law of LAPLACE**. Still other names of this distribution are **double, bilateral or two–tailed exponential distribution**[20] because it is a combination of an exponential distribution (on the right–hand side of $a$) and a reflected exponential distribution (on the left–hand side of $a$). Thus, the LAPLACE DF has a cusp at $a$, see Fig. 5/25.

We list the following related distributions:

- When $X \sim LA(a,b)$ then
  - $Y = \dfrac{|X-a|}{b} \sim EX(0,1)$,
  - $V = |X-a| \sim EX(0,b)$.
- When $X_1,\ X_2 \overset{\text{iid}}{\sim} LA(a,b)$, then $X = |X_1|\big/|X_2|$ is $F$–distributed with $\nu_1 = \nu_2 = 2$ degrees of freedom.
- When $X_1,\ X_2 \overset{\text{iid}}{\sim} EX(0,b)$, then $X = X_1 - X_2 \sim LA(0,b)$.

We have the following functions and characteristics of the general LAPLACE distribution:

$$f(x|a,b) = \frac{1}{2\,b}\exp\left[-\frac{|x-a|}{b}\right];\ \ x\in\mathbb{R},\ a\in\mathbb{R},\ b>0 \tag{5.24a}$$

$$F(x|a,b) = \frac{1}{2}\left\{1+\text{sign}(x-a)\left[1-\exp\left(-\frac{|x-a|}{b}\right)\right]\right\} \tag{5.24b}$$

$$R(x|a,b) = 1-\frac{1}{2}\left\{1+\text{sign}(x-a)\left[1-\exp\left(-\frac{|x-a|}{b}\right)\right]\right\} \tag{5.24c}$$

$$h(x|a,b) = \frac{\exp\left(-\dfrac{|x-a|}{b}\right)}{b\left\langle 2-\left\{1+\text{sign}(x-a)\left[1-\exp\left(-\dfrac{|x-a|}{b}\right)\right]\right\}\right\rangle} \tag{5.24d}$$

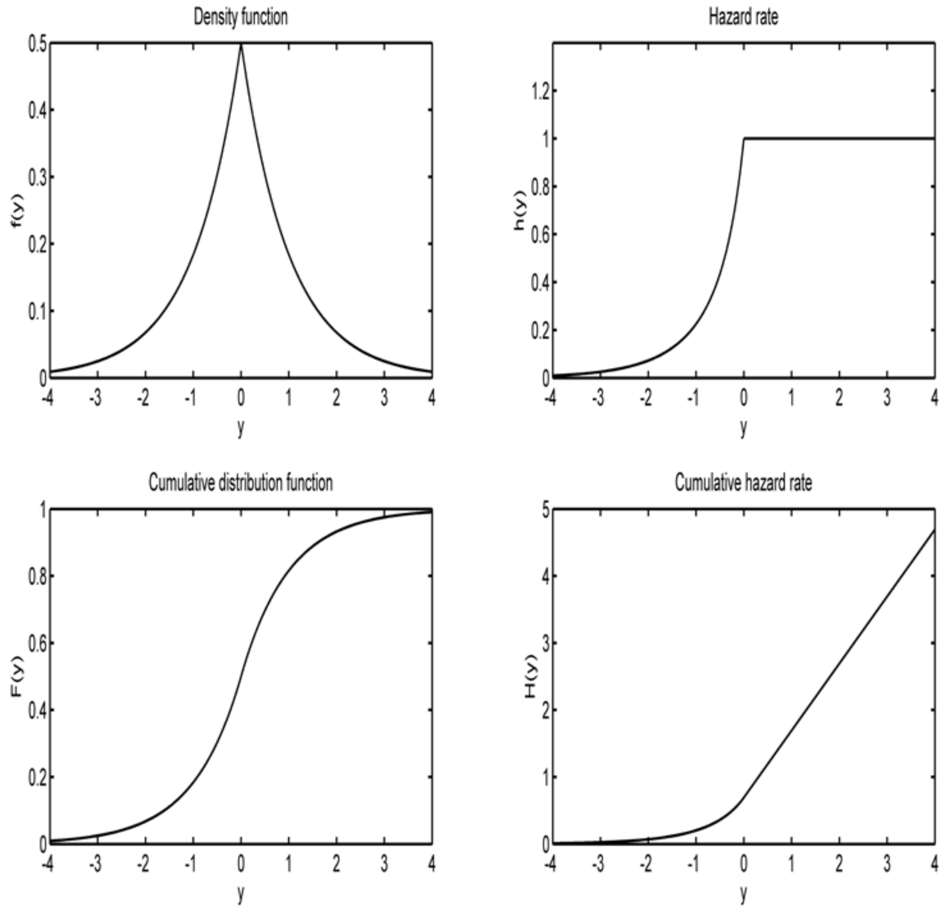
$$H(x|a,b) = -\ln\left\langle 1-\frac{1}{2}\left\{1+\text{sign}(x-a)\left[1-\exp\left(-\frac{|x-a|}{b}\right)\right]\right\}\right\rangle \tag{5.24e}$$

The hazard rate is increasing for $x<a$ and constant $(=1\big/b)$ for $x\geq a$.

---

[20] The double exponential distribution must not be confused with the **doubly exponential distribution** which is another name for the type I extreme value distributions which are exponentiated twice.

Figure 5/25: Several functions for the reduced LAPLACE distribution




$$
\begin{aligned}
F_X^{-1}(P) &= x_P = a - b\,\mathrm{sign}(P-0.5)\,\ln\big[1 - 2\,|P-0.5|\big],\;\; 0 < P < 1 && \text{(5.24f)}\\
a &= x_{0.5} && \text{(5.24g)}\\
b &\approx x_{0.8161} - a && \text{(5.24h)}\\
x_M &= a && \text{(5.24i)}\\
M_X(t) &= \exp(a\,t)\,\big(1 - b^2\,t^2\big)^{-1} && \text{(5.24j)}\\
C_X(t) &= \exp(i\,a\,t)\,\big(1 + b^2\,t^2\big)^{-1} && \text{(5.24k)}\\
\mu'_r(X) &= \frac{1}{2}\sum_{i=0}^{r}\binom{r}{i} b^i\,a^{r-i}\,i!\,\big[1 + (-1)^i\big] && \text{(5.24l)}\\
\mu'_1(X) &= \mathrm{E}(X) = a && \text{(5.24m)}\\
\mu_r(X) &= \left\{\begin{array}{cl} 0 & \text{if } r \text{ is odd} \\ r!\,b^r & \text{if } r \text{ is even} \end{array}\right\} && \text{(5.24n)}
\end{aligned}
$$

$$\mu_2(X) = \mathrm{Var}(X) = 2\,b^2 \tag{5.24o}$$

$$\alpha_3 = 0 \tag{5.24p}$$

$$\alpha_4 = 6 \tag{5.24q}$$

$$\kappa_r(X) = \left\{ \begin{array}{cl} 0 & \text{if } r \text{ is odd} \\ 2\,b^r\,(r-1)! & \text{if } r \text{ is even} \end{array} \right\} \tag{5.24r}$$

$$I(X) = \frac{1+\ln(2\,b)}{\ln 2} \tag{5.24s}$$

$$F_Y^{-1}(P) = \left\{ \begin{array}{cl} \ln(2\,P) & \text{for } 0 < P \le 0,5 \\ -\ln[2\,(1-P)] & \text{for } 0.5 \le P < 1 \end{array} \right\} \tag{5.24t}$$

$$f_Y(y_P) = \left\{ \begin{array}{cl} P & \text{for } 0 < P \le 0.5 \\ 1-P & \text{for } 0.5 \le P < 1 \end{array} \right\} \tag{5.24u}$$

$$F_Y^{-1(r)}(P) = \left\{ \begin{array}{cl} \dfrac{(-1)^{r-1}\,(r-1)!}{P^r} & \text{for } 0 < P \le 0.5 \\ \dfrac{(r-1)!}{(1-P)^r} & \text{for } 0.5 \le P < 1 \end{array} \right\},\; r = 1, 2, \ldots \tag{5.24v}$$

**Excursus: GOVINDARAJULU's idea of deriving moments of order statistics from a symmetric distribution**

The moments of order statistics from the LAPLACE distribution have been found by GOVINDARAJULU (1963) using a method that in fact applies to all symmetric distributions. Let $Y_{r:n}$ $(r = 1, \ldots, n)$ denote the order statistics in a sample from a distribution which is symmetric about zero with $F_Y(y)$. Let $V_{r:n}$ $(r = 1, \ldots, n)$ denote the order statistics from the corresponding folded distribution (folded at zero) with CDF $F_V(y) = 2\,F_Y(y) - 1,\; y \ge 0$. Then, GOVINDARAJULU showed that we have the following relations:

$$\alpha_{r:n}^{(k)} = \mathrm{E}\big(Y_{r:n}^k\big) = 2^{-1}\left\{\sum_{i=0}^{r-1}\binom{n}{i}\mathrm{E}\big(V_{r-i:n-i}^k\big) + (-1)^k\sum_{i=r}^{n}\binom{n}{i}\mathrm{E}\big(V_{i-r+1:i}^k\big)\right\},\; 1 \le r \le n, \tag{5.25a}$$

$$\begin{aligned}\alpha_{r,s:n} = \mathrm{E}\big(Y_{r:n}\,Y_{s:n}\big) &= 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}\mathrm{E}\big(V_{r-i:n-i}\,V_{s-i:n-i}\big) - \sum_{i=r}^{s-1}\binom{n}{i}\mathrm{E}\big(V_{i-r+1:i}\,V_{s-i:n-i}\big)\right. \\ &\quad \left. + \sum_{i=s}^{n}\binom{n}{i}\mathrm{E}\big(V_{i-s+1:i}\big)\,\mathrm{E}\big(V_{i-r+1:i}\big)\right\},\; 1 \le r < s \le n.\end{aligned} \tag{5.25b}$$

(5.25a,b) are extremely helpful in the case of $Y \sim LA(0,1)$ because the corresponding folded distribution is the reduced exponential distribution whose moments of order statistics can be given in closed form, see (5.9l–n). Thus we arrive at the following explicit expressions of the moments of the LAPLACE order statistics:

$$\begin{aligned}\alpha_{r:n} = \mathrm{E}\big(Y_{r:n}\big) &= 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}S_1(r-i,n-i)\right.\\ &\quad\left. - \sum_{i=r}^{n}\binom{n}{i}S_1(i-r+1,i)\right\},\ 1 \le r \le n, \qquad (5.26a)\end{aligned}$$

$$\begin{aligned}\alpha_{r:n}^{(2)} = \mathrm{E}\big(Y_{r:n}^2\big) &= 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}S_2(r-i,n-i)\right.\\ &\quad\left. + \sum_{i=r}^{n}\binom{n}{i}S_2(i-r+1,i)\right\},\ 1 \le r \le n, \qquad (5.26b)\end{aligned}$$

$$\begin{aligned}\alpha_{r,s:n} &= 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}S_3(r-i,s-i,n-i)\right.\\ &\quad - \sum_{i=r}^{s-1}\binom{n}{i}S_1(i-r+1,i)\,S_1(s-i,n-i)\\ &\quad\left. + \sum_{i=s}^{n}\binom{n}{i}S_3(i-s+1,i-r+1,i)\right\},\ \ 1 \le r < s \le n, \qquad (5.26c)\end{aligned}$$

where

$$S_1(r,n) = \sum_{i=n-r+1}^{n}\frac{1}{i} = \sum_{j=1}^{r}\frac{1}{n-j+1} \qquad (5.26d)$$

$$S_2(r,n) = \sum_{i=n-r+1}^{n}\frac{1}{i^2} + \big[S_1(r,n)\big]^2 = \sum_{j=1}^{r}\frac{1}{(n-j+1)^2} + \big[S_1(r,n)\big]^2 \qquad (5.26e)$$

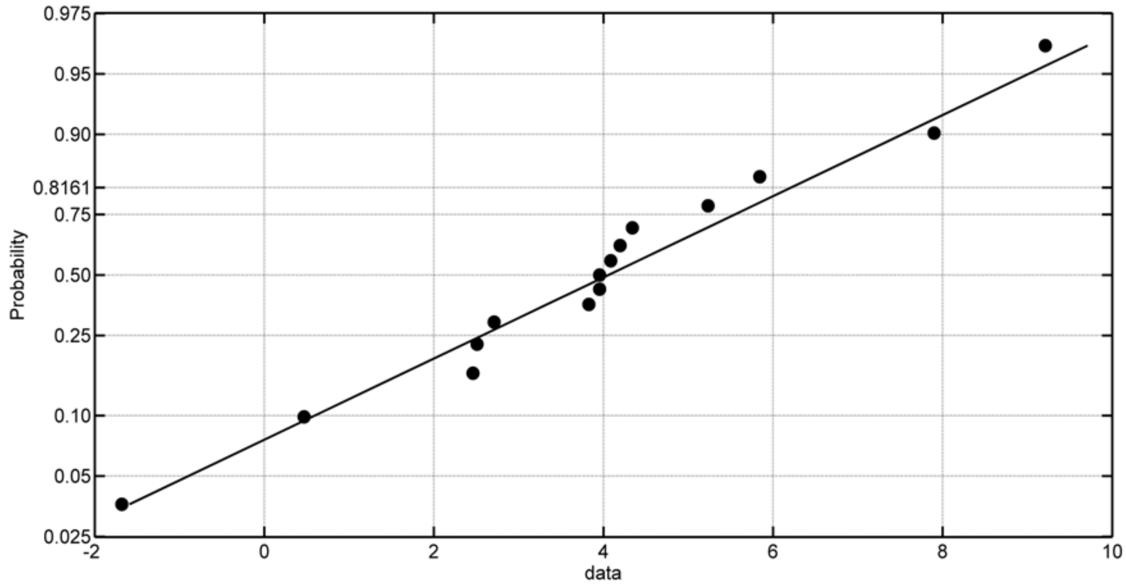
$$S_3(r,s,n) = S_2(r,n) + S_1(r,n)\,S_1(s,n). \qquad (5.26f)$$

The elements of the variance–covariance matrix follow as

$$\left.\begin{aligned}\beta_{r,r:n} &= \alpha_{r:n}^{(2)} - \big[\alpha_{r:n}\big]^2,\ 1 \le r \le n,\\ \beta_{r,s:n} &= \mathrm{E}\big(Y_{r:n}\,Y_{s:n}\big) - \alpha_{r:n}\,\alpha_{s:n},\ 1 \le r < s \le n.\end{aligned}\right\} \qquad (5.26g)$$

Linear estimation in LEPP for order statistics input is done by LLOYD's estimator with means and variance-covariance matrix computed by (5.26a–g).

Figure 5/26: LAPLACE probability paper with data and regression line




### 5.2.9 Logistic distribution — $X \sim LO(a,b)$

The logistic distribution[21] shares many properties with the normal distribution, both have the same support $\mathbb{R}$ and are symmetric and bell–shaped. Contrary to the normal distribution the logistic distribution has explicit formulas for the CDF and the percentiles, thus, it is a strong competitor to the normal distribution. But the logistic distribution has a greater kurtosis $(\alpha_4 = 4.2)$ compared with $\alpha_4 = 3$ of the normal distribution, i.e. it has a higher peak at $a$ and longer tails. As a consequence we note the following features:

- The normal DF has points of inflection at $x = a \pm b$, while those of the logistic DF are at $x = a \pm \sqrt{3}\,\ln[2+\sqrt{3}\,]\,b/\pi \approx a \pm 0.7261\,b$.
- The probabilities of $X$ to be in the central $2\,\sigma$–interval are $0.6827$ for the normal distribution and $0.7196$ for the logistic distribution.

We mention the following related distributions:

- When $Y_1 \sim EX(0,1)$, then $Y_2 = -\ln\left[\dfrac{\exp(-Y_1)}{1+\exp(-Y_1)}\right] \sim LO(0,1)$.
- When $Y_1,\ Y_2 \overset{\text{iid}}{\sim} EX(0,1)$, then $Y = -\ln\big(Y_1/Y_2\big) \sim LO(0,1)$.
- When $Y_i \overset{\text{iid}}{\sim} EMX1(0,1);\ i = 1,\ldots,n$; then $Y = \sum\limits_{i=1}^{n} Y_i/i \sim LO(0,1)$.
- When $X_1, X_2 \overset{\text{iid}}{\sim} EMX1(a,b)$, then $X = X_1 - X_2 \sim LO(0,b)$.

[21] Suggested reading for this section: BALAKRISHNAN (1992), JOHNSON/KOTZ/BALAKRISHNAN (1995, Chapter 23).

- When $Y_i \overset{\text{iid}}{\sim} LO(0,1);\ i = 1,\ldots,n$; then $Y = \Big(\max_i Y_i - \ln n\Big) \sim EMX1(0,1)$.
- When $X \sim PA(0,b,c)$, then $Y = -\ln\left[\left(\dfrac{X}{b}\right)^c - 1\right] \sim LO(0,1)$.
- When $X \sim PO(0,1,c)$, then $Y = -\ln\big(Y^{-c} - 1\big) \sim LO(0,1)$.

Concerning applications of the logistic distribution the interested reader is referred to BALKRISHNAN (1992).

The DF and CDF of the logistic distribution can be written in different ways and — as a consequence — the CCDF, hazard function and cumulative hazard function, too.

$$f(x|a,b)\left\{\begin{aligned} &= \frac{\exp\left(\dfrac{x-a}{b}\right)}{b\left[1+\exp\left(\dfrac{x-a}{b}\right)\right]^2},\quad a\in\mathbb{R},\ b>0,\ x\in\mathbb{R}\\ &= \frac{\exp\left(-\dfrac{x-a}{b}\right)}{b\left[1+\exp\left(-\dfrac{x-a}{b}\right)\right]^2}\\ &= \frac{1}{4\,b}\,\text{sech}^2\left(\frac{x-a}{2\,b}\right)\end{aligned}\right\} \tag{5.27a}$$

$$F(x|a,b)\left\{\begin{aligned} &= 1-\left[1+\exp\left(\frac{x-a}{b}\right)\right]^{-1}\\ &= \left[1+\exp\left(-\frac{x-a}{b}\right)\right]^{-1}\\ &= \frac{1}{2}\left[1+\tanh\left(\frac{x-a}{b}\right)\right]\end{aligned}\right\} \tag{5.27b}$$

The DF and CDF are related as

$$f(x|a,b) = \frac{1}{b}\,F(x|a,b)\,\big[1-F(x|a,b)\big]. \tag{5.27c}$$

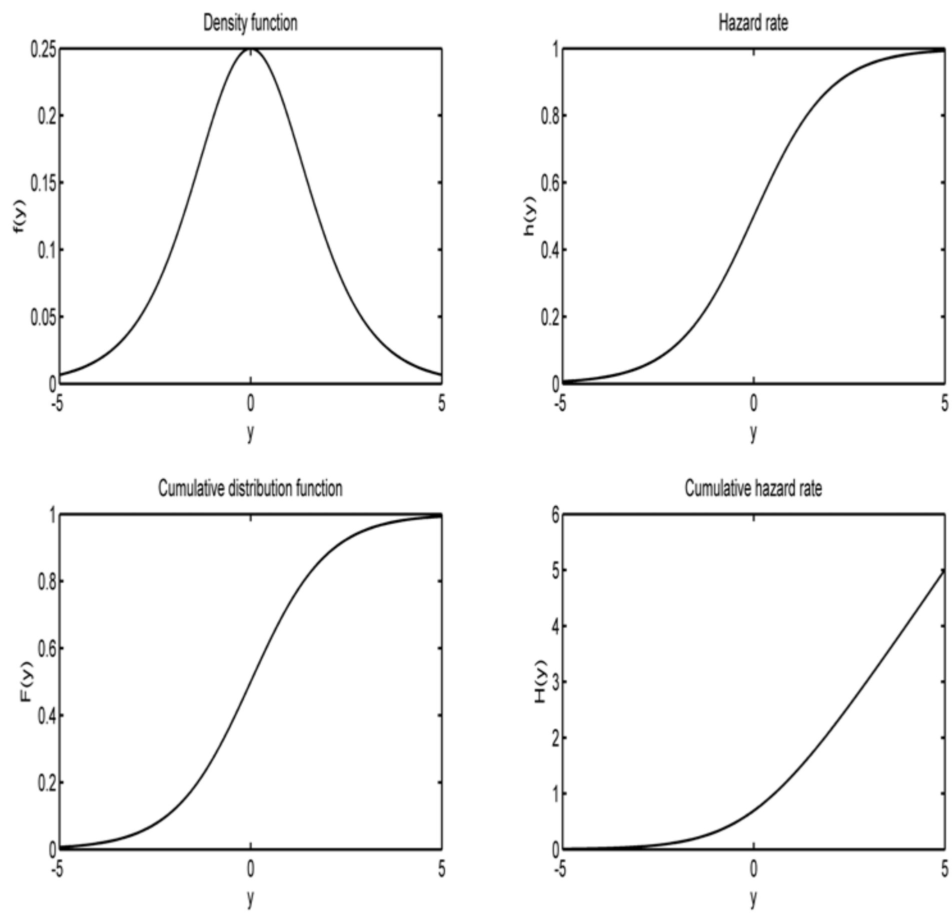
$$R(x|a,b)\left\{\begin{aligned} &= \left[1+\exp\left(\frac{x-a}{b}\right)\right]^{-1}\\ &= \frac{\exp\left(-\dfrac{x-a}{b}\right)}{1+\exp\left(-\dfrac{x-a}{b}\right)}\\ &= \frac{1}{2}\left[1-\tanh\left(\frac{x-a}{b}\right)\right]\end{aligned}\right\} \tag{5.27d}$$

$$h(x|a,b) = \frac{1}{b}\left[1+\exp\left(-\frac{x-a}{b}\right)\right]^{-1} \tag{5.27e}$$

The hazard function is proportional to the CDF.

$$H(x|a,b) = \ln\left[1+\exp\left(\frac{x-a}{b}\right)\right] \tag{5.27f}$$

Figure 5/27: Several functions for the reduced logistic distribution




$$F_X^{-1}(P) \;=\; x_P \;=\; a + b\,\ln\left(\frac{P}{1-P}\right),\;\; 0 < P < 1 \text{ [22]} \tag{5.27g}$$

$$a \;=\; x_{0.5} \tag{5.27h}$$

$$b \;\approx\; x_{0.7311} - a \tag{5.27i}$$

$$x_M \;=\; a \tag{5.27j}$$

[22] The term $\ln[P/(1-P)]$ is called **logit**.

$$M_X(t) \left\{ \begin{aligned} &= \exp(a\,t)\,B(1-b\,t, 1+b\,t)\ ^{23} \\ &= \exp(a\,t)\,\Gamma(1-b\,t)\,\Gamma(1+b\,t) \\ &= \exp(a\,t)\,\pi\,b\,t\ \csc(\pi\,b\,t) \end{aligned} \right\} \tag{5.27k}$$

$$C_X(t) = \exp(i\,a\,t)\,\pi\,b\,i\,t\ \csc(\pi\,b\,i\,t) \tag{5.27l}$$

$$\mu_1'(X) = \mathrm{E}(X) = a \tag{5.27m}$$

$$\mu_2(X) = \mathrm{Var}(X) = b^2\,\frac{\pi^2}{3} \tag{5.27n}$$

$$\alpha_3 = 0 \tag{5.27o}$$

$$\alpha_4 = 4.2 \tag{5.27p}$$

$$\kappa_r(X) = \left\{ \begin{array}{cl} 0 & \text{if } r \text{ is odd} \\ 6\,b^r\,(2^r-1)\,\mathfrak{B}_r\ ^{24} & \text{if } r \text{ is even} \end{array} \right\} \tag{5.27q}$$

$$I(X) = \frac{1}{\ln 2}\big(\ln b + 2\big) \tag{5.27r}$$

$$F_Y^{-1}(P) = y_P = \ln\!\left(\frac{P}{1-P}\right),\quad 0 < P < 1 \tag{5.27s}$$

$$f_Y(y_P) = P\,(1-P) \tag{5.27t}$$

$$F_Y^{-1(r)}(P) = (r-1)!\left[(-1)^{r-1}\frac{1}{P^r} + \frac{1}{(1-P)^r}\right],\quad r = 1, 2, \ldots \tag{5.27u}$$

With respect to the reduced order statistics we have the following results:

$$\mathrm{E}\big(e^{t\,Y_{r:n}}\big) = \frac{\Gamma(r+t)\,\Gamma(n-r+1-t)}{\Gamma(r)\,\Gamma(n-r+1)},\quad r \le r \le n, \tag{5.28a}$$

$$\alpha_{r:n} = \psi(r) - \psi(n-r+1),\ ^{25} \tag{5.28b}$$

$$\beta_{r,r:n} = \psi'(r) + \psi'(n-r+1).\ ^{26} \tag{5.28c}$$

The product moments $\alpha_{r,s:n} = \mathrm{E}(Y_{r:n}\,Y_{s:n}),\ 1 \le r < s \le n$, have to be computed via the following recursion formulas taken from BALAKRISHNAN/RAO (eds.) (1998a, 91–102):

[23] $B(\cdot,\cdot)$ is the complete beta function.

[24] $\mathfrak{B}_r$ is the $r$–th **BERNOULLI number**, see ABRAMOWITZ/STEGUN (1965, p. 804).

[25] $\psi(\cdot)$ is the **digamma function**.

[26] $\psi'(\cdot)$ is the **trigamma function**.

$$\alpha_{r,r+1:n+1} = \frac{n+1}{n-r+1}\left[\alpha_{r,r+r:n} - \frac{r}{n+1}\,\alpha^{(2)}_{r+1:n+1} - \frac{1}{n-r}\,\alpha_{r:n}\right],\ 1 \le r \le n-1, \tag{5.29a}$$

$$\alpha_{n,n+1:n+1} = \frac{n+1}{n}\left[\frac{1}{n-1}\,\alpha_{n:n} + \alpha_{n-1,n:n} - \frac{1}{n+1}\,\alpha^{(2)}_{n:n+1}\right],\ n \ge 2 \tag{5.29b}$$

$$\alpha_{r,s:n+1} = \alpha_{r,s-1:n+1} + \frac{n+1}{n-s+2}\left[\alpha_{r,s:n} - \alpha_{r,s-1:n} - \frac{1}{n-s+1}\alpha_{r:n}\right], \tag{5.29c}$$
$$1 \le r < s \le n,\ s-r \ge 2,$$

$$\alpha_{r+1,n+1:n+1} = \alpha_{r+2,n+1:n+1} + \frac{n+1}{r+1}\left[\frac{1}{r}\,\alpha_{n:n} + \alpha_{r,n:n} - \alpha_{r-1,n:n}\right], 1 \le r \le n-2, \tag{5.29d}$$

$$\alpha_{1,n+1:n+1} = \alpha_{1,n:n+1} - \frac{1}{n-1}\left[\alpha_{2,n+1:n+1} - \alpha_{2,n:n+1}\right] + \frac{n+1}{n-1}\,\alpha_{1:n},\ n \ge 3. \tag{5.29e}$$

A special formula is needed for $\alpha_{1,3:3}$:

$$\alpha_{1,3:3} = \alpha_{1,2:3} - \alpha^{(2)}_{2:3} + 3\,\alpha_{1:2}. \tag{5.29f}$$

The single moments of order one are directly given by (5.28b) and those of order two are derived from (5.28c) using (5.28b) as

$$\alpha^{(2)}_{r:n} = \beta_{r,r:n} - \alpha^2_{r:n}.$$

The starting value for the product moments' recursion is

$$\alpha_{1,2:3} = \alpha^2_{1:1} = 0.$$

For $n+1=3$ we proceed as follows:

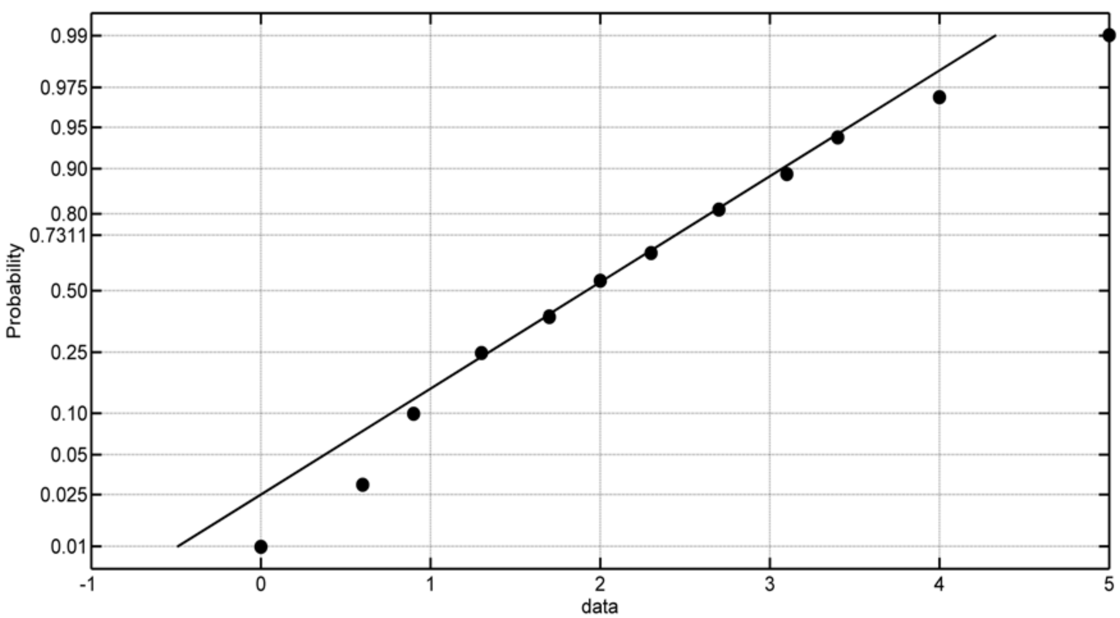
- $\alpha_{1,2:3}$ from (5.29a), then
- $\alpha_{2,3:3}$ from (5.29b) and finally
- $\alpha_{1,3:3}$ from (5.29f).

For $n+1=4$ we compute:

- $\alpha_{1,2:4}$ and $\alpha_{2,3:4}$ by (5.29a), then
- $\alpha_{3,4:4}$ by (5.29b), then
- $\alpha_{1,3:4}$ by (5.29c), then
- $\alpha_{2,4:4}$ by (5.29d) and finally
- $\alpha_{1,4:4}$ by (5.29e).

This process may be followed similarly to determine $\alpha_{r,s:n+1}$ for $n+1 = 5, 6, \ldots$. Unfortunately, we cannot restrict the computation by recursion to only a portion of the upper triangle in the product matrix $(\alpha_{r,s:n})$ as given by (2.13b). The off–diagonal elements of the variance–covariance matrix follow in the usual way: $\beta_{r,s:n} = \alpha_{r,s:n} - \alpha_{r:n}\,\alpha_{s:n}$. To prevent the results to be distorted by the cumulation of rounding errors in the course of recursion we have decided in LEPP to only apply the recursion formulas up to sample size $35$. For $n > 35$ the moments are computed by the approximating formulas (2.23) through (2.25). Tables of covariances are given by SHAH(1966) for $n \leq 10$ and by GUPTA et al. (1967) for $n \leq n \leq 25$.

Figure 5/28: Logistic probability paper with data and regression line




## 5.2.10 MAXWELL–BOLTZMANN distribution — $X \sim MB(a, b)$

This distribution, which is named after the Scottish theoretical physicist and mathematician JAMES CLERK MAXWELL (1831 – 1879) and the Austrian physicist LUDWIG EDUARD BOLTZMANN (1844 – 1906), has its main application in chemistry and physics where it describes the distribution of speeds of molecules in thermal equilibrium. In statistics the distribution can be thought of as the magnitude of a random three–dimensional vector whose components $X_i$; $i = 1, 2, 3$; are independent and distributed as $NO(0, b)$. Then $X = \sqrt{X_1^2 + X_2^2 + X_3^2} \sim MB(0, b)$. Whereas in this case the sum $X_1 + X_2 + X_3$ has a $\chi^2$–distribution with $\nu = 3$ degrees of freedom, the variate $X$ has a **$\chi$–distribution** with $\nu = 3$ degrees of freedom.[27] Thus, the results for the MAXWELL–BOLTZMANN

[27] Other special cases of the $\chi$–distribution are the **half–normal distribution** ($\nu = 1$) and the **RAYLEIGH distribution** ($\nu = 2$).

distribution follow from those of a $\chi$–distribution.

$$f(x|a,b) = \frac{1}{b}\sqrt{\frac{2}{\pi}}\left(\frac{x-a}{b}\right)^2 \exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right], \quad x \ge a,\ a \in \mathbb{R},\ b > 0 \tag{5.30a}$$

$$F(x|a,b) = \frac{2}{\sqrt{\pi}}\,\gamma\left[\frac{3}{2}\,\middle|\,\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right] \tag{5.30b}$$

$\gamma(z|u) = \int\limits_0^u e^{-t}\,t^{z-1}\,\mathrm{d}t$ is the **incomplete gamma function**. As the MAXWELL–BOLTZMANN distribution is nothing but the $\chi_3$–distribution its CDF can be given in terms of the CDF of a $\chi_3^2$ variate. Let $F_{\chi_3^2}(\cdot)$ denote the CDF of $\chi_3^2$ then $F_{\chi_3}(z) = F_{\chi_3^2}(z^2)$. This feature is extremely useful because a routine for the $\chi_\nu^2$–CDF is implemented in MATLAB as well as in other statistical software packages. We can thus express (5.30b) as

$$F(x|a,b) = F_{\chi_3^2}\left[\left(\frac{x-a}{b}\right)^2\right]. \tag{5.30c}$$

It is also possible to write the MAXWELL–BOLTZMANN CDF in terms of the **error function**

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}}\int\limits_0^z \exp(-t^2)\,\mathrm{d}t$$

as

$$F(x|a,b) = \text{erf}\left(\frac{x-a}{b\sqrt{2}}\right) - \frac{x-a}{b}\,\exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right]\sqrt{\frac{2}{\pi}}. \tag{5.30d}$$

Because the error function and the CDF $\Phi(z)$ of the standard normal distribution are linked as

$$\text{erf}(z) = 2\,\Phi(z\sqrt{2}) - 1$$

we may even write the MAXWELL–BOLTZMANN CDF in terms of the normal CDF:

$$F(x|a,b) = 2\,\Phi\left(\frac{x-a}{b}\right) - \frac{x-a}{b}\,\exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right]\sqrt{\frac{2}{\pi}} - 1. \tag{5.30e}$$

The other functions and characteristics of the MAXWELL–BOLTZMANN distribution are:

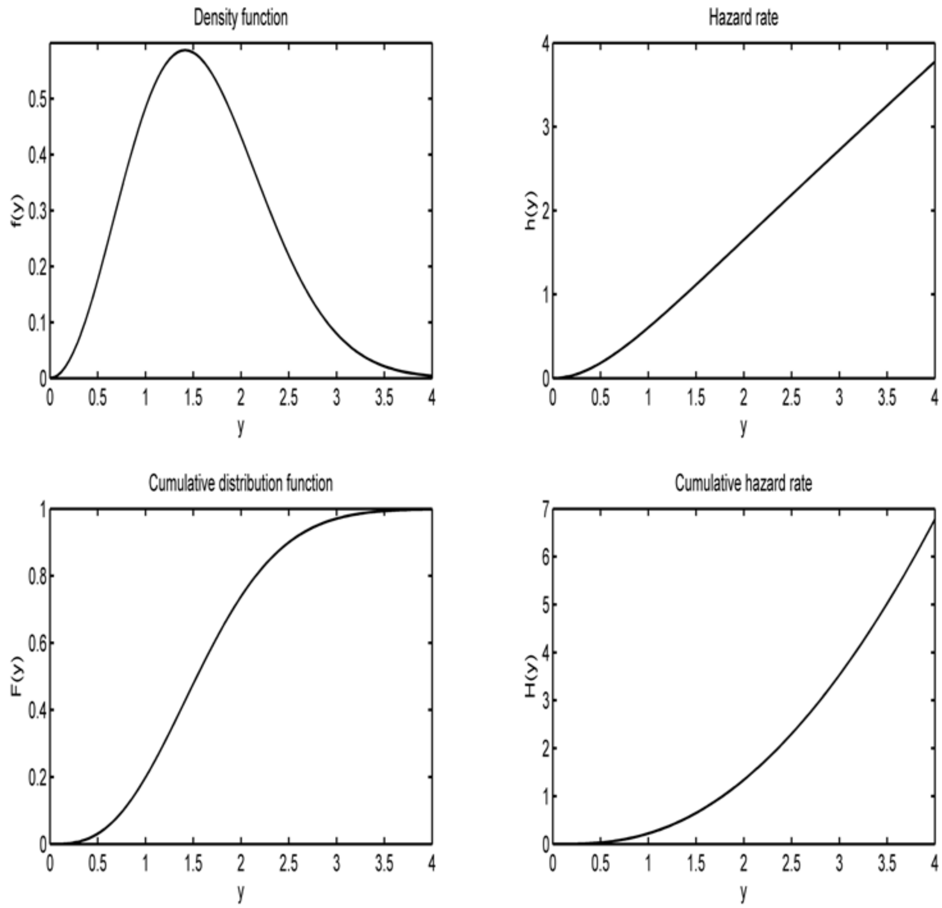
$$R(x|a,b) = 1 - F_{\chi_3^2}\left[\left(\frac{x-a}{b}\right)^2\right] \tag{5.30f}$$

$$h(x|a,b) = \frac{f(x|a,b)}{R(x|a,b)} \tag{5.30g}$$

$$H(x|a,b) = -\ln\left[R(x|a,b)\right] \tag{5.30h}$$

The hazard function $h(x|a,b)$ and the cumulative hazard function $H(x|a,b)$ cannot be given in closed form.

Figure 5/29: Several functions for the reduced MAXWELL–BOLTZMANN distribution




The percentiles of the MAXWELL–BOLTZMANN distribution can be expressed by the $\chi_3^2$–percentiles $\chi_3^2(P)$ as:[28]

$$F_X^{-1}(P) \;=\; x_P \;=\; a + b\sqrt{\chi_3^2(P)} \;=\; a + b\,\chi_3(P), \;\; 0 \le P < 1. \tag{5.30i}$$

$$a \;=\; x_0 \tag{5.30j}$$

$$b \;\approx\; x_{0.1987} - a \tag{5.30k}$$

$$x_M \;=\; a + b\sqrt{2} \;\approx\; a + 1.4142\,b \tag{5.30l}$$

$$C_X(t) = i\exp(iat)\left\{bt\sqrt{\frac{2}{\pi}} - \exp\left(-\frac{b^2t^2}{2}\right)(b^2t^2-1)\left[\text{sign}(t)\,\text{erfi}\left(\frac{b|t|}{\sqrt{2}}\right) - i\right]\right\}\;^{29} \tag{5.30m}$$

[28] The percentile function or the inverse CDF of $\chi_\nu^2$ is also implemented in MATLAB and other statistical software.

[29] $\text{erfi}(z) = i\,\text{erf}(i\,z)$ is the **imaginary error function**.

$$\mu'_r(Y) = 2^{1+r/2}\,\Gamma\left(\frac{3+r}{2}\right) \Big/ \sqrt{\pi};\quad r = 1, 2, \ldots \tag{5.30n}$$

$$\mu'_1(Y) = 2\sqrt{\frac{2}{\pi}} \approx 1.5958 \tag{5.30o}$$

$$\mu'_2(Y) = 3 \tag{5.30p}$$

$$\mu'_3(Y) = 8\sqrt{\frac{2}{\pi}} \approx 6.3831 \tag{5.30q}$$

$$\mu'_4(Y) = 15 \tag{5.30r}$$

$$\mu'_1(X) = \mathrm{E}(X) = a + 2\,b\sqrt{\frac{2}{\pi}} \approx a + 1.5958\,b \tag{5.30s}$$

$$\mu_2(X) = \mathrm{Var}(X) = b^2\,\frac{3\,\pi - 8}{\pi} \approx 0.4535\,b^2 \tag{5.30t}$$

$$\alpha_3 = \frac{2\sqrt{2}\,(16 - 5\,\pi)}{(3\,\pi - 8)^{3/2}} \approx 0.4857 \tag{5.30u}$$

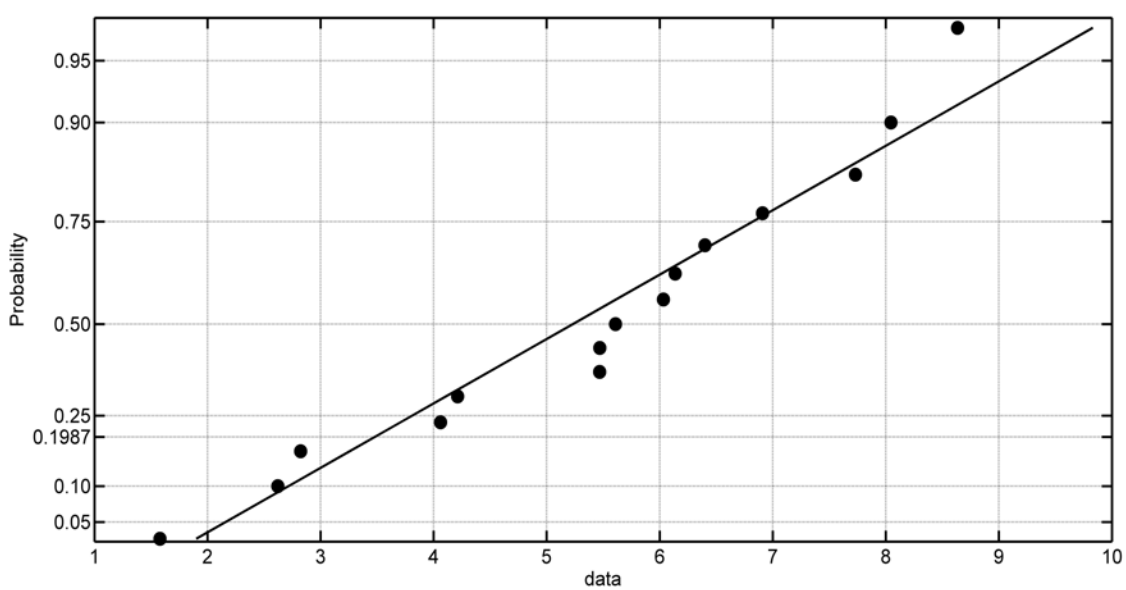
$$\alpha_4 = \frac{15\,\pi^2 + 16\,\pi - 192}{(3\,\pi - 8)^2} \approx 3.1082 \tag{5.30v}$$

$$F_Y^{-1}(P) = \chi_3(P),\quad 0 \le P < 1 \tag{5.30w}$$

$$f_Y(y_P) = \sqrt{\frac{2}{\pi}}\,\chi_3^2(P)\,\exp\left[-\frac{\chi_3^2(P)}{2}\right] \tag{5.30x}$$

In LEPP linear estimation for order statistics input is done by GLS with percentiles of the empirical CDF as regressors and the asymptotic variance–covariance matrix (4.53c,d).

Figure 5/30: MAXWELL–BOLTZMANN probability paper with data and regression line



### 5.2.11 Normal distribution — $X \sim NO(a,b)$

The normal distribution[30] — sometimes called **GAUSS distribution** — is the most popular distribution in statistics where it plays a central role.

- Many distributions converge to a normal distribution for certain limit processes of their parameters.
- A great number of sample statistics, the most import being the sum, the mean and the portion, are exactly or approximately normally distributed.
- Distributions like the $\chi^2$–, the $t$– and the $F$–distributions, which are important in statistical inference, are based on the the normal distribution.

The earliest published derivation of the normal distribution as an approximation to a binomial distribution, when $n \to \infty$ and $P$ held constant, is a pamphlet of 1733 written by ABRAHAM DE MOIVRE (1667 – 1754). In the nineteenth century and the beginning of the twentieth century the normal distribution was established on forms of **central limit theorems**, e.g. by CARL FRIEDRICH GAUSS (1777 – 1855) as the resultant of a large number of additive and independent errors when doing astronomical observations, and in a more rigorous way by A.M. LJAPUNOV (1857 – 1918), P. LÉVY (1896 – 1971), W. FELLER (1906 – 1970), J.W. LINDEBERG (1876 – 1932) and B.V. GNEDENKO (1912 – 1995). Generally, a central limit theorem states conditions under which the distribution of the standardized sum of random variables tends to the unit normal distribution $NO(0,1)$ as the number variables in the sum increases.

Among the great number of distributions related to the normal we only mention the following ones:

- The normal distribution is **reproductive with respect to summation**, i.e. $X_i \overset{\text{iid}}{\sim} NO(a_i, b_i);\ i = 1, 2, \ldots, n$; then $X = c \pm \sum\limits_{i=1}^{n} d_i\, X_i \sim NO(a,b)$ with $a = c \pm \sum\limits_{i=1}^{n} d_i\, a_i$ and $b^2 = \sum\limits_{i=1}^{n} d_i^2\, b_i^2$.
- $Z_1,\ Z_2 \overset{\text{iid}}{\sim} NO(0,1)$, then
  - $X = Z_1\, Z_2$ is distributed with DF $f(x) = \dfrac{1}{\pi} \int\limits_0^{\infty} \left(1 + t^2\right) \cos(x\, t)\, \mathrm{d}t$,
  - $X = Z_1 \big/ Z_2 \sim CA(0,1)$.
- $X_i \overset{\text{iid}}{\sim} NO(a,b);\ i = 1, 2, \ldots, n$; and independent, then
  - $X = \sum\limits_{i=1}^{n} \left(\dfrac{X_i - a}{b}\right)^2 \sim \chi^2_n$,

[30] There exist numerous monographs on this distribution. A good overview is JOHNSON/KOTZ/BALAKRISHNAN (1994, Chapter 13).

- $X = \sum\limits_{i=1}^{n} \left( \dfrac{X_i - \overline{X}}{b} \right)^2 \sim \chi^2_{n-1}$, where $\overline{X} = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$.
- Furthermore: $\chi^2_\nu \to NO(\nu, 2\,\nu)$ for $\nu \to \infty$.

- $X_i \overset{\text{iid}}{\sim} NO(a, b);\ i = 1, 2, \ldots, n;$, then
  - $X = \dfrac{\overline{X} - a}{S} \sqrt{n} \sim t_{n-1}$ ($t$–distribution with $n-1$ degrees of freedom), where
    $S = \sqrt{\dfrac{1}{n-1} \sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2}$.
  - Furthermore: $t_\nu \to NO(0, 1)$ for $n \to \infty$.

- $X_i \overset{\text{iid}}{\sim} NO(a_X, b_X);\ i = 1, 2, \ldots, n;\ Y_j \overset{\text{iid}}{\sim} NO(a_Y, b_Y);\ j = 1, 2, \ldots, m;$ both series independent, then $X = \dfrac{\dfrac{1}{n-1} \sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{\dfrac{1}{m-1} \sum\limits_{j=1}^{m} \left(Y_j - \overline{Y}\right)^2} \sim F_{n-1,m-1}$.
  ($F_{\nu_1,\nu_2}$ is the $F$– or **FISHER distribution** with $\nu_1, \nu_2$ degrees of freedom.)

- $X$ binomially distributed with parameters $n$ and $P$, then $Z = \dfrac{X - n\,P}{\sqrt{n\,P\,(1-P)}} \overset{\text{approx}}{\sim} NO(0, 1)$ for $n \to \infty$.

- $X$ POISSON distributed with parameter $\lambda$, then $Z = \dfrac{X - \lambda}{\sqrt{\lambda}} \overset{\text{approx}}{\sim} NO(0, 1)$ for $\lambda \to \infty$.

- Relation to lognormal distributions
  - $X_1 \sim LNL(a, b, c)$, then $V_1 = \ln(X_1 - a) \sim NO(b, c)$,
  - $X_2 \sim LNU(a, b, c)$, then $V_2 = \ln(a - X_2) \sim NO(b, c)$.
  - Furthermore, $LNL(0, b, c)$ and $LNU(0, b, c) \to NO(b, c)$ for $c \to 0$.

Functions and characteristics of the normal distribution are as follows:

$$f(x|a, b) = \frac{1}{b\sqrt{2\,\pi}} \exp\left[-\frac{(x-a)^2}{2\,b^2}\right];\ \ x \in \mathbb{R},\ a \in \mathbb{R},\ b > 0 \tag{5.31a}$$

$$\varphi(z) = \frac{1}{\sqrt{2\,\pi}} \exp\left(-\frac{z^2}{2}\right) \text{ – DF of the \textbf{standardized} normal distribution} \tag{5.31b}$$

$$\varphi(z) = \varphi(-z) \tag{5.31c}$$

$$f(x|a, b) = \frac{1}{b}\,\varphi\left(\frac{x-a}{b}\right) \tag{5.31d}$$

$$\varphi(z) = b\,f(a + b\,z|a, b) \tag{5.31e}$$

$$F(x|a,b) = \int\limits_{-\infty}^{x} f(v|a,b)\,\mathrm{d}v - \text{no closed form possible} \tag{5.31f}$$

$$\Phi(z) = \int\limits_{-\infty}^{z} \varphi(t)\,\mathrm{d}t - \text{no closed form possible} \tag{5.31g}$$

$$\Phi(z) = 1 - \Phi(-z) \tag{5.31h}$$

$$F(x|a,b) = \Phi\left(\frac{x-a}{b}\right) \tag{5.31i}$$

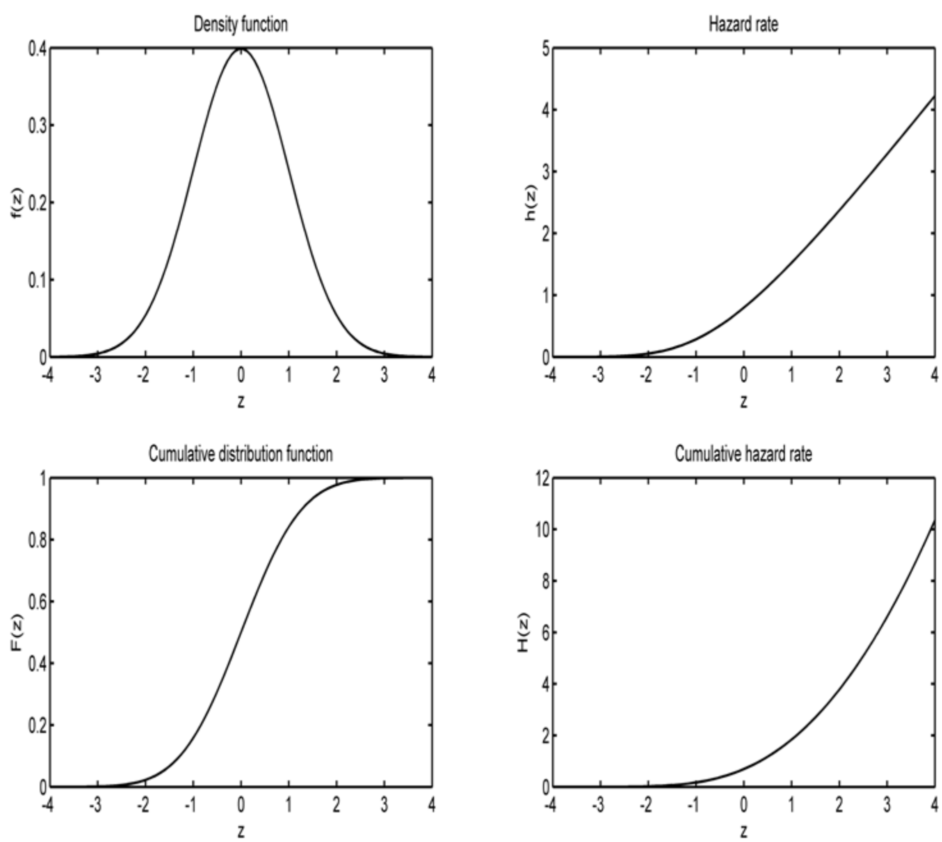
$$\Phi(z) = F(a + b\,z|a,b) \tag{5.31j}$$

$$R(x|a,b) = 1 - \Phi\left(\frac{x-a}{b}\right) \tag{5.31k}$$

$$h(x|a,b) = \frac{f(x|a,b)}{R(x|a,b)} \tag{5.31l}$$

$$H(x|a,b) = -\ln\big[R(x|a,b)\big] \tag{5.31m}$$

Figure 5/31: Several functions for the reduced (standardized) normal distribution



$$x_P = F_X^{-1}(P),\;\; 0 < P < 1,\;\; \text{no closed form possible} \tag{5.31n}$$

$$z_P = \Phi^{-1}(P),\;\; 0 < P < 1,\;\; \text{no closed form possible} \tag{5.31o}$$

$$x_P = a + b\,z_P \tag{5.31p}$$

$$z_P = -z_{1-P}\;\;{}^{31} \tag{5.31q}$$

$$a = x_{0.5} \tag{5.31r}$$

$$b \approx x_{0.8413} - a \tag{5.31s}$$

$$x_M = a \tag{5.31t}$$

$$M_X(t) = \exp\bigl[a\,t + b^2\,t^2/2\bigr] \tag{5.32a}$$

$$C_X(t) = \exp\bigl[i\,a\,t - b^2\,t^2/2\bigr] \tag{5.32b}$$

$$\mu_1'(X) = \mathrm{E}(X) = a \tag{5.32c}$$

$$\mu_2'(X) = a^2 + b^2 \tag{5.32d}$$

$$\mu_3'(X) = 3\,a^2\,b^2 + a^3 \tag{5.32e}$$

$$\mu_4'(X) = 3\,b^4 + 6\,a^2\,b^2 + a^4 \tag{5.32f}$$

$$\mu_r(X) = \left\{\begin{array}{cl} 0 & \text{for } r \text{ odd} \\ b^r\,(r-1)\,(r-3)\cdot\ldots\cdot 3\,,1 & \text{for } r \text{ even} \end{array}\right\} \tag{5.32g}$$

$$\mu_2(X) = \mathrm{Var}(X) = b^2 \tag{5.32h}$$

$$\mu_4(X) = 3\,b^4 \tag{5.32i}$$

$$\mu_6(X) = 15\,b^6 \tag{5.32j}$$

$$\alpha_3 = 0 \tag{5.32k}$$

$$\alpha_4 = 3 \tag{5.32l}$$

$$\kappa_1(X) = a \tag{5.32m}$$

$$\kappa_2(X) = b^2 \tag{5.32n}$$

$$\kappa_r(X) = 0 \;\text{ for }\; r \ge 3 \tag{5.32o}$$

$$I(X) = \mathrm{ld}\bigl[b\sqrt{2\,\pi\,e}\,\bigr] \tag{5.32p}$$

$$\Phi^{-1}(P) = z_P,\;\; 0 < P < 1 \tag{5.32q}$$

[31] $z_P$ — sometimes $5 + z_P$ to avoid negative values — is called **probit**.

$$\left.\begin{aligned}
\Phi^{-1(1)}(P) &= \frac{1}{\varphi\big[\Phi^{-1}(P)\big]},\\
\Phi^{-1(2)}(P) &= \frac{\Phi^{-1}(P)}{\Big\{\varphi\big[\Phi^{-1}(P)\big]\Big\}^2},\\
\Phi^{-1(3)}(P) &= \frac{1+2\left\{\Phi^{-1}(P)\right\}^2}{\Big\{\varphi\big[\Phi^{-1}(P)\big]\Big\}^3},\\
\Phi^{-1(4)}(P) &= \frac{\Phi^{-1}(P)\Big\{7+6\left[\Phi^{-1}(P)\right]^2\Big\}}{\Big\{\varphi\big[\Phi^{-1}(P)\big]\Big\}^4},\\
\Phi^{-1(5)}(P) &= \frac{7+46\left[\Phi^{-1}(P)\right]^2+24\left[\Phi^{-1}(P)\right]^4}{\Big\{\varphi\big[\Phi^{-1}(P)\big]\Big\}^5},\\
\Phi^{-1(6)}(P) &= \frac{\Phi^{-1}(P)\Big\{127+326\left[\Phi^{-1}(P)\right]^2+120\left[\Phi^{-1}(P)\right]^4\Big\}}{\Big\{\varphi\big[\Phi^{-1}(P)\big]\Big\}^6}
\end{aligned}\right\} \tag{5.32r}$$

For smaller sample sizes the single and product moments of normal order statistics can be given in terms of some elementary functions.

$\boldsymbol{n=2}$

$$\left.\begin{aligned}
&\alpha_{2:2} = -\alpha_{1:2} = \frac{1}{\sqrt{\pi}} \approx 0.5641\\
&\alpha_{1,2:2} = 0
\end{aligned}\right\} \tag{5.33a}$$

$\boldsymbol{n=3}$

$$\left.\begin{aligned}
&\alpha_{3:3} = -\alpha_{1:3} = \frac{3}{2\sqrt{\pi}} \approx 0.8463, \quad \alpha_{2:3} = 0,\\
&\alpha_{1,2:3} = \alpha_{2,3:3} = \frac{\sqrt{3}}{2\,\pi} \approx 0.2757, \quad \alpha_{1,3:3} = -\frac{\sqrt{3}}{\pi} \approx -0.5513
\end{aligned}\right\} \tag{5.33b}$$

$\boldsymbol{n=4}$

$$\left.\begin{aligned}
&\alpha_{4:4} = -\alpha_{1:4} = \frac{6}{\pi\sqrt{\pi}}\ \arctan(\sqrt{2}\,) \approx= 1.0294,\\
&\alpha_{3:4} = -\alpha_{2:4} = \frac{6}{\sqrt{\pi}} - \frac{18}{\pi\,,\,\sqrt{\pi}}\ \arctan(\sqrt{2}\,) \approx 0.2970,\\
&\alpha_{1,2:4} = \alpha_{3,4:4} = \frac{\sqrt{3}}{\pi} \approx 0.5513,\\
&\alpha_{1,3:4} = \alpha_{2,4:4} = -\frac{\sqrt{3}}{\pi}\,(2-\sqrt{3}\,) \approx -0.1477,\\
&\alpha_{1,4:4} = -\frac{3}{\pi} \approx -0.9549, \quad \alpha_{2,3:4} = \frac{\sqrt{3}}{\pi}\,(2-\sqrt{3}) \approx 0.1477
\end{aligned}\right\} \tag{5.33c}$$

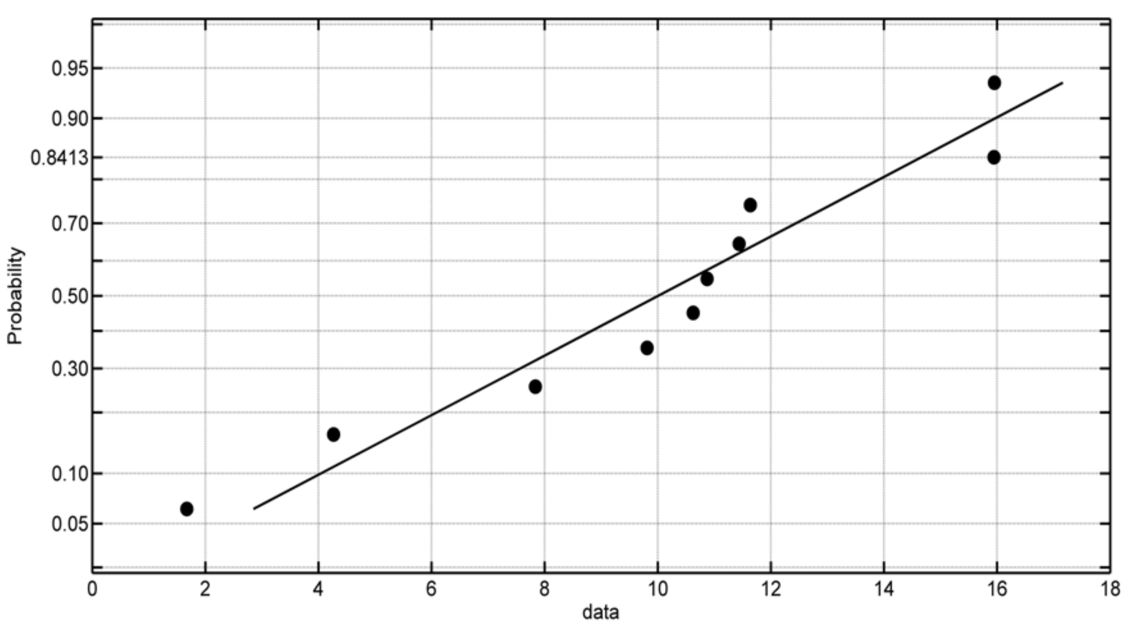
$\boldsymbol{n = 5}$

$$\left.\begin{aligned}
\alpha_{5:5} &= -\alpha_{1:5} = \frac{10}{\sqrt{\pi}}\left(\frac{3}{2\,\pi}\ \arctan(\sqrt{2}\,) - \frac{1}{4}\right) \approx 1.1630 \\
\alpha_{4:5} &= -\alpha_{2:5} = \frac{30}{\sqrt{\pi}}\left(\frac{1}{\pi}\ \arctan(\sqrt{2}\,)\right) - \frac{40}{\sqrt{\pi}}\left(\frac{3}{2\,\pi}\ \arctan(\sqrt{2}\,) - \frac{1}{4}\right) \approx 0.4950 \\
\alpha_{3:5} &= 0
\end{aligned}\right\} \quad (5.33d)$$

The above results for single moments are due to BOSE/GUPTA (1959), but their method fails for sample size six or more. The general approach for the derivation of product moments' expressions is due to GODWIN (1949). A more up–to–date presentation is to be found in BALAKRISHNAN/COHEN (1991, Section 3.9). These authors also give a lot of recursion formulas for single as well as for product moments. ROYSTON (1982) has developed an algorithm for computing the expected values of normal order statistics for sample sizes up to 1000. Among the numerous tables of moments of reduced normal order statistics we mention:

- TEICHROEW (1956) — means and product moments for $n \leq 20$,
- HARTER (1961) — means for $n = 2(1)100(25)250(50)400$,
- SARHAN/GREENBERG (1962) — variances and covariances for $n \leq 20$,
- TIETJEN et. al. (1977) — variances and covariances for $n \leq 50$.

Figure 5/32: Normal probability paper with data and regression line




Concerning normal probability plotting, especially the choice of plotting positions, we mention BARNETT (1976) and BROWN/HETTMANSPERGER (1996). In LEPP linear estimation for order statistics input is realized as follows:

- $3 \leq n \leq 10$ — LLOYD's estimator with tabulated means and variance–covariance matrix,
- $n > 10$ — LLOYD's estimator with approximated means and variance–covariance matrix, using the short versions of the approximating formulas (2.23) – (2.25).

### 5.2.12 Parabolic distributions of order 2

In this section we will present two types of distributions whose DF is a parabola, either opened to the top (Sect. 5.2.12.1) or opened to the bottom (Sect. 5.2.12.2). The U–shaped type of Sect. 5.2.12.1 leads to a power–function distribution when folded about the mean $\mathrm{E}(X) = a$.

#### 5.2.12.1 U–shaped parabolic distribution — $X \sim PAU(a, b)$

Generally, a parabola opened to the top is a power–function of even order: $y = x^{2k}$; $k = 1, 2, \ldots$; and $x \in \mathbb{R}$. To arrive at a parabola which is symmetric around $a \in \mathbb{R}$ and has a scale parameter $b > 0$, we have to modify the power–function to $y = \dfrac{2k+1}{2b}\left(\dfrac{x-a}{b}\right)^{2k}$. We will assume $k$ to be known and to be equal to 1.[32] This distribution is a model for variates with nearly impossible realizations near the mean and high probabilities for outlying realizations.

$$f(x|a,b) = \frac{3}{2b}\left(\frac{x-a}{b}\right)^2, \quad a-b \leq x \leq a+b,\ a \in \mathbb{R},\ b > 0 \tag{5.34a}$$

$$F(x|a,b) = \frac{1}{2}\left[\left(\frac{x-a}{b}\right)^3 + 1\right] \tag{5.34b}$$

$$R(x|a,b) = \frac{1}{2}\left[1 - \left(\frac{x-a}{b}\right)^3\right] \tag{5.34c}$$

$$h(x|a,b) = \frac{3\,(a-x)^2}{b^3 + (a-x)^3} \tag{5.34d}$$

$$H(x|a,b) = \ln 2 - \ln\left[1 - \left(\frac{x-a}{b}\right)^3\right] \tag{5.34e}$$

$$F_X^{-1}(P) = x_P = a + b\,\mathrm{sign}(2P-1)\left(2\,|P-0.5|\right)^{1/3}, \quad 0 \leq P \leq 1 \tag{5.34f}$$

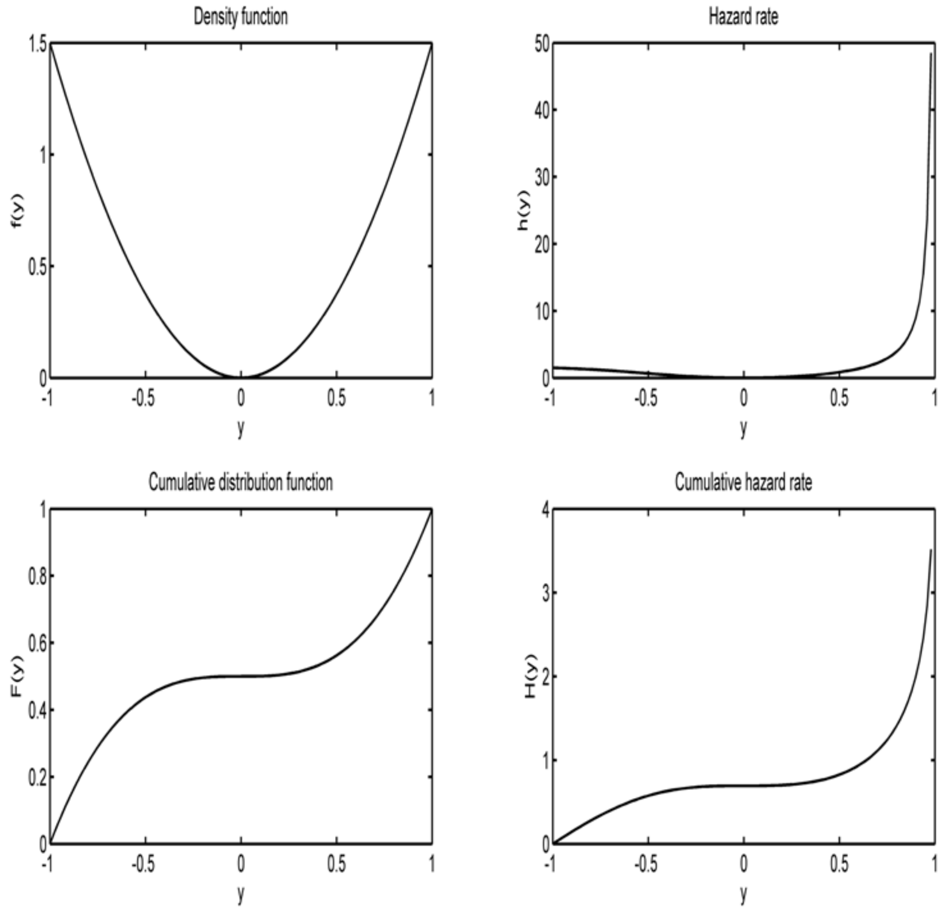
$$a = x_{0.5} \tag{5.34g}$$

$$b = x_1 - a \tag{5.34h}$$

$$x = a \ \text{ is an anti–mode} \tag{5.34i}$$

[32] The greater $k$ the more the DF approaches to the perpendiculars drawn at the edges $a \pm b$ of the support and the more probability mass is moved towards the borders.

Figure 5/33: Several functions for the reduced U–shaped parabolic distribution




$$\mu'_r(Y) = \mu_r(Y) = \frac{3\big[1+(-1)^r\big]}{2\,(3+r)}, \quad Y = \frac{X-a}{b} \tag{5.34j}$$

$$\mu'_1(X) = \mathrm{E}(X) = a \tag{5.34k}$$

$$\mu_2(X) = \mathrm{Var}(X) = 0.6\,b^2 \tag{5.34l}$$

$$\alpha_3 = 0 \tag{5.34m}$$

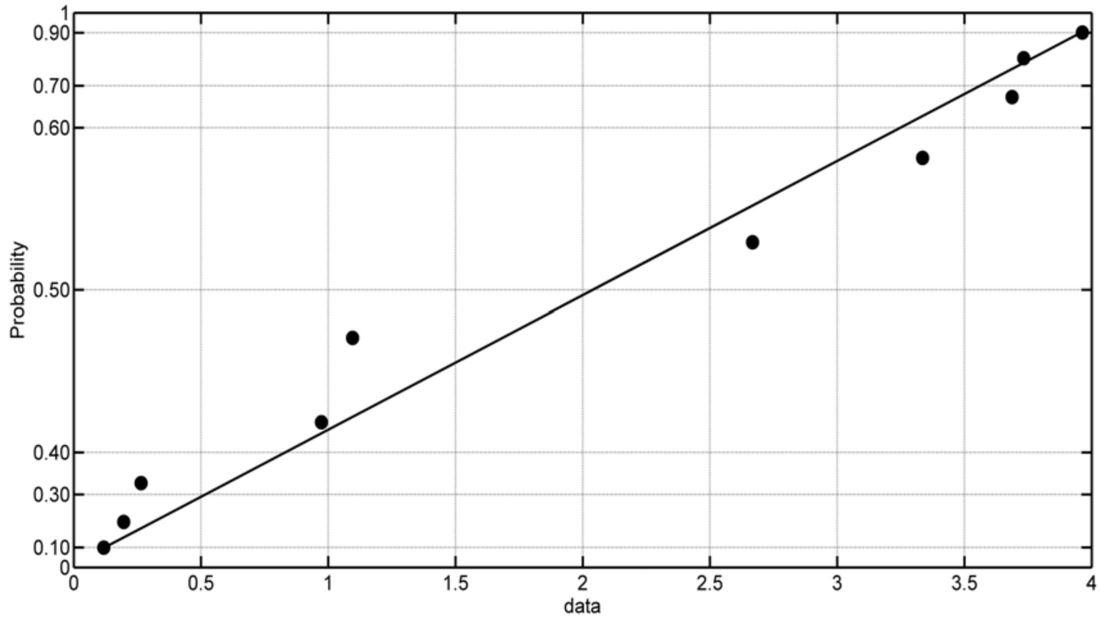
$$\alpha_4 = \frac{25}{12} \approx 1.1905 \tag{5.34n}$$

$$F_Y^{-1}(P) = y_P = \mathrm{sign}(2\,P-1)\,\Big(2\,|P-0.5|\Big)^{1/3}, \quad 0 \le P \le 1 \tag{5.34o}$$

$$f_Y(y_P) = 1.5\,\Big(2\,|P-0.5|\Big)^{2/3} \tag{5.34p}$$

The support of the U–shaped parabolic distribution is finite, so it is no problem to compute the moments of the reduced order statistics by evaluating the integrals (2.9b) and (2.11a). Thus, in LEPP LLOYD's estimator with computed means and variance-covariance matrix is used. For greater sample sizes the evaluation of the double integral for the product moments can last some minutes even when regarding the symmetry of the distribution.

Figure 5/34: U–shaped parabolic probability paper with data and regression line




### 5.2.12.2 Inverted U–shaped parabolic distribution — $X \sim PAI(a,b)$

We will present results for an inverted U–shaped parabolic distribution where the order of the parabola is equal to two.

$$f(x|a,b) = \frac{3}{4\,b}\left[1-\left(\frac{x-a}{b}\right)^2\right],\;\; a-b \le x \le a+b,\; a \in \mathbb{R},\; b>0 \tag{5.35a}$$

$$F(x|a,b) = \frac{1}{2}+\frac{1}{4}\left[\frac{3\,(x-a)}{b}-\left(\frac{x-a}{b}\right)^3\right] \tag{5.35b}$$

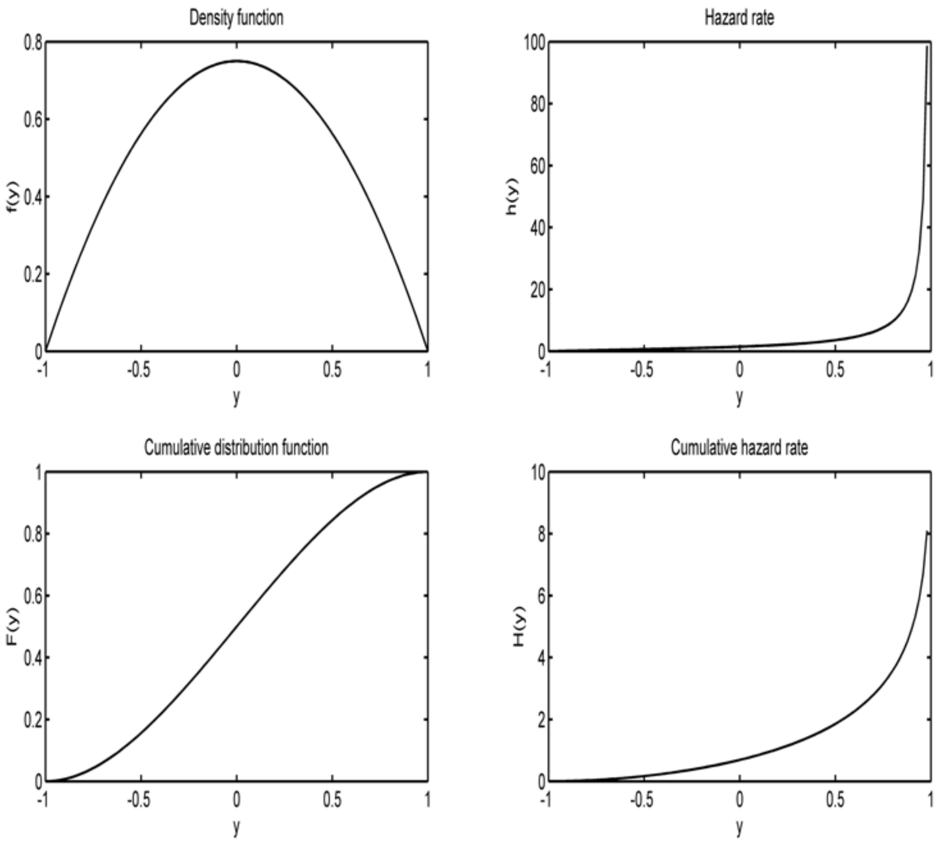
$$R(x|a,b) = \frac{1}{2}-\frac{1}{4}\left[\frac{3\,(x-a)}{b}-\left(\frac{x-a}{b}\right)^3\right] \tag{5.35c}$$

$$h(x|a,b) = \frac{3\left[1-\left(\frac{x-a}{b}\right)^2\right]}{b\left[2-\frac{3\,(x-a)}{b}+\left(\frac{x-a}{b}\right)^3\right]} \tag{5.35d}$$

$$H(x|a,b) = -\ln\big[R(x|a,b)\big] \tag{5.35e}$$

$$F_Y^{-1}(P) = y_P \left\{\begin{array}{l}\text{is the admissible solution, i.e. } -1 \le y_P \le 1, \text{ of} \\ 4\,(P-0.5) = 3\,y_P - y_P^3,\;\; 0 \le P \le 1.\end{array}\right\} \tag{5.35f}$$

Figure 5/35: Several functions for the reduced inverted U–shaped parabolic distribution




$$x_P = a + b\,y_P \tag{5.35g}$$

$$a = x_{0.5} \tag{5.35h}$$

$$b = x_1 - a \tag{5.35i}$$

$$x_M = a \tag{5.35j}$$

$$\mu'_r(Y) = \mu_r(Y) = \frac{1.5\big[1+(-1)^r\big]}{(1+r)\,(3+r)},\quad Y = (X-a)\big/b \tag{5.35k}$$

$$\mu'_1(X) = \mathrm{E}(X) = a \tag{5.35l}$$

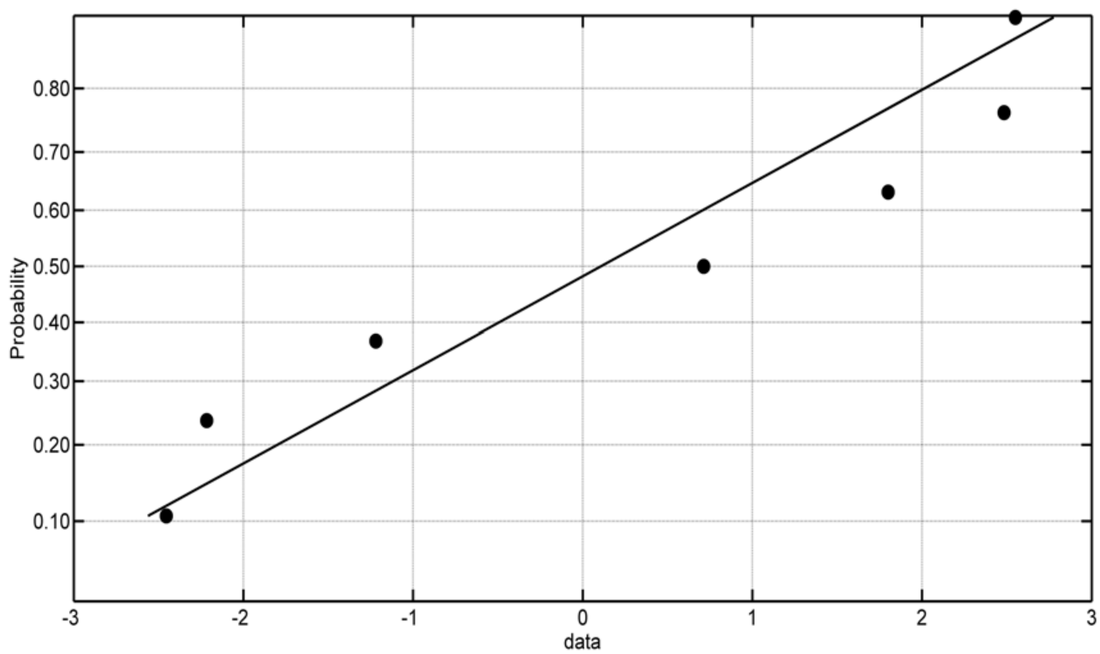
$$\mu_2(X) = \mathrm{Var}(X) = 0.2\,b^2 \tag{5.35m}$$

$$\alpha_3 = 0 \tag{5.35n}$$

$$\alpha_4 = \frac{75}{21} \approx 3.5714 \tag{5.35o}$$

Moments of reduced order statistics can be determined by evaluating the integrals (2.9b) and (2.11a). Thus, in LEPP LLOYD's estimator with computed means and variance-covariance matrix is used. For greater sample sizes the evaluation of the double integral for the product moments can last some minutes even when regarding the symmetry of the distribution.

Figure 5/36: Inverted U–shaped parabolic probability paper with data and regression line




### 5.2.13 RAYLEIGH distribution — $X \sim RA(a, b)$

This distribution is named after the British physicist JOHN WILLIAM STRUTT, 3rd BARON RAYLEIGH (1842 – 1919), who discovered the element argon, which earned him the Nobel Prize for Physics in 1904. He derived this distribution in the field of acoustics. Generally, it is a model for distributions that involve wave propagation, radiation and related phenomena. In meteorology it is a model for wind–speed. If a particles velocities in two orthogonal directions are two independent normal variates with zero means and equal variances, then the distance the particle travels per unit time is RAYLEIGH distributed, or stated more formally:

$$X_1,\ X_2 \overset{\text{iid}}{\sim} NO(0, b) \ \Rightarrow\ X = \sqrt{X_1^2 + X_2^2} \sim RA(0, b).$$

Thus, when $X_1$ and $X_2$ are regarded as the elements of a two–dimensional vector, the length or magnitude $X$ of this vector has a RAYLEIGH distribution. When $X_1$ and $X_2$ are the real and imaginary components, respectively, of a complex random variable, the absolute value of this complex number is RAYLEIGH distributed.

The RAYLEIGH distribution is related to several other distribution, e.g.:

- It is a **$\chi$–distribution** with $\nu = 2$ degrees of freedom.
- It is a special **WEIBULL distribution** (see Sect. 5.3.24) with shape parameter $c = 2$.
- It also is a **linear hazard rate distribution**, see (5.36d).

- It is a special case of **STACY's generalized gamma distribution** with DF
$$f(x|a,b,c,d) = \frac{c\,(x-a)^{c\,d-1}}{b^{c\,d}\,\Gamma(d)}\,\exp\left\{-\left(\frac{x-a}{b}\right)^c\right\},\ x \geq a,\ a \in \mathbb{R},\ b,c,d > 0,$$
when $c = 2$ and $d = 1$.

- It is a special case of **CREEDY–MARTIN's generalized gamma distribution** with DF
$$f(x|\theta_1,\theta_2,\theta_3,\theta_4) = \exp\left\{\theta_1\,\ln x + \theta_2\,x + \theta_3\,x^2 + \theta_4\,x^3 - \eta\right\},\ x > 0,$$
where $\eta$ is a normalizing constant to be determined numerically. $\theta_2 = \theta_4 = 0$ gives a RAYLEIGH distribution.

- The **RICE distribution** with DF
$$f(x|b,\nu) = \frac{x}{b^2}\,\exp\left\{-\frac{(x^2-\nu^2)}{2\,b^2}\right\} I_0\left(\frac{x\,\nu}{b^2}\right),$$
where $I_0(\cdot)$ is the **BESSEL function of the first kind** with zero order, leads to the RAYLEIGH distribution $RA(0,b)$ when $\nu = 0$.

We give the following functions and characteristics of the general RAYLEIGH distribution:

$$f(x|a,b) = \frac{x-a}{b^2}\,\exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right],\ x \geq a,\ a \in \mathbb{R},\ b > 0 \tag{5.36a}$$

$$F(x|a,b) = 1 - \exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right] \tag{5.36b}$$

[33]

$$R(x|a,b) = \exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right] \tag{5.36c}$$

$$h(x|a,b) = \frac{x-a}{b^2} \tag{5.36d}$$

$$H(x|a,b) = \frac{(x-a)^2}{2\,b^2} \tag{5.36e}$$

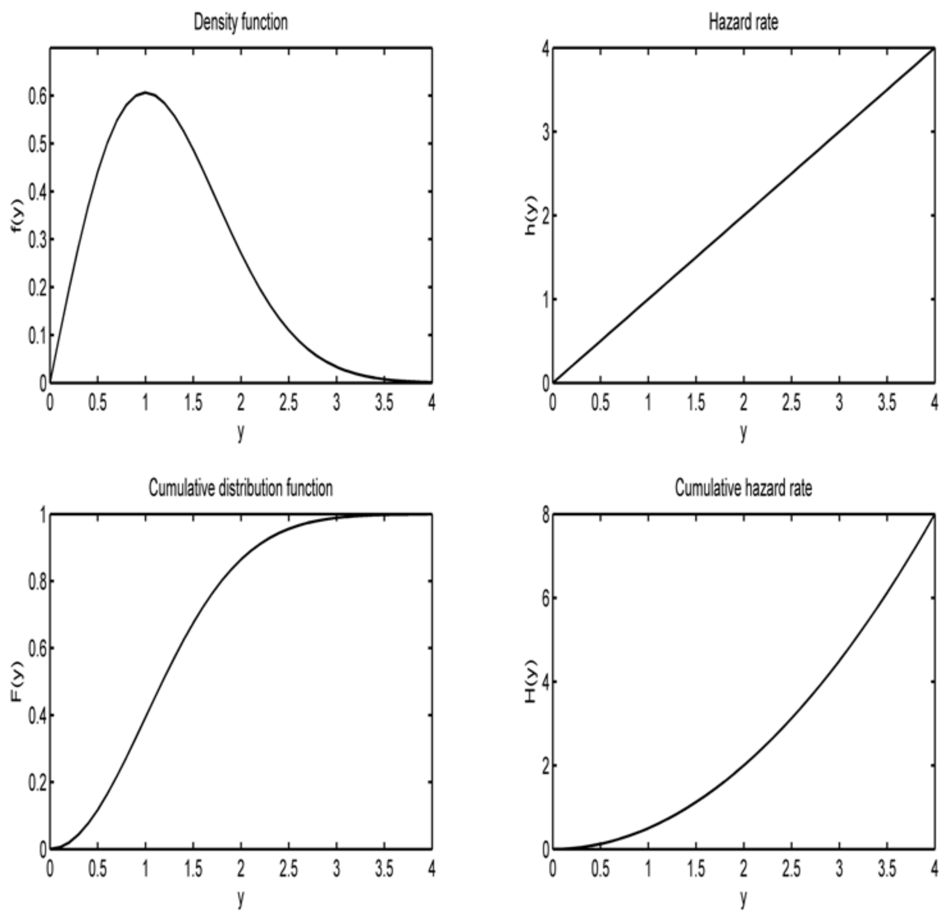
$$F_X^{-1}(P) = x_P = a + b\,\sqrt{-2\,\ln(1-P)},\ 0 \leq P < 1 \tag{5.36f}$$

$$a = x_0 \tag{5.36g}$$

$$b \approx x_{0.3935} - a \tag{5.36h}$$

[33] We notice that
$$f(x|a,b) = \frac{x-a}{b^2}\left[1 - F(x|a,b)\right]$$
which gives rise to several recurrence relations for single and product moments of order statistics, see BALKRISHNAN/RAO (1998a, p. 199).

Figure 5/37: Several functions for the reduced RAYLEIGH distribution




$$x_{0.5} = a + b\sqrt{\ln 4} \approx a + 1.1774\, b \tag{5.36i}$$

$$x_M = a + b \tag{5.36j}$$

$$M_Y(t) = 1 + t\,\exp\left(t^2/2\right)\sqrt{\pi/2}\left[\operatorname{erf}\left(\frac{t}{\sqrt{2}}\right) + 1\right] \tag{5.36k}$$

$$C_Y(t) = 1 - t\,\exp\left(-t^2/2\right)\sqrt{\pi/2}\left[\operatorname{erfi}\left(\frac{t}{\sqrt{2}}\right) - i\right] \tag{5.36l}$$

$$\mu'_r(Y) = 2^{r/2}\Gamma\left(1 + \frac{r}{2}\right),\ \ Y = (X-a)/b \tag{5.36m}$$

$$\mu'_1(Y) = \operatorname{E}(Y) = \sqrt{\pi/2} \approx 1.2533 \tag{5.36n}$$

$$\mu'_2(Y) = 2 \tag{5.36o}$$

$$\mu'_3(Y) = 3\sqrt{\pi/2} \approx 3.7599 \tag{5.36p}$$

$$\mu'_4(Y) = 8 \tag{5.36q}$$

$$\mu_2(Y) = \operatorname{Var}(Y) = \frac{4-\pi}{2} \approx 0.4292 \tag{5.36r}$$

$$\mu_3(Y) = (\pi - 3)\sqrt{\pi/2} \approx 0.1775 \tag{5.36s}$$

$$\mu_4(Y) = \frac{32 - 3\pi^2}{4} \approx 0.5978 \tag{5.36t}$$

$$\mu_1'(X) = \mathrm{E}(X) = a + b\sqrt{\pi/2} \approx a + 1.2533\,b \tag{5.36u}$$

$$\mu_2(X) = \mathrm{Var}(X) = b^2\,\frac{4-\pi}{2} \approx 0.4292\,b^2 \tag{5.36v}$$

$$\alpha_3 = \frac{2\,(\pi-3)\sqrt{\pi}}{(4-\pi)^{3/2}} \approx 0.6311 \tag{5.36w}$$

$$\alpha_4 = \frac{32 - 3\pi^2}{(4-\pi)^2} \approx 3.2450 \tag{5.36x}$$

$$f_Y^{-1}(P) = y_P = \sqrt{-2\,\ln(1-P)},\quad 0 \le P < 1 \tag{5.36y}$$

$$f_Y(y_P) = (1-P)\sqrt{-2\,\ln(1-P)} \tag{5.36z}$$

Because the WEIBULL distribution is reproductive with respect to forming the sample minimum, the first order statistic in a sample of size $n$ from a RAYLEIGH distribution, which is a special WEIBULL distribution, is another RAYLEIGH distribution with DF

$$f_{1:n}\left(x\,\middle|\,a, \frac{b}{\sqrt{n}}\right) = \frac{n}{b^2}\,(x-a)\,\exp\left[-\frac{n}{2}\left(\frac{x-a}{b}\right)^2\right]. \tag{5.37a}$$

Thus, we have the following single moments of the first reduced RAYLEIGH order statistic $Y_{1:n} = (X_{1:n} - a)\big/b$:

$$\alpha_{1:n} = \mathrm{E}(Y_{1:n}) = \sqrt{\frac{\pi}{2\,n}}, \tag{5.37b}$$

$$\alpha_{1:n}^{(2)} = \mathrm{E}(Y_{1:n}^2) = \frac{2}{n}. \tag{5.37c}$$

DYER/WHISENAND (1973, pp. 28, 29) give the following exact and explicit expressions for the first two single moments and also the product moments of reduced RAYLEIGH order statistics. In LEPP we use these formulas for $n \le 25$ to compute the input to LLOYD's estimator, whereas for $n > 25$ linear estimation is based on BLOM's unbiased, nearly best linear estimator.

$$\alpha_{r:n} = \sqrt{\frac{\pi}{2}}\,n\binom{n-1}{r-1}\sum_{i=0}^{r-1}(-1)^{r-1-i}\binom{r-1}{i}\Big/(n-i)^{3/2}, \tag{5.38a}$$

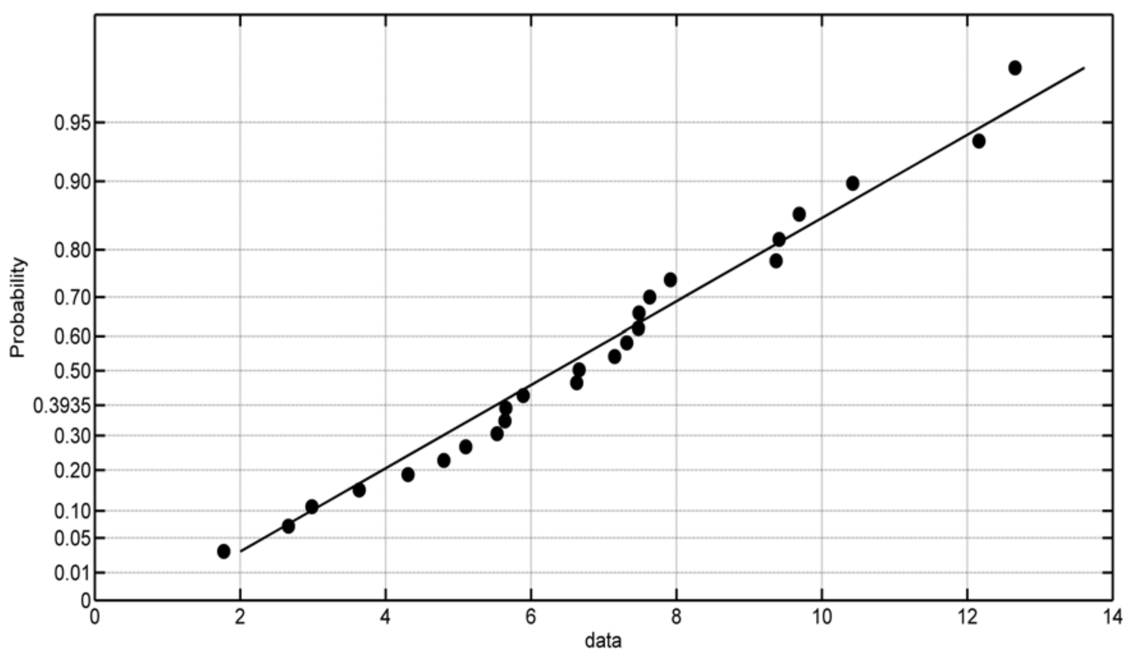
$$\alpha_{r:n}^{(2)} = 2\,n\binom{n-1}{r-1}\sum_{i=0}^{r-1}(-1)^{r-1-i}\binom{r-1}{i}\Big/(n-i)^2, \tag{5.38b}$$

$$\alpha_{r,s:n} = \left\{ \begin{array}{l} \dfrac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} \displaystyle\sum_{i=0}^{s-r-1} \binom{s-r-1}{i} \sum_{j=0}^{r-1} (-1)^{i+j} \binom{r-1}{j} \\ \times\, H(s-r-i+j, n-s+i+1) \end{array} \right\}, \; r < s, \tag{5.38c}$$

where

$$H(a,b) = \frac{\dfrac{\pi}{2} - \arctan\left(\sqrt{\dfrac{b}{a}}\right) + \sqrt{a\,b}\,\dfrac{a-b}{(a+b)^2}}{(a\,b)^{3/2}}\,. \tag{5.38d}$$

Figure 5/38: RAYLEIGH probability paper with data and regression line




## 5.2.14 Reflected exponential distribution — $X \sim RE(a, b)$

We reflect the exponential distribution with DF $f(x|a,b) = (1/b)\,\exp\big[-(x-a)/b\big]$, $x \geq a$, around $x = a$ and arrive at $f(x|a,b) = (1/b)\exp\big[-(a-x)/b\big],\ x \leq a$. By this transformation the lower threshold of the exponential distribution has turned into an upper threshold for the reflected exponential distribution. When we start from a **power–function distribution** with DF $f(x|a,b,c) = (c/b)\big[(x-a)/b\big]^{c-1}$ and introduce $\widetilde{X} = \ln(X-a)$, $\widetilde{a} = \ln b$ and $\widetilde{b} = 1/c$ the distribution of $\widetilde{X}$ is a reflected exponential distribution with DF $f(\widetilde{x}|\widetilde{a},\widetilde{b}\,) = (1/\widetilde{b})\,\exp\big[-(\widetilde{a}-\widetilde{x})/\widetilde{b}\big]$.

The functions and characteristics of the reflected exponential distribution easily follow from those of the exponential distribution.

$$f(x|a,b) = \frac{1}{b}\exp\left(-\frac{a-x}{b}\right) = \frac{1}{b}\exp\left(\frac{x-a}{b}\right),\ x \le a,\ a \in \mathbb{R},\ b > 0 \quad (5.39a)$$

$$F(x|a,b) = \exp\left(\frac{x-a}{b}\right) \quad (5.39b)$$

$$R(x|a,b) = 1 - \exp\left(\frac{x-a}{b}\right) \quad (5.39c)$$

$$h(x|a,b) = \frac{1}{b\left[\exp\left(\frac{a-x}{b}\right) - 1\right]} \quad (5.39d)$$

$$H(x|a,b) = -\ln\left[1 - \exp\left(\frac{x-a}{b}\right)\right] \quad (5.39e)$$

$$F_X^{-1}(P) = x_P = a + b\,\ln P, \quad 0 < P \le 1 \quad (5.39f)$$

Figure 5/39: Several functions for the reduced reflected exponential distribution

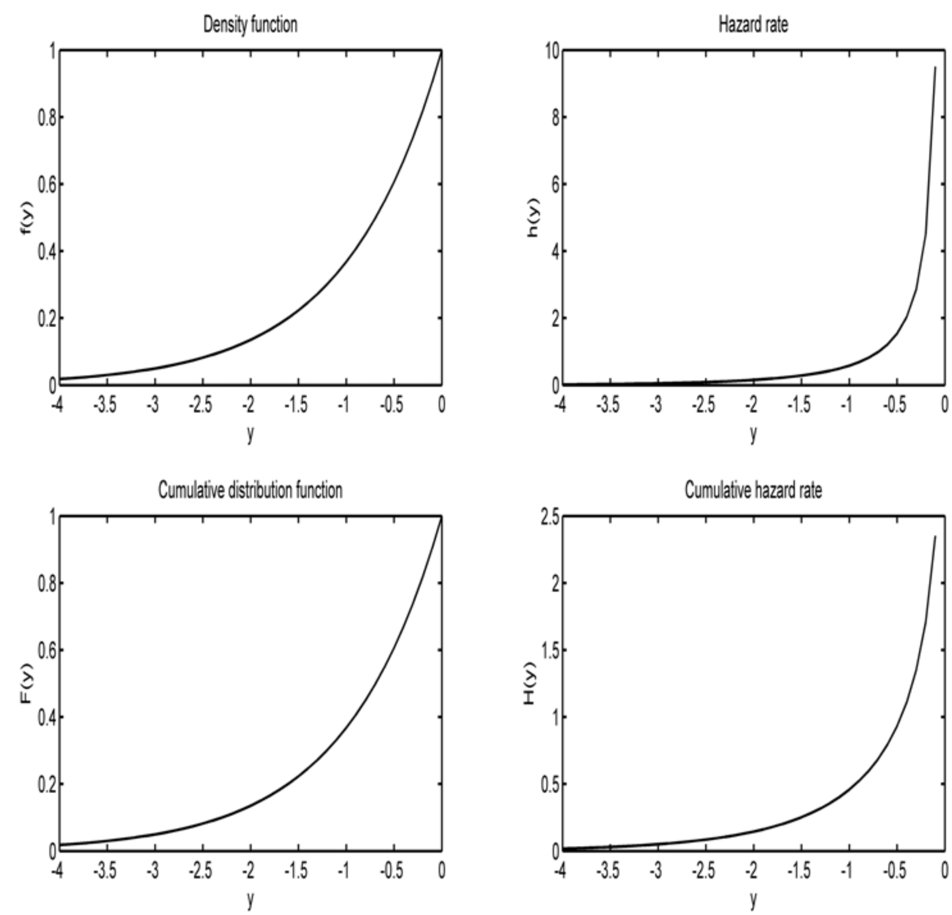
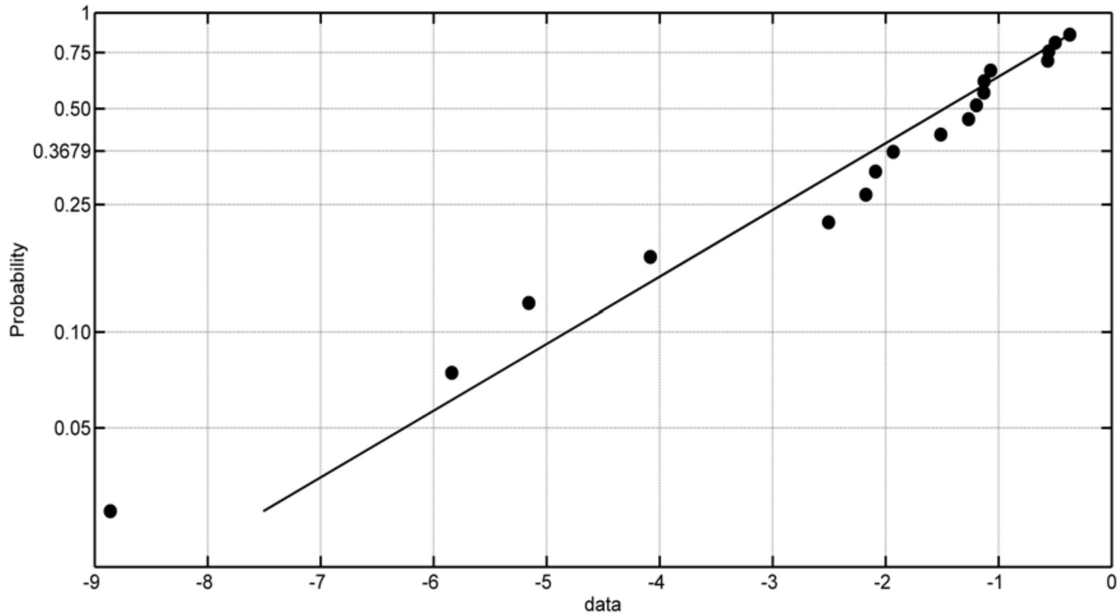


Figure 5/40: Reflected exponential probability paper with data and regression line




$$
\begin{aligned}
x_{0.5} &= a - b\,\ln 2 \;\approx\; a - 0.6931\,b && (5.39g)\\
a &= x_1 && (5.39h)\\
b &\approx a - x_{0.3679} && (5.39i)\\
x_M &= a \;\text{ with }\; f(x_M|a,b) = 1/b && (5.39j)\\
M_X(t) &= \frac{\exp(a\,t)}{1+b\,t} && (5.39k)\\
C_X(t) &= \frac{\exp(i\,a\,t)}{1+i\,b\;t}\,, \quad i=\sqrt{-1} && (5.39l)\\
\mu'_r(X) &= (-b)^r\,\exp\Big(-\frac{a}{b}\Big)\;\Gamma\Big(1+r,-\frac{a}{b}\Big)) && (5.39m)\\
\mu'_1(X) &= \mathrm{E}(X) \;=\; a-b && (5.39n)\\
\mu'_2(X) &= b^2 + (a-b)^2 && (5.39o)\\
\mu'_3(X) &= (a-b)^3 + b^2\,(3\,a - 5\,b) && (5.39p)\\
\mu'_4(X) &= (a-b)^4 + b^2\,(6\,a^2 - 20\,a\,b + 23\,b^2) && (5.39q)\\
\mu_2(X) &= \mathrm{Var}(X) \;=\; b^2 && (5.39r)\\
\mu_3(X) &= -2\,b^3 && (5.39s)\\
\mu_4(X) &= 9\,b^4 && (5.39t)\\
\alpha_3 &= -2 && (5.39u)\\
\alpha_4 &= 9 && (5.39v)
\end{aligned}
$$

$$I(X) = \operatorname{ld} 2 + \operatorname{ld} b \approx 1.442695\,(1+\ln b) \tag{5.39w}$$

$$F_Y^{-1}(P) = y_P = \ln P \tag{5.39x}$$

$$f_Y(y_P) = P \tag{5.39y}$$

$$F^{-1(r)}(P) = (-1)^{r-1}\,\frac{(r-1)!}{P^r}, \quad r = 1, 2, \ldots \tag{5.39z}$$

The moments of the reduced reflected exponential order statistics can be derived from those of the exponential distribution. Let $\widetilde{\alpha}_{r:n}$, $\widetilde{\beta}_{r,r:n}$ and $\widetilde{\beta}_{r,s:n}$ denote the moments of the reduced exponential distribution as given by (5.9l–n) then those of the reduced reflected exponential distribution follow from (5.13a–c). These computed moments are the input to LLOYD's estimator in LEPP.

### 5.2.15 Semi–elliptical distribution — $X \sim SE(a, b)$

This distribution is known as **WIGNER's semicircle distribution**, named after the Hungarian–American physicist and mathematician EUGENE PAUL WIGNER (1902 – 1995), winner of the NOBEL prize in physics 1963. The distribution arises as the limiting distribution of eigenvalues of many random symmetric matrices. The graph of the density function is a semi–ellipse centered at $x = a$ and a horizontal half–axis equal to $b$, i.e. the support is $[a-b, a+b]$. For $b = \sqrt{2/\pi} \approx 0.7979$ the graph of the density is a semicircle, and for $b < \sqrt{2/\pi}$ $\left(b > \sqrt{2/2}\right)$ the semi–ellipse is compressed (stretched), see Fig. 5/41.

$$f(x|a,b) = \frac{2}{b\,\pi}\sqrt{1-\left(\frac{x-a}{b}\right)^2}, \quad a-b \le x \le a+b,\ a \in \mathbb{R},\ b > 0 \tag{5.40a}$$

$$F(x|a,b) = \frac{1}{2} + \frac{1}{\pi}\left\{\frac{x-a}{b}\sqrt{1-\left(\frac{x-a}{b}\right)^2} + \arcsin\left(\frac{x-a}{b}\right)\right\} \tag{5.40b}$$

$$R(x|a,b) = \frac{1}{2} - \frac{1}{\pi}\left\{\frac{x-a}{b}\sqrt{1-\left(\frac{x-a}{b}\right)^2} + \arcsin\left(\frac{x-a}{b}\right)\right\} \tag{5.40c}$$

$$h(x|a,b) = \frac{f(x|a,b)}{R(x|a,b)} \tag{5.40d}$$

$$H(x|a,b) = -\ln\big[R(x|a,b)\big] \tag{5.40e}$$

$$F_Y^{-1}(P) = y_P = \left\{\begin{array}{l}\text{is the admissible solution, i.e. } -1 \le y_P \le 1, \text{ of}\\ \pi(P-0.5) - y_P\sqrt{1-y_P^2} - \arcsin(y_P) = 0,\ 0 \le P \le 1.^{34}\end{array}\right\} \tag{5.40f}$$

[34] The MATLAB root–finding procedure `fzero` returns `NaN` for $P < 0.05$ and for $P > 0.95$. Thus, in these cases we linearly interpolate $y_P$ between $y_0 = -1$ and $y_{0.05} = -0.8054$ and between $y_{0.95} = 0.8054$ and $y_1 = 1$.

Figure 5/41: Densities of semi–elliptical distributions

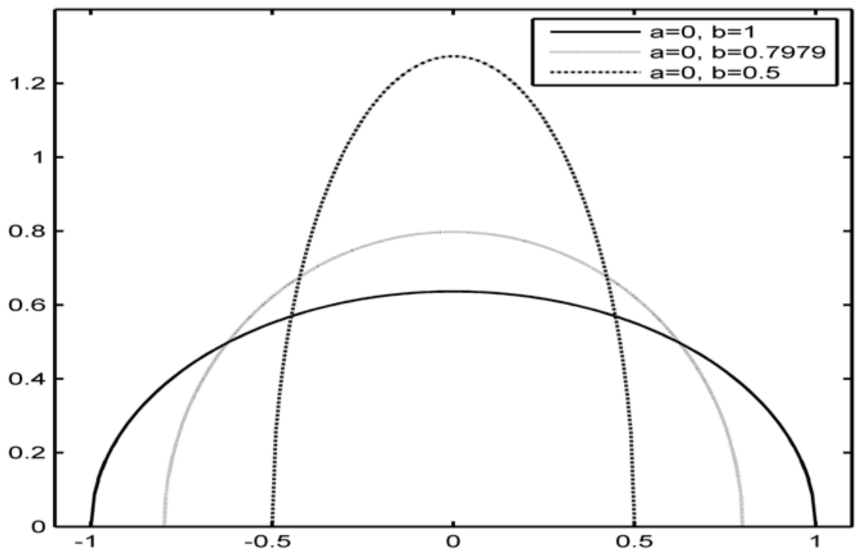
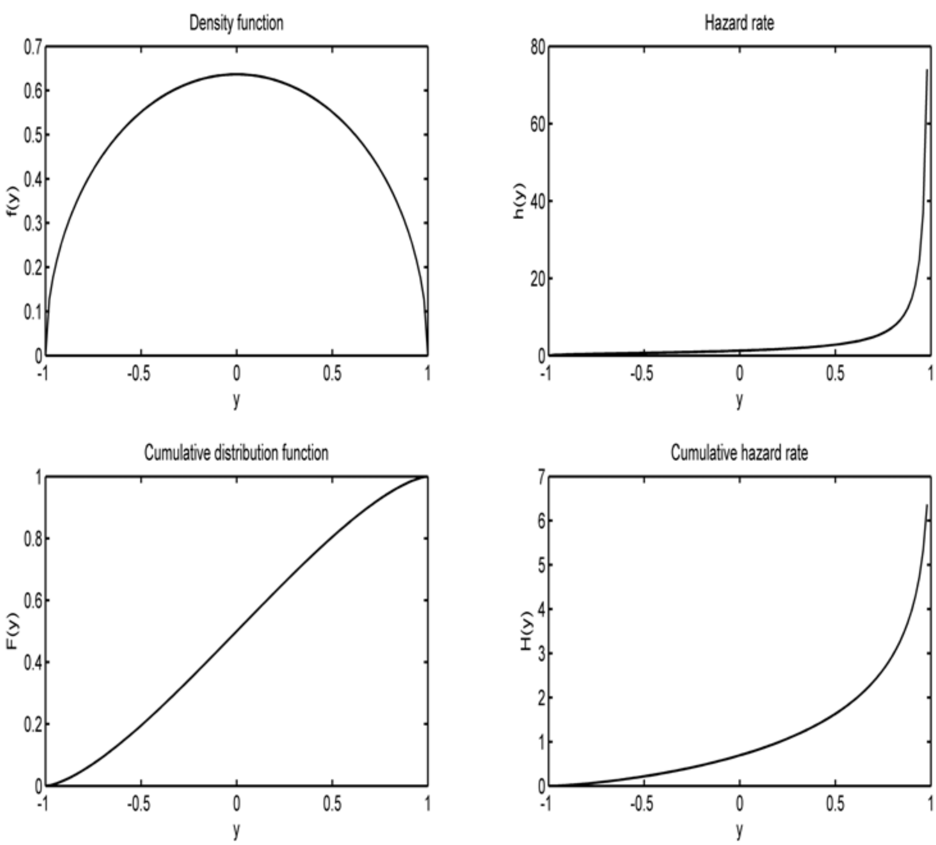



Figure 5/42: Several functions for the reduced semi–elliptical distribution



$$x_P = a + b\,y_P \tag{5.40g}$$

$$a = x_{0.5} \tag{5.40h}$$

$$b = x_1 - a \tag{5.40i}$$

$$x_M = a \tag{5.40j}$$

$$M_X(t) = 2\,\exp(a\,t)\,\frac{I_1(b\,t)}{b\,t}\ ^{35} \tag{5.40k}$$

$$C_X(t) = 2\,\exp(i\,a\,t)\,\frac{J_1(b\,t)}{b\,t}\ ^{36} \tag{5.40l}$$

$$\mu'_r(Y) = \mathrm{E}\big(Y^r\big) = \left\{\begin{array}{cl} 0 & \text{for } r \text{ odd} \\ \left(\dfrac{1}{2}\right)^r C_{r/2} & \text{for } r \text{ even}^{37} \end{array}\right\},\ Y = (X-a)\big/b \tag{5.40m}$$

$$\mu'_1(X) = \mathrm{X} = a \tag{5.40n}$$

$$\mu_2(X) = \mathrm{Var}(X) = \frac{b^2}{4} \tag{5.40o}$$

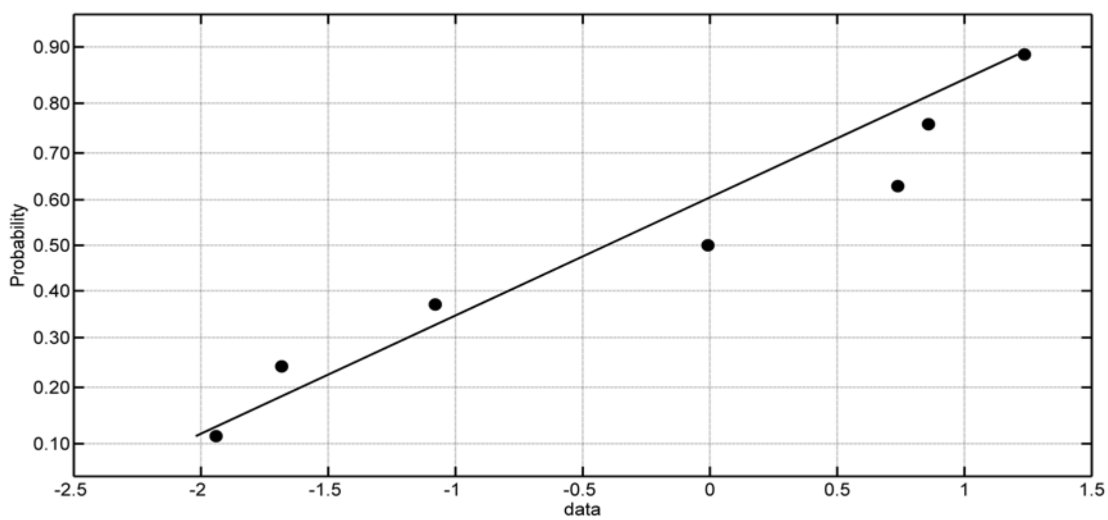
$$\alpha_3 = 0 \tag{5.40p}$$

$$\alpha_4 = 1 \tag{5.40q}$$

$$I(X) = b\,\pi - \frac{1}{2\,\ln 2} \approx 3.1416\,b - 0.7213 \tag{5.40r}$$

The moments of the order statistics can be computed by integration, which lasts some time for large $n$. They are input to LLOYD's estimator in LEPP.

Figure 5/43: Semi–elliptical probability paper with data and regression line




---

[35] $I_1(\cdot)$ is the modified BESSEL function, see ABRAMOWITZ/STEGUN (1972, p. 376).

[36] $J_1(\cdot)$ is the BESSEL function, see ABRAMOWITZ/STEGUN (1972, p. 360).

[37] $C_n = \binom{2\,n}{n}\big/(n+1)$ is the **CATALAN number**.

### 5.2.16 TEISSIER distribution — $X \sim TE(a, b)$

In 1934 the French biologist GEORGES TEISSIER proposed this distribution as a life–time distribution which is characterized by an exponentially declining **mean residual life**.[38] The mean residual life is the expected time to be lived by an individual aged $x$:

$$\mu(x) = \mathrm{E}(X - x|X > x) = \frac{\int_x^\infty R(v)\,\mathrm{d}v}{R(x)} \tag{5.41a}$$

where $R(x) = \Pr(X > x)$ is the survival function. $\mu(x)$ satisfies:

$$\mu(x) \geq 0, \qquad \frac{\mathrm{d}\mu(x)}{\mathrm{d}} x \geq -1, \qquad \int_0^\infty \frac{1}{\mu(x)}\,\mathrm{d}x = \infty. \tag{5.41b}$$

For $X \sim TE(a, b)$ we have

$$\mu(x) = (a + b)\, \exp\left(-\frac{x - a}{b}\right), \tag{5.41c}$$

where $a + b = \mathrm{E}(X) = \mathrm{E}(X - a|X > a)$.

We give the following functions and characteristics:

$$f(x|a, b) = \frac{1}{b}\left[\exp\left(\frac{x-a}{b}\right)-1\right]\exp\left[1+\frac{x-a}{b}-\exp\left(\frac{x-a}{b}\right)\right], \quad x \geq a,\ a \in \mathbb{R},\ b > 0 \tag{5.42a}$$

$$F(x|a, b) = 1 - \exp\left[1 + \frac{x - a}{b} - \exp\left(\frac{x - a}{b}\right)\right] \tag{5.42b}$$

$$R(x|a, b) = \exp\left[1 + \frac{x - a}{b} - \exp\left(\frac{x - a}{b}\right)\right] \tag{5.42c}$$

$$h(x|a, b) = \frac{1}{b}\left[\exp\left(\frac{x - a}{b}\right) - 1\right] \tag{5.42d}$$

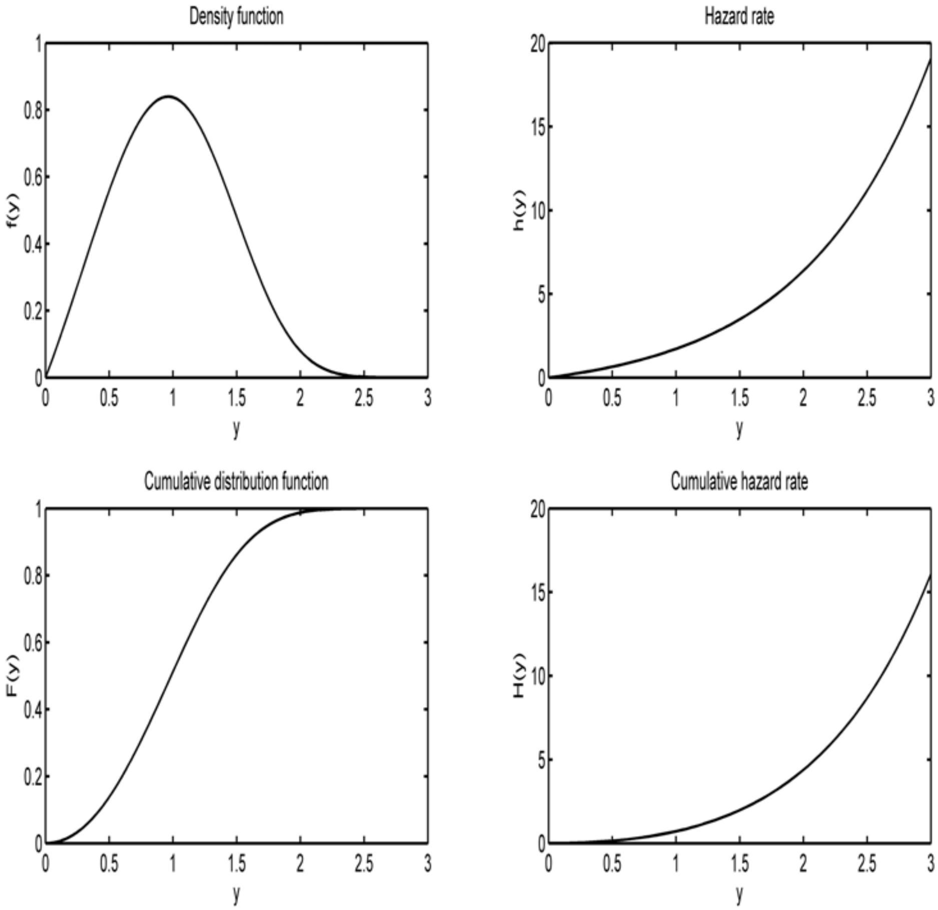
$$H(x|a, b) = \exp\left(\frac{x - a}{b}\right) - \frac{x - a}{b} - 1 \tag{5.42e}$$

$$F_Y^{-1}(P) = y_P = \left\{\begin{array}{l}\text{is the admissible solution, i.e. } y_P > 0, \text{ of} \\ 1 + y_P - \exp(y_P) - \log(1 - P) = 0, \;\; 0 \leq P < 1.\end{array}\right\} \tag{5.42f}$$

---

[38] Later on, this model has been discussed by LAURENT (1975). RINNE (1985) used the TEISSIER model to estimate the life–time distribution, with life–time expressed in kilometers, for German motorcars based on prices of used cars. These prices could be approximated sufficiently well by an exponentially falling function of the kilometers traveled.

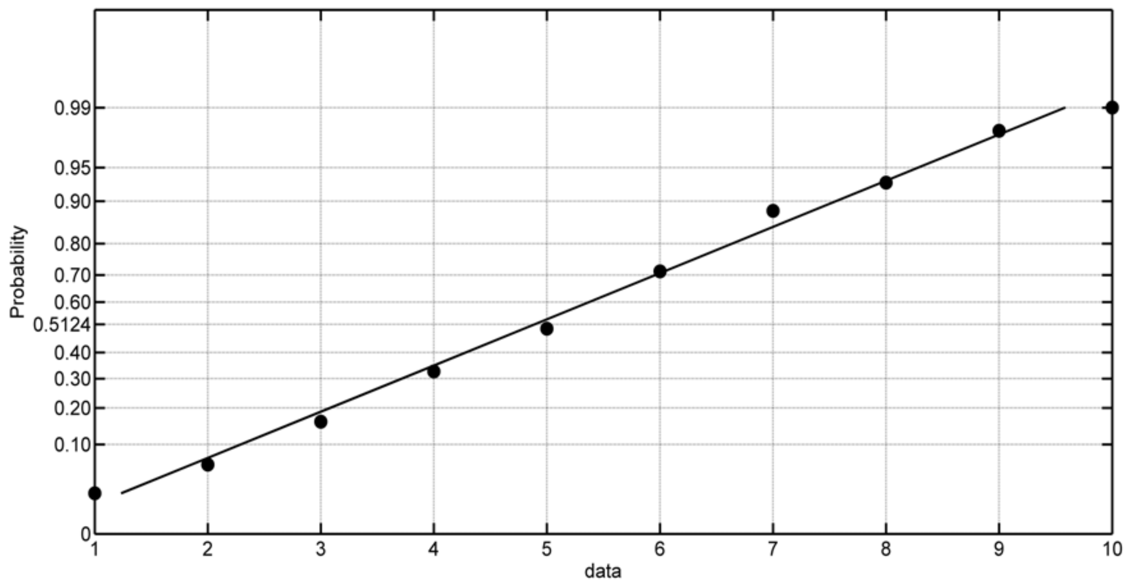
Figure 5/44: Several functions for the reduced TEISSIER distribution




$$x_P = a + b\,y_P \tag{5.42g}$$
$$a = x_0 \tag{5.42h}$$
$$b \approx x_{0.5124} - a \tag{5.42i}$$
$$x_{0.5} \approx a + 0.9852\,b \tag{5.42j}$$
$$x_M \approx a + 0.9624\,b \tag{5.42k}$$
$$\mu_1'(Y) = \text{E}(Y) = 1, \quad Y = (X-a)\big/b \tag{5.42l}$$
$$\mu_2'(Y) \approx 1.1927 \tag{5.42m}$$
$$\mu_3'(Y) \approx 1.5958 \tag{5.42n}$$
$$\mu_4'(Y) \approx 2.3227 \tag{5.42o}$$
$$\mu_1'(X) = \text{E}(X) = a + b \tag{5.42p}$$
$$\mu_2(X) = \text{Var}(X) \approx 0.1927\,b^2 \tag{5.42q}$$
$$\alpha_3 \approx 0.0177 \tag{5.42r}$$
$$\alpha_4 \approx 2.5781 \tag{5.42s}$$

In LEPP linear estimation for order statistics input is realized by BLOM's unbiased, nearly best linear estimator with means evaluated by numerical integration.

Figure 5/45: TEISSIER probability paper with data and regression line




## 5.2.17 Triangular distributions

In the following section we present results for three types of triangular distributions which are of location–scale type and which are parameterized in such a way that the location parameter $a$ is equal to the mode:

- the **symmetric version** with DF

$$f(x|a,b) = \frac{1}{b}\left(1 - \frac{|x-a|}{b}\right) = \frac{b-|x-a|}{b^2}, \quad a-b \le x \le a+b \tag{5.43a}$$

  or equivalently written as

$$f(x|a,b) = \left\{\begin{array}{ll} \dfrac{x-(a-b)}{b^2} & \text{for} \quad a-b \le x \le a \\ \dfrac{a+b-x}{b^2} & \text{for} \quad a \le x \le a+b \end{array}\right\}, \tag{5.43b}$$

- the **right–angled and positively skew version** with DF

$$f(x|a,b) = \frac{2}{b}\left(1 - \frac{x-a}{b}\right) = \frac{2}{b^2}\,(a+b-x), \quad a \le x \le a+b, \tag{5.43c}$$

- the **right–angled and negatively skew version** with DF

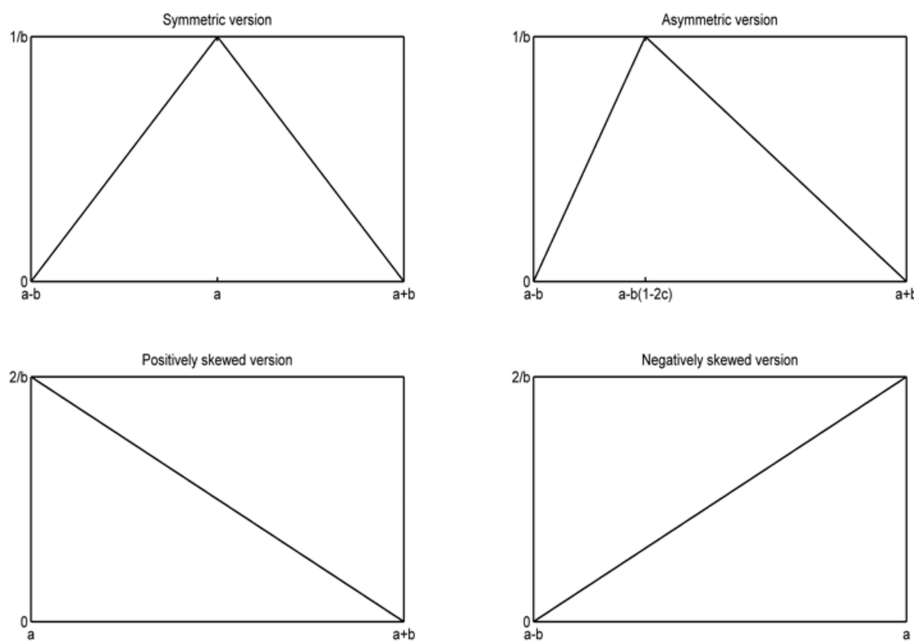
$$f(x|a,b) = \frac{2}{b}\left(\frac{x-a}{b} + 1\right) = \frac{2}{b^2}\,[x-(a-b)], \quad a-b \le x \le a. \tag{5.43d}$$

These three triangular distributions are special cases of the **general triangular distribution** which — besides $a$ and $b$ — has a third parameter $c,\ 0 < c < 1$, which is responsible for the skewness and the location of the mode[39] on the abscissa:

$$f(x|a,b) = \left\{ \begin{array}{ll} \dfrac{x-(a-b)}{2\,c\,b^2} & \text{for} \quad a-b \le x \le a-b\,(1-2\,c) \\ \dfrac{a+b-x}{2\,(1-c)\,b^2} & \text{for} \quad a-b\,(1-2\,c) \le x \le a+b. \end{array} \right\} \tag{5.43e}$$

(5.43e) turns into the symmetric version (5.43a) for $c = 0.5$. With $c \to 0$ (5.43e) goes to a right–angled and positively skew triangular distribution, but — contrary to (5.43c) — with support $[a-b, a+b]$ and DF $f(x|a,b) = (a+b-x)\big/(2\,b^2)$. With $c \to 1$ (5.43e) goes to a right–angled and negatively skew triangular distribution, but — contrary to (5.43d) — with support $[a-b, a+b]$ and DF $f(x|a,b) = (x-a+b)\big/(2\,b^2)$.

Figure 5/46: Densities of several triangular distributions




#### 5.2.17.1 Symmetric version — $X \sim TS(a,b)$

The symmetric triangular distribution, sometimes called **tine distribution** or **SIMPSON's distribution**[40] is the distribution of the sum of two independent and identically distributed **uniform variables** (= convolution of two uniform distributions) , more precisely:

$$X_1,\ X_2 \overset{\text{iid}}{\sim} UN(a,b) \ \Rightarrow\ X = X_1 + X_2 \sim TS(2\,a+b, b).$$

[39] This mode is $x_M = a - b\,(1-2\,c)$.

[40] THOMAS SIMPSON (1710 – 1761) seems to be the first statistician who had suggested this distribution.

We give the following functions and characteristics of the symmetric triangular distribution:

$$f(x|a,b) = \frac{b-|x-a|}{b^2},\; a-b \le x \le a+b,\; a \in \mathbb{R},\; b>0 \tag{5.44a}$$

$$F(x|a,b) = \left\{\begin{array}{ll} \frac{1}{2}\left(1+\frac{x-a}{b}\right)^2, & a-b \le x \le a \\ 1-\frac{1}{2}\left(1-\frac{x-a}{b}\right)^2, & a \le x \le a+b \end{array}\right\} \tag{5.44b}$$

$$R(x|a,b) = \left\{\begin{array}{ll} 1-\frac{1}{2}\left(1+\frac{x-a}{b}\right)^2, & a-b \le x \le a \\ \frac{1}{2}\left(1-\frac{x-a}{b}\right)^2, & a \le x \le a+b \end{array}\right\} \tag{5.44c}$$

$$h(x|a,b) = \left\{\begin{array}{ll} \frac{2\,(x-a+b)}{2\,b^2-(x-a+b)^2}, & a-b \le x \le a \\ \frac{2}{a+b-x}, & a \le x \le a+b \end{array}\right\} \tag{5.44d}$$

$$H(x|a,b) = \left\{\begin{array}{ll} -\ln\left[1-\frac{1}{2}\left(\frac{x-a+b}{b}\right)^2\right], & a-b \le x \le a \\ \ln 2-\ln\left[\left(\frac{a+b-x}{b}\right)^2\right], & a \le x \le a+b \end{array}\right\} \tag{5.44e}$$

$$F_Y^{-1}(P) = y_P = \left\{\begin{array}{ll} \sqrt{2\,P}-1, & 0 \le P \le 0.5 \\ 1-\sqrt{2\,(1-P)}, & 0.5 \le P \le 1 \end{array}\right\},\; Y=(X-a)\big/b \tag{5.44f}$$

$$x_P = a+b\,y_P \tag{5.44g}$$

$$a = x_{0.5} \tag{5.44h}$$

$$b = x_{0.5}-x_0 \;=\; x_1-x_{0.5} \tag{5.44i}$$

$$x_M = a \tag{5.44j}$$

$$M_X(t) = \exp(a\,t)\,\frac{2\left[\cosh(b\,t)-1\right]}{(b\,t)^2} \tag{5.44k}$$

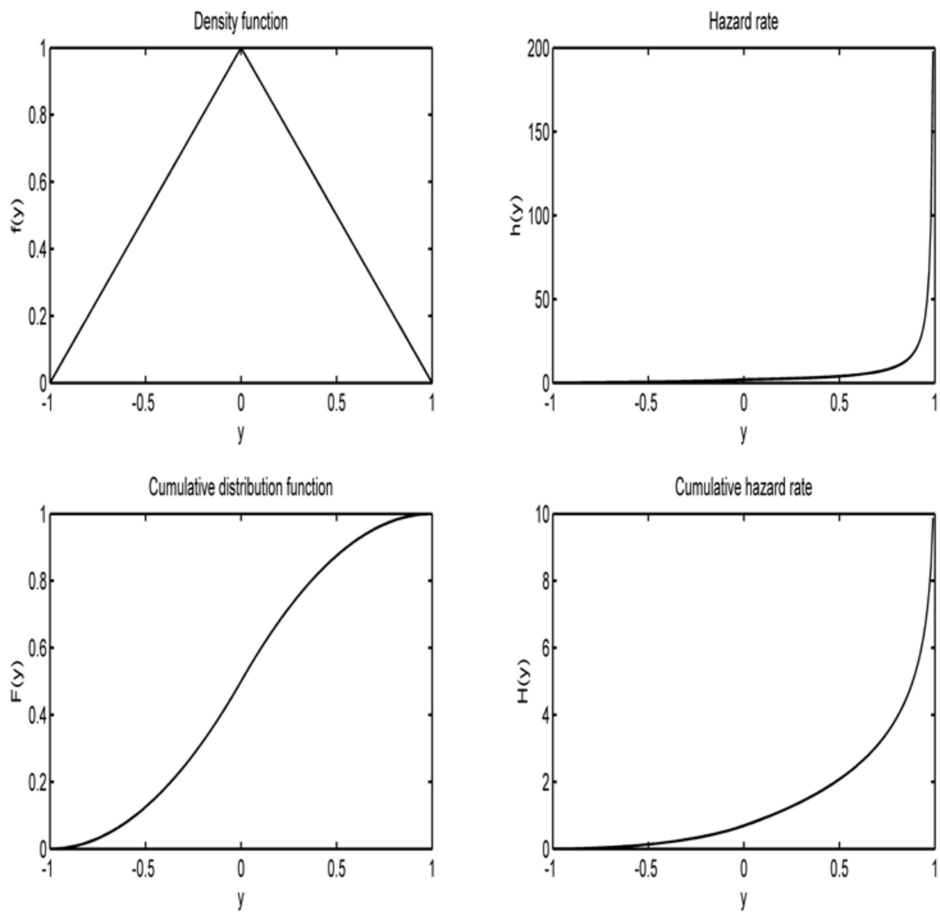
$$= \exp(a\,t)\,\frac{\exp(b\,t)+\exp(-b\,t)-2}{(b\,t)^2} \tag{5.44l}$$

$$C_X(t) = \exp(i\,a\,t)\,\frac{2\left[1-\cosh(b\,t)\right]}{(b\,t)^2} \tag{5.44m}$$

$$= \exp(i\,a\,t)\,\frac{\left[2-\exp(-i\,b\,t)-\exp(i\,b\,t)\right.}{(b\,t)^2} \tag{5.44n}$$

$$\mu_r'(Y) = \left\{\begin{array}{ll} 0 & \text{for } r \text{ odd} \\ \frac{1+(-1)^r}{(1+r)\,(2+r)} & \text{for } r \text{ even} \end{array}\right\} \tag{5.44o}$$

Figure 5/47: Several functions for the reduced symmetric triangular distribution




$$\mu_2'(Y) = 1/6 \tag{5.44p}$$

$$\mu_4'(Y) = 1/15 \tag{5.44q}$$

$$\mu_1'(X) = \mathrm{E}(X) = a \tag{5.44r}$$

$$\mu_2(X) = \mathrm{Var})X) = b^2/6 \tag{5.44s}$$

$$\alpha_3 = 0 \tag{5.44t}$$

$$\alpha_4 = 2.4 \tag{5.44u}$$

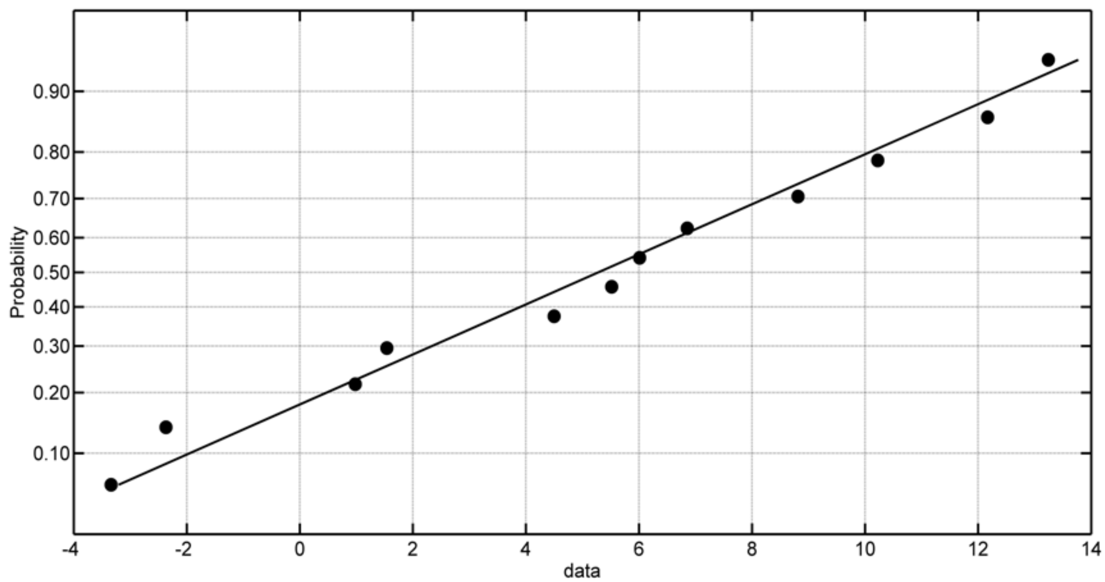
$$I(X) = \mathrm{ld}\, b + \frac{1}{\ln 4} \approx 0.7213 + \mathrm{ld}\, b \tag{5.44v}$$

$$F_Y^{-1}(P) = y_P = \left\{ \begin{array}{cl} \sqrt{2\,P} - 1, & 0 \le P \le 0.5 \\ 1 - \sqrt{2\,(1-P)}, & 0.5 \le P \le 1 \end{array} \right\} \tag{5.44w}$$

$$f_Y(y_P) = \left\{ \begin{array}{cl} \sqrt{2\,P}, & 0 \le P \le 0.5 \\ \sqrt{2\,(1-P)}, & 0 \le P \le 1 \end{array} \right\} \tag{5.44x}$$

The moments of the reduced order statistics, which are computed by numerical integration, are the input to LLOYD's estimator in LEPP.

Figure 5/48: Symmetric triangular probability paper with data and regression line




#### 5.2.17.2 Right–angled and negatively skew version — $X \sim TN(a,b)$

When the symmetric triangular distribution is folded to the left about $x = a$ we obtain the negatively skew and right–angled triangular distribution with $x_M = a$ and support $[a-b, a]$, i.e. we have a left half–triangular distribution. We may also think of this distribution as a reflected right–angled and positively skew triangular distribution. The negatively skew version is a special case of:

- the **beta distribution** (1.12a) for $c = 2,\ d = 1$ and $a$ substituted with $a - b$,
- the **power–function distribution** (1.49a) for $c = 2$ and $a$ substituted with $a - b$.

$$f(x|a,b) = \frac{2}{b}\left(\frac{x-a}{b}+1\right) = \frac{2}{b^2}\big[x-(a-b)\big],\ a-b \le x \le a,\ a \in \mathbb{R},\ b>0 \quad (5.45a)$$

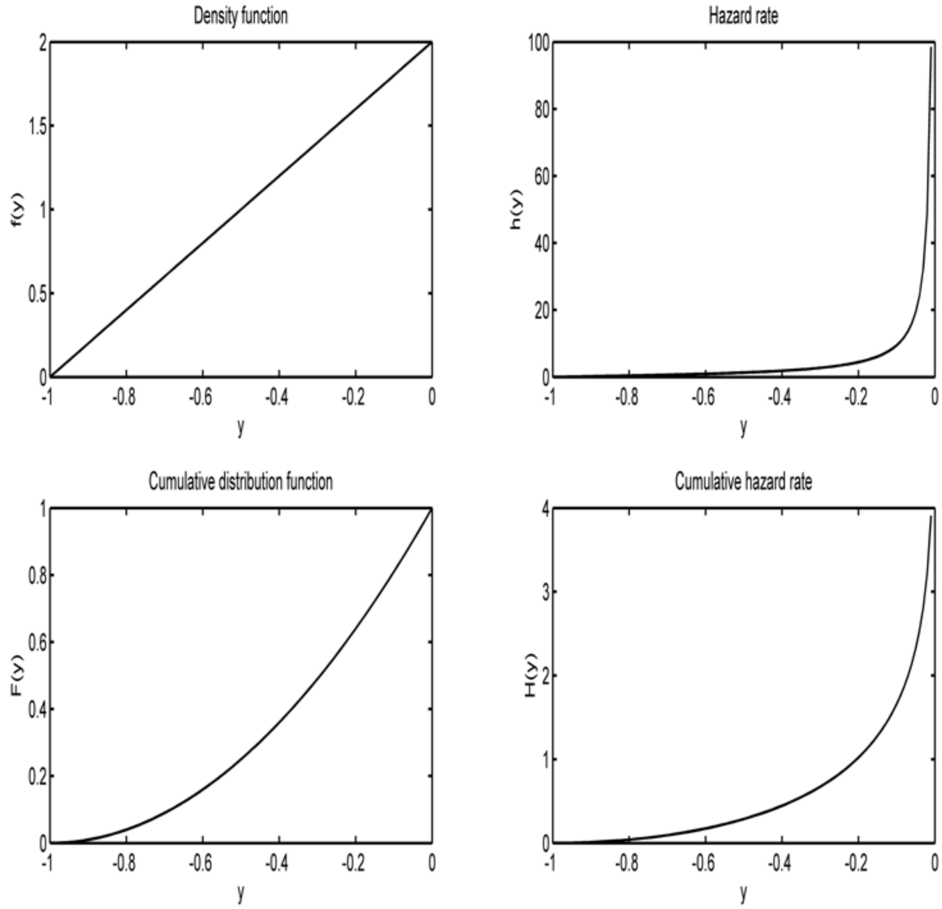
$$F(x|a,b) = \left(\frac{x-a+b}{b}\right)^2 \quad (5.45b)$$

$$R(x|a,b) = 1-\left(\frac{x-a+b}{b}\right)^2 = \frac{a-x}{b}\left(\frac{x-a}{b}+2\right) \quad (5.45c)$$

$$h(x|a,b) = \frac{1}{a-x}+\frac{1}{a-2\,b-x} \quad (5.45d)$$

$$H(x|a,b) = 2\ln b - \ln\big[(a-x)\,(x+2\,b-a)\big] \quad (5.45e)$$

Figure 5/49: Several functions for the reduced right–angled and negatively skew triangular distribution




$$F_X^{-1}(P) = x_P = a + b\left(\sqrt{P} - 1\right),\ \ 0 \le P \le 1 \tag{5.45f}$$

$$a = x_1 \tag{5.45g}$$

$$b = x_1 - x_0 \tag{5.45h}$$

$$x_{0.5} \approx a - 0.2929\,b \tag{5.45i}$$

$$x_M = a \tag{5.45j}$$

$$M_X(t) = \frac{2\,\exp(a\,t)\left[\exp(-b\,t) + b\,t - 1\right]}{(b\,t)^2} \tag{5.45k}$$

$$C_X(t) = \frac{2\,\exp(i\,a\,t)\left[1 - \exp(-i\,b\,t) - i\,b\,t\right]}{(b\,t)^2} \tag{5.45l}$$

$$\mu'_r(Y) = \frac{2\,(-1)^r}{(1+r)\,(2+r)};\ \ r = 0, 1, 2\ldots;\ Y = (X-a)\big/b \tag{5.45m}$$

$$\mu_1'(Y) = \mathrm{E}(Y) = -\frac{1}{3} \tag{5.45n}$$

$$\mu_2'(Y) = \frac{1}{6} \tag{5.45o}$$

$$\mu_3'(Y) = -\frac{1}{10} \tag{5.45p}$$

$$\mu_4'(Y) = \frac{1}{15} \tag{5.45q}$$

$$\mu_1'(X) = \mathrm{E}(X) = a - \frac{b}{3} \tag{5.45r}$$

$$\mu_2(X) = \mathrm{Var}(X) = \frac{b^2}{18} \tag{5.45s}$$

$$\alpha_3 \approx -0.5657 \tag{5.45t}$$

$$\alpha_4 = 2.4 \tag{5.45u}$$

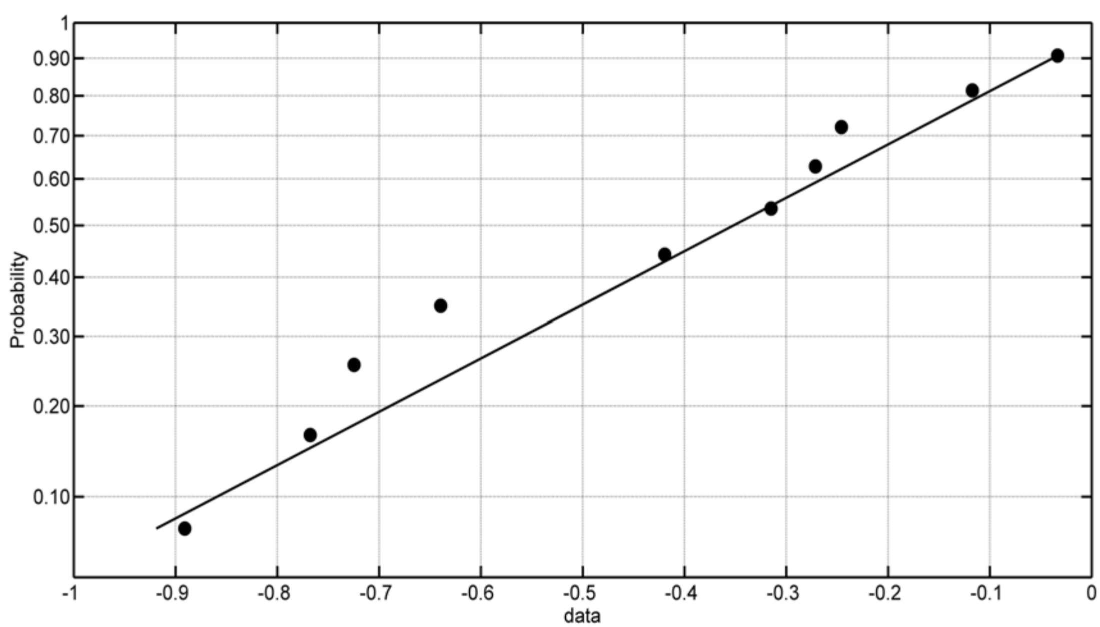
$$I(X) = \mathrm{ld}\, b + \frac{1}{\ln 4} - 1 \approx -0.2787 + \mathrm{ld}\, b \tag{5.45v}$$

$$F_Y^{-1}(P) = y_P = \sqrt{P} - 1 \tag{5.45w}$$

$$f_Y(y_P) = 2\sqrt{P} \tag{5.45x}$$

The moments of the reduced order statistics, which are computed by numerical integration, are the input to LLOYD's estimator in LEPP.

Figure 5/50: Right–angled, negatively skew triangular probability paper with data and regression line



#### 5.2.17.3 Right–angled and positively skew version — $X \sim TP(a, b)$

This distribution is the right–half symmetric triangular distribution, i.e. the latter distribution is folded to the right–hand side about $x = a$. We can also regard it as a reflected right–angled and negatively skew triangular distribution. The positively skew version is a special **beta distribution** with $c = 1$ and $d = 2$.

$$f(x|a,b) = \frac{2}{b}\left(1-\frac{x-a}{b}\right) = \frac{2}{b^2}\,(a+b-x)\big],\ a \le x \le a+b,\ a \in \mathbb{R},\ b > 0 \quad (5.46a)$$

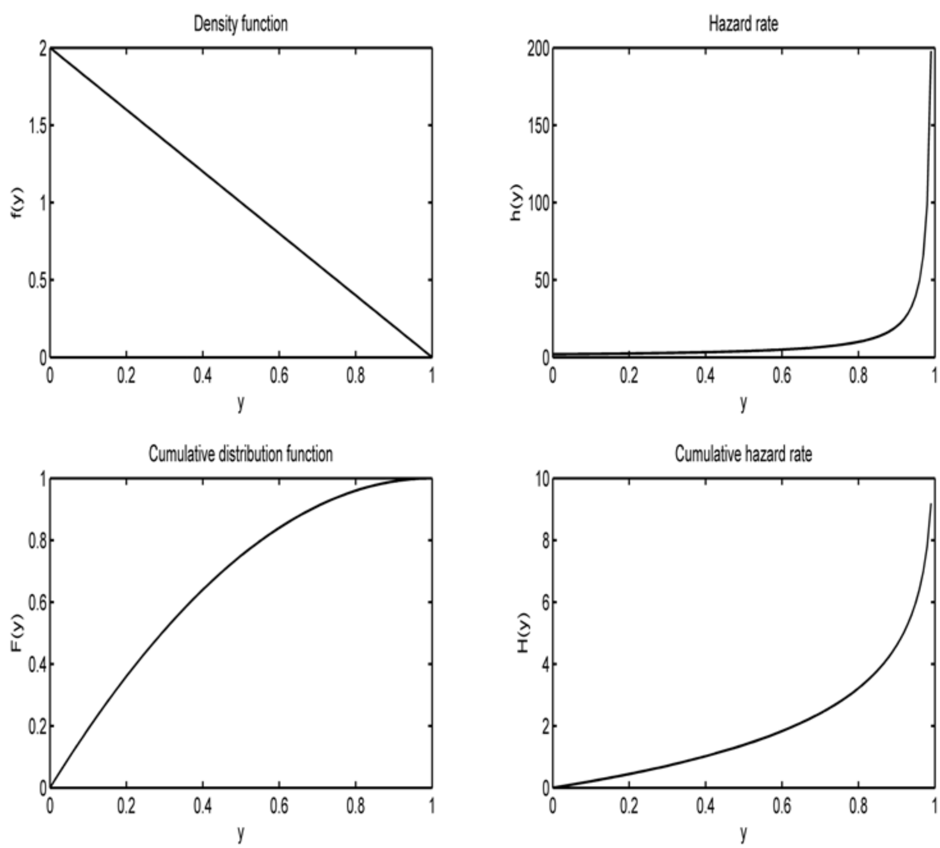
$$F(x|a,b) = 1-\left(\frac{a+b-x}{b}\right)^2 = \frac{a-x}{b}\left(\frac{x-a}{b}-2\right) \quad (5.46b)$$

$$R(x|a,b) = \left(\frac{a+b-x}{b}\right)^2 \quad (5.46c)$$

$$h(x|a,b) = \frac{2}{a+b-x} \quad (5.46d)$$

$$H(x|a,b) = 2\left[\ln b - \ln\left[(a+b-x)\right]\right] \quad (5.46e)$$

Figure 5/51: Several functions for the reduced right–angled and positively skew triangular distribution



$$F_X^{-1}(P) = x_P = a + b\left(1 - \sqrt{1-P}\right), \quad 0 \le P \le 1 \tag{5.46f}$$

$$a = x_0 \tag{5.46g}$$

$$b = x_1 - x_0 \tag{5.46h}$$

$$x_{0.5} \approx a + 0.2929\, b \tag{5.46i}$$

$$x_M = a \tag{5.46j}$$

$$M_X(t) = \frac{2\,\exp(a\,t)\left[\exp(b\,t) - b\,t - 1\right]}{(b\,t)^2} \tag{5.46k}$$

$$C_X(t) = \frac{2\,\exp(i\,a\,t)\left[1 - \exp(i\,b\,t) + i\,b\,t\right]}{(b\,t)^2} \tag{5.46l}$$

$$\mu_r'(Y) = \frac{2}{(1+r)\,(2+r)}; \quad r = 0, 1, 2\ldots; \; Y = (X-a)\big/b \tag{5.46m}$$

$$\mu_1'(Y) = \mathrm{E}(Y) = \frac{1}{3} \tag{5.46n}$$

$$\mu_2'(Y) = \frac{1}{6} \tag{5.46o}$$

$$\mu_3'(Y) = \frac{1}{10} \tag{5.46p}$$

$$\mu_4'(Y) = \frac{1}{15} \tag{5.46q}$$

$$\mu_1'(X) = \mathrm{E}(X) = a + \frac{b}{3} \tag{5.46r}$$

$$\mu_2(X) = \mathrm{Var}(X) = \frac{b^2}{18} \tag{5.46s}$$

$$\alpha_3 \approx 0.5657 \tag{5.46t}$$

$$\alpha_4 = 2.4 \tag{5.46u}$$

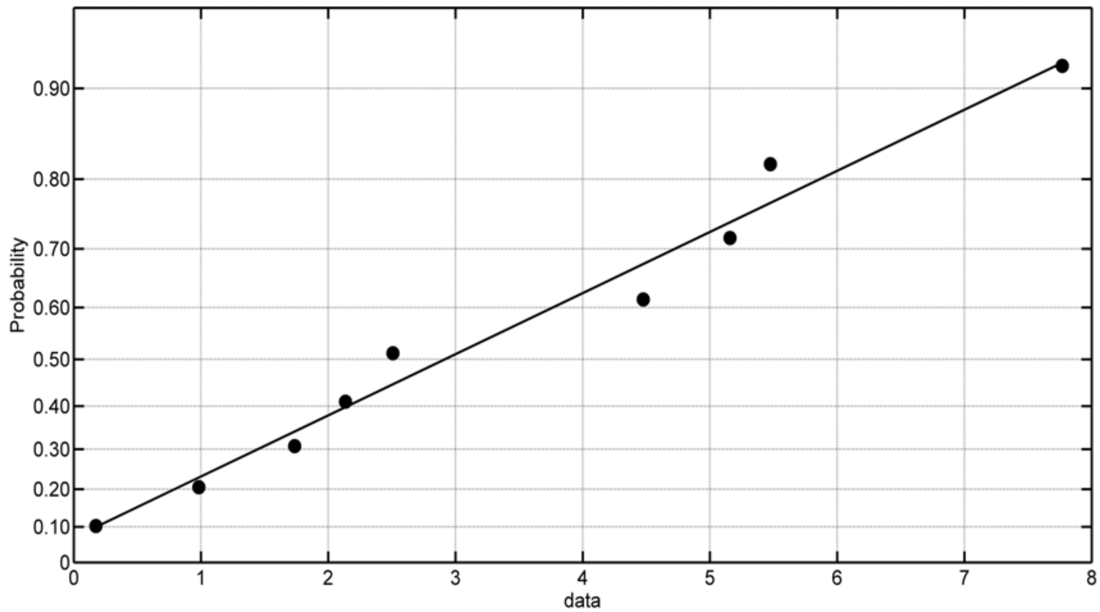
$$I(X) = \mathrm{ld}\, b + \frac{1}{\ln 4} \approx 0.7213 + \mathrm{ld}\, b \tag{5.46v}$$

$$F_Y^{-1}(P) = y_P = 1 - \sqrt{1-P} \tag{5.46w}$$

$$f_Y(y_P) = 2\sqrt{1-P} \tag{5.46x}$$

The moments of the reduced order statistics, which are computed by numerical integration, are the input to LLOYD's estimator in LEPP.

Figure 5/52: Right–angled, positively skew triangular probability paper with data and regression line




## 5.2.18 Uniform or rectangular distribution — $X \sim UN(a,b)$

The uniform distribution exits in different parameterizations. We will use

$$f(x|a,b) = \frac{1}{b} \text{ for } a \le x \le x + b. \tag{5.47a}$$

Other notations are:

$$f(x|a,b^*) = \frac{1}{b^* - a} \text{ for } a \le x \le b^*, \tag{5.47b}$$

$$f(x|a,\Delta) = \frac{1}{2\,\Delta} \text{ for } a - \Delta \le x \le a + \Delta. \tag{5.47c}$$

When a continuous variate $X$ with support $[a, a+b]$ is assumed to have a probability of falling into any sub–interval $[a^{**}, a^{**}+b^{**}]$ of $[a, a+b]$ and which is the same for all these sub–intervals and proportional to the interval–length $b^{**}$, then the variate $X$ has a uniform distribution. Thus, the uniform distribution expresses the principle insufficient reason in probability theory. It is a popular **prior distribution** in BAYESIAN statistics.

Uniform distributions are special cases of the **beta distribution** (1.12a), obtained by setting $c = d = 1$. The uniform distribution (5.47c) with $a = 0$ and $\Delta = 0.5 \cdot 10^{-k}$ is often used to represent the distribution of **roundoff errors** in values tabulated to the nearest $k$ decimal places. A rectangular distribution also arises as a result of the **probability integral transformation**. If $X$ is a continuous variate with CDF $F(x)$, then $U = F(X)$ is distributed according to the reduced uniform distribution of (5.47a) with DF $f(u|0,1)=1$ for $0 \le u \le 1$. The **distribution of the sum**

$$X = X_1 + \ldots + X_n,\ X_i \overset{\text{iid}}{\sim} UN(0,b),$$

is the so–called **IRWIN–HALL distribution** with DF

$$f(x) = \frac{1}{b^n\,(n-1)!}\sum_{j=0}^{k}(-1)^j\binom{n}{j}(x-j\,b)^{n-1},\ \ k \le \frac{x}{b} \le k+1;\ \ 0 \le k \le n-1. \tag{5.48a}$$

From (5.48a) we find — by dividing the sum $X$ by the sample size $n$ — the **distribution of the arithmetic mean** $T = X/n$. For $b=1$ in (5.48a) we then have the so–called **BATES distribution** with DF

$$f(t) = \frac{n^n}{(n-1)!}\sum_{j=0}^{[t\,n]}(-1)^j\binom{n}{j}\left(t-\frac{j}{n}\right)^{n-1},\ \ 0 \le t \le 1. \tag{5.48b}$$

When $U \sim UN(0,1)$, then $V = -\ln U$ has a reduced **exponential distribution**, and conversely, if $V \sim EX(0,1)$, then $U = \exp(-V) \sim UN(0,1)$. Furthermore, when $U \sim UN(0,1)$, then $V = -2\,\ln U$ has a **$\chi^2$–distribution** with $\nu = 2$ degrees of freedom.

We give the following functions and characteristics of the uniform distribution:

$$f(x|a,b) = \frac{1}{b},\quad a \le x \le a+b;\ a \in \mathbb{R},\ b>0 \tag{5.49a}$$

$$F(x|a,b) = \frac{x-a}{b} \tag{5.49b}$$

$$R(x|a,b) = 1-\frac{x-a}{b} = \frac{a+b-x}{b} \tag{5.49c}$$

$$h(x|a,b) = \frac{1}{a+b-x} \tag{5.49d}$$

$$H(x|a,b) = \ln b - \ln(a+b-x) \tag{5.49e}$$

$$F_X^{-1}(P) = x_P = a + b\,P \tag{5.49f}$$

$$x_{0.5} = a + b/2 \tag{5.49g}$$

$$a = x_0 \tag{5.49h}$$

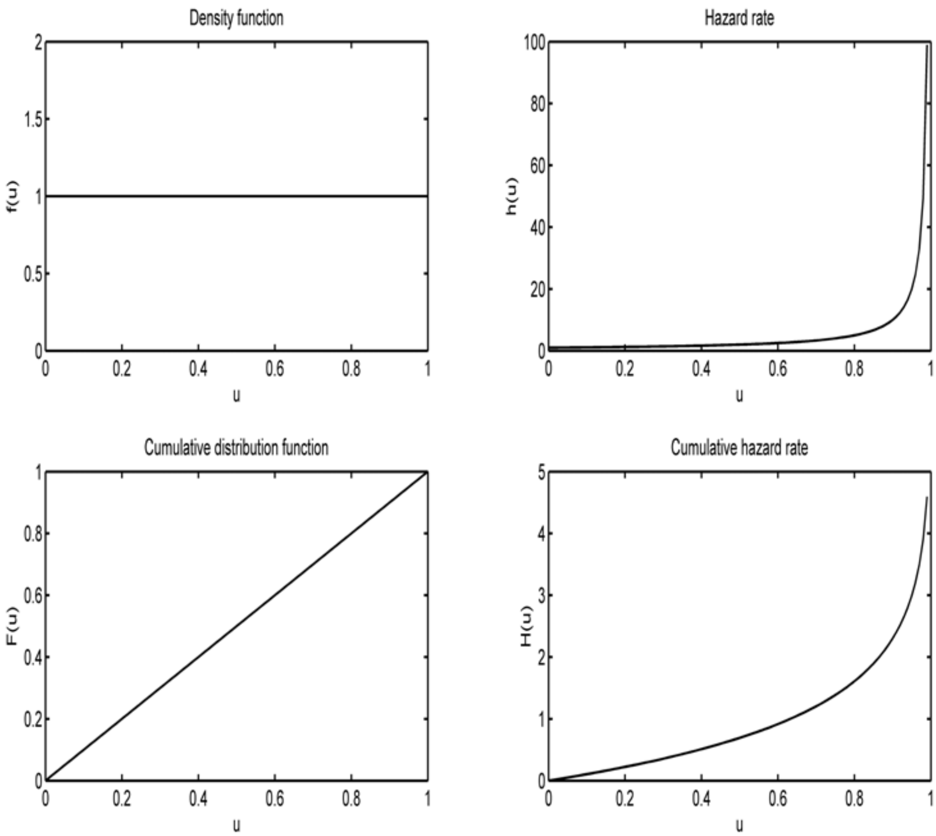
$$b = x_1 - x_0 \tag{5.49i}$$

$$x_M - \text{not defined} \tag{5.49j}$$

$$M_X(t) = \exp(a\,t)\,\frac{\exp(b\,t)-1}{b\,t} \tag{5.49k}$$

$$C_X(t) = \frac{\exp(i\,a\,t)\big[\exp(i\,b\,t)-1\big]}{i\,b\,t} \tag{5.49l}$$

Figure 5/53: Several functions for the reduced uniform or rectangular distribution




$$\mu'_r(U) = \frac{1}{1+r}; \quad r = 0, 1, 2, \ldots; \; U = (X-a)/b \tag{5.49m}$$

$$\mu'_r(X) = \frac{(a+b)^{r+1} - a^{r+1}}{b\,(r+1)}; \quad r = 0, 1, 2, \ldots \tag{5.49n}$$

$$\mu'_1(X) = \mathrm{E}(X) \;=\; a + b/2 \tag{5.49o}$$

$$\mu'_2(X) = a^2 + a\,b + b^2/3 \tag{5.49p}$$

$$\mu'_3(X) = a^3 + 3\,a^2\,b/2 + a\,b^2 + b^3/4 \tag{5.49q}$$

$$\mu'_4(X) = a^4 + 2\,a^3\,b + 2\,a^2\,b^2 + a\,b^3 + b^4/5 \tag{5.49r}$$

$$\mu_r(X) = \left\{ \begin{array}{cl} \dfrac{b^r}{2^r\,(1+r)} & \text{for } r \text{ even} \\ 0 & \text{for } r \text{ even} \end{array} \right\} \tag{5.49s}$$

$$\mu_2(X) = \mathrm{Var}(X) \;=\; b^2/12 \tag{5.49t}$$

$$\mu_4(X) = b^4/80 \tag{5.49u}$$

$$\alpha_3 = 0 \tag{5.49v}$$

$$\alpha_4 = 1.6 \tag{5.49w}$$

$$\kappa_r = \left\{ \begin{array}{cc} 0 & \text{for } r \text{ odd} \\ b^r\, \mathfrak{B}_r/b & \text{for } r \text{ even}^{41} \end{array} \right\} \tag{5.50a}$$

$$I(X) = \operatorname{ld} b = \frac{\ln b}{\ln 2} \tag{5.50b}$$

$$F_U^{-1}(P) = u_P = P,\ 0 \le P \le 1 \tag{5.50c}$$

$$f_U(u_P) = 1 \tag{5.50d}$$

For the reduced uniform variable we have $f_U(u) = 1,\ 0 \le u \le 1$, and $F_U(u) = u$, $0 \le u \le 1$. Thus, the DF of the $r$–th reduced order statistics $Y_{r:n}$ is

$$f_{r:n}(u) = \frac{1}{(r-1)!\,(n-r)!}\, u^{r-1}\,(1-u)^{n-r},\ \ 0 \le u \le 1, \tag{5.51a}$$

which is nothing but the DF of the reduced beta distribution with parameters $c = r$ and $d = n - r + 1$. So, the single raw moments easily follow as:

$$\begin{aligned} \alpha_{r:n}^{(k)} &= \mathrm{E}\big(U_{r:n}^k\big) = \int_0^1 u^k\, f_{r:n}(u)\,\mathrm{d}u \\ &= \frac{B(r+k, n-r+1)}{B(r, n-r+1)} \\ &= \frac{n!}{(n+k)!}\,\frac{(r+k-1)!}{(r-1)!};\ \ r = 1,\ldots,n;\ k = 0,1,2,\ldots; \end{aligned} \tag{5.51b}$$

especially

$$\alpha_{r:n} = \mathrm{E}(U_{r:n}) = \frac{r}{n+1} =:\ p_r,^{42} \tag{5.51c}$$

$$\alpha_{r:n}^{(2)} = \mathrm{E}(U_{r:n}^2) = \frac{r+1}{n+2}\,\frac{r}{n+1}. \tag{5.51d}$$

The variance of $U_{r:n}$ follows as

$$\beta_{r,r:n} = \mathrm{Var}(U_{r:n}) = \frac{p_r\,(1-p_r)}{n+2}. \tag{5.51e}$$

The joint DF of $U_{r:n}$ and $U_{s:n},\ 1 \le r < s \le n$, according to (2.6a) is

$$f_{r,s:n}(u,v) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}\,u^{r-1}\,(v-u)^{s-r-1}\,(1-v)^{n-s}. \tag{5.51f}$$

---

41 $\mathfrak{B}_r$ is a BERNOULLI number.

42 $p_r = r/(n+1)$ is equal to WEIBULL's plotting position (3.6b).

From (5.51f) we obtain the $(k_r, k_s)$–th product moment of $(U_{r:n}, U_{s:n})$ as

$$\begin{aligned}
\alpha_{r,s:n}^{(k_r,k_s)} &= \mathrm{E}\big(U_{r:n}^{k_r}\, U_{s:n}^{k_s}\big) = \int\limits_0^1 \int\limits_0^v u^{k_r}\, v^{k_s}\, f_{r,s:n}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} B(r+k_r, s-r)\, B(s+k_r+k_s, n-s+1) \\
&= \frac{n!}{(n+k_r+k_s)!}\,\frac{(r+k_r-1)!}{(r-1)!}\,\frac{(s+k_r+k_s-1)!}{(s+k_r-1)!} \qquad (5.51g)
\end{aligned}$$

From (5.51g) we only need the special case $k_r = k_s = 1$:

$$\alpha_{r,s:n} = \mathrm{E}\big(U_{r:n}\, U_{s:n}\big) = \frac{r\,(s+1)}{(n+1)\,(n+2)}, \qquad (5.51h)$$

which is — in combination with $\mathrm{E}(U_{r:n}) = p_r = r\big/(n+1)$ and $\mathrm{E}(U_{s:n}) = p_s = s\big/(n+1)$ — gives the covariance

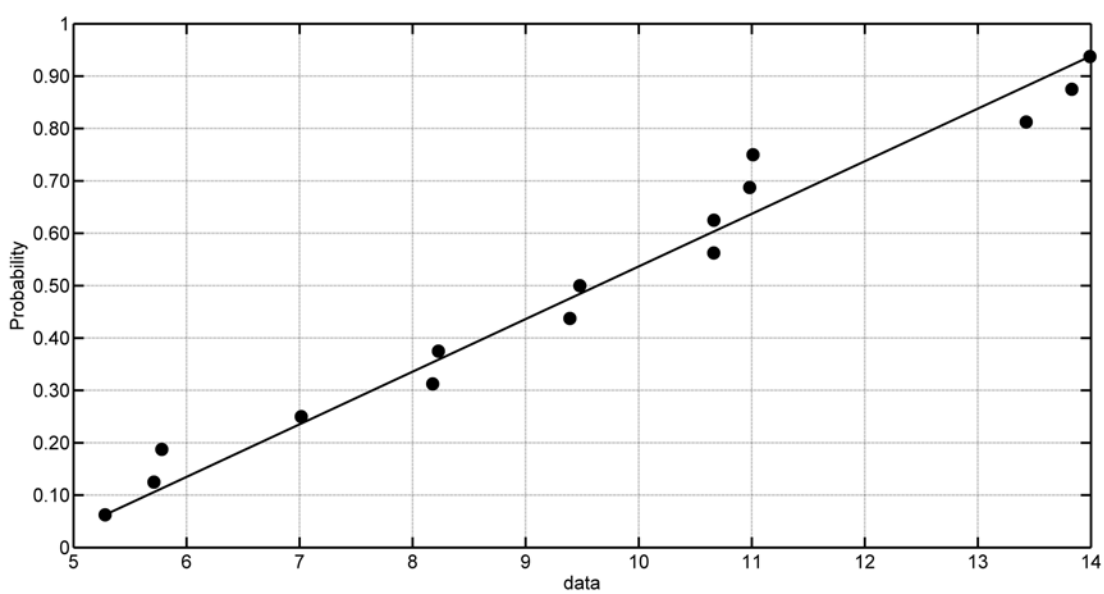
$$\begin{aligned}
\beta_{r,s:n} &= \frac{r\,(s+1)}{(n+1)\,(n+2)} - \frac{r}{n+1}\,\frac{s}{n+1} \\
&= \frac{p_r\,(1-p_s)}{n+2},\ \ 1 \le r < s \le n. \qquad (5.51i)
\end{aligned}$$

We may combine the variance formula (5.51e) and the covariance formula (5.51i) into one formula:

$$\beta_{r,s:n} = \frac{p_r\,(1-p_s)}{n+2},\ \ 1 \le r \le s \le n. \qquad (5.51j)$$

In LEPP we have used (5.51c) and (5.51j) together with LLOYD's estimator.

Figure 5/54: Uniform or rectangular probability paper with data and regression line



### 5.2.19 U–shaped beta distribution — $X \sim UB(a,b)$

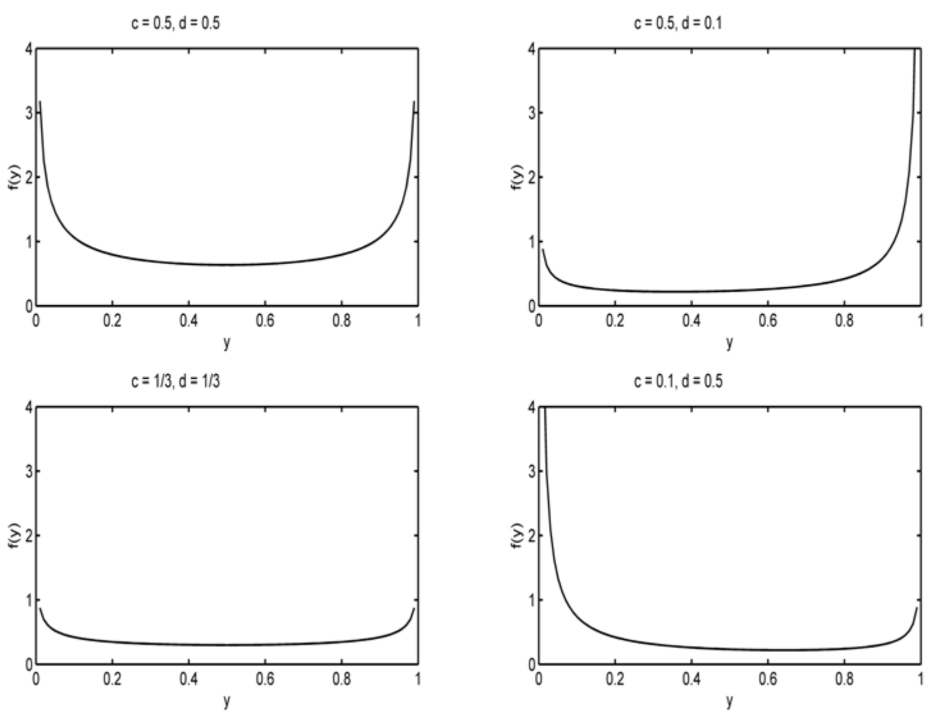
Due to its two shape parameters $c$ and $d$ the beta DF

$$f(x|a,b,c,d) = \frac{\left(\dfrac{x-a}{b}\right)^{c-1}\left(1-\dfrac{x-a}{b}\right)^{d-1}}{b\,B(c,d)};\, a \le x \le a+b;\; a \in \mathbb{R};\; b,c,d > 0;$$

can take a great variety of shapes:

- for $c = d$ it is symmetric,
- for $c > d$ $(c < d)$ it is positively (negatively) skew,
- for $c = 1$ and/or $d = 1$ it is linear,
- for $c \neq 1$ and/or $d \neq 1$ it is curved,
- for $c \le 1$ and $d > 1$ $(d \le 1$ and $c > 1)$ it has a mode at the left–hand (right–hand) border of the support,
- for $c > 1$ and $d > 1$ it has a mode between $a$ and $a + b$,
- for $0 < c < 1$ and $0 < d < 1$ it is U–shaped with an antimode, see Fig. 5/55, and
  - either symmetric for $0 < c = d < 1$,
  - or asymmetric for $0 < c \neq d < 1$.

Figure 5/55: Density functions of several U–shaped reduced beta distributions



We will discuss and present results for the symmetric U–shaped beta distribution with $c = d = 0.5$, which is similar to the U–shaped parabolic distribution of Sect. 5.2.12.1. An evident difference between these two U–shaped distributions is the height of the density above the antimode which is zero for the parabolic case and greater than zero for the beta case. The symmetric U–shaped beta density goes to a rectangular density for $c = d \to 1$, and for $c = d \to 0$ the central part of the density comes down to the abscissa whereas the flanks become very steep.

The following results are special cases of the beta distribution, see JOHNSON/KOTZ/BALAKRISHNAN (1995, Chapter 25), when $c = d = 0.5$.

$$\begin{aligned}
f(x|a,b) &= \frac{\left(\dfrac{x-a}{b}\right)^{-0.5}\left(1-\dfrac{x-a}{b}\right)^{-0.5}}{b\,B(0.5,0.5)} \\
&= \frac{1}{\pi\sqrt{(x-a)(a+b-x)}},\; a \le x \le a+b,\; a \in \mathbb{R},\; b > 0 \qquad (5.52a)
\end{aligned}$$

$$F(x|a,b) = \frac{2\,\arcsin\left(\sqrt{\dfrac{x-a}{b}}\right)}{\pi} \qquad (5.52b)$$

$$R(x|a,b) = \frac{2\,\arccos\left(\sqrt{\dfrac{x-a}{b}}\right)}{\pi} \qquad (5.52c)$$

$$h(x|a,b) = \frac{1}{2\,\arccos\left(\sqrt{\dfrac{x-a}{b}}\right)\sqrt{(x-a)\,(a+b-x)}} \qquad (5.52d)$$

$$H(x|a,b) = \ln\pi - \ln 2 - \ln\left[\arccos\left(\frac{x-a}{b}\right)\right] \qquad (5.52e)$$

$$F_X^{-1}(P) = a + b\,\sin^2\left(\frac{\pi\,P}{2}\right),\; 0 \le P \le 1 \qquad (5.52f)$$

$$a = x_0 \qquad (5.52g)$$

$$b = x_1 - x_0 \qquad (5.52h)$$

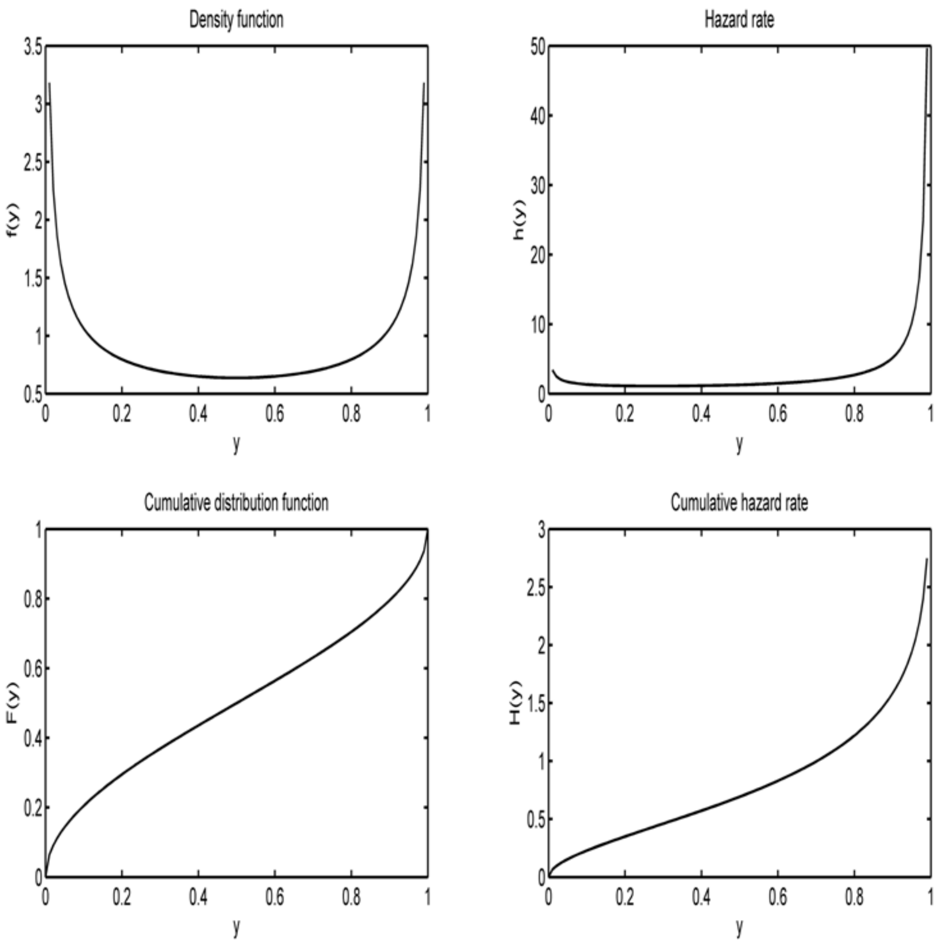
$$x_{0.5} = a + \frac{b}{2} \qquad (5.52i)$$

$$x = a + \frac{b}{2} \text{ is an antimode} \qquad (5.52j)$$

$$M_X(t) = \exp\left[\left(a+\frac{b}{2}\right)t\right] I_0\left(\frac{b\,t}{2}\right)\;{}^{43} \qquad (5.52k)$$

[43] $I_\bullet(\bullet)$ is the modified BESSEL function, see ABRAMOWITZ/STEGUN (1972, p. 376).

Figure 5/56: Several functions for the reduced U–shaped beta distribution




$$\mu'_r(Y) = \frac{1 \cdot 3 \cdot \ldots \cdot (2\,r-1)}{2^r\,r!}; \quad r = 1, 2, \ldots; Y = (X-a)\big/b \tag{5.52l}$$

$$\mu'_1(Y) = \mathrm{E}(Y) = \frac{1}{2} \tag{5.52m}$$

$$\mu'_2(Y) = \frac{3}{8} \tag{5.52n}$$

$$\mu'_3(Y) = \frac{5}{16} \tag{5.52o}$$

$$\mu'_4(Y) = \frac{35}{128} \tag{5.52p}$$

$$\mu'_1(X) = \mathrm{E}(X) = a + \frac{b}{2} \tag{5.52q}$$

$$\mu_2(X) = \mathrm{Var}(X) = \frac{b^2}{18} \tag{5.52r}$$

$$\alpha_3 = 0 \tag{5.52s}$$

$$\alpha_4 = 1.5 \tag{5.52t}$$

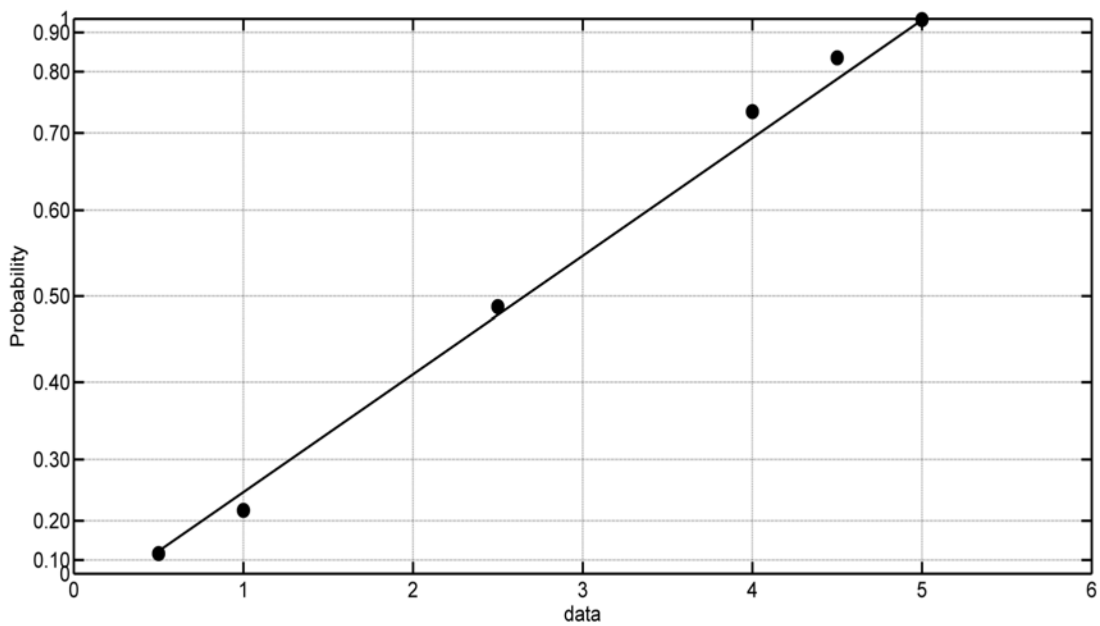
$$I(X) \approx \mathrm{ld}\, b - 0.3485 \tag{5.52u}$$

$$F_Y^{-1}(P) = y_P = \sin^2\left(\frac{\pi\, P}{2}\right), \ \ 0 \le P \le 1 \tag{5.52v}$$

$$f_Y(y_P) = \frac{1}{\pi\, \sin\left(\frac{\pi\, P}{2}\right)\sqrt{1-\sin^2\left(\frac{\pi\, P}{2}\right)}} \tag{5.52w}$$

The moments of the reduced order statistics can be computed by numerical integration and are the input to LLOYD's estimator in LEPP.

Figure 5/57: U–shaped beta probability paper with data and regression line




## 5.2.20 V–shaped distribution — $X \sim VS(a, b)$

This distribution with a density function, whose graph is given by the letter V, may be regarded as a linear approximation to the U–shaped parabolic distribution of Sect. 5.2.12.1. The functions and characteristics are:

$$f(x|a,b) = \left\{\begin{array}{ll} \dfrac{a-x}{b^2} & \text{for } a-b \le x \le a \\ \dfrac{x-a}{b^2} & \text{for } a \le x \le a+b \end{array}\right\}, \ \ a \in \mathbb{R},\ b > 0 \tag{5.53a}$$

$$F(x|a,b) = \left\{ \begin{array}{ll} \dfrac{1}{2}\left[1 - \left(\dfrac{a-x}{b}\right)^2\right] & \text{for } a-b \le x \le a \\[2ex] \dfrac{1}{2}\left[1 + \left(\dfrac{a-x}{b}\right)^2\right] & \text{for } a \le x \le a+b \end{array} \right\} \tag{5.53b}$$

$$R(x|a,b) = \left\{ \begin{array}{ll} \dfrac{1}{2}\left[1 + \left(\dfrac{a-x}{b}\right)^2\right] & \text{for } a-b \le x \le a \\[2ex] \dfrac{1}{2}\left[1 - \left(\dfrac{a-x}{b}\right)^2\right] & \text{for } a \le x \le a+b \end{array} \right\} \tag{5.53c}$$

$$h(x|a,b) = \left\{ \begin{array}{ll} \dfrac{2\,(a-x)}{b^2 + (a-x)^2} & \text{for } a-b \le x \le a \\[2ex] \dfrac{2\,(a-x)}{(a-x)^2 - b^2} & \text{for } a \le x \le a+b \end{array} \right\} \tag{5.53d}$$

$$H(x|a,b) = \left\{ \begin{array}{ll} -\ln\left\{0.5\left[1 + \left(\dfrac{a-x}{b}\right)^2\right]\right\} & \text{for } a-b \le x \le a \\[2ex] -\ln\left\{0.5\left[1 - \left(\dfrac{a-x}{b}\right)^2\right]\right\} & \text{for } a \le x \le a+b \end{array} \right\} \tag{5.53e}$$

$$F_X^{-1}(P) = x_P = \left\{ \begin{array}{ll} a - b\sqrt{1-2\,P} & \text{for } 0 \le P \le 0.5 \\[1ex] a + b\sqrt{2\,P-1} & \text{for } 0.5 \le P \le 1 \end{array} \right\} \tag{5.53f}$$

$$a = x_{0.5} \tag{5.53g}$$

$$b = x_{0.5} - x_0 = x_1 - x_{0.5} \tag{5.53h}$$

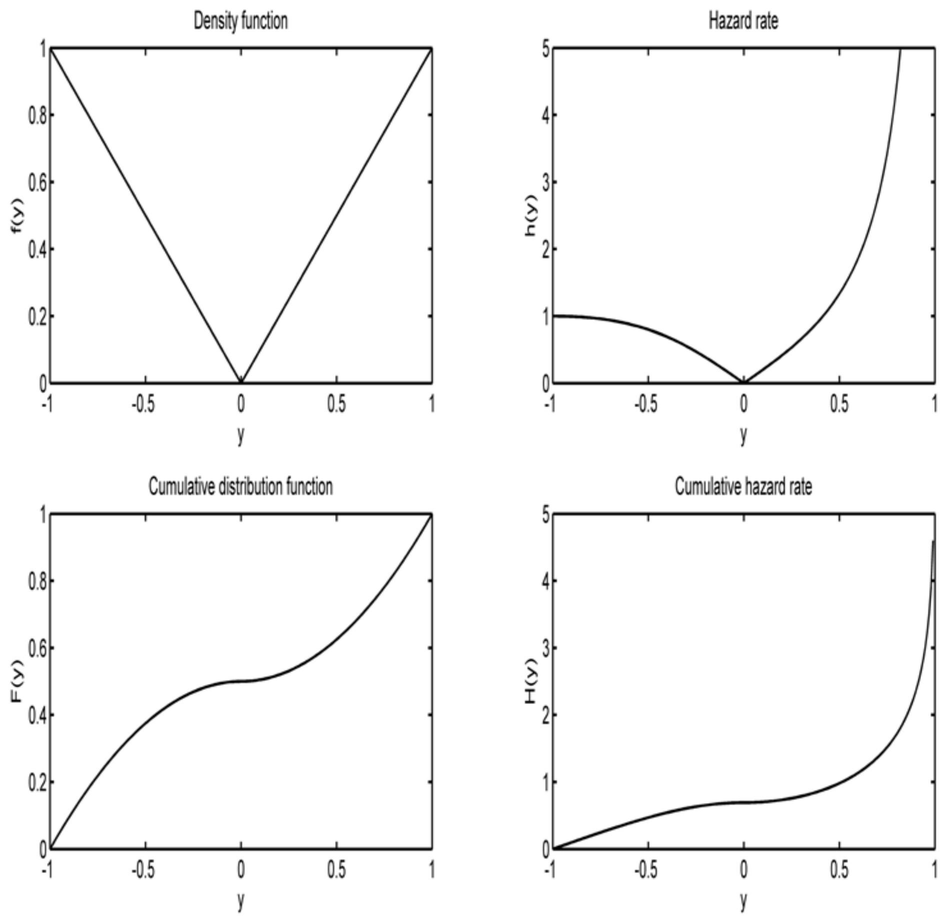
$$x = a \text{ is an antimode} \tag{5.53i}$$

$$M_X(t) = \frac{\exp(a\,t)\big\{2 + \exp(b\,t)\,[b\,t-1] - \exp(-b\,t)\,[b\,t+1]\big\}}{(b\,t)^2} \tag{5.53j}$$

$$\mu_r'(Y) = \left\{ \begin{array}{ll} 0 & \text{for } r \text{ odd} \\[1ex] \dfrac{2}{2+r} & \text{for } r \text{ even} \end{array} \right\},\quad Y = (X-a)/b \tag{5.53k}$$

$$\mu_r'(X) = \frac{2\,a^{r+2} + (a+b)^{r+1}\,(b + r\,b - a) - (a-b)^{r+1}\,(b + r\,b + a)}{b^2\,(1+r)\,(2+r)} \tag{5.53l}$$

Figure 5/58: Several functions for the reduced V–shaped distribution




$$\mu_1'(X) = \mathrm{E}(X) = a \tag{5.53m}$$

$$\mu_2'(X) = a^2 + \frac{b^2}{2} \tag{5.53n}$$

$$\mu_2(X) = \mathrm{Var}(X) = \frac{b^2}{2} \tag{5.53o}$$

$$\alpha_3 = 0 \tag{5.53p}$$

$$\alpha_4 = \frac{4}{3} \tag{5.53q}$$

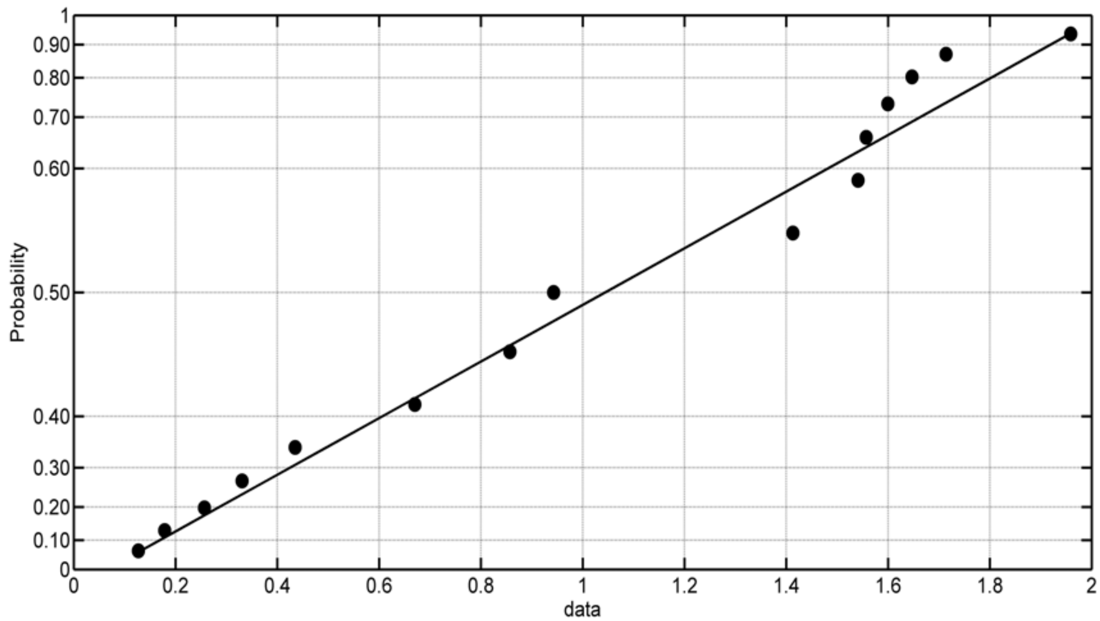
$$I(X) = \mathrm{ld}\, b + \frac{1}{2\,\ln 2} \approx 0.7213 + \mathrm{ld}\, b \tag{5.53r}$$

$$F_Y^{-1}(P) = y_P = \left\{ \begin{array}{ll} -\sqrt{1-2\,P} & \text{for } 0 \le P \le 0.5 \\ \sqrt{2\,P-1} & \text{for } 0.5 \le P \le 1 \end{array} \right\} \tag{5.53s}$$

$$f_Y(y_P) = \sqrt{|1-2\,P|} \ \text{ for } 0 \le P \le 1 \tag{5.53t}$$

Moments of the reduced order statistics can be computed by numerical integration and are the input to LLOYD's estimator in LEPP.

Figure 5/59: V–shaped probability paper with data and regression line




## 5.3 The case of ln–transformable distributions

In Sect. 1.3 we have presented eight distributions which — after a ln–transformation of its variable — will be of location–scale type. These distributions have three parameters, one of them being a location or shift parameter in the role of either a lower threshold (This is the case for the maximum type II and minimum type III extreme value distributions, the lognormal, the PARETO and the power–function distributions.) or an upper threshold (This is the case for the minimum type II and the maximum type III extreme value distributions and also another form of the lognormal distribution.). The shift parameter $a$ has to be known or it has to be estimated when linear estimates of the remaining two suitably transformed parameters based on the ln–transformed sample data have to be found.

### 5.3.1 Estimation of the shift parameter

The estimation of a shift parameter which is a threshold value or bound is particularly inaccurate and sometimes difficult as might be seen when reading the monographs of the distributions, which will be discussed further down. A very popular approach, which is of maximum likelihood type, but biased, is to take $\widehat{a}_\ell = X_{1:n}$, the sample minimum when, $a$ is a lower threshold, or to take $\widehat{a}_u = X_{n:n}$, the sample maximum, for $a$ being an upper bound. These estimators are always admissible, i.e. no observation is smaller (or greater) than the estimate, but these estimators do not fully take into account the complete sample values and thus are insufficient.

We will present an approach which is based on all sample values[44] and is working for all bounded distributions whether bounded on only one side or on both sides. The approach is due to COOKE (1979) and has a higher asymptotic efficiency than the popular bias–corrected estimators $\widehat{a}_u = X_{n:n} + \big(X_{n:n} - X_{n-1:n}\big)$ for the upper bound or $\widehat{a}_\ell = X_{1:n} - \big(X_{2:n} - X_{1:n}\big)$ for the lower bound.

COOKE's estimator is derived by expressing the expectation of the first (last) order statistic in terms of an integral, the integrand being $x_{1:n}$ $(x_{n:n})$ times its DF. Then one uses integration by parts, replaces $\mathrm{E}(X_{1:n})$ $\big[\mathrm{E}(X_{n:n})\big]$ by $X_{1:n}$ $\big[X_{n:n}\big]$ and finally replaces the distribution function by the sample or empirical distribution function and solves for $\widehat{a}_\ell$ $\big[\widehat{a}_u\big]$. The following excursus gives the details in case of an upper threshold.

**Excursus: COOKE's estimator for an upper threshold $\boldsymbol{a_u}$**

Let $F(x)$ be the CDF of a random variable $X$ which is assumed to be unbounded on the left.[45] The sample maximum $X_{n:n}$ has CDF

$$F_{n:n}(x) = \big[F(x)\big]^n \tag{5.54a}$$

and mean

$$\mu_{n:n} = \mathrm{E}(X_{n:n}) = \int\limits_{-\infty}^{a_u} x \,\mathrm{d}F_{n:n}(x). \tag{5.54b}$$

We integrate (5.54b) by parts

$$\begin{aligned} \mu_{n:n} &= \Big[x\, F_{n:n}(x)\Big]_{-\infty}^{a_u} - \int\limits_{-\infty}^{a_u} F_{n:n}(x)\mathrm{d}x \\ &= a_u - \int\limits_{-\infty}^{a_u} F_{n:n}(x)\,\mathrm{d}x, \end{aligned} \tag{5.54c}$$

and write

$$a_u = \mu_{n:n} + \int\limits_{-\infty}^{a_u} F_{n:n}(x)\,\mathrm{d}x. \tag{5.54d}$$

This suggests — observing (5.54a) — the estimator

$$\widehat{a}_u = X_{n:n} + \int\limits_{x_{1:n}}^{x_{n:n}} \big[\widehat{F}(x)\big]^n\,\mathrm{d}x \tag{5.54e}$$

[44] This procedure presupposes an uncensored sample with non–grouped data.

[45] The estimator to be developed also holds for a left–bounded distribution with either a known or an unknown bound.

where $\widehat{F}(x)$ is the empirical CDF:

$$\widehat{F}(x) = \left\{ \begin{array}{ll} 0 & \text{for } x < x_{1:n} \\ \dfrac{i}{n} & \text{for } x_{i:n} \leq x < x_{i+1:n};\ i = 1, 2, \ldots, n-1 \\ 1 & \text{for } x \geq x_{n:n}. \end{array} \right\} \tag{5.54f}$$

The integral in (5.54e) turns into

$$\int\limits_{x_{1:n}}^{x_{n:n}} \big[\widehat{F}(x)\big]^n \,\mathrm{d}x = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^n \big(X_{i+1:n} - X_{i:n}\big) \tag{5.54g}$$

which after some manipulation gives

$$\int\limits_{x_{1:n}}^{x_{n:n}} \big[\widehat{F}(x)\big]^n \,\mathrm{d}x = X_{n:n} - \sum_{i=0}^{n-1} \left[\left(1 - \frac{i}{n}\right)^n - \left(1 - \frac{i+1}{n}\right)^n\right] X_{n-i:n}. \tag{5.54h}$$

So, the suggested estimator (5.54e) is

$$\widehat{a}_u = 2\, X_{n:n} - \sum_{i=0}^{n-1} \left[\left(1 - \frac{i}{n}\right)^n - \left(1 - \frac{i+1}{n}\right)^n\right] X_{n-i:n}. \tag{5.54i}$$

For large sample size $n$ (5.54i) can be approximated by

$$\widehat{a}_u = 2\, X_{n:n} - \big(1 - e^{-1}\big) \sum_{i=0}^{n-1} e^{-i}\, X_{n-i:n}. \tag{5.54j}$$

For estimating a lower threshold $a_\ell$ we proceed in a similar way. Starting with

$$\mu_{1:n} = \mathrm{E}(X_{1:n}) = \int\limits_{a_\ell}^{\infty} x\,\mathrm{d}F_{1:n}(x), \tag{5.55a}$$

where $F_{1:n}(x) = \big[1 - F(x)\big]^n$, we find after integration by parts:

$$\mu_{1:n} = a_\ell + \int\limits_{a_\ell}^{\infty} \big[1 - F(x)\big]^n \,\mathrm{d}x. \tag{5.55b}$$

The resulting estimator is

$$\widehat{a}_\ell = 2\, X_{1:n} - \sum_{i=1}^{n} \left[\left(1 - \frac{i-1}{n}\right)^n - \left(1 - \frac{i}{n}\right)^n\right] X_{i:n}, \tag{5.55c}$$

which — for large $n$ — turns into

$$\widehat{a}_\ell = 2\, X_{1:n} - (e - 1) \sum_{i=1}^{n} e^{-i}\, X_{i:n}. \tag{5.55d}$$

COOKE's estimators have been implemented in LEPP for uncensored and non–grouped sample data, but the user is asked by the program whether he wants it to estimate the threshold or he wants to input his own threshold value. For all other types of data input the user has to provide a threshold value!

We remind the reader that the estimates of the the scale and shape parameters are very sensitive with respect to the threshold. A good idea is to try several threshold values and take that one which maximizes the coefficient of correlation between the transformed data and the regressor.

**Example 5/4: WEIBULL parameter estimation using different threshold values**

The following 16 WEIBULL order statistics have been simulated with $a = b = c = 2$:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2.3745 | 2.4348 | 2.6393 | 2.9097 | 3.1388 | 3.2352 | 3.3864 | 3.5200 |
| 3.5360 | 3.6406 | 4.0390 | 4.1789 | 4.4086 | 4.5285 | 4.6338 | 5.4652 |

Using the ln-transformed order statistics, transformed with different values for $a$, and applying the estimation procedure of Sect. 5.3.2.4 we find the following results:

| $a$ | $\widehat{b}$ | $\widehat{c}$ | $r\big(\alpha_{r:n}, \widetilde{x}_{r:n}\big)$ |
|---|---|---|---|
| 0 | 4.0054 | 4.5111 | 0.9732 |
| 0.5 | 3.4905 | 3.9034 | 0.9765 |
| 1 | 2.9695 | 3.2832 | 0.9805 |
| 1.5 | 2.4368 | 2.6388 | 0.9853 |
| 2 | 1.8740 | 1.9292 | 0.9888 |
| 2.1 | 1.7518 | 1.7658 | 0.9881 |
| 2.2 | 1.6209 | 1.5830 | 0.9854 |
| 2.3164 | 1.4389 | 1.3034 | 0.9731 |
| 2.35 | 1.3624 | 1.1707 | 0.9599 |
| 2.37 | 1.2723 | 1.0020 | 0.9269 |

$\widehat{a} = 2.3164$ is COOKE's estimate. With $a$ approaching its input value 2 from either below or above the the correlation coefficient increases and the estimates of $b$ and $c$ come closer to their input values.

### 5.3.2 Extreme value distributions of types II and III

The extreme value distributions of types II and III as well as the PARETO and power–function distributions have — besides $a$ and $b$ — a third parameter $c$ which is responsible for the shape. For known $c$ we may design a special probability paper and then linearly estimate the parameters $a$ and $b$. We have decided to take a different approach in parameter estimation for these distributions. This approach is based on the suitably transformed sample data using either a known or a separately estimated location parameter $a$ resulting in a location–scale distribution, whose parameters $\widetilde{a} = \ln b$ — or $\widetilde{a} = -\ln b$ — and $\widetilde{b} = 1/c$ will be estimated along the lines of Section 5.2. The estimates may be solved for estimates of the original parameters $b$ and $c$. In LEPP we give the estimates of $\widetilde{a}$ and $\widetilde{b}$ as well as their re–transformations. The type II and III extreme value distribution can be transformed to type I distributions as has been stated in Sect. 1.3 so that we can fall back on the procedures of Sect. 5.2.5 to estimate the transformation of the parameters $b$ and $c$.

#### 5.3.2.1 Type II maximum or inverse WEIBULL distribution — $X \sim EMX2(a, b, c)$

This FRÉCHET–type distribution is the limiting distribution of the sample maximum when the sample comes from a distribution that is unlimited from above and is of CAUCHY–type, i.e. it has a fat tail on the right–hand side such that for some positive $k$ and $A$ we have

$$\lim_{x\to\infty} x^k\big[1 - F(x)\big] = A.$$

A consequence will be that moments will not always exist for a type II maximum extreme value distribution. In

$$f(x|a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{-c-1}\exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right],\quad \left\{\begin{array}{l} x \geq a,\ a \in \mathbb{R} \\ b > 0,\ c > 0 \end{array}\right\} \tag{5.56a}$$

the parameter $a$ is a lower threshold, $b$ is a scale parameter and $c$ is responsible for the shape.

$$F(x|a,b,c) = \exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right] \tag{5.56b}$$

$$R(x|a,b,c) = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right] \tag{5.56c}$$

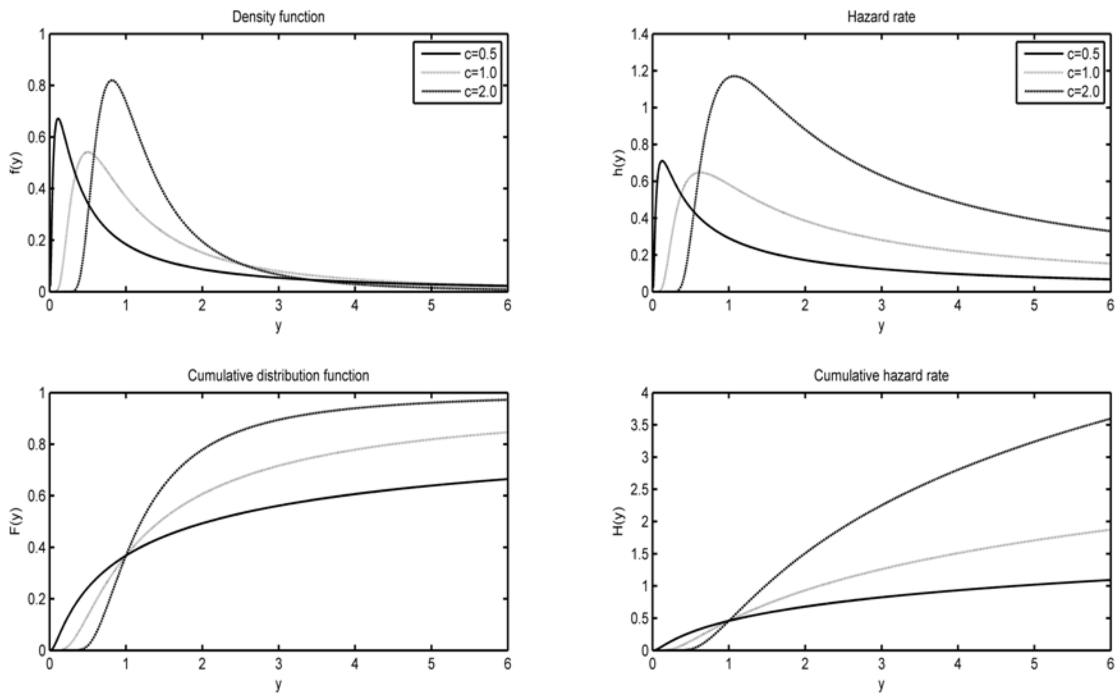
$$h(x|a,b,c) = \frac{c\left(\dfrac{x-a}{b}\right)^{-c}}{(x-a)\left\{1 - \exp\left[-\left(\dfrac{x-a}{b}\right)^{-c}\right]\right\}} \tag{5.56d}$$

$$H(x|a,b,c) = -\ln\left\{1 - \exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right]\right\} \tag{5.56e}$$

$$F_X^{-1}(P) = x_P = a + b\big(-1/\ln P\big)^{1/c},\ \ 0 \leq P < 1 \tag{5.56f}$$

$$
\begin{aligned}
a &= x_0 && \text{(5.56g)}\\
b &\approx x_{0.3679} - a && \text{(5.56h)}\\
x_{0.5} &\approx a + 1.4427^{1/c}\, b && \text{(5.56i)}\\
x_M &= a + b\left(\frac{c}{1+c}\right)^{1/c} && \text{(5.56j)}
\end{aligned}
$$

Figure 5/60: Several functions for the reduced type II maximum extreme value distribution



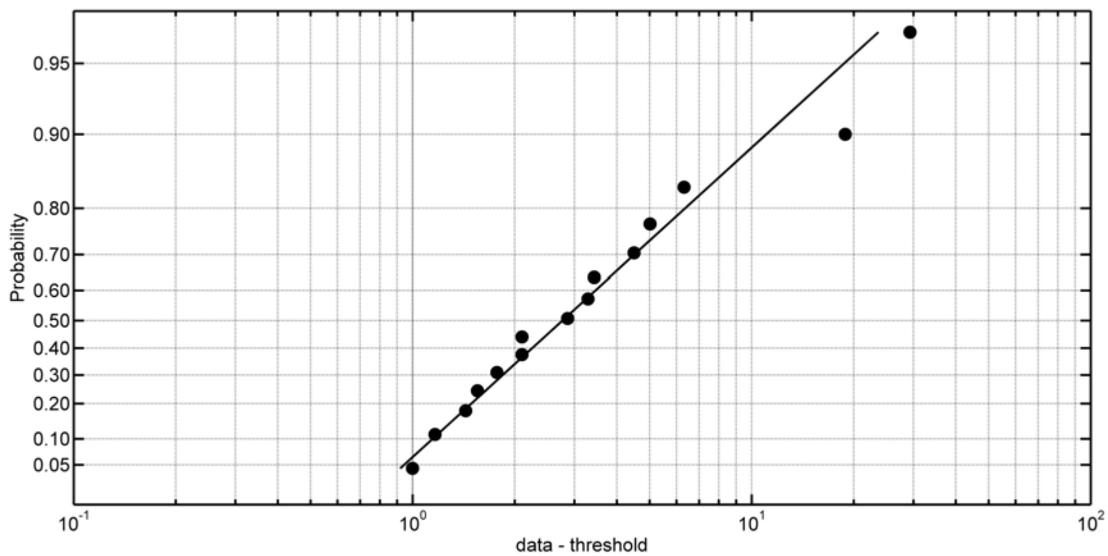

In order to write the formulas for the moments of the maximum and the minimum type II extreme value distributions as concise as possible we introduce the abbreviation:

$$
{}_{-}\Gamma_r := \Gamma\Big(1 - \frac{r}{c}\Big).
$$

$$
\begin{aligned}
\mu_r'(Y) &= {}_{-}\Gamma_r,\;\; c > r;\;\; Y = (X-a)\big/ b && \text{(5.56k)}\\
\mu_1'(X) &= \mathrm{E}(X) \;=\; a + b\,{}_{-}\Gamma_1,\;\; c > 1 && \text{(5.56l)}\\
\mu_2(X) &= \mathrm{Var}(X) \;=\; b^2\big[{}_{-}\Gamma_2 - {}_{-}\Gamma_1^2\big],\;\; c > 2 && \text{(5.56m)}\\
\alpha_3 &= \frac{{}_{-}\Gamma_3 - 3\,{}_{-}\Gamma_2\,{}_{-}\Gamma_1 + 2\,{}_{-}\Gamma_1}{\big[{}_{-}\Gamma_2 - {}_{-}\Gamma_1^2\big]^{3/2}},\;\; c > 3 && \text{(5.56n)}\\
\alpha_4 &= \frac{{}_{-}\Gamma_4 - 4\,{}_{-}\Gamma_3\,{}_{-}\Gamma_1 + 6\,{}_{-}\Gamma_2\,{}_{-}\Gamma_1^2 - 3\,{}_{-}\Gamma_1^4}{\big[{}_{-}\Gamma_2 - {}_{-}\Gamma_1^2\big]^2},\;\; c > 4 && \text{(5.56o)}\\
F_Y^{-1}(P) &= y_P \;=\; \big(-1\big/\ln P\big)^{1/c},\;\; 0 \le P < 1 && \text{(5.56p)}\\
f_Y(y_P) &= c\,P\,(-\ln P)\,(-\ln P)^{1/c} && \text{(5.56q)}
\end{aligned}
$$

$\widetilde{X} = \ln(X - a)$ has a type I maximum extreme value distribution with parameters $\widetilde{a} = \ln b$ and $\widetilde{b} = 1/c$. Thus, we apply the procedure of Sect. 5.2.5.1 to find linear estimates of $\widetilde{a}$ and $\widetilde{b}$, which finally may be solved for estimates of $b$ and $c$.

Figure 5/61: Type II maximum extreme value probability paper with data and regression line




#### 5.3.2.2 Type III maximum or reflected WEIBULL distribution — $X \sim EMX3(a, b, c)$

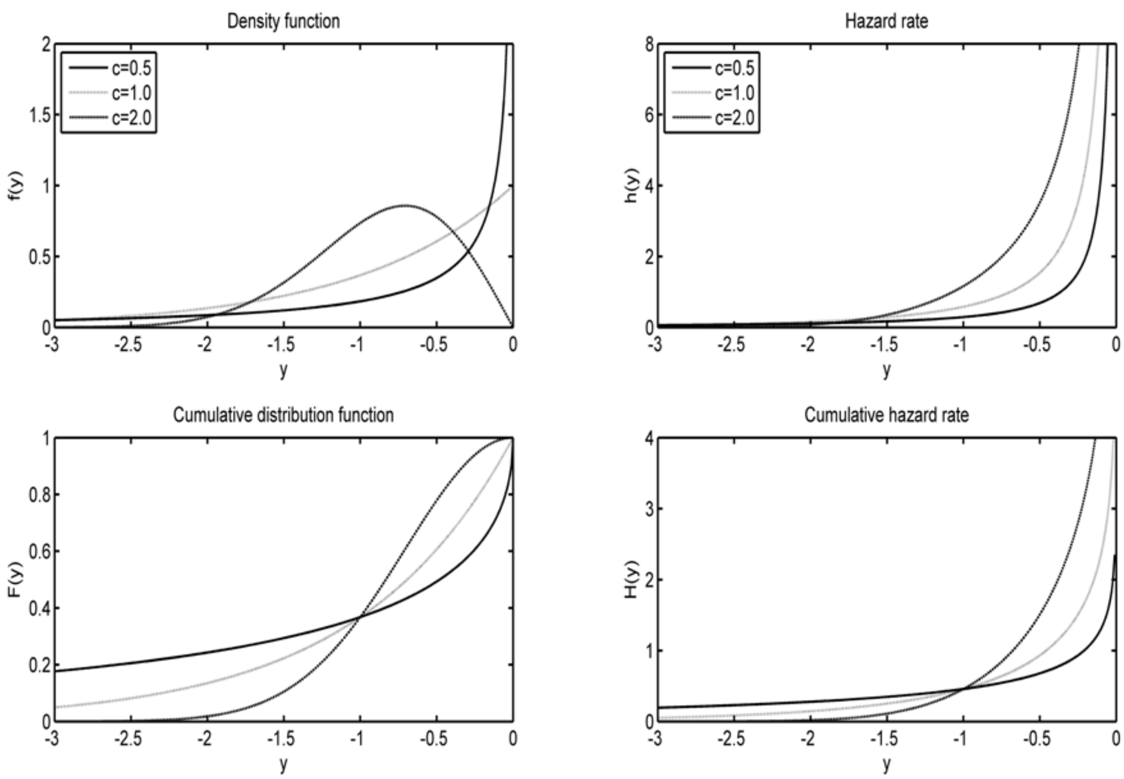
When the **WEIBULL distribution** with DF given by (1.45a), which is of minimum–type, is **reflected** about $x = a$ we arrive at the maximum type III distribution. This distribution is the limiting distribution of the sample maximum when the sampled distribution has a support bounded from above by $x_u$ and a CDF which behaves like $\beta\,(x_u - x)^\alpha$ for some $\alpha,\ \beta > 0$ as $x \to x_u^-$. A prototype of such a distribution is the uniform distribution over some interval $[x_\ell, x_u]$. A maximum type III distribution with $c = 1$ is a **reflected exponential distribution**.

$$f(x|a,b,c) = \frac{c}{b}\left(\frac{a-x}{b}\right)^{c-1} \exp\left[-\left(\frac{a-x}{b}\right)^c\right], \quad \left\{\begin{array}{l} x \le a,\ a \in \mathbb{R} \\ b > 0,\ c > 0 \end{array}\right\} \tag{5.57a}$$

$$F(x|a,b,c) = \exp\left[-\left(\frac{a-x}{b}\right)^c\right] \tag{5.57b}$$

$$R(x|a,b,c) = 1 - \exp\left[-\left(\frac{a-x}{b}\right)^c\right] \tag{5.57c}$$

Figure 5/62: Several functions for the reduced type III maximum extreme value distribution




$$h(x|a,b,c) = \frac{c\left(\dfrac{a-x}{b}\right)^c}{(a-x)\left\{\exp\left[\left(\dfrac{a-x}{b}\right)^c\right]-1\right\}} \tag{5.57d}$$

$$H(x|a,b,c) = -\ln\left\{1-\exp\left[-\left(\frac{a-x}{b}\right)^c\right]\right\} \tag{5.57e}$$

$$F_X^{-1}(P) = x_P = a - b\left(-\ln P\right)^{1/c},\ \ 0 < P \le 1 \tag{5.57f}$$

$$a = x_1 \tag{5.57g}$$

$$b \approx a - x_{0.3679} \tag{5.57h}$$

$$x_{0.5} \approx a - 0.6931^{1/c}\, b \tag{5.57i}$$

$$x_M = a - b\left(\frac{c-1}{c}\right)^{1/c},\ \ c \ge 1 \tag{5.57j}$$

In order to write the formulas for the moments of the maximum and the minimum type III extreme value distributions as concise as possible we introduce the abbreviation

$$_+\Gamma_r := \Gamma\left(1+\frac{r}{c}\right).$$

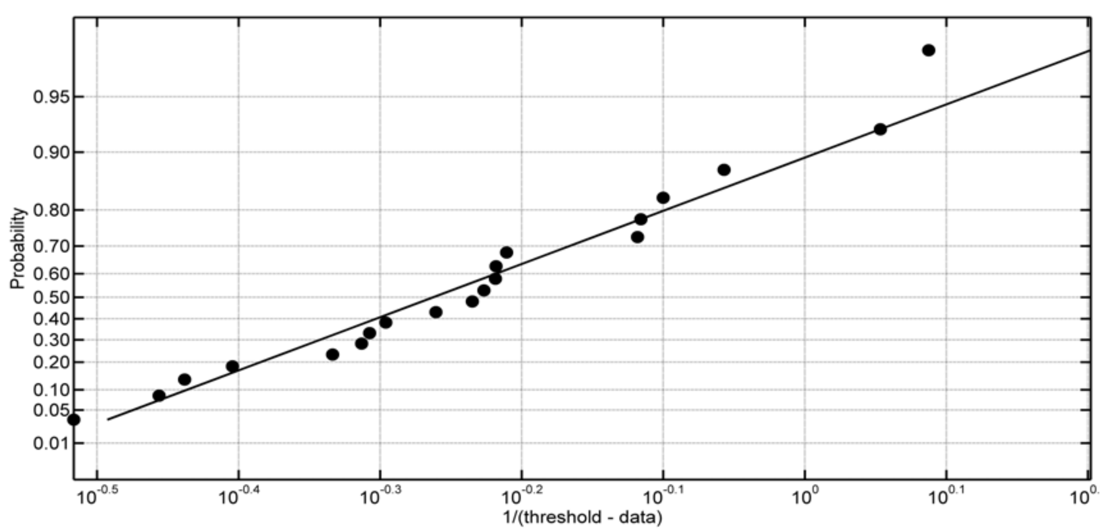
$$
\begin{aligned}
\mu'_r(Y) &= (-1)^r\,{}_+\Gamma_r;\;\; r = 0,1,2,\ldots;\;\; Y = (X-a)/b && (5.57k)\\
\mu'_1(X) &= \mathrm{E}(X) \;=\; a - b\,{}_+\Gamma_1 && (5.57l)\\
\mu_2(X) &= \mathrm{Var}(X) \;=\; b^2\big[{}_+\Gamma_2 -{}_+\Gamma_1^1\big] && (5.57m)\\
\alpha_3 &= -\frac{{}_+\Gamma_3 - 3\,{}_+\Gamma_2\,{}_+\Gamma_1 + 2\,{}_+\Gamma_1^3}{\big[{}_+\Gamma_2 -{}_+\Gamma_1^2\big]^{3/2}} && (5.57n)\\
\alpha_4 &= \frac{{}_+\Gamma_4 - 4\,{}_+\Gamma_3\,{}_+\Gamma_1 + 6\,{}_+\Gamma_2\,{}_+\Gamma_1^2 - 3\,{}_+\Gamma_1^4}{\big[{}_+\Gamma_2 -{}_+\Gamma_1^2\big]^2} && (5.57o)\\
F_Y^{-1}(P) &= y_P \;=\; -\big(-\ln P\big)^{1/c},\;\; 0 < P \le 1 && (5.57p)\\
f_Y(y_P) &= c\,P\,(-\ln P)\,\big(-1/\ln P\big)^{1/c} && (5.57q)
\end{aligned}
$$

Moments of the reduced order statistics can be traced back to those the type III minimum or WEIBULL distribution. Let $Y^*_{r:n},\; Y^*_{s:n}$ be the transformed WEIBULL order statistics as given in Sect. 5.3.2.4. Then the moments of the reduced maximum type III or reflected WEIBULL order statistics $Y_{r:n},\; Y_{s:n}$ follow as:

$$
\begin{aligned}
\mathrm{E}\big(Y_{r:n}\big) &= -\mathrm{E}\big(Y^*_{n-r+1:n}\big)\\
\mathrm{Var}\big(Y_{r:n}\big) &= \mathrm{Var}\big(Y^*_{n-r+1:n}\big)\\
\mathrm{Cov}\big(Y_{r:n}, Y_{s:n}\big) &= \mathrm{Cov}\big(Y^*_{n-s+1:n}, Y^*_{n-r+1:n}\big)
\end{aligned}
$$

These moments depend — among others — on the shape parameter $c$ and thus are not helpful in linear estimating. We remember that $\widetilde{X} = -\ln(a-X)$ has a type I maximum distribution with parameters $\widetilde{a} = -\ln b$ and $\widetilde{b} = 1/c$. Thus, we apply the procedure of Sect. 5.2.5.1 to linearly estimate $\widetilde{a}$ and $\widetilde{b}$. The estimates may be solved for estimates of $b$ and $c$.

Figure 5/63: Type III maximum extreme value probability paper with data and regression line



### 5.3.2.3 Type II minimum or FRÉCHET distribution — $X \sim EMN2(a, b, c)$

The FRÉCHET distribution is the limiting distribution of the sample minimum when the sampled distribution is unlimited from below and is of CAUCHY–type with a fat tail on the left–hand side, i.e. for some positive $k$ and $A$ we have

$$\lim_{x \to -\infty} (-x)^k \, F(x) = A.$$

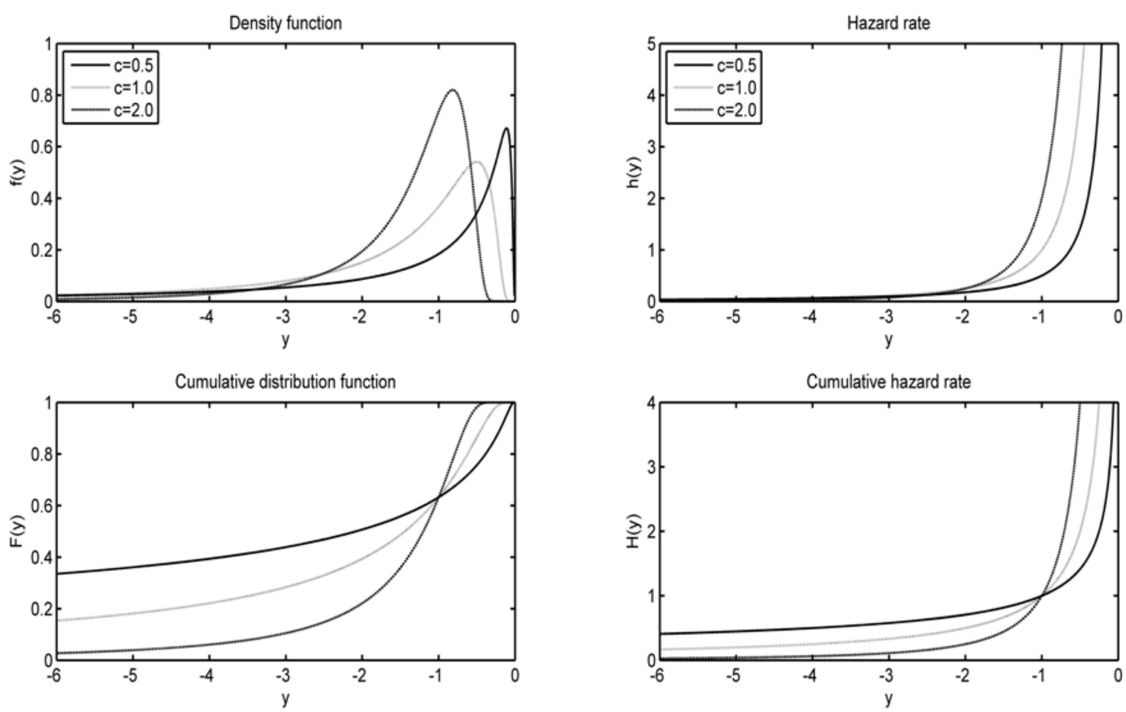
We also arrive at the minimum type II extreme value distribution by reflecting the maximum type II distribution about $x = a$. Thus, all the functions and characteristics easily follow from those of the maximum type II distribution, especially, moments will not always exist.

$$f(x|a,b,c) = \frac{c}{b}\left(\frac{a-x}{b}\right)^{-c-1} \exp\left[-\left(\frac{a-x}{b}\right)^{-c}\right], \quad \left\{ \begin{array}{c} x \le a,\ a \in \mathbb{R} \\ b > 0,\ c > 0 \end{array} \right\} \tag{5.58a}$$

$$F(x|a,b,c) = 1 - \exp\left[-\left(\frac{a-x}{b}\right)^{-c}\right] \tag{5.58b}$$

$$R(x|a,b,c) = \exp\left[-\left(\frac{a-x}{b}\right)^{-c}\right] \tag{5.58c}$$

Figure 5/64: Several functions for the reduced type II minimum extreme value distribution



$$h(x|a,b,c) = \frac{c}{b}\left(\frac{a-x}{b}\right)^{-c-1} \tag{5.58d}$$

$$H(x|a,b,c) = \left(\frac{a-x}{b}\right)^{-c} \tag{5.58e}$$

$$F_X^{-1}(P) = x_P = a - b\left[-\frac{1}{\ln(1-P)}\right]^{1/c}, \quad 0 < P \leq 1 \tag{5.58f}$$

$$a = x_1 \tag{5.58g}$$

$$b \approx a - x_{0.6321} \tag{5.58h}$$

$$x_{0.5} \approx a - 1.4427^{1/c}\, b \tag{5.58i}$$

$$x_M = a - b\left(\frac{c}{1+c}\right)^{1/c} \tag{5.58j}$$

$$\mu_r'(Y) = (-1)^r\, {}_{-}\Gamma_r, \quad c > r;\ Y = (X-a)\big/b \tag{5.58k}$$

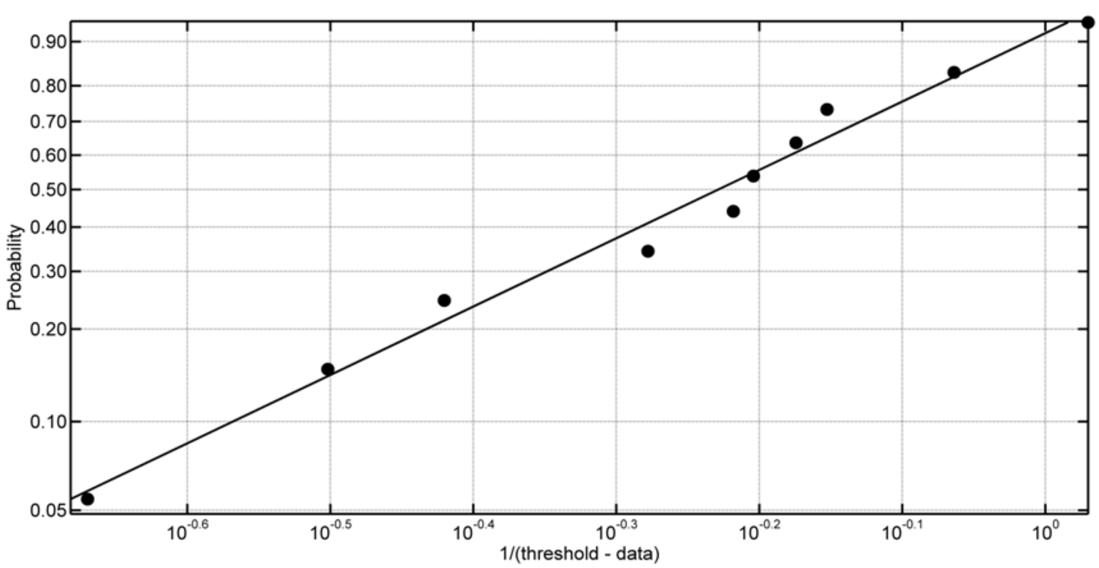
$$\mu_1'(X) = \mathrm{E}(X) = a - b\, {}_{-}\Gamma_1, \quad c > 1 \tag{5.58l}$$

$$\mu_2(X) = \mathrm{Var}(X) = b^2\big[{}_{-}\Gamma_2 - {}_{-}\Gamma_1^2\big], \quad c > 2 \tag{5.58m}$$

$$\alpha_3 = -\frac{{}_{-}\Gamma_3 - 3\,{}_{-}\Gamma_2\,{}_{-}\Gamma_1 + 2\,{}_{-}\Gamma_1^3}{\big[{}_{-}\Gamma_2 - {}_{-}\Gamma_1^2\big]^{3/2}}, \quad c > 3 \tag{5.58n}$$

$$\alpha_4 = \frac{{}_{-}\Gamma_4 - 4\,{}_{-}\Gamma_3\,{}_{-}\Gamma_1 + 6\,{}_{-}\Gamma_2\,{}_{-}\Gamma_1 - 3\,{}_{-}\Gamma_1^4}{\big[{}_{-}\Gamma_2 - {}_{-}\Gamma_1^2\big]^2}, \quad c > 4 \tag{5.58o}$$

Figure 5/65: Type II minimum extreme value probability paper with data and regression line



$$F_Y^{-1}(P) = y_P = -\left[-\frac{1}{\ln(1-P)}\right]^{1/c}, \quad 0 < P \le 1 \tag{5.58p}$$

$$f_Y(y_P) = c\,(1-P)\,\big[-\ln(1-P)\big]\,\big[(-\ln(1-P)\big]^{1/c} \tag{5.58q}$$

Introducing $\widetilde{X} = -\ln(X-a)$ we have a type I minimum extreme value distribution with parameters $\widetilde{a} = -\ln b$ and $\widetilde{b} = 1\big/c$ whose linear estimates can be found by the procedure of Sect. 5.2.5.2.

#### 5.3.2.4 Type III minimum or WEIBULL distribution — $X \sim EMN3(a, b, c)$

In recent years the WEIBULL distribution has become one of the most popular distributions in statistics. The latest and most comprehensive documentation of all its aspects — theory, applications, estimation and testing — is RINNE (2009). Other monographs on this distribution are MURTHY/XIE/JIANG (2004) and DODSON (1994).

Depending on the value of its shape parameter $c$ the WEIBULL density may look very different:

- For $0 < c < 1$ it is inversely J–shaped.
- For $c = 1$ we have the DF of an **exponential distribution**.
- For $c < 3.6$ it is positively skew.
- For $c \approx 3.6$ it is approximately symmetric.
- For $c > 3.5$ it is negatively skew.
- For $c \to 0$ the DF concentrates at $x = a$.
- For $c \to \infty$ the DF concentrates at $x = a + b$.

Also depending on $c$ the WEIBULL distribution shows different types of aging as measured by the hazard function (5.59d).

- For $0 < c < 1$ the hazard rate is decreasing (negative aging).
- For $c = 1$ the hazard rate is constant (no aging).
- For $c > 1$ the hazard rate is increasing (positive aging).

The WEIBULL distribution is related to the following distributions:

- For $c = 1$ it is an **exponential distribution**.
- For $c = 2$ it as a **RAYLEIGH distribution**.

- For $c \approx 3.6$ it is approximately a **normal distribution**.

Furthermore, it is a special case of

- the **polynomial hazard rate distribution**,
- the **$\chi$–distribution**,
- the **three–** and **four–parameter gamma distributions**.

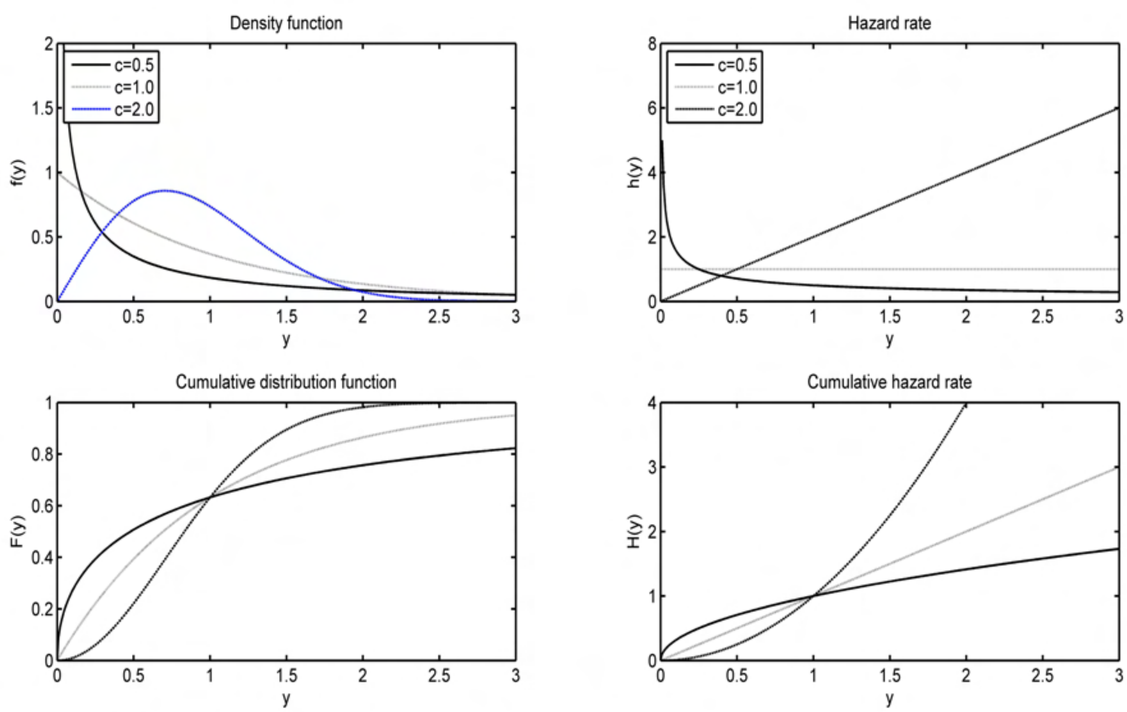
Its role within the family of extreme value distributions has been discussed in Sect. 1.3.

$$f(x|a,b,c) = \frac{c}{b}\left(\frac{a-x}{b}\right)^{c-1}\exp\left[-\left(\frac{x-a}{b}\right)^{c}\right], \left\{\begin{array}{l} x \geq a,\ a \in \mathbb{R} \\ b > 0,\ c > 0 \end{array}\right\} \tag{5.59a}$$

$$F(x|a,b,c) = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^{c}\right] \tag{5.59b}$$

$$R(x|a,b,c) = \exp\left[-\left(\frac{x-a}{b}\right)^{c}\right] \tag{5.59c}$$

Figure 5/66: Several functions for the reduced type III minimum extreme value or WEIBULL distribution



$$h(x|a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \tag{5.59d}$$

$$H(x|a,b,c) = \left(\frac{x-a}{b}\right)^{c} \tag{5.59e}$$

$$F_X^{-1}(P) = x_P = a + b\,[-\ln(1-P)]^{1/c}, \quad 0 \le P < 1 \tag{5.59f}$$

$$a = x_0 \tag{5.59g}$$

$$b \approx x_{0.6321} - a \;^{46} \tag{5.59h}$$

$$x_{0.5} \approx a + 0.6931^{1/c}\, b$$

$$x_M = a + b\left(\frac{c-1}{c}\right)^{1/c}, \; c \ge 1 \tag{5.59i}$$

$$\mu_r'(Y) = {}_{+}\Gamma_r, \;\; c > r, \; Y = (X-a)\big/b \tag{5.59j}$$

$$\mu_r(Y) = \sum_{j=0}^{r}\binom{r}{j}(-1)^j\,{}_{+}\Gamma_1^j\,{}_{+}\Gamma_{r-j} \tag{5.59k}$$

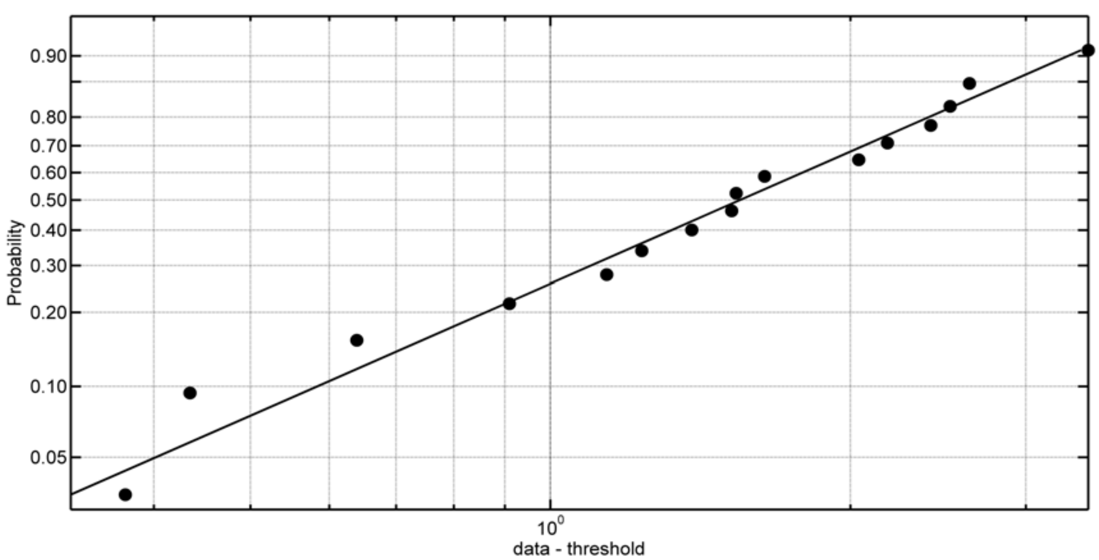
$$\mu_1'(X) = \mathrm{E}(X) = a - b\,{}_{+}\Gamma_1 \tag{5.59l}$$

$$\mu_2(X) = \mathrm{Var}(X) = b^2\left[{}_{+}\Gamma_2 - {}_{+}\Gamma_1^2\right] \tag{5.59m}$$

$$\alpha_3 = \frac{{}_{+}\Gamma_3 - 3\,{}_{+}\Gamma_2\,{}_{+}\Gamma_1 + 2\,{}_{+}\Gamma_1^3}{\left[{}_{+}\Gamma_2 - {}_{+}\Gamma_1^2\right]^{3/2}} \tag{5.59n}$$

$$\alpha_4 = \frac{{}_{+}\Gamma_4 - 4\,{}_{+}\Gamma_3\,{}_{+}\Gamma_1 + 6\,{}_{+}\Gamma_2\,{}_{+}\Gamma_1^2 - 3\,{}_{+}\Gamma_1^4}{\left[{}_{+}\Gamma_2 - {}_{+}\Gamma_1^2\right]^{2}} \tag{5.59o}$$

Figure 5/67: Type III minimum extreme value or WEIBULL probability paper with data and regression line




46 The percentile $x_{0.6321}$ is called **characteristic life.**

$$
\begin{aligned}
F_Y^{-1}(P) &= y_P = \big[-\ln(1-P)\big]^{1/c}, \;\; 0 \le P < 1 && (5.59p)\\
f_Y(y_P) &= c\,(1-P)\,\big[-\ln(1-P)\big]\left[-\frac{1}{\ln(1-P)}\right]^{1/c} && (5.59q)
\end{aligned}
$$

Due to LIEBLEIN (1955) there exist closed form expressions for the moments of reduced WEIBULL order statistics, but they do not help in finding linear estimate as they depend on $c$. Instead, we use the transformation $\widetilde{X} = \ln(X - a)$, which has a type I minimum extreme value distribution with parameters $\widetilde{a} = \ln b$ and $\widetilde{b} = 1/c$, and estimate with the procedure of Sect. 5.2.5.2.

The DF of the reduced WEIBULL order statistic $Y_{r:n},\; 1 \le r \le n$, is

$$
f_{r:n}(y) = r\binom{n}{r}\,c\,y^{c-1}\,\exp\big\{-y^{(n-r+1)\,c}\big\}\,\left[1-\exp(-y^c)\right]^{r-1}, \tag{5.60a}
$$

from where we obtain the $k$–th raw moment of $Y_{r:n}$ to be

$$
\begin{aligned}
\mathrm{E}\big(Y_{r:n}^k\big) &= r\binom{n}{r}\int\limits_0^\infty y^k\,c\,y^{c-1}\,\exp\big\{-y^{(n-r+1)\,c}\big\}\,\big[1-\exp\big(-y^c\big)\big]^{r-1}\,\mathrm{d}y\\
&= r\binom{n}{r}\sum_{i=0}^{r-1}(-1)^i\binom{r-1}{i}\int\limits_0^\infty y^k\,c\,y^{c-1}\,\exp\big\{-(n-r+i+1)\,y^c\big\}\,\mathrm{d}y\\
&= r\binom{n}{r}\Gamma\Big(1+\frac{k}{c}\Big)\sum_{i=0}^{r-1}\frac{(-1)^i\binom{r-1}{i}}{(n-r+i+1)^{1+(k/c)}}\,.
\end{aligned} \tag{5.60b}
$$

(5.60b) is due to LIEBLEIN (1955).

The joint DF of $Y_{r:n}$ and $Y_{s:n}$ $(1 \le r < s \le n)$ is

$$
f_{r,s:n}(u,v) = \left\{
\begin{array}{l}
\dfrac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}\,c^2\,(u\,v)^{c-1}\,\exp\big\{-u^c\big\}\times\\[2ex]
\exp\big\{-(n-s+1)\,v^c\big\}\big[1-\exp\big\{-u^c\big\}\big]^{r-1}\times\\[2ex]
\big[\exp\big\{-u^c\big\}-\exp\big\{-v^c\big\}\big]^{s-r-1}, \;\; 0 < u < v < \infty.
\end{array}
\right\} \tag{5.61a}
$$

From (5.61a) we obtain the product moment of $Y_{r:n}$ and $Y_{s:n}$ as

$$\begin{aligned}
\mathrm{E}\big(Y_{r:n}\,Y_{s:n}\big) &= \int_0^\infty\!\!\int_0^v u\,v\,f_{r,s:n}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{n!\,c^2}{(r-1)!\,(s-r-1)!\,(n-s)!}\sum_{i=0}^{r-1}\sum_{j=0}^{s-r-1}(-1)^{s-r-1-j+i}\times \\
&\quad \binom{r-1}{i}\binom{s-r-1}{j}\int_0^\infty\!\!\int_0^v \exp\big\{-(i+j+1)\,u^c\big\}\times \\
&\quad \exp\big\{-(n-r-j)\,v^c\big\}\,(u\,v)^c\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}\sum_{i=0}^{r-1}\sum_{j=0}^{s-r-1}(-1)^{s-r-1-j+i}\times \\
&\quad \binom{r-1}{i}\binom{s-r-1}{j}\phi_c(i+j+1,\ n-r-j). \qquad (5.61b)
\end{aligned}$$

This result is due to LIEBLEIN (1955) and $\phi_c(a,b)$ is LIEBLEIN's $\phi$–function defined by

$$\phi_c(a,b) = c^2\int_0^\infty\int_0^y \exp\big\{-a\,x^c - b\,y^c\big\}\,x^2\,y^2\,\mathrm{d}x\,\mathrm{d}y. \qquad (5.62a)$$

Through a differential equation approach, LIEBLEIN has derived an explicit algebraic formula for the $\phi$–function:

$$\phi_c(a,b) = \frac{\Gamma^2\left(1+\dfrac{1}{c}\right)}{(a\,b)^{1+(1/c)}}\,I_{a/(a+b)}\left(1+\frac{1}{c},\ 1+\frac{1}{c}\right)\ \text{ for } a \geq b, \qquad (5.62b)$$

where $I_p(c,d)$ is **PEARSON's incomplete beta function** defined as

$$I_p(c,d) = \frac{\Gamma(c+d)}{\Gamma(c)\,\Gamma(d)}\int_0^p t^{c-1}\,(1-t)^{d-1}\,\mathrm{d}t,\ \ 0 < p \leq 1.$$

When $a < b,\ \phi_c(a,b)$ may be computed from the identity

$$\phi_c(a,b) + \phi_c(b,a) = \frac{\Gamma^2\left(1+\dfrac{1}{c}\right)}{(a\,b)^{1+(1/c)}}\,. \qquad (5.62c)$$

(5.61b) together with $\mathrm{E}\big(Y_{r:n}\big)$ and $\mathrm{E}\big(Y_{s:n}\big)$ lead to the covariance $\mathrm{Cov}\big(Y_{r:n},Y_{s:n}\big) = \mathrm{E}\big(Y_{r:n}\,Y_{s:n}\big) - \mathrm{E}\big(Y_{r:n}\big)\,\mathrm{E}\big(Y_{s:n}\big)$.

There exist a number of tables giving the means, variances and covariances of order statistics from the reduced WEIBULL distribution:

- WEIBULL (1959) used LIEBLEIN's expression for $\mathrm{E}\big(Y_{r:n}\big)$ to tabulate the means, variances and covariances to five decimal places for $n = 1(1)15,\ r = 1(1)n$ and $c = 1/\alpha$ with $\alpha = 0.1(0.1)0.6(0.2)1.0$.
- WEIBULL (1967) presents means, variances and covariances of all order statistics for $n = 5(5)20$ and $c^{-1} = 0.1(0.1)1.0$ to five decimal places.
- GOVINDARAJULU/JOSHI (1968), based on LIEBLEIN's results, tabulated the means, variances and covariances to five decimal places for $n = 1(1)10$ and for $c = 1.0(0.5)3.0(1.0)10$.
- MCELHONE/LARSEN (1969) tabulated $\mathrm{E}\big(Y_{r:n}\big)$ to six significant figures for $n = 1(1)25$ and $c = 1(1)10$, $\mathrm{Var}\big(Y_{r:n}\big)$ and $\mathrm{Cov}\big(Y_{r:n}, Y_{s:n}\big)$ for $c = 1(1)5$ for the same set of $n$–values.
- HARTER (1970) tabulated $\mathrm{E}\big(Y_{r:n}\big)$ to five decimal places for $n = 1(1)40,\ r = 1(1)n$ and $c = 0.5(0.5)4.0(1.0)8.0$.
- BALAKRISHNAN/CHAN (1993) proposed tables for the means, variances and covariances of all order statistics for $n$ up to 20 and $c = 1/5,\ 1/4,\ 1/3,\ 1/2,\ 1.5(0.5)3,\ 4(2)10$.

### 5.3.3 Lognormal distributions

We have two types of lognormal distributions:[47]

- one with a lower threshold, which is the most popular version, and
- the other one with an upper threshold, which is reflection of the first one about $x = a$.

Whereas the normal distribution originates from an additive superimposition of a great number of variates (**central limit theorem**), the lognormal distribution[48] is the result of a superimposition of a great number of variates which are proportional to one another, i.e. the result is a product and not a sum as is the case with the normal distribution.

#### 5.3.3.1 Lognormal distribution with lower threshold — $X \sim LNL(a, \widetilde{a}, \widetilde{b})$

A variate $X$ is said to be lognormal distributed with lower threshold when there is a real number $a$ such that $\widetilde{X} = \ln(X - a)$ is normally distributed. There exist several ways or

[47] Suggested reading for this section: AITCHISON/BROWN (1957), CROW/SHIMIZU (1988) and JOHNSON/KOTZ/ BALAKRISHNAN (1994, Chapter 14).

[48] Sometimes this distribution is called by the names of the pioneers of its development: GALTON, KAPTEYN, GIBRAT or COBB–DOUGLAS. For logical reasons the distribution of $X$ should really be called **antilognormal** distribution, because it is not the distribution of a logarithm of a normal variate but of an exponential, i.e. antilogarithm, function of such a variate.

parameterizations to write the lognormal distribution. JOHNSON/KOTZ/BALAKRISHNAN (1994, p. 207) define the distribution of $X$ by[49]

$$Z = \eta + \delta\, \ln(X - a), \;\; a \in \mathbb{R},\; \delta > 0,\; \eta \in \mathbb{R}, \tag{5.63a}$$

where $Z$ is the standardized normal variate. Thus, the DF of $X$ reads

$$f(x|a,\eta,\delta) = \frac{\delta}{(x-a)\sqrt{2\,\pi}}\, \exp\left\{-\frac{1}{2}\big[\eta + \delta\, \ln(x-a)\big]^2\right\},\;\; x > a. \tag{5.63b}$$

An alternative notation replaces $\eta$ and $\delta$ by the mean $\widetilde{a}$ and the standard deviation $\widetilde{b}$ of $\widetilde{X} = \ln(X - a)$:

$$\widetilde{a} \;=\; \mathrm{E}\big[\ln(X-a)\big], \tag{5.64a}$$

$$\widetilde{b}^2 \;=\; \mathrm{Var}\big[\ln(X-a)\big]. \tag{5.64b}$$

The two sets of parameters (5.63a) and (5.64a,b) are linked as

$$\widetilde{a} \;=\; -\frac{\eta}{\delta} \tag{5.65a}$$

$$\widetilde{b} \;=\; \frac{1}{\delta}, \tag{5.65b}$$

so that (5.63a) turns into

$$Z = \frac{\ln(X-a) - \widetilde{a}}{\widetilde{b}}, \tag{5.66a}$$

and(5.63b) becomes

$$f\big(x|a,\widetilde{a},\widetilde{b}\,\big) = \frac{1}{(x-a)\,\widetilde{b}\,\sqrt{2\,\pi}}\, \exp\left\{-\frac{\big[\ln(x-a) - \widetilde{a}\big]^2}{2\,\widetilde{b}^2}\right\},\;\; x > a. \tag{5.66b}$$

$a$, the lower threshold of $X$, is the location parameter of $X$, $\widetilde{a}$ is its scale parameter and $\widetilde{b}$ is its shape parameter.[50]

BALAKRISHNAN/COHEN (1991, p. 278) suggest a third set of parameters:

$$\beta = \exp(\widetilde{a}) \tag{5.67a}$$

as the lognormal scale parameter and

$$\omega = \exp\left(\widetilde{b}^2\right) \tag{5.67b}$$

as the lognormal shape parameter.

---

49 We have chosen another set of letters for the parameters.

50 For $\widetilde{b} \to 0$ the DF $f(x|a,\widetilde{a},\widetilde{b})$ approaches the normal DF, and the greater $\widetilde{b}$ the higher the degree of positive skewness.

The DF of $X$ is positively skew, but $\widetilde{X} = \ln(X - a)$ has a normal distribution with DF

$$f\big(\widetilde{x}|\widetilde{a},\widetilde{b}\,\big) = \frac{1}{\widetilde{b}\,\sqrt{2\,\pi}}\,\exp\left\{-\frac{(\widetilde{x}-\widetilde{a})^2}{2\,\widetilde{b}^2}\right\}. \tag{5.68}$$

Thus, we can apply the procedure od Sect. 5.2.11 to $\widetilde{X}$ and directly find estimates of the lognormal parameters $\widetilde{a}$ and $\widetilde{b}$ without any need of re–transformation, provided $a$ is known or has been estimated.

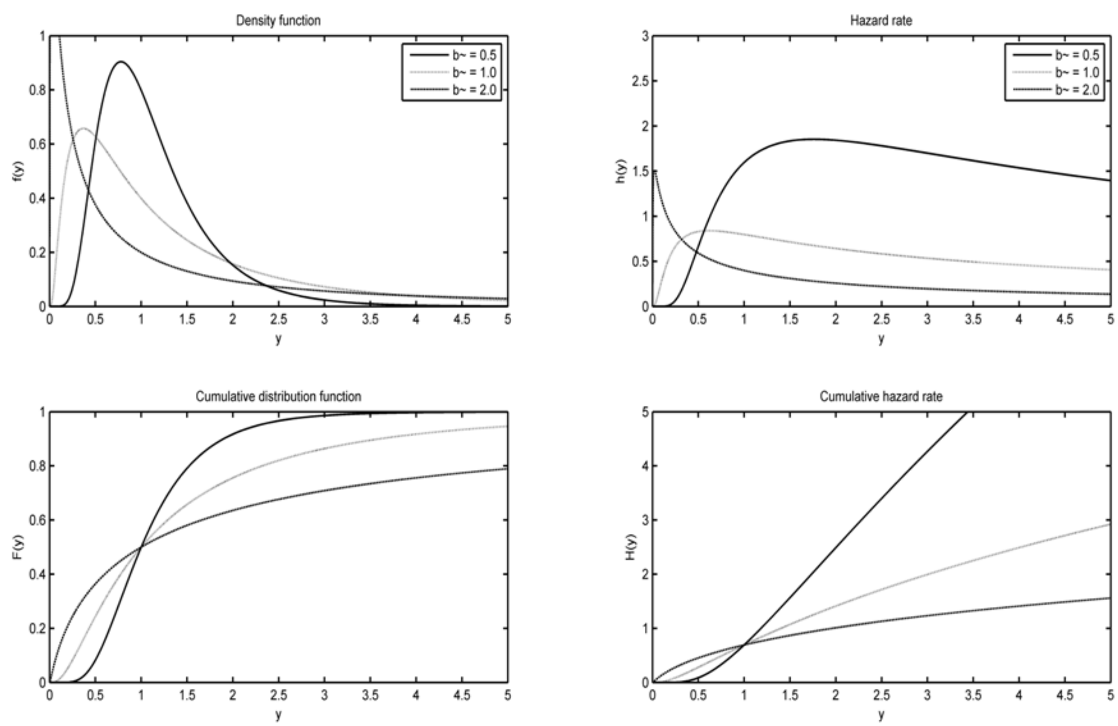
We give the following functions and characteristics of $X$:

$$f\big(x|\widetilde{a},\widetilde{b}\,\big) = \frac{1}{\widetilde{b}\,(x-a)\,\sqrt{2\,\pi}}\,\exp\left\{-\frac{\big[\ln(x-a)-\widetilde{a}\big]^2}{2\,\widetilde{b}^2}\right\},\ \left\{\begin{matrix} x > a,\ a \in \mathbb{R} \\ \widetilde{a} \in \mathbb{R},\ \widetilde{b} > 0\end{matrix}\right\} \tag{5.69a}$$

$$f\big(\widetilde{x}|\widetilde{a},\widetilde{b}\big) = \frac{1}{\widetilde{b}\,(x-a)}\,\phi\left[\frac{\ln(x-a)-\widetilde{a}}{\widetilde{b}}\right] \text{ [51]} \tag{5.69b}$$

$$F\big(x|a,\widetilde{a},\widetilde{b}\big) = \Phi\left[\frac{\ln(x-a)-\widetilde{a}}{\widetilde{b}}\right] \text{ [52]} \tag{5.69c}$$

Figure 5/68: Several functions for the reduced lognormal distribution with lower threshold




[51] $\phi(\cdot)$ is the DF of the standard normal distribution, see (5.31c).

[52] $\Phi(\cdot)$ is the CDF of the standard normal distribution, see (5.31g).

$$R\big(x|a,\widetilde{a},\widetilde{b}\,\big) = 1-\Phi\left[\frac{\ln(x-a)-\widetilde{a}}{\widetilde{b}}\right] = \Phi\left[-\frac{\ln(x-a)-\widetilde{a}}{\widetilde{b}}\right] \tag{5.69d}$$

$$h\big(x|a,\widetilde{a},\widetilde{b}\,\big) = \frac{\phi\left[\dfrac{\ln(x-a)-\widetilde{a}}{\widetilde{b}}\right]}{\widetilde{b}\,(x-a)\,\Phi\left[-\dfrac{\ln(x-a)-\widetilde{a}}{\widetilde{b}}\right]} \tag{5.69e}$$

$$H\big(x|a,\widetilde{a},\widetilde{b}\,\big) = -\ln\left\{\Phi\left[-\frac{\ln(x-a)-\widetilde{a}}{\widetilde{b}}\right]\right\} \tag{5.69f}$$

$$F_X^{-1}(P) = x_P = a+\exp\big(\widetilde{a}+z_P\,\widetilde{b}\,\big),\;\; 0<P<1 \;^{53} \tag{5.69g}$$

$$x_{0.5} = a+\exp\big(\widetilde{a}\,\big) \tag{5.69h}$$

$$\widetilde{a} \approx \frac{1}{2}\big[\ln(x_{0.8413}-a)+\ln(x_{0.1587}-a)\big] \tag{5.69i}$$

$$\widetilde{b} \approx \frac{1}{2}\big[\ln(x_{0.8413}-a)-\ln(x_{0.1587}-a)\big] \tag{5.69j}$$

$$x_M = a+\exp\big(\widetilde{a}-\widetilde{b}^2\,\big) \tag{5.69k}$$

$$\mathrm{E}\big[(X-a)^r\big] = \mathrm{E}\big\{\exp\big[r\,(\widetilde{a}+\widetilde{b}\,Z)\big]\big\} = \exp\left(r\,\widetilde{a}+\frac{r^2}{2}\,\widetilde{b}^2\right) \tag{5.69l}$$

$$\mu_1'(X) = \mathrm{E}(X) = a+\exp\left(\widetilde{a}+\frac{\widetilde{b}^2}{2}\right) = a+\beta\,\omega^{1/2} \tag{5.69m}$$

$$\mu_2(X) = \mathrm{Var}(X) = \exp\big(2\,\widetilde{a}+\widetilde{b}^2\,\big)\big[\exp\big(\widetilde{b}^2\,\big)-1\big] = \beta^2\,\omega\,(\omega-1) \tag{5.69n}$$

$$\begin{aligned}\mu_r(X) &= \mathrm{E}\big\{\big[X-\mathrm{E}(X)\big]^2\big\}\\ &= \exp\big(r\,\widetilde{a}\,\big)\,\omega^{r/2}\sum_{j=0}^{r}(-1)^j\binom{r}{j}\omega^{(r-j)\,(r-j-1)/2}\end{aligned} \tag{5.69o}$$

$$\mu_3(X) = \beta^3\,\omega^{3/2}\,(\omega-1)^2\,(\omega+2) \tag{5.69p}$$

$$\mu_4(X) = \beta^4\,\omega^2\,(\omega-1)^2\,(\omega^4+2\,\omega^3+3\,\omega^2-3) \tag{5.69q}$$

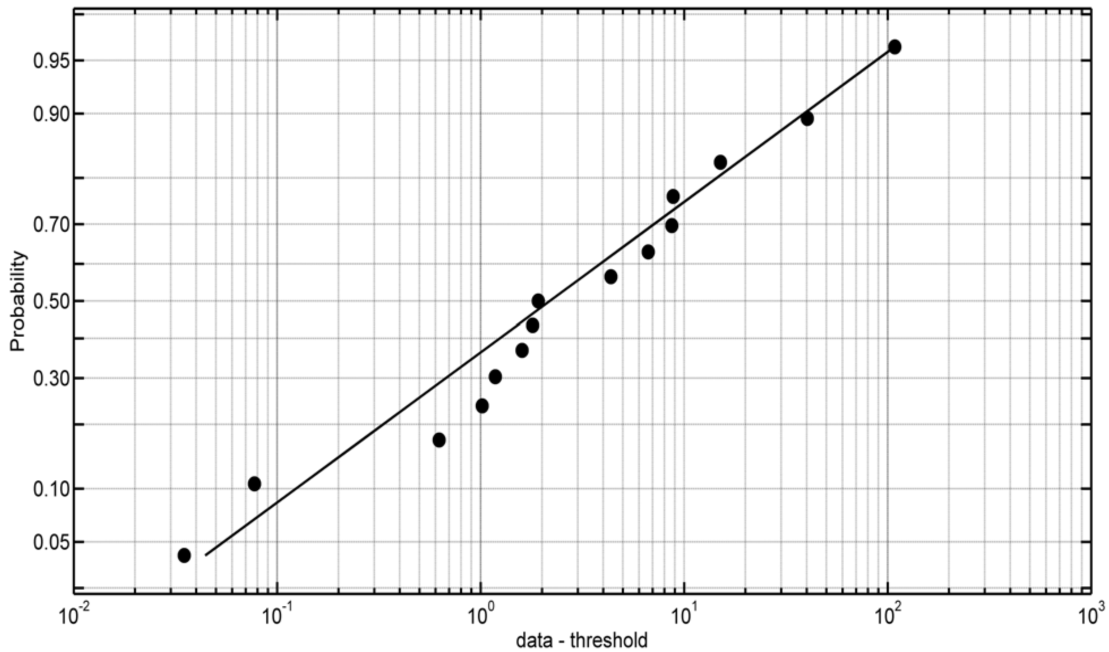
$$\alpha_3 = (\omega-1)^{1/2}\,(\omega+2) \tag{5.69r}$$

$$\alpha_4 = \omega^4+2\,\omega^3+3\,\omega^2-3 \tag{5.69s}$$

The first four moments of $X_{r:n}$ and all product moments $\mathrm{E}(X_{r:n}\,X_{s:n}),\ r\le s$, for the reduced lognormal distribution, i.e. with $a=\widetilde{a}=0$ and $\widetilde{b}=1$, have been tabulated by GUPTA/MCDONALD/GALARNEAU (1974) for sample sizes up to 20.

[53] $z_P$ is the percentile of order $P$ of the standard normal distribution, see (5.31o).

Figure 5/69: Lognormal (lower threshold) probability paper with data and regression line




#### 5.3.3.2 Lognormal distribution with upper threshold — $X \sim LNU(a, \widetilde{a}, \widetilde{b})$

A variate $X$ is said to be lognormal distributed with upper threshold when there is a number $a,\ a \in \mathbb{R}$, such that $\widetilde{X} = \ln(a - X)$ is normally distributed. The functions and characteristics of this type of lognormal distribution easily follow from those of the lower-threshold lognormal distribution because both distributions are related by a reflection about $x = a$. The parameters $\widetilde{a}$ and $\widetilde{b}$ of the upper–threshold lognormal distribution are

$$\widetilde{a} = \mathrm{E}\big[\ln(a - X)\big], \tag{5.70a}$$

$$\widetilde{b} = \mathrm{Var}\big[\ln(a - X)\big]. \tag{5.70b}$$

We further have:

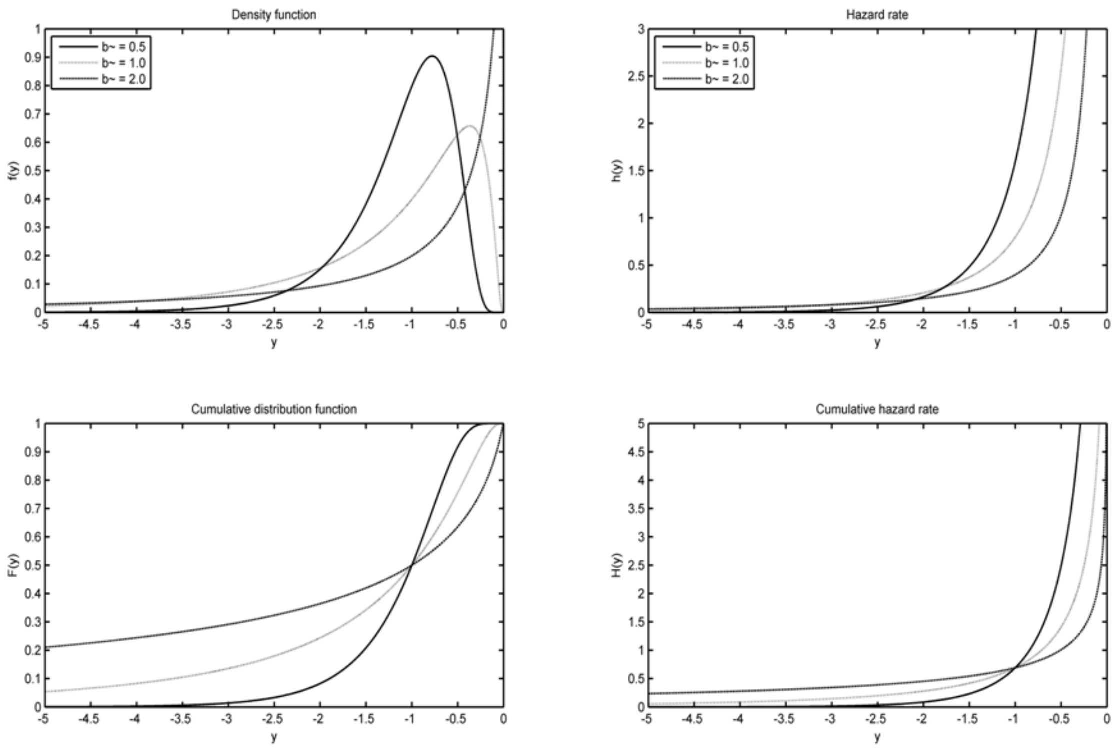
$$f\big(x|\widetilde{a},\widetilde{b}\,\big) = \frac{1}{\widetilde{b}\,(a-x)\,\sqrt{2\,\pi}}\,\exp\left\{-\frac{\big[\ln(a-x)-\widetilde{a}\big]^2}{2\,\widetilde{b}^2}\right\},\ \left\{\begin{matrix} x < a,\ a \in \mathbb{R} \\ \widetilde{a} \in \mathbb{R},\ \widetilde{b} > 0 \end{matrix}\right\} \tag{5.71a}$$

$$f\big(\widetilde{x}|\widetilde{a},\widetilde{b}\big) = \frac{1}{\widetilde{b}\,(a-x)}\,\phi\left[\frac{\ln(a-x)-\widetilde{a}}{\widetilde{b}}\right] \tag{5.71b}$$

$$F\big(x|a,\widetilde{a},\widetilde{b}\big) = \Phi\left[\frac{\ln(a-x)-\widetilde{a}}{\widetilde{b}}\right] \tag{5.71c}$$

$$R\big(x|a,\widetilde{a},\widetilde{b}\big) = 1 - \Phi\left[\frac{\ln(a-x)-\widetilde{a}}{\widetilde{b}}\right] = \Phi\left[-\frac{\ln(a-x)-\widetilde{a}}{\widetilde{b}}\right] \tag{5.71d}$$

Figure 5/70: Several functions for the reduced lognormal distribution with upper threshold




$$h\big(x|a,\widetilde{a},\widetilde{b}\big) = \frac{\phi\left[\dfrac{\ln(a-x)-\widetilde{a}}{\widetilde{b}}\right]}{\widetilde{b}\,(a-x)\,\Phi\left[-\dfrac{\ln(a-x)-\widetilde{a}}{\widetilde{b}}\right]} \tag{5.71e}$$

$$H\big(x|a,\widetilde{a},\widetilde{b}\big) = -\ln\left\{\Phi\left[-\frac{\ln(a-x)-\widetilde{a}}{\widetilde{b}}\right]\right\} \tag{5.71f}$$

$$F_X^{-1}(P) = x_P = a-\exp\big(\widetilde{a}+z_P\,\widetilde{b}\big),\ \ 0<P<1 \tag{5.71g}$$

$$x_{0.5} = a-\exp\big(\widetilde{a}\big) \tag{5.71h}$$

$$\widetilde{a} \approx \frac{1}{2}\big[\ln(x_{0.8413}-a)+\ln(x_{0.1587}-a)\big] \tag{5.71i}$$

$$\widetilde{b} \approx \frac{1}{2}\big[\ln(x_{0.8413}-a)-\ln(x_{0.1587}-a)\big] \tag{5.71j}$$

$$x_M = a-\exp\big(\widetilde{a}-\widetilde{b}^2\big) \tag{5.71k}$$

$$\mu_1'(X) = \mathrm{E}(X) = a-\exp\left(\widetilde{a}+\frac{\widetilde{b}^2}{2}\right) = a-\beta\,\omega^{1/2} \tag{5.71l}$$

$$\mu_2(X) = \mathrm{Var}(X)=\exp\big(2\,\widetilde{a}+\widetilde{b}^2\big)\big[\exp\big(\widetilde{b}^2\big)-1\big]=\beta^2\,\omega\,(\omega-1) \tag{5.71m}$$

$$\mu_3(X) = -\beta^3\,\omega^{3/2}\,(\omega-1)^2\,(\omega+2) \tag{5.71n}$$

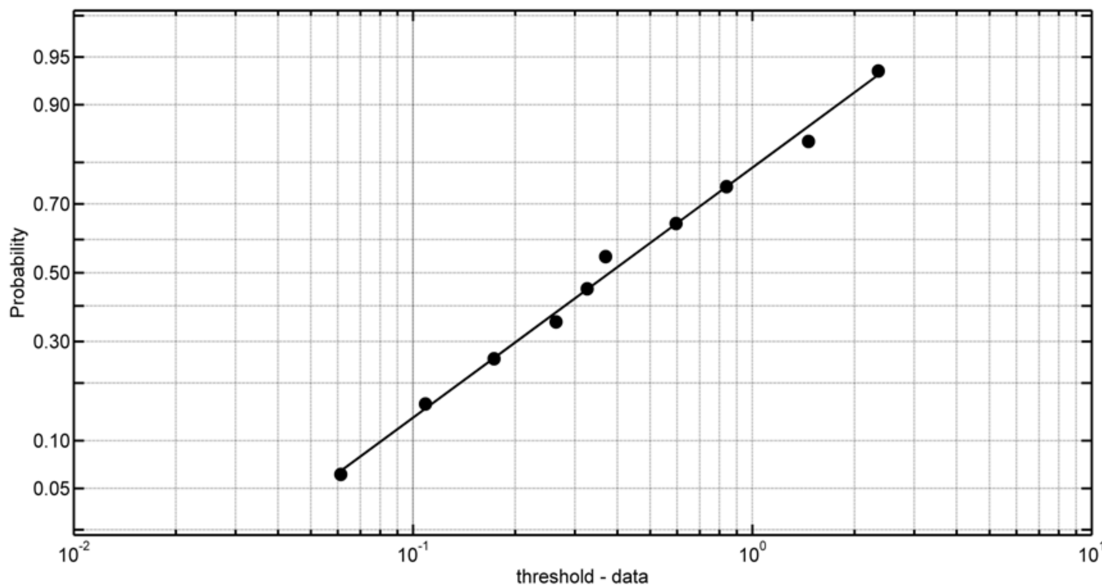
$$\mu_4(X) = \beta^4\,\omega^2\,(\omega-1)^2\,(\omega^4+2\,\omega^3+3\,\omega^2-3) \tag{5.71o}$$

$$\alpha_3 = -(\omega-1)^{1/2}\,(\omega+2) \tag{5.71p}$$

$$\alpha_4 = \omega^4+2\,\omega^3+3\,\omega^2-3 \tag{5.71q}$$

The estimates of $\widetilde{a}$ and $\widetilde{b}$ follow from the procedure for the normal distribution of Sect. 5.2.11 with $\widetilde{x} = \ln(a - x)$ and $a$ known or estimated.

Figure 5/71: Lognormal (upper threshold) probability paper with data and regression line




### 5.3.4 PARETO distribution — $X \sim PA(a, b, c)$

This distribution[54] has been proposed by the Swiss professor of economics VILFREDO PARETO (1848 – 1923) as a model for the distribution of income in an economy and is known as the **PARETO distribution of the first kind**.[55] We will call it PARETO distribution for short and note its following related distributions:[56]

- The **generalized PARETO distribution** with CDF

$$F(x|\alpha, \beta, c, d) = 1 - \left(\frac{\beta + \alpha}{x + \alpha}\right)^c \exp\big[-d\,(x - \beta)\big],\ \ x \geq \beta,\ d \in \mathbb{R},$$

  results in the PARETO distribution $PA(0, 1, c)$ for $\alpha = d = 0$ and $\beta = 1$.

---

[54] Suggested reading for this section: ARNOLD (1983), JOHNSON/KOTZ/BALAKRISHNAN (1994, Chapter 14).

[55] The **PARETO distribution of the second kind**, also known as **LOMAX distribution**, has CDF

$$F(x|\alpha, K) = 1 - \frac{K^\alpha}{(x + K)^\alpha},\ \ x \geq 0.$$

The **PARETO distribution of the third kind** has CDF

$$F(x|\alpha, \beta, K) = 1 - \frac{K\ \exp(-\beta\, x)}{(x + K)^\alpha},\ \ x > 0.$$

[56] The parameter $c$ is known as **PARETO's constant**.

- $X \sim PA(a,b,c) \;\Rightarrow\; \widetilde{X} = \ln(X-a) \sim EX(\ln b, 1/c)$ This relation will be used for linear estimation of $\widetilde{a} = \ln b$ and $\widetilde{b} = 1/c$.
- $X \sim PA(0,1,c) \;\Rightarrow\; X^{-1} \sim PO(0,1,c)$.
- $X \sim PA(0,b,c) \;\Rightarrow\; V = -\ln\left[\left(\dfrac{X}{b}\right)^c - 1\right] \sim LO(0,1)$.
- $X_i \overset{\text{iid}}{\sim} PA(0,b,c);\ i=1,2,\ldots,n;\ \Rightarrow\ X = 2\,b \sum\limits_{i=1}^{n} \ln\left(\dfrac{X_i}{c}\right) \sim \chi^2(2\,n)$
- $Y \sim EMX1(0,1) \;\Rightarrow\; X = b\left\{1 - \exp\left[-\exp(-X)\right]\right\}^{1/c} \sim PA(0,b,c)$

We note the following functions and characteristics of the PARETO distribution:

$$f(x|a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{-c-1} = \frac{c}{b}\left(\frac{b}{x-a}\right)^{c+1},\ \left\{\begin{array}{l} x \geq a+b,\ a \in \mathbb{R}, \\ b > 0,\ c > 0 \end{array}\right\} \tag{5.72a}$$

$$F(x|a,b,c) = 1 - \left(\frac{x-a}{b}\right)^{-c} = 1 - \left(\frac{b}{x-a}\right)^{c} \tag{5.72b}$$

$$R(x|a,b,c) = \left(\frac{x-a}{b}\right)^{-c} = \left(\frac{b}{x-a}\right)^{c} \tag{5.72c}$$

$$h(x|a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{-1} = \frac{c}{b}\left(\frac{b}{x-a}\right) \tag{5.72d}$$

$$H(x|a,b,c) = c\,\ln\left(\frac{x-a}{b}\right) \tag{5.72e}$$

$$F_X^{-1}(P) = x_P = a + b\,(1-P)^{-1/c},\ \ 0 \leq P < 1 \tag{5.72f}$$

$$x_0 = a + b \tag{5.72g}$$

$$x_{0.5} = a + 2^{1/c}\,b \tag{5.72h}$$

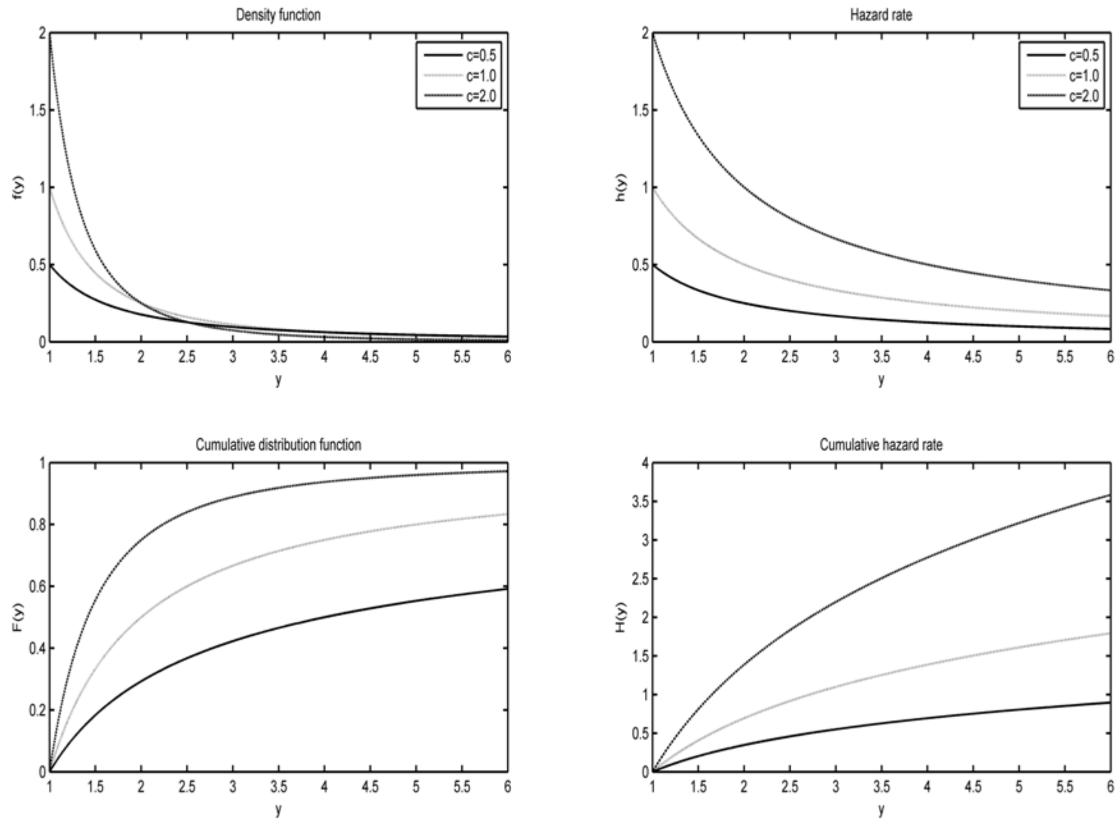
$$x_M = a + b \tag{5.72i}$$

$$\mu'_r(X) = \left\{\begin{array}{ll} \displaystyle\sum_{i=0}^{r} \binom{r}{i} \frac{c}{c-(r-i)}\, b^{r-i}\, a^i & \text{for}\quad c > r \\ \qquad\quad \text{not defined} & \text{for}\quad c \leq r \end{array}\right\} \tag{5.72j}$$

$$\mu'_1(X) = \mathrm{E}(X) = a + b\,\frac{c}{c-1},\ \ c > 1 \tag{5.72k}$$

$$\mu_2(X) = \mathrm{Var}(X) = b^2\,\frac{c}{(c-1)^2\,(c-2)},\ \ c > 2 \tag{5.72l}$$

Figure 5/72: Several functions for the reduced PARETO distribution




$$\alpha_3 = 2\,\frac{c+1}{c-3}\sqrt{\frac{c-2}{c}},\ \ c>3 \tag{5.72m}$$

$$\alpha_4 = \frac{3\,(c-2)\,(3\,c^2+c+2}{c\,(c-3)\,(c-4)},\ \ c>4 \tag{5.72n}$$

We notice that

$$\lim_{c\to\infty}\alpha_3 = 2 \ \text{ and } \ \lim_{c\to\infty}\alpha_4 = 9.$$

Moments of reduced PARETO order statistics $Y_{r:n}$ can be given in closed form, see ARNOLD/BALAKRISHNAN/NAGARAJA (1992, p. 36).

$$\alpha_{r:n}^{(k)} = \mathrm{E}\big(Y_{r:n}^k\big) = \frac{\Gamma(n+1)\,\Gamma\left(n+1-r-\dfrac{k}{c}\right)}{\Gamma(n+1-r)\,\Gamma\left(n+1-\dfrac{k}{c}\right)} \ \text{ for } \ c>\frac{k}{n+1-r} \tag{5.73a}$$

Especially we have

$$\alpha_{1:n}^{(k)} = \frac{n\,c}{n\,c-k} \ \text{ for } \ c>\frac{k}{n}. \tag{5.73b}$$

The following recurrence relation holds for $2 \le r \le n$ and $k = 1, 2, \ldots$:

$$\alpha_{r:n}^{(k)} = \frac{n\,c}{n\,c - k}\,\alpha_{r-1:n-1}^{(k)}, \quad c > \frac{k}{n+1-r}. \tag{5.73c}$$

Product moments for $1 \le r < s \le n$ and $k = 1, 2, \ldots$ are given by

$$\alpha_{r,s:n}^{(k_r,k_s)} = \mathrm{E}\big(Y_{r:n}^{k_r}\,Y_{s:n}^{k_s}\big) = \frac{\Gamma(n+1)}{\Gamma(n+1-s)}\,\frac{\Gamma\left(n+1-s-\dfrac{k_s}{c}\right)}{\Gamma\left(n+1-r-\dfrac{k_s}{c}\right)}\,\frac{\Gamma\left(n+1-r-\dfrac{k_r+k_s}{c}\right)}{\Gamma\left(n+1-\dfrac{k_r+k_s}{c}\right)}$$

$$\text{for} \quad c > \max\left[\frac{k_s}{n+1-s}, \frac{k_r+k_s}{n+1-r}\right], \tag{5.74a}$$

and especially for $1 \le r \le n-1$ by

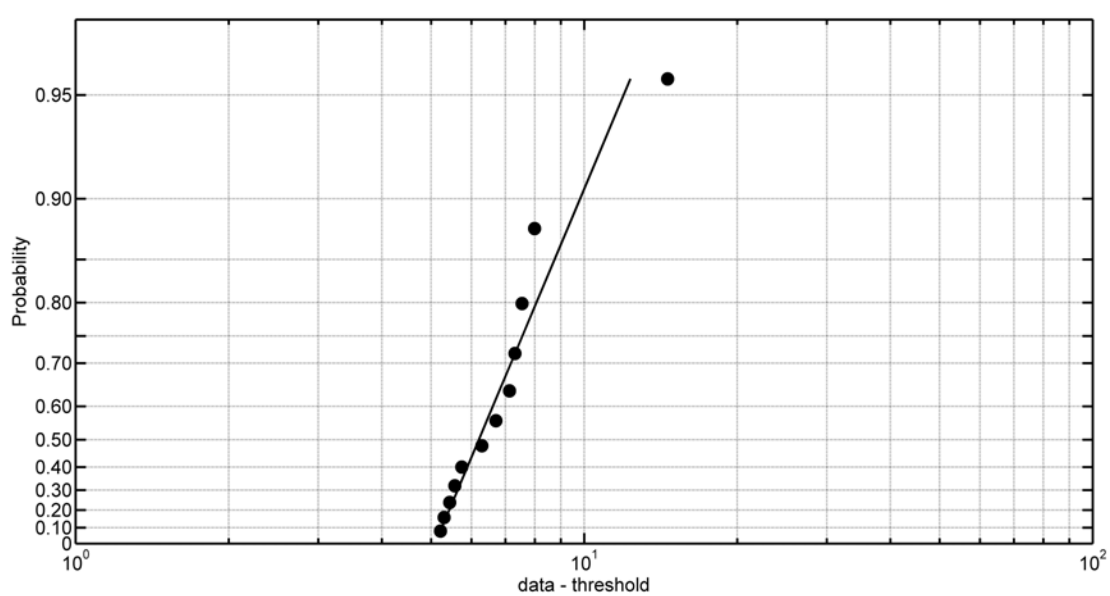
$$\alpha_{r,r+1:n} = \frac{c\,(n-r)}{(n-r)\,c-1}\,\alpha_{r:n}^{(2)} \quad \text{for} \quad c > \frac{2}{n+1-r}. \tag{5.74b}$$

and the recurrence relation for $1 \le r < s \le n$ and $s - r > 2$:

$$\alpha_{r,s:n} = \frac{c\,(n+1-s)}{c\,(n+1-s)-1}\,\alpha_{r,s-1:n}. \tag{5.74c}$$

MALIK (1966) has tabulated the means and covariances for $n \le 8$ and $c = 2.5(0.5)5.0$. They have been further discussed by HUANG (1975) and KABE(1972). We find estimates for $b$ and $c$ when $a$ is known or estimated using $\widetilde{X} = \ln(X - a)$ which is exponentially distributed with $\widetilde{a} = \ln b$ and $\widetilde{b} = 1/c$.

Figure 5/73: PARETO probability paper with data and regression line



### 5.3.5 Power–function distribution — $X \sim PO(a, b, c)$

Starting with the **beta distribution**

$$f(x|a,b,c,d) = \frac{\left(\dfrac{x-a}{b}\right)^{c-1}\left(1-\dfrac{x-a}{b}\right)^{d-1}}{b\,B(c,d)}$$

we arrive at the power–function distribution by setting $d = 1$. There are relationships of the power–function distribution with several other distributions:

- $X \sim PO(a,b,c) \;\Rightarrow\; V = \ln(X-a) \sim RE(\ln b, 1/c)$.
  This relation is used for linear estimating $\widetilde{a} = \ln b$ and $\widetilde{b} = 1/c$.
- $X \sim PO(0,1,c) \;\Rightarrow\; Y = -\ln[-c\,\ln X] \sim EMX1(0,1)$ .
- $X \sim PO(0,1,c) \;\Rightarrow\; Y = -\ln\big(X^{-c}-1\big) \sim LO(0,1)$.
- $X \sim PO(0,1,c) \;\Rightarrow\; X^{-1} \sim PA(0,1,c)$.
- $X \sim PO(a,b,1) \;\Rightarrow\; X \sim UN(a,b)$.
- $X \sim PO(0,1,c) \;\Rightarrow\; V = \big[\ln\big(X^c\big)\big]^{1/d} \sim EMN3(0,1,d)$.
- $X_1,\, X_2 \overset{\text{iid}}{\sim} PO(0,1,c) \;\Rightarrow\; V = X_1/X_2 \sim LA(0,1)$.
- $X_i \overset{\text{iid}}{\sim} PO(0,1,c);\; i = 1,2,\ldots,n;\; \Rightarrow\; V = \sum\limits_{i=1}^{n} \ln X_i \sim \text{Gamma}(c,n)$.

We note the following functions and characteristics:

$$f(x|a,b,c) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1}, \quad \left\{\begin{array}{l} a \le x \le a+b,\; a \in \mathbb{R} \\ b > 0,\; c > 0 \end{array}\right\} \tag{5.75a}$$

$$F(x|a,b,c) = \left(\frac{x-a}{b}\right)^{c} \tag{5.75b}$$

$$R(x|a,b,c) = 1-\left(\frac{x-a}{b}\right)^{c} \tag{5.75c}$$

$$h(x|a,b,c) = \frac{c\left[1+\dfrac{1}{\left(\dfrac{x-a}{b}\right)^{c}-1}\right]}{a-x} \tag{5.75d}$$

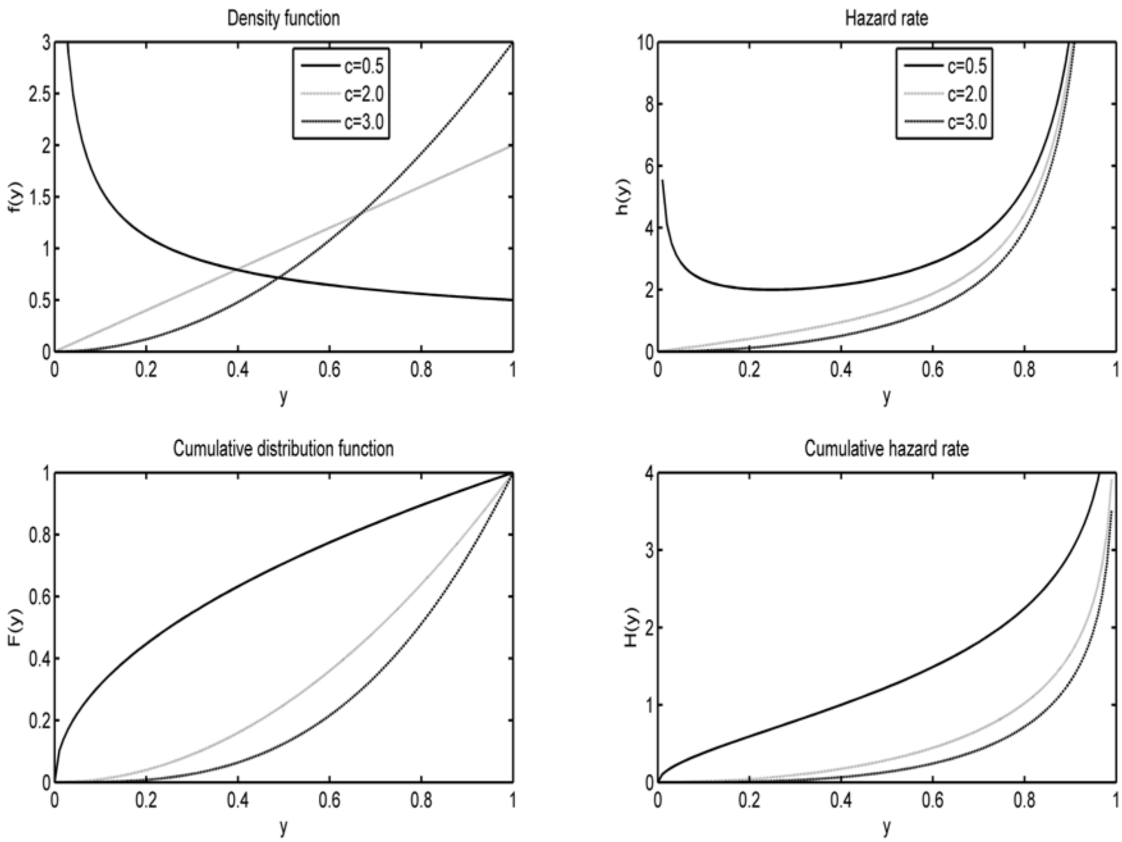
$$H(x|a,b,c) = -\ln\left[1-\left(\frac{x-a}{b}\right)^{c}\right] \tag{5.75e}$$

$$F_X^{-1}(P) = x_P = a + b\,P^{1/c}, \quad 0 \le P \le 1 \tag{5.75f}$$

$$a = x_0 \tag{5.75g}$$

$$b = x_1 - x_0 \tag{5.75h}$$

Figure 5/74: Several functions for the reduced power–function distribution




$$x_{0.5} = a + 0.5^{1/c}\,b \tag{5.75i}$$

$$x_M = \left\{ \begin{array}{ccl} a & \text{for} & 0 < c < 1 \\ \text{not defined} & \text{for} & c = 1 \\ a + b & \text{for} & c > 1 \end{array} \right\} \tag{5.75j}$$

$$\mu'_r(Y) = \frac{c}{c + r},\;\; Y = (X - a)\big/b \tag{5.75k}$$

$$\mu'_r(X) = \sum_{i=0}^{r} \frac{c}{c + r - i}\, b^{r-i}\, a^i \tag{5.75l}$$

$$\mu'_1(X) = \mathrm{E}(X) \;=\; a + b\,\frac{c}{c+1} \tag{5.75m}$$

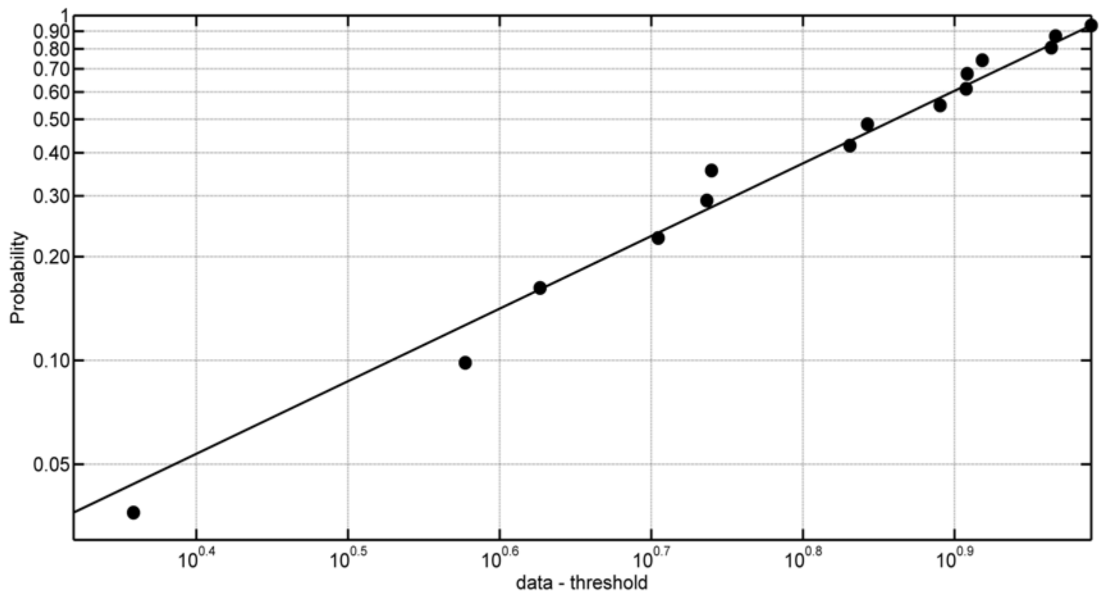
$$\mu_2(X) = \mathrm{Var}(X) \;=\; b^2\,\frac{c}{(c+1)^2\,(c+2)} \tag{5.75n}$$

$$\alpha_3 = \frac{2\,(1-c)}{c-3}\sqrt{\frac{c+2}{c}} \tag{5.75o}$$

$$\alpha_4 = \frac{3\,(c+2)\,(3\,c^2 - c + 2)}{c\,(c+3)\,(c+4)} \tag{5.75p}$$

Moments of the reduce power–function order statistics $Y_{r:n}$ can be given in closed form, see (2.15a–e). BALAKRISHNAN/RAO (1998a, p. 191) give a set of recurrence relations to compute all the product moments for all sample sizes.

Figure 5/75: Power–function probability paper with data and regression line



# 6 The program LEPP

Based on MATLAB we have designed the menu–driven program LEPP. It executes linear estimation by different approaches and plots the data together with the fitted regression line on probability paper for those 35 distributions, which have been presented in Chapter 5. The reader will find this program together with all the supporting function–subprograms in the folder ”LEPP-Program”. The folder ”LEPP-Data” contains Monte Carlo simulated data sets, three sets for each distribution, for testing the program by the reader.

In Sect. 6.1 we say how LEPP is organized and what is going on in its three parts. Then, in Sect. 6.2 we describe the input-format for the sample–data and of the moments–matrix of reduced order statistics. The moments–matrix is not compulsory, but may be provided by those users who have access to tabulated moments. Finally, in Sect. 6.3 we give hints to how to use LEPP and to what will be the output of LEPP. The most important hints for installing and running LEPP are also given in the read–me-file `program.txt` in the folder ”LEPP-Read-Me”. The second read–me-file named `data.txt` in that folder describes the data sets with the following items: sample size, censoring mode and values of the parameters used in Monte Carlo simulation.

## 6.1 Structure of LEPP

The main program, which is called by typing `LEPP`, comes as a script M–file and has about 1,000 lines of commands. These are organized in three parts:

1. checking the input and preparing the sample–data,
2. computing the regressor and the variance–covariance matrix, preparing the ordinate axis of the probability paper and displaying a chart with several functions for the chosen distribution,
3. processing the sample–data of part 1, the regressor and the variance–covariance matrix of part 2, displaying the numerical results (estimates together with their variance–covariance matrix, correlation of data and regressor) and information about the method and data used and last but not least the probability paper with the plot of the sample–data and the fitted regression line.

The purpose of the **first part** is to check whether the matrix `data` with the sample–data and whether — in case tabulated moments of reduced order statistics should be used — the matrix `moments` have been stored in the directory containing LEPP and its supporting function–subprograms and whether these matrices are such that the program will run properly. The format of these matrices and their contents are described in Sect. 6.2. When there is something out of order with the matrices LEPP prints an error–message on the screen and stops running. With respect to **data–type 1** (ordered observations with corresponding ranks) LEPP checks whether there are at least three differing observations and whether their ranks are between $1$ and the sample size $n$, which is part of the contents of the matrix `data` and — in case `moments` is provided — this matrix matches `data`. Regarding **data–type 2** (grouped data) LEPP looks if there are at least classes with non–zero frequencies and if these frequencies sum to the sample size. Furthermore, when the last (greatest) class limit is equal to $\infty$, this class is eliminated as it cannot be depicted on the probability paper. For this data–type LEPP asks the user to make his choice among one of the following plotting positions:

1. midpoint,
2. BLOM,
3. WEIBULL,
4. median.

Regarding **data–type 3** (randomly or multiply censored data) LEPP checks the conformity of the sample size and the number of input data and if there are enough uncensored observations to ensure estimation. Finally, LEPP computes the KAPLAN–MEIER estimate of the CDF with the MATLAB command `ecdf`.

The **second part** starts by asking the user for his choice of distribution. We have implemented the `switch` environment of MATLAB to execute all the commands that are specific to a distribution:

- preparation of the ordinate axis of the probability paper,
- chart with DF, CDF, HF and CHF of the reduced form of the chosen distribution,
- computing the regressor vector together with the variance–covariance to execute linear estimation in part three.

The latter task is executed by a function–subprogram named `..-reg(.,.,.,.,.,.)` which may call a lot of other supporting function–subprograms. For data–types 2 and 3 we always make use of the methods presented in Sections 4.3.1 and 4.3.2, but for data–type 1 the method is dependent on the distribution chosen and is one of the methods given in Sect. 4.2. The result–screen of part three tells the user which method has been chosen. With respect to data–type 1 a lot of other function–subprograms may be called for executing integration, forming derivatives, approximating moments and so on. Altogether, LEPP is accompanied by about 160 function–subprograms.

**Part three** of LEPP takes the regressor vector and the variance–covariance matrix of part two and the sample data of part one to compute the linear estimates of the location and scale parameters together with their variance–covariance matrix and the correlation between the sample date and the regressor values. These results along with some other information are displayed in the command window. Another output of part three is a figure with the probability paper showing the data and the fitted regression line.

## 6.2 Input to LEPP

Before calling `LEPP` the user has to provide his sample–data in a matrix `data` which has to be stored in the LEPP–directory. We distinguish between three types of matrices, each named `data`, each having two rows and identified by one of the digits $1,\ 2,\ 3,$ which is an entry in `data`. Also, when the sample–data of type 1 shall be processed with tabulated moments of reduced order statistics, these moments have to be provided in a matrix `moments` in the LEPP–directory.

The first row of `data`, when the sample–data are **order statistics**, has as its first entry the data–type identifier `1` which is followed by the sample data, arranged in ascending order. The second row starts with the sample size followed by the ranks of the order statistics in the row above. The following example shows `data` for a doubly censored sample — two observations censored on each side — of size $n = 12$, which has been taken from the $COO$–data of the data files:

$$\texttt{data}: \quad \begin{array}{cccccc} 1 & 0.8936 & 1.0100 & \dots & 2.9575 & 3.2921 \\ 12 & 3 & 4 & \dots & 9 & 10 \end{array}$$

In case the reader wants to use tabulated moments of reduced order statistics he has to provide — besides `data` — the matrix `moments`. This matrix has in its first row the same ranks as those in `data`, in its second row the means $\alpha_{r:n}$ and in the following rows the upper triangle of the variance–covariance matrix $(\beta_{r,s:n}),\ r \le s$. The following example first shows `data` for selected order statistics in a sample of size $n = 8$ and then the corresponding matrix `moments`, both taken from the $EMN1$–data of the data files:

$$\texttt{data}: \quad \begin{array}{cccccc} 1 & -12.6597 & -10.9792 & 1.0039 & 6.1577 & 6.3672 \\ 8 & 1 & 2 & 5 & 7 & 8 \end{array}$$

$$\texttt{moments}: \quad \begin{array}{ccccc} 1 & 2 & 5 & 7 & 8 \\ -2.6567 & -1.5884 & -0.2312 & 0.4528 & 0.9021 \\ 1.6449 & 0.6118 & 0.1626 & 0.0834 & 0.0543 \\ 0 & 0.6564 & 0.1736 & 0.0893 & 0.0583 \\ 0 & 0 & 0.2317 & 0.1210 & 0.0797 \\ 0 & 0 & 0 & 0.1835 & 0.1223 \\ 0 & 0 & 0 & 0 & 0.1936 \end{array}$$

When the CAUCHY, the exponential or the uniform distributions are to be fitted to the data `moments` is not allowed!

For data of type 2 the sample consists consists of **grouped observations** and the entries to `data` are

- in the first row the identifier `2` followed by the upper class limits in ascending order,
- in the second row the sample size followed by the class frequencies. (LEPP checks if these class frequencies sum to the sample size.)

The following example, taken from the $COO$–data of the data–file, has $n = 80$ observations grouped in 9 classes.

`data :`

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | −3.5 | −2.5 | −1.5 | −0.5 | 0.5 | 1.5 | 2.5 | 3.5 | 5.0 |
| 80 | 3 | 8 | 9 | 9 | 15 | 16 | 11 | 6 | 3 |

For data of type 3 the sample is **multiply or randomly censored** and the entries to `data` are

- in the first row the identifier `3` followed by the observations in ascending order,
- in the second row the sample size followed by the indicators 0 or 1, where 1 stands for a censored observation. (LEPP checks if the number of observations is equal to the sample size!)

The following example shows `data` for the $n = 12$ observations taken from the $TE$–data of the data–file.

`data :`

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3 | 0.1925 | 0.5939 | 0.5968 | 0.7013 | 0.7802 | 0.8122 | 0.8157 | ... |
| 12 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | ... |

## 6.3 Directions for using LEPP

The user should create a directory, perhaps named LEPP, where to copy the program files from the folder ”LEPP-Program”. He may also copy the data files from the folder ”LEPP-Data” into this directory when he wants to run LEPP with one these data files. When using data of his own the reader has to create his file `data` along the specifications given in the preceding section. When he wants to try one of the data–sets of the data files, he should open the desired MAT–file and rename one of the three matrices of this file as `data`.

The next step is to type `LEPP` to start the program. Then, the user follows the instructions given by the program. The sequence of instructions depends on the type of data to be processed. When the user chooses one of the ln–transformable distributions, he will be asked by LEPP to enter the threshold for this distribution. When the data are of type 1 and uncensored, the user can cause LEPP to estimate the threshold or he can enter a value of his own. During a session, there will appear two figures on the screen, a chart with the DF, CDF, HF and CHF of the chosen distribution and the probability paper with data and the regression line. This figure has to be moved to see what is in the command window behind the figure.

# Mathematical and statistical notations

| | |
|---|---|
| $\mathbf{1}$ | vector of ones |
| $\sim$ | distributed as |
| $\stackrel{\text{approx}}{\sim}$ | approximately distributed |
| $\stackrel{\text{asym}}{\sim}$ | asymptotically distributed |
| $\stackrel{\text{iid}}{\sim}$ | independently and identically distributed |
| $\stackrel{\text{d}}{=}$ | equality in distribution |
| $:=$ | equal by definition |
| $\equiv$ | identical or equivalent |
| $\forall$ | for all |
| $\propto$ | proportional to |
| $\widehat{\ }$ | indicating an estimator or estimate |
| $a$ | location parameter |
| $\boldsymbol{a}$ | vector of weights to produce $\widehat{a}$ |
| $\widetilde{a}$ | location parameter, transformed ($\widetilde{a} = \ln b$) |
| $\boldsymbol{\alpha}$ | vector of $\alpha_{r:n}$ |
| $\boldsymbol{\alpha}^*$ | vector of $\alpha^*_{r:n}$ |
| $\alpha_3$ | index of skewness |
| $\alpha_4$ | index of kurtosis |
| $\alpha_{r:n}$ | mean of the $r$–th reduced order statistic $Y_{r:n}$ |
| $\alpha^{(k)}_{r:n}$ | crude single moment of $Y^k_{r:n}$ |
| $\alpha^{(k,\ell)}_{r,s:n}$ | crude product moment of $Y^k_{r:n}$ and $Y^\ell_{s:n}$ |
| $\alpha_{r,s:n}$ | crude product moment of $Y_{r:n}$ and $Y_{s:n}$ |
| $\alpha^*_{r:n}$ | TAYLOR approximation of $\alpha_{r:n}$ |
| $\arcsin$ | arc–sine function or inverse sine function |
| $\arctan$ | arc–tangent function or inverse tangent function |
| ARE | asymptotic relative efficiency |
| $AS(a,b)$ | arc–sine distribution |
| $b$ | scale parameter |

| | |
|---|---|
| $\boldsymbol{b}$ | vector of weights to produce $\widehat{b}$ |
| $\widetilde{b}$ | scale parameter, transformed $(\widetilde{b} = 1/c)$ |
| $\boldsymbol{B}$ | variance–covariance matrix of reduced order statistics |
| $\boldsymbol{B}^*$ | TAYLOR approximation of $\boldsymbol{B}$ |
| $\mathfrak{B}_r$ | $r$–th BERNOULLI number |
| $B(\cdot,\cdot)$ | complete beta function |
| $\beta^{rs}$ | element of $\boldsymbol{B}^{-1}$ |
| $\beta_{r,r:n}$ | variance of $Y_{r:n}$ |
| $\beta_{r,s:n}$ | covariance of $Y_{r:n}$ and $Y_{s:n}$ |
| $\beta^*_{r,r:n}$ | TAYLOR approximation of $\beta_{r,r:n}$ |
| $\beta^*_{r,s:n}$ | TAYLOR approximation of $\beta_{r,s:n}$ |
| BLIE | best linear invariant estimator or estimate |
| BLUE | best linear unbiased estimator or estimate |
| $c$ | shape parameter |
| $C_n$ | CATALAN number |
| $C(t)$ | characteristic function $\big(C(t) = \mathrm{E}(\exp[i\,t\,X])\big)$ |
| $CA(a,b)$ | CAUCHY distribution |
| CCDF | complementary cumulative distribution (reliability or survival function) |
| CDF | cumulative distribution function |
| CHF | cumulative hazard function (cumulative hazard rate) |
| $\chi^2(\nu)$ | $\chi^2$– distribution with $\nu$ degrees of freedom |
| $\chi(\nu)$ | $\chi$– distribution with $\nu$ degrees of freedom |
| cos | cosine function |
| $COO(a,b)$ | ordinary cosine distribution |
| $COR(a,b)$ | raised cosine distribution |
| Cov$(\cdot)$ | covariance |
| csc | cosecant function |
| d | differential operator |
| $\delta_i$ | censoring indicator; $\delta_i = 1$ $(0)$ – censored (uncensored) observation |

| | |
|---|---|
| DF | density function |
| DFR | decreasing failure function (decreasing failure rate) |
| $\mathrm{diag}(\boldsymbol{M})$ | diagonal matrix from square matrix $\boldsymbol{M}$ |
| $\varepsilon$ | random error variate with $\mathrm{E}(\varepsilon) = 0$ |
| $\boldsymbol{\varepsilon}$ | random vector with $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{o}$ |
| $\mathrm{E}(\cdot)$ | expectation or mean |
| $\mathrm{erf}(\cdot)$ | error function |
| $\mathrm{erfi}(\cdot)$ | imaginary error function |
| $EMN1(a, b)$ | extreme value distribution of type I for the minimum |
| $EMN2(a, b, c)$ | extreme value distribution of type II for the minimum |
| $EMN3(a, b, c)$ | extreme value distribution of type III for the minimum |
| $EMX1(a, b)$ | extreme value distribution of type I for the maximum |
| $EMX2(a, b, c)$ | extreme value distribution of type II for the maximum |
| $EMX3(a, b, c)$ | extreme value distribution of type III for the maximum |
| $EX(a, b)$ | exponential distribution |
| $f(\cdot)$ | general density function (DF) |
| $f_X(\cdot)$ | density function of $X$ |
| $f_Y(\cdot)$ | reduced density function |
| $f_{r:n}(\cdot)$ | DF of the $r$–th order statistic |
| $f_{r,s:n}(\cdot)$ | joint DF of the $r$–th and $s$–th order statistics |
| $F(\cdot)$ | general cumulative distribution function (CDF) |
| $F_X(\cdot)$ | CDF of $X$ |
| $F_Y(\cdot)$ | reduced CDF |
| $F^{-1}(P)$ | percentile function |
| $F_{r:n}(\cdot)$ | CDF of the $r$–th order statistic |
| $F_{r,s:n}(\cdot)$ | joint DF of the $r$–th and $s$–th order statistics |
| ${}_pF_q(\boldsymbol{a}; \boldsymbol{b}; z)$ | generalized hypergeometric function |
| $\gamma$ | EULER–MASCHERONI's constant |
| $\gamma(\cdot\vert\cdot)$ | incomplete gamma function (lower part) |

| | |
|---|---|
| $\Gamma(\cdot\|\cdot)$ | incomplete gamma function (upper part) |
| $\Gamma(\cdot)$ | complete gamma function |
| GLS | general least squares |
| $h(\cdot)$ | hazard function (hazard rate) |
| $H(\cdot)$ | cumulative hazard function (cumulative hazard rate) |
| $HC(a,b)$ | half–CAUCHY distribution |
| HF | hazard function (hazard rate) |
| $HL(a,b)$ | half–logistic distribution |
| $HN(a,b)$ | half–normal distribution |
| $HS(a,b)$ | hyperbolic secant distribution |
| $\boldsymbol{I}$ | identity or unity matrix |
| $I(\cdot)$ | entropy |
| $I_0(\cdot)$ | BESSEL function |
| $I_p(c,d)$ | PEARSON's incomplete beta function |
| IFR | increasing failure function (increasing failure rate) |
| iid | identically and independently distributed |
| $K(t)$ | cumulant generating function $\big(K(t) = \ln M(t)\big)$ |
| $\kappa_r(\cdot)$ | $r$–th cumulant |
| $LA(a,b)$ | LAPLACE distribution |
| LEPP | MATLAB program for linear estimation and probability plotting |
| $\lim$ | limit |
| $LNL(a,b,c)$ | lognormal distribution with lower threshold |
| $LNU(a,b,c)$ | lognormal distribution with upper threshold |
| $LO(a,b)$ | logistic distribution |
| $L(t)$ | LAPLACE transform $\big(L(t) = \mathrm{E}(\exp -t\,X)\big)$ |
| $M(t)$ | crude moment generating function $\big(M(t) = \mathrm{E}\{\exp(t\,X)\}\big)$ |
| $M(\cdot,\cdot,\cdot)$ | KUMMER's function |
| $MB(a,b)$ | MAXWELL–BOLTZMANN distribution |
| ML | maximum likelihood |

| | |
|---|---|
| MLE | maximum likelihood estimator or estimate |
| MSE | mean squared error |
| $\mu$ | mean or expectation $(\mu := \mu'_1)$ |
| $\mu_r(\cdot)$ | $r$–th central moment (= moment about the mean) |
| $\mu'_r(\cdot)$ | $r$–th crude moment (= moment about zero or uncorrected moment) |
| $\mu_{r:n}$ | mean or expectation of $X_{r:n}$ |
| $\mu_{r,s:n}$ | crude product moment of $X_{r:n}$ and $X_{s:n}$ |
| $\mu_{r:n}^{(k)}$ | crude moment of $X_{r:n}^k$ |
| $\mu_{r,s:n}^{(k,\ell)}$ | crude product moment of $X_{r:n}^k$ and $X_{s:n}^\ell$ |
| $n_j^u$ | cumulated frequencies up to and including class $j$ |
| $NO(a,b)$ | normal distribution |
| $\boldsymbol{o}$ | vector of zeros |
| $\mathbf{0}$ | matrix of zeros |
| OLS | ordinary least squares |
| $\boldsymbol{\Omega}$ | inverse of $\boldsymbol{B}$ |
| $\boldsymbol{\Omega}^*$ | inverse of $\boldsymbol{B}^*$ |
| $PA(a,b,c)$ | PARETO distribution |
| $PAI(a,b)$ | inverted U–shaped parabolic distribution of order 2 |
| $PAU(a,b)$ | U–shaped parabolic distribution of order 2 |
| $\text{PD}_X(P_2 - P_1)$ | percentile distance of percentiles $x_{P_1},\ x_{P_2}$ |
| $PO(a,b,c)$ | power–function distribution |
| $\varphi(z)$ | DF of the standard normal distribution |
| $\Phi(z)$ | CDF of the standard normal distribution |
| plim | probability limit |
| $p_r$ | mean of $U_{r:n}$ $\big(p_r = r/(n+1)\big)$ |
| $\Pr(\cdot)$ | probability of |
| $\psi(\cdot)$ | digamma function |
| $\psi'(\cdot)$ | trigamma function |
| $q_r$ | complement of $p_r$ $(q_r = 1 - p_r)$ |

| | |
|---|---|
| $\mathbb{R}$ | set of real numbers |
| $R(\bullet)$ | complementary distribution function (reliability or survival function) |
| $RA(a,b)$ | RAYLEIGH distribution |
| $RE(a,b)$ | reflected exponential distribution |
| RMSE | root mean squared error |
| RSS | residual sum of squares |
| $SE(a,b)$ | semi–elliptical distribution |
| sech | hyperbolic secant function |
| $\sigma(\bullet)$ | standard deviation |
| $\sigma^2(\bullet)$ | variance |
| $\sigma_{r,r:n}$ | variance of $X_{r:n}$ |
| $\sigma_{r,s:n}$ | covariance of $X_{r:n}$ and $X_{s:n}$ |
| $\mathbf{\Sigma}$ | variance–covariance matrix of order statistics |
| $\sin$ | sinus function |
| $\sinh$ | hyperbolic sinus function |
| $\tanh$ | hyperbolic tangent function |
| $TE(a,b)$ | TEISSIER distribution |
| $\boldsymbol{\theta}$ | location–scale parameter vector containing $a$ and $b$ |
| $\widetilde{\boldsymbol{\theta}}$ | vector containing $\widetilde{a}$ and $\widetilde{b}$ |
| $TN(a,b)$ | triangular distribution, right–angled and negatively skew |
| $TP(a,b)$ | triangular distribution, right–angled and positively skew |
| $TS(a,b)$ | symmetric triangular distribution |
| TTT | total time on test |
| $u$ | realization of $U$ |
| $U$ | reduced uniform variate |
| $UB(a,b)$ | U–shaped beta distribution |
| $UN(a,b)$ | uniform or rectangular distribution |
| Var$(\bullet)$ | variance for scalar argument or variance–covariance matrix for vector argument |

| | |
|---|---|
| $VS(a,b)$ | V–shaped distribution |
| $x$ | realization of $X$ |
| $X$ | general location–scale distributed variate |
| $x_M$ | mode of $X$ |
| $x_P$ | percentile of order $P,\ \ 0 \leq P \leq 1$ |
| $x_j^u$ | upper class limit of class $j$ |
| $x_j^\ell$ | lower class limit of class $j$ |
| $\overline{X}$ | sample mean of $X$ |
| $\widetilde{X}$ | variate, ln–transformed to location–scale type |
| $X_{r:n}$ | $r$–th order statistic of $X$ in a sample of size $n$ |
| $y$ | realization of $Y$ |
| $Y$ | reduced location–scale distributed variate $\big(Y = (X-a)/b\big)$ |
| $z$ | realization of $Z$ |
| $Z$ | standardized variate $\big(Z = (X-\mu_X)/\sigma_X\big)$ |
| $\zeta(s)$ | RIEMANN's zeta function |
| $Z(t)$ | central moment generating function $\big(\mathrm{E}\big[\exp\{t\,(X-\mu_X)\}\big]\big)$ |

# Bibliography


**A-A-A-A-A**

ABRAMOWITZ, M. / STEGUN, I.A. (eds.) (1965): *Handbook of Mathematical Functions*; Dover, New York

AITCHISON, J. / BROWN, J.A.C. (1957): *The Lognormal Distribution*; Cambridge University Press, Cambridge

AITKEN, A.C. (1935): On least squares and linear combinations of observations; *Proc. Roy. Soc. Edinb.* **55**, 42–48

ALI, M.M. / UMBACH, D. (1998): Optimal linear inference using selected order statistics in location–scale models; in BALAKRISHNAN, N. / RAO, C.R.: *Order Statistics — Applications* (Handbook of Statistics, Vol. 17), Amsterdam etc., North–Holland, 187–213

ANDERSON, E. (1957): A semigraphical method for the analysis of complex problems; *Proceedings of the National Academy of Sciences* **13**, 923–927 (Reprinted in *Technometrics* **2**, 125–136.)

ANDREWS, D.F. (1972): Plot of high dimensional data; *Biometrics* **28**, 125–136

ANSCOMBE, F.J. (1973): Graphs in statistical analysis; *The American Statistician* **27**, 17–21

ARNOLD, B.C. (1983): *Pareto Distributions*; International Co–operative Publishing House, Fairland

ARNOLD, B.C. / BALAKRISHNAN, N. (1989): *Relations, Bounds and Approximations for Order Statistics*; Lecture Notes in Statistics **53**, Springer, New York

ARNOLD, B.C. / BALAKRISHNAN, N. / NAGARAJA, H.N. (1992): *A First Course in Order Statistics;* Wiley, New York etc.

ARNOLD, B.C. / GROENEVELD, R.A. (1980): Some properties of the arcsine distribution; *Journal of the American Statistical Association* **75**, 173–175

**B-B-B-B-B**

BALAKRISHNAN, N. (1985): Order Statistics from the half–logistic distribution; *J. Statist. Comput. Simul.*, **20**, 287–309

BALAKRISHNAN, N. (ed.) (1992): *Handbook of the Logistic Distribution;* Marcel Dekker, New York

BALAKRISHNAN, N. / CHAN, P.S. (1992a): Order statistics from extreme value distributions, I: Tables of means, variances and covariances; *Communications in Statistics — Simulation and Computation*, **11**, 1199–1217

BALAKRISHNAN, N. / CHAN, P.S. (1992b): Extended tables of means, variances and covariances of order statistics from the extreme value distribution for sample sizes up to 30; *Report, Department of Mathematics and Statistics, McMaster University,* Hamilton, Canada

BALAKRISHNAN, N. / CHAN, P.S. (1993): Extended tables of means, variances and covariances of Weibull order statistics; *Report, Department of Mathematics and Statisticss, McMaster University,* Hamilton, Ontario

BALAKRISHNAN, N. / COHEN, A.C. (1991): *Order Statistics and Inference: Estimation Methods;* Academic Press, San Diego

BALAKRISHNAN, N. / PUTHENPURA, S. (1986): Best linear unbiased estimators of location and scale parameters of the half–logistic distribution; *J. Statist. Comput. Simul.*, **25**, 193–204

BALAKRISHNAN, N. / RAO, C.R. (eds.) (1998a): *Order Statistics: Theory and Methods*; Handbook of Statistics, Vol. 16, Elsevier, Amsterdam

BALAKRISHNAN, N. / RAO, C.R. (eds.) (1998b): *Order Statistics: Applications*; Handbook of Statistics, Vol. 17, Elsevier, Amsterdam

BARLOW, R.E. / CAMPO, R. (1975): Time on test processes and applications to failure data analysis; in BARLOW / FUSSSELL / SINGPURWALLA (eds.): *Reliability and Fault Tree Analysis*, SIAM, Philadelphia, 451–481

BARNETT, V. (1966): Order statistics estimators of the location of the Cauchy distribution; *Journal of the American Statistical Association* **61**, 1205–1217; Corrigendum, Ibid., **63**, 383–385

BARNETT, V. (1975): Probability plotting methods and order statistics; *Applied Statistics* **24**, 95–108

BARNETT, V. (1976): Convenient probability plotting positions for the normal distribution; *Applied Statistics* **25**, 47–50

BENIGER, J.R. / ROBYN, D.L. (1978): Quantitative graphics in statistics — A brief history; *The American Statistician* **32**, 1–11

BERGMAN, B. / KLEFSJÖ, B. (1984): The total time on test concept and its use in reliability theory; *Operations Research* **31**, 506–606

BLOM, G. (1958): *Statistical Estimates and Transformed Beta Variables;* Stockholm

BLOM, G. (1962): Nearly best linear estimates of location and scale parameters, in: SARHAN/GREENBERG (eds.): *Contributions to Order Statistics*; Wiley, New York, 34–46

BOSE, R.C. / GUPTA, S.S. (1959): Moments of order statistics from a normal population; *Biometrika* **46**, 433–440

BROWN, B.M. / HETTMANSPERGER, T.P. (1996): Normal scores, normal plots, and tests for normality; *Journal of the American Statistical Association* **91**, 1668–1675

**C-C-C-C-C**

CHERNOFF, H. (1973): The use of faces to represent points in k–dimensional space graphically; *Journal of the American Statistical Association* **70**, 548–554

CHERNOFF, H. / LIEBERMAN, G.J. (1954): Use of normal probability paper; *Journal of the American Statistical Association* **49**, 778–785

CLEVELAND, W.S. (1993): *Visualizing Data*; Hobart Press, Summit, NJ

CLEVELAND, W.S. (1994): *The Elements of Graphic Data*; Hobart Press, Summit, NJ

CLEVELAND, W.S. / MCGILL, R. (1984): Graphical perception: Theory, experimentation, and application to the development of graphical methods; *Journal of the American Statistical Association* **79**, 532–554

COOKE, P. (1979): Statistical inference for bounds of random variables; *Biometrika* **66**, 367–374

COX, D.R. (1978): Some remarks on the role in statistics of graphical methods; *Applied Statistics* **27**, 4–9

CROME, A.W.F. (1782): *Producten–Karte von Europa*; Dessau

CROW, E.L. / SHIMIZU, K. (eds., 1988): *Lognormal Distributions*; Marcel Dekker, New York

**D-D-D-D-D**

DAVID, H.A. (1981): *Order Statistics, 2nd* ed.; Wiley, New York etc.

DESCARTES, R. (1637): *La Géométrie*, Appendix to “Discours de la Méthode”, Paris

DODSON, B. (1994): *Weibull Analysis*; ASQC Quality Press, Milwaukee

DOWNTON, F. (1966b): Linear estimates with polynomial coefficients; *Biometrika* **53**, 129-141

DOWNTON, F. (1966b): Linear estimates of parameters in the extreme value distribution; *Technometrics* **8**, 3–17

DYER, D.D. / WHISENAND, C.W. (1973): Best linear unbiased estimator of the parameter of the Rayleigh distribution – I: Small sample theory for censored order statistics; *IEEE Trans. on Reliab.* **22**, 27–34

**E-E-E-E-E**

EPSTEIN, B. / SOBEL, M. (1953): Life testing; *Journal of the American Statistical Association* **48**, 486–502

**F-F-F-F-F**

FISHER, R.A. / TIPPETT, L.H.C. (1928): Limiting forms of the frequency distribution of the largest and smallest member of a sample; *Proceedings of the Cambridge Philosophical Society* **24**, 180–90

FOURIER, J.B.J. (1821): Notions générales sur la population; *Récherches Statistiques sur la Ville de Paris et le Département de la Seine* **1**, 1–70

FUNKHOUSER, H.G. (1938): Historical development of the graphical representation of statistical data; *Osiris* **3**, 269–404

**G-G-G-G-G**

GALAMBOS, J. (1978): *The Asymptotic Theory of Extreme Order Statistics;* Wiley, New York etc.

GALAMBOS, J. (1998): Univariate extreme value theory and applications; in BALAKRISHNAN/RAO (eds.): *Handbook of Statistics — Order Statistics: Theory and Methods*, North–Holland, Amsterdam etc., 315–333

GALTON, F. (1899): A geometric determination of the median value of a system of normal variants from two of its centiles; *Nature* **16**, 12–104

GODWIN, H.J. (1949): Some low moments of order statistics; *Ann. Math. Statist.* **20**, 279–285

GOVINDARAJULU, Z. (1963): Relationships among moments of order statistics in samples from two related distributions; *Technometrics* **5**, 514–518

GOVINDARAJULU, Z. / EISENSTAT, S. (1965): Best linear estimates of location and scale parameters of a chi (1 d.f.) distribution, using ordered observation; *Rep. Stat. Appl. Res.*, JUSE **12**, 149–164

GOVINDARAJULU, Z. / JOSHI, M. (1968): Best linear unbiased estimation of location and scale parameters of Weibull distribution using ordered observations; *Reports of Statistical Application Research, Japanese Union of Scientists and Engineers* **15**, 1–14

GUERRY, A.M. (1833): *Essai sur la statistique morale de la France*; Paris

GUMBEL, E.J. (1958): *Statistics of Extremes;* Columbia University Press, New York

GUPTA, A.K. (1952): Estimation of the mean and standard deviation of a normal population from a censored sample; *Biometrika* **39**, 260–273

GUPTA, S.S. / MCDONALD, G.G. / GALARNEAU, D.I. (1974): Moments, product moments and percentage points of the order statistics from the lognormal distribution for samples of size twenty and less; *Sankhyā*, **36**, 230–260

GUPTA, S.S. / QUEREISHI, A.S. / SHAH, B.K. (1967): Best linear unbiased estimators of the logistic distribution using order statistics; *Technometrics* **9**, 43–56

**H-H-H-H-H**

HALLEY, E. (1686): On the height of mercury in the barometer at different elevation above the surface of the earth, and on the rising and falling of the mercury on the change of weather; *Philosophical Transactions*, 104–115

HARTER, H.L. (1961): The best linear estimates for the parameters of the normal distribution; in SARHAN/GREENBERG (eds.): *Contributions to Order Statistics*; Wiley, New York; 206–269

HARTER, H.L. (1970): *Order Statistics and Their Use in Testing and Estimation;* two volumes, U.S. Government Printing Office, Washington D.C.

HARTER, H.L. (1984): Another look at plotting positions; *Communications in Statistics – Theory and Methods*; **13**, 1633

HOAGLIN, D.A. (1980): A Poissonness plot; *The American Statistician* **34**, 146–149

HUANG, J.S. (1975): A note on order statistics from the Pareto distribution; *Skand. Aktuarietidskr.* **58**, 187–190

HUFF, D. (1954): *How to Lie with Statistics*, 1st ed.; Norton, New York

**J-J-J-J-J**

JEVONS, W.S. (1863): *A Serious Fall in the Value of Gold Ascertained, and its Social Effects Set Forth*; London

JOHNSON, L.G. (1964): *The Statistical Treatment of Fatigue Experiments;* Amsterdam

JOHNSON, N.L. / KOTZ, S. / BALAKRISHNAN, N. (1994): *Continuous Univariate Distributions*, Vol. 1, 2nd ed., Wiley, New York etc.

JOHNSON, N.L. / KOTZ, S. / BALAKRISHNAN, N. (1995): *Continuous Univariate Distributions*, Vol. 2, 2nd ed., Wiley, New York etc.

JOHNSON, N.L. / KOTZ, S. / KEMP, A. (1992): *Univariate Discrete Distributions*, 2nd ed., Wiley, New York etc.

**K-K-K-K-K**

KABE, D.G. (1972): On moments of order statistics from the Pareto distribution; *Skand. Aktuarietidskr.* **55**, 179–182

KAPLAN, E.L. / MEIER, P. (1958): Nonparametric estimation from incomplete data; *Journal of the American Statistical Association* **53**, 457–481

KEMPERMAN, J.H.B. / SKIBINSKY, M. (1982): On the characterization of an interesting property of the arcsine distribution; *Pacific Journal of Mathematics* **103**, 457–465

KIMBALL, B.F. (1960): On the choice of plotting positions on probability paper; *Journal of the American Statistical Association* **55**, 546–560

KRÄMER, W. (1994): *So überzeugt man mit Statistik*; Campus, Frankfurt/New York

**L-L-L-L-L**

LALANNE, L. (1846): *Mémoire sur les tables graphiques et sur la géométrie anamorphique appliquées à diverses questions se rattachent à l'art de l'ingénieur*; Paris

LAMBERT, J.H. (1760): *Photometrie*; Augustae Vindelicorum

LAMBERT, J.H. (1765): Theorie der Zuverlässigkeit; *Beiträge zum Gebrauch der Mathematik und deren Anwendung* **I**, 424–488, Berlin

LAMBERT, J.H. (1779): *Pyrometrie*, Berlin

LAURENT, A.G. (1975): Failure and mortality from wear and tear — The Teissier–model; in PATIL et al., (eds.): *Statistical Distributions in Scientific Work*, Vol. 2, Reidel Publishing Company, Dordrecht, 301–320

LIEBLEIN, J. (1953): On the exact evaluation of the variances and covariances of order statistics in samples from the extreme–value distribution; *Annals of Mathematical Statistics* **24**, 282–287

LIEBLEIN, J. (1955): On moments of order statistics from the Weibull distribution; *Annals of mathematical Statistics* **26**, 330–333

LLOYD, E.H. (1952): Least–squares estimation of location and scale parameters using order statistics; *Biometrika* **39**, 88–95

LOCKHART, R.A. / STEPHENS, M.A. (1998): The probability plot: Tests of fit based on the correlation coefficient; in BALAKRISHNAN/RAO (eds.): *Handbook of Statistics, Vol. 17 (Order Statistics: Applications)*, North–Holland, Amsterdam etc., 453–473

LOONEY, S.W. / GULLEDGE, T.R. (1985): Probability plotting positions and goodness–of–fit for the normal distribution; *The Statistician* **34**, 297–303

LORENZ, M.O. (1905): Methods of measuring the concentration of wealth; *Journal of the American Statistical Association* **9**, 209–219

LUKIESCH, M. (1965): *Visual Illusions*; Dover, New York

**M-M-M-M-M**

MALIK, H.J. (1966): Exact moments of order statistics from the Pareto distribubution; *Skand. Aktuarietidskr.* **49**, 144–157

MANN, N.R. (1969): Optimum estimators for linear functions of location and scale parameters; *Annals of Mathematical Statistics* **40**, 2149–2155

MARTIN, F.C. / LEAVENS, D.H. (1931): A new grid for fitting a normal probability curve to a given frequency distribution; *Journal of the American Statistical Association* **9**, 209–219

MCELHONE, D.H. / LARSEN, N.C. (1969): Expected values, variances and covariances of order statistics drawn from a Weibull density with integer shape parameter; *U.S. Atomic Energy Commission R&D Report IS–2156*, Ames Laboratory, Iowa State University

MONMONIER, M. (1991): *How to Lie with Maps*; Chicago University Press, Chicago

MOSTELLER, F. (1946): On some useful “inefficient” statistics; *Ann. Math. Statist.* **17**, 377–408

MOSTELLER, F. / TUKEY, J.W. (1949): The uses and usefulness of binomial probability paper; *Journal of the American Statistical Association* **44**, 174–212

MURTHY, D.N.P. / XIE, M. / JIANG, R. (2004): *Weibull Models*; Wiley, New York

**N-N-N-N-N**

NELSON, W. (1972): Theory and applications of hazard plotting for censored failure data; *Technometrics* **14**, 945–966

NORTON, R.M. (1975): On properties of the arc sine law; *Sankhyā* **A 37**, 306–308

NORTON, R.M. (1978): Moment properties and the arc sine law; *Sankhyā* **A 40**, 192–198

NORTON, R.M. (1981): Arc sine distribution; in JOHNSON/KOTZ/BALAKRISHNAN (eds.): *Encyclopedia of Statistical Sciences* ,Vol. I, Wiley, New York etc., 122–124

**O-O-O-O-O**

OGAWA, J. (1951): Contributions to the theory of systematic statistic, I; *Osaka Math. Jour.* **3**, 175–213

OGAWA, J. (1962): Optimum spacing and grouping for the exponential distribution; in SARHAN/GREENBERG (eds.): *Contributions to Order Statistics*; Wiley, New York etc.; 371–380

**P-P-P-P-P**

PATEL, J.K. / KAPADIA, C.H. / OWEN, D.B. (1976): *Handbook of Statistical Distributions*; Marcel Dekker, New York

PEROZZO, L (1880): Della repprasentazione graphica di una collectivà di individui nella successione del tempo; *Annali di Statisica* **12**, 1–16

PLAYFAIR, W. (1788): *Commercial and Political Atlas*; London

PLAYFAIR, W. (1801): *Statistical Breviary; Shewing, on a Principle Entirely New, the Resources of Every State and Kingdom in Europe*; London

**R-R-R-R-R**

RIEDWYL, H. (1987): *Graphische Gestaltung von Zahlenmaterial*; Haupt, Bern

RINNE, H. (1985): Estimation of car-lifetime from prices of used cars — The Teissier–model (in German); in: GÜNTER BUTLER et al. (eds): *Statistik zwischen Theorie und Praxis*, Vandenhoeck & Ruprecht, Göttingen, 172 – 184

RINNE, H. (2009): *The Weibull Distribution — A Handbook*, CRC Press, Boca Raton etc.

ROYSTON, E. (1956): A note on the history of the graphical representation of data (Studies in the history of statistics III); *Biometrika* **3**, 241–247

ROYSTON, J.P. (1982) Expected normal order statistics (exact and approximate); *Applied Stattistics* **31**, 161–165

**S-S-S-S-S**

SARHAN, A.E. / GRENNBERG, B.G. (1956): Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Part I. The normal distribution up to samples of size 10; *Ann. Math. Statist.* **27**, 427–451; *Corrigendum* **40**, 325

SARHAN, A.E. / GREENBERG, B.G. (eds.) (1962): *Contributions to Order Statistics;* Wiley, New York etc.

SCHMID, C.F. / SCHMID, S.E. (1979): *Handbook of Graphic Presentation*, 2nd ed.; Wiley, New York

SHAH, B.K. (1966): On the bivariate moments of order statistics from a logistic distribution; *Ann. Math. Statist.* **37**, 1002–1010

SHANTARAM, R. (1978): On the stochastic equation $X + Y = XY$; in TAILLIE/PATIL/KOTZ (eds.): *Distributions in Statistical Work*, Vol. 4, Reidel, Dordrecht, 419–431

SHANTARAM, R. (1981): A characterization of the arcsine law; *Sankhyā* **A 40**, 190–207

SHAPIRO, S.S. / WILK, M.B. (1965): An analysis–of–variance test for normality (complete samples); *Biometrika* **52**, 591–611

STIGLER, S.M. (1974): Cauchy and the witch of Agnesi: An historical note on the Cauchy distribution, (Studies in the history of probability and statistics, XXXIII); *Biometrika* **61**, 375–379

**T-T-T-T-T**

TALACKO, J. (1956): Perk's distributions and their role in the theory of Wiener's stochastic variables; *Trabajos de Estadistica* **7**, 159–174

TEICHROEW, D. (1956): Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution; *Ann. Math. Statist.* **27**, 410–426

TEISSIER, G. (1934): Recherces sur le vieillissement et sur les lois de la mortalité; *Annales Physiologiques et de Physicochimie Biologique*, **10**, 237–258, 260–284

TIETJEN, G.L. / KAHANER, D.K. / BECKMAN, R.J. (1977): Variances and covariances of the normal order statistics for sample sizes 2 to 50; in AMER. MATH. SOC. (ed.): *Selected Tables in Mathematical Statistics*, vol. 5; 1–74

TILLING, L. (1975): Early experimental graphs; *British Journal for the History of Science* **8**, 193–213

TOLANSKY, S. (1964): *Optical Illusions*; Pergamon Press, Oxford

TUFTE, E.R. (2001): *The Visual Display of Quantitative Information*; Sec. ed., Graphics Press, Cheshire

TUFTE, E.R. (1992): *Envisioning Information*; Graphics Press, Cheshire

TUKEY, J.W. (1977): *Exploratory Data Analysis*, Addison–Wesley, Reading

**V-V-V-V-V**

VAUGHAN, D.C. (1994): The exact values of the expected values, variances and covariances of the order statistics from the Cauchy distribution; *J. Stat. Comput. Simul.* **49**, No. 1–2, 21–32

**W-W-W-W-W**

WALKER, F.A. (1874): *Statistical Atlas of the United States, Based on the Results of the Ninth Census 1870*; U.S. Bureau of the Census, Washington D.C.

WANG, P.C.C. (1978): *Graphical Representation of Multivariate Date*; Academic Press, New York

WEIBULL, W. (1939): The phenomenon of rupture in solids; *Ingeniörs Vetenskaps Akademiens Handligar Report No. 153,* Stockholm

WEIBULL, W. (1959): Statistical evaluation of data from fatigue and creep–rupture tests; Part I, Fundamental concepts an general methods; *WADC TR59–400, Part I, Wright Air Development Center*, Wright–Patterson Air Force Base, Dayton, OH

WEIBULL, W. (1967): Estimation of distribution parameters by a combination of the best linear order statistic method and maximum likelihood; *Air Force Materials Laboratory Technical Report AFML-TR-67-105*, Wright–Patterson Air Force Base, Dayton, OH

WESTBERG, U. / KLEFSJÖ, B. (1994): TTT–plotting for censored data based on piecewise exponential estimator; *Int. Jour. Reliability, Quality and Safety Engineering* **1**, 1–13

WHITE, J.S. (1964): Least–squares unbiased censored linear estimation for the log Weibull extreme value) distribution; *Journal of Industrial Mathematics* **14**, 21–60

WHITE, J.S. (1969): The moments of the log Weibull order statistics; *Technometrics* **11**, 373–386

**Z-Z-Z-Z-Z**

ZEUNER, G. (1869): *Abhandlungen aus der mathematischen Statistik*; Leipzig

# Author Index

# Subject Index