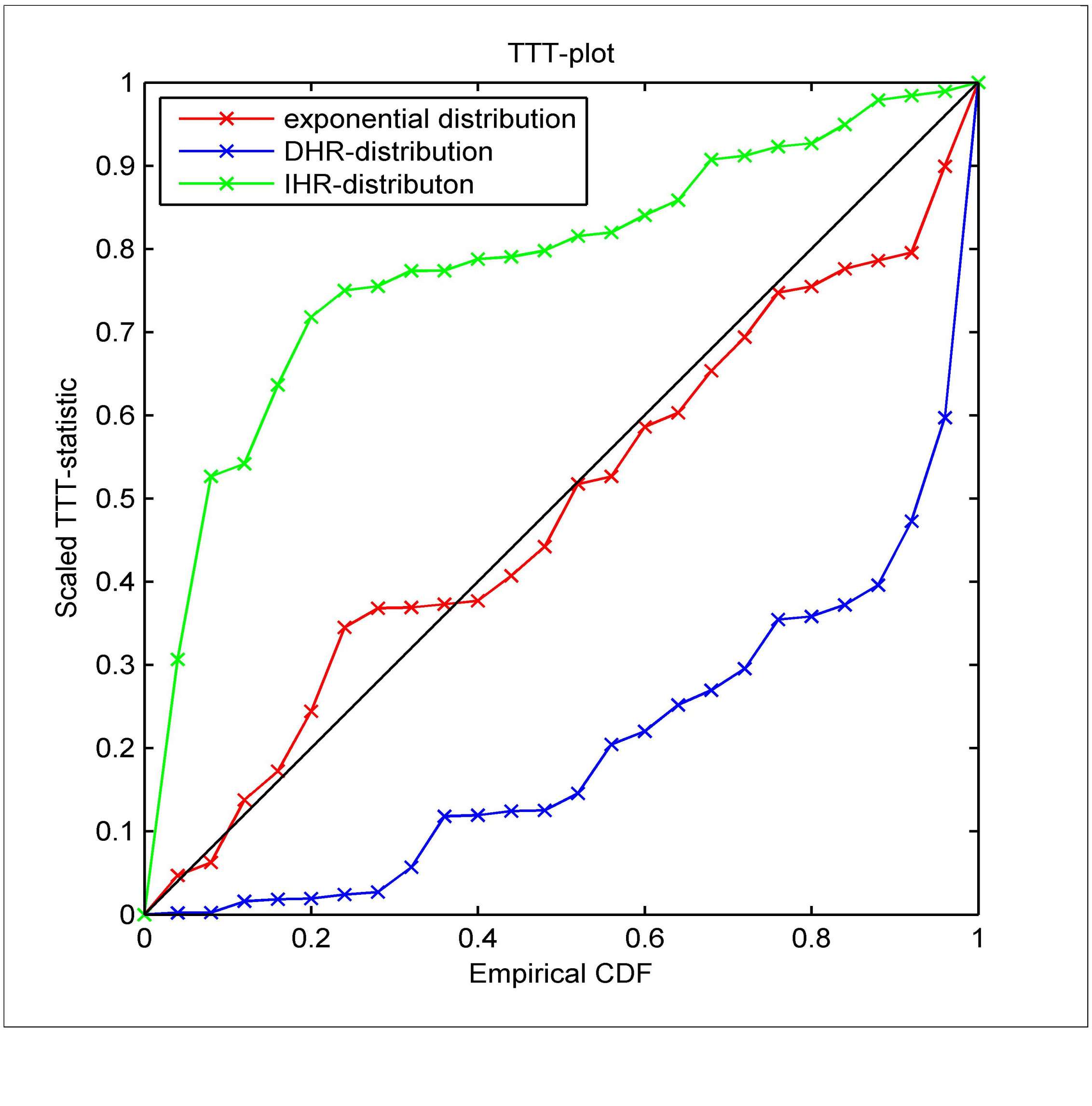
The Hazard Rate

— Theory and Inference —

(With supplementary MATLAB–Programs)

Horst Rinne



Department of Economics and Management Science
Chair of Econometrics and Statistics
Justus–Liebig–University, D 35394 Giessen, Germany

Preface

When we look at biological organisms like human beings, animals, plants or at technical devices like motorcars, aircrafts, television sets and parts thereof or at economic and socio–economic units like enterprizes, corporations, labor unions or at social units like families, parties or even states, we observe that in every moment of their existence we will find them in a well–defined state. A patient, after some medical treatment, may be alive, an adult person may be out of work, a piece of machinery may either be down or functioning, a labor union may be on strike or a state may be at war with some other state. The sojourn–time in a given state for such a unit ends by the occurrence of some random event. In this book, the time–to–event since entering into a give state will generally be called lifetime and the terminating event will be called failure. Since the terminating event is random the lifetime is a random variable with realizations which — in general — will be non–negative.

The main body of this monograph is given by Parts I and II. The final Part III, entitled 'Appendices', gives the usual ingredients of a scientific opus: the bibliography and an author index as well as a subject index. Here one will also find information about the MATLAB–programs which have been written to facilitate practical working with the methods and procedures described in the monograph. Part I is descriptive in the sense that gives definitions for the various functions of the variate 'lifetime' and explores how these functions are related to one another. This is done in Chapter 1 where we also differentiate between the univariate and the multivariate cases and between the continuous and the discrete cases. Chapter 2 introduces several classes of lifetime distributions with respect to aging. Chapter 3 is devoted to univariate parametric distributions and enumerates important continuous and discrete distributions known in probability theory. Here the reader will find the formulas for the four most important representatives of a lifetime variate: its probability density in the continuous case or its probability mass function in the discrete case, its survival function, its hazard rate function and its mean residual life function. In order to have a graph of these four functions for any set of the function parameters the reader can revert to the MATLAB–programs stored in 'Distributions.zip' and described in Part III.

Part II is devoted to the inference of the hazard rate. The focus is on non–BAYESIAN and non–parametric inferential procedures of — unless stated otherwise — univariate continuous lifetime distributions. The non–parametric approach to estimate hazard rates from lifetime data is flexible, model–free and data–driven. No shape assumption is imposed other than that the hazard rate is a smooth function, or occasionally in Chapter 7, a monotone function. Such an approach typically involves smoothing of an initial and discrete hazard rate estimate, with arbitrary choice of the smoother.

In Chapters 5 and 6 we present estimation techniques to find such initial estimates for non–grouped and grouped data after having introduced — in Chapter 4 — sampling techniques for lifetime data with the pertaining denotation of the quantities coming up. The core chapter of Part II is Chapter 8 presenting smoothing techniques. Emphasis here is on smoothing with kernels, a technique that is most elaborated, explored and used in practice, but we also look at some other techniques.

Chapter 9 on hazard plotting is an exception from the non–parametric nature of this second part when we plot on special graph paper. For each location–scale distribution we can design an especially scaled grid so that the data points will lie on a straight line when coming from that distribution. So the graph is a means of testing for a special distribution. Estimates of the pertaining parameters of this distribution can be found as special hazard quantiles. Hazard plotting thus serves as an instrument for estimating and testing and is the bridge to the testing procedures of Chapter 10.

Testing procedures in Chapter 10 are not devoted to hypotheses on parameters of some parametric lifetime distribution, but they are concerned with the presence or absence of certain aging properties. These properties have been described in Chapter 2 and will be tested here by means of non–parametric methods. The testing of the no–aging property (= constant hazard rate) may be seen and taken as testing for exponentially distributed lifetime and thus is parametric. Most of the testing procedures are of numerical type, but with the total–time–on–test plot we also have a graphical approach. After presentation of the prerequisites like order statistics, spacings and TTT–transform we test for properties of the hazard rate, i.e., its shape and behavior, and we test for aging classes.

Contents

List of Figures

List of Tables

Part I

Theoretical and Probabilistic Concepts

1 The Hazard Rate and its Relatives

In the course of time several functions have evolved in probability theory that define and describe a random variable. Here, we will focus on those functions and concepts which are related to lifetime as a random variable.

1.1 The Univariate Continuous Case

The emphasis in this section is on those functions which are of special interest in describing the evolution of the risks to which a given unit is subjected over time. Thus, the variate under study is the **lifetime** X — or more generally —the duration or sojourn–time of the unit spent in a given state. Lifetimes or survival times are data that measure the time between two events, namely that of entering into the state and that of escaping from that state. The latter event will be called **failure**. Generally, this time time between events is a one–dimensional and continuous variate, defined on $[0, \infty)$ unless stated otherwise. Multidimensional variates will be discussed in Sect. 1.3 and discrete variates in Sect. 1.2. There are situations where lifetime can be thought of to be negative, e.g., **shelf–aging** meaning that a unit might fail before its real usage will start and this starting is taken as origin of time.

1.1.1 Functions Describing Lifetime[1]

Six representatives of a lifetime distribution will be discussed, each of them bearing different names depending on their field of application:

1. the probability density function $f(x)$, abbreviated by **PDF**, density function for short, also known as failure density, or as failure rate;

2. the cumulative distribution function $F(x)$, abbreviated by **CDF**, distribution function for short, also known as failure function or as lifetime distribution function;

3. the complementary cumulative distribution function $S(x)$, abbreviated by **CCDF**, also known as survival function or as reliability function;

4. the hazard rate function $h(x)$, abbreviated by **HR**, hazard rate for short, also known as instantaneous failure rate or as force of mortality;

5. the cumulative hazard rate $H(x)$, abbreviated by **CHR**, also known as integrated hazard rate;

6. the mean residual life function $\mu(x)$, abbreviated by **MRL**, also known as life expectancy of an x–survivor or as mean future life of an x–survivor.

The origins of some of these functions date back to the 17^{th} century when the first life tables came into existence, whereas their application in the engineering sciences and in the life sciences only started in the 1950's. Each of these six functions completely describes the distribution of lifetime, and any of these functions determines the other five, see Tab. 1/1 at the end of Sect. 1.1.1.6. The six functions are answering different questions with respect to the lifetime variable. The choice further depends on whether

[1] Suggested reading for this section: BAIN/ENGELHARDT (1991, Chapter 1), LEEMIS (1986; 1995, Chapter 3), RINNE (2009, Chapter 2), SMITH (2002, Chapters 1 and 2).

- the mathematical representation has a tractable form or
- intuition is gained concerning the distribution by seeing a plot of the representative.

The six representatives are not the only ways to define the distribution of a random variable X. Other concepts include, e.g.:

- the moment generating function $\mathrm{E}\big(e^{s\,X}\big)$ with E as expectation operator,
- the probability generating function $\mathrm{E}\big(Z^{X}\big)$ for discrete X,
- the characteristic function $\mathrm{E}\big(e^{i\,s\,X}\big),\ i := \sqrt{-1}$,
- the MELLIN transform $\mathrm{E}\big(X^{s-1}\big)$, R. H. MELLIN (1854 – 1933),
- the LAPLACE transform $\mathrm{E}\big(e^{-s\,X}\big)$, P. S. DE LAPLACE (1749 – 1827),
- the exponential FOURIER transform $\mathrm{E}\big(e^{-i\,s\,X}\big)$, J. B. J. FOURIER (1768 – 1830),
- the density quantile function $f\big[F^{-1}(P)\big],\ 0 \le P \le 1$;
- the total–time–on–test transform $\int\limits_0^{F^{-1}(P)} S(x)\,dx,\ 0 \le P \le 1$.

The six representatives used here have been chosen because of their special meaning for lifetime data, for their intuitive appeal, for their usefulness in lifetime data analysis, and — last but not least — for their popularity in probability theory and in statistics.

1.1.1.1 The Failure Density Function

The first lifetime distribution representative to be described is the **failure density** PDF — or in a more general context — the probability density function, defined as

$$f(x) := \lim_{\Delta \to 0} \frac{\Pr\left(x - \dfrac{\Delta}{2} < X \le x + \dfrac{\Delta}{2}\right)}{\Delta}. \tag{1.1a}$$

Thus, for Δ small the product $f(x)\,\Delta$ approximates the probability of failure in the time interval $(x - \Delta/2,\ x + \Delta/2]$ or — roughly speaking — the probability of failure around an age of x. The probability of reaching an age between x_l and $x_u,\ x_l < x_u$, is

$$\Pr(x_l \le X \le x_u) = \int\limits_{x_l}^{x_u} f(x)\,dx. \tag{1.1b}$$

Especially for a newly born organism or creature or a just produced unit, e.g., for a unit starting at age $x = 0$, the probability to fail up to an age $x > 0$ is given by

$$\Pr(X \le x) = \int\limits_0^{x} f(u)\,du. \tag{1.1c}$$

For varying x formula (1.1c) gives the lifetime distribution function, see (1.2a).

Theorem 1: All probability density functions for the variate 'lifetime' must satisfy two conditions:

1. $$f(x) \geq 0, \;\; \forall\, x \geq 0, \tag{1.1d}$$
2. $$\int\limits_0^\infty f(x)\,dx = 1. \qquad \blacksquare \tag{1.1e}$$

Remarks:

1. When X has a parametric distribution with a shift parameter $a \in \mathbb{R}$ (1.1d,e) turn into
$$f(x) \geq 0 \;\; \forall\, x \geq a \quad \text{and} \quad \int\limits_a^\infty f(x)\,dx = 1.$$
a is called **safe life** when $a > 0$, i.e., failing before the age a is impossible. When $a < 0$ we have **shelf–aging**.
2. Usually, when describing a particular PDF only its non–zero part will be explicitly stated, and it should be understood that PDF is zero over any unspecified region of $\mathbb{R}$.
3. Many characteristics such as age, length, weight etc. are true continuous variables, at least conceptually, although it could be said that due to the physical limitations of measuring devices, the characteristic can be observed only as a discrete variable. However, the measurement restrictions are usually insignificant to other sources of error, and the continuous model is mathematically and conceptually much more convenient.
4. For continuous variables we have the following numerical equivalences:
$$\Pr(x_l \leq X \leq x_u) = \Pr(x_l < X \leq x_u) = \Pr(x_l \leq X < x_u) = \Pr(x_l < X < x_u),$$
i.e., the probability of failing within a given interval is the same whether we include or exclude none, one or both of its end–points.

Generally, for lifetime X the density function is **positively skewed** (skewed to the right or steep on the left–hand side), see Fig. 1/1 below. Thus, $f(x)$ has a flat and relatively long right–hand tail, meaning that longer lifetimes are less probable than shorter lifetimes and that the mean life (life expectancy) is greater than the median life, see Sect. 1.1.1.2.

1.1.1.2 The Lifetime Distribution Function

The second lifetime distribution representative is the failure function or lifetime function CDF, defined as
$$F(x) := \Pr(X \leq x), \; x \geq 0, \tag{1.2a}$$
giving the probability of failing up to age x or of having a life span of at most length x.

Theorem 2: Any function $F(x)$ may be the CDF of a lifetime variable if its satisfies the following properties:

1. $$\lim_{x\to 0} F(x) = 0, \tag{1.2b}$$
2. $$\lim_{x\to\infty} F(x) = 1, \tag{1.2c}$$
3. $$F(x_u) \geq F(x_l) \;\; \forall\, x_u > x_l, \text{ i.e., } F(x) \text{ is a non–decreasing function of } x, \tag{1.2d}$$
4. $$F(x) \text{ is continuous, i.e., } \lim_{\Delta\to 0} F(x+\Delta) = \lim_{\Delta\to 0} F(x-\Delta) = F(x), \; \Delta > 0. \quad \blacksquare \tag{1.2e}$$

Remarks:

1. Because $F(x)$ is a probability, see (1.2a), we have
$$0 \leq F(x) \leq 1. \tag{1.2f}$$

2. CDF and PDF are related as
$$f(x) = \frac{dF(x)}{dx} \text{ and} \tag{1.2g}$$
$$F(x) = \int_0^x f(u)\,du. \tag{1.2h}$$

Because $F(x)$ is a monotone and increasing[2] function, see Fig. 1/1 below, the inverse function $F^{-1}(\boldsymbol{.})$ exists and is called **percentile function** or **quantile function**:
$$F(x_P) = P \quad \Longrightarrow \quad x_P = F^{-1}(P),\ 0 \leq P \leq 1. \tag{1.3}$$

x_P is called the percentile or quantile of order P. The special percentile $x_{0.5}$ is called **median life**, i.e., there are equal chances of failing before or surviving beyond the age $x_{0.5}$. Because of the positive skewness of most lifetime densities the median life is more popular than the **mean life** $\mu := \mathrm{E}(X)$ in measuring the central tendency by a single number. For positive skewness of PDF we find $_{0.5} < \mu$.

1.1.1.3 The Survival Function

Another lifetime distribution representative is CCDF, the **survival function** or **reliability function**, defined as
$$S(x) := \Pr(X > x),\ x \geq 0, \tag{1.4a}$$
indicating the probability of surviving an age of x or becoming older than x. From (1.2a) and (1.4a) we see that the lifetime distribution and the survival function are complementary functions:
$$S(x) = 1 - F(x) \text{ and } F(x) = 1 - S(x). \tag{1.4b}$$

Thus, $S(x)$ is the probability of *exceeding* x and $F(x)$ is the probability of *reaching* x, or — stated for a technical unit — $S(x)$ gives the probability of its functioning at time x and $F(x)$ is the probability of its being down at time x. PDF and CCDF are related as
$$f(x) = -\frac{d}{dx}\,S(x) \text{ and} \tag{1.4c}$$
$$S(x) = \int_x^\infty f(u)\,du. \tag{1.4d}$$

The study of $S(x)$ is at the heart of survival analysis and reliability theory. The survival function is important in describing systems of components, i.e., in calculating systems' reliability, see Sect. 1.1.2.3.

[2] In this text we always use **increasing** in the sense of **non–decreasing**, and **decreasing** has the meaning of **non–increasing**.

From (1.4d) we can establish the following relations, simply because PDFs integrate to one:

$$S(0) = \int_0^\infty f(u)\,du = 1. \tag{1.4e}$$

Furthermore,

$$S(\infty) = \lim_{x\to\infty} S(x) = \lim_{x\to\infty} \int_x^\infty f(u)\,du = 0. \tag{1.4f}$$

Finally, for $x_u \geq x_l$:

$$S(x_l) - S(x_u) = \int_{x_l}^{x_u} f(u)\,du \geq 0. \tag{1.4g}$$

These properties establish the following

Theorem 3: The survival function $S(x)$ is monotone and decreasing over its support $[0,\ \infty)$. Furthermore, $S(x)$ satisfies $S(0) = 1,\ S(\infty) = 0$. ■

In fact, any monotone decreasing function $S(x)$ with support $[0,\ \infty)$ and $S(0) = 1,\ S(\infty) = 0$ is the survival function of some lifetime variate, for an example see Fig. 1/1 further down. The matching random variable is the one having PDF as $f(x) = -dS(x)/dx$.

The **raw moments** $\mu'_k = \mathrm{E}\big(X^k\big);\ k = 0, 1, 2, \ldots;$ of random lifetime X may be expressed in terms of its survival function $S(\cdot)$ as stated in the following

Theorem 4: Let X be a variate with $F(x),\ S(x) = 1 - F(x)$ and $f(x) = F'(x) = -S'(x)$, all functions defined on $[a,\ \infty)$. Then

$$\begin{aligned}\mu'_k \;:=\; \mathrm{E}\big(X^k\big) \;&:=\; \int_a^\infty x^k\, f(x)\,dx \\ &=\; a^k + k\int_a^\infty x^{k-1}\, S(x)\,dx;\ k = 0, 1, 2, \ldots;\end{aligned} \tag{1.5a}$$

if and only if

$$\lim_{x\to\infty}\big[x^{k-1}\, f(x)\big] = 0 \text{ for } k < \infty. \quad ■ \tag{1.5b}$$

Proof of Theorem 4: Integrating by parts the last term on the right–hand side of (1.5a) gives

$$\begin{aligned}\int_a^\infty S(x)\,k\,x^{k-1}\,dx &= \Big(S(x)\,x^k\Big)\Big|_a^\infty - \int_a^\infty -f(x)\,x^k\,dx \\ &= \lim_{x\to\infty}\Big[S(x)\,x^k\Big] - S(a)\,a^k + \int_a^\infty x^k\,f(x)\,dx \\ &= \lim_{x\to\infty}\Big[S(x)\,x^k\Big] - a^k + \mu'_k, \text{ because } S(a) = 1.\end{aligned}$$

Applying once L' HOSPITALS's rule to the indeterminate form $\lim_{x\to\infty}\Big[S(x)\,x^k\Big]$ gives

$$\begin{aligned}\lim_{x\to\infty}\frac{S(x)}{x^{-k}} &= \lim_{x\to\infty}\frac{-f(x)}{-k}\,x^{-k-1} \\ &= \frac{1}{k}\lim_{x\to\infty}\Big[x^{k+1}\,f(x)\Big] \\ &= 0, \text{ because of (1.5b).} \quad ■\end{aligned}$$

Two special cases of (1.5a) for a lifetime variate with $a = 0$ are:

1. the mean, often called **mean time to failure** and abbreviated by **MTTF**,

$$\mu := \mu_1' := \mathrm{E}(X) = \int\limits_0^\infty S(x)\,dx; \tag{1.5c}$$

   (Therefore, to find the average lifespan, we integrate the survival function over its support, i.e., the mean life is equal to the area beneath the survival function.) and

2. the **variance**

$$\begin{aligned} \sigma^2 := \mathrm{Var}(X) &= \mathrm{E}\big(X^2\big) - \mu^2 \\ &= 2\int\limits_0^\infty x\,S(x)\,dx - \left[\int\limits_0^\infty S(x)\,dx\right]^2. \end{aligned} \tag{1.5d}$$

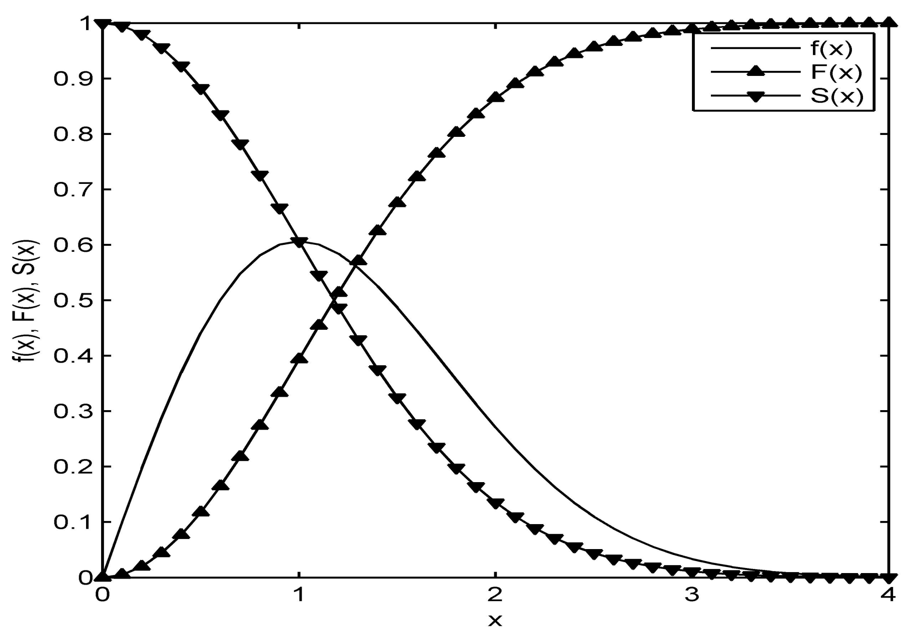
Example 1/1: PDF, CDF, and CCDF of the linear hazard rate distribution

A distribution with the linear hazard rate $h(x) = a + b\,x;\ x \ge 0,\ a \ge 0,\ b > 0$; has:

$$\begin{aligned} f(x) &= (a + b\,x)\,\exp\Big\{-a\,x - \frac{b}{2}\,x^2\Big\}, \\ F(x) &= 1 - \exp\Big\{-a\,x - \frac{b}{2}\,x^2\Big\}, \\ S(x) &= \exp\Big\{-a\,x - \frac{b}{2}\,x^2\Big\}. \end{aligned}$$

Fig. 1/1 shows the functions $f(x),\ F(x)$, and $S(x)$ for $a = 0$ and $b = 1$, which is nothing but the reduced RAYLEIGH distribution, a special case of the WEIBULL distribution.

Figure 1/1: PDF, CDF, and CCDF of the linear hazard rate distribution with $a = 0$ and $b = 1$



$$
\begin{aligned}
x_{\text{mode}} &= -\frac{a}{b} + \sqrt{\frac{1}{b}} \\
x_P &= -\frac{a}{b} + \sqrt{\left(\frac{a}{b}\right)^2 - \frac{2}{b}\,\ln(1-P)}, \quad 0 \le P \le 1 \\
\mu_r &= \mathrm{E}\big(x^r\big) \;=\; \sum_{i=0}^{\infty} \frac{(-b/2)^i}{i!} \left[\frac{\Gamma(2\,i+r+1)}{a^{2\,i+r}} + b\,\frac{\Gamma(2+r+2\,i)}{a^{2\,i+r+2}}\right]
\end{aligned}
$$

In the special case $a = 0$ we have $\mu_r = \Gamma(1+r/2)\big/(b/2)^{r/2}$.

1.1.1.4 The Hazard Rate Function

The reliability (survival) function examines the chance that breakdowns of organisms, of technical units etc. occur beyond a given point in time. To monitor the lifetime of a unit across the support of its lifetime distribution, the **hazard rate** $h(x)$ is used.

In fact, the hazard rate usually is more informative about the underlying mechanism of failure than the other representatives of a lifetime distribution. For this reason, consideration of the hazard rate may be the dominant method for summarizing survival data. COX/OAKES (1984, p. 16) give the following number of reasons why consideration of the hazard rate may be a good idea:

> “(i) it may be physically enlightening to consider the immediate ‘risk’ attaching to an individual known to be alive at age t,
>
> (ii) comparison of groups of individuals are sometimes intensively made via the hazard,
>
> (iii) hazard–based models are often convenient when there is censoring or there are several types of failure,
>
> (iv) comparison with an exponential distribution is particular simple in terms of the hazard,
>
> (v) the hazard is the special form for the ‘single failure’ system of the complete intensity function for more elaborate point processes, i.e., systems in which several point events can occur for each individual.”

The hazard rate is perhaps the most popular of the six representatives modeling and analyzing lifetime data. This is due to its intuitive interpretation as the amount of risk to fail associated with a unit at age x. Another reason for its popularity is that it is a special case of the **intensity function** for a non–homogeneous POISSON process. A hazard rate function models the occurrence of only one, namely the first event (= failure), whereas the intensity function models the occurrence of a sequence of events over time.

The hazard rate goes by several aliases.

- In the engineering sciences it is known as the failure rate.[3]
- In actuarial science it is known as the **force of mortality** or **force of decrement**
- In vital statistics and in the life sciences it is known as the **age–specific death rate**.

[3] This name gives reason for confusion with the failure density.

- In economics its *reciprocal* is known as **MILLS' ratio**, see MILLS (1926).

- In point process and extreme value theory it is known as the **rate function** or **intensity function**.

The hazard rate can be derived using the concept of *conditional probability*. Let A and B be two random events with $\Pr(A) > 0$, than the probability of the conditional event $B \,|\, A$ (= event B happens, given event A has happened) is defined as

$$\Pr(B \,|\, A) = \frac{\Pr(A \cap B)}{\Pr(A)}, \tag{1.6a}$$

where $A \cap B$ means that events A and B happen simultaneously. Now let $A :=$ '$X > x$' (= lifetime is greater than x) and $B :=$ '$X > x + y$', then, evidently, $A \cap B =$ '$X > x + y$'. As $\Pr(A) = \Pr(X > x) = S(x)$ and $\Pr(A \cap B) = \Pr(B) = \Pr(X > x + y) = S(x + y)$ the conditional survival probability according to (1.6a) results as

$$\Pr(X > x + y \,|\, X > x) = \frac{S(x + y)}{S(x)}. \tag{1.6b}$$

The conditional event '$X > x + y \,|\, X > x$' can be transformed to '$X - x > y \,|\, X > x$' and the corresponding conditional variate

$$Y \,|\, X > x := X - x \,|\, X > x$$

is called **future lifetime** or **remaining lifetime** of an x–survivor.[4] We may write (1.6b) as

$$S(y \,|\, X > x) = \frac{S(x + y)}{S(x)} \tag{1.6c}$$

which is called **conditional survival function**. Its complement is the **conditional distribution function**

$$\begin{aligned} F(y \,|\, X > x) &= 1 - \frac{S(x + y)}{S(x)} \\ &= \frac{S(x) - S(x + y)}{S(x)} \\ &= \frac{F(x + y) - F(x)}{1 - F(x)}. \end{aligned} \tag{1.6d}$$

Differentiating (1.6d) with respect to y gives the **conditional failure density**:

$$\begin{aligned} f(y \,|\, X > x) &= \frac{dF(y \,|\, X > x)}{dy} \\ &= \frac{d}{dy}\left(\frac{F(x + y) - F(x)}{1 - F(x)}\right) \\ &= \frac{f(x + y)}{1 - F(x)} = \frac{f(x + y)}{S(x)}. \end{aligned} \tag{1.6e}$$

[4] This variate plays a role in Sect. 1.1.1.6 and is discussed in more detail under the heading of truncated lifetime distributions in Sect. 1.1.2.5.

$f(y \mid X > x)$ really is a density function, as the two conditions (1.1d,e) are fulfilled:

$$
\begin{array}{lrcl}
1. & f(y \mid X > x) & \geq & 0, \;\text{ because }\; f(x+y) \geq 0 \;\text{ and }\; S(x) > 0. \\
2. & \int\limits_0^\infty f(y \mid X > x)\, dy & = & \frac{1}{S(x)} \int\limits_0^\infty f(x+y)\, dy \\
 & & = & \frac{1}{S(x)} \int\limits_x^\infty f(u)\, du, \;\; u = x+y \\
 & & = & \frac{1}{S(x)} S(x) \;\; = 1.
\end{array}
$$

For small Δ we have

$$
\begin{aligned}
\Pr(x < X \leq x + \Delta \mid X > x) &\approx f(\Delta \mid X > x)\, \Delta \\
&= \frac{f(x+\Delta)}{S(x)}\, \Delta.
\end{aligned} \tag{1.7a}
$$

This is an approximation of an x–survivor's chance to fail within the small time span Δ adjacent to x. Now, the hazard rate follows from (1.6e) and (1.7a) with $\Delta \to 0$:

$$
\begin{aligned}
h(x) &= \lim_{\Delta \to 0} f(\Delta \mid X > x) \\
&= \lim_{\Delta \to 0} \frac{f(x+\Delta)}{S(x)} \\
&= \frac{f(x)}{S(x)}, \; S(x) > 0.
\end{aligned} \tag{1.7b}
$$

In other words, for a small increment in time, Δ, the conditional probability that an x–survivor fails in the time interval $(x,\; x+\Delta]$ is roughly equal to the product $h(x)\,\Delta$. Another possible interpretation of $h(x)$ is the rate at which failures occur per unit of time relative to the portion of the population which has not yet failed.

When we want to predict the chance of failure at age x for a newly born or produced unit having $F(x)$ as its CDF we have to use $f(x)$, i.e., $f(x)$ is an **unconditional predictor** for risk to fail at x. When we know that a unit has survived up to x, we have to use $h(x)$ which is a **conditional predictor**. Comparing numerically $f(x)$ to $h(x)$ we notice:

- $f(0) = h(0)$,
- $f(x) \geq h(x) \;\; \forall\, x > 0$, because $S(x) \leq 1 \;\; \forall\, x > 0$.

There is a fundamental difference between the hazard rate function $h(x)$ and the conditional failure density $f(y \mid X > x)$.

1. $h(x)$ is a function of x, the age reached, whereas $f(y \mid X > x)$ is a function of the future lifetime y following a given age x.
2. Both, $h(x)$ and $f(y \mid X > x)$ are non–negative, but $h(x)$ is not a density function as it is not normalized, instead we have $\int\limits_0^\infty h(x)\, dx = \infty$, see Theorem 5.

Summarizing we can state the following

Theorem 5: Any function $h(x)$ is a HR if and only if it satisfies the following properties:

$$1. \quad h(x) \geq 0 \;\; \forall\, x \geq 0, \tag{1.7c}$$

$$2. \quad \int_0^\infty h(x)\,dx = \infty. \qquad \blacksquare \tag{1.7d}$$

Proof of Theorem 5: The properties necessarily hold since

1. $f(x) \geq 0$ and $S(x) > 0$, thus $h(x) = f(x) \big/ S(x) \geq 0$,

2. $$\begin{aligned} \int_0^\infty h(x)\,dx &= -\int_0^\infty d\big[\ln S(x)\big], \quad \text{see (1.8c)} \\ &= -\ln S(x)\Big|_0^\infty \\ &= \ln S(0) - \ln S(\infty) \\ &= \ln 1 - \ln 0 \\ &= \infty. \qquad \blacksquare \end{aligned}$$

The hazard rate measures the propensity to fail or to die depending on the age reached and it thus plays a key role in characterizing the process of aging and in classifying lifetime distributions, see Sect. 2. Generally, HR more precisely describes the stochastic regularity of the variate ‘lifetime’ than the positively skewed course of PDF or the monotone courses of CDF or CCDF. We will distinguish between

- monotone hazard rates, either increasing, when the unit is wearing out with age, or decreasing, when the unit is improving with age, and
- non–monotone hazard rates either U–shaped (= bathtub–shaped) as, e.g., is the case with the age–specific death rate in human life tables, or having any other non–monotone course, e.g., an inverted bathtub–shape.

It is easily possible to express the hazard rate of a population by its PDF, CDF, and CCDF.[5]

$$h(x) = \frac{f(x)}{\int_x^\infty f(u)\,du}, \tag{1.8a}$$

$$= \frac{F'(x)}{1 - F(x)}, \tag{1.8b}$$

$$= -\frac{S'(x)}{S(x)} = -\frac{d\ln S(x)}{dx}. \tag{1.8c}$$

Conversely, we may write the PDF, CD, and CCDF of a population in terms of its HR. Integrating in (1.8c) yields

$$\begin{aligned} \int_0^x h(u)\,du &= -\ln S(x)\Big|_0^x \\ &= -\ln S(x) + \ln S(0) \\ &= -\ln S(x), \quad \text{because} \quad S(0) = 1. \end{aligned} \tag{1.9a}$$

[5] It is also possible to write the moments $\mathrm{E}\big(X^k\big)$ in terms of the hazard rate, see MUTH (1974).

Upon exponentiating (1.9a) turns into

$$S(x) = \exp\left\{ -\int\limits_0^x h(u)\,du \right\}, \tag{1.9b}$$

so that

$$F(x) = 1 - \exp\left\{ -\int\limits_0^x h(u)\,du \right\}. \tag{1.9c}$$

Finally, differentiating (1.9c) yields $f(x)$ in terms of $h(x)$:

$$\begin{aligned} f(x) = \frac{dF(x)}{dx} &= \frac{d\left[1 - \exp\left\{ -\int\limits_0^x h(u)\,du \right\}\right]}{dx} \\ &= h(x)\,\exp\left\{ -\int\limits_0^x h(u)\,du \right\}. \end{aligned} \tag{1.9d}$$

Excursus: Defining distributions by their hazard rate function

Applying (1.9b–d) we want to see what distribution results from four different models of the hazard rate.

1. The **constant hazard rate model**
   From
   $$h(x) = \lambda \;\; \forall\, x \geq 0,\; \lambda > 0,$$
   we find
   $$\begin{aligned} f(x) &= \lambda\,\exp\left\{ -\int\limits_0^x \lambda\,du \right\} = \lambda\,e^{-\lambda\,x}, \\ F(x) &= 1 - \exp\left\{ -\int\limits_0^x \lambda\,du \right\} = 1 - e^{-\lambda\,x}, \\ S(x) &= \exp\left\{ -\int\limits_0^x \lambda\,du \right\} = e^{-\lambda\,x}. \end{aligned}$$
   Thus, the constant hazard rate model gives the **exponential distribution**.
2. The **linear hazard rate model**
   From
   $$h(x) = a + b\,x \;\; \forall\, x \geq 0,\; a \geq 0,\; b > 0,$$
   we find, see Example 1:
   $$\begin{aligned} f(x) &= (a + b\,x)\,\exp\left\{ -a\,x - \frac{b}{2}\,x^2 \right\}, \\ F(x) &= 1 - \exp\left\{ -a\,x - \frac{b}{2}\,x^2 \right\}, \\ S(x) &= \exp\left\{ -a\,x - \frac{b}{2}\,x^2 \right\}. \end{aligned}$$
   For $a = 0$ this is a **RAYLEIGH distribution**.

3. The **power hazard rate model**
   The HR
$$h(x) = c\,x^{c-1} \;\; \forall\, x \geq 0,\; c > 0,$$
   leads to
$$\begin{aligned} f(x) &= c\,x^{c-1}\exp(-x^c)\,, \\ F(x) &= 1 - \exp(-x^c)\,, \\ S(x) &= \exp(-x^c)\,. \end{aligned}$$
   This is the **reduced WEIBULL distribution**, see RINNE (2009).
4. The **exponential hazard rate model**
   Setting
$$h(x) = e^x,\; x \geq 0,$$
   gives
$$\begin{aligned} f(x) &= e^x\,\exp\{-e^x + 1\}\,, \\ F(x) &= 1 - \exp\{-e^x\}\,, \\ S(x) &= \exp\{-e^x\}\,. \end{aligned}$$
   This is recognized as a **GOMPERTZ distribution**.

Based on $h(x) = f(x)\big/S(x)$ it appears that the approximate *unconditional* probability of failure in $(x,\, x + dx]$, $\Pr(x < X \leq x + dx) \approx f(x)\,dx$, is equal to the product of the probability of surviving beyond x and the approximate *conditional* probability of failure in $(x,\, x + dx]$:

$$f(x)\,dx = S(x)\,h(x)\,dx,$$

from which we can define the survival probability as

$$S(x) = \int\limits_x^\infty f(u)\,du = \int\limits_x^\infty S(u)\,h(u)\,du,$$

expressing the survival probability in terms of the *future lifetime*. Looking at $S(x) = \exp\big\{-\int_0^x h(u)\,du\big\}$ in (1.9b) we see the survival probability expressed in terms of the *past lifetime*.

Excursus: Reversed hazard rate

The **reversed (reverse) hazard rate**, also named **retro hazard**, was first mentioned by the name 'dual of the hazard rate' in BARLOW et al. (1963). The name 'reversed hazard rate' was first used by LAGAKOS et al. (1988).[6] It extends the concept of hazard rate to a reverse time direction and is defined as:

$$\begin{aligned} rh(x) &:= \lim_{\Delta \to 0} \frac{\Pr(x - \Delta < X \leq x \,|X \leq x)}{\Delta} \\ &= \frac{f(x)}{F(x)},\;\; F(x) > 0. \end{aligned} \tag{1.10a}$$

(1.10a) can be derived along the lines of (1.6a) – (1.7b) with $A :=$ '$X \leq x$' and '$B := X \leq x - y$', so that $A \cap B =$ '$X \leq x - y$'. From (1.10a) it is seen that $rh(x)$ describes the probability of an *immediate past*

[6] Newer papers on the topic are: CHANDRA/ROY (2001, 2005), GUPTA/NANDA (2001), KUNDU et al. (2009), NANDA/GUPTA (2001, 2004), and SANKARAN et al. (2007).

failure, given that the unit *has* already failed at time x, as opposed to the *immediate future failure*, given that the unit *has not* failed at time x, described by $h(x)$.

$h(x)$ and $rh(x)$ are related as

$$rh(x) = h(x)\,\frac{S(x)}{1-S(x)}, \quad S(x) < 1, \tag{1.10b}$$

$$h(x) = rh(x)\,\frac{F(x)}{1-F(x)}, \quad F(x) < 1. \tag{1.10c}$$

Both rates are equal to one another for $x = x_{0.5}$, otherwise we have

$$h(x) \left\{ \begin{array}{lll} < rh(x) & \text{for} & x < x_{0.5}, \\ > rh(x) & \text{for} & x > x_{0.5}. \end{array} \right\} \tag{1.10d}$$

The reversed hazard rate may be expressed by the PDF, CDF, and CCDF of X as

$$rh(x) = \frac{f(x)}{\int\limits_0^x f(u)\,du} = \frac{F'(x)}{F(x)} = \frac{-S'(x)}{1-S(x)}. \tag{1.10e}$$

It is also possible to express the PDF, CDF, and CCDF in terms of the reversed hazard rate. From (1.10a) we have

$$rh(x) = \frac{d\ln F(x)}{dx}. \tag{1.10f}$$

Integrating (1.10f) yields

$$\begin{aligned} \int\limits_x^\infty rh(u)\,du &= \ln F(x)\Big|_x^\infty \\ &= -\ln F(x), \text{ because } F(\infty) = 1. \end{aligned} \tag{1.10g}$$

Upon exponentiating (1.10g) turns into

$$F(x) = \exp\left\{ -\int\limits_x^\infty rh(u)\,du \right\}, \tag{1.10h}$$

so that

$$S(x) = 1 - \exp\left\{ -\int\limits_x^\infty rh(u)\,du \right\} \tag{1.10i}$$

and

$$\begin{aligned} f(x) = \frac{dF(x)}{dx} &= \frac{d\exp\left\{ -\int\limits_x^\infty rh(u)\,du \right\}}{dx} \\ &= rh(x)\,\exp\left\{ -\int\limits_x^\infty rh(u)\,du \right\}. \end{aligned} \tag{1.10j}$$

In the previous excursion we have seen that the exponential distribution has a constant hazard rate, but the reversed hazard of that distribution is decreasing. From (1.10h–j) we see that we cannot find a distribution defined on $[0,\ \infty)$ having a constant reversed hazard rate, but the **reflected exponential distribution** defined on $(-\infty,\ 0]$ with

$$\begin{aligned} f(x) &= \lambda\,e^{\lambda\,x}; \ x \le 0, \ \lambda > 0; \\ F(x) &= e^{\lambda\,x} \end{aligned}$$

has a constant reversed hazard rate: $rh(x) = \lambda$.

1.1.1.5 The Cumulative Hazard Rate Function

The **cumulative hazard rate** or **integrated hazard rate** CHR is defined as

$$H(x) := \int_0^x h(u)\,du, \tag{1.11a}$$

and satisfies three conditions:

1. $H(0) = 0$,
2. $\lim_{x\to\infty} H(x) = \infty$,
3. $H(x)$ is increasing (= non–decreasing).

From (1.9b) and (1.11a) we easily find

$$S(x) = \exp\big[-H(x)\big] \tag{1.11b}$$

and furthermore

$$F(x) = 1 - \exp\big[-H(x)\big], \tag{1.11c}$$

$$f(x) = -\frac{d\exp\big[-H(x)\big]}{dx}. \tag{1.11d}$$

Vice versa we have the following relations between PDF, CDF, and CCDF on the one hand and CHR on the other hand:

$$\begin{aligned} H(x) &= -\ln\left\{\int_x^\infty f(u)\,du\right\}, &(1.11e)\\ &= -\ln S(x), &(1.11f)\\ &= -\ln[1-F(x)]\,. &(1.11g)\end{aligned}$$

But what is the meaning of $H(x)$? — Whereas $h(x)\,\Delta$ can be given an intuitive interpretation as Pr$(x < X \leq x+\Delta \,|\, X > x)$, $H(x)$ cannot. $H(x)$ is not the sum or the integral of conditional probabilities because the conditioning event changes with x, and there is no law of probability leading to $H(x)$. Thus, $H(x)$ does not have a probabilistic connotation. Yet $H(x)$ plays a key role in reliability and survival analysis, because of the exponentiation formula (1.11b) which says that with $H(x)$ specified we have

$$\Pr(X > x) = e^{-H(x)}, \;\; x \geq 0. \tag{1.12}$$

Excursus: Hazard potential

Based on (1.12) SINGPURWALLA (2006) introduced a new notion, the **hazard potential**. Turning to the right–hand side of (1.12), we note that $e^{-H(x)}$ is the survival function of an exponentially distributed variate, say Z, with scale parameter equal to one, evaluated at $H(x)$, i.e.:

$$\Pr(X > x) = e^{-H(x)} = \Pr\big[Z > H(x)\big]. \tag{1.13}$$

To appreciate the physical connotation of (1.13), we note that because of

$$\Pr(X \leq x) = \Pr\big[Z \leq H(x)\big],$$

we may claim that the time to failure, X, of a unit coincides with the time at which its cumulative hazard $H(x)$ crosses a random threshold Z, where Z has an exponential distribution with scale parameter equal to one, i.e., $X = H^{-1}(Z)$. The random threshold Z, where $Z = H(X)$, is defined as the **hazard potential** of the unit. We may interpret Z as an unknown resource with which the unit is endowed at the time of its inception. With Z considered a resource, $H(x)$ can be interpreted as the amount of resource consumed by time x and the HR, $h(x) = -dH(x)\big/dx$, can be considered the rate at which this resource is consumed. The unit fails when this resource becomes depleted. The term 'potential' refers to a feature parallel to that of **life potential**, see Sect. 1.1.2.6. The difference here is that we are alluding to a unit's resistance to failure rather than its capacity for work.

Another possibility giving insight into (1.12) is the provision of an indifference principle for reliability and survival analysis. Corresponding to every non–negative variate X having an absolutely continuous survival function $S(x) = \Pr(X > x)$, there exists a variate Z taking values $H(x),\ 0 \le H(x) < \infty$, whose survival function is an exponential with scale parameter equal to one. The survival function of X is indexed by $x,\ x \ge 0$, whereas that of Z is indexed by $H(x) = -\int\limits_0^x dS(u)\big/S(u)$.

We finally introduce the notion **hazard quantile**, denoted x_Λ^H and defined by

$$H(x_\Lambda) = \Lambda,\ \ \Lambda \ge 0,\ \ \text{or} \tag{1.14a}$$

$$x_\Lambda^H = H^{-1}(\Lambda),\ \ \Lambda \ge 0. \tag{1.14b}$$

The hazard quantile plays a role in hazard plotting and in designing hazard papers, see Sections 9.2 and 9.3. Based on (11c,g) we see that the ordinary quantile or percentile $x_P,\ 0 \le P \le 1$, and the hazard quantile $x_\Lambda^H,\ \Lambda \ge 0$, are linked as

$$x_P = x^H_{-\ln(1-P)},\ \ \text{and} \tag{1.14c}$$

$$x_\Lambda^H = x_{1-\exp(\Lambda)}. \tag{1.14d}$$

1.1.1.6 Mean Residual Life Function[7]

In Sect. 1.1.1.4 we have introduced the conditional lifetime variate $Y \,|\, X > x := X - x \,|\, X > x$, called **future lifetime** or **remaining lifetime** of an x–survivor. The pertaining PDF reads

$$f(y \,|\, X > x) = \frac{f(x+y)}{S(x)},\ \ y \ge 0.$$

The mean of this variate, denoted $\mu(x)$,[8] as it depends on the age reached, is called **mean residual life** (MRL):

$$\begin{aligned}
\mu(x) &= \mathrm{E}(Y \,|\, X > x) \\
&= \frac{1}{S(x)} \int\limits_0^\infty y\, f(x+y)\, dy \\
&= \frac{1}{S(x)} \int\limits_x^\infty (u-x)\, f(u)\, du,\ \ x + y = u,
\end{aligned}$$

[7] Suggested reading for the section: GUESS/PROSCHAN (1988), MUTH (1980), OAKES/DASU (1990), SWARTZ (1973).

[8] In actuarial science and life tables it is denoted $\overset{o}{e}_x$, see Sect. 6.1.

$$= \frac{1}{S(x)} \left\{ \int_x^\infty u\, f(u)\, du - x \int_x^\infty f(u)\, du \right\}$$

$$= \frac{1}{S(x)} \int_x^\infty u\, f(u)\, du - x. \tag{1.15a}$$

Upon application of Theorem 4 to (1.15a) we find

$$\mu(x) = \frac{1}{S(x)} \int_x^\infty S(u)\, du. \tag{1.15b}$$

For $x = 0$ we have the unconditional mean life (MTTF):

$$\mu(0) = \mu = \mathrm{E}(X). \tag{1.15c}$$

Looking at (1.15b) we see that $\mu(x)$ is the area beneath the survival function to the right of of x divided by the ordinate $S(x)$ at x, corresponding to the fraction surviving x.

The mean residual life $\mu(x)$ must not be confused with the **mean age of an x–survivor**:

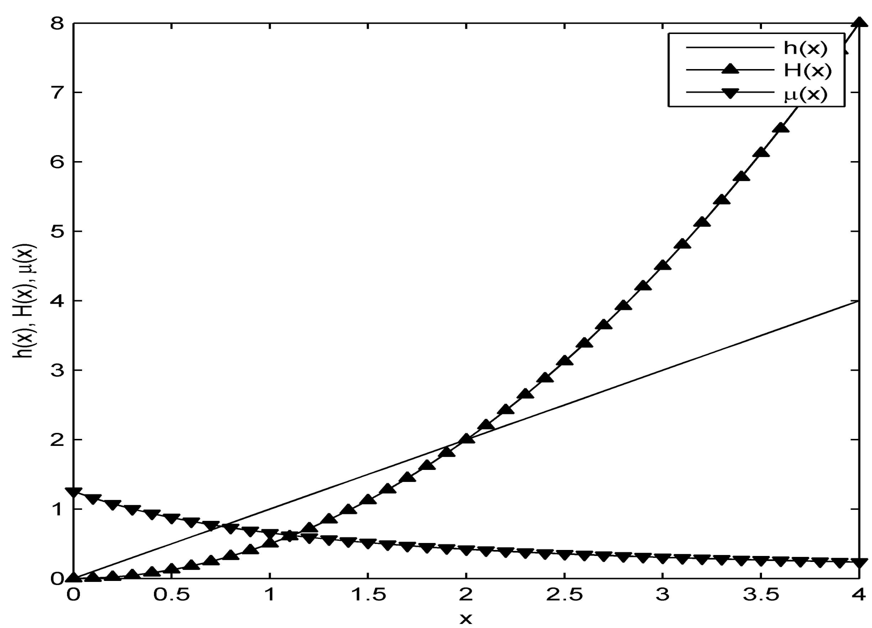
$$\mathrm{E}(X \mid X > x) = \frac{1}{S(x)} \int_x^\infty u\, f(u)\, du. \tag{1.16a}$$

From (1.15a) we see that both means are related as

$$\mu(x) = \mathrm{E}(X \mid X > x) - x. \tag{1.16b}$$

Example 1/2: HR, CHR, and MRL of the linear hazard rate distribution

Figure 1/2: HR, CHR, and MRL of the linear hazard rate distribution with $a = 0$ and $b = 1$



For the linear hazard rate distribution defined in Example 1/1 we have

$$\begin{aligned} h(x) &= a + b\,x, \\ H(x) &= a\,x + \frac{b}{2}\,x^2, \\ \mu(x) &= \frac{\exp\left(\dfrac{a^2}{2\,b}\right)\sqrt{\dfrac{\pi}{2\,b}}\left[1-\operatorname{erf}\left(\sqrt{\dfrac{(a+b\,x)^2}{2\,b}}\right)\right]}{\exp\left\{-a\,x-\dfrac{b}{2}\,x^2\right\}} \end{aligned}$$

The following theorem gives the properties of MRL, see SWARTZ (1973).

Theorem 6: If $\mu(x)$ is the MRL of a survival function $S(x)$ with finite mean $\mathrm{E}(X) = \mu$ then:

1) $\mu(x) \ge 0 \;\; \forall\, x \ge 0,$ (1.17a)

2) $\mu(0) = \mathrm{E}(X),$ (1.17b)

3) if $S(x)$ is absolutely continuous, then $\mu'(x) \ge -1,$ (1.17c)

4) $\displaystyle\int_0^\infty \frac{1}{\mu(x)}\,dx$ diverges, (1.17d)

5) $\displaystyle S(x) = \frac{\mu(0)}{\mu(x)}\,\exp\left\{-\int_0^x \frac{1}{\mu(u)}\,du\right\}.$ ■ (1.17e)

Proof of Theorem 6: It is fairly obvious why property 1) would be necessary since MRL is a conditional expectation of a non–negative variate. Part 1) further follows from (1.15b) because $S(x) \ge 0 \;\; \forall\, x \ge 0$.

2) follows from (1.15b) with (1.5c) observing $S(0) = 1$.

To proof 3) we take a closer look at the derivative of $\mu(x)$, starting with (1.15b):[9]

$$\begin{aligned} \mu'(x) &= \frac{-S^2(x) + f(x)\int_x^\infty S(u)\,du}{S^2(x)} \\ &= h(x)\,\mu(x) - 1. \end{aligned} \tag{1.18}$$

As $h(x)$ and $\mu(x)$ are non–negative we have $\mu'(x) \ge -1$.

For showing 4) and 5) we once more begin with (1.15b), then by simplifying the expression for $-1/\mu(x)$ we find that

$$\begin{aligned} -\frac{1}{\mu(x)} &= -\frac{S(x)}{\int_x^\infty S(u)\,du}, \\ &= \frac{\dfrac{d}{dx}\displaystyle\int_x^\infty S(u)\,du}{\int_x^\infty S(u)\,du}, \end{aligned}$$

[9] By differentiating the numerator and denominator of $\mu(x) = \int_0^x S(u)\,du / S(x)$ it can be shown that $\lim_{x\to\infty}\mu(x) = \lim_{x\to\infty}\left[-\frac{d}{dx}\,\ln f(x)\right]^{-1}$.

$$= \frac{d}{dx}\left[\ln\int\limits_x^\infty S(u)\,du\right]. \tag{1.19a}$$

Integrate each side of (1.19a) between 0 and u to obtain

$$-\int\limits_0^z \frac{1}{\mu(x)}\,dx = \int\limits_0^z d\left[\ln\int\limits_x^\infty S(u)\,du\right] \tag{1.19b}$$

$$= \ln\int\limits_z^\infty S(u)\,du - \ln\int\limits_0^\infty S(u)\,du \tag{1.19c}$$

$$= \ln\int\limits_z^\infty S(u)\,du - \ln\mu(0). \tag{1.19d}$$

The limit of (1.19d) for $z \to \infty$ — after multiplication by (-1) — is

$$\begin{aligned}
\lim_{z\to\infty}\int\limits_0^z \frac{1}{\mu(x)}\,dx &= \lim_{z\to\infty}\left[\ln\mu(0) - \ln\int\limits_z^\infty S(u)\,du\right] \\
&= \ln\mu(0) - \ln\left[\lim_{z\to\infty}\int\limits_z^\infty S(u)\,du\right] \\
&= \ln\mu(0) - \ln 0, \text{ as } S(\infty) = 0 \\
&= \ln\mu(0) + \infty.
\end{aligned}$$

Thus 4) has been proven.

Exponentiating each side of (1.19d) and using (1.15b) gives

$$\exp\left\{-\int\limits_0^x \frac{1}{\mu(u)}\,du\right\} = \frac{\int_x^\infty S(u)\,du}{\mu(0)} = \frac{\mu(x)}{\mu(0)}\,S(x). \tag{1.19e}$$

Finally, 5) follows from cross multiplication of each side of (1.19e) by $\mu(0)/\mu(x)$. ∎

(1.17d) is known as the **inversion formula**[10] which serves as a starting point in expressing the other four representatives of a lifetime distribution in terms of $\mu(x)$:

$$f(x) = \frac{1+\mu'(x)}{\mu^2(x)}\,\mu(0)\,\exp\left\{-\int\limits_0^x \frac{1}{\mu(u)}\,du\right\}, \tag{1.20a}$$

$$= \frac{1+\mu'(x)}{\mu(x)}\,\exp\left\{-\int\limits_0^x \frac{1+\mu'(u)}{\mu(u)}\,du\right\}, \tag{1.20b}$$

$$F(x) = 1 - \frac{\mu(0)}{\mu(x)}\,\exp\left\{-\int\limits_0^x \frac{1}{\mu(u)}\,du\right\}, \tag{1.20c}$$

[10] MEILIJSON (1972) gives another proof of this formula based on the LAPLACE transform.

$$h(x) = \frac{1 + \mu'(x)}{\mu(x)}, \tag{1.20d}$$

$$H(x) = \ln\left\{\frac{\mu(x)}{\mu(0)}\right\} + \int_0^x \frac{1}{\mu(u)}\,du, \tag{1.20e}$$

$$= \int_0^x \frac{1 + \mu'(u)}{\mu(u)}\,du. \tag{1.20f}$$

Example 1/3: Finding PDF, CDF, CCDF, HR, and CHR from a given MRL

What are the five representatives of a lifetime distribution when its MRL is given by

$$\mu(x) = a + b\,x;\quad x \geq 0,\ a > 0,\ b > 0?$$

The resulting distribution may be called **linear mean residual lifetime distribution**.[11] Applying (1.17e) and (1.20a–f) we find after some manipulation:

$$\begin{aligned}
h(x) &= \frac{1+b}{a+b\,x},\\
H(x) &= \frac{1+b}{b}\ln\left(\frac{a+b\,x}{a}\right),\\
S(x) &= \frac{a}{a+b\,x}\left(\frac{a}{a+b\,x}\right)^{1/b},\\
F(x) &= 1 - \frac{a}{a+b\,x}\left(\frac{a}{a+b\,x}\right)^{1/b},\\
f(x) &= \frac{a\,(1+b)\left(\dfrac{a}{a+b\,x}\right)^{1/b}}{(a+b\,x)^2}.
\end{aligned}$$

MRL may also be written in terms of PDF, CDF, CCDF, HR, and CHR:[12]

$$\mu(x) = \frac{\int_0^\infty u\,f(x+u)\,du}{\int_0^\infty f(u)\,du}, \tag{1.21a}$$

$$= \frac{\int_x^\infty u\,f(u)\,du}{\int_x^\infty f(u)\,du} - x, \tag{1.21b}$$

$$\mu(x) = \frac{\int_x^\infty \big[1 - F(u)\big]\,du}{1 - F(x)}, \tag{1.21c}$$

[11] BARLOW/PROSCHAN (1975) have shown that any mixture of exponential distributions yields a distribution with decreasing HR what — see Sect. 2.3 – is equivalent to an increasing MRL. Based on this result MORRISON (1978) proved that when taking the gamma as the mixing distribution the result is a distribution with a linearly increasing MRL which can be identified as the **PARETO distribution of the second kind**.

[12] Note that it is possible for the MRL to exist but for the hazard rate function not to exist and vice versa. If, e.g., we modify the CAUCHY distribution to a **half–CAUCHY distribution** having $f(x) = 2/[\pi\,(1+x^2)],\ x \geq 0$, the MRL does not exist whereas $h(x) = 2/[(1+x^2)\,(\pi - 2\,\arctan x)]$.

$$\mu(x) = \frac{\int_x^\infty S(u)\,du}{S(x)}, \tag{1.21d}$$

$$\mu(x) = \frac{\int_x^\infty \exp\Big\{-\int_0^z h(u)\,du\Big\}\,dz}{\exp\Big\{-\int_0^x h(u)\,du\Big\}}, \tag{1.21e}$$

$$\mu(x) = \frac{\int_x^\infty \exp\{-H(u)\}\;du}{\exp\{-H(x)\}} = \int_0^\infty \exp\big[H(x) - H(u+x)\big]\,du. \tag{1.21f}$$

GUESS/PROSCHAN (1978) stated several bounds for MRL depending on the moments, the CDF and the percentile function of X. GUPTA (1981) showed how to express the moments of X in terms of the mean residual lifetime. He also stated that MRL is the reciprocal of the hazard rate of the asymptotic forward and backward recurrence times of a renewal process. Both recurrence times have the same asymptotic distribution with PDF

$$f^*(x) = \frac{1 - F(x)}{\mathrm{E}(X)} \tag{1.22a}$$

and CCDF

$$S^*(x) = \frac{1}{\mathrm{E}(X)} \int_x^\infty S(u)\,du. \tag{1.22b}$$

The corresponding HR is

$$h^*(x) = \frac{f^*(x)}{S^*(x)}$$

which upon inserting (1.22a,b) and taking the reciprocal gives

$$\frac{S^*(x)}{f^*(x)} = \frac{1}{S(x)} \int_x^{[\infty} S(u)\,du = \mu(x). \tag{1.22c}$$

Two survival functions $S_0(x)$ and $S_1(x)$ are said to have **proportional mean residual life** if

$$\mu_1(x) = \Theta\,\mu_0(x)\ \ \forall\, x \ge 0\ \text{ and }\ \Theta > 0, \tag{1.23a}$$

where $\mu_0(x)$ and $\mu_1(x)$ are the respective mean residual lives at time x. It can be shown that if $S_0(x)$ and $S_1(x)$ have proportional mean residual life, then

$$S_1(x) = S_0(x) \left[\int_x^\infty \frac{S_0(u)\,du}{\mu_0(0)}\right]^{1/\Theta - 1}. \tag{1.23b}$$

The hazard rate and the mean residual life are conditional concepts, both are conditioned on survival to time x. An essential difference between HR and MRL is that the former accounts only for the immediate future in assessing the event ‘unit failure’, whereas the latter accounts for the whole future. This is readily seen if we multiply both $h(x)$ and $\mu(x)$ by $S(x)$:

$$h(x)\,S(x) = -\frac{dS(x)}{dx}, \tag{1.24a}$$

$$\mu(x)\,S(x) = \int_x^\infty S(u)\,du. \tag{1.24b}$$

The right–hand side of (1.24a) depends on the probability law at the point x only, whereas the right–hand side of (1.24b) depends on the probability law of X at all points in (x, ∞). This intuition explains the difference between the two. Both, MRL and HR are needed in practice. In theory we define classes of distributions depending on the behavior of MRL and HR, see Chapter 2. The MRL function has a tremendous range of applications. For example, WATSON/WELLS (1961) use MRL in studying burn–in. Actuaries apply MRL to setting rates of benefits for life insurance. Distributions with increasing MRL have been found useful as models in the social science for the duration of wars and strikes or of jobs, a phenomenon called 'inertia'.

In Tab. 1/1 we have summarized the most important relationships between the six representatives of the lifetime distribution scattered in this and the preceding sections. The table has been arranged as an input–output table showing how to switch over from one representative to another one.

1.1.2 The Hazard Rate for Special Cases

Assuming a continuous variate we will give results on the behavior of the hazard rate when we apply special operations to the variate and its distribution. The results of this section may be generalized to the discrete and the multivariate cases, see Sect. 1.2 and 1.3.

1.1.2.1 Transformation of Random Variables

Suppose we have a continuous random variable X with known representatives of its distribution, and we consider a new random variable Y which is some function of X, i.e., let

$$y = g(x) \tag{1.25a}$$

be a function of x such that its inverse

$$x = g^{-1}(y) \tag{1.25b}$$

exists. When seeking the representatives of the Y–distribution in terms of those of the X–distribution we have to distinguish between two cases.

In the first case let $y = g(x)$ be a *strictly increasing function*. Then, if X is less than or equal to x it follows that Y is less than or equal to the unique value of y that corresponds to the given value of x. Thus, if $X \leq x$, then $Y \leq g(x)$. Conversely, if $Y \leq y$, then $X \leq g^{-1}(y)$, and the probabilities of these events are equal, i.e.,

$$\begin{aligned} \Pr(Y \leq y) &= \Pr\big[X \leq g^{-1}(y)\big] \\ F(y) &= F\big[g^{-1}(y)\big]. \end{aligned} \tag{1.26a}$$

(1.26a) can be confusing since the CDFs on opposite sides of the equation are not the same functions. The one on the left–hand side is the CDF of Y, whereas the one on the right–hand side is for the random variable X. To clarify this, we write (1.26a) as

$$F_Y(y) = F_X\big[g^{-1}(y)\big]. \tag{1.26b}$$

From (1.26b), which relates the CDFs of X and Y, we can derive relationships for the PDFs, CCDFs, HRs, and CHRs as well.[13] Since the CCDF is the complement of the CDF it follows from (1.26b) that $1 - S_Y(y) = 1 - S_X\big[g^{-1}(y)\big]$ or that

$$S_Y(y) = S_X\big[g^{-1}(y)\big]. \tag{1.26c}$$

[13] Generally, the MRL of Y cannot be given easily and as an exact function of the X–MRL, see the excursus at the end of this section.

Table 1/1: Relations among the six functions describing a continuously distributed stochastic lifetime

from \ to	$f(x)$	$F(x)$	$S(x)$	$h(x)$	$H(x)$	$\mu(x)$
$f(x)$	–	$\int_0^x f(u)\,du$	$\int_x^\infty f(u)\,du$	$\dfrac{f(x)}{\int_x^\infty f(u)\,du}$	$-\ln\left\{\int_x^\infty f(u)\,du\right\}$	$\dfrac{\int_0^\infty u\,f(x+u)\,du}{\int_x^\infty f(u)\,du}$
$F(x)$	$F'(x)$	–	$1-F(x)$	$\dfrac{F'(x)}{1-F(x)}$	$-\ln\{1-F(x)\}$	$\dfrac{\int_x^\infty [1-F(u)]\;du}{1-F(x)}$
$S(x)$	$-R'(x)$	$1-S(x)$	–	$\dfrac{-S'(x)}{S(x)}$	$-\ln[S(x)]$	$\dfrac{\int_x^\infty S(u)\,du}{S(x)}$
$h(x)$	$h(x)\,\exp\left\{-\int_0^x h(u)\,du\right\}$	$1-\exp\left\{-\int_0^x h(u)\,du\right\}$	$\exp\left\{-\int_0^x h(u)\,du\right\}$	–	$\int_0^x h(u)\,du$	$\dfrac{\int_x^\infty \exp\left\{-\int_0^u h(v)\,dv\right\}du}{\exp\left\{-\int_0^x h(u)\,du\right\}}$
$H(x)$	$-\dfrac{d\,\{\exp[-H(x)]\}}{dx}$	$1-\exp\{-H(x)\}$	$\exp\{-H(x)\}$	$H'(x)$	–	$\dfrac{\int_x^\infty \exp\{-H(u)\}\,du}{\exp\{-H(x)\}}$
$\mu(x)$	$\dfrac{1+\mu'(x)}{\mu^2(x)}\times\mu(0)\times$ $\times\,\exp\left\{-\int_0^x \dfrac{1}{\mu(u)}\,du\right\}$	$1-\dfrac{\mu(0)}{\mu(x)}\times$ $\times\,\exp\left\{-\int_0^x \dfrac{1}{\mu(u)}\,du\right\}$	$\dfrac{\mu(0)}{\mu(x)}\times$ $\times\,\exp\left\{-\int_0^x \dfrac{1}{\mu(u)}\,du\right\}$	$\dfrac{1}{\mu(x)}\,\{1+\mu'(x)\}$	$\ln\left\{\dfrac{\mu(x)}{\mu(0)}\right\}+$ $+\int_0^x \dfrac{1}{\mu(u)}\,du$	–

Next, the PDF is the derivative of the CDF, so we differentiate both side of (1.26b) with respect to y, obtaining

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\,F_Y(y) \\ &= \frac{d}{dy}\,F_X\big[g^{-1}(y)\big] \\ &= f_X\big[g^{-1}(y)\big]\,\frac{d}{dy}g^{-1}(y), \end{aligned} \tag{1.26d}$$

using the chain rule of differentiation. Since $g^{-1}(y)$ is simply x we can simply write

$$f_Y(y) = f_X\big[g^{-1}(y)\big]\,\frac{dx}{dy}. \tag{1.26e}$$

Furthermore, HR is the ratio of the PDF and the CCDF. Thus,

$$\begin{aligned} h_Y(y) &= \frac{f_Y(y)}{S_Y(y)} \\ &= \frac{f_X\big[g^{-1}(y)\big]\,\dfrac{dx}{dy}}{S_X\big[g^{-1}(y)\big]} \\ &= h_X\big[g^{-1}(y)\big]\,\frac{dx}{dy}. \end{aligned} \tag{1.26f}$$

Finally, the CHR is the negative of the $\ln$–transformed CCDF, see (1.11f). So we have

$$\begin{aligned} H_Y(y) &= -\ln S_Y(y) \\ &= -\ln S_X\big[g^{-1}(y)\big] \\ &= H_X\big[g^{-1}(y)\big]. \end{aligned} \tag{1.26g}$$

Example 1/4: Increasing transformation of the reduced exponentially distributed variate

The reduced exponential distribution has

$$f_X(x) = e^{-x},\;\; F_X(x) = 1 - e^{-x},\;\; S_X(x) = e^{-x},\;\; h_X(x) = 1,\;\; H_X(x) = x;\;\; x \ge 0.$$

Let $y = g(x) = x^2$. We first have

$$x = g^{-1}(y) = y^{1/2} \;\text{ and }\; \frac{dx}{dy} = \frac{1}{2}\,y^{-1/2}.$$

The resulting representatives of the Y–distribution follow as

$$\begin{aligned} F_Y(y) &= 1 - e^{-y^{1/2}}, \\ S_Y(y) &= e^{-y^{1/2}}, \\ f_Y(y) &= \frac{1}{2}\,y^{-1/2}\,e^{-y^{1/2}}, \\ h_Y(y) &= \frac{1}{2}\,y^{-1/2}, \\ H_Y(y) &= y^{1/2}. \end{aligned}$$

For MRL we find $\mu_X(x) = 1,\;\; \mu_Y(y) = 2\left(1+y^{1/2}\right)$, so $\mu_X(x)$ is constant, whereas $\mu_Y(y)$ is increasing. We see that Y has a reduced WEIBULL distribution with shape parameter equal to $1/2$.

In the second case $y = g(x)$ is a *strictly decreasing function* and the reasoning and the results change a bit. In this case we see that if X is less than x, then Y will be greater than the value of y which corresponds to the given value of x, conversely, if $Y > y$, then $X < g^{-1}(y)$. In terms of probabilities we have

$$\begin{aligned} \Pr(Y > y) &= \Pr\big[X < g^{-1}(y)\big] = \Pr\big[X \leq g^{-1}(y)\big] \text{ or} \\ S_Y(y) &= F_X\big[g^{-1}(y)\big] = 1 - S_X\big[g^{-1}(y)\big]. \end{aligned} \tag{1.27a}$$

Then

$$F_Y(y) = 1 - S_Y(y) = 1 - F_X\big[g^{-1}(y)\big] = S_X\big[g^{-1}(y)\big] \tag{1.27b}$$

and

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\,F_Y(y) = -\frac{d}{dy}\,F_X\big[g^{-1}(y)\big] \\ &= -f_X\big[g^{-1}(y)\big]\,\frac{dx}{dy} \end{aligned} \tag{1.27c}$$

by the chain rule. Since $x = g^{-1}(y)$ is a decreasing function the derivative dx/dy in (1.27c) will be negative and the PDF $f_Y(y)$ will be positive, as required. The HR of Y is

$$\begin{aligned} h_Y(y) &= \frac{f_Y(y)}{S_Y(y)} = -\frac{f_X\big[g^{-1}(y)\big]\,\dfrac{dx}{dy}}{1 - S_X\big[g^{-1}(y)\big]} \\ &= -h_X\big[g^{-1}(y)\big]\,\frac{S_X\big[g^{-1}(y)\big]}{1 - S_X\big[g^{-1}(y)\big]}\,\frac{dx}{dy}. \end{aligned} \tag{1.27d}$$

In general, the CHR of Y cannot be written in terms of the X–CHR, but in can be expressed in terms of the X–CDF as

$$H_Y(y) = -\ln S_Y(y) = -\ln F_X\big[g^{-1}(y)\big]. \tag{1.27e}$$

Example 1/5: Decreasing transformation of the reduced exponentially distributed variate

We take the transformation $y = g(x) = x^{-1}$ and have

$$x = g^{-1}(y) = y^{-1},\ \ \frac{dx}{dy} = -y^{-2}.$$

From the representatives of the X–distribution in Example 1/4 and using (1.27a–e) we find

$$\begin{aligned} S_Y(y) &= 1 - e^{-y^{-1}}, \\ F_Y(y) &= e^{-y^{-1}}, \\ f_Y(y) &= y^{-2}\,e^{-y^{-1}}, \\ h_Y(y) &= y^{-2}\,\frac{e^{-y^{-1}}}{1 - e^{-y^{-1}}} = \frac{y^{-2}}{e^{-y^{-1}} - 1}, \\ H_Y(y) &= -\ln\Big[1 - e^{-y^{-1}}\Big]. \end{aligned}$$

The distribution of Y is recognized as the type–II maximum extreme value distribution, also known as **inverse WEIBULL distribution**.

We now explore two special transformations. The first one is the *linear transformation:*

$$y = g(x) = a + b\,x, \quad b \neq 0. \tag{1.28a}$$

If the transformation is *increasing* $(b > 0)$ we have from (1.26f,g) with $x = g^{-1}(y) = (y-a)\big/b$ and $dx\big/dy = b^{-1}$:

$$h_Y(y) = \frac{1}{b}\,h_X\left(\frac{y-a}{b}\right), \tag{1.28b}$$

i.e., the HR at a given value of the new variable is b^{-1} times the hazard at the value of the original variable corresponding to the given value of the new variable, and

$$H_Y(y) = H_X\left(\frac{y-a}{b}\right), \tag{1.28c}$$

i.e., the CHR of Y is simply the CHR of X evaluated at the retransformed y–value. For a *decreasing* linear transformation, i.e., for $y = a + b\,c,\ b < 0$, we have from (1.27d,e):

$$h_Y(y) = \frac{1}{b}\,h_X\left(\frac{y-a}{b}\right)\frac{S_X\left(\dfrac{y-a}{b}\right)}{F_X\left(\dfrac{y-a}{b}\right)}, \tag{1.28d}$$

$$H_Y(y) = -\ln F_X\left(\frac{y-a}{b}\right). \tag{1.28e}$$

The second special case is the **probability integral transformation**:

$$y = g(x) = F_X(x), \tag{1.29a}$$

where $F_X(x)$ is the increasing CDF of X. We note that $x = g^{-1}(y) = F_X^{-1}(y),\ 0 \leq y \leq 1$, i.e., x is given by the percentile function of X. From (1.26b) we have

$$F_Y(y) = F_X\big[g^{-1}(y)\big] = F_X\Big[F_X^{-1}(y)\Big] = y, \tag{1.29b}$$

therefore

$$f_Y(y) = \frac{d}{dy}\,F_Y(y) = 1. \tag{1.29c}$$

Thus, Y has the **reduced uniform distribution** with

$$f_Y(y) = 1 \ \text{ for } \ 0 \leq y \leq 1. \tag{1.29d}$$

The other representatives of the Y–distribution are:

$$S_Y(y) = 1 - y, \tag{1.29e}$$

$$h_Y(y) = \frac{1}{1-y}, \tag{1.29f}$$

$$H_Y(y) = -\ln(1-y). \tag{1.29g}$$

Excursus: Moments of transformed variates

With the only exception of the linear transformation the moments of a transformed variate cannot be given as exact functions of the moments of the original variate. The following approximations are based on the **delta method** (method of statistical differentials). For the mean of $Y = g(X)$ we have

$$\mu_Y = \mathrm{E}(Y) \approx g\big[\mathrm{E}(X)\big] + \frac{\mathrm{Var}(X)}{2}\,\frac{d^2 g(x)}{dx^2}\bigg|_{x=\mathrm{E}(X)} \tag{1.30a}$$

and for the variance of Y

$$\sigma_Y^2 = \text{Var}(y) \approx \text{Var}(X) \left[\frac{dg(x)}{dx}\Big|_{x=\text{E}(X)} \right]^2. \tag{1.30b}$$

For the linear transformation $Y = a + b\,x$ we have exact relationships:

$$\text{E}(a + b\,X) = a + b\,\text{E}(X), \tag{1.30c}$$
$$\text{Var}(a + b\,X) = b^2\,\text{Var}(X). \tag{1.30d}$$

With $b > 0$ in $Y = a + b\,X$ the MRLs of X and Y are related as

$$\mu_Y(y) = a + b\,\mu_X\left(\frac{y-a}{b}\right). \tag{1.30e}$$

1.1.2.2 Mixing and Compounding[14]

In some situations, units may not come from a homogeneous population. A demographer who is to construct a nation's life table might encounter several ethnic groups having different patterns of mortality. A reliability engineer, for instance, might have a component that has been manufactured in one of two facilities, but is not certain which one the unit comes from. In **finite mixture models**, a unit is assumed to be from one of m populations. The case $m = \infty$ is called **countable mixture**. When there is a single population that is mixed by a continuous parameter Θ (for example, the amount of impurities present in a raw material or the temperature of solder applied in a circuit board), a stochastic parameter model (= **continuous mixture model**) is appropriate.

Suppose that $F(x \,|\, \theta)$ represents the lifetime CDF given that $\Theta = \theta$ and that $G(\theta)$ represents the CDF of the random parameter Θ. The function $F(x)$, defined by

$$F(x) = \int\limits_{\text{all }\theta} F(x \,|\, \theta\, dG(\theta), \tag{1.31a}$$

which is the marginal CDF of X, is called **compound distribution** of $F(\cdot)$ and $G(\cdot)$. $F(x \,|\, \theta)$ is known as the **kernel** and $G(\cdot)$ is the **mixing (or compounding) distribution**. If the entire mass of the corresponding measure of $G(\cdot)$ is confined to a countable number of points $\theta_1,\,\theta_2,\ldots$ and the masses at $\theta_j;\,j = 1,\,2,\ldots;$ are $G(\theta_j)$, then (1.31a) takes the form

$$F(X) = \sum_{j=1}^{\infty} F(x \,|\, \theta_j)\,G(\theta_j), \tag{1.31b}$$

which is a countable mixture CDF.[15] If the entire mass of the corresponding measure $G(\cdot)$ is confined to only a finite number of finite points $\theta_1,\,\theta_2,\ldots,\theta_m$, then (1.31a) becomes a finite mixture of m components whose CDF is given by

$$F(x) = \sum_{j=1}^{m} F(x \,|\, \theta_j)\,G(\theta_j). \tag{1.31c}$$

To simplify notation in (1.31b,c), we write

$$p_j := G(\theta_j) \;\text{ and }\; F_j(x := F(x \,|\, \theta_j),$$

[14] Suggested reading for this section: AL–HUSSAINI/SULTAN (2001).

[15] For example, the non–central χ^2–distribution is a countable mixture of **POISSON** and **χ^2–distributions**.

so that (1.31b) turns into

$$F(x) = \sum_{j=1}^{\infty} F_j(x)\, p_j, \tag{1.32a}$$

where $p_j \geq 0 \;\forall\, j$ and $\sum\limits_{j=1}^{\infty} p_j = 1$. Also, (1.31c) becomes

$$F(x) = \sum_{j=1}^{m} F_j(x)\, p_j, \tag{1.32b}$$

where $p_j \geq 0 \;\forall\, j$ and $\sum\limits_{j=1}^{m} p_j = 1$. In (1.32a,b) p_j is known as the j–th **mixing proportion** and $F_j(x)$ as the j–th **component** in the mixture. It may be noticed that the choice $F_j(x) := F(x\,|\,\theta_j)$ restricts the CDF $F(x\,|\,\theta_j)$ for all values of j to belong to the same family of distributions. However, formulas (1.32a,b) are written in the most general forms on which each of the CDFs $F_j(x)$ could belong to a distinct family. The only requirement here is that, for any j, $F_j(x)$ is a CDF.

If, in (1.31a), $G(\theta)$ is absolutely continuous, a PDF $g(\theta)$ exists such that $g(\theta) = G'(\theta)$ and if $f(x)$ and $f(x\,|\,\theta)$ are the PDFs corresponding to the CDFs $F(x)$ and $F(x\,|\,\theta)$, than from (1.31a) we have

$$f(x) = \int\limits_{\text{all } \theta} f(x\,|\,\theta)\, g(\theta)\, d\theta. \tag{1.33a}$$

Similarly, the PDFs corresponding to (1.32a,b) are given by

$$f(x) \;=\; \sum_{j=1}^{m} f_j(x)\, p_j \;\text{ and} \tag{1.33b}$$

$$f(x) \;=\; \sum_{j=1}^{\infty} f_j(x)\, p_j, \tag{1.33c}$$

where $f_j(x)$ is the j–th component density function corresponding to $F_j(x)$.

Having found the mixed CDF and PDF by one of the foregoing formulas, we can find the HR, CHR, and MRL, respectively, of the mixed population by applying (1.7b), (1.11f), and (1.15c), respectively.

Example 1/6: Continuous mixture of exponential distributions[16]

Let the lifetime X have an exponential distribution with a positive parameter (= scaling factor) θ :

$$f(x\,|\,\theta) = \theta\, e^{-\theta\, x}, \;\; x \geq 0.$$

Now suppose that $\theta > 0$ is a realization of a random variable Θ which also has an exponential distribution, but with parameter λ :

$$g(\theta) = \lambda\, e^{-\lambda\, \theta}, \;\; \lambda > 0.$$

In this case the compound distribution of X is found using the integration by parts:

$$\begin{aligned} f(x) &= \int\limits_0^{\infty} \theta\, e^{-\theta\, x}\, \lambda\, e^{-\lambda\, \theta}\, d\theta \\ &= \lambda \int\limits_0^{\infty} \theta\, e^{-\theta\,(x+\lambda)}\, d\theta \\ &= \left[-\frac{\lambda\, e^{-\theta\,(x+\lambda)}}{x+\lambda} + \frac{\lambda}{x+\lambda} \int e^{-\theta\,(x+\lambda)}\, d\theta \right]_0^{\infty} \end{aligned}$$

$$
\begin{aligned}
&= \left[-\frac{\lambda\,\theta\,e^{-\theta\,(x+\lambda)}}{x+\lambda} - \frac{\lambda}{(x+\lambda)^2}\,e^{-\theta\,(x+\lambda)}\right]_0^\infty \\
&= \frac{\lambda}{(x+\lambda)^2},\;\; x \geq 0.
\end{aligned}
$$

This is recognized as a special case of the **log–logistic distribution**, see Sect. 3.1. The corresponding CDF and CCDF are

$$
\begin{aligned}
F(x) &= 1 - \frac{\lambda}{x+\lambda}, \\
S(x) &= \frac{\lambda}{x+\lambda}.
\end{aligned}
$$

So the HR and CHR follow as

$$
\begin{aligned}
h(x) &= \frac{1}{x+\lambda}, \\
H(x) &= \ln\left[\frac{x+\lambda}{\lambda}\right].
\end{aligned}
$$

A MRL $\mu(x)$ does not exist. We notice that, while the HR of each $f(x\,|\,\theta)$ is a constant, namely $h(x\,|\,\theta) = \theta$, the HR of the mixture is decreasing. This result holds for finite mixtures of exponential distributions, see below.

It can be shown that in the case of a finite mixture the HR and the MRL of the compound distribution may be written in terms of the HRs $h_j(x)$ and of the MRLs $\mu_j(x)$ of the m mixed distributions:

$$
h(x) = \frac{\sum\limits_{j=1}^{m} p_j\,S_j(x)\,h_j(x)}{\sum\limits_{j=1}^{m} p_j\,S_j(x)}, \tag{1.34a}
$$

$$
\mu(x) = \frac{\sum\limits_{j=1}^{m} p_j\,S_j(x)\,\mu_j(x)}{\sum\limits_{j=1}^{m} p_j\,S_j(x)}. \tag{1.34b}
$$

Thus, the HR and MRL of the mixed model may be considered as a weighted average of the HRs and MRLs of the individual populations, the weights being $p_j\,S_j(x)$. One interesting property of a *mixed exponential model* is that it has a decreasing HR. Suppose, $X_j;\ j = 1, 2, \ldots, m;$ are exponentially distributed with scale parameter $b_j,\ b_j > 0$, respectively, then

$$
h(x) = \frac{f(x)}{S(x)} = \frac{\sum\limits_{j=1}^{m} p_j\,\frac{1}{b_j}\,\exp\Big(-\frac{x}{b_j}\Big)}{\sum\limits_{j=1}^{m} p_j\,\exp\Big(-\frac{x}{b_j}\Big)},\;\; x \geq 0.
$$

It can be shown that this HR is a decreasing function, decreasing from the average of the failure rates, $\sum_{j=1}^{m} p_j\,\frac{1}{b_j}$, at $x = 0$, to the minimum of the failure rates, $1\big/\max(b_j)$, as $x \to \infty$. This suggests one possible justification for a decreasing HR model.

1.1.2.3 Formation of Systems[17]

Up to now we have considered modeling lifetime of single units, components, people etc. by using hazards and its relatives. However, it is especially true in the engineering sciences that pieces of

[17] Suggested reading for this section: BARLOW/PROSCHAN (1975), CROWDER et al. (1991, Chapter 9), LEEMIS (1995, Chapter 2), MEEKER/ESCOBAR (1998, Chapter 15), SMITH (2002, Chapter 3).

equipment consist of many — possibly different — interacting components. The term ‘reliability’ is commonly used to describe the ‘survival’ of such components and of such a system. Essentially, the reliability of a component is the probability that it is operational. The primary concern of engineers when looking at a system of components is its reliability and how the reliabilities of individual components affect the reliability of the entire system. Once the the system reliability has been found we can calculate the system hazard applying (1.8c). We will only give a short introduction into the theory of reliability of systems; more details may be found in the suggested readings.

It is certainly true that the reliability of components may change with time. However, initially we make the assumption that at some instant in time we are able to observe the components and know whether they are functioning or not. Let $C_i; i = 1, 2, \ldots, m$; denote component i and suppose that each component has one of two operational states: ‘functioning’ and ‘not functioning’. For each i the indicator z_i associated with C_i is defined by

$$z_i = \left\{ \begin{array}{l} 1 \text{ if } C_i \text{ is functioning} \\ 0 \text{ if } C_i \text{ is not functioning} \end{array} \right\}; \quad i = 1, 2, \ldots, m. \tag{1.35a}$$

A **structure function** is a useful tool in describing the way m components are related to form a system. The structure function defines the system state as a function of the component states and is given by

$$\phi(z_1, \ldots, z_m) = \left\{ \begin{array}{l} 1 \text{ if the system is functioning} \\ 0 \text{ if is not functioning.} \end{array} \right\} \tag{1.35b}$$

Since there are m components there are 2^m different values that the **system state vector**

$$\boldsymbol{z} = (z_1, z_2, \ldots, z_m)$$

can assume and $\binom{m}{j}$ of these vectors correspond to exactly j functioning components; $j = 0, 1, \ldots, m$. The structure function $\phi(\boldsymbol{z})$, maps the system state vector $\boldsymbol{z}$ to 0 or 1, yielding the state of the system. The most common system structures are the series and parallel systems and most other complicated structures can be reduced to these two types.

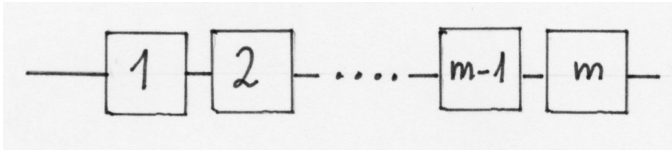
A **series system** functions when all its components function. Thus $\phi(\boldsymbol{z})$ assumes the value 1 when $z_1 = z_2 = \ldots = z_m = 1$, and 0 otherwise. Therefore, its structure function $\phi_S(\boldsymbol{z})$ is given by

$$\begin{aligned} \phi_S(\boldsymbol{z}) &= \left\{ \begin{array}{l} 0 \text{ if there exists an } i \text{ such that } z_i = 0 \\ 1 \text{ if } z_i = 1 \,\forall\, i, \end{array} \right\} && (1.36a) \\ &= \min(z_1, z_2, \ldots, z_m), && (1.36b) \\ &= \prod_{i=1}^{m} z_i. && (1.36c) \end{aligned}$$

These three different ways of expressing the value of the structure function are equivalent, although (1.36c) is preferred because of its compactness. The **block–diagram** in Fig. 1/3 visualizes a series system of m components.[18] Systems that function only when all their components function should be modeled as series systems.

[18] A block–diagram is a graphic device for expressing the arrangement of the components to form a system. If a path can be traced through functioning components from left to right on a block–diagram, then the system functions. The boxes represent the components, and either component numbers i or probabilities P_i are placed inside the boxes.

Figure 1/3: Series system block–diagram



A **parallel system** functions when one or more of its components function. Its structure function $\phi_P(\boldsymbol{z})$ assumes the value 0 when $z_1 = z_2 = \ldots = 0$, and 1 otherwise. Therefore,

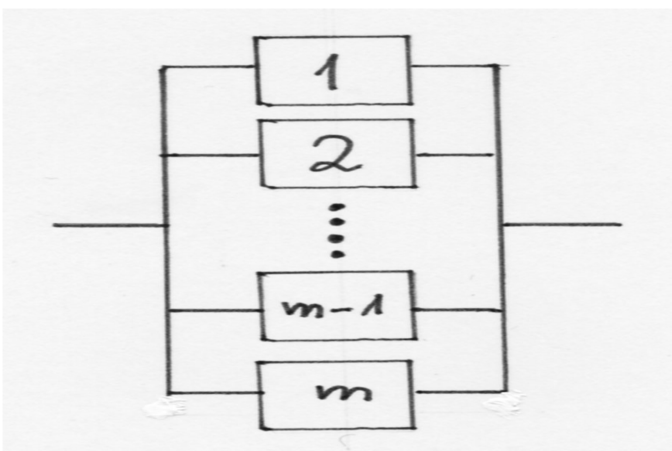
$$\phi_P(\boldsymbol{z}) = \left\{ \begin{array}{l} 0 \text{ if } z_i = 0 \,\forall\, i, \\ 1 \text{ if there exists an } i \text{ such that } z_i = 1 \end{array} \right\} \tag{1.37a}$$

$$= \max(z_1,\, z_2, \ldots, z_m), \tag{1.37b}$$

$$= 1 - \prod_{i=1}^{m} (1 - z_i). \tag{1.37c}$$

See Fig. 1/4 for a block–diagram of a parallel arrangement of m components. Such an arrangement is appropriate when all components must fail for the system to fail.

Figure 1/4: Parallel system block–diagram



To avoid studying structure functions that are unreasonable, a subset of all possible system of m components, that is, **coherent systems**, has been defined. A system is coherent if

1. its **structure function** is **non–decreasing** in $\boldsymbol{z}$, i.e.,

$$\phi(z_1, \ldots, z_{i-1},\, 0,\, z_{i+1}, \ldots, z_m) \leq \phi(z_1, \ldots, z_{i-1},\, 1,\, z_{i+1}, \ldots, z_m) \tag{1.38a}$$

and

2. there are no **irrelevant components**, i.e., component C_i is irrelevant if, for all states of the other components in the system (that is, for all values of z_j for $j \neq i$)

$$\phi(z_1, \ldots, z_{i-1},\, 1,\, z_{i+1}, \ldots, z_m) = \phi(z_1, \ldots, z_{i-1},\, 0,\, z_{i+1}, \ldots, z_m). \tag{1.38b}$$

The structure function of a coherent system may be quite difficult to describe in simple terms.[19] However, it can be shown that the structure function of any coherent system is bounded above and

[19] Some techniques in this context are the formation of path vectors, minimal path vector, cut vectors, and minimal cut vector.

below by the structure functions of parallel and series systems what inevitably leads to bounds on the reliability of coherent systems, see (1.44c).

Theorem 7: If $\phi_C(\boldsymbol{z})$ is the structure function of a coherent system of m components in the state vector $\boldsymbol{z} = (z_1,\ z_2, \ldots, z_m)$, then

$$\phi_S(\boldsymbol{z}) = \prod_{i=1}^{m} z_i \leq \phi_C(\boldsymbol{z}) \leq \phi_P(\boldsymbol{z}) = 1 - \prod_{i=1}^{m}(1 - z_i). \qquad \blacksquare \tag{1.39}$$

Series and parallel systems are coherent systems. Before showing how the structure function is related to the system reliability we present some other types of coherent systems.

1. *Systems with components in series–parallel*

   Methods for evaluating the reliability of structures with components in both series and parallel provide the basis for evaluating more complicated structures. There are two types of simple (rectangular) series–parallel structures.

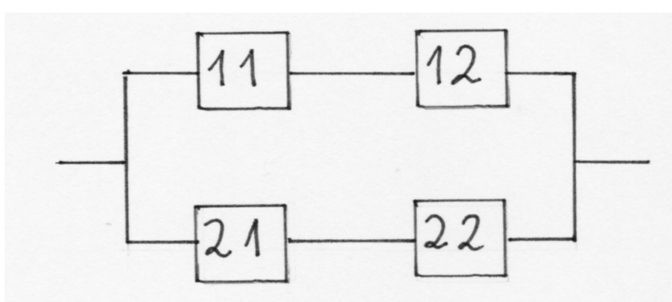
   1.1 *Series–parallel system structure with system–level redundancy*

   In some applications it is more cost effective to achieve higher reliability by using two or more copies of a series system rather than having to improve the reliability of the single system itself. This idea leads to a $r \times k$ series–parallel system–level redundancy structure having r parallel sets, each of k components in series. The structure function reads

   $$\phi(\boldsymbol{z}) = 1 - \prod_{i=1}^{r}\left[1 - \prod_{j=1}^{k} z_{ij}\right], \tag{1.40a}$$

   if the components are independent. Fig. 1/5 shows such a 2×2 structure.

   Figure 1/5: Block–diagram of a 2×2 series–parallel system with system–level redundancy

   For the system depicted in Fig. 1/5 the structure function is

   $$\phi(\boldsymbol{z}) = 1 - \big[1 - z_{11}\, z_{12}\big]\,\big[1 - z_{21}\, z_{22}\big],$$

   and there are — out of 16 possible vectors — seven system state vectors leading to $\phi(\boldsymbol{z}) = 1$.

   1.2 *Series–parallel system structure with component–level redundancy*

   Component redundancy is an important method for improving system reliability. A $r \times k$ component–level redundant structure has k series structures, each one made of r components in parallel. If it is necessary to have only one path through the system such a structure is, for a given number of identical components, more reliable than the series–parallel system with system–level redundancy. The structure function reads

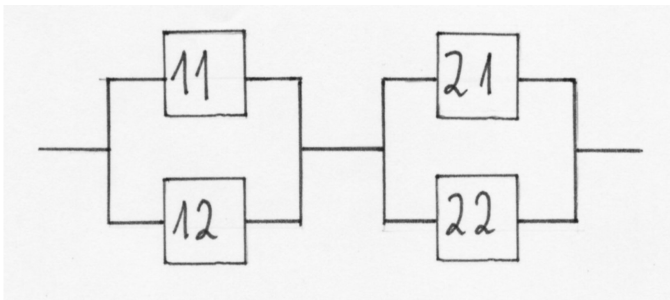
   $$\phi(\boldsymbol{z}) = \prod_{i=1}^{k}\left[1 - \prod_{j=1}^{r} z_{ij}\right], \tag{1.40b}$$

if the components are independent. The structure function is the product of the structure functions of the k parallel subsystems each consisting of r components. Fig. 1/6 shows such a 2×2 structure where (1.40b) results into

$$\phi(\boldsymbol{z}) = \big[1 - (1 - z_{11})\,(1 - z_{12})\big]\,\big[1 - (1 - z_{21})\,(1 - z_{22})\big],$$

and there are nine system state vectors — out of 16 possible vectors — leading to $\phi(\boldsymbol{z}) = 1$.

Figure 1/6: Block–diagram of a 2×2 series–parallel system with component–level redundancy



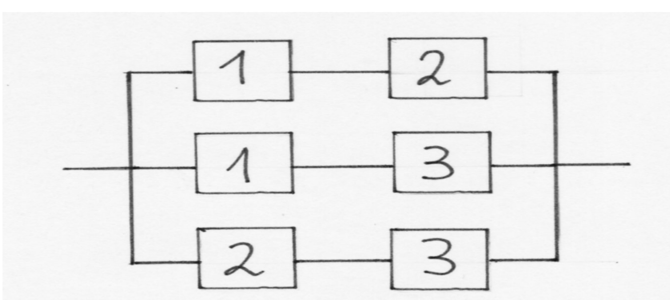
2. *k–out–of–m system*

Another way of increasing system reliability consists in supplying more components than are necessary for functioning. By a k–out–of–m system $(k \leq m)$ we mean a system of m components which will function provided at least k of its components are functioning. This means that the structure function is

$$\phi(\boldsymbol{z}) = \left\{ \begin{array}{ll} 1 \text{ if } & \sum\limits_{i=1}^{k} z_i \geq k, \\ 0 \text{ if } & \sum\limits_{i=1}^{k} z_i < k. \end{array} \right\} \tag{1.41a}$$

Fig. 1/7 shows the block–diagram of a 2–out–of–3 system which looks like that of a series–parallel system with system–level redundancy. Note that this diagram does not reflect the physical layout, but rather the paths through the system that will allow operation of the system.

Figure 1/7: Block–diagram of a 2–out–of–3 system



The structure function of the 2–out–of–3 system is

$$\phi(\boldsymbol{z}) = 1 - (z_1\, z_2)\,(1 - z_1\, z_3)\,(1 - z_2\, z_3). \tag{1.41b}$$

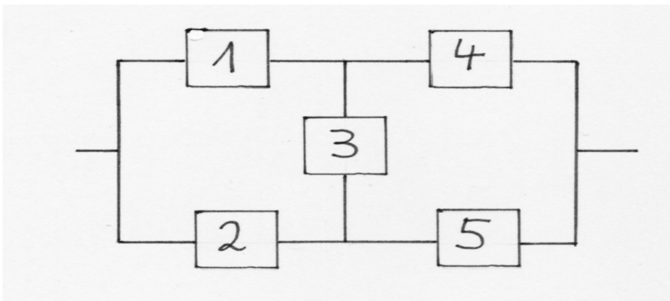
There are two border cases of the k–out–of–m system:

1) the series system is a m–out–of–m system,
2) the parallel system is a 1–out–of–m system.

3. *Bridge system*

   Bridge–structure systems provide another useful way of improving the reliability of certain systems. Fig. 1/8 illustrates a simple bridge system where component 3 is the bridge. If component 3 is working (not working), this system has the same structure as Fig. 1/6 (Fig. 1/5).

   Figure 1/8: Block–diagram of a bridge system



   A moment's reflection on the diagram in Fig. 1/8 reveals that the system functions if any one of the following sets of components functions:

$$\{1,\, 4\},\ \{1,\, 3,\, 5\},\ \{2,\, 5\},\ \{2,\, 3,\, 4\}.$$

   These sets are referred to as **minimal path sets**. Since one or more of these sets of components must function for the system to function, the block–diagram may be written as a parallel arrangement of these sets, each set being a series arrangement of its members. Thus, the structure function corresponding to Fig. 1/8 is

$$\phi(\boldsymbol{z}) = 1 - (1 - z_1\, z_2)\,(1 - z_1\, z_3\, z_5)\,(1 - z_2\, z_5)\,(1 - z_2\, z_3\, z_4). \tag{1.42}$$

We now turn to a technique of finding the reliability[20] of a coherent system of m *independent components*. We introduce the notation Z_i to denote the random state of C_i at a given point in time:

$$Z_i = \left\{ \begin{array}{l} 0 \ \text{ if } \ C_i \ \text{ has failed} \\ 1 \ \text{ if } \ C_i \ \text{ is functioning} \end{array} \right\};\ i = 1,\, 2, \ldots, m.$$

These m variates can be written as a random state vector $\boldsymbol{Z}$. The probability that C_i is functioning *at a certain time* is given by

$$P_i = \Pr(Z_i = 1). \tag{1.43a}$$

These m values can be written as a reliability vector:

$$\boldsymbol{P} = (P_1,\, P_2, \ldots, P_m). \tag{1.43b}$$

The **system reliability** at a certain time is defined as

$$R(\boldsymbol{P}) = \Pr\big[\phi(\boldsymbol{Z}) = 1\big], \tag{1.43c}$$

because R is a quantity that can be calculated from the vector $\boldsymbol{P}$. The method of calculation used here is based on the fact that $\Pr\big[\phi(\boldsymbol{Z}) = 1\big]$ is equal to $\mathrm{E}\big[\phi(\boldsymbol{Z})\big]$, since $\phi(\boldsymbol{Z})$ is a BERNOULLI random vector. Consequently, the expected value of $\phi(\boldsymbol{Z})$ is the system reliability:

$$R(\boldsymbol{P}) = \mathrm{E}\big[\phi(\boldsymbol{Z})\big]. \tag{1.43d}$$

[20] There exist several techniques, see LEEMIS (1995, pp. 28ff.), each having its special advantages and disadvantages.

For instance, applying (1.43d) to the structure function $\phi_S(\cdot)$ of a series system — see (1.36c) — we find

$$\begin{aligned} R_S(\boldsymbol{P}) &= \mathrm{E}\big[\phi_S(\boldsymbol{Z})\big] \\ &= \mathrm{E}\left[\prod_{i=1}^{m} Z_i\right] \\ &= \prod_{i=1}^{m} \mathrm{E}\big(Z_i\big), \text{ because of independence} \\ &= \prod_{i=1}^{m} P_i. \end{aligned} \tag{1.44a}$$

Likewise we find the reliability of a parallel system using $\phi_P(\cdot)$ of (1.37c) as

$$\begin{aligned} R_P(\boldsymbol{P}) &= \mathrm{E}\big[\phi_P(\boldsymbol{Z})\big] \\ &= \mathrm{E}\left[1 - \prod_{i=1}^{m}(1 - Z_i)\right] \\ &= 1 - \mathrm{E}\left[\prod_{i=1}^{m}(1 - Z_i)\right] \\ &= 1 - \prod_{i=1}^{m} \mathrm{E}\big(1 - Z_i\big), \text{ because of independence} \\ &= 1 - \prod_{i=1}^{m}(1 - P_i). \end{aligned} \tag{1.44b}$$

We can state the following **rule to find the reliability of a coherent system** with independent components:

1. Determine the system structure function $\phi(\boldsymbol{Z})$.
2. Replace each Z_i by $P_i = \Pr(Z_i = 1)$.
3. The result is $R(\boldsymbol{P})$.

Applying this rule to (1.39) we can state that for any coherent system with structure function $\phi_C(\boldsymbol{Z})$, $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_m)$, its reliability $R(\boldsymbol{P})$ is bounded above and below by the reliabilities of a series system and a parallel system, each having the same components:

$$R_S(\boldsymbol{P}) = \prod_{i=1}^{m} P_i \le R(\boldsymbol{P}) \le R_P(\boldsymbol{P}) \le 1 - \prod_{i=1}^{m}(1 - P_i). \tag{1.44c}$$

This inequality is not especially sharp. For instance, having $m = 5$ identical components acting independently with $P_1 = P_2 = \ldots = P_5 = 0.9$ we find

$$R_S(\boldsymbol{P}) = 0.59049 \le R(\boldsymbol{P}) \le R_P(\boldsymbol{P}) = 0.99999.$$

In order to introduce *time dependency* into the reliability function we have to substitute P_i by the survival function $S_i(x)$. Let

$$\boldsymbol{S}(x) = \big(S_1(x), S_2(x), \ldots, S_m(x)\big),$$

then the **time dependent system reliability function** is denoted $R\big[\boldsymbol{S}(x)\big]$ and the **system hazard rate** will be

$$h(x) = -\frac{dR\big[\boldsymbol{S}(x)\big]\big/dx}{R\big[\boldsymbol{S}(x)\big]}. \tag{1.45}$$

In general, (1.45) will not result into a handsome formula, even when we assume identical components so that $S_i(x) = S(x)\,\forall\, i$. In most cases $dR\big[\boldsymbol{S}(x)\big]\big/dx$ has to be determined by numerical differentiation. Therefore, we only take a look at the hazard rate functions of the two most simple systems, i.e., the series and the parallel systems.

The reliability function of a *series system* is

$$R_S\big[\boldsymbol{S}(x)\big] = \prod_{i=1}^{m} S_i(x). \tag{1.46a}$$

Now, by

$$S_i(x) = \exp\left\{-\int_0^x h_i(u)\,du\right\}, \tag{1.46b}$$

see (1.9b), where $h_i(.)$ is the HR of component C_i, we first have

$$R_S\big[\boldsymbol{S}(x)\big] = \exp\left\{-\sum_{i=1}^{m}\int_0^x h_i(u)\,du\right\} \tag{1.46c}$$

and upon interchanging summation and integration we find

$$R_S\big[\boldsymbol{S}(x)\big] = \exp\left\{-\int_0^x \sum_{i=1}^{m} h_i(u)\,du\right\}. \tag{1.46d}$$

Because $R_S(\boldsymbol{S}(x)\big]$ is the survival function of the series system (1.9b) holds and

$$h_S(x) = \sum_{i=1}^{m} h_i(x) \tag{1.46e}$$

must be the **hazard rate of the series system** which is the sum of the component hazard rates. Thus, the series system hazard rate is the higher the more components are linked together. This is in accordance with the fact that the series system reliability is a decreasing function of the number of components. When the components have identical lifetime distributions we have $h_i(x) = h(x)\;\forall\, i$ and (1.46e) turns into

$$h_S(x) = m\,h(x). \tag{1.46f}$$

Because of (1.36b) this is the **hazard rate of the minimum order statistic**.

The reliability function of a *parallel system* is

$$R_P\big[\boldsymbol{S}(x)\big] = 1 - \prod_{i=1}^{m} F_i(x). \tag{1.47a}$$

Assuming identical components — $F_i(x) = F(x)\;\forall\, i$ — (1.47a) turns into

$$R_P\big[\boldsymbol{S}(x)\big] = 1 - \big[F(x)\big]^m \tag{1.47b}$$

so that (1.45) gives the **hazard rate of the parallel system**

$$h_P(x) = \frac{m\,f(x)\,\big[F(x)\big]^{m-1}}{1-\big[F(x)\big]^m}, \tag{1.47c}$$

which — because of (1.37b) — is nothing but the **hazard rate of the maximum order statistic**. Let

$$h(x) = \frac{f(x)}{1 - F(x)}$$

be the hazard rate of a component, then (1.47c) can be transformed into

$$\begin{aligned} h_P(x) &= \frac{m\,\big[F(x)\big]^{m-1}}{\sum\limits_{i=0}^{m-1}\big[F(x)\big]^{i}}\,h(x),\ \big[F(x)\big]^0 = 1, &\qquad (1.47\text{d})\\ &= \frac{m}{\sum\limits_{i=0}^{m-1}\big[F(x)\big]^{-i}}\,h(x). &\qquad (1.47\text{e})\end{aligned}$$

(1.47e) follows from (1.47d) when dividing the numerator and the denominator on the right–hand side by $\big[F(x)\big]^{m-1}$. The factor $\Big(m\Big/\sum\limits_{1=0}^{m-1}\big[F(x)\big]^{-i}\Big)$ goes to 0 as $x \to 0$, and it goes to 1 as $x \to \infty$, thus the hazard rate of the parallel system is always less than the hazard rate of an individual component. This is in accordance with the fact that the reliability of a parallel system is always higher than that of an individual component.

1.1.2.4 Acceleration and Proportional Hazards[21]

The accelerated life model and the proportional hazards model are designed to include a vector $\boldsymbol{z}$ of **covariates** (= explanatory variables) $z_i;\ i = 1, \dots, k$; in a lifetime model. z_i influences the lifetime X of the unit under study, and the z_i are non–random. Covariates may account for the fact that the population of units is not truly homogeneous. Other possibilities for the elements of $\boldsymbol{z}$ include cumulative load applied, time–varying stress, and environmental factors. The difference between accelerated life models and proportional hazards models is that in the first case the covariates affect the rate at which the unit ages, and in the second case the covariates increase or decrease the hazard rate. So, in accelerated life models the survival function has to be modeled and in proportional hazards models the hazard rate has to be modified.

We first give a short introduction to **accelerated life models**. The question here is how to link the covariates to the survival function. One approach is to define one lifetime model when $\boldsymbol{z} = \mathbf{0}$, called the **baseline model** and other models for $\boldsymbol{z} \neq \mathbf{0}$. Analysis is simplified when there is only a single model appropriate for all values of $\boldsymbol{z}$. The survivor function of X in the accelerated life model is

$$S(x) = S_0\big[x\,\psi(\boldsymbol{z})\big],\ \ x \ge 0, \qquad (1.48\text{a})$$

where $S_0(\cdot)$ is a **baseline survival function** and $\psi(\boldsymbol{z})$ is a **link function**. The covariates are linked to the lifetime by $\psi(\boldsymbol{z})$, satisfying

$$\psi(\mathbf{0}) = 1 \ \text{ and }\ \psi(\boldsymbol{z}) > 0\ \forall\, \boldsymbol{z} \neq \mathbf{0}. \qquad (1.48\text{b})$$

With these attributes of $\psi(\boldsymbol{z}),\ \boldsymbol{z} = \mathbf{0}$ implies $S_0(x) = S(x)$. A very popular choice for $\psi(\boldsymbol{z})$ is the **log–linear link function**

$$\psi(\boldsymbol{z}) = \exp\big(\boldsymbol{\beta}'\,\boldsymbol{z}\big). \qquad (1.48\text{c})$$

The vector $\boldsymbol{\beta}$ represents regression coefficients and $\boldsymbol{z}$ is a vector of non–random regressors. With (1.48c) the covariates accelerate $\big(\boldsymbol{\beta}'\,\boldsymbol{z} > 0\big)$ or decelerate $\big(\boldsymbol{\beta}'\,\boldsymbol{z} < 0\big)$ the rate at which a unit

[21] Suggested reading for ‘acceleration’: NELSON (1990), and for ‘proportional hazards’: COX/OAKES (1984).

moves through time with respect to the baseline case. Other, less popular choices for the link function are

$$\psi(\boldsymbol{z}) = \boldsymbol{\beta}'\,\boldsymbol{z} \ \text{ and } \ \psi(\boldsymbol{z}) = \big(\boldsymbol{\beta}'\,\boldsymbol{z}\big)^{-1}.$$

With these two specifications it may happen that $\psi(\boldsymbol{z}) < 0$ for some value of $\boldsymbol{\beta}$ resulting into a negative lifetime. The other lifetime distribution representatives for an accelerated lifetime model with $S(x) = S_0\big[x\,\psi(\boldsymbol{z})\big]$ are

$$\begin{aligned}
F(x) &= 1 - S(x) = S_0\big[x\,\psi(\boldsymbol{z})\big], && (1.48d)\\
f(x) &= \psi(\boldsymbol{z})\,f_0\big[x\,\psi(\boldsymbol{z})\big], && (1.48e)\\
h(x) &= \frac{f(x)}{S(x)} = \psi(\boldsymbol{z})\,\frac{f_0\big[x\,\psi(\boldsymbol{z})\big]}{S_0\big[x\,\psi(\boldsymbol{z})\big]} = \psi(\boldsymbol{z})\,h_0\big[x\,\psi(\boldsymbol{z})\big], && (1.48f)\\
H(x) &= -\ln\Big\{S_0\big[x\,\psi(\boldsymbol{z})\big]\Big\} = H_0\big[x\,\psi(\boldsymbol{z})\big]. && (1.48g)
\end{aligned}$$

We recognize that these formulas resemble those of the variable of the variable transformation of Sect. 1.1.2.1.

Whereas accelerated lifetime models modify the rate that the unit moves through time, **proportional hazard models** modify the hazard rate by the factor $\psi(\boldsymbol{z})$:

$$h(x) = \psi(\boldsymbol{z})\,h_0(x), \ \ x \ge 0. \qquad (1.49a)$$

$h_0(x)$ is called the **baseline hazard**, representing the hazard rate for a unit having $\psi(\boldsymbol{z}) = 1$. As before, a popular choice for the link function here is the log–linear form (1.48c), and the hazard rate increases when $\boldsymbol{\beta}'\,\boldsymbol{z} > 0$ and decreases when $\boldsymbol{\beta}'\,\boldsymbol{z} < 0$. The 'proportional' terminology arises in a perfectly natural way. If two units 1 and 2 have lifetimes depending on respective vectors of covariate values $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, then

$$\frac{h_1(x)}{h_2(x)} = \frac{h_0(x)\,\psi(\boldsymbol{z}_1)}{h_0(x)\,\psi(\boldsymbol{z}_2)} = \frac{\psi(\boldsymbol{z}_1)}{\psi(\boldsymbol{z}_2)},$$

showing clearly how the baseline hazards cancel from this ratio, so that the hazard ratio for the two units does not depend on lifetime x. The other lifetime distribution representatives can be determined from (1.49a):

$$\begin{aligned}
H(x) &= \int\limits_0^x h(u)\,du\\
&= \int\limits_0^x \psi(\boldsymbol{z})\,h_0(u)\,du\\
&= \psi(\boldsymbol{z})\,H_0(x), && (1.49b)\\
S(x) &= \exp\left\{-\int\limits_0^x \psi(\boldsymbol{z})\,h_0(u)\,du\right\}\\
&= \exp\left\{-\psi(\boldsymbol{z})\int\limits_0^x h_0(u)\,du\right\}\\
&= \left\{\exp\left[-\int\limits_0^x h_0(u)\,du\right]\right\}^{\psi(\boldsymbol{z})}\\
&= \big[S_0(x)\big]^{\psi(\boldsymbol{z})}, && (1.49c)
\end{aligned}$$

$$F(x) = 1 - \big[S_0(x)\big]^{\psi(\boldsymbol{z})}, \tag{1.49d}$$

$$f(x) = f_0(x)\,\psi(\boldsymbol{z})\,\big[S_0(x)\big]^{\psi(\boldsymbol{z})-1}. \tag{1.49e}$$

Example 1/7: Accelerated life model and proportional hazards model for WEIBULL baseline

The WEIBULL baseline survival function for the *accelerated life model* is

$$S_0(x) = \exp\Big[-\Big(\frac{x}{b}\Big)^c\Big];\ x \geq 0;\ b,\ c > 0,$$

where b is a scale parameter and c is a shape parameter. Introducing a link function $\psi(\boldsymbol{z})$ into $S_0(\cdot)$ according to (1.48a) gives

$$\begin{aligned} S_A(x) &= S_0\big[x\,\psi(\boldsymbol{z})\big] \\ &= \exp\left[-\left(\frac{x\,\psi(\boldsymbol{z})}{b}\right)^c\right]. \end{aligned} \tag{1.50a}$$

Thus, the accelerated lifetime has a WEIBULL distribution as well, but the scale parameter has changed from b to $b_A = b\big/\psi(\boldsymbol{z})$. The hazard rate belonging to (1.50a) is

$$\begin{aligned} h_A(x) &= \psi(\boldsymbol{z})\,h_0(x) \\ &= \psi(\boldsymbol{z})\,\frac{c}{b}\left[\frac{x\,\psi(\boldsymbol{z})}{b}\right]^{c-1} \\ &= \frac{c}{b\big/\psi(\boldsymbol{z})}\left[\frac{x\,\psi(\boldsymbol{z})}{b}\right]^{c-1}. \end{aligned} \tag{1.50b}$$

For the *proportional hazards model* the WEIBULL baseline hazard is

$$h_0(x) = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1}.$$

So, according to (1.49a) the hazard rate for a unit with covariate vector $\boldsymbol{z}$ and link function $\psi(\boldsymbol{z})$ is

$$h_P(x) = h_0(x)\,\psi(\boldsymbol{z} = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1}\psi(\boldsymbol{z}). \tag{1.51a}$$

Inserting the usual baseline $S_0(x) = \exp\big[-\big(\frac{x}{b}\big)^c\big]$ for the WEIBULL distribution into (1.49c) the survival function corresponding to (1.51a) is

$$\begin{aligned} S_P(x) &= \Big\{\exp\Big[-\Big(\frac{x}{b}\Big)^c\Big]\Big\}^{\psi(\boldsymbol{z}} \\ &= \exp\Big[-\Big(\frac{x}{b}\Big)^c\,\psi(\boldsymbol{z})\Big] \\ &= \exp\left[-\left(\frac{x\,\psi(\boldsymbol{z})^{1/c}}{b}\right)^c\right]. \end{aligned} \tag{1.51b}$$

(1.51b) can be recognized as a WEIBULL distribution as well, but contrary to (1.50a) the scale parameter is $b_P = b\big/\psi(\boldsymbol{z})^{1/c}$. The WEIBULL distribution is the only baseline distribution where the accelerated life and the proportional hazards models coincide in this fashion. (1.50a) and (1.51b) are identical for $c = 1$ which is an exponential distribution.

1.1.2.5 Truncated Distributions[22]

Truncation and censoring are two operations which often are confused in statistical literature, but there is a clear distinction between these two concepts. **Truncation** is confined to the distribution

[22] Suggested reading for this section: COHEN (1991).

or to the population whereas **censoring** is related to samples. So we talk about *truncated distributions* and *censored samples*, and we may have a censored sample from a truncated distribution. Truncation means that the original or natural support of a distribution has been shrunken so that the portion of the population in the truncated part can never be observed and a certain part of the original probability mass will be cut off. Thus, truncation modifies the distribution and leads to conditional distributions. Censoring, either from an unmodified distribution or from a truncated distribution, modifies the selection of the random variables and is thus related to the sampling process. Censoring means that — for one reason or the other — the statistician refrains from measuring the exact value of a unit's characteristic when this value falls inside a certain area, e.g., is greater or smaller than a given threshold. Censored samples produce two types of observations, those with known and complete value of the characteristic under study, and those which fall into a special region of the characteristic and are thus only known by their frequency and not by their value. A censored observation is distinct from a **missing observation** in that the order of the censored observation relative to some of the uncensored observations is known and conveys information regarding the distribution being sampled.

There are three common **types of truncation**:

- **left truncation**, also known as **lower truncation** or **truncation from below**,
- **right truncation**, also known as **upper truncation** or **truncation from above**,
- **double truncation**, also known as **truncation on both sides** or **truncation from below and above**.

Regarding censoring we have substantial more types which are fully described and discussed in RINNE (2009, p. 291–312) and in Sect. 4. Truncation of a lifetime distribution leads to different types of lifetime:

- Left truncation gives **future lifetime** or remaining lifetime, i.e., lifetime greater than x_l, the **lower point of truncation**. Left truncation may be realized by burn–in or by preselecting of apparently weak looking units (freaks).
- Right truncation gives **early lifetime** or young lifetime, i.e., lifetime less than x_u, the **upper point of truncation**. Right truncation will be met with when for economic or safety reasons items will be in operation for at most x_u units of time.
- Double truncation gives **interim lifetime**, i.e., lifetime which is less than x_u, but greater than x_l and thus is in the interval $[x_l,\, x_u]$. We may think of interim lifetime as either future lifetime truncated on the right or as early lifetime truncated on the left.

We first present *results for the future lifetime* which has already been introduced in Sect. 1.1.1.4 in conjunction with the introduction of the hazard rate. Future lifetime beyond x_l will be denoted

$$X - x_l \,|\, X \geq x_l := Y \,|\, x_l$$

and its distributional representatives — expressed in terms of the original distribution —are

$$f_l(y \,|\, x_l) = \frac{f(x_l + y)}{1 - F(x_l)} = \frac{f(x_l + y)}{S(x_l)}, \;\; x_l \geq 0,\; y \geq 0, \tag{1.52a}$$

$$F_l(y \,|\, x_l) = \frac{F(x_l + y) - F(x_l)}{1 - F(x_l)} = \frac{S(x_l) - S(x_l + y)}{S(x_l)}, \tag{1.52b}$$

$$S_l(y \,|\, x_l) = \frac{S(x_l + y)}{S(x_l)} = \frac{1 - F(x_l + y)}{1 - F(x_l)}, \tag{1.52c}$$

$$h_l(y \,|\, x_l) \;=\; \frac{f_l(y \,|\, x_l)}{S_l(y \,|\, x_l)} \;=\; \frac{f(x_l + y)}{S(x_l + y)} \;= h(x_l + y). \tag{1.52d}$$

The hazard rate for the future lifetime Y of a distribution truncated on the left at x_l is identical to that of the original distribution at $x = x_l + y$, i.e., the courses of these two hazard rates only differ by a translation.

$$\begin{aligned} H_l(y \,|\, x_l) &= \int_0^y h_l(u \,|\, x_l)\, du, \\ &= \int_0^y h(x_l + u)\, du, \\ &= \int_{x_l}^{x_l+y} h(v)\, dv, \quad v = x_l + u, \\ &= H(x_l + y) - H(x_l). \end{aligned} \tag{1.52e}$$

The raw moments of $Y \,|\, x_l$ are

$$\mathrm{E}\big(Y^k \,|\, x_l\big) = \frac{k}{S(x_l)} \int_0^\infty y^{k-1}\, S(x_l + y)\, dy. \tag{1.52f}$$

Especially, for $k = 1$ we have

$$\mathrm{E}(Y \,|\, x_l) = \frac{1}{S(x_l)} \int_0^\infty S(x_l + y)\, dy \tag{1.52g}$$

which — in Sect. 1.1.1.e — was denoted $\mu(x_l)$ and called the MRL of X. The MRL of the future lifetime — after some manipulations — is

$$\mu_l(y) = \frac{1}{S(x_l + y)} \int_y^\infty S(x_l + v)\, dv = \mu(x_l + y), \tag{1.52h}$$

i.e., the MRL of the left–truncated variable, truncated at x_l, after y units of time is the same as that of the original variable at age $x_l + y$. Like the HR the MRL is translated. The percentile function of $Y \,|\, x_l$ follows from

$$F\big(y_P \,|\, x_l\big) = \frac{F(x_l + y_P) - F(x_l)}{1 - F(x_l)} = P, \;\; 0 \leq P \leq 1,$$

as

$$y_P = F^{-1}\big[p + (1 - P)\, F(x_l)\big] - x_l, \tag{1.52i}$$

i.e., y_P is equal to the percentile of order $P + (1 - P)\, F(x_l)$ of the original distribution, but reduced by the value of the truncation point x_l. Sometimes (particularly, when the distribution is highly skewed), the median is preferred to the mean, in which case, the quantity 'median residual life' at x_l is preferred to the mean residual lifetime. The median residual lifetime at x_l is the length of the interval from x_l to that time where one–half of the units alive at x_l will still be alive.

Now we give *results for the early lifetime*, denoted

$$x_u - X \,|\, X \leq x_u =: Y \,|\, x_u.$$

$$\begin{aligned}
f_u(y\,|\,x_l) &= \frac{f(y)}{F(x_u)} = \frac{f(y}{1-S(x_u}, \;\; x_u>0, \; 0\le y\le x_u, &(1.53a)\\
F_u(y\,|\,x_l) &= \frac{F(y)}{F(x_u)} = \frac{1-S(y)}{1-S(x_u)}, &(1.53b)\\
S_u(y\,|\,x_l) &= \frac{S(y)-S(x_u)}{1-S(x_u)} = \frac{F(x_u)-F(y)}{F(x_u)}, &(1.53c)\\
h_u(y\,|\,x_l) &= \frac{f_u(y\,|\,x_u)}{S_u(y\,|\,x_u)} = \frac{f(y)}{S(y)-S(x_u)} = h(y)\,\frac{S(y)}{S(y)-S(x_u)}. &(1.53d)
\end{aligned}$$

Since $S(y) > S(y) - S(x_u)$ for $y < x_u$ the hazard rate of the right–truncated distribution is greater than that of the original distribution. (1.53d) shows that truncation from below markedly affects the course of the HR, whereas truncation from above only shifts the HR, see (1.52d) and Fig. 1/9. Since the HR at y is conditional on survival to y, truncation below y is immaterial. However, truncation above y has an effect on the HR at y, since the time interval remaining for failing is shortened. It should be clear that as $y \to x_u$ from below, $h_u(y\,|\,x_u)$ becomes indefinitely large since the interval for failing approaches zero.

$$H_u(y\,|\,x_u) = -\ln\big[S_u(y\,|\,x_u)\big] = \ln F(x_u) - \ln\big[S(y)-S(x_u)\big]. \tag{1.53e}$$

For the mean of $Y \mid x_u$ we find

$$\begin{aligned}
\mathrm{E}(Y\,|\,x_u) &= \int\limits_0^{x_u} S_u(v\,|\,x_u)\,dv\\
&= \frac{1}{1-S(x_u)}\left\{\int\limits_0^{x_u} S(v)\,dv - x_u\,S(x_u)\right\}, &(1.53f)
\end{aligned}$$

and for the MRL we have

$$\begin{aligned}
\mu_u(y) &= \frac{1}{S_u(y\,|\,x_u)}\int\limits_0^{x_u} S_u(v\,|\,x_u)\,dv\\
&= \frac{1}{S(y)-S(x_u)}\left\{\int\limits_0^{x_u} S(v)\,dv - S(x_u)\big[x_u-y\big]\right\}. &(1.53g)
\end{aligned}$$

We notice that $\mu_u(y)$ approaches zero with $x \to x_u$. The percentile function of $Y\,|\,x_u$ follows from

$$F_u(y_P\,|\,x_u) = \frac{F(y_P)}{F(x_u)} = P, \;\; 0\le P\le 1,$$

as

$$y_P = F^{-1}\big[P\,F(x_u)\big] \tag{1.53h}$$

i.e., y_P is equal to the percentile of order $P\,F(x_u)$ of the original distribution.

The results for the *interim lifetime*

$$\big[(x_u-x_l)-(x_u-X)\,\big|\,x_l\le X\le x_u\big] = \big[X-x_l\,|\,x_l\le X\le x_u\big] =: Y\,|\,x_l;x_u$$

are the following

$$\begin{aligned}
f_{l,u}(y\,|\,x_l;x_u) &= \frac{f(x_l+y)}{F(x_u)-f(x_l)}\\
&= \frac{f(x_l+y)}{S(x_l)-S(x_u)}, \;\; 0\le x_l<x_u, \;\; 0\le y\le x_u-x_l, &(1.54a)
\end{aligned}$$

$$F_{l,u}(y\,|\,x_l;x_u) \;=\; \frac{F(x_l+y)-F(x_l)}{F(x_u)-F(x_l)} \;=\; \frac{S(x_l)-S(x_l+y)}{S(x_l)-S(x_u)}, \tag{1.54b}$$

$$S_{l,u}(y\,|\,x_l;x_u) \;=\; \frac{F(x_u)-F(x_l+y)}{F(x_u)-F(x_l)} \;=\; \frac{S(x_l+y)-S(x_u)}{S(x_l)-S(x_u)}, \tag{1.54c}$$

$$\begin{aligned} h_{l,u}(y\,|\,x_l;x_u) &= \frac{f_{l,u}(y\,|\,x_l;x_u)}{S_{l,u}(y\,|\,x_l;x_u)} = \frac{f(x_l+y)}{S(x_l+y)-S(x_u)} \\ &= h(x_l+y)\,\frac{S(x_l+y)}{S(x_l+y)-S(x_u)}, \end{aligned} \tag{1.54d}$$

$$\begin{aligned} H_{l,u}(y\,|\,x_l;x_u) &= -\ln\big[S_{l,u}(y\,|\,x_l;x_u)\big] \\ &= \ln\big[F(x_u)-F(x_l)\big]-\ln\big[S(x_l+y)-S(x_u)\big], \end{aligned} \tag{1.54e}$$

$$\begin{aligned} \mathrm{E}(Y\,|\,x_l;x_u) &= \int\limits_{x_l}^{x_u} S_{l,u}(v\,|\,x_l;x_u)\,dv \\ &= \frac{1}{S(x_l)-S(x_u)}\left\{\int\limits_0^{x_u-x_l} S(x_l+v)-[x_u-x_l]\,S(x_u)\right\}, \end{aligned} \tag{1.54f}$$

$$\begin{aligned} \mu_{l,u}(y) &= \frac{1}{S_{l,u}(y\,|\,x_l;x_u)}\int\limits_y^{x_u-x_l} S_{l,u}(v\,|\,x_l;x_u)\,dv \\ &= \frac{1}{S(x_l+y)-S(x_u)}\left\{\int\limits_y^{x_u-x_l} S(x_l+v)\,dv-[x_u-x_l-y]\,S(x_u)\right\}. \end{aligned} \tag{1.54g}$$

The percentile function of $Y\,|\,x_l,x_u$ follows from

$$F_{l,u}(y_P\,|\,x_l;x_u) = \frac{F(x_l+y_P)-F(x_l)}{F(x_u)-F(x_l)} = P$$

as

$$y_P = F^{-1}\big\{P\,[F(x_u)-F(x_l)]+F(x_l)\big\}-x_l. \tag{1.54h}$$

We notice that

- by setting $x_u=\infty$ the formulas (1.54a–h) turn into those for the case of the left–truncation, i.e., into formulas (1.52a–i),
- by setting $x_l=0$ the formulas (1.54a–h) turn into those for the case of right–truncation i.e., into formulas (1.53a–h) and
- by setting $x_l=0$ and $x_u=\infty$ (1.54a–h) give the results of Sections 1.1.1.3 to 1.1.1.6.

Example 1/8: Truncation of the reduced RAYLEIGH distribution

The reduced RAYLEIGH distribution, equivalent to the reduced WEIBULL distribution with shape parameter equal to 2, has

$$\begin{aligned} f(x) &= 2\,x\,\exp\big(-x^2\big),\;\; x\ge 0, \\ F(x) &= 1-\exp\big(-x^2\big), \\ S(x) &= \exp\big(-x^2\big), \end{aligned}$$

$$
\begin{aligned}
h(x) &= 2\,x,\\
H(x) &= x^2,\\
\mathrm{E}\big(X^k\big) &= \Gamma\left(1+\frac{k}{2}\right),\\
\mathrm{E}(X) &= \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2} \approx 0.88623,\\
\mu(x) &= \frac{1}{2}\sqrt{\pi}\exp\big(x^2\big)\,\mathrm{erfc}(x).^{23}
\end{aligned}
$$

Truncation from below at x_l gives

$$
\begin{aligned}
f_l(y\,|\,x_l) &= 2\,(x_l+y)\,\exp\big[-y\,(2\,x_l+y)\big];\;\; x_l>0,\; y\;0;\\
F_l(y\,|\,x_l) &= 1-\exp\big[-y\,(2\,x_l+y)\big],\\
S_l(y\,|\,x_l) &= \exp\big[-y\,(2\,x_l+y)\big],\\
h_l(y\,|\,x_l) &= 2\,(x_l+y),\\
H_l(y\,|\,x_l) &= y\,(2\,x_l+y^2),\\
\mathrm{E}(Y\,|\,x_l) &= \frac{1}{2}\sqrt{\pi}\,\exp\big(x_l^2\big)\,\mathrm{erfc}(x_l),\\
\mu_l(y) &= \mu(x_l+y) = \frac{1}{2}\sqrt{\pi}\,\exp\big[(x_l+y)^2\big]\,\mathrm{erfc}(x_l+y),\\
y_P &= \sqrt{x_l^2-\ln(1-P)}-x_l.
\end{aligned}
$$

Truncation from above at x_u gives

$$
\begin{aligned}
f_u(y\,|\,x_u) &= \frac{2\,y\,\exp\big(-y^2\big)}{1-\exp\big(-x_u^2\big)};\;\; x_u>0,\; 0\le y\le x_u;\\
F_u(y\,|\,x_u) &= \frac{1-\exp\big(-y^2\big)}{1-\exp\big(-x_u^2\big)},\\
S_u(y\,|\,x_u) &= \frac{\exp\big(-y^2\big)-\exp\big(-x_u^2\big)}{1-\exp\big(-x_u^2\big)},\\
h_u(y\,|\,x_u) &= \frac{2\,y}{1-\exp\big(y^2-x_u^2\big)},\\
H_u(y\,|\,x_u) &= -\ln\left[\frac{\exp\big(-y^2\big)-\exp\big(-x_u^2\big)}{1-\exp\big(-x_u^2\big)}\right],\\
\mathrm{E}(Y\,|\,x_u) &= \frac{2\,x_u-\exp\big(-x_u^2\big)\sqrt{\pi}\,\mathrm{erf}(x_u)}{2\big[1-\exp\big(-x_u^2\big)\big]},\\
\mu_u(y) &= \frac{\exp\big(y^2\big)\,\big\{2\,(x_u-y)+\exp\big(x_u^2\big)\sqrt{\pi}\,\big[\mathrm{erf}(x_u)+\mathrm{erf}(y)\big]\big\}}{2\,\big[\exp\big(y^2\big)-\exp\big(x_u^2\big)\big]},\\
y_P &= \sqrt{-\ln\big\{1-P\,\big[1-\exp\big(-x_u^2\big)\big]\big\}}.
\end{aligned}
$$

Truncation from both sides gives

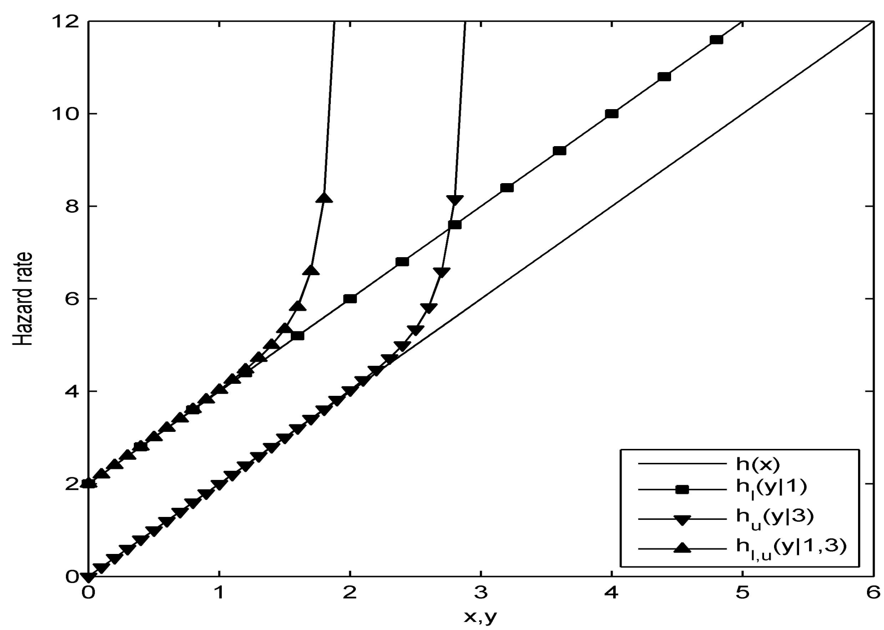
$$
\begin{aligned}
f_{l,u}(y\,|\,x_l;x_u) &= \frac{2\,(x_l+x_u)\,\exp\big[-(x_l+y)^2\big]}{\exp\big(-x_l^2\big)-\exp\big(-x_u^2\big)};\;\; 0<x_l<x_u,\; 0\le y\le x_u-x_l;\\
F_{l,u}(y\,|\,x_l;x_u) &= \frac{\exp\big(-x_l^2\big)-\exp\big[-(x_l+y)^2\big]}{\exp\big(-x_l^2\big)-\exp\big(-x_u^2\big)},\\
S_{l,u}(y\,|\,x_l;x_u) &= \frac{\exp\big[-(x_l+y)^2\big]-\exp\big(-x_l^2\big)}{\exp\big(-x_l^2\big)-\exp\big(-x_u^2\big)},
\end{aligned}
$$

[23] $\mathrm{erf}(x) = 1-\frac{2}{\sqrt{\pi}}\int\limits_0^x \exp\big(-u^2\big)\,du$ is the **error function** and erfc$=1-$erf the **complementary error function**.

$$\begin{aligned}
h_{l,u}(y\,|\,x_l;x_u) &= \frac{2\,(x_l+x_u)}{1-\exp\big[(x_l+y)^2-x_u^2\big]},\\
H_{l,u}(y\,|\,x_l;x_u) &= -\ln\left\{\frac{\exp\big[-(x_l+y)^2\big]-\exp\big(-x_u^2\big)}{\exp\big(-x_l^2\big)-\exp\big(-x_u^2\big)}\right\},\\
\mathrm{E}(Y\,|\,x_l;x_u) &= \frac{\exp\big(x_l^2\big)\,\big\{2\,(x_u-x_l)+\exp\big(x_u^2\big)\sqrt{\pi}\,\big[\mathrm{erf}(x_l)-\mathrm{erf}(x_u)\big]\big\}}{2\big[\exp\big(x_l^2\big)-\exp\big(x_u^2\big)\big]},\\
\mu_{l,u}(y) &= \frac{\exp\big[(x_l+y)^2\big]\,\big\{2\,(x_l-x_u-y)+\exp\big(x_u^2\big)\sqrt{\pi}\,\big[\mathrm{erf}(x_l+y)-\mathrm{erf}(2\,x_l-x_u)\big]\big\}}{2\big[\exp\big[(x_l+y)^2\big]-\exp\big(x_u^2\big)\big]},\\
y_P &= \sqrt{-\ln\big[P\,\exp\big(-x_u^2\big)+(1-P)\,\exp\big(-x_l^2\big)\big]}-x_l.
\end{aligned}$$

Fig. 1/9 shows the hazard rates for the original distribution together with those of the three types of truncation where the truncation points are $x_l=1$ and $x_u=3$, respectively.

Figure 1/9: Hazard rates of the reduced RAYLEIGH distribution, non–truncated and truncated at $x_l=1$ and/or $x_u=3$



1.1.2.6 Life Potential

The integral

$$\Pi(x) = \int\limits_x^\infty S(u)\,du \tag{1.55a}$$

$$= \mu(x)\,S(x) \tag{1.55b}$$

is the **life potential**, i.e., the total number of expected time units to be spend by the fraction of units in the population which survive the age x. Let N be the number of persons at age $x=0$, i.e., N is the size of the population, then in life tables and in the actuarial sciences the product $N\times\Pi(x)$ is known as 'the expected total number of years lived beyond age x by persons alive at age x'. There, this quantity is denoted T_x and is measured in 'population units $\times$ time units', i.e., person–years. When the population is mechanical equipment the quantity $N \times \Pi(x)$ will be called or machine–hours. In the engineering sciences the complement of $\Pi(x)$ to $\mu = \mathrm{E}(X)$:

$$\mu - \Pi(x) = \int\limits_0^x S(u)\,du \tag{1.55c}$$

is known as **total–time–on–test up to age x**.[24] We see that

$$\Pi(0) = \int\limits_0^\infty S(u)\,du = \mu = \mathrm{E}(X). \tag{1.55d}$$

The **scaled life potential**

$$\begin{aligned} \widetilde{S}(x) &= \frac{\Pi(x)}{\Pi(0)},\ \ 0 \le \widetilde{S}(x) \le 1, \\ &= \frac{\mu(x)}{\mu}\,S(x), \\ &= \frac{1}{\mu}\int\limits_x^\infty S(u)\,du \end{aligned} \tag{1.56a}$$

tells what fraction of the original life potential $\Pi(0) = \mu$ is still available at age x. Whereas $S(x)$ says what *fraction of the initial size of population units* has survived the age x, $\widetilde{S}(x)$ tells what *fraction of the initial number of lifetime units* has not been 'consumed' up to x. $\widetilde{S}(x)$ can be regarded as a survival function, but not as one of population units, but as one of lifetime units. Thus, we may call $\widetilde{S}(x)$ the **potential survival function**. Looking at (1.56a) we see that

$$\widetilde{S}(x) \ge S(x) \ \text{ for increasing MRL}$$

and

$$\widetilde{S}(x) \le S(x) \ \text{ for decreasing MRL.}$$

$\widetilde{S}(x)$ is — like $S(x)$ — a monotone and decreasing function with

$$\widetilde{S}(0) = 1 \ \text{ and } \ \widetilde{S}(\infty) = 0.$$

Besides $\widetilde{S}(x)$ we can define the following representatives of the life potential distribution:

- **potential distribution function**

$$\widetilde{F}(x) := 1 - \widetilde{S}(x) = \frac{1}{\mu}\int\limits_0^x S(u)\,du, \tag{1.56b}$$

- **potential density function**

$$\widetilde{f}(x) := \frac{d\widetilde{F}(x)}{dx} = \frac{S(x)}{\mu}, \tag{1.56c}$$

[24] We mention that the **total–time–on–test transform** of a lifetime distribution is a function of P, $0 \le P \le 1$, the portion failing, and the upper limit of integration is expressed in terms of the percentile $x_P = F^{-1}(P)$:

$$H_F^{-1}(P) = \int\limits_0^{F^{-1}(P)} S(u)\,du.$$

- **potential hazard rate**

$$\widetilde{h}(x) := \frac{\widetilde{f}(x)}{\widetilde{S}(x)} = \frac{S(x)}{\int\limits_x^\infty S(u)\,du} = \frac{1}{\mu(x)}, \tag{1.56d}$$

- **cumulative potential hazard rate**

$$\widetilde{H}(x) := \int\limits_0^x \widetilde{h}(u)\,du = \int\limits_0^x \frac{1}{\mu(u)}\,du = \ln\mu - \ln\mu(x) - \ln S(x), \quad \text{see (1.19e).} \tag{1.56e}$$

The potential hazard rate $\widetilde{h}(x)$ may be regarded as the *natural rate of depreciation* for a population having lifetime distribution $F(x)$, when this rate is applied to the stock of units having age x. The only source of depreciation is 'death' or failure of population units.[25] $\widetilde{h}(x)$ is the velocity with which the stock of lifetime units decreases at age x and $\widetilde{h}(x)\,\Delta,\ \Delta$ small, is the amount of lifetime units vanishing in $[x,\,x+\Delta]$.

Example 1/9: Life potential of the exponential and the reduced RAYLEIGH distributions

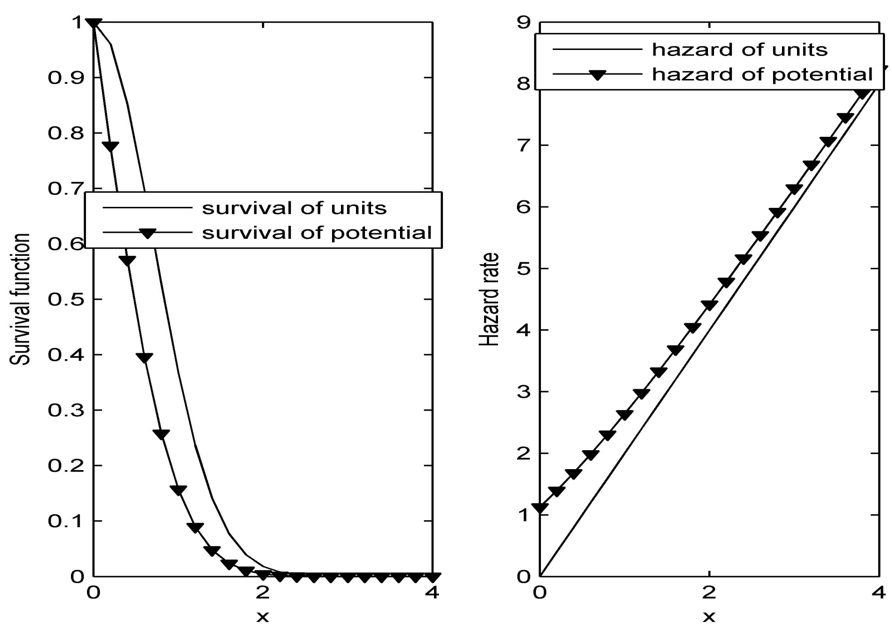
The exponential distribution with $f(x) = \frac{1}{b}\exp\big(-\frac{x}{b}\big)$ is the only continuous distribution where $S(x) = \widetilde{S}(x)\ \forall\, x \geq 0$. This is a consequence of its constant MRL: $\mu(x) = b\ \forall\, x \geq 0$.

For the reduced RAYLEIGH distribution of Example 1/8 we have

$$\Pi(x) = \frac{1}{2}\sqrt{\pi}\,\mathrm{erfc}(x),\ \ x \geq 0,\ \ \widetilde{S}(x) = \frac{\Pi(x)}{\mathrm{E}(X)} \ = \ \mathrm{erfc}(x),\ \ \widetilde{h}(x) = \frac{2\,\exp\big(-x^2\big)}{\sqrt{\pi}\,\mathrm{erfc}(x)}.$$

From Fig. 1/10 we see that $\widetilde{S}(x) < S(x)$, thus, the stock of lifetime units decreases faster than the stock of population units. We further see that $\widetilde{h}(x) > h(x)$, but both rates approach to one another with $x \to \infty$.

Figure 1/10: Survival functions $S(x),\ \widetilde{S}(x)$ and hazard rates $h(x),\ \widetilde{h}(x)$ of the reduced RAYLEIGH distribution



[25] In general, this natural rate of depreciation will be different from the rate of depreciation applied in accounting because there we have to allow for economic and fiscal reasons.

1.2 The Univariate Discrete Case[26]

A large amount of research has been devoted to continuous lifetimes. Many of the concepts that apply to continuous distributions also apply to discrete distributions, but discrete failure time distributions are applied less frequently than continuous distributions since there are fewer situations for which failures can only occur at discrete points in time. However, discrete lifetimes have several important applications. Actuaries and biostatisticians are interested in lifetimes of persons or organisms, measured in years, months, weeks, or days, i.e., time is grouped and counted in some units or time intervals. A reliability engineer will monitor a device only once per time period (hour, day etc.) and will count the time periods successfully completed prior to failure of the device.There are situations in reliability theory where clock time is not the best scale on which to describe lifetime. 'Time' can also be the number of times or cycles that a piece of equipment is operated successfully, or the number of miles that an airplane is flown or a car is run.

1.2.1 General Results

We assume that lifetime X can take a discrete set of values $\{x_i\};\ i = 1,\, 2, \ldots;$ at which failure may occur; generally, we take $x_i = i$. When the x_i's are equidistant, i.e., $x_i \in M = \{a,\, a + 1,\, a + 2, \ldots\}$ for some $a \in \mathbb{R}$, then the transformation $x_i - a$ leads to the set of non–negative integers $\mathbb{N}_0$. We will restrict lifetimes to the set $\mathbb{N}_0$ with **probability mass function PMF** $\{P_i\}$, where

$$P_i := \Pr(X = i);\ i = 0,\, 1,\, 2,\, \ldots$$

Often, we have $P_0 = 0$, but, for generality, we assume that P_0 is not necessarily equal to zero. The case $P_0 \neq 0$ corresponds to a non–zero portion of dud units in a reliability context or to a non–zero probability that a fetus dies at birth or to an egg failing to hatch in a biostatistics context. Different representatives for such discrete lifetime distributions will be described next.

The most basic representative of a discrete lifetime distribution is its PMF

$$\Pr(X = i) = P_i;\ i = 0,\, 1,\, 2,\, \ldots; \tag{1.57}$$

where X is the random time to failure or to death and i is the observed number of time units to this event.

The **survival** or **reliability function** is

$$\begin{aligned} S_i &= 1 - \Pr(X < i);\ i = 0,\, 1,\, 2,\, \ldots; \\ &= \Pr(X \geq i), \\ &= \sum_{j \geq i} P_j. \end{aligned} \tag{1.58a}$$

This is a non–increasing step function which is left–continuous[27] since

$$\lim_{\epsilon \to 0} \big[S_{i-\epsilon} - S_i\big] = 0,\ \ \epsilon > 0,\ i \geq 0.$$

The reliability function is a unique function of the probabilities P_i, and similarly the P_i are determined uniquely by the S_i :

$$P_i = S_i - S_{i+1};\ i = 0,\, 1,\, 2,\, \ldots \tag{1.58b}$$

[26] Suggested reading for the section: KEMP (2004), LAI (2013), SALVIA/BOLLINGER (1982), SHAKED et al. (1995).

[27] This definition is different from that of continuous variate, see (1.4a).

The **distribution function** or **failure function** is the complement to the survival function

$$\begin{aligned} F_i &= \Pr(X < i). \\ &= 1 - S_i;\;\; i = 0,\, 1,\, 2,\, \ldots \end{aligned} \tag{1.59}$$

The **hazard rate** in the discrete case is defined as[28]

$$h_i = \frac{\Pr(X = i)}{\Pr(X \geq i)};\; i = 0,\, 1,\, 2,\, \ldots \tag{1.60a}$$

This is a *conditional probability*, i.e., the probability that a unit fails at time i, given that it has survived to at least time i. Thus we have

$$0 \leq h_i \leq 1,$$

whereas in the continuous case we have $h(x) \geq 0$. We may write h_i in terms of P_i and/or S_i:

$$h_i = \frac{P_i}{\sum\limits_{j \geq i} P_j}, \tag{1.60b}$$

$$= \frac{P_i}{S_i}, \tag{1.60c}$$

$$= 1 - \frac{S_{i+1}}{S_i}. \tag{1.60d}$$

Conversely, we may express P_i and S_i in terms of h_i. Because $S_0 = 1$ we may write S_i for $i = 1,\, 2,\, \ldots$ as the telescope product

$$S_i = \frac{S_i}{S_{i-1}}\,\frac{S_{i-1}}{S_{i-2}}\,\cdots\,\frac{S_2}{S_1}\,\frac{S_1}{S_0}. \tag{1.61a}$$

From (1.60d) we have

$$1 - h_i = \frac{S_{i+1}}{S_i};\; i = 0,\, 1,\, 2,\, \ldots \tag{1.61b}$$

Combining (1.61a) and (1.61b) we find[29]

$$S_i = \prod_{j=0}^{i-1} (1 - h_j);\; i = 0,\, 1,\, 2,\, \ldots \tag{1.61c}$$

From (1.60c) we have

$$P_i = h_i\, S_i, \tag{1.61d}$$

and upon inserting (1.61c) into (1.61d) we find

$$P_i = h_i \prod_{j=0}^{i-1} (1 - h_j);\; i = 0,\, 1,\, 2,\, \ldots \tag{1.61e}$$

In the continuous case we have with respect to the cumulative hazard function

$$\begin{aligned} H(x) &= \int\limits_0^x h(u)\, du = \int\limits_0^x \frac{f(u)}{S(u)}\, du \\ &= -\ln S(x). \end{aligned}$$

[28] Some authors define the discrete hazard rate as $\Pr(X = i) \big/ \Pr(X > i)$.

[29] Remember: $\prod\limits_{i=j}^{k} a_i = 1$ for $k < j$.

In the discrete case we have a dilemma when defining the cumulative hazard function, because, in general, summing the hazard rate values h_i — as analogue to integrating $h(x)$ in the continuous case — is not equal to taking the negative of the logarithm of the survival function S_i. Thus, two possible, but different choices for the discrete case exist, which give rise to two estimators in Sect. 6.2:

$$ {}_1H_i := -\ln S_i, \tag{1.62a} $$

called **cumulative hazard function** by KEMP (2004), and

$$ {}_2H_i := \sum_{j=0}^{i} h_j, \tag{1.62b} $$

called **accumulated hazard function** by KEMP.

Excursus: The pseudo–hazard rate

By defining an alternative hazard rate h_i^*, called **pseudo–hazard rate** by ROY/GUPTA (1992), it is possible to have

$$ S_i = \exp(-H_i^*) = \exp\left(-\sum_{j=0}^{i} h_j^*\right). $$

This pseudo–hazard rate reads

$$ h_i^* = \ln\left(\frac{S_{i-1}}{S_i}\right);\ i = 0,\,1,\,2,\,\ldots; $$

where $S_{-1} = 1$. Then

$$ H_i^* = \sum_{j=0}^{i} = \ln S_{-1} - \ln S_i = -\ln S_i. $$

The pseudo–hazard rate h_i^* and the hazard rate h_i are linked as

$$ h_i^* = -\ln(1 - h_{i-1}). $$

COX/OAKES (1984, p. 15) prefer to define the cumulative hazard rate for discrete lifetimes as

$$ H(x) = \sum_{x_j<x} \ln[1 - h(x_j)], $$

because the relationship $S(x) = \exp[-H(x)]$ will be presumed for discrete lifetimes. If the $h(x_j)$ are small, we have for the COX/OAKES definition

$$ H(x) \approx \sum_{x_j<x} h(x_j). $$

${}_1H_i$ and ${}_2H_i$ are linked as follows: From (1.62a) using (1.58a) and (1.61c) we find

$$ {}_1H_i = -\ln\left(\sum_{j\geq i} P_i\right) = -\ln\left[\prod_{j=0}^{i-1}(1-h_j)\right] = -\sum_{j=0}^{i-1}\ln\Big[1 - {}_2H_j + {}_2H_{j-1}\Big]. \tag{1.63a} $$

From (1.62b) using (1.60b) together with (1.58a,b) we have

$$ {}_2H_i = \sum_{j=0}^{i}\left(\frac{P_i}{\sum\limits_{k\geq j} P_k}\right) = \sum_{j=0}^{i}\left(\frac{S_j - S_{j+1}}{S_j}\right) = \sum_{j=0}^{i}\Big[1 - \exp\Big({}_1H_j - {}_1H_{j+1}\Big)\Big]. \tag{1.63b} $$

Expressing S_i P_i and h_i by ${}_1H_i$ we find

$$S_i = \exp(-\,{}_1H_i), \tag{1.64a}$$
$$P_i = \exp(-\,{}_1H_i) - \exp(-\,{}_1H_{i+1}), \tag{1.64b}$$
$$h_i = 1 - \exp({}_1H_i -\,{}_1H_{i+1}), \tag{1.64c}$$

and expressing S_i P_i and h_i by ${}_2H_i$ we have

$$S_i = \prod_{j=0}^{i-1} \bigl(1 -\,{}_2H_j +\,{}_2H_{j-1}\bigr),\ \ {}_2H_{-1} = 0, \tag{1.64d}$$
$$P_i = \bigl({}_2H_i -\,{}_2H_{i-1}\bigr) \prod_{j=0}^{i-1} \bigl(1 -\,{}_2H_j +\,{}_2H_{j-1}\bigr), \tag{1.64e}$$
$$h_i = {}_2H_i -\,{}_2H_{i-1}. \tag{1.64f}$$

For h_i small, we have

$$1 - h_i \approx \exp(-h_i). \tag{1.65a}$$

From this (1.61e) and (1.61c) become

$$P_i \approx h_i \exp(-\,{}_2H_{i-1}), \tag{1.65b}$$
$$S_i \approx \exp(-\,{}_2H_{i-1}). \tag{1.65c}$$

These approximations correspond to the well–known results for continuous variates:

$$f(x) = h(x) \exp\left[-\int_0^x h(u)\,du\right],$$
$$S(x) = \exp\left(-\int_0^x h(u)\,du\right).$$

For a discrete distribution with *increasing hazard rate* the condition $h_0 \le h_1 \le \ldots$ applied to (1.61c) yields the inequality

$$S_i \le (1 - h_0)^i = (1 - P_0)^i \approx \exp(-i\,h_0) = \exp(-i\,P_0). \tag{1.66}$$

This inequality is reversed for a *decreasing hazard rate* distribution.

The mean residual life function is defined by KALBFLEISCH/PRENTICE (1980, p. 7), LAWLESS (1982, p. 44) and LEEMIS (1995, p. 57) as

$$L_i := \mathrm{E}(X - i \,|\, X \ge i),\ \ i \ge 0. \tag{1.67a}$$

Therefore, in the discrete case we have[30]

$$L_i = \frac{\sum\limits_{j\ge i} j\,P_j}{\sum\limits_{j\ge i} P_j} - i, \tag{1.67b}$$
$$= \frac{\sum\limits_{j>i} S_j}{S_i}, \tag{1.67c}$$
$$= \sum_{j\ge i} \prod_{k=i}^{j} (1 - h_k) \tag{1.67d}$$

[30] In (1.67b) the term $\sum\limits_{j\ge i} j\,P_j \Big/ \sum\limits_{j\ge i} P_j$ is the mean age at death of an i–survivor.

$$L_i = \sum_{j>i} \exp\big({}_1H_i - {}_1H_j\big), \tag{1.67e}$$

$$= \sum_{j\geq i} \prod_{k=i}^{j} \big(1 - {}_2H_k + {}_2H_{k-1}\big). \tag{1.67f}$$

From (1.67b) we see that, since the MRL function is defined for all $i \geq 0$, and everything in this expression except i is constant between mass function values, MRL decreases with a slope of -1 at all time values for which there is no mass.

Reverting to (1.66), which holds for an *increasing hazard rate distribution*, we find — see (1.72a) — from $\mathrm{E}(X) = \sum\limits_{j>0} S_j$ the inequality

$$\mathrm{E}(X) \leq \frac{1-h_0}{h_0} = \frac{1-P_0}{P_0}. \tag{1.68}$$

This inequality is reversed for a *decreasing hazard rate distribution*.

What can be said about $L = \lim_{i\to\infty} L_i$? — Let

$$h = \lim_{i\to\infty} h_i,\;\; P = \lim_{i\to\infty}\left\{\frac{P_{i+1}}{P_i}\right\},\;\; S = \lim_{i\to\infty}\left\{\frac{S_i}{S_{i+1}}\right\},$$

then SALVIA/BOLLINGER (1982) stated and proved the following

Theorem 8: If $0 < h < 1$, then

- $L = (S-1)^{-1}$,
- $P = S^{-1}$,
- $h = (L+1)^{-1}$. ■

For $h = 0$ we find

$$S = P = 1 \quad \text{and} \quad L = \infty,$$

and for $h = 1$ we have

$$P = L = 0 \quad \text{and} \quad S = \infty.$$

These extremes do in fact occur. For example, the YULE distribution

$$P_i = \rho\, B(i, \rho+1) = \frac{\rho\,\Gamma(\rho+1)\,\Gamma(i)}{\Gamma(i+\rho+1)};\;\; i = 1,\,2,\ldots;\; \rho \in \mathbb{R}^+;$$

has $h = 0$, and the POISSON distribution

$$P_i = \frac{\lambda^i}{i!}\, e^{-\lambda};\;\; i = 0,\,1,\ldots$$

has $h = 1$, see Example 1/10.

The discrete lifetime distribution representatives $P_i,\; S_i,\; h_i,\; {}_1H_i$, and ${}_2H_i$ may be expressed in terms of L_i. From (1.67c) we have

$$L_i\, S_i = S_{i+1} + S_{i+2} + \ldots$$

and furthermore

$$L_{i-1}\, S_{i-1} = S_i + S_{i+1} + \ldots$$

with difference $L_{i-1}\, S_{i-1} - L_i\, S_i = S_i$. Solving for S_i gives

$$S_i = S_{i-1}\,\frac{L_{i-1}}{1+L_i}, \tag{1.69a}$$

which upon substituting S_{i-1} by $S_{i-2}\,L_{i-2}\big/(1+L_{i-1})$, and so forth until $S_0\,L_0\,(1-L_1) = L_0\,(1-L_1)$, results into

$$S_i = \prod_{j=0}^{i-1} \frac{L_j}{1+L_{j+1}}. \tag{1.69b}$$

From (1.69b) we have

- together with (1.61b):

$$h_i = 1 - \frac{L_i}{1+L_{i+1}}, \tag{1.69c}$$

- together with (1.69c) and (1.61d):

$$P_i = \left(1 - \frac{L_i}{1+L_{i+1}}\right) \prod_{j=0}^{i-1} \frac{L_j}{1+L_{j+1}}, \tag{1.69d}$$

- together with (1.62a):

$${}_1H_i = -\sum_{j=0}^{i-1} \ln\left(\frac{L_j}{1+L_{j+1}}\right), \tag{1.69e}$$

- and together with (1.69c) and (1.62b):

$${}_2H_i = (i+1) - \sum_{j=0}^{i} \frac{L_j}{1+L_{j+1}}. \tag{1.69f}$$

(1.69c) is the basis for a recursion formula

$$L_{i+1} = \frac{L_i}{1-h_i} - 1. \tag{1.70}$$

In the continuous case it does not matter whether we define MRL by $\mathrm{E}(X-x\,|\,X>x)$ or by $\mathrm{E}(X-x\,|\,X \geq x)$, but in the discrete case there is a difference. MRL defined as

$$\mu(i) = \mathrm{E}(X-i\,|\,X>i),\;\; i>0, \tag{1.71a}$$

is different from(1.67a). We have

$$\mu(i) = L_{i+1} + 1,\;\; i>0. \tag{1.71b}$$

With (1.67a) the MRL function at time $i=0$ is the mean of the lifetime distribution:

$$L_0 = \sum_{j\geq 0} j\,P_j = \sum_{j>0} S_j = \mathrm{E}(X) \tag{1.72a}$$

whilst

$$\mu(0) = L_1 + 1 = \frac{\mathrm{E}(X)}{1-P_0}. \tag{1.72b}$$

For this reason L_i should be preferred over $\mu(i)$.

In Tab. 1/2 we have collected all the relations between the six representatives of a discrete lifetime distribution.

Table 1/2: Relations among the six functions describing a discrete stochastic lifetime

from \ to	P_i	S_i	h_i	${}_1H_i$	${}_2H_i$	L_i
P_i	–	$\sum\limits_{j\ge i} P_j$	$\dfrac{P_i}{\sum\limits_{j\ge i} P_j}$	$-\ln\Big(\sum\limits_{j\ge i} P_j\Big)$	$\sum\limits_{j=0}^{i} \dfrac{P_j}{\sum\limits_{k\ge j} P_k}$	$\dfrac{\sum\limits_{j\ge i} j\,P_j}{\sum\limits_{j\ge i} P_j} - i$
S_i	$S_i - S_{i+1}$	–	$1 - \dfrac{S_{i+1}}{S_i}$	$-\ln S_i$	$\sum\limits_{j=0}^{i} \dfrac{S_j - S_{j+1}}{S_j}$	$\dfrac{\sum\limits_{j>i} S_j}{S_i}$
h_i	$h_i \prod\limits_{j=0}^{i-1}(1-h_j)$	$\prod\limits_{j=0}^{i-1}(1-h_j)$	–	$-\sum\limits_{j=0}^{i-1}\ln(1-h_j)$	$\sum\limits_{j=0}^{i} h_j$	$\sum\limits_{j\ge i}\prod\limits_{k=i}^{j}(1-h_k)$
${}_1H_i$	$\exp(-\,{}_1H_i) - \exp(-\,{}_1H_{i+1})$	$\exp(-\,{}_1H_i)$	$1-\exp({}_1H_i - {}_1H_{i+1})$	–	$-\sum\limits_{j=0}^{i-1}\big[1-\exp({}_1H_j - {}_1H_{j+1})\big]$	$\sum\limits_{j>i}\exp({}_1H_i - {}_1H_j)$
${}_2H_i$	$({}_2H_i - {}_2H_{i-1}) \times \prod\limits_{j=0}^{i-1}(1 - {}_2H_j + {}_2H_{j-1})$	$\prod\limits_{j=0}^{i-1}(1 - {}_2H_j + {}_2H_{j-1})$	${}_2H_i - {}_2H_{i-1}$	$-\sum\limits_{j=0}^{i-1}\ln(1 - {}_2H_j + {}_2H_{j-1})$	–	$\sum\limits_{j\ge i}\prod\limits_{k=i}^{j}(1 - {}_2H_k + {}_2H_{k-1})$
L_i	$\Big(1-\dfrac{L_i}{1+L_{i+1}}\Big) \times \prod\limits_{j=0}^{i-1}\Big(\dfrac{L_j}{1+L_{j+1}}\Big)$	$\prod\limits_{j=0}^{i-1}\dfrac{L_j}{1+L_{j+1}}$	$1-\dfrac{L_i}{1+L_{i+1}}$	$-\sum\limits_{j=0}^{i-1}\ln\Big(\dfrac{L_j}{1+L_{j+1}}\Big)$	$(i+1)-\sum\limits_{j=0}^{i}\dfrac{L_j}{1+L_{j+1}}$	–

Example 1/10: Hazard rate of the POISSON distribution

The POISSON distribution has PMF

$$P_i = \frac{\lambda^i}{i!}\,e^{-\lambda}; i = 0,\,1,\,\ldots;\ \lambda > 0. \tag{1.73a}$$

Applying (1.60b) to (1.73a) we find

$$h_i = \left(1 + \frac{\lambda}{i+1} + \frac{\lambda^2}{i+2} + \ldots\right)^{-1};\ i = 0,\,1,\,\ldots, \tag{1.73b}$$

so

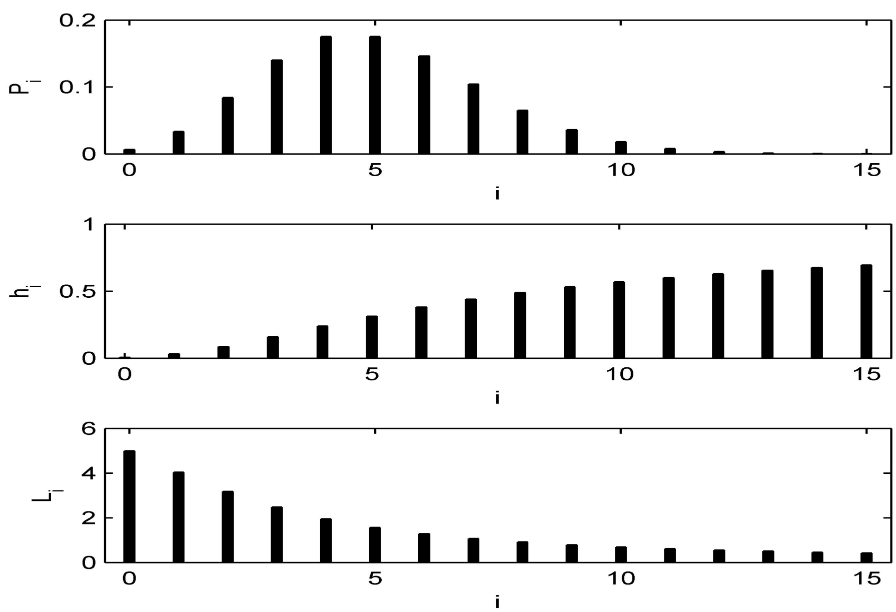
$$\lim_{i\to\infty}\{h_i\} = 1 \tag{1.73c}$$

and

$$e^{-\lambda} \le h_i \le 1. \tag{1.73d}$$

Fig. 1/11 shows the probabilities P_i, the hazard rate values h_i and the mean residual life values L_i for $\lambda = 5$. We have an increasing HR and a decreasing MRL, a result which does not depend on the value of λ. For computing MRL we have used the recursion (1.70) starting with $L_0 = \mathrm{E}(X) = \lambda$.

Figure 1/11: PMF, HR and MRL of the POISSON distribution with $\lambda = 5$



Excursus: Results for a distribution having both discrete and continuous components

A lifetime distribution may have both discrete and continuous components, e.g., a device has a non–zero probability of failing on switching on and/or off whereas otherwise the chance of failing has a continuous density. In the mixed case the density is a sum of discrete and continuous parts. Specifically, if $h_c(x)$ denotes the hazard rate for the continuous part and mass points occur at $x_i;\ i = 0,\,1,\ldots$; then the overall survivor function can be written by the so–called ‘pseudo–integral’ as

$$S(x) = \exp\left[-\int_0^x h_c(u)\,du\right]\prod_{j=0}^{i-1}(1 - h_j). \tag{1.74a}$$

The corresponding HR is

$$h(x)\,dx = h_c(x)\,dx + \sum_j h_j\,\delta(x - x_j) \tag{1.74b}$$

where $\delta(x - x_j)$ is the **DIRAC function**

$$\delta(y)\,dy = \left\{\begin{array}{ll} 1 & \text{for} \quad y = 0 \\ 0 & \text{otherwise.} \end{array}\right\} \tag{1.74c}$$

The cumulative hazard function is now defined as

$$H(x) = \int_0^x h(u)\,du = \int_0^x h_c(u)\,du + \sum_{x_i < x} h_i. \tag{1.74d}$$

1.2.2 Special Results[31]

As in the continuous case it is possible to specify a discrete distribution by defining a suitable hazard rate. In the section we will present four such approaches.

The first approach with a **constant discrete hazard rate** takes

$$h_i = c \;\; \forall\, i,\; 0 < c < 1. \tag{1.75a}$$

From (1.61e) we find

$$P_i = c\,(1 - c)^i;\; i = 0,\; 1,\; \ldots \tag{1.75b}$$

This is recognized as the **geometric distribution**, and it is, of course the discrete analogue of the exponential distribution, both having a constant hazard rate and the memoryless property. For the geometric distribution we note

$$S_i = (1 - c)^i, \tag{1.75c}$$

$${}_1H_i = c\,(i + 1), \tag{1.75d}$$

$${}_2H_i = -i\,\ln(1 - c), \tag{1.75e}$$

$$L_i = \frac{1}{c}. \tag{1.75f}$$

SALVIA/BOLLINGER (1982) proposed the following **decreasing discrete hazard rate**

$$h_i = \frac{c}{i + 1};\;\; i = 0,\; 1, \ldots;\; 0 < c < 1.$$

PADGETT/SPURRIER(1985) generalized this model by introducing a second parameter $\alpha,\ \alpha \geq 0$:

$$h_i = \frac{c}{\alpha\,i + 1};\;\; i = 0,\; 1, \ldots;\; 0 < c < 1;\; \alpha \geq 0, \tag{1.76a}$$

resulting into

$$P_i = c\,\frac{\prod\limits_{j=0}^{i-1}(\alpha\,j - c + 1)}{\prod\limits_{j=0}^{i}(\alpha\,j + 1)}, \tag{1.76b}$$

$$S_i = \prod_{j=0}^{i-1}\frac{\alpha\,j - c + 1}{\alpha\,j + 1}. \tag{1.76c}$$

[31] Suggested reading for this section: PADGETT/SPURRIER (1985), RINNE (2009, pp. 119–125), SALVIA/BOLLINGER (1982).

As $c \to 1$, the model approaches the **degenerate distribution** with $P_0 = 1$. α is a shape parameter and with respect to the parameter c we have: $c = \Pr(X = 0)$. For $\alpha = 1$ we have the original model of SALVIA/BOLLINGER, and for $\alpha = 0$ we are back to the geometric distribution with constant hazard rate $h_i = c \ \forall\, i$.

The **increasing discrete hazard rate distribution** of SALVIA/BOLLINGER has

$$h_i = 1 - \frac{c}{i+1};\ \ i = 0,\ 1, \ldots;\ 0 < c \leq 1,$$

which has been generalized by PADGETT/SPURRIER to

$$h_i = 1 - \frac{c}{\alpha\, i + 1};\ \ i = 0,\ 1, \ldots;\ 0 < c < 1;\ \alpha \geq 0. \tag{1.77a}$$

Corresponding to (1.77a) we have

$$P_i = \frac{c^i\,(\alpha\, i - c + 1}{\prod\limits_{j=0}^{i}(\alpha\, j + 1)}, \tag{1.77b}$$

$$S_i = \frac{c^i}{\prod\limits_{j=0}^{i}(\alpha\, j + 1)}. \tag{1.77c}$$

As $\alpha \to \infty$ this model approaches the **BERNOULLI distribution** with parameter $P = 1 - c$ and $Q = 1 - P = c$. As $c \to 0$ the model approaches the **degenerate distribution** with $P_0 = 1$. α is a shape parameter and c is corresponds to $\Pr(X > 0)$. $\{P_i\}$ is a non–decreasing series if

- $c > 0.5$ and
- $\alpha \leq \dfrac{(c-1)^2}{2\,c - 1}$.

Otherwise, the PMF first increases and then decreases. For $\alpha = 1$ we have the original model of SALVIA/BOLLINGER, and for $\alpha = 0$ we are back to the geometric distribution, but with hazard rate $h_i = 1 - c \ \forall\, i$.

There are several discrete models which — depending on the value of a certain parameter — have a constant, an increasing or a decreasing hazard rate, respectively, like the continuous WEIBULL distribution with hazard rate $h(x) = \dfrac{c}{b}\left(\dfrac{x}{b}\right)^{c-1}$; $x \geq 0,\ b > 0,\ c > 0$; see RINNE (2009). The **type–I discrete WEIBULL model** of NAKAGAWA/OSAKI (1975) has

$$h_i = 1 - q^{(i+1)^\beta - i^\beta};\ i = 0,\ 1,\ \ldots;\ 0 < q < 1;\ \beta > 0; \tag{1.78a}$$

with corresponding

$$P_i = q^{i^\beta} - q^{(i+1)^\beta}, \tag{1.78b}$$

$$S_i = q^{i^\beta}. \tag{1.78c}$$

The parameter β plays the same role as the shape parameter c in the continuous WEIBLL distribution.

- h_i is constant with value $1 - q$ for $\beta = 1$,
- h_i is decreasing for $0 < \beta < 1$,
- h_i is increasing for $\beta > 1$.

The **type–II discrete WEIBULL distribution** of STEIN/DATTERO (1984) directly mimics the continuous hazard rate $\frac{c}{b}\left(\frac{x}{b}\right)^{c-1}$ by

$$h_i = \left\{ \begin{array}{cl} \alpha\, i^{\beta-1} & \text{for } i = 1,\, 2, \dots \\ 0 & \text{for } i = 0 \text{ and } i > m \end{array} \right\},\ \alpha > 0,\ \beta > 0. \tag{1.79a}$$

m is a truncation value given by

$$m = \left\{ \begin{array}{cl} \text{int}\big[\alpha^{-1/(\beta-1)}\big] & \text{for } \beta > 1 \\ \infty & \text{for } \beta \le 1 \end{array} \right\}, \tag{1.79b}$$

which is necessary to ensure $h_i \le 1$. The corresponding PMF and CCDF are

$$P_i = \alpha\, i^{\beta-1} \prod_{j=1}^{i-1} \Big(1 - \alpha\, j^{\beta-1}\Big);\ i = 1,\, 2, \dots \tag{1.79c}$$

$$S_i = \prod_{j=1}^{i-1} \Big(1 - \alpha\, j^{\beta-1}\Big). \tag{1.79d}$$

The hazard rate is

- constant with value α for $\beta = 1$,
- decreasing for $0 < \beta < 1$,
- increasing for $\beta > 1$.

The **type–III discrete WEIBULL distribution** of PADGETT/SPURRIER (1985) has

$$h_i = 1 - \exp\Big[-d\,(i+1)^{\beta}\Big];\ i = 0,\, 1, \dots;\ d > 0,\ \beta \in \mathbb{R} \tag{1.80a}$$

with corresponding

$$P_i = \exp\left(-d \sum_{j=1}^{i} j^{\beta}\right) \Big\{1 - \exp\Big[-d\,(i+1)^{\beta}\Big]\Big\}, \tag{1.80b}$$

$$S_i = \exp\left[-d \sum_{j=1}^{i} j^{\beta}\right]. \tag{1.80c}$$

The hazard rate is

- constant with value $1 - \exp(-d)$ for $\beta = 0$,
- decreasing for $\beta < 0$,
- increasing for $\beta > 0$.

1.3 The Multivariate Cases[32]

We will only give short comments on the case of multivariate lifetime distributions for several reasons:

1. Realistic and tractable multivariate lifetime models are scarce.
2. The hazard rate concept and the mean residual life concept are somewhat difficult to extend to the multivariate situation.
3. A third difficulty is that often the sample is not big enough in relation to the dimension of the model in order to find 'good' estimates of the model and its parameters. Furthermore, many data are censored in such a way that one cannot determine whether or not the variates are independent.

Sometimes two or more lifetime variables $X_1, \ldots, X_m$ are of interest simultaneously and a multivariate model is required. For example, a device may have two or more integral parts and it may be desired to model the joint lifetime distribution of these parts. Let

$$\boldsymbol{X} = (X_1, \ldots, X_m)'; \; m = 2, 3, \ldots$$

be the (column) vector of variates and

$$\boldsymbol{x} = (x_1, \ldots, x_m)'$$

a vector of its realizations. Then, a multivariate distribution can be specified either in terms of the **joint survival function**

$$S(\boldsymbol{x}) := \Pr(\boldsymbol{X} > \boldsymbol{x}) = \Pr(X_1 > x_1, \ldots, X_m > x_m) \tag{1.81a}$$

or in terms of **joint failure (distribution) function**

$$F(\boldsymbol{x}) := \Pr(\boldsymbol{X} \le \boldsymbol{x}) = \Pr(X_1 \le x_1, \ldots, X_m \le x_m) \tag{1.81b}$$

or — in the continuous case — in terms of the **joint failure density**

$$f(\boldsymbol{x}) := \frac{\partial^m F(\boldsymbol{x})}{\partial x_1 \ldots \partial x_m} = -\frac{\partial^m S(\boldsymbol{x})}{\partial x_1 \ldots \partial x_m} \tag{1.81c}$$

or — in the discrete case — in terms of the **joint probability mass function**

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \Pr(X_1 = x_1, \ldots X_m = x_m); \; (x_1, \ldots x_m) \in \mathbb{N}_0^m. \tag{1.81d}$$

In fortunate circumstances $X_1, \ldots, X_m$ can be assumed to be independent and the joint functions in (1.81a–d) can be written as products of the one–dimensional marginal functions $S_i(x_i),\; F_i(x_i),\; f_i(x_i)$ or $\Pr(X_i = x_i)$, respectively. In this case one is effectively back in the univariate framework.

[32] Suggested reading for the section: ASADIA (1999), COX (1972); JOHNSON/KOTZ (1975), MCGILL (1992), SHAKED et al. (1995).

1.3.1 Continuous Distributions

In the context of defining the concepts IHR and DHR (increasing and decreasing hazard rate), see Sect. 2.1, for multivariate distributions we find several attempts in the pertaining literature. Some authors like HARRIS (1970) or BRINDLEY/THOMPSON (1972) give no explicit definition of a multivariate hazard rate. So BRINDLEY/THOMPSON state that a multivariate distribution with $S(\boldsymbol{x})$ defined on the positive orthant is IHR (DHR) if

$$\Pr(x_1 > x + \Delta, \ldots, X_m > x_m + \Delta \,|\, X_1 > x_1, \ldots, X_m > x_m) = \frac{S(\boldsymbol{x} + \Delta)}{S(\boldsymbol{x})}$$

is decreasing (increasing) in $\boldsymbol{x}$ for each $\Delta > 0$ and all $\boldsymbol{x} \geq \mathbf{0}$ such that $S(\boldsymbol{x}) > 0$. Other authors like BASU (1971) or PURI/RUBIN (1974) define the multivariate hazard rate as a scalar quantity. In the bivariate case BASU gives the hazard rate as

$$\begin{aligned} h(x_1, x_2) &= \frac{f(x_1, x_2)}{\Pr(X_1 > x_1, X_2 > x_2)} \\ &= \frac{f(x_1, x_2)}{1 - F(x_1, \infty) - F(\infty, x_2) + F(x_1, x_2)}. \end{aligned} \tag{1.82a}$$

For independent variates $X_1,\ X_2$ (1.82a) turns into

$$\begin{aligned} h(x_1, x_2) &= \frac{f(x_1, x_2)}{\Pr(X_1 > x_1)\Pr(X_2 > x_2)} \\ &= \frac{f(x_1)}{S(x_1)}\,\frac{f(x_2)}{S(x_2)} \\ &= h(x_1)\,h(x_2) \end{aligned} \tag{1.82b}$$

where $f(\cdot),\ S(\cdot),\ h(\cdot)$ are the marginal functions, respectively. (1.82a) may easily be extended to the case of more than two variates.

Some authors like JOHNSON/KOTZ (1975) take the point of view that, for a concept such as 'multivariate hazard rate', it is unreasonable to expect a single value to represent this aspect of a multivariate distribution. The basic idea underlying the univariate definition is that of rate of decrease of 'survivors' with increase in value x of X as, e.g., in a life table where the hazard rate is in fact the force of mortality. When there are two or more variates this rate depends on which variate is changed and we need a different rate for each variate. So, JOHNSON/KOTZ defined the **joint multivariate hazard rate** of m absolutely continuous variables $X_1, \ldots, X_m$ as the vector

$$\begin{aligned} h_{\boldsymbol{X}}(\boldsymbol{x}) &:= \big[-(\partial/\partial x_1), \ldots, -(\partial/\partial x_m)\big] \ln S(\boldsymbol{x}) \\ &= -\mathrm{grad} \ln S(\boldsymbol{x}). \end{aligned} \tag{1.83a}$$

Sometimes $h_{\boldsymbol{X}}(\boldsymbol{x})$ is called the **hazard gradient** of $\boldsymbol{X}$. For convenience we will write a component of the vector $h_{\boldsymbol{X}}(\boldsymbol{x})$ as

$$h_i(\boldsymbol{x}) := -\left(\frac{\partial}{\partial x_i}\right) \ln S(\boldsymbol{x});\ i = 1, \ldots, m. \tag{1.83b}$$

(1.83a,b) are motivated by the fact that in the univariate case we have

$$h(x) = \frac{dH(x)}{dx} = -\frac{d \ln S(x)}{dx},$$

see (1.11a,f).

If the multivariate hazard rate (1.83a) is constant, i.e., does not vary with *any* of $x_1, \ldots, x_m$, so that $h_{\boldsymbol{X}}(\boldsymbol{x}) = \boldsymbol{c}$, this means that, whenever the hazard rate exists, we have $\partial \ln S(\boldsymbol{x})/\partial x_i =$

$-c_i$ $(i = 1, \ldots, m)$. Hence, $S(\boldsymbol{x}) = \exp(-c_i\,x_i)\,s_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m)$; $i = 1, \ldots, m$; whence

$$S(\boldsymbol{x}) \propto \exp\biggl(-\sum_{i=1}^{m} c_i\,x_i\biggr).$$

Thus, the X_i are mutually independent exponential variables if and only if the multivariate hazard rate is constant. We may distinguish between **strictly constant vector hazard rates** $\bigl[h_{\boldsymbol{X}}(\boldsymbol{x}) = \boldsymbol{c}\bigr]$ as defined above, and **locally constant vector hazard rates** for which $h_i(\boldsymbol{x})$ does not depend on x_i, though it may depend on other x's.

Example 1/11: Bivariate exponential distributions[33]

A bivariate exponential distribution, **BED** for short, has both marginal distributions as exponential. As KOTZ/BALAKRISHNAN/JOHNSON (2000) show, many of such BEDs exist. We will present these BEDs in their standard form, but location and scale parameters can easily be introduced, if needed, through appropriate linear transformations.

GUMBEL's BED has the joint survival function

$$S(\boldsymbol{x}) = \exp(-x_1 - x_2 - \theta\,x_1\,x_2);\; x_1,\, x_2 \ge 0,\; \theta > 0; \tag{1.84a}$$

the joint density function

$$f(\boldsymbol{x}) = \exp(-x_1 - x_2 - \theta\,x_1\,x_2)\,\bigl[(1+\theta\,x_1)\,(1+\theta\,x_2) - \theta\bigr]; \tag{1.84b}$$

and the conditional PDF of X_2, given $X_1 = x_1$,

$$f(x_2\,|\,x_1) = \exp\bigl[-(1+\theta\,x_1)\,x_2\bigr]\bigl\{(1+\theta\,x_1)\,(1+\theta\,x_2) - \theta\bigr\}. \tag{1.84c}$$

The latter is not exponential whereas the marginal distributions of each X_1 and X_2 are standard exponential. If $\theta = 0$, then X_1 and X_2 are mutually independent.

The joint multivariate hazard rate of GUMBEL's BED is

$$h_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{pmatrix} h_1(\boldsymbol{x}) \\ h_2(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} 1+\theta\,x_2 \\ 1+\theta\,x_1 \end{pmatrix}. \tag{1.84d}$$

The components in (1.84d) are constant with respect to variation in the corresponding variable, i.e., $h_1(\boldsymbol{x})$ does not depend on x_1 nor does $h_2(\boldsymbol{x})$ on x_2, but not with respect to variation in the other variable. So, the distribution of $\boldsymbol{X} = (X_1,\, X_2)$ has a locally, but not strictly constant bivariate hazard rate, see the graphs on the right–hand side of Fig. 1/12.

The scalar multivariate hazard rate (1.82a) for GUMBEL's BED using

$$\begin{aligned} F(x_1, x_2) &= 1 - \exp(-x_1) - \exp(-x_2) + \exp[-x_2 - x_1\,(1+\theta\,x_2)], \\ F(x_1, \infty) &= 1 - \exp(-x_1), \\ F(\infty, x_2) &= 1 - \exp(-x_2), \end{aligned}$$

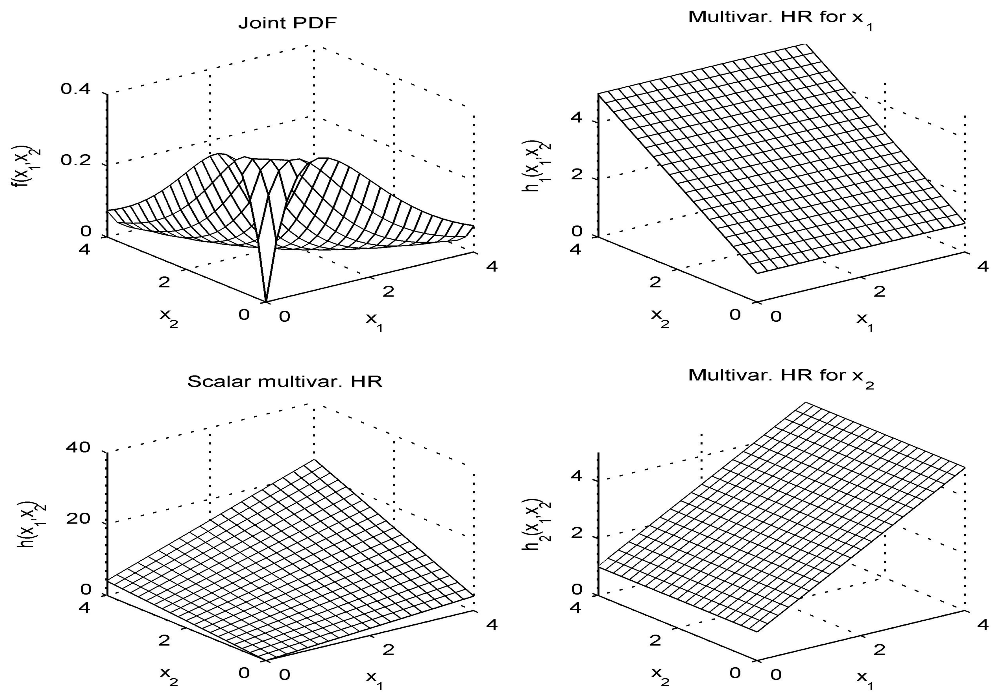
results in

$$h(x_1, x_2) = (1+\theta\,x_1)\,(1+\theta\,x_2) - \theta. \tag{1.84e}$$

Fig. 1/12 displays four functions of GUMBEL's BED with parameter $\theta = 1$. In the upper left corner we have the PDF of (1.84b) and in the lower left corner the scalar multivariate HR of (1.84e). On the right–hand side we see the first (upper graph) and the second (lower graph) component of the joint multivariate HR of (1.84d).

[33] Results for the multivariate normal distribution can be found in GUPTA/GUPTA (1997), MA (2000), MCGILL (1992), NAVARRO/RUIZ (2004).

Figure 1/12: GUMBEL's bivariate exponential distribution with $\theta = 1$



Another BED is that of MARSHALL/OLKIN (2007). The physical model consists of two components, subjected to shocks that are always fatal. These shocks are assumed to be governed by independent POISSON processes with parameters $\lambda_1,\ \lambda_2$ and λ_{12}, according as the shock applies to component 1 only, component 2 only, or both components, respectively. The joint survival function of the lifetimes X_1 and X_2 of the two components is

$$S(\boldsymbol{x}) = \exp\Big[-\lambda_1\,x_1-\lambda_2\,x_2-\lambda_{12}\,\max(x_1,\,x_2)\Big];\ \lambda_1,\,\lambda_2,\,\lambda_{12}>0;\ x_1,\,x_2\ge 0 \qquad (1.85a)$$

$$= \left\{\begin{array}{ll}\exp\big[-\lambda_1\,x_1-(\lambda_2+\lambda_{12})\,x_2\big] & \text{for } 0\le x_1\le x_2\\ \exp\big[-(\lambda_1+\lambda_{12})\,x_1-\lambda_2\,x_2\big] & \text{for } 0\le x_2\le x_1.\end{array}\right\} \qquad (1.85b)$$

The marginal distributions are genuine one–dimensional exponential distributions:

$$\left.\begin{array}{rcl}S(x_1,\infty) & = & \exp\big[-(\lambda_1+\lambda_{12})\,x_1\big],\\ S(\infty,x_2) & = & \exp\big[-(\lambda_2+\lambda_{12})\,x_2\big].\end{array}\right\} \qquad (1.85c)$$

The probability that a failure on component i occurs first is

$$\Pr(X_i<X_j)=\frac{\lambda_i}{\lambda_1+\lambda_2+\lambda_{12}};\ i,\,j=1,\,2;\ i\ne j; \qquad (1.85d)$$

and we have a positive probability that both components fail simultaneously:

$$\Pr(X_1=X_2)=\frac{\lambda_{12}}{\lambda_1+\lambda_2+\lambda_{12}}. \qquad (1.85e)$$

(1.85e also is the correlation coefficient between X_1 and X_2. The joint PDF of this BED — using $S(\boldsymbol{x})$ of (1.85a) — is

$$f(\boldsymbol{x})=\left\{\begin{array}{ll}\lambda_2\,(\lambda_1+\lambda_{12})\,S(\boldsymbol{x}) & \text{for}\quad 0<x_2<x_1\\ \lambda_1\,(\lambda_2+\lambda_{12})\,S(\boldsymbol{x}) & \text{for}\quad 0<x_1<x_2\\ \lambda_{12}\,S(\boldsymbol{x}) & \text{for}\quad x_1=x_2>0.\end{array}\right\} \qquad (1.85f)$$

The joint multivariate hazard rate results as

$$h_{\boldsymbol{X}}(\boldsymbol{x}) = \left\{ \begin{array}{ll} \begin{pmatrix} \lambda_1 \\ \lambda_2 + \lambda_{12} \end{pmatrix} & \text{for} \quad x_1 < x_2 \\ \begin{pmatrix} \lambda_1 + \lambda_{12} \\ \lambda_2 \end{pmatrix} & \text{for} \quad x_1 > x_2. \end{array} \right\} \tag{1.85g}$$

These hazard rates are strictly increasing, but for x_2 (x_1) fixed, the first (second) component is a non–decreasing function of x_1 (x_2).

RINNE (2009, p. 173 ff.) has extended the two models above and further BEDs by power transformation to bivariate WEIBULL distributions.

The **joint multivariate mean residual life**, defined by ARNOLD/ZAHEDI (1988), is the vector

$$\mu_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{pmatrix} \mu_1(\boldsymbol{x}) \\ \vdots \\ \mu_m(\boldsymbol{x}) \end{pmatrix} = \mathrm{E}(\boldsymbol{X} - \boldsymbol{x} \,|\, \boldsymbol{X} \geq \boldsymbol{x}) \tag{1.86a}$$

where

$$\begin{aligned} \mu_i(\boldsymbol{x}) &= \mathrm{E}(X_i - x_i \,|\, \boldsymbol{X} \geq \boldsymbol{x}) \\ &= \frac{\int\limits_0^\infty S(x_1, \ldots, x_{i-1}, x_i + u, x_{i+1}, \ldots, x_m)\, du}{S(\boldsymbol{x})};\ i = 1, 2, \ldots, m \end{aligned} \tag{1.86b}$$

whenever $S(\boldsymbol{x}) > 0$, see also (1.15b). It can be shown easily that the following relationship holds between the $h_i(\boldsymbol{x})$'s of $h_{\boldsymbol{X}}(\boldsymbol{x})$, see (1.83b), and the $\mu_i(\boldsymbol{x})$'s of $\mu_{\boldsymbol{X}}(\boldsymbol{x})$:

$$\frac{\partial}{x_i}\mu_i(\boldsymbol{x}) = \mu_i(\boldsymbol{x})\, h_i(\boldsymbol{x}) - 1;\ i, 2, \ldots, m. \tag{1.86c}$$

Looking at GUMBEL's BED of Example(1/11) we find

$$\mu_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{pmatrix} \dfrac{1}{1 + \theta\, x_2} \\ \dfrac{1}{1 + \theta\, x_1} \end{pmatrix},$$

showing that $\mu_i(\boldsymbol{x})$ is constant with respect to x_i, but decreases with respect to the other variable. For GUMBEL's BED we have

$$h_1(\boldsymbol{x})\, \mu_1(\boldsymbol{x}) = h_2(\boldsymbol{x})\, \mu_2(\boldsymbol{x}) = 1.$$

Another multivariate hazard rate concept has been proposed by COX (1972) viewing the multivariate lifetime as a point process. In the bivariate case we have the following four components of the hazard rate vector:

$$\lambda_i = \lim_{\Delta x \to 0} \frac{\Pr(x \leq X_i < x + \Delta x \,|\, X_1 \geq x\, X_2 \geq x)}{\Delta x};\ i = 1, 2; \tag{1.87a}$$

$$\lambda_{12}(x_1 \,|\, x_2) = \lim_{\Delta x \to 0} \frac{\Pr(x_1 \leq X_1 < x_1 + \Delta x \,|\, X_1 \geq x_1\, X_2 = x_2)}{\Delta x};\ x_1 > x_2; \tag{1.87b}$$

$$\lambda_{21}(x_2 \,|\, x_1) = \lim_{\Delta x \to 0} \frac{\Pr(x_2 \leq X_2 < x_2 + \Delta x \,|\, X_1 = x_1\, X_2 \geq x_2)}{\Delta x};\ x_1 < x_2. \tag{1.87c}$$

In terms of the joint survivor function $S(x_1, x_2)$ for X_1 and X_2 it is readily seen that

$$\lambda_1(x) = -\frac{\partial S(x_1,x_2)/\partial x_1}{S(x_1,x_2)}\bigg|_{x_1=x_2=x}, \tag{1.87d}$$

$$\lambda_{12}(x_1)\,|\,x_2) = -\frac{\partial^2 S(x_1,x_2)/\partial x_1 \partial x_2}{\partial S(x_1,x_2)/\partial x_2},\ x_1 > x_2, \tag{1.87e}$$

with similar expressions for $\lambda_2(x)$ and $\lambda_{21}(x_2\,|\,x_1)$. The functions (1.87a–c) completely specify the joint distribution of X_1 and X_2. The joint PDF of X_1 and X_2 can be shown to be

$$\left.\begin{array}{l} \lambda_2(x_2)\,\lambda_{12}(x_1\,|\,x_2)\,\exp\left\{-\int\limits_0^{x_2}\big[\lambda_1(u)+\lambda_2(u)\big]\,du - \int\limits_{x_2}^{x_1}\lambda_{12}(u\,|\,x_2)\,du\right\},\ x_1 \geq x_2 \\ \lambda_1(x_1)\,\lambda_{21}(x_2\,|\,x_2)\,\exp\left\{-\int\limits_0^{x_1}\big[\lambda_1(u)+\lambda_2(u)\big]\,du - \int\limits_{x_1}^{x_1}\lambda_{12}(u\,|\,x_1)\,du\right\},\ x_1 \leq x_2. \end{array}\right\} \tag{1.87f}$$

This can be verified by viewing the process as a point process. For example, with $x_1 \geq x_2$, the probability of having no failures in $[0,\,x_2)$ and then the event $X_2 \in [x_2,\,x_2+\Delta x_2)$ is

$$\lambda_2(x_2)\,\Delta x_2\,\exp\left\{-\int\limits_0^{x_2}[\lambda_1(u)+\lambda_2(u)]\,du\right\}.$$

Conditional on this, the probability of no further failures in $[x_2,\,x_1)$ and the event $x_1 \in [x_1,\,x_1 + \Delta x_1)$ is

$$\lambda_{12}(x_1\,|\,x_2)\,\Delta x_1\,\exp\left\{-\int\limits_{x_2}^{x_1}\lambda_{12}(u\,|\,x_2)\,du\right\}.$$

Multiplying these probabilities we get the first line of (1.87f).

1.3.2 Discrete Distributions

We will briefly comment on bivariate distributions. Results for the case of more than two discrete variates can be found in SHAKED et al. (1995). Let $\boldsymbol{X} = (X_1,\,X_2)$ be a random vector with support in $\mathbb{N}_0^2$ and denote its joint PMF by

$$\Pr(x_1,\,x_2) = \Pr(X_1 = x_1,\,X_2 = x_2),\ (x_1,\,x_2) \in \mathbb{N}_0^2. \tag{1.88a}$$

Remember, that in the discrete case the hazard rat is a conditional probability. The discrete multivariate conditional hazard rate functions of $(X_1,\,X_2)$ are defined as

$$\lambda_1(x) = \Pr(X_1 = x,\,X_2 > x\,|\,X_1 \geq x_1,\,X_2 \geq x),\ x \in \mathbb{N}_0, \tag{1.88b}$$

$$\lambda_2(x) = \Pr(X_2 = x,\,X_1 > x\,|\,X_1 \geq x_1,\,X_2 \geq x),\ x \in \mathbb{N}_0, \tag{1.88c}$$

$$\lambda_{12}(x) = \Pr(X_1 = x,\,X_2 = x\,|\,X_1 \geq x_1,\,X_2 \geq x),\ x \in \mathbb{N}_0, \tag{1.88d}$$

$$\lambda_1(x\,|\,x_2) = \Pr(X_1 \geq x\,|\,X_2 \geq x,\,X_2 = x_2),\ x > x_2,\ (x_1,\,x_2) \in \mathbb{N}_0^2, \tag{1.88e}$$

$$\lambda_2(x\,|\,x_1) = \Pr(X_2 \geq x\,|\,X_1 = x_1,\,X_2 \geq x),\ x > x_1,\ (x_1,\,x_2) \in \mathbb{N}_0^2, \tag{1.88f}$$

provided the conditions in the above conditional probabilities have positive probabilities. Otherwise, these functions are set to 1.

The meaning of these functions is as follows. The functions $\lambda_1(x)$, $\lambda_2(x)$ and $\lambda_{12}(x)$ describe the *initial* hazard rates, i.e., the hazard rates before a failure of any component. Given that no failure has occurred before time x, then, at time x one of the following four events must occur:

1. only component 1 fails, the probability being $\lambda_1(x)$,
2. only component 1 fails, the probability being $\lambda_2(x)$,
3. both components fail, the probability being $\lambda_{12}(x)$,
4. no component fails, the probability being $1 - \lambda_1(x) - \lambda_2(x) - \lambda_{12}(x)$.

Now suppose that one component failed at x_1 (or x_2) and that the other component stayed alive at that time. Then, conditional on $X_1 = x_1$ (or $X_2 = x_2$), the hazard rate of the live component at time $x > x_1$ (or $x > x_2$) is given by $\lambda_2(x \mid x_1)$ [or $\lambda_1(x \mid x_2)$].

The hazard rates given in (1.88b–f) are the discrete analogs of the bivariate hazard rate functions described in COX (1972), see (1.87a–c) for the absolute continuous case. But in COX there is no analog of (1.88d) since absolute continuity of the distribution of $(X_1,\, X_2)$ is assumed there. Failure of both components at the same time has zero probability in the absolute continuous case, but in the discrete case it may be positive and is given by $\lambda_{12}(x)$.

From (1.88b–f) we see that the joint distribution of X_1 and X_2 determines the conditional hazard rate functions. But also the converse is true, i.e., (1.88b–f) determine Pr$(x_1,\, x_2)$ of (1.88a). For more details see SHAKED et al. (1995) who also give necessary and sufficient conditions on the functions (1.88b–f) which ensure that they are hazard rate functions of some random vector $(X_1,\, X_2)$.

2 Aging Criteria and Classes of Univariate Lifetime Distributions[1]

It is quite natural and obvious to classify lifetime distributions by using so–called aging criteria. In the context of lifetime analysis **aging** does not mean that a statistical unit becomes older in the sense of chronological calendarian time, rather aging is a notion pertaining to the behavior of residual life. Aging is thus the phenomenon that a chronological older unit has a shorter residual life in some statistical sense than a newer or chronological younger unit. We may distinguish between

- **positive** (true or adverse) **aging** indicating a decline — in some way or the other — of residual life with growing age x.
- **negative** (inverse or beneficial) **aging** when residual life is increasing with x in some way or the other.

Lifetime distributions are mostly characterized with respect to aging by the behavior of

- their hazard rate $h(x)$ or
- their mean residual life $\mu(x)$.

Hazard rate classes will be discussed in Sections 2.1 and 2.2. Mean residual life classes are the topic of Section 2.3. But there are more statistical concepts used in classifying lifetime distributions. These will be presented in Section 2.4 where we will also show how all the aging criteria are linked.

Classes of lifetime distributions based on notions of aging afford statisticians an opportunity to consider problems of a character somewhat different from the usual. Instead of assuming that he knows nothing about the underlying lifetime distribution, the statistician assumes that he does not know the parametric form of the distribution, but that he does know, for example, that the hazard rate is increasing. More generally, he knows that some type of aging property holds for the lifetime distribution; this aging property give rise to a corresponding geometric property of the distribution. Knowing that a lifetime distribution belongs to a certain class, it is possible by using certain additional information to give approximations and bounds of the percentiles, moments and survival probabilities of this distribution. Of course, it is possible to test whether certain hypotheses on aging hold or not, see Sect. 10.3.

This chapter only present results for univariate distributions. Readers interested in aging criteria for multivariate distributions are referred to BLOCK/ SAVITS (1982, 1988), HARRIS (1970) or SHAKED/SHANTHIKUMAR (1987, 1988).

2.1 Monotone Hazard Rate Distributions

Since most materials, structures and devices wear out with time, the class of failure distributions for which the hazard rate is increasing is one of special interest. The phenomenon of work hardening of certain materials and the debugging of complex systems make the class of failure

[1] Suggested reading for this chapter: HOLLANDER/PROSCHAN (1984), MARSHALL/OLKIN (2007).

distributions with decreasing hazard rate also of some interest. Here, the terms 'increasing' and 'decreasing' are not used in the strict sense, but increasing (decreasing) stands for non–decreasing (non–increasing). Note that with this convention the continuous exponential distribution and the discrete geometric distribution with constant hazard rates belong to both classes. There are, of course, examples such as dynamic loading of structures, where a non–monotonic hazard rate function would be appropriate. Structures undergoing adjustment and modification also tend to have a non–monotonic hazard rate.

The assumption that a lifetime distribution has a monotone hazard rate is quite strong as we shall show, but such distributions possess many useful and interesting properties. Most results on monotone hazard rates hold for the continuous as well as for the discrete case, but there are some differences, especially in the way how to detect whether the distribution's hazard rate is increasing or decreasing. So we have decided to present the continuous and the discrete cases in two separate Sections 2.1.1 and 2.1.2. In Section 2.1.3 we will introduce the related concept of the hazard rate average and see when this is increasing or decreasing.

2.1.1 Continuous IHR and DHR Distributions[2]

We start by defining the properties **IHR (increasing hazard rate)** and **DHR (decreasing hazard rate)** without assuming that the distribution $F(\cdot)$ has a density and thus has a hazard rate $h(x) = f(x)/S(x)$. This definition is quite general and has its origin in the conditional failure function (1.6d), defined as the probability of failure in a finite time interval of length y, given the age x. If $F(\cdot)$ denotes the failure function, then the failure rate by this definition would be

$$\frac{F(x+y)-F(x)}{1-F(x)}. \tag{2.1}$$

We mention that when we divide this quantity by y and let $y \to 0$ we will obtain the familiar hazard rate $h(x)$.

Definition 1: A continuous distribution $F(\cdot)$ is IHR (DHR) if and only if

$$\frac{F(x+y)-F(x)}{1-F(x)}$$

is increasing (decreasing) in x for $y > 0$, where $x \geq 0$ such that $F(x) < 1$. ■

We could have defined IHR without restricting x to non–negative values; however, for DHR, we cannot extend x towards $-\infty$. Note that if $F(\cdot)$ is DHR, then $F(x) > 0$ for $x > 0$, The following theorem shows for distributions with support $[0, \infty)$ the equivalence of IHR (DHR) distribution with a density and distributions for which $h(x)$ is increasing (decreasing).

Theorem 9: Assume $F(\cdot)$ has a density $f(\cdot)$ with $F(0^-) = 0$. Then $F(\cdot)$ is IHR (DHR) if and only if $h(x)$ is increasing (decreasing). ■

Proof: Note that if in (2.11) we divide by y and let y approach to zero, we obtain $h(x) = f(x)/[1-F(x)]$. Hence, we need only show that $h(x)$ increasing (decreasing) in x implies (2.1) increasing (decreasing) in x. For $x_1 \leq x_2$

$$h(x_1) \underset{(\geq)}{\leq} h(x_2)$$

implies

$$\int\limits_0^x h(x_1+u)\,du \underset{(\geq)}{\leq} \int\limits_0^x h(x_2+u)\,du.$$

[2] Suggested reading for this section: BARLOW/MARSCHALL (1964), BARLOW/MARSHALL/PROSCHAN (1963), BARLOW/PROSCHAN (1965, 1975).

That is,

$$\exp\left\{-\int\limits_{x_2}^{x_2+x} h(u)\,du\right\} \underset{(\geq)}{\leq} \exp\left\{-\int\limits_{x_1}^{x_1+x} h(u)\,du\right\}$$

implying

$$\frac{F(x_2+x)-F(x_2)}{1-F(x_2)} \underset{(\leq)}{\geq} \frac{F(x_1+x)-F(x_1)}{1-F(x_1)}. \qquad \blacksquare$$

We will now show how the IHR (DHR) property is related to the future lifetime of x–survivors. Let x be the age of a statistical unit, then its survival function of future life $Y\,|\,x$ is given, see (1.52c), as

$$S(y\,|\,x) = \frac{S(x+y)}{S(x)} \tag{2.2a}$$

which can be written in terms of the hazard rate $h(x)$:

$$\begin{aligned} S(y\,|\,x) &= \frac{\exp\left\{-\int\limits_0^{x+y} h(u)\,du\right\}}{\exp\left\{-\int\limits_0^{x} h(u)\,du\right\}} \\ &= \exp\left\{-\int\limits_x^{x+y} h(u)\,du\right\}. \end{aligned} \tag{2.2b}$$

From (2.2b) we see that the conditional survival probability is an increasing (decreasing) function of x, the age reached, if and only if the hazard rate is decreasing (increasing). Thus, complementary to Definition 1 of IHR and DHR given above, we can state the following

Definition 2: $F(\cdot)$ is IHR (DHR) if and only if $S(y\,|\,x)$ is decreasing (increasing) in x for any $y>0,\ x\geq 0$ such that $S(x)>0$. $\blacksquare$

Introducing the notions **stochastically larger (smaller)**, we may characterize IHR (DHR) in still another way. A variate X_1 with distribution $F_1(x)$ is called stochastically smaller (larger) than a variate X_2 with $F_2(x)$, abbreviated

$$X_1 \underset{(\geq)}{\overset{\text{st}}{\leq}} X_2 \tag{2.3a}$$

if

$$F_1(x) \underset{(\leq)}{\geq} F_2(x)\ \ \forall\, x. \tag{2.3b}$$

Evidently we have

$$Y\,|\,x_1 \underset{(\leq)}{\overset{\text{st}}{\geq}} Y\,|\,x_2 \ \text{ for }\ x_1 \leq x_2 \tag{2.3c}$$

if the underlying lifetime distribution $F(\cdot)$ is IHR (DHR), i.e., the future lifetimes $Y\,|\,x$ become stochastically smaller (larger) with growing x.

The most important *geometric properties* of the IHR (DHR) lifetime distributions are stated in the following two Theorems 10 and 11.

Theorem10: $F(\cdot)$ is IHR (DHR) if and only if its logarithmic survival function $\ln S(x)$ is concave (convex). $\blacksquare$

Because $H(x) = -\ln S(x)$, see (1.11f), we may express Theorem 10 in terms of the cumulative hazard rate as well:

$F(\cdot)$ is IHR (DHR) if and only if its CHR $H(x) = \int\limits_0^x h(u)\,du$ is convex (concave).

We will first give a proof of Theorem 10 based on Definition 1 not assuming the existence of a density function.

Proof 1: Let $S(x) = 1 - F(x) = \exp\big[-H(x)\big]$. Then

$$\frac{F(x+y) - F(x)}{1 - F(x)} = 1 - \exp\Big\{-\big[H(x+y) - H(x)\big]\Big\},$$

and $F(\cdot)$ is IHR (DHR) if and only if $H(x+y) - H(x)$ is increasing (decreasing) in x for all $y > 0$. Thus $F(\cdot)$ is IHR (DHR) if and only if $H(x)$ is convex (concave). ■

Assuming a continuous distribution with existing density we have

Proof 2: $F(\cdot)$ is IHR (DHR) if its hazard rate has a non–negative (non–positive) first derivative, i.e.,

$$\frac{dh(x)}{dx} = \frac{d}{dx}\left[\frac{f(x)}{S(x)}\right] = \frac{S(x)\,f'(x) - \big[f(x)\big]^2}{\big[S(x)\big]^2} \underset{(\le)}{\ge} 0. \tag{2.4a}$$

For $\ln S(x)$ to be concave (convex) its second derivative has to be non–positive (non–negative), i.e.,

$$\frac{d^2\big[\ln S(x)\big]}{dx^2} = \frac{d}{dx}\left[-\frac{f(x)}{S(x)}\right] = -\frac{S(x)\,f'(x) - \big[f(x)\big]^2}{\big[S(x)\big]^2} \underset{(\ge)}{\le} 0. \tag{2.4b}$$

Multiplying (2.4b) by -1 gives (2.4a). ■

The second theorem on geometrical properties of IHR (DHR) distributions refers to their density functions.[3]

Theorem 11: Let $f(x)$ be a PDF defined on $\mathbb{R}^+$. The accompanying distribution is IHR (DHR) if and only if $\ln f(x)$ is concave (convex). ■

BARLOW/PROSCHAN (1965) give another statement on IHR and DHR distributions:

1. $F(\cdot)$ is IHR if and only if $S(x) = 1 - F(x)$ is a PÓLYA frequency function of order 2.
2. $F(\cdot)$ is DHR if and only if $S(x+y)$ is totally positive of order 2 in x and y for $x + y \ge 0$.

Example 2/1: IHR and DHR property of the WEIBULL distribution

The reduced[4] WEIBULL distribution has

$$\begin{aligned}
f(x) &= c\,x^{c-1}\exp\big(-x^c\big),\ x \ge 0,\ c > 0,\\
F(x) &= 1 - \exp\big(-x^c\big),\\
S(x) &= \exp\big(-x^c\big),\\
h(x) &= c\,x^{c-1},\\
H(x) &= x^c.
\end{aligned}$$

[3] Two more geometrical properties of these classes are:

1. The density function of a DHR distribution is a decreasing function.
2. The density function of an IHR distribution need not be unimodal.

[4] The reduced WEIBULL distribution has a location parameter set to 0 and a scale parameter set to 1.

1. With respect to the hazard rate $h(x)$ we have:

$$\frac{dh(x)}{dx} = c\,(c-1)\,x^{c-2}\begin{cases}\le 0 \text{ for } 0 < c \le 1 \;\Rightarrow\; \text{DHR},\\ \ge 0 \text{ for } c \ge 1 \;\Rightarrow\; \text{IHR}.\end{cases}$$

2. With respect to $\ln S(x) = -x^c$ we have:

$$\frac{d^2 \ln S(x)}{dx^2} = -c\,(c-1)\,x^{c-2}\begin{cases}\ge 0 \text{ (convex) for } 0 < c \le 1 \;\Rightarrow\; \text{DHR}\\ \le 0 \text{ (concave) for } c \ge 1 \;\Rightarrow\; \text{IHR}.\end{cases}$$

3. With respect to $\ln f(x) = \ln c + (c-1)\,\ln x - x^c$ we have:

$$\frac{d^2 \ln f(x)}{dx^2} = -\frac{c-1}{x^2} - c\,(c-1)\,x^{c-2}\begin{cases}\ge 0 \text{ (convex) for } 0 < c \le 1 \;\Rightarrow\; \text{DHR}\\ \le 0 \text{ (concave) for } c \ge 1 \;\Rightarrow\; \text{IHR}.\end{cases}$$

An immediate consequence of Theorem 10 is the following

<u>Lemma:</u> If $F(\cdot)$ is IHR (DHR) then $\big[S(x)\big]^{1/x}$ is increasing (decreasing). ■

By this lemma, $S(x) \le \big[S(x)\big]^{t/x}$ for $t > x$ and

$$\int\limits_x^\infty u^r\,S(u)\,du \le \int\limits_x^\infty u^r\,\big[S(u)\big]^{u/x}\,du < \infty$$

when $S(x) < 1$ and $r \ge 0$. Hence, IHR distributions have finite moments of all orders. DHR distributions necessarily must not have finite moments. For example, $F(x) = 1 - \dfrac{1}{1+x}$, $x \ge 0$, is DHR, but the mean does not exist.

BARLOW/PROSCHAN (1965) have given and proved a lot of theorems on IHR (DHR) distributions which rest on the fact that the exponential distribution is the boundary distribution between these two classes. We only cite four of their results giving bounds for the survival probability and for the moments.

1. If $F(\cdot)$ is IHR (DHR) with known percentile x_P of order P, $0 < P < 1$, i.e., $F(x_P) = P$, then

$$S(x)\begin{Bmatrix}\underset{(\le)}{\ge} \exp(-\alpha\,x) \;\text{ for }\; x \le x_P,\\ \underset{(\ge)}{\le} \exp(-\alpha\,x) \;\text{ for }\; x \ge x_P,\end{Bmatrix} \tag{2.5a}$$

where

$$\alpha = -\frac{\ln(1-P)}{x_P}. \tag{2.5b}$$

2. If $F(\cdot)$ is IHR with known mean μ, a sharp *lower bound* for $S(\cdot)$ is

$$S(x) \ge \begin{Bmatrix}\exp(-x/\mu) & \text{for } x < \mu\\ 0 & \text{for } x \ge \mu,\end{Bmatrix} \tag{2.6a}$$

and a sharp *upper bound* is

$$S(x) \le \begin{Bmatrix}1 & \text{for } x \le \mu\\ \exp(-\omega\,x) & \text{for } x > \mu,\end{Bmatrix} \tag{2.6b}$$

ω being the solution of $1 - \omega\,\mu = \exp(-\omega\,x)$.

3. If $F(\cdot)$ is DHR with known mean μ a sharp *upper bound* for $S(\cdot)$ is

$$S(x) \leq \left\{ \begin{array}{cc} \exp(-x/\mu) & \text{for } x \leq \mu, \\ \dfrac{\mu}{x\,e} & \text{for } x \geq \mu. \end{array} \right\} \tag{2.7}$$

   The sharp *lower bound* for DHR distributions is zero.

4. For an IHR distribution with $\mu = \mathrm{E}(X)$ we have the following inequality on moments:

$$\mu_r = \mathrm{E}\big(X^r\big) \leq r!\,\mu^r;\ r = 1,\,2,\,\ldots \tag{2.8a}$$

   whereas for DHR distributions with existing moments we have

$$\mu_r = \mathrm{E}\big(X^r\big) \geq r!\,\mu^r;\ r = 1,\,2,\,\ldots \tag{2.8b}$$

   A consequence of (2.8a,b) is the following inequality on the coefficient of variation:

$$\frac{\sqrt{\mathrm{Var}(X)}}{\mu} \underset{(\geq)}{\leq} 1 \tag{2.8c}$$

   for IHR (DHR) distributions.

An interesting question is whether or not the monotone hazard rate is preserved under certain operation with variates. We assume that the involved random variables are statistically independent.

1. A **mixture**

$$F(x) = \sum_{i=1}^{m} \alpha_i\,F_i(x);\ \alpha_i \geq 0,\ \sum_{i=1}^{m} \alpha_i = 1; \tag{2.9}$$

   of m DHR distributions is a DHR distribution. Mixing of IHR distributions does not necessarily result in an IHR distribution.

2. A **convolution** of IHR distributions also is IHR, especially if X_1 and X_2 are IHR with hazard rates $h_1(x)$ and $h_2(x)$, respectively, then $Y = X_1 + X_2$ has hazard rate $h_Y(x) \leq \min\big[h_1(x),\, h_2(x)\big]$. The sum of DHR variates is not DHR.

3. A **coherent structure** of IHR (DHR) components does not necessarily have an IHR (DHR) lifetime distribution, i.e., the IHR (DHR) class is not closed under formation of coherent systems. But, parallel and series systems of *identical* IHR components are IHR. For series systems the components do not have to be identical.

4. **Order statistics** from IHR distributions also have IHR distributions. However, this is not true for **spacings** from an IHR distribution. Order statistics from a DHR distribution are not necessarily DHR. However, spacings from a DHR distribution are DHR.

Tab. 2/1 in Sect. 2.4 summarizes preservation results for other classes of lifetime distributions.

2.1.2 Discrete IHR and DHR Distributions[5]

The results of the preceding section on continuous IHR and DHR distributions also hold for the discrete case unless given in terms of a PDF. A discrete distribution with PMF

$$P_i = \Pr(X = i);\ i = 0,\,1,\,\ldots$$

[5] Suggested reading for this section: GUPTA et al. (1997), KEMP (2004), LANGBERG et al. (1980), SHAKED et al. (1995).

is said to be IHR (DHR) if its hazard rate

$$h_i = \frac{P_i}{\sum\limits_{j=i}^{\infty} P_j};\ i = 0,\, 1,\, \ldots$$

is non–decreasing (non–increasing). Notice that $h_i \leq 1$. As the survival function $\sum\limits_{j \geq i} P_j$ rarely can be given in closed form it is not easy to determine the monotonicity of the hazard rate by using the difference $h_{i+1} - h_i$. In Theorem 11 we have seen that the IHR (DHR) property of a continuous distribution can be determined by the curvature of $\ln f(x)$. Because the PMF P_i plays the same role for the CDF (CCDF) as the PDF $f(x)$ in the continuous case, i.e., giving the increment in CDF (decrement in CCDF), a simple criterion for determining the monotonicity can be based on the curvature of the PMF, see GUPTA et al. (1997).

Define

$$\eta_i := \frac{P_i - P_{i+1}}{P_i} = 1 - \frac{P_{i+1}}{P_i} \tag{2.10a}$$

and

$$\Delta\eta_i = \eta_{i+1} - \eta_i = \frac{P_{i+1}}{P_i} - \frac{P_{i+2}}{P_{i+1}}. \tag{2.10b}$$

Recalling that **PMF** is **log convex** if

$$P_i\, P_{i+2} > P_{i+1}^2 \ \ \forall\, i$$

and **log concave** if

$$P_i\, P_{i+2} < P_{i+1}^2 \ \ \forall\, i,$$

we can say that log convexity is equivalent to $\Delta\eta_i < 0$ and log concavity is equivalent to $\Delta\eta_i > 0$. Thus, analogous to Theorem 11 we can state:

$$\left.\begin{array}{l} 1)\ \text{If }\ \Delta\eta_i < 0,\ \text{ then }\ h_i\ \text{ is non–increasing (DHR).} \\ 2)\ \text{If }\ \Delta\eta_i > 0,\ \text{ then }\ h_i\ \text{ is non–decreasing (IHR).} \\ 3)\ \text{If }\ \Delta\eta_i = 0,\ \text{ then }\ \dfrac{P_{i+1}}{P_i} = \dfrac{P_{i+2}}{P_{i+1}}\ \ \forall\, i. \end{array}\right\} \tag{2.10c}$$

The difference $\Delta\eta_i = 0$ in 3) implies

$$P_i = c^i\, P_0,$$

where c is a positive constant, and three distributions with this property are possible, see GUPTA et al. (1997):

a) $P_i = P_0\,(1 - P_0)^i;\ i = 0,\, 1,\, \ldots$

This is the geometric distribution which has a constant hazard rate and is IHR as well as DHR.

b) $P_i = P_0 = 1/(m+1);\ i = 0,\, 1,\, \ldots,\, m$

This is the discrete uniform distribution which is IHR.

c) $P_i = \dfrac{c^i}{1 + c + c^2 + \ldots + c^m};\ i = 0,\, 1,\, \ldots,\, m$

This distribution is IHR, too.

Thus, in order to find out whether a discrete distribution is IHR or DHR or not monotone we just have to study the behavior of the ratio of two adjacent probabilities.

Example 2/2: Monotonicity of well–known discrete distributions[6]

Binomial distribution

$$\begin{aligned} P_i &= \binom{n}{i} P^i\,(1-P)^{n-i};\ i = 0,\,1,\,\ldots,\,n,\ n\in\mathbb{N},\ 0<P<1; \\ \frac{P_{i+1}}{P_i} &= \frac{n-i}{i+1}\,\frac{P}{1-P} \\ \Delta\eta_i &= \frac{n+1}{(i+1)\,(i+2)}\,\frac{P}{1-P} > 0 \ \Rightarrow\ \text{IHR} \end{aligned}$$

Logarithmic series distribution

$$\begin{aligned} P_i &= \frac{a\,P^i}{i};\ i = 1,\,2,\,\ldots,\ 0<P<1,\ a = -\big[\ln(1-P)\big]^{-1}; \\ \frac{P_{i+1}}{P_i} &= \frac{i}{i+1}\,P \\ \Delta\eta_i &= -\frac{P}{(i+1)\,(i+2)} < 0 \ \Rightarrow\ \text{DHR} \end{aligned}$$

Negative binomial distribution

$$\begin{aligned} P_i &= \binom{k+i-1}{i} q^k\,(1-q)^i;\ i = 0,\,1,\,\ldots,\ 0<q<1,\ k>0; \\ \frac{P_{i+1}}{P_i} &= \frac{k+i}{(i+1)}\,(1-q) \\ \Delta\eta_i &= \frac{k-1}{(i+1\,(i+2)}\,(1-q) \begin{cases} <0 \ \text{for}\ 0<k<1 \ \Rightarrow\ \text{DHR} \\ =0 \ \text{for}\ k=1 \ \Rightarrow\ \text{geometric distribution} \\ >0 \ \text{for}\ k>1 \ \Rightarrow\ \text{IHR} \end{cases} \end{aligned}$$

POISSON distribution

$$\begin{aligned} P_i &= \frac{\lambda^i}{i!}\,e^{-\lambda};\ i = 0,\,1,\,\ldots,\ \lambda>0; \\ \frac{P_{i+1}}{P_i} &= \frac{\lambda}{i+1} \\ \Delta\eta_i &= \frac{\lambda}{(i+1)\,(i+2)} > 0 \ \Rightarrow\ \text{IHR} \end{aligned}$$

We have just seen that the ratio of two adjacent probabilities serves to investigate the behavior of the hazard rate. GUPTA et al. (1997) also show that it is possible to compute the hazard rate when these ratios are known. The fundamental equation, which has to be evaluated, gives the reciprocal

[6] For more discrete distributions see Sect. 3.2.

of the hazard rate as follows:

$$\begin{aligned}\frac{1}{h_i} &= \frac{\Pr(X \geq i)}{\Pr(X = i)} \\ &= \frac{P_i + P_{i+1} + P_{i+2} + \ldots}{P_i} \\ &= 1 + \frac{P_{i+1}}{P_i} + \frac{P_{i+2}}{P_{i+1}}\,\frac{P_{i+1}}{P_i} + \frac{P_{i+3}}{P_{i+2}}\,\frac{P_{i+2}}{P_{i+1}}\,\frac{P_{i+1}}{P_i} + \ldots \\ &= 1 + \sum_{j=0}^{\infty}\prod_{k=i}^{i+j}\frac{P_{k+1}}{P_k}. \end{aligned} \tag{2.11}$$

2.1.3 IHRA and DHRA Distributions[7]

Introducing the so–called **hazard rate average (HRA)** lessens the requirement of a monotone course of the hazard rate when monotonicity is only asked for its average. Thus, the class of IHRA (DHRA) distributions — **IHRA (DHRA)** $\hat{=}$ **increasing (decreasing) hazard rate average** — is wider than the class of IHR (DHR) distributions. In defining HRA we have to distinguish between continuous and discrete distributions.

Definition 1: A *continuous* distribution is IHRA (DHRA) if

$$HRA(x) = -\frac{1}{x}\,\ln S(x) \tag{2.12a}$$

is increasing (decreasing) in $x,\ x \geq 0$, or, equivalently, if

$$\big[S(x_1)\big]^{1/x_1} \underset{(\leq)}{\geq} \big[S(x_2)\big]^{1/x_2} \ \text{ for } \ 0 \leq x_1 \leq x_2. \qquad \blacksquare \tag{2.12b}$$

Recall from (1.11f) that $-\ln S(x)$ represents the cumulative hazard rate $H(x) = \int\limits_0^x h(u)\,du$, when the hazard rate exists and we then have

$$HRA(x) = \frac{1}{x}\int\limits_0^x h(u)\,du = \frac{H(x)}{x}. \tag{2.12c}$$

Thus, we may give a second and equivalent

Definition 2: A continuous distribution is IHRA (DHRA) if

$$\frac{H(x_2)}{x_2} \underset{(\leq)}{\geq} \frac{H(x_2)}{x_2} \ \text{ for } \ 0 \leq x_1 \leq x_2. \qquad \blacksquare \tag{2.12d}$$

Based on (2.12c) IHRA (DHRA) means that

$$\frac{d}{dx}\left[\frac{H(x)}{x}\right] \underset{(\leq)}{\geq} 0 \ \ \forall\, x \geq 0 \tag{2.13a}$$

or

$$h(x) \underset{(\leq)}{\geq} \frac{H(x)}{x} \ \ \forall\ \geq 0, \tag{2.13b}$$

[7] Suggested reading for this section: BARLOW/PROSCHAN (1975), LANGBERG et al. (1982).

i.e., the increment $h(x)$ of $H(x)$ has to be greater (smaller) than the hazard rate average $HRA(x)$ for all x. For example, looking at the reduced WEIBULL distribution we have

$$H(x) = x^c,\ \ h(x) = c\,x^{c-1},\ \ \frac{H(x)}{x} = HRA(x) = x^{c-1}$$

and

$$h(x) = c\,x^{c-1} \begin{cases} \leq \dfrac{H(x)}{x} = x^{c-1} \ \text{ for } \ 0 < c \leq 1 \Rightarrow \ \text{DHRA and DHR} \\ \geq \dfrac{H(x)}{x} = x^{c-1} \ \text{ for } \ c \geq 1 \Rightarrow \ \text{IHRA and IHR.} \end{cases}$$

It is obvious that an IHRA distribution is characterized by decreasing $\big[S(x)\big]^{1/x}$ on $[0,\,\infty)$, while a DHRA distribution is characterized by increasing $\big[S(x)\big]^{1/x}$ on $[0,\,\infty)$. Hence, we can formulate

Theorem 12: A distribution is IHRA (DHRA) if and only if

$$S(a\,x) \underset{(\leq)}{\geq} \big[S(x)\big]^{1/a},\ 0 < a < 1,\ x \geq 0. \qquad \blacksquare \qquad (2.14)$$

Another theorem on IHRA (DHRA) is

Theorem 13: If a distribution $F(\cdot)$ is IHR (DHR) then $F(\cdot)$ is IHRA (DHR). $\blacksquare$

The reverse of Theorem 13 necessarily does not always hold. For example,

$$F(x) = \big(1 - e^{-x}\big)\,\big(1 - e^{-c\,x}\big);\ x \geq 0,\ c > 1,$$

is IHRA but not IHR.

Furthermore we can state

Theorem 14: A distribution $F(\cdot)$ is IHRA (DHRA) if and only if the difference $S(x) - \exp(\lambda\,x)$ has exactly one change in sign from $+$ to $-$ (from $-$ to $+$) for all $\lambda > 0$. $\blacksquare$

From Theorem 14 we can easily deduce the following bounds for the survival probability of an IHRA (DHRA) distribution with known percentile X_P of order $P,\ 0 < P < 1:$

$$S(x) \left\{ \begin{array}{ll} \underset{(\leq)}{\geq} \exp\left[-\dfrac{x\,\ln(1-P)}{x_P}\right] & \text{for } \ x \leq x_P \\ \underset{(\geq)}{\leq} \exp\left[-\dfrac{x\,\ln(1-P)}{x_P}\right] & \text{for } \ x \geq x_P. \end{array} \right\} \qquad (2.15)$$

BARLOW/PROSCHAN (1975, pp. 91 ff.) give the following stochastic model leading to an IHRA lifetime distribution. A device is subject to shocks occurring randomly in time according to a POISSON process, each shock independently causes random damage to the device. The damages accumulate until a critical threshold or capacity is exceeded, at which time the device fails. This time to failure is governed by an IHRA distribution. For closure and inheritance of IHRA (DHRA) distributions see Tab. 2/1 in Sect. 2.4.

In the *discrete case* we have — see (1.62a,b) — two different CHRs, the *cumulative* hazard rate function

$${}_1H_i = -\ln S_i = -\ln\left(\sum_{j\geq i} P_j\right);\ i = 0,\,1,\,\ldots$$

and the *accumulated* hazard rate

$${}_2H_i = \sum_{j=0}^{i} h_j;\ i = 0,\,1,\,\ldots$$

Thus, we can define two discrete hazard rate averages:

$$ {}_1HRA_i = -\frac{1}{i+1}\ln S_i;\; i = 0,\, 1,\, \ldots; \tag{2.16a} $$

$$ {}_2HRA_i = \frac{1}{i+1}\sum_{j=0}^{i} h_j;\; i = 0,\, 1,\, \ldots \tag{2.16b} $$

Definition: A discrete distribution is IHRA (DHRA) *in the sense of the cumulative hazard rate* if

$$ {}_1HRA_i \underset{(\leq)}{\geq} {}_1HRA_{i-1};\; i = 1,\, 2,\, \ldots \tag{2.17a} $$

or equivalently if

$$ S_i^{1/(i+1)} \underset{(\geq)}{\leq} S_{i-1}^{1/i};\; i = 1,\, 2,\, \ldots \tag{2.17b} $$

and it is IHRA (DHRA) *in the sense of the accumulated hazard rate* if

$$ {}_2HRA_i \underset{(\leq)}{\geq} {}_2HRA_{i-1};\; i = 1,\, 2,\, \ldots \tag{2.17c} $$

or equivalently if

$$ \frac{i}{i+1} \underset{(\leq)}{\geq} \frac{{}_2H_{i-1}}{{}_2H_i};\; i = 1,\, 2,\, \ldots \qquad \blacksquare \tag{2.17d} $$

Theorem 15: If a discrete lifetime distribution is IHR (DHR) then it is IHRA (DHRA) in the sense of the cumulative hazard rate as well as in the sense of the accumulated hazard rate. ■

Proof: We first look at ${}_1HRA_i$. If

$$ h_0 \leq h_1 \leq \ldots \leq h_{i-1} \leq h_i \leq \ldots $$

then, because of $0 \leq h_i \leq 1 \;\; \forall\, i$,

$$ \ln(1-h_0) \geq \ln(1-h_1) \geq \ldots \geq \ln(1-h_{i-1}) \geq \ln(1-h_i) \geq \ldots $$

and

$$ \begin{aligned} {}_1HRA_i - {}_1HRA_{i-1} &= -\frac{1}{i+1}\sum_{j=0}^{i-1}\ln(1-h_j) + \frac{1}{i}\sum_{j=0}^{i-2}\ln(1-h_j) \\ &= \frac{(i+1)\sum\limits_{j=0}^{i-2}\ln(1-h_j) - i\sum\limits_{j=0}^{i-1}\ln(1-h_j)}{i\,(i+1)} \\ &= \frac{\sum\limits_{j=0}^{i-2}\big[\ln(1-h_j) - \ln(1-h_{i-1})\big] - \ln(1-h_{i-1})}{i\,(i+1)} \geq 0 \end{aligned} $$

and the distribution is IHRA. Similarly, if

$$ h_0 \geq h_1 \geq \ldots \geq h_{i-1} \geq h_i \geq \ldots $$

then

$$ \ln(1-h_0) \leq \ln(1-h_1) \leq \ldots \leq \ln(1-h_{i-1}) \leq \ln(1-h_i) \leq \ldots $$

and

$$_1HRA_i - {}_1HRA_{i-1} \leq 0$$

and the distribution is DHRA.

We now look at $_2HRA_i$. If

$$h_0 \leq h_1 \leq \ldots \leq h_{i-1} \leq h_i \leq \ldots$$

then

$$\begin{aligned} i\,{}_2H_i - (i+1)\,{}_2H_{i-1} &= i\sum_{j=0}^{i} h_j - (i+1)\sum_{j=0}^{i-1} h_j \\ &= \sum_{j=0}^{i-1}(h_i - h_j) \geq 0 \end{aligned}$$

and the distribution is IHRA. Similarly, if

$$h_0 \geq h_1 \geq \ldots \geq h_{i-1} \geq h_i \geq \ldots$$

then

$$i\,{}_2H_i - (i+1)\,{}_2H_{i-1} \leq 0$$

and the distribution is DHRA. ■

When the distribution is not IHR (DHR) Theorem 15 must not hold. We recommend to express the behavior of the hazard rate average in the discrete sense by means of $_2HRA_i$, because (2.16b) is the average by construction, i.e., a sum divided by the number of its summands.

2.2 Non–monotone Hazard Rate Distributions[8]

The IHR (IHRA) property is characteristic for devices that consistently (on the average) deteriorate with age, whereas the DHR (DHRA) property is characteristic for devices that consistently (on the average) improve with age. But many physical phenomenon exhibit hazard rates that are not monotone. Of special interest are hazard rate which first decrease and afterwards increase and look like a **bathtub** or which first increase and then decrease and look like an **inverted (upside–down) bathtub**.

A common description of bathtub–shaped hazard rates which is appropriate for modeling human lifetimes by means of the lifetable[9] shows three phases: an initial phase during which the hazard rate (here: the one–year age specific death rate) decreases, followed by a middle phase during which the hazard rate is approximately constant, concluded by a final phase during which the hazard rate increases. For human beings, the first phase (infant mortality) shows death due typically to hereditary defects, whose impact diminishes with age. The middle phase (chance failure) shows death due typically to sudden jolts such as accidents. The final phase (wear–out) shows death resulting from the natural accumulation of negative health effects. The logical counterpart to bathtub–shaped hazard rates is the three phase situation in which the hazard rate initially increases, then becomes essentially constant, and ultimately decreases. This upside–down shape can be found in accelerated life testing, in which the units tested are subjected to abnormally high stress levels. We will abbreviate the bathtub property by **DIHR (decreasing–increasing hazard rate)** and the upside–down bathtub property by **IDHR (increasing–decreasing hazard rate)**.

[8] Suggested reading for this section: DHILLON (1979, 1981), GLASER (1980), GRIFFITH (1982), HJORTH (1980), LAI et al. (2001), SILVA et al. (2010).

[9] The lifetable is presented and discussed in Chapter 6.

Whether a hazard rate is DIHR or IDHR can be found out by investigating its first derivative in the case of a continuous variate or its first difference in the case of a discrete variate. A quite general definition which extends the idea of DIHR and IDHR to situations where the hazard rate itself does not exist is

Definition: A lifetime distribution $F(x)$ with $x \in [0, \infty)$ is said to be DIHR (IDHR) if there exists a $x_0 > 0$ such that $H(x) = -\ln[1 - F(x)]$ is concave (convex) on $[0, x_0)$ and convex (concave) on $[x_0, \infty)$. ■

GLASER (1980) has given sufficient conditions to characterize a given lifetime distribution as being IHR, DHR, IDHR, and DIHR, assuming that its PDF $f(x)$ is continuous and twice differentiable on $[0, \infty)$. These conditions rest upon the reciprocal of the hazard rate

$$g(x) := \frac{1}{h(x)} = \frac{S(x)}{f(x)} \tag{2.18a}$$

with first derivative

$$g'(x) = g(x)\,\zeta(x) - 1 \tag{2.18b}$$

where

$$\zeta(x) = -\frac{f'(x)}{f(x)} \tag{2.18c}$$

and

$$\zeta'(x) = \big[\zeta(x)\big]^2 - \frac{f''(x)}{f(x)}. \tag{2.18d}$$

These conditions are stated in the following

Theorem 16:

a) If $\zeta'(x) > 0 \;\; \forall\, x \geq 0$, then IHR.

b) If $\zeta'(x) < 0 \;\; \forall\, x \geq 0$, then DHR.

c) Suppose there exists $x_0 > 0$ such that

$$\zeta'(x) < 0 \;\; \forall\, x \in [0, x_0), \quad \zeta'(x_0) = 0 \;\; \text{and} \;\; \zeta'(x) > 0 \;\; \forall\, x > x_0. \tag{2.19a}$$

   ca) If there exists $y_0 > 0$ such that $g'(y_0) = 0$, then DIHR.

   cb) If there does not exist $y_0 > 0$ such that $g'(y_0) = 0$, then IHR.

d) Suppose there exists $x_0 > 0$ such that

$$\zeta'(x) > 0 \;\; \forall\, x \in [0, x_0), \quad \zeta'(x_0) = 0 \;\; \text{and} \;\; \zeta'(x) < 0 \;\; \forall\, x > x_0. \tag{2.19b}$$

   da) If there exists $y_0 > 0$ such that $g'(y_0) = 0$, then IDHR.

   db) If there does not exist $y_0 > 0$ such that $g'(y_0) = 0$, then DHR. ■

GLASER (1980) has supplemented Theorem 16 by the following lemma that helps to avoid finding a root y_0 of $g'(\cdot)$.

Lemma: Suppose(2.19a) or (2.19b) hold in Theorem 16.

a) Suppose $\varepsilon = \lim_{x \to 0^+} f(x)$ exists, possibly equal to 0 or ∞.

   (i) If $\varepsilon = \infty$ and (2.19a) holds, then DIHR.

   (ii) If $\varepsilon = 0$ and (2.19a) holds, then IHR.

(iii) If $\varepsilon = \infty$ and (2.19b) holds, then DHR.

(iv) If $\varepsilon = 0$ and (2.19b) holds, then IDHR.

b) Suppose $\delta = \lim_{x\to 0^+} g(x)\,\zeta(x)$ exists, possibly equal to 0 or $-\infty$.

(i) If $\delta > 1$ and (2.19a) holds, then DIHR.

(ii) If $\delta < 1$ and (2.19a) holds, then IHR.

(iii) If $\delta > 1$ and (2.19b) holds, then DHR.

(iv) If $\delta < 1$ and (2.19b) holds, then IDHR. ■

Among the popular and classical distributions we seldom find one which is DIHR or IDHR. Two prominent exceptions are the **log–normal distribution** with[10]

$$
\begin{aligned}
f(x) &= \frac{1}{b\sqrt{2\,\pi}}\,\frac{1}{x}\,\exp\left[-\frac{(\ln x - a)^2}{2\,b^2}\right],\ x>0,\ a = \mathrm{E}(\ln X)\in\mathbb{R},\ b = \sqrt{\mathrm{Var}(\ln X)},\\
&= \frac{1}{b}\,\frac{1}{x}\,\phi\left(\frac{\ln x - a}{b}\right), \qquad (2.20a)\\
F(x) &= \Phi\left(\frac{\ln x - a}{b}\right), \qquad (2.20b)
\end{aligned}
$$

and the **inverse GAUSSIAN distribution**, also called **WALD distribution** with

$$
\begin{aligned}
f(x) &= \sqrt{\frac{b}{2\,\pi\,x^3}}\,\exp\left[-\frac{b\,(x-a)^2}{2\,a^2\,x}\right],\ x>0,\ a = \mathrm{E}(X)>0,\ b = \frac{a^3}{\mathrm{Var}(X)}>0, \qquad (2.21a)\\
F(x) &= \Phi\left[\sqrt{\frac{b}{x}}\left(\frac{x}{a}-1\right)\right] + \exp\left(\frac{2\,b}{a}\right)\,\Phi\left[-\sqrt{\frac{b}{x}}\left(\frac{x}{a}+1\right)\right]. \qquad (2.21b)
\end{aligned}
$$

These two distributions are IDHR irrespective of their parameter values. Figures 2/1 and 2/2 depict the hazard rates of these distributions for several combinations of parameter values.

There are several approaches to construct DIHR and IDHR distributions. We only mention:[11]

1. directly specifying a hazard rate that has a bathtub or inverted bathtub shape and then recovering its CDF and PDF using (1.9c,d),

2. mixing or compounding of distributions,

3. generalizing a familiar distribution by introducing additional parameters.

We will give examples for each of these three approaches.

[10]

1. $\phi(u) = \frac{1}{\sqrt{2\,\pi}}\,\exp\big(-u^2/2\big)$ is the PDF of the **standardized normal distribution**.

2. $\Phi(u) = \int\limits_{-\infty}^{u} \phi(z)\,dz$ is the CDF of the standardized normal distribution which cannot be given in closed form.

[11] For more approaches see LAI et al. (2001).

Figure 2/1: Hazard rates of several log–normal distributions

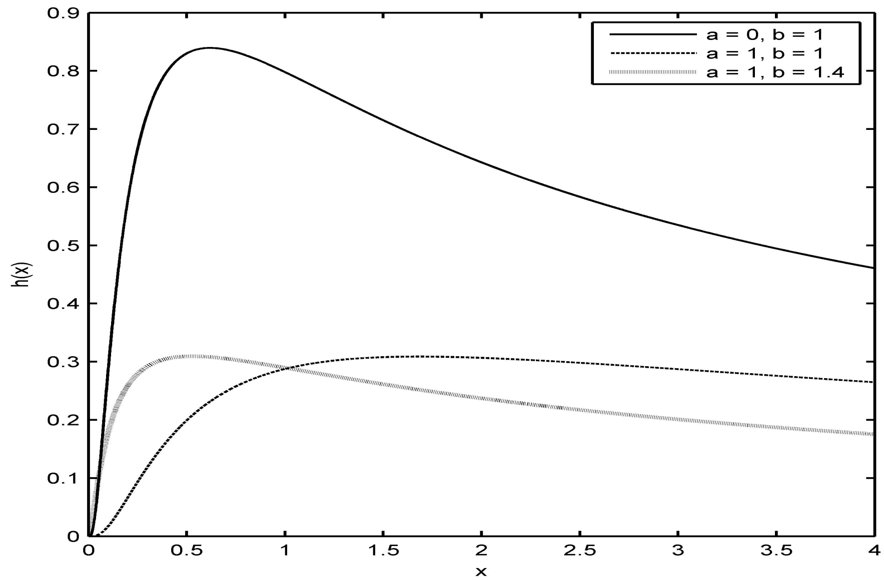
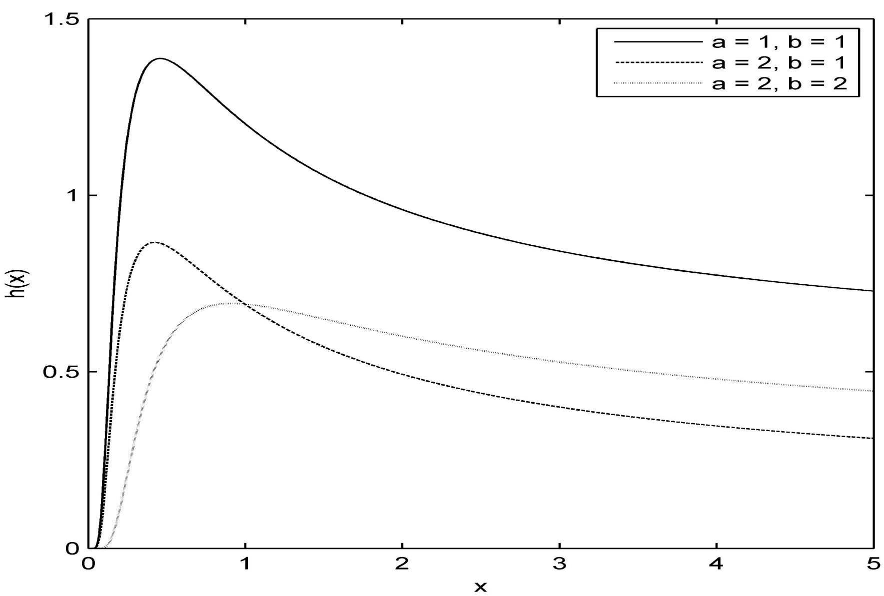


Figure 2/2: Hazard rates of several inverse GAUSSIAN distributions



A simple bathtub–shaped hazard rate is realized by a polynomial of second degree:[12]

$$h(x) = a + b\,x + c\,x^2;\ x \geq 0,\ a > 0,\ b < 0,\ c > 0. \tag{2.22a}$$

The change point (minimum) is at $x = -b\big/(2\,c)$. PDF and CDF corresponding (2.22a) are

$$f(x) \;=\; (a + b\,x + c\,x^2)\,\exp\biggl(-a\,x - \frac{b}{2}\,x^2 - \frac{c}{3}\,x^3\biggr), \tag{2.22b}$$

$$F(x) \;=\; 1 - \exp\biggl(-a\,x - \frac{b}{2}\,x^2 - \frac{c}{3}\,x^3\biggr). \tag{2.22c}$$

[12] GLASER (1980) shows that it is not possible to create an upside–down bathtub–shaped hazard rate distribution on $[0,\,\infty)$ by a polynomial.

Another possibility of directly specifying a bathtub–shaped hazard rate is to add or superimpose a decreasing and an increasing function. The approach of HJORTH (1980) rests upon this idea:

$$h(x) = \delta\, x + \frac{\theta}{1 + \beta\, x};\ x \geq 0,\ \delta,\ \theta,\ \beta \geq 0. \tag{2.23a}$$

$\delta\, x$ is an increasing term and $\theta\big/(1 + \beta\, x)$ is a decreasing term. For $\beta > 0$ (2.23a) results in

$$f(x) = \frac{\theta + \delta\, x\,(1 + \beta\, x)}{(1 + \beta\, x)^{\theta/\beta+1}}\ \exp\!\left(-\frac{\delta\, x^2}{2}\right), \tag{2.23b}$$

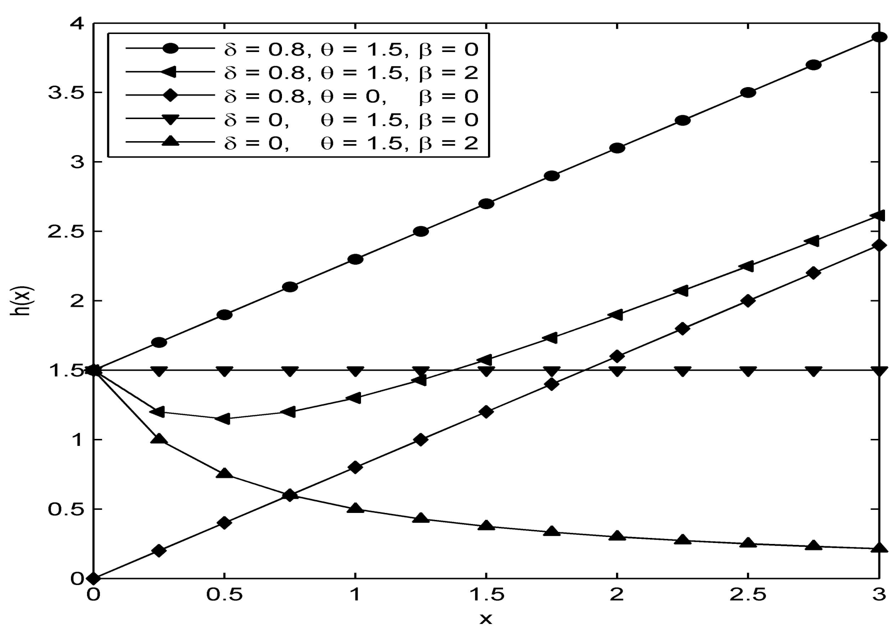
$$F(x) = 1 - \frac{\exp\!\big(-\delta\, x^2/2\big)}{(1 + \beta\, x)^{\theta/\beta}}, \tag{2.23c}$$

$$H(x) = \frac{\delta/x^2}{2} + \frac{\delta\, \ln(1 + \beta\, x)}{\beta}. \tag{2.23d}$$

For $\beta = 0$ we have to take the limit of $f(x)$ and $F(x)$ as $\beta \to 0$. Special cases of (2.23a) — see Fig. 2/3 — are:

- $\theta = 0 \;\Rightarrow\;$ increasing hazard rate $\delta\, x$ of a RAYLEIGH distribution,
- $\delta = 0 \;\Rightarrow\;$ decreasing hazard rate of a distribution with $F(x) = 1 - (1 + \beta\, x)^{-\theta/\beta}$, which is the reduced LOMAX distribution when $\beta = 1$.
- $\delta = \beta = 0 \;\Rightarrow\;$ constant hazard rate of an exponential distribution,
- $\delta \geq \theta\, \beta \;\Rightarrow\;$ increasing hazard rate,
- $0 < \delta < \theta\, \beta \;\Rightarrow\;$ bathtub–shaped hazard rate with change point (minimum) at $x = \dfrac{1}{\beta}\left(\sqrt{\dfrac{\theta\, \beta}{\delta}} - 1\right)$.

Figure 2/3: Hazard rates of several HJORTH distributions



An example for a compound distribution with possible bathtub–shaped hazard rate is the exponential power distribution of DHILLON (1981), called DHILLON–I distribution. The hazard rate reads:

$$h(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \exp\left[\left(\frac{x-a}{b}\right)^{c}\right];\ x \ge a,\ a \in \mathbb{R},\ b,\ c > 0. \tag{2.24a}$$

This hazard rate is displayed in Fig. 2/5 further down. The following functions belong to (2.24a):

$$f(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \exp\left\{1 - \exp\left[\left(\frac{x-a}{b}\right)^{c}\right] + \left(\frac{x-a}{b}\right)^{c}\right\}, \tag{2.24b}$$

$$F(x) = 1 - \exp\left\{1 - \exp\left[\left(\frac{x-a}{b}\right)^{c}\right]\right\}, \tag{2.24c}$$

$$H(x) = \exp\left\{1 - \exp\left[\left(\frac{x-a}{b}\right)^{c}\right]\right\} - 1. \tag{2.24d}$$

When $b = 1$ we have a log–WEIBULL distribution, also known as type–I extreme value distribution of the minimum, and the hazard is increasing for $b \ge 1$. The bathtub–shaped hazard rate comes up for $0 < b < 1$ with a change point (minimum) at $x = \frac{1}{a}\left(\frac{1-b}{b}\right)^{1/b}$.

A first example for a generalized distribution is STACY's (1962) generalized gamma distribution[13] with[14]

$$f(x) = \frac{d\,x^{d\,c-1}}{b^{d\,c}\,\Gamma(c)} \exp\left[-\left(\frac{x}{b}\right)^{d}\right];\ x \ge 0;\ b,\ c,\ d > 0. \tag{2.25}$$

This distribution includes many other distributions as special cases, see RINNE (2009, pp. 111 ff.). The behavior of the hazard rate does not depend on the scaling parameter b, but it depends on $c\,d - 1$ as follows:

- $c\,d - 1 < 0$
  - ⋆ $d \le 1 \Rightarrow$ DHR
  - ⋆ $d > 1 \Rightarrow$ DIHR
- $c\,d - 1 > 0$
  - ⋆ $d \ge 1 \Rightarrow$ IHR
  - ⋆ $d < 1 \Rightarrow$ IDHR
- $c\,d - 1 = 0$
  - ⋆ $d = 1 \Rightarrow$ constant hazard rate
  - ⋆ $d < 1 \Rightarrow$ DHR
  - ⋆ $d > 1 \Rightarrow$ IHR.

---

13 The ordinary gamma distribution with $f(x) = \frac{1}{b\,\Gamma(d)}\left(\frac{x}{b}\right)^{d-1}\exp\left(-\frac{x}{b}\right);\ x \ge 0,\ b,\ d > 0$; is IHR for $d \ge 1$ and IHR for $0 < d \le 1$.

14 $\Gamma(z) = \int\limits_0^\infty u^{z-1}\,e^{-u}\,du$ is the complete gamma function.

A second example is the generalized exponential geometric distribution given by SILVA et al. (2010) with

$$F(x) = \left(\frac{1-\exp(-b\,x)}{1-p\,\exp(-b\,x)}\right)^{c};\ x \ge 0,\ b,\,c > 0,\ p \in (0,\,1), \tag{2.26a}$$

$$f(x) = \frac{c\,b\,(1-p)\,\exp(-b\,x)\,\big[1-\exp(-b\,x)\big]^{c-1}}{\big[1-p\,\exp(-b\,x)\big]^{c+1}}, \tag{2.26b}$$

$$h(x) = \frac{c\,b\,(1-p)\,\exp(-b\,x)\,\big[1-\exp(-b\,x)\big]^{c-1}\,\big[1-p\,\exp(-b\,x)\big]^{-1}}{\big[1-p\,\exp(-b\,x)\big]^{c}-\big[1-\exp(-b\,x)\big]^{c}} \tag{2.26c}$$

This distribution is:

- IDHR for $p \in \left(\dfrac{c-1}{c+1},\,1\right)$ and $c > 1$,
- IHR for $p \in \left(0,\,\dfrac{c-1}{c+1}\right)$ and $c > 1$,
- DHR otherwise.

2.3 MRL classes of Distributions[15]

The behavior of the mean residual life function MRL

- $\mu(x) = \mathrm{E}(X - x \,|\, X \ge x) = \dfrac{1}{S(x)} \displaystyle\int_x^\infty S(u)\,du$ in the continuous case and
- $L_i = \mathrm{E}(X - i \,|\, X \ge i) = \dfrac{1}{S_i} \displaystyle\sum_{j>i} S_j$ in the discrete case

may also be used to characterize aging and to classify lifetime distributions. While the hazard rate function at x provides information about a *small interval* just after x, the MRL function at x considers information about the *whole interval* after x (all after x). This intuition explains the difference between the two.

When MRL is monotone and increasing (decreasing) — abbreviated **IMRL (DMRL)** — we have beneficial (adverse) aging. But there are also distributions with non–monotone MRL. Of special interest in this case are the **DIMLR** class where MRL has a bathtub shape (first decreasing, afterwards increasing) and the **IDMRL** class with an upside–down bathtub shape (first increasing, afterwards decreasing).

What functions are MRL functions? — Several characterizations are possible which answer this. We cite the characterization given by GUESS/PROSCHAN (1988, p. 217).[16]

Theorem 17: Consider the following conditions:

(i) $\mu(x):\ [0,\,\infty)\ \to\ [0,\,\infty)$.

(ii) $\mu(0) > 0$.

[15] Suggested reading for the section: BRYSON/SIDDIQUI (1969), EBRAHIMI (1986), GUESS/PARK (1988), GUESS/PROSCHAN (1988), KEMP (2004), KLEFSJÖ (1982b), LAI et al. (2001), WATSON/WELLS (1961).

[16] A characterization with examples of DMRL (IMRL) in the discrete case may be found in EBRAHIMI (1986).

(iii) $\mu(x)$ is right–continuous (not necessarily continuous).

(iv) $m(x) := \mu(x) + x$ is increasing on $[0, \infty)$.

(v) When there exists x_0 such that $\mu(x_0^-) = \lim_{x \to 0^-} \mu(x) = 0$, then $\mu(x) = 0$ holds for $x \in [x_0, \infty)$. Otherwise, when there does not exist such a x_0 with $\mu(x_0^-) = 0$, then $\int\limits_0^\infty \frac{1}{\mu(u)}\, du = \infty$ holds.

A function $\mu(x)$ satisfies (i) – (v) if and only if $\mu(x)$ is the MRL function of a non–degenerate at $x = 0$ lifetime distribution. ■

Note that condition (ii) rules out the the degenerate at $x = 0$ distribution. (iv) is a statement on the expected time of death given that a unit has survived to time x, see (1.16a,b). Theorem 17 delineates which functions can serve as MRL functions, and hence, it provides models for life–lengths. For recovering the other representatives of a lifetime distribution see (1.20a–f).

We now look at the relationship between the HR classes and the MRL classes.

Theorem 18: A lifetime distribution that is IHR (DHR) has a decreasing (increasing) MRL function, i.e., the IHR (DHR) class is contained in the DMRL (IMRL) class. ■

The implication IHR ⇒ DMRL (DHR ⇒ IMRL) of Theorem 18 cannot be reversed in general. We give a proof of Theorem 18 for a discrete lifetime distribution:[17]

$$\begin{aligned} L_i - L_{i+1} &= \sum_{j>i} \left(\frac{S_j}{S_i} - \frac{S_{j+1}}{S_{i+1}} \right) \\ &= \sum_{j>i} \left[(h_j - h_i) \prod_{k=i+1}^{j-1} (1 - h_k) \right] \underset{(\leq)}{\geq} 0, \end{aligned} \tag{2.27}$$

accordingly as $h_j \underset{(\leq)}{\geq} h_i,\ j > i$. Therefore, when $h_i < h_{i+1}\ \forall\, i$ (= IHR), the MRL function decreases and when $h_i > h_{i+1}\ \forall\, i$ (= DHR), the MRL function increases.

Example 2/3: Distributions with monotone HR and monotone MRL

The reduced WEIBULL distribution has

$$h(x) = c\, x^{c-1};\ x \geq 0,\ c > 0;$$

which is DHR for $0 < c \leq 1$ and IHR for $c \geq 1$. With

$$S(x) = \exp\big(- x^c \big)$$

we find with the help of (1.21d)

$$\mu(x) = \frac{\int\limits_x^\infty \exp\big(- u^c \big)\, du}{\exp\big(- x^c \big)} = \frac{\gamma\left(\frac{1}{c},\, x^c \right)}{c \exp\big(- x^c \big)}, \tag{2.28}$$

where

$$\gamma(a,\, z) = \int\limits_z^\infty u^{a-1}\, e^{-u}\, du$$

[17] For a proof in the continuous case see BRYSON/SIDDIQUI (1969).

is the complementary incomplete gamma function.[18] We have

$$\frac{d\mu(x)}{dx} = \frac{x^c \exp\big(x^c\big)\,\gamma\Big(\dfrac{1}{c},\,x^c\Big) - \big(x^c\big)^{1/c}}{x} \begin{cases} \leq 0 \text{ for } c \geq 1 \;\Rightarrow\; \text{DMRL}, \\ \geq 0 \text{ for } 0 < c \leq 1 \;\Rightarrow\; \text{IMRL}. \end{cases}$$

Fig. 2/4 shows the hazard rate functions and the corresponding mean residual life functions.

A discrete distribution analyzed by KEMP (2004) is the zero–inflated geometric distribution with $0 < \lambda < 1$ and $0 < \alpha < 1$. We have

$$h_0 = 1 - \alpha\,\lambda, \quad h_i = 1 - \lambda \text{ for } i \geq 1, \text{ and} \tag{2.29a}$$

$$S_0 = 1, \quad S_1 = \alpha\,\lambda, \quad S_i = \alpha\,\lambda^i \text{ for } i > 1. \tag{2.29b}$$

Thus

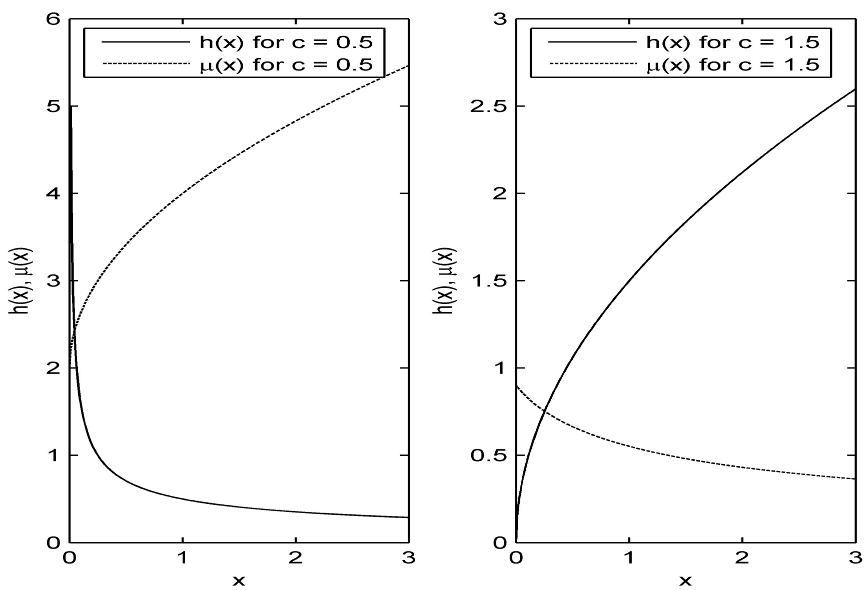
$$\left.\begin{aligned} L_0 &= \frac{\sum\limits_{j>0} S_j}{S_0} = \frac{\alpha\,\lambda}{1-\lambda}, \\ L_i &= \frac{\sum\limits_{j>i} S_j}{S_i} = \frac{\lambda}{1-\lambda}, \quad i > 0. \end{aligned}\right\} \tag{2.29c}$$

Now

$$h_0 > h_1 = h_2 = \ldots \text{ and } L_0 < L_1 = L_2 = \ldots$$

Therefore, the zero–inflated geometric distribution is DHR and IMRL.

Figure 2/4: HR and corresponding MRL of two reduced WEIBULL distributions



From the fact that IHR ⇒ DMRL and DHR ⇒ IMRL one may conjecture that DIHR implies IDMRL, i.e., a bathtub–shaped hazard rate implies an upside–down MRL. The following Theorem 19 is given by and proved by MI (1995).

Theorem 19: If the hazard rate function $h(x)$ has a bathtub shape, then the associated MRL has an upside–down bathtub shape. ■

Fig. 2/5 demonstrate Theorem 19 for the exponential power distribution of DHILLON (1981) with the hazard rate given in (2.24a), which is DIHR for $0 < b < 1$.

[18] $\Gamma(a,\,z) = \int\limits_0^z u^{a-1}\,e^{-u}\,du$ is the incomplete gamma function.

Figure 2/5: HR function and MRL function of a DHILLON–I distribution $(a = 0,\ b = 1,\ c = 0.5)$

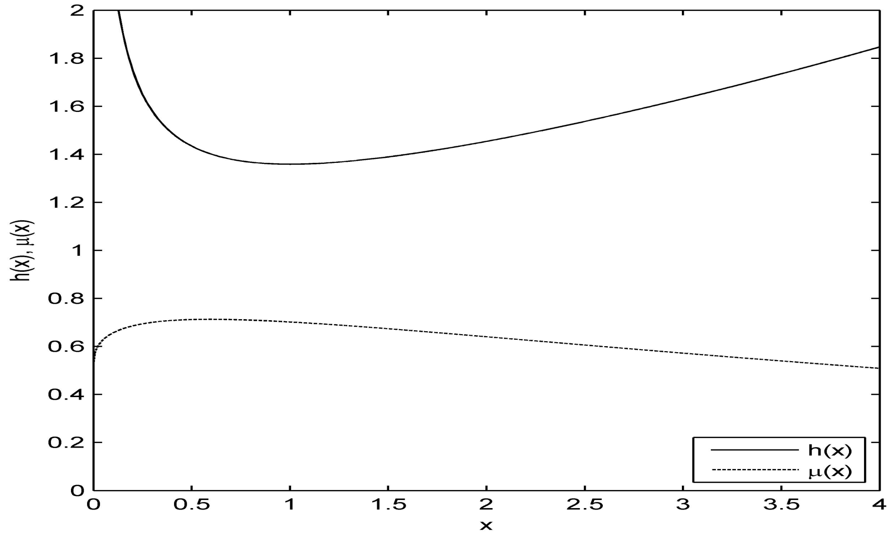
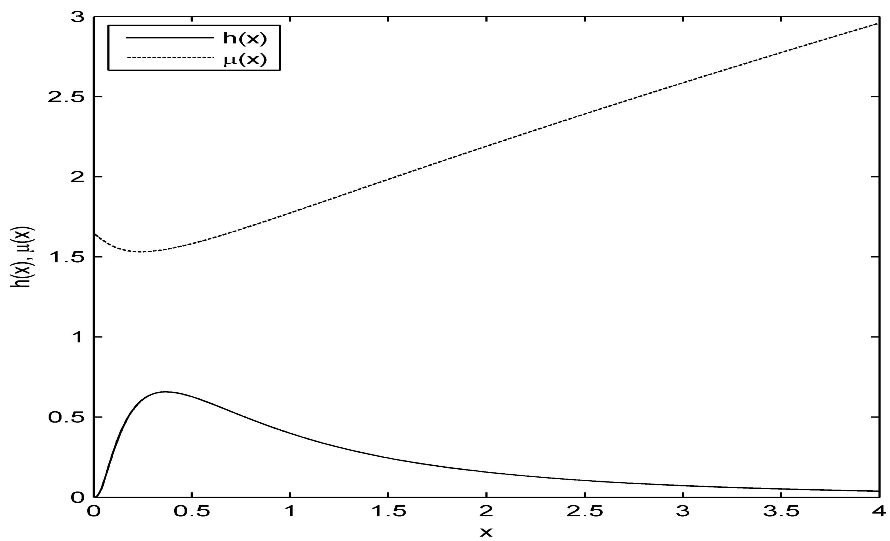


Figure 2/6: HR function and MRL function of a log–normal distribution $(a = 0,\ b = 1)$



The converse of Theorem 19 is not necessarily true as demonstrated by the following example of MI (1995):

$$\mu(x) = \begin{cases} x^2 + 1 & \text{for } 0 \le x < 1, \\ 2\,x & \text{for } 1 \le x < 2, \\ 4\,\exp\big[-0.25\,(x-2)\big] & \text{for } x \ge 2. \end{cases}$$

This $\mu(x)$ has an upside–down bathtub shape and is a MRL function of a certain lifetime distribution. Applying (1.20d) we find the corresponding hazard rate function:

$$h(x) = \begin{cases} \dfrac{2\,x+1}{x^2+1} & \text{for } 0 \le x < 1, \\ \dfrac{3}{2\,x} & \text{for } 1 \le x < 2, \\ 0.25\,\big\{\exp\big[0.25\,(x-2)\big] - 1\big\} & \text{for } x \ge 2. \end{cases}$$

This hazard rate is not bathtub–shaped, instead it is bathtub–shaped over $[0,\, 2)$, drops down to $0.25/e \approx 0.0920$ at $x = 2$ and increases to infinity over $(2,\, \infty)$.

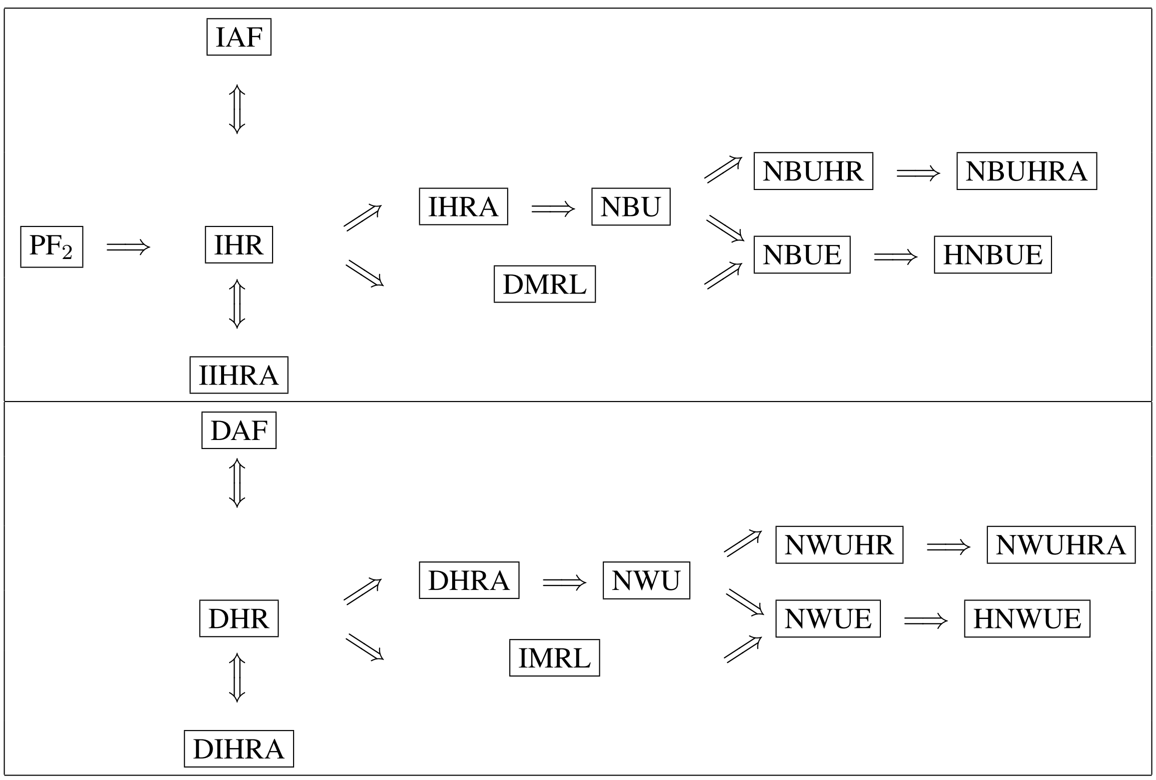
From Theorem 19 we may conjecture that a distribution with bathtub–shaped hazard rate might have a bathtub–shaped mean residual life function. As Fig. 2/6 shows this conjecture holds at least for the log–normal distribution.

The reader interested in DIHR and IDHR residual life functions of discrete distributions should consult GUESS/PARK (1988).

2.4 Classification According to other Aging Criteria[19]

We start this section by a figure showing the most common aging criteria and the way they are linked, i.e., which criterion is implied by which other criterion, or stated otherwise, which criterion defines a subclass of distributions of what other criterion's subclass. The chain in the upper part refers to positive aging while the chain in the lower part refers to negative aging. A proof of the implications in Fig. 2/7 can be found in BRYSON/SIDDIQUI (1969) and in KLEFSJÖ (1982a).

Figure 2/7: Chains of implications for several aging criteria



PF_2 means PÓLYA density of order 2 and is the strongest aging criterion.

We now present and discuss those criteria in Fig. 2/7 which have not been described in the preceding sections and we start on the left–hand side and move to the right. The criteria **IAF** and **DAF** have been introduced by BRYSON/SIDDIQUI (1969), and IAF (DAF) stands for **increasing (decreasing) specific aging factor**. The specific aging factor is defined as

$$A(x,\, y) := \frac{S(x)\, S(y)}{S(x+y)};\ x,\, y \geq 0. \tag{2.30a}$$

[19] Suggested reading for the section: BARLOW/PROSCHAN (1975), BRYSON/SIDDIQUI (1969), JOHNSON/KOTZ/BALAKRISHNAN (1995, pp. 663 ff.), KEMP (2004), KLEFSJÖ (1982a,b), MARSHALL/PROSCHAN (1972).

Notice the interchangeability of the arguments x and y and the relationship to NBU and NWU in (2.32). If a distribution is NBU (NWU), its specific aging factor results as $A(x,\,y) \underset{(\leq)}{\geq} 1$. The motivation for $A(x,\,y)$ may be seen as follows: Consider two units with lifetimes described by one and the same distribution $F(\cdot)$, and let x denote the chronological age of one of them. The other unit is 'new', i.e., has a chronological age of zero. Then $S(y)$ is the probability that the new unit will survive for at least a duration y. Correspondingly, the ratio $S(x+y)\big/S(x)$ is the probability that the older unit will survive for that same duration, given its prior survival up to time x. The specific aging factor is the a comparison of these two survival probabilities. It will be strictly greater than unity if and only if the older unit has 'aged' in that it has less chance of surviving for duration y than does a new unit. The range of $A(x,\,y)$ is the extended positive real line; however, it is undefined if either numerator factor vanishes.

Definition: A distribution function $F(\cdot)$ is called IAF if

$$A(x_2,\,y) \geq A(x_1,\,y)\ \ \forall\, y \geq 0,\ x_2 \geq x_1 \geq 0, \tag{2.30b}$$

and DAF if

$$A(x_2,\,y) \leq A(x_1,\,y)\ \ \forall\, y \geq 0,\ x_2 \geq x_1 \geq 0. \quad \blacksquare \tag{2.30c}$$

We now prove that criteria IAF and IHR are equivalent.

Proof: IAF specifies that $A(x,\,y)$ is an increasing function of x for all y. Differentiating,

$$\frac{dA(x,\,y)}{dx} = \frac{S(x+y)\,S'(x) - S(x)\,S'(x+y)}{S(x)\,S(x+y)} \geq 0.$$

Hence,

$$\frac{S'(x+y)}{S(x+y)} \leq \frac{S'(x)}{S(x)}$$

or

$$h(x+y) \geq h(x).$$

Since this holds for all $y \geq 0,\ h(\cdot)$ is increasing and IHR holds. Conversely, if IHR holds, then the foregoing steps may be reversed to show that $A(x,\,y)$ is an increasing function of x. $\blacksquare$
The equivalence of DHR and DAF can be proved along the same lines.

An obvious generalization of the hazard rate average $HRA(x)$, see (2.12a), is the **specific interval–average hazard rate**

$$HRA(x,\,y) := \frac{1}{x}\int\limits_{y}^{x+y} h(u)\,du = \frac{H(x+y) - H(x)}{x}. \tag{2.31a}$$

We have

$$HRA(x,\,0) = HRA(x) = \frac{1}{x}\int\limits_{0}^{x} h(u)\,du. \tag{2.31b}$$

Definition: A distribution $F(\cdot)$ is called **IIHRA (increasing interval–average hazard rate)** if

$$HRA(x_2,\,y) \geq HRA((x_1,\,y)\ \ \forall\, x_2 \geq x_1 \geq 0,\ y \geq 0, \tag{2.31c}$$

and **DIHRA (decreasing interval–average hazard rate)** if

$$HRA(x_2,\,y) \leq HRA(x_1,\,y)\ \ \forall\, x_2 \geq x_1 \geq 0,\ y \geq 0. \quad \blacksquare \tag{2.31d}$$

To prove the equivalence of criteria IHR and IIHRA we need the following

Lemma: Let $h(x)$ be integrable with no more than finitely many discontinuities in any finite interval. Then $h(x)$ is monotone increasing for all $x > 0$ if and only if

$$h(y) \le \frac{1}{x} \int\limits_{y}^{y+x} h(u)\,du \le h(y+x) \;\; \forall\, y \ge 0,\; x > 0. \qquad \blacksquare$$

Proof of IHR $\Leftrightarrow$ IIHRA: In accordance with IHR, $h(x_2) \ge h(x_1) \;\; \forall\, x_2 \ge x_1 \ge 0$, let $h(\cdot)$ be monotone increasing, and choose $x_2 \ge x_1$. Then

$$\begin{aligned} HRA(x_2,\,y) - HRA(x_1,\,y) &= \left(\frac{1}{x_2} - \frac{1}{x_1}\right) \int\limits_{y}^{y+x_2} h(u)\,du + \frac{1}{x_2} \int\limits_{y+x_1}^{y+x_2} h(u)\,du \\ &= \frac{x_2 + x_1}{x_2} \left(\frac{1}{x_2 - x_1} \int\limits_{y+x_1}^{y+x_2} h(u)\,du - \frac{1}{x_1} \int\limits_{y}^{y+x_1} h(u)\,du \right) \\ &\ge \frac{x_2 - x_1}{x_2} \big[h(y+x_1) - h(y+x_1) \big] \end{aligned}$$

by the above Lemma. Hence

$$HRA(x_2,\,y) \ge HRA(x_1,\,y).$$

Conversely, suppose $HRA(x_2,\,y) \ge HRA(x_1,\,y) \;\; \forall\, y \ge 0$ and $x_2 \ge x_1$. Then

$$\frac{1}{x_2} \int\limits_{y}^{x+x_2} h(u)\,du \ge \frac{1}{x_1} \int\limits_{y}^{y+x_1} h(u)\,du.$$

In particular, if $x_1 \to 0$, then the right–hand side becomes

$$\lim_{x \to 0} \frac{\int\limits_{x}^{x+x} h(u)\,du}{x} = h(y^+),$$

so that, for any positive x_2

$$\frac{1}{x_2} \int\limits_{y}^{y+x_2} h(u)\,du \ge h(y^+).$$

Since y and x_2 are arbitrary, the above Lemma applies to prove the monotonicity of $h(\cdot)$. $\blacksquare$

The equivalence of DHR and DIHRA in the lower chain of Fig. 2/7 can be proved along the same lines.

The probability of an x–survivor living another y units of time is

$$S(y \,|\, x) = \Pr(Y > y \,|\, X \ge x) = \frac{S(x+y)}{S(x)},$$

whereas the probability of a new unit living more than y units of time is

$$\Pr(Y > y \,|\, X \ge 0) = S(y).$$

Definition:[20] A lifetime distribution is said to be **NBU (new better than than used)** or **NWU (new worse than used)** accordingly as the conditional survival probability $S(y\,|\,x)$ is is less (or greater) than the unconditional survival probability $S(y)$, i.e.,

$$S(y) \underset{(\leq)}{\geq} \frac{S(x+y)}{S(x)} \;\; \forall\, x \geq 0,\; y \geq 0. \qquad \blacksquare \qquad (2.32)$$

We will prove the implications IHR ⇒ NBU and DHR ⇒ NWU for a discrete distribution, where NBU (NWU) means

$$\frac{S_{i+j}}{S_i} \underset{(\geq)}{\leq} S_j,\;\; i \neq j = 0,\, 1,\, \ldots$$

Proof: From (1.61c) we have

$$\frac{S_{i+j}}{S_i\,S_j} = \frac{\prod\limits_{k=0}^{i+j-1}(1-h_k)}{\prod\limits_{k=0}^{i-1}(1-h_k)\prod\limits_{k=0}^{j-1}(1-h_k)} = \prod_{k=0}^{j-1}\frac{1-h_{i+k}}{1-h_k}.$$

If h_i increases with i, then

$$\frac{1-h_{i+j}}{1-h_j} < 1$$

and

$$\frac{S_{i+j}}{S_i} < S_j.$$

Similarly, if h_i decreases with i, then

$$\frac{S_{i+j}}{S_i} > S_j.$$

It is to be noticed that $F(\cdot)$ is NBU (NWU) if and only if $Y\,|\,x$ is stochastically smaller (greater) than X. Furthermore, we have

Theorem 20:

a) $F(\cdot)$ is NBU if and only if $H(\cdot)$ is **superadditive**, i.e.,

$$H(x+y) \geq H(x) + H(y) \;\text{ for }\; x,\, y > 0.$$

b) $F(\cdot)$ is NWU if and only if $H(\cdot)$ is **subadditive**, i.e.,

$$H(x+y) \leq H(x) + H(y) \;\text{ for }\; x,\, y > 0. \qquad \blacksquare$$

For a continuous distribution it is easily shown that NBU (NWU) can be equivalently stated as

$$S(y) \underset{(\leq)}{\geq} \frac{S(x+y)}{S(x)}$$

[20] HOLLANDER/PARK/PROSCHAN (1986) introduced subclasses of of the NBU (NWU) distributions, called new better (worse) than used of age x_0. For these classes the survival probability at age 0 is greater (smaller) than or equal to the conditional survival probability at specified age $x_0 > 0$:

$$S(x+x_0) \underset{(\geq)}{\leq} S(x)\,S(x_0) \;\; \forall\, x > 0.$$

They presented preservation and non–preservation properties of the two classes under various reliability operations and also showed how to test whether or not a distribution is new better than than used at age x_0.

and

$$\int_0^x h(u)\,du \underset{(\geq)}{\leq} \int_y^{x+y} h(u)\,du;\ \ x,\,y>0.$$

But for discrete NBU (NWU) distributions the corresponding equivalence of

$$S_j \underset{(\leq)}{\geq} \frac{S_{i+j}}{S_i};\ \ i,\,j=0,\,1,\,\ldots$$

and

$$\sum_{k=0}^{j} h_k \underset{(\geq)}{\leq} \sum_{k=i}^{i+j} h_k$$

does not hold.

From the mean residual life of an x–survivor

$$\mu(x)=\mathrm{E}(Y\,|\,x\geq x)=\frac{1}{S(x)}\int_x^\infty S(u)\,du,$$

and the mean life of a new unit

$$\mu(0)=\mathrm{E}(X)=\int_0^\infty S(u)\,du$$

we have the following

Definition: A distribution $F(\cdot)$ is called **NBUE (new better than used in expectation)** if

$$\left.\begin{array}{l}\mu(0)\geq\mu(x),\ \ x>0,\\ \text{or equivalently}\\ \int_x^\infty S(u)\,du\leq\mu(0)\,S(x),\ \ x>0.\end{array}\right\}\qquad(2.33a)$$

A distribution $F(\cdot)$ is called **NWUE (new better than used in expectation)** if

$$\left.\begin{array}{l}\mu(0)\leq\mu(x),\ \ x>0,\\ \text{or equivalently}\\ \int_x^\infty S(u)\,du\geq\mu(0)\,S(x),\ \ x>0.\end{array}\right\}\qquad\blacksquare\qquad(2.33b)$$

We will prove the implication NBU $\Rightarrow$ NBUE (NWU $\Rightarrow$ NWUE) for a discrete distribution.

Proof: If

$$S_{i+j} \underset{(\geq)}{\leq} S_i\,S_j;\ \ j=0,\,1,\,\ldots$$

then

$$\sum_{j\geq 0} S_{i+j} \underset{(\geq)}{\leq} S_i \sum_{j\geq 0} S_j.\qquad\blacksquare$$

Definition: A continuous distribution $F(\cdot)$ with finite mean $\mu=\int_0^\infty S(u)\,du$ is said to be **HNBUE (harmonic new better than used in expectation)** if

$$\int_x^\infty S(u)\,du\leq\mu\,\exp\!\left(-\frac{x}{\mu}\right)\ \ \text{for}\ \ x\geq 0.\qquad\blacksquare\qquad(2.34a)$$

If the reversed inequality is true, $F(\cdot)$ is said to be **HNWUE (harmonic new worse than used in expectation)**. This gives a dual class to the HNBUE class of distributions in the same way as the IHR, IHRA, NBU, NBUE, and DMRL classes have their duals. The HNBUE and HNWUE classes have been introduced by ROLSKI (1975). The names HNBUE resp. HNWUE are to be explained as follows. Starting from the mean residual life

$$\mu(x) = \frac{1}{S(x)} \int_x^\infty S(u)\,du$$

then the inequality (2.34a) can be written as[21]

$$\frac{1}{\dfrac{1}{x}\displaystyle\int_0^x \frac{1}{\mu(u)}\,du} \le \mu \ \text{ for } \ x > 0. \tag{2.34b}$$

This inequality says that the integral harmonic mean of $\mu(u),\ 0 \le u \le x$, is less than or equal to the integral harmonic mean of μ. For HNBUE distributions we have

$$S(x) \le \left\{ \begin{array}{cl} 1 & \text{for } x \le \mu \\ \exp\left(\dfrac{\mu - x}{\mu}\right) & \text{for } x > \mu \end{array} \right\} \tag{2.34c}$$

and for HNWUE distributions

$$S(x) \le \frac{\mu}{x}\left[1 - \exp\left(-\frac{x}{\mu}\right)\right] \ \text{ for } \ x \ge 0 \tag{2.34d}$$

holds.

Definition: For an absolutely continuous distribution $F(\cdot)$ with hazard rate $h(\cdot)$, we say that $F(\cdot)$ is **NBUHR (NWUHR) — new better (worse) than used in hazard rate** — if

$$h(x) \underset{(\le)}{\ge} h(0) \ \ \forall\, x \ge 0. \qquad \blacksquare \tag{2.35}$$

Definition: For an absolutely continuous distribution $F(\cdot)$, we say that $F(\cdot)$ is **NBUHRA (NWUHRA) — new better (worse) than used in hazard rate average** — if

$$h(0) \underset{(\ge)}{\le} \frac{1}{x}\int_0^x h(u)\,du = -\frac{\ln\left[1 - F(x)\right]}{x} \ \ \forall\, x \ge 0. \qquad \blacksquare \tag{2.36}$$

[21] Notice that from (2.34a) we first have

$$S(x)\,\frac{\mu(x)}{\mu} \le \exp\left(-\frac{x}{\mu}\right).$$

Using (1.19e) we then find

$$\exp\left\{-\int_0^x \frac{1}{\mu(u)}\,du\right\} \le \exp\left(-\frac{x}{\mu}\right)$$

and

$$-\int_0^x \frac{1}{\mu(u)}\,du \le -\frac{x}{\mu}.$$

Rearranging the latter inequality results into (2.34a).

In the following Table 2/1 we have compiled from HOLLANDER/PROSCHAN (1984) and KLEFSJÖ (1982a) the results pertaining to the closure of classes of lifetime distributions under three reliability operations: mixture and convolution of distributions and formation of coherent systems.

Table 2/1: Closure and inheritance of classes of lifetime distributions under reliability operations

Class	Mixture of distributions	Convolution of distributions	Formation of coherent systems
IHR	not closed	closed	not closed
IHRA	”	”	closed
NBU	”	”	”
NBUE	”	”	not closed
DMRL	”	not closed	”
HNBUE	”	closed	”
DHR	closed	not closed	not closed
DHRA	”	”	”
NWU	not closed	”	”
NWUE	”	”	”
IMRL	closed	”	”
HNWUE	”	”	”

3 Presentation of Univariate Parametric Distributions

In this chapter we have compiled for a great number of univariate parametric distributions

- the PDF $f(x)$ in the continuous case or the PMF P_i in the discrete case,
- the CCDF $S(x)$ or S_i,
- the HR $h(s)$ or h_i, and
- the MRL $\mu(x)$ or L_i, if existing.

We also indicate the aging class. Not all distributions may serve as lifetime distributions.

3.1 Continuous Distributions[1]

For those distributions defined on $\mathbb{R}$ or with negative argument x we have not given results for the variant resulting from lower truncation at $x = 0$. The formulas for such a truncated distribution easily follow by applying (1.52a-e). We have applied the following notation for parameters:

- $a \in \mathbb{R}$ for a location parameter which causes a shift of the distribution along the abscissa,
- $b > 0$ a scale parameter causing a stretching or a shrinkage of the distribution,
- any other Latin or Greek letter for a parameter that affects the shape of the distribution.

The aging property does not depend on a and b.

We use the following mathematical and statistical functions:

$$B(c,d) = \frac{\Gamma(c)\,\Gamma(d)}{\Gamma(c+d)} = \int\limits_0^1 t^{c-1}\,(1-t)^{d-1}\,dt \text{ – complete beta function}$$

$$B(P,c,d) = \int\limits_0^P t^{c-1}\,(1-t)^{d-1}\,dt \text{ – incomplete beta function}$$

$$\Gamma(a) = \int\limits_0^\infty t^{a-1}\,e^{-t}\,dt \text{ – complete gamma function}$$

$$\Gamma(a,\,u) = \int\limits_0^u t^{a-1}\,e^{-t}\,dt \text{ – incomplete gamma function}$$

$$\gamma(a,\,u) = \int\limits_u^\infty t^{a-1}\,e^{-t}\,dt \text{ – complementary incomplete gamma function}$$

[1] Suggested reading for this section: JOHNSON/KOTZ/BALAKRISHNAN (1994, 1995), LEEMIS(1995), MEEKER/ESCOBAR (1998), PATEL (1973), RINNE (2009, 2010). The interactive program *ContDist* which is written in MATLAB and which is included in the accompanying file ‘Distributions.zip’ displays for all distributions presented here a graph of the functions $f(x)$, $S(x)$, $h(x)$ and — if existing — $\mu(x)$ for any set of user–chosen parameter values.

$$\Phi(u) = \int\limits_{-\infty}^{u} \frac{1}{\sqrt{2\,\pi}} \exp[-t^2/2]\,dt - \text{CDF of the standardized normal distribution}$$

$$\phi(u) = \frac{1}{\sqrt{2\,\pi}} \exp[-u^2/2]\,du - \text{PDF of the standardized normal distribution}$$

$$\operatorname{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int\limits_{0}^{x} \exp\big(-u^2\big)\,du - \text{error function}$$

Alpha distribution

This distribution is related to the normal distribution in the following way: Consider $Y \sim No(\mu;\sigma)$, truncated to the left of $y = 0$. Then, $X = 1/Y$ has an alpha distribution with parameters $\alpha = \mu/\sigma$ and $b = 1/\sigma$. The alpha distribution has been applied to tool wear and has also been suggested in modeling lifetimes under accelerated life testing, see SALVIA (1985):

$$f(x) = \frac{b\,\exp\left[-0.5\,\left(\alpha - \dfrac{b}{x}\right)^2\right]}{\sqrt{2\,\pi}\,\Phi(\alpha)\,x^2};\ x \geq 0,\ \alpha \in \mathbb{R},\ b > 0$$

$$S(x) = 1 - \frac{\Phi\left(\alpha - \dfrac{b}{x}\right)}{\Phi(\alpha)}$$

$$h(x) = \frac{b\,\exp\left[-0.5\,\left(\alpha - \dfrac{b}{x}\right)^2\right]}{\sqrt{2\,\pi}\,\left[\Phi(\alpha) - \Phi\left(\alpha - \dfrac{b}{x}\right)\right]\,x^2} \Rightarrow \text{IDHR}$$

$$\mu(x) - \text{does not exist}$$

The PDF has its mode at $x = b\left(\sqrt{\alpha^2 + 8} - \alpha\right)/4$ which moves to the left (right) as α (b) increases.

Arcsine distribution

This distribution is a special case of the beta distribution (see below) with shape parameters $c = d = 0.5$.[2] The name is derived from the fact that its CDF and CCDF are written in terms of the arcsine function, the inverse of the sine function. The arcsine distribution with $a = 0$ and $b > 0$ having support $[-b;\ b]$ gives the position at random time of a particle engaged in simple harmonic motion with amplitude $b > 0$.

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

$$S(x) = \frac{1}{2} - \frac{\arcsin\left(\dfrac{x-a}{b}\right)}{\pi}$$

[2] A beta distribution with $c + d = 1$, but $c \neq 0.5$ is sometimes called a generalized arcsine distribution.

$$h(x) = \frac{1}{b\sqrt{1-\left(\dfrac{x-a}{b}\right)^2}\,\arccos\left(\dfrac{x-a}{b}\right)} \Rightarrow \text{DIHR}$$

$$\mu(x) = (a-x) + \frac{\sqrt{a+b-x)\,(x+b-a)}}{\arccos\left(\dfrac{x-a}{b}\right)} \Rightarrow \text{IDMRL}$$

The arcsine distribution is a location–scale distribution. The PDF is symmetric and U–shaped, its minimum is at $x = a$ with $f(a) = (b\,\pi)^{-1}$. The bathtub–shaped hazard rate has its minimum at $x^* \approx a - 0.4421\,b$ with $h(x^*) \approx (1.8197\,b)^{-1}$. The upside–down shaped MRL has its maximum at x^+ being the solution of $\sqrt{(a+b-x^+)\,(x^+ + b - a)} - (a - x^+)\,\arccos\left(\frac{x^+-a}{b}\right) = \sqrt{(a+b-x^+)\,(x^+ + b - a)}\,\left[\arccos\left(\frac{x^+-a}{b}\right)\right]^2$.

Beta distribution

The name of this distribution has its origin in the complete beta function $B(c, d)$ which is part of the formulas for the PDF and other distribution representatives. The PDF of a beta distribution can take on a great variety of shapes depending on its two shape parameters $c,\ d \in \mathbb{R}$:

- symmetric for $c = d$,
- unimodal for $c > 1$ and $d > 1$,
- U–shaped for $c < 1$ and $d < 1$,
- J–shaped for $d \le 1 \le c$, but $c \ne d$,
- inversely (or reflected) J–shaped for $c \le 1 \le d$, but $c \ne d$.

The beta distribution also includes several other distributions as special cases, e.g.,

- the uniform distribution for $c = d = 1$,
- the right–angled negatively (positively) skewed triangular distribution for $d = 1$ and $c = 2$ ($c = 1$ and $d = 2$),
- the arcsine distribution for $c = d = 0.5$,
- the power function distribution for $d = 1,\ c > 0$.

$$f(x) = \frac{1}{B(c,d)}\,\frac{(x-a)^{c-1}\,(a+b-x)^{d-1}}{b^{c+d-1}};\ a \le x \le a+b;\ a \in \mathbb{R};\ b,\ c,\ d > 0$$

$$\begin{aligned} S(x) &= 1 - \frac{1}{B(c,d)} \int\limits_0^{(x-a)/b} u^{c-1}\,(1-u)^{d-1}\,du \\ &= 1 - \mathrm{I}_{\left(\frac{x-a}{b}\right)}(c,d) \ = \ \mathrm{I}_{\left(1-\frac{x-a}{b}\right)}(d,c) \end{aligned}$$

The function $\mathrm{I}_z(c, d)$ is the incomplete beta function ratio which has to be evaluated numerically.

$$h(x) \quad - \quad \text{no closed form} \Rightarrow \begin{cases} \text{DIHR for } 0 < c \lesssim 0.8 \text{ and } d \text{ arbitrarily} \\ \text{IHR for all other combinations of } c \text{ and } d \end{cases}$$

$$\mu(x) \quad - \quad \text{no closed form} \Rightarrow \begin{cases} \text{IDMRL for } 0 < c \lesssim 0.8 \text{ and } d \text{ arbitrarily} \\ \text{DMRL for all other combinations of } c \text{ and } d \end{cases}$$

BIRNBAUM–SAUNDERS distribution

This distribution has been suggested by BIRNBAUM/SAUNDERS (1968, 1969) as a lifetime model for materials subject to cyclic patterns of stress where the ultimate failure comes from the growth of prominent flaws.

$$f(x) = \frac{\sqrt{\dfrac{x}{b}} + \sqrt{\dfrac{b}{x}}}{2\,c\,x\,\sqrt{2\,\pi}} \exp\left[-\frac{1}{2\,c^2}\left(\sqrt{\frac{x}{b}} - \sqrt{\frac{b}{x}}\right)^2\right];\; x \geq 0;\; b,\,c > 0$$

The variate $Y = \left(\sqrt{x/b} - \sqrt{b/x}\,\right)/c$ has a standard normal distribution, so

$$S(x) = \Phi\left[-\frac{1}{c}\left(\sqrt{\frac{x}{b}} - \sqrt{\frac{b}{x}}\right)\right].$$

$h(x) = f(x)/S(x)$ has no closed form, but it is IDHR with a maximum at $x^* \approx b/(-0.4604 + 1.8417\,c)^2$ which is close to zero for b small and c large, so that the upside–down bathtub shape does not come up clearly and the hazard rate seems to be DHR. The IDHR pattern is best seen for b and c around 1. For $h(x)$ we further notice:

- $h(0) = 0$ and
- $\displaystyle\lim_{x\to\infty} h(x) = \frac{1}{2\,b\,c^2}$.

$\mu(x)$ has no closed form and it is DIMRL for b and c around 1, otherwise IMRL.

BURR distribution of type XII

BURR (1942) has suggested a number of forms of a CDF that might be useful for purposes of graduation. Special attention has been devoted to type XII.

$$f(x) = \frac{c\,d}{b}\left(\frac{x-a}{b}\right)^{d-1}\left[1 + \left(\frac{x-a}{b}\right)^d\right]^{-c-1};\; x \geq a;\; a \in \mathbb{R};\; b,\,c,\,d > 0$$

$$S(x) = \left[1 + \left(\frac{x-a}{b}\right)^d\right]^{-c};$$

$$h(x) = \frac{c\,d}{b}\left(\frac{x-a}{b}\right)^{d-1}\left[1 - \left(\frac{x-a}{b}\right)^d\right]^{-1} \Rightarrow \begin{cases} \text{DHR for } 0 < d \leq 1, \\ \text{IDHR for } d > 1 \end{cases}$$

$$\mu(x) - \text{no closed form} \Rightarrow \begin{cases} \text{IMRL for } 0 < d \leq 1 \\ \text{DIMRL for } d > 1 \\ \text{does not exist for } c\,d \leq 1 \end{cases}$$

The hazard rate has — for $d > 1$ — a maximum at $x^* = a + b\,(d-1)^{1/d}$ with $h(x^*) = (c/b)\,(d-1)^{1-1/d}$.

CAUCHY distribution

The CAUCHY distribution is a symmetric distribution defined on $\mathbb{R}$. Moments and thus MRL do not exist.

$$
\begin{aligned}
f(x) &= \left\{\pi\, b\left[1+\left(\frac{x-a}{b}\right)^2\right]\right\}^{-1};\ x\in\mathbb{R}; a\in\mathbb{R},\, b>0\\
S(x) &= \frac{1}{2}-\frac{1}{\pi}\,\arctan\left(\frac{x-a}{b}\right)\\
h(x) &= \frac{1}{b\left[1+\left(\dfrac{x-a}{b}\right)^2\right]\left[\dfrac{\pi}{2}-\arctan\left(\dfrac{x-a}{b}\right)\right]} \Rightarrow \text{IDHR}
\end{aligned}
$$

The hazard rate has its maximum at $x^*\approx a+0.428978\,b$ with $h(x^*)\approx 0.7246/b$.

χ distribution

The χ distribution is related to the χ^2 distribution. For $\nu=2$ the χ distribution is equal to the RAYLEIGH distribution.

$$
\begin{aligned}
f(x) &= \frac{x^{\nu-1}\,\exp\left(-\dfrac{x^2}{2}\right)}{2^{\nu/2-1}\Gamma\left(\dfrac{\nu}{2}\right)};\ x\geq 0;\ \nu>0\\
S(x) &= 1-\frac{\Gamma(\nu/2,\,x^2/2)}{\Gamma\left(\dfrac{\nu}{2}\right)};\\
h(x) &= \frac{x^{\nu-1}\,\exp\left(-\dfrac{x^2}{2}\right)}{2^{\nu/2-1}\left[\Gamma\left(\dfrac{\nu}{2}\right)-\Gamma(\nu/2,\,x^2/2)\right]} \Rightarrow \begin{cases}\text{DIHR for } 0<\nu<1\\ \text{IHR for } \nu\geq 1\end{cases}\\
\mu(x) &- \text{no closed form} \Rightarrow \begin{cases}\text{IDMRL for } 0<\nu<1\\ \text{DMRL for } \nu\geq 1\end{cases}
\end{aligned}
$$

χ^2 distribution

The χ^2 distribution is a special case of the gamma distribution (see below) with $b=2$.

$$
\begin{aligned}
f(x) &= \frac{x^{\nu/2-1}\,\exp\left(-\dfrac{x}{2}\right)}{2^{\nu/2}\Gamma\left(\dfrac{\nu}{2}\right)};\ x\geq 0;\ \nu>0\\
S(x) &= 1-\frac{\Gamma(\nu/2,\,x/2)}{\Gamma\left(\dfrac{\nu}{2}\right)};\\
h(x) &= \frac{x^{\nu/2-1}\,\exp\left(-\dfrac{x}{2}\right)}{2^{\nu/2}\left[\Gamma\left(\dfrac{\nu}{2}\right)-\Gamma(\nu/2,\,x/2)\right]} \Rightarrow \begin{cases}\text{DHR for } 0<\nu<2\\ \text{0.5 for } \nu=2 \text{ (exponential distribution)}\\ \text{IHR for } \nu>2\end{cases}
\end{aligned}
$$

$$\mu(x) \quad - \quad \text{no closed form} \Rightarrow \begin{cases} \text{IMRL for } 0 < \nu < 2 \\ 2 \text{ for } \nu = 2 \\ \text{DMRL for } \nu > 2 \end{cases}$$

We notice that $\lim_{x\to\infty} h(x) = 0.5$ and $\lim_{x\to\infty} \mu(x) = 2$.

Cosine distribution, ordinary

This distribution has a convex PDF and thus is no good approximation to the bell–shaped PDF of the normal distribution.

$$f(x) = \frac{1}{2\,b}\cos\!\left(\frac{x-a}{b}\right);\; a - \frac{b\,\pi}{2} \le x \le a + \frac{b\,\pi}{2};\; a \in \mathbb{R},\, b > 0$$

$$S(x) = 0.5\left[1 - \sin\!\left(\frac{x-a}{b}\right)\right]$$

$$h(x) = \frac{1}{b\left[\sec\!\left(\frac{x-a}{b}\right) - \tan\!\left(\frac{x-a}{b}\right)\right]} \Rightarrow \text{IHR}$$

$$\mu(x) = \frac{a + \frac{b\,\pi}{2} - x - b\,\cos\!\left(\frac{x-a}{b}\right)}{1 + \sin\!\left(\frac{a-x}{b}\right)} \Rightarrow \text{DMRL}$$

Cosine distribution, raised

This variant of the cosine distribution is bell–shaped and resembles the PDF of the normal distribution.

$$f(x) = \frac{1}{2\,b}\left[1 + \cos\!\left(\pi\,\frac{x-a}{b}\right)\right];\; a - b \le x \le a + b;\; a \in \mathbb{R},\, b > 0$$

$$S(x) = 0.5\left[1 - \frac{x-a}{b} - \frac{1}{\pi}\,\sin\!\left(\pi\,\frac{x-a}{b}\right)\right]$$

$$h(x) = \frac{1 + \cos\!\left(\pi\,\frac{x-a}{b}\right)}{b\left[1 - \frac{x-a}{b} - \frac{1}{\pi}\,\sin\left(\pi\,\frac{x-a}{b}\right)\right]} \Rightarrow \text{IHR}$$

$$\mu(x) \approx b\,\frac{0.199339 - \frac{x-a}{2\,b} + \left(\frac{x-a}{b}\right)^2 - 0.0506606\,\cos\!\left(\pi\,\frac{x-a}{b}\right)}{0.5\left[1 - \frac{x-a}{b} - \frac{1}{\pi}\,\sin\!\left(\pi\,\frac{x-a}{b}\right)\right]} \Rightarrow \text{DMRL}$$

DHILLON–I distribution

DHILLON (1979) has proposed the following hazard rate

$$h(x) = k\,\lambda\,c\,x^{c-1} + (1-k)\,\beta\,x^{\beta-1}\,b\,\exp\!\left(b\,x^{\beta}\right)$$

which is a linear combination of two components, $0 \le k \le 1$ being the combining linear factor. λ and b are the scale factors in the first and second component, respectively, while c and β are shape parameters in the two components, where $\lambda,\ b,\ c,\ \beta > 0$. This model includes several other distributions, e.g., the GOMPERTZ–MAKEHAM distribution for $c = \beta = 1$, the WEIBULL distribution for $k = 1$ and the Log–WEIBULL distribution (= extreme value distribution of type I for the minimum) for $k = 0,\ \beta = 1$, and is capable of representing different courses of the hazard rate. The DHILLON–I distribution, proposed in DHILLON (1981) results when $k = 0$ and thus is less complicated, but still models increasing and bathtub–shaped hazard rates. Introducing a location parameter $a,\ a \in \mathbb{R}$, the DHILLON–I distribution has:

$$
\begin{aligned}
f(x) &= \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1}\exp\left\{1-\exp\left[\left(\frac{x-a}{b}\right)^{c}\right]+\left(\frac{x-a}{b}\right)^{c}\right\};\ x \geq a;\ a \in \mathbb{R};\ b,\ c > 0\\
S(x) &= \exp\{1-\exp\left[\left(\frac{x-a}{b}\right)^{c}\right]\\
h(x) &= \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1}\exp\left[\left(\frac{x-a}{b}\right)^{c}\right] \Rightarrow \begin{cases}\text{DIHR for } 0 < c < 1\\ \text{IHR for } c \geq 1\end{cases}\\
\mu(x) &- \text{ no closed form} \Rightarrow \begin{cases}\text{IDMRL for } 0 < c < 1\\ \text{DMRL for } c \geq 1\end{cases}
\end{aligned}
$$

For $0 < c < 1$ the hazard rate hast its maximum at $x^* = a + b\,[(x-a)/b]^{1/c}$ with $h(x^*) = (c/b)\,[(1-c)/c]^{(c-1)/c}\,\exp[1-c)/c]$.

DHILLON–II distribution

This distribution of DHILLON (1981) is capable of generating decreasing or upside–down bathtub–shaped hazard rates.

$$
\begin{aligned}
f(x) &= \frac{c+1}{x-a+b}\left[\ln\left(\frac{x-a}{b}+1\right)\right]^{c}\exp\left\{-\left[\ln\left(\frac{x-a}{b}+1\right)\right]^{c+1}\right\};\ x \geq a;\ a \in \mathbb{R};\ b,\ c \geq 0\\
S(x) &= \exp\left\{-\left[\ln\left(\frac{x-a}{b}+1\right)\right]^{c+1}\right\}\\
h(x) &= \frac{c+1}{x-a+b}\left[\ln\left(\frac{x-a}{b}+1\right)\right]^{c} \Rightarrow \begin{cases}\text{DHR for } c = 0\\ \text{IDHR for } c > 0\end{cases}\\
\mu(x) &- \text{ no closed form} \Rightarrow \begin{cases}\text{IMRL for } c = 0\\ \text{DIMRL for } 0 < c \lesssim 2\\ \text{DMRL for } c \gtrsim 2\end{cases}
\end{aligned}
$$

For $c > 0$ the hazard rate has its maximum at $x^* = a + \big(e^c - 1\big)$ with $h(x^*) = (c/e)^c\,(c+1)\big/b$.

Double WEIBULL distribution

Upon combining the (common) WEIBULL distribution and the reflected WEIBULL distribution into one distribution we arrive at the double WEIBULL distribution. A special case of the latter distribution (for $c = 2$) is the LAPLACE distribution.

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

$$S(x) = \begin{cases} 1 - 0.5\,\exp\left[-\left(\dfrac{a-x}{b}\right)^c\right] & \text{for } x \leq a \\[2ex] 0.5\,\exp\left[-\left(\dfrac{x-a}{b}\right)^c\right] & \text{for } x \geq a \end{cases}$$

$$h(x) = \begin{cases} \dfrac{c\left(\dfrac{a-x}{b}\right)^{c-1} \exp\left[-\left(\dfrac{a-x}{b}\right)^c\right]}{b\left\{2 - \exp\left[-\left(\dfrac{a-x}{b}\right)^c\right]\right\}} & \text{for } x \leq a \\[4ex] \dfrac{c\left(\dfrac{x-a}{b}\right)^{c-1} \exp\left[-\left(\dfrac{x-x}{b}\right)^c\right]}{b\,\exp\left[-\left(\dfrac{x-a}{b}\right)^c\right]} & \text{for } x \geq a \end{cases}$$

$$\mu(x) \quad - \quad \text{no closed form}$$

The hazard rate is — except for $c = 2$ — far from being monotone; it is asymmetric around $x = a$ in any case. The mean residual life function generally decreases, but for $c \gtrsim 3$ it is not monotone.

Exponential distribution

Formerly, the exponential distribution was regarded as the prototype of a lifetime distribution.

$$\begin{aligned} f(x) &= \frac{1}{b}\exp\left(-\frac{x-a}{b}\right);\; x \geq a;\; a \in \mathbb{R},\, b > 0 \\ S(x) &= \exp\left(-\frac{x-a}{b}\right) \\ h(x) &= \frac{1}{b} \;\Rightarrow\; \text{IHR and DHR} \\ \mu(x) &= b \;\Rightarrow\; \text{IMRL and DMRL} \end{aligned}$$

Exponentiated exponential distribution

The name of this distribution is derived from the fact that its CDF is the exponentiated CDF of the exponential distribution. This distribution also goes by the name generalized exponential distribution. A special case (for $c = 1$) is the exponential distribution.

$$\begin{aligned} f(x) &= \frac{c}{b}\exp\left(-\frac{x-a}{b}\right)\left[1 - \exp\left(\frac{x-a}{b}\right)\right]^{c-1};\; x \geq a;\; a \in \mathbb{R},\, b > 0 \\ S(x) &= 1 - \left[1 - \exp\left(-\frac{x-a}{b}\right)\right]^c \\ h(x) &\;-\; \text{no closed form} \Rightarrow \begin{cases} \text{DHR for } 0 < c \leq 1 \\ \text{IHR for } c \geq 1 \end{cases} \\ \mu(x) &\;-\; \text{no closed form} \Rightarrow \begin{cases} \text{IMRL for } 0 < c \leq 1 \\ \text{DMRL for } c \geq 1 \end{cases} \end{aligned}$$

F distribution

This distribution, also known as FISHER distribution, who discovered it in the context of variance analysis, is the distribution of the ratio of two independently distributed χ^2 variables. More precisely, if $X_1 \sim \chi^2_{\nu_1}$ and $X_2 \sim \chi^2_{\nu_2}$, then $X = \dfrac{X_1/\nu_1}{X_2/\nu_2} \sim F_{\nu_1,\nu_2}$. The parameters $\nu_1,\ \nu_2$ are called degrees of freedom, but nevertheless they are not restricted to integer values.

$$f(x) = \frac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{x^{(\nu_1-2)/2}}{\left(1+\dfrac{\nu_1}{\nu_2}\,x\right)^{(\nu_1+\nu_2)/2}};\ x \ge 0;\ \nu_1,\ \nu_2 > 0$$

There are no closed formulas for $S(x),\ h(x)$ and $\mu(x)$. As $\mathrm{E}(X) = \nu_2/(\nu_2-2)$ only exists for $\nu_2 > 2$, MRL also exists only for $\nu_2 > 2$. The hazard rate is either IDHR or IHR and the mean residual life function is either DIMRL or DMRL, depending on the ν_1–ν_2–combination.

FRÉCHET distribution

This distribution is also known as the extreme value distribution of type II for the minimum.

$$\begin{aligned}
f(x) &= \frac{c}{b}\left(\frac{a-x}{b}\right)^{-c-1} \exp\left[-\left(\frac{a-x}{b}\right)^{-c}\right];\ x \le a;\ a \le 0;\ b,\ c > 0 \\
S(x) &= \exp\left[-\left(\frac{a-x}{b}\right)^{-c}\right] \\
h(x) &= \frac{c}{b}\left(\frac{a-x}{b}\right)^{-c-1} \quad \Rightarrow \ \text{IHR} \\
\mu(x) &- \ \text{no closed form} \ \Rightarrow \ \text{DMRL}
\end{aligned}$$

Gamma distribution

For $c \in \mathbb{N}$ this distribution is called ERLANG distribution.

$$\begin{aligned}
f(x) &= \frac{(x-a)^{c-1}\exp\left(-\dfrac{x-a}{b}\right)}{b^c\,\Gamma(c)};\ x \ge a;\ a \in \mathbb{R};\ b,\ c > 0 \\
S(x) &= \frac{\gamma\left[c, \dfrac{x-a}{b}\right]}{\Gamma(c)} \\
h(x) &= \frac{(x-a)^{c-1}\exp\left(-\dfrac{x-a}{b}\right)}{b^c\,\gamma\left[c, \dfrac{x-a}{b}\right]} \ \Rightarrow \ \begin{cases} \text{DHR for } \ 0 < c \le 1 \\ \text{IHR for } \ c \ge 1 \end{cases} \\
\mu(x) &- \ \text{no closed form} \ \Rightarrow \ \begin{cases} \text{IMRL for } \ 0 < c \le 1 \\ \text{DMRL for } \ c \ge 1 \end{cases}
\end{aligned}$$

Generalized exponential geometric distribution

This rather complicated looking distribution has been proposed by SILVA et al. (2010) to model different types of aging. We have introduced an additional location parameter a and changed the original scaling factor b to a scaling parameter $1/b$.

$$f(x) = \frac{c\,(1-p)\exp\left(-\dfrac{x-a}{b}\right)\left[1-\exp\left(-\dfrac{x-a}{b}\right)\right]^{c-1}}{b\left[1-p\,\exp\left(-\dfrac{x-a}{b}\right)\right]^{c-1}};\ x \ge a;\ a \in \mathbb{R};\ b,\ c > 0,\ p \in (0,1)$$

$$S(x) = 1 - \left[\frac{1-\exp\left(-\dfrac{x-a}{b}\right)}{1-p\,\exp\left(-\dfrac{x-a}{b}\right)}\right]^{c}$$

$$h(x) = \frac{c\,(1-p)\,\exp\left(-\dfrac{x-a}{b}\right)\left[1-\exp\left(-\dfrac{x-a}{b}\right)\right]^{c-1}}{b\left[1-p\,\exp\left(-\dfrac{x-a}{b}\right)\right]^{c+1}-\left[1-p\,\exp\left(-\dfrac{x-a}{b}\right)\right]^{c}\left[1-\exp\left(-\dfrac{x-a}{b}\right)\right]^{c}}$$

$\mu(x)$ – no closed form

We have

$$h(x) = \begin{cases} \text{IDHR for } p \in \left(\frac{c-1}{c+1}, 1\right) \text{ and } c > 1, \\ \text{IHR for } p \in \left(0, \frac{c-1}{c+1}\right) \text{ and } c > 1, \\ \text{DHR otherwise.} \end{cases}$$

For the MRL we have:

$$\mu(x) = \begin{cases} \text{DIMRL or multiply bended when IDHR or DHR} \\ \text{DMRL when IHR.} \end{cases}$$

Generalized gamma distribution

This generalization of the three–parameter gamma distribution by introducing a fourth parameter d goes back to STACY (1962). The generalization allows for four types of aging with respect to the hazard rate.

$$f(c) = \frac{d\,(x-a)^{c\,d-1}}{b^{c\,d}\,\Gamma(c)}\,\exp\left[-\left(\frac{x-a}{b}\right)^{d}\right];\ x \ge a;\ a \in \mathbb{R};\ b,\ c,\ d > 0$$

$$S(x) = \frac{\gamma\left[c, \left(\dfrac{x-a}{b}\right)^{d}\right]}{\Gamma(c)}$$

$$h(x) = \frac{d\,(x-a)^{c\,d-1}\,\exp\left[-\left(\dfrac{x-a}{b}\right)^{d}\right]}{b^{c\,d}\,\gamma\left[c, \left(\dfrac{x-a}{b}\right)^{d}\right]}$$

$\mu(x)$ – no closed form

The behavior of $h(x)$ and $\mu(x)$ is as follows:

- constant HR and constant MRL for $c\,d-1=0$ and d = 1,
- DHR for $\begin{cases} c\,d-1=0 \text{ and } 0<d<1 \Rightarrow \text{ IMRL} \\ c\,d-1<0 \text{ and } 0<d\le 1 \Rightarrow \text{ IMRL,} \end{cases}$
- DIHR for $c\,d-1<1$ and $d>1 \Rightarrow$ IDMRL,
- IHR for $\begin{cases} c\,d-1=0 \text{ and } d>1 \Rightarrow \text{ DMRL} \\ c\,d-1>0 \text{ and } d\ge 1 \Rightarrow \text{ DMRL,} \end{cases}$
- IDHR for $c\,d-1>0$ and $0<d<1 \Rightarrow$ IMRL, DMRL or DIMRL.

Generalized linear hazard rate distribution

This generalization of the linear hazard rate distribution (see below) includes the linear hazard rate as the special case for $c=1$. When $c=1$ and $\beta=0$ we have an exponential distribution. For other values of c the hazard rate is not linear.

$$\begin{aligned}
f(x) &= c\,(\alpha+\beta\,x)\left\{1-\exp\left[-\left(\alpha\,x+\frac{\beta}{2}\,x^2\right)\right]\right\}^{c-1}\exp\left[-\left(\alpha\,x+\frac{\beta}{2}\,x^2\right)\right];\\
&\quad x\ge 0;\ \alpha,\ \beta\ge 0,\ \text{but not } \alpha=\beta=0;\ c>0\\
S(x) &= 1-\left\{1-\exp\left[-\left(\alpha\,x+\frac{\beta}{2}\,x^2\right)\right]\right\}^{c}\\
h(x) &= \frac{c\,(\alpha+\beta\,x)\,\exp\left[-\left(\alpha\,x+\frac{\beta}{2}\,x^2\right)\right]}{\left\{1-\exp\left[-\left(\alpha\,x+\frac{\beta}{2}\,x^2\right)\right]\right\}\left\langle\left\{1-\exp\left[-\left(\alpha\,x+\frac{\beta}{2}\,x^2\right)\right]\right\}^{-c}-1\right\rangle}\\
\mu(x) &- \text{ no closed form}
\end{aligned}$$

With respect to $h(x),\ \mu(x)$ we have:

- $0<c<1,\ \beta=0 \Rightarrow$ DHR and IDMRL,
- $0<c<1,\ \beta>0 \Rightarrow$ DIHR and IDMRL,
- $c>0,\ \alpha>0 \Rightarrow$ IHR and DMRL,
- $c=1,\ \beta=0,\ \alpha>0 \Rightarrow h(x)=\alpha,\ \mu(x)=1/\alpha.$

Generalized logistic distribution

The broadest generalized logistic distribution is that of type IV.

$$\begin{aligned}
f(x) &= \frac{1}{B(c,d)}\,\frac{\exp(-d\,x/b)}{b\big[1+\exp(-x/b)\big]^{c+d}};\ x\in\mathbb{R};\ b,\ c,\ d>0\\
S(x) &= 1-\mathrm{I}_{\left(\frac{1}{1+\exp(-x/b)}\right)}(c,d),\ \text{see Beta distribution}\\
h(x) &- \text{ no closed form } \Rightarrow \text{ IHR with } \lim_{x\to\infty} h(x)=\frac{d}{b}\\
\mu(x) &- \text{ no closed form } \Rightarrow \text{ DMRL with } \lim_{x\to\infty}\mu(x)=\frac{b}{d}
\end{aligned}$$

Interchanging c and d gives the type IV generalized logistic distribution for $-X$. For $c=d$ we have the type III generalized logistic distribution. For $d=1$ we have the type I generalization, and type II results for $d=1$ and $-X$.

Generalized LOMAX distribution

This distribution is also known as generalized PARETO distribution of the second kind.

$$f(x) = \frac{1}{b}\left[1+c\,\frac{x-a}{b}\right]^{-(1+1/c)}\begin{cases} a \le x \le a-b/c \text{ for } c<0 \\ x \ge a \text{ for } c>0 \end{cases}; \; a\in\mathbb{R};\; b>0,\; c\in\mathbb{R}\setminus\{0\}$$

$$\lim_{c\to 0} f(x) = \frac{1}{b}\exp\left(-\frac{x-a}{b}\right) - \text{exponential distribution}$$

$$S(x) = \left(1+c\,\frac{x-a}{b}\right)^{-1/c} \quad \text{for } \; c\neq 0$$

$$h(x) = \frac{1}{b}\left(1+c\,\frac{x-a}{b}\right)^{-1} \;\Rightarrow\; \begin{cases}\text{IHR for } c<0 \\ \text{DHR for } c>0\end{cases}$$

$$\mu(x) = \begin{cases} \dfrac{\int_x^{\infty} S(u)\,du}{S(x)} \text{ or } \dfrac{\int_x^{a-b/c} S(u)\,du}{S(x)} \;\Rightarrow\; \begin{cases}\text{DMRL for } \; c<0 \\ \text{IMRL or IDMRL for } \; 0<c<1\end{cases} \\ \text{does not exist for } c\ge 1 \end{cases}$$

Generalized RAYLEIGH distribution

The additional shape parameter c causes different types of aging compared to the ordinary RAYLEIGH distribution which comes up for $c = 1$.

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

$$S(x) = 1-\left\{\exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right]\right\}^c$$

$$h(x) = \frac{f(x)}{S(x)} \;\Rightarrow\; \begin{cases}\text{DIHR for } \; 0<c<0.5 \\ \text{IHR for } \; c\ge 0.5\end{cases}$$

$$\mu(x) \;-\; \text{no closed form} \;\Rightarrow\; \begin{cases}\text{IDMRL for } \; 0<c<0.5 \\ \text{DMRL for } \; c\ge 0.5\end{cases}$$

GOMPERTZ distribution

This distribution has been suggested by GOMPERTZ to smooth the course of mortality rates in human life tables for higher ages.

$$f(x) = \alpha\,\exp(\beta\,x)\,\exp\left\{\frac{\alpha}{\beta}\,[1-\exp(\beta\,x)]\right\};\; x\ge 0;\; \alpha,\,\beta>0$$

$$S(x) = \exp\left\{\frac{\alpha}{\beta}\,[1-\exp(\beta\,x)]\right\}$$

$$h(x) = \alpha\,\exp(\beta\,x) \;\Rightarrow\; \text{IHR}$$

$$\mu(x) \;-\; \text{no closed form} \;\Rightarrow\; \text{DMRL}$$

GOMPERTZ–MAKEHAM distribution

This distribution has an extra parameter γ allowing for pure random failures whereas the common GOMPERTZ distribution only models failure by wear and tear.

$$\begin{aligned}
f(x) &= \gamma\left\{\frac{\alpha}{\beta}\,[1-\exp(\beta\,x)]-\gamma\,x\right\}+\alpha\,\exp(\beta\,x)\,\exp\left\{\frac{\alpha}{\beta}\,[1-\exp(\beta\,x)]-\gamma\,x\right\};\\
&\quad x\geq 0;\ \alpha,\,\beta,\,\gamma>0\\
S(x) &= \exp\left\{\frac{\alpha}{\beta}\,[1-\exp(\beta\,x)]-\gamma\,x\right\}\\
h(x) &= \gamma+\alpha\,\exp(\beta\,x)\ \Rightarrow\ \text{IHR}\\
\mu(x) &- \text{no closed form}\ \Rightarrow\ \text{DMRL}
\end{aligned}$$

GUMBEL distribution

This is one of the extreme value distributions, namely, type I for the maximum. Because of the PDF formula it is also known as double exponential distribution.

$$\begin{aligned}
f(x) &= \frac{1}{b}\,\exp\left[-\frac{x-a}{b}-\exp\left(-\frac{x-a}{b}\right)\right];\ x\in\mathbb{R};\ a\in\mathbb{R},\,b>0\\
S(x) &= 1-\exp\left[-\exp\left(-\frac{x-a}{b}\right)\right]\\
h(x) &= \frac{\exp\left(-\dfrac{x-a}{b}\right)}{b\,\left\{\exp\left[\exp\left(-\dfrac{x-a}{b}\right)\right]-1\right\}}\ \Rightarrow\ \text{IHR with }\lim_{x\to\infty}h(x)=\frac{1}{b}\\
\mu(x) &- \text{no closed form}\ \Rightarrow\ \text{DMRL with }\lim_{x\to\infty}\mu(x)=b
\end{aligned}$$

Half–CAUCHY distribution

Left truncation of the CAUCHY distribution at its median $x=a$ gives the half–CAUCHY distribution.

$$\begin{aligned}
f(x) &= 2\left\{b\,\pi\left[1+\left(\frac{x-a}{b}\right)^2\right]\right\}^{-1};\ x\geq a;\ a\in\mathbb{R},\,b>0\\
S(x) &= 1-\frac{2}{\pi}\,\arctan\left(\frac{x-a}{b}\right)\\
h(x) &= \frac{2}{b\left[1+\left(\dfrac{x-a}{b}\right)^2\right]\left[\pi-2\,\arctan\left(\dfrac{x-a}{b}\right)\right]}\ \Rightarrow\ \text{IDHR}
\end{aligned}$$

$h(x)$ has a maximum at x^* being the solution of $b-\pi\,(x^*-a)+2\,(x^*-a)\,\arctan\left(\dfrac{x^*-a}{b}\right)=0$. $\mu(x)$ does not exist.

Half–logistic distribution

Left truncation of the logistic distribution at its mean (= mode = median) $x = a$ gives the half–logistic distribution.

$$\begin{aligned}
f(x) &= \frac{2\,\exp\left(\dfrac{x-a}{b}\right)}{b\left[1+\exp\left(\dfrac{x-a}{b}\right)\right]^2};\ x \ge a;\ a \in \mathbb{R},\ b > 0\\
S(x) &= \frac{2}{1+\exp\left(\dfrac{x-a}{b}\right)}\\
h(x) &= \frac{1}{b\left[1+\exp\left(\dfrac{x-a}{b}\right)\right]} \ \Rightarrow\ \text{IHR}\\
\mu(x) &- \ \text{no closed form} \ \Rightarrow\ \text{DMRL}
\end{aligned}$$

Half–normal distribution

Left truncation of the normal distribution at its mean (= mode = median) $x = a$ gives the half–normal distribution.

$$\begin{aligned}
f(x) &= \frac{1}{b}\sqrt{\frac{2}{\pi}}\,\exp\left[-\frac{(x-a)^2}{2\,b^2}\right];\ x \ge a;\ a \in \mathbb{R},\ b > 0\\
S(x) &= 2\,\Phi\left(\frac{a-x}{b}\right)\\
h(x) &= \frac{1}{b\sqrt{2\,\pi}}\,\frac{\exp\left[-\dfrac{(x-a)^2}{2\,b^2}\right]}{\Phi\left(\dfrac{a-x}{b}\right)} \ \Rightarrow\ \text{IHR}\\
\mu(x) &- \ \text{no closed form} \ \Rightarrow\ \text{DMRL}
\end{aligned}$$

HJORTH distribution

This distribution is capable to show different types of aging.

$$\begin{aligned}
f(x) &= \frac{\theta+\delta\,x\,(1+\beta\,x)}{(1+\beta\,x)^{\theta/\beta+1}}\,\exp\left(-\delta\,\frac{x^2}{2}\right);\ x \ge 0;\ \beta,\ \delta,\ \theta > 0\\
S(x) &= \frac{\exp\left(-\delta\,\dfrac{x^2}{2}\right)}{(1+\beta\,x)^{\theta/\beta}}
\end{aligned}$$

For $\beta = 0$ we have to take the limits of $f(x)$ and $S(x)$ leading to

$$\begin{aligned}
f(x) &= \exp\left(-\theta\,x-\frac{\delta}{2}\,x^2\right)\,(\theta+\delta\,x),\\
S(x) &= \exp\left(-\theta\,x-\frac{\delta}{2}\,x^2\right).
\end{aligned}$$

$$h(x) = \delta\,x + \frac{\theta}{1+\beta\,x} \;\Rightarrow\; \begin{cases} \text{IHR for } \theta = 0 \\ \text{DHR for } \delta = 0 \\ \text{DIHR for } 0 < \delta < \theta\,\beta \\ \text{IHR for } \delta \geq \theta\,\beta \\ \text{constant for } \beta = \delta = 0 \end{cases}$$

$$\mu(x) \;-\; \text{no closed form} \;\Rightarrow\; \begin{cases} \text{does not exist for } \delta = \theta = 0 \\ \text{DMRL for } \theta = 0 \\ \text{IMRL for } \delta = 0 \\ \text{IDMRL or DMRL for } 0 < \delta < \theta\,\beta \\ \text{DMRL for } \delta \geq \theta\,\beta \\ \text{constant for } \beta = \delta = 0 \end{cases}.$$

In case of DIHR the hazard rate has its minimum at $x^* = \big[\sqrt{\theta\,\beta/\delta} - 1\big]\big/\beta$ with $h(x^*) = 2\,\theta\sqrt{\delta/(\theta\,\beta)} - \delta/\beta$.

Hyperbolic secant distribution

For $a = 0$ and $b = 2/\pi$ this distribution shares many properties with the standardized (= reduced) normal distribution.

$$f(x) = \frac{1}{b\,\pi}\,\text{sech}\left(\frac{x-a}{b}\right);\; x \geq \mathbb{R} <,\; a \in \mathbb{R},\, b > 0$$

$$S(x) = 1 - \frac{2}{\pi}\arctan\left[\exp\left(\frac{x-a}{b}\right)\right]$$

$$h(x) = \frac{\text{sech}\left(\dfrac{x-a}{b}\right)}{b\left\{\pi - 2\,\arctan\left[\exp\left(\dfrac{x-a}{b}\right)\right]\right\}} \;\Rightarrow\; \text{IHR with } \lim_{x\to\infty} h(x) = \frac{1}{b}$$

$$\mu(x) \;-\; \text{no closed form} \;\Rightarrow\; \text{DMRL with } \lim_{x\to\infty}\mu(x) = b$$

Inverse GAUSSIAN distribution

This distribution is also known as WALD distribution.

$$f(x) = \sqrt{\frac{b}{2\,\pi\,x^3}}\,\exp\left[-\frac{b\,(x-a)^2}{2\,a^2\,x}\right];\; x > 0;\; a,\, b > 0$$

$$S(x) = \Phi\left[\sqrt{\frac{b}{x}}\,\left(1 - \frac{x}{a}\right)\right] - \exp\left(\frac{2\,b}{a}\right)\,\Phi\left[-\sqrt{\frac{b}{x}}\,\left(\frac{x}{a}+1\right)\right]$$

$$h(x) \;-\; \text{no closed form} \;\Rightarrow \text{IDHR}$$

$$\mu(x) \;-\; \text{no closed form} \;\Rightarrow\; \text{DIMRL}$$

Inverse RAYLEIGH distribution

$$f(x) = \frac{2\,b}{(x-a)^3}\,\exp\left[-\frac{b}{(x-a)^3}\right];\; x \geq a,\; a \in \mathbb{R},\; b > 0$$

$$S(x) = 1 - \exp\left[-\frac{b}{(x-a)^3}\right]$$

$$h(x) = \frac{2\,b}{(x-a)^3\left\{\exp\left[\frac{b}{(x-a)^2}\right]-1\right\}} \Rightarrow \text{IDHR}$$

$$\mu(x) = u\left[\exp\left(-b/u^2\right)-1\right] + \sqrt{b\,\pi}\,\text{erf}\left(\sqrt{b}/u\right) \text{ with } u = x-a; \Rightarrow \text{DIMRL or DMRL}$$

$h(x)$ has a maximum at x^* which is the solution of $2\,\exp\left[b/(x^*-a)^2\right]\left[2\,b - 3\,(x^*-a)^2\right] + 6\,(x^*-a)^2 = 0$.

Inverse WEIBULL distribution

This distribution is also known as extreme value distribution of type II for the maximum.

$$f(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{-c-1}\exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right];\; x \geq a;\; a \in \mathbb{R};\; b,\, c > 0$$

$$S(x) = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right]$$

$$h(x) = \frac{c\left(\frac{x-a}{b}\right)^{-c-1}\exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right]}{b\left\{1-\exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right]\right\}} \Rightarrow \text{IDHR}$$

$$\mu(x) - \begin{cases} \text{no closed form} \;\Rightarrow\; \text{DIMRL} \\ \text{does not exist for } \; 0 < c \leq 1 \end{cases}$$

$h(x)$ has a maximum at x^* which is the solution of $\left(\frac{b}{x^*-a}\right)^c \Big/ \left\{1-\exp\left[-\left(\frac{b}{x^*-a}\right)^c\right]\right\} = \frac{c+1}{c}$.

LAPLACE distribution

This distribution is also known as double, bilateral or two–tailed exponential distribution.

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

$$S(x) = 1 - \frac{1}{2}\left\{1+\text{sign}(x-a)\,\left[1-\exp\left(-\frac{|x-a|}{b}\right)\right]\right\}$$

$$h(x) = \frac{\exp\left(-\frac{|x-a|}{b}\right)}{b\left\langle 2-\left\{1+\text{sign}(x-a)\,\left[1-\exp\left(-\frac{|x-a|}{b}\right)\right]\right\}\right\rangle}$$

$$\mu(x) - \text{no closed form}$$

$h(x)$ is increasing over $(-\infty, a)$ and constant with $h(x) = 1/b$ over $[a, \infty)$. $\mu(x)$ is decreasing over (∞, a) and constant with $\mu(x) = b$ over $[a, \infty)$.

Linear hazard rate distribution

This distribution includes the exponential distribution for $\alpha \neq 0,\ \beta = 0$ and the RAYLEIGH distribution for $\alpha = 0,\ \beta \neq 0$. For $\alpha = \beta = 0$ we have a distribution with everlasting life.

$$\begin{aligned}
f(x) &= (\alpha + \beta\, x)\, \exp\!\left[-\alpha\, x - \frac{\beta}{2}\, x^2\right];\ x \geq 0;\ \alpha,\, \beta \geq 0 \\
S(x) &= \exp\!\left[-\alpha\, x - \frac{\beta}{2}\, x^2\right] \\
h(x) &= \alpha + \beta\, x\ \Rightarrow\ \text{IHR} \\
\mu(x) &- \text{no closed form}\ \Rightarrow \begin{cases} \text{constant with}\ 1/\alpha\ \text{ for }\ \alpha \neq 0,\ \beta = 0 \\ \text{DMRL with}\ \lim\limits_{x\to\infty} \mu(x) = 0\ \text{ for } \alpha = 0,\ \beta \neq 0 \end{cases}
\end{aligned}$$

Logistic distribution

The logistic distribution shares many properties with the normal distribution.

$$\begin{aligned}
f(x) &= \frac{1}{b}\, \frac{\exp\left(\dfrac{x-a}{b}\right)}{\left[1 + \exp\left(\dfrac{x-a}{b}\right)\right]^2};\ x \in \mathbb{R}; a \in \mathbb{R},\, b > 0 \\
S(x) &= \left[1 + \exp\left(\frac{x-a}{b}\right)\right]^{-1} \\
h(x) &= \frac{1}{b}\left[1 + \exp\left(-\frac{x-a}{b}\right)\right]^{-1}\ \Rightarrow\ \text{IHR with}\ \lim_{x\to\infty} h(x) = \frac{1}{b} \\
\mu(x) &- \text{no closed form}\ \Rightarrow\ \text{DMRL with}\ \lim_{x\to\infty} \mu(x) = b
\end{aligned}$$

Log–gamma distribution

We present the version of the log–gamma distribution given in JOHNSON/KOTZ/BALAKRISHNAN (1995, pp. 89ff.),[3] which — for $c \neq 1$ — is a generalization of the extreme value distribution of type I for the minimum (= log–WEIBULL distribution).

$$\begin{aligned}
f(x) &= \frac{1}{b\,\Gamma(c)}\, \exp\!\left[c\, \frac{x-a}{b} - \exp\left(\frac{x-a}{b}\right)\right];\ x \in \mathbb{R};\ a \in \mathbb{R},\ b,\ c > 0 \\
S(x) &= \frac{\gamma\left[c, \exp\left(\dfrac{x-a}{b}\right)\right]}{\Gamma(c)} \\
h(x) &- \text{no closed form}\ \Rightarrow\ \text{IHR} \\
\mu(x) &- \text{no closed form}\ \Rightarrow\ \text{DMRL}
\end{aligned}$$

[3] JOHNSON/KOTZ/ BALAKRISHNAN (1994, p. 383) define another log–gamma function which is the distribution of a variate X when $-\ln X$ has a gamma distribution with $a = 0$ and $b,\ c$. The PDF of this version is

$$f(x) = \frac{1}{b^c\,\Gamma(c)}\, \frac{(-\ln x)^{c-1}}{x^{1+1/b}};\ 0 < x < 1.$$

Log–LAPLACE distribution

If Y has a LAPLACE distribution, then $X = \exp(Y)$ is said to have a log–LAPLACE distribution. We look at the version with a scale parameter b and two shape parameters c and d. For $c = d$ the PDF is symmetric in the sense that the variate and its reciprocal have the same distribution.

$$
\begin{aligned}
f(x) &= \frac{1}{b}\,\frac{c\,d}{c+d} \times \left\{\begin{array}{ll} \left(\dfrac{x}{b}\right)^{c-1} & \text{for } 0 \le x \le b \\[2ex] \left(\dfrac{b}{x}\right)^{d+1} & \text{for } x \ge b \end{array}\right\};\ b,\ c,\ d > 0 \\
S(x) &= \begin{cases} 1 - \dfrac{d}{c+d}\left(\dfrac{x}{b}\right)^{c} & \text{for } 0 \le x \le b \\[2ex] \dfrac{c}{c+d}\left(\dfrac{b}{x}\right)^{d} & \text{for } x \ge b \end{cases} \\
h(x) &= \begin{cases} \dfrac{\dfrac{c\,d}{b}\left(\dfrac{x}{b}\right)^{c-1}}{c+d-d\left(\dfrac{x}{b}\right)^{c}} & \text{for } 0 \le x \le b \\[4ex] \dfrac{\dfrac{c\,d}{b}\left(\dfrac{b}{x}\right)^{d+1}}{c\left(\dfrac{b}{x}\right)^{d}} & \text{for } x \ge b \end{cases} \\
\mu(x) &- \text{complicated closed form}
\end{aligned}
$$

HR and MRL have rather different courses depending on the special c–d–value combination, e.g., for $c > 1$ and $d > 1$ we have IHR over $[0, b)$ and DHR over $[b, \infty)$ with DIMRL or IMRL over $[0, \infty)$. MRL does not exist for $d \le 1$.

Log–logistic distribution

In economics this distribution is known as FISK distribution where it describes the distribution of income.

$$
\begin{aligned}
f(x) &= \frac{\dfrac{c}{b}\left(\dfrac{x-a}{b}\right)^{c-1}}{\left[1 + \left(\dfrac{x-a}{b}\right)^{c}\right]^2};\ x \ge a;\ a \in \mathbb{R},\ b,\ c > 0 \\
S(x) &= \frac{1}{1 + \left(\dfrac{x-a}{b}\right)^{c}} \\
h(x) &= \frac{\dfrac{c}{b}\left(\dfrac{x-a}{b}\right)^{c-1}}{1 + \left(\dfrac{x-a}{b}\right)^{c}} \Rightarrow \begin{cases} \text{DHR for } 0 < c \le 1 \\ \text{IDHR for } c > 1 \end{cases} \\
\mu(x) &- \text{no closed form} \Rightarrow \begin{cases} \text{does not exist for } 0 < c \le 1 \\ \text{IMRL or DIMRL for } c > 1 \end{cases}
\end{aligned}
$$

In case of IDHR the hazard rate has its maximum at $x^* = a + b\,(c-1)^{1/c}$ with $h(x^*) = (1/b)\,(c-1)^{(c-1)/c}$.

Log–normal with lower threshold

The log–normal distribution with lower threshold is more popular than that with an upper threshold. A variate X is said to be log–normally distributed with lower threshold when there is a real number a such that $X = \ln(X - a)$ is normally distributed. In economics it is known as GALTON or GIBRAT distribution.[4] We have $\mathrm{E}[\ln(X-a)] = \alpha$ and $\mathrm{Var}[\ln(X-a)] = \beta^2$.

$$
\begin{aligned}
f(x) &= \frac{1}{\beta\,(x-a)\sqrt{2\,\pi}}\,\exp\left\{-\frac{[\ln(x-a)-\alpha]^2}{2\,\beta^2}\right\};\ x \geq a;\ a \in \mathbb{R},\ \alpha \in \mathbb{R},\ \beta > 0 \\
S(x) &= \Phi\left[-\frac{\ln(x-a)-\alpha}{\beta}\right] \\
h(x) &= \frac{\phi\left[-\dfrac{\ln(x-a)-\alpha}{\beta}\right]}{\beta\,(x-a)\,\Phi\left[-\dfrac{\ln(x-a)-\alpha}{\beta}\right]} \ \Rightarrow\ \text{IDHR} \\
\mu(x) &- \ \text{no closed form} \ \Rightarrow\ \text{DIMRL}
\end{aligned}
$$

$h(x)$ has its maximum at $x^* = a + \exp(\alpha + \beta\,u)$ where u is the solution of $\phi(u) = u\,[1 - \Phi(u)]\,(1+\beta)/2$.

Log–normal with upper threshold

This distribution is defined on $(-\infty, a]$, a being the upper threshold.

$$
\begin{aligned}
f(x) &= \frac{1}{\beta\,(a-x)\sqrt{2\,\pi}}\,\exp\left\{-\frac{[\ln(a-x)-\alpha]^2}{2\,\beta^2}\right\};\ x < a;\ a \in \mathbb{R},\ \alpha \in \mathbb{R},\ \beta > 0 \\
S(x) &= \Phi\left[-\frac{\ln(a-x)-\alpha}{\beta}\right] \\
h(x) &= \frac{\phi\left[-\dfrac{\ln(a-x)-\alpha}{\beta}\right]}{\beta\,(a-x)\,\Phi\left[-\dfrac{\ln(a-x)-\alpha}{\beta}\right]} \ \Rightarrow\ \text{IHR} \\
\mu(x) &- \ \text{no closed form} \ \Rightarrow\ \text{DMRL}
\end{aligned}
$$

Log–WEIBULL distribution

When Y is WEIBULL distribute then $X = \ln X$ has a log–WEIBULL distribution, also known as extreme value distribution of type I for the minimum.

$$
\begin{aligned}
f(x) &= \frac{1}{b}\,\exp\left[\frac{x-a}{b} - \exp\left(\frac{x-a}{b}\right)\right];\ x \in \mathbb{R};\ a \in \mathbb{R},\ b > 0 \\
S(x) &= \exp\left[-\exp\left(\frac{x-a}{b}\right)\right] \\
h(x) &= \frac{1}{b}\,\exp\left(\frac{x-a}{b}\right) \ \Rightarrow\ \text{IHR} \\
\mu(x) &- \ \text{no closed form} \ \Rightarrow\ \text{DMRL}
\end{aligned}
$$

[4] The hazard rate of this log–normal distribution is discussed in detail by SWEET (1990).

LOMAX distribution

This distribution is also known as PARETO distribution of the second kind.

$$\begin{aligned}
f(x) &= \frac{c}{b}\left(1+\frac{x-a}{b}\right)^{-c-1}\ ;\ x \ge a;\ a \in \mathbb{R};\ b,\ c>0 \\
S(x) &= \left(1+\frac{x-a}{b}\right)^{-c} \\
h(x) &= \frac{c}{x-a-b}\ \Rightarrow\ \text{DHR} \\
\mu(x) &= \begin{cases} \text{does not exist for } 0<c\le 1 \\ \dfrac{x-a-b}{c-1} \ \text{ for } c>1\ \Rightarrow\ \text{IMRL} \end{cases}
\end{aligned}$$

MAXWELL–BOLTZMANN distribution

This distribution is nothing but the χ–distribution with 3 degrees of freedom, see also the χ–distribution above.

$$\begin{aligned}
f(x) &= \frac{1}{b}\sqrt{\frac{2}{\pi}}\left(\frac{x-a}{b}\right)^2 \exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right];\ x\ge a;\ a\in\mathbb{R},\ b>0 \\
S(x) &= 1-F_{\chi_3^2}\left[\left(\frac{x-a}{b}\right)^2\right];\ F_{\chi_3^2}(.) - \text{CDF of the } \chi^2\text{–distribution with } \nu=3 \\
h(x) &- \text{no closed form}\ \Rightarrow\ \text{IHR} \\
\mu(x) &- \text{no closed form}\ \Rightarrow\ \text{DMRL}
\end{aligned}$$

MUTH distribution

$$\begin{aligned}
f(x) &= \frac{1}{b}\left[\exp\left(c\,\frac{x-a}{b}\right)-c\right]\exp\left[-\frac{1}{c}\exp\left(c\,\frac{x-a}{b}\right)+c\,\frac{x-a}{b}+\frac{1}{c}\right]; \\
&\quad x\ge a;\ a\in\mathbb{R},\ b>0,\ 0<c\le 1 \\
S(x) &= \exp\left[-\frac{1}{c}\exp\left(c\,\frac{x-a}{b}\right)+c\,\frac{x-a}{b}+\frac{1}{c}\right] \\
h(x) &= \frac{1}{b}\left[\exp\left(c\,\frac{x-a}{b}\right)-c\right]\ \Rightarrow\ \text{IHR with } h(a)=\frac{1-c}{b}\ \forall\, c \\
\mu(x) &- \text{no closed form}\ \Rightarrow\ \text{DMRL}
\end{aligned}$$

Normal distribution

The normal or GAUSS distribution is of utmost importance in statistics. We have parameterized this distribution by a which is the mean $\mu = \mathrm{E}(X)$ and by b which is the standard deviation $\sigma = \sqrt{\mathrm{Var}(X)}$.

$$\begin{aligned}
f(x) &= \frac{1}{b\sqrt{2\,\pi}}\exp\left[-\frac{(x-a)^2}{2\,b^2}\right] = \frac{1}{b}\,\phi\left(\frac{x-a}{b}\right);\ x\in\mathbb{R},\ a\in\mathbb{R},\ b>0 \\
S(x) &= 1-\Phi\left(\frac{x-a}{b}\right) = \Phi\left(\frac{a-x}{b}\right) \\
h(x) &= \frac{1}{b}\,\frac{\phi\left(\dfrac{x-a}{b}\right)}{\Phi\left(\dfrac{a-x}{b}\right)}\ \Rightarrow\ \text{IHR} \\
\mu(x) &= a+b^2\,h(x)-x\ \Rightarrow\ \text{DMRL}
\end{aligned}$$

Parabolic U–shaped distribution

$$
\begin{aligned}
f(x) &= \frac{3}{2\,b}\left(\frac{x-a}{b}\right)^2;\ a-b \le x \le a+b;\ a \in \mathbb{R},\ b>0 \\
S(x) &= \frac{1}{2}\left[1-\left(\frac{x-a}{b}\right)^3\right] \\
h(x) &= \frac{3\,(a-x)^2}{b^3+(a-x)^3} \Rightarrow \text{ DIHR with minimum at } x^* = a \text{ and } h(x^*)=0 \\
\mu(x) &= \frac{0.75\,b^4 - b^3\,x + 0.25\,x^4}{b^3 - x^3} \Rightarrow \text{ IDMRL with maximum at } x^* \approx a - 0.596072\,b
\end{aligned}
$$

Parabolic inverted U–shaped distribution

$$
\begin{aligned}
f(x) &= \frac{3}{4\,b}\left[1-\left(\frac{x-a}{b}\right)^2\right];\ a-b \le x \le a+b;\ a \in \mathbb{R},\ b>0 \\
S(x) &= \frac{1}{2}-\frac{1}{4}\left[3\,\frac{x-a}{b}-\left(\frac{x-a}{b}\right)^3\right] \\
h(x) &- \text{ complicated form } \Rightarrow \text{ IHR} \\
\mu(x) &= b\,\frac{0.75-2\left(\frac{x-a}{b}\right)^2-0.25\left(\frac{x-a}{b}\right)^4}{\left[2+\left(\frac{x-a}{b}\right)\right]\left[\left(\frac{x-a}{b}\right)-1\right]^2} \Rightarrow \text{ DMRL}
\end{aligned}
$$

PARETO distribution of the first kind

In economics this distribution serves as model for the distribution of income.

$$
\begin{aligned}
f(x) &= \frac{c}{b}\left(\frac{x-a}{b}\right)^{-c-1};\ x \ge a+b;\ a \in \mathbb{R};\ b,\, c>0 \\
S(x) &= \left(\frac{x-a}{b}\right)^{-c} \\
h(x) &= \frac{c}{x-a} \Rightarrow \text{ DHR} \\
\mu(x) &= \begin{cases} \text{does not exist for } c \le 1 \\ \dfrac{b}{c-1}\left(\dfrac{x-a}{b}\right) \Rightarrow \text{ IMRL} \end{cases}
\end{aligned}
$$

Power function distribution

This distribution gives the uniform distribution for $c = 1$ and the right–angled negatively skew triangular distribution for $c = 2$ as special cases.

$$
\begin{aligned}
f(x) &= \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1};\ a \le x \le a+b;\ a \in \mathbb{R};\ b,\, c>0 \\
S(x) &= 1-\left(\frac{x-a}{b}\right)^c
\end{aligned}
$$

$$h(x) = \frac{c}{a-x}\left[1 + \frac{1}{\left(\frac{x-a}{b}\right)^c - 1}\right] \Rightarrow \begin{cases} \text{DIHR for } 0 < c < 1 \\ \text{IHR for } c \geq 1 \end{cases}$$

$$\mu(x) = \frac{\dfrac{c\,(x-a-b)}{\left(\frac{x-a}{b}\right)^c - 1} - (x-a)}{c+1} \Rightarrow \begin{cases} \text{IDMRL for } 0 < c \lesssim 0.815 \\ \text{DMRL for } c \gtrsim 0.815 \end{cases}$$

The DIHR hazard rate its minimum at $x^* = a + b\,(1-c)^{1/c}$ with $h(x^*) = (1/b)\,(1-c)^{(c-1)/c}$.

RAYLEIGH distribution

This distribution is a special case of the χ–distribution when $\nu = 2$, a special case of the WEIBULL distribution when $c = 2$ and a special case of the generalized gamma distribution when $c = 1$ and $d = 2$. It is also a linear hazard rate distribution.

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

$$S(x) = \exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right]$$

$$h(x) = \frac{x-a}{b} \Rightarrow \text{IHR}$$

$$\mu(x) - \text{no closed form} \Rightarrow \text{DMRL}$$

Reflected exponential distribution

Upon reflecting the exponential distribution with $f(x) = (1/b)\,\exp[-(x-a)/b]$ around $x = a$ we arrive at the reflected exponential distribution whereby the lower threshold turns into an upper threshold and the constant hazard rate property gets lost.

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

$$S(x) = 1 - \exp\left(\frac{x-a}{b}\right)$$

$$h(x) = \frac{1}{b\left[\exp\left(\frac{a-x}{b}\right) - 1\right]} \Rightarrow \text{IHR}$$

$$\mu(x) - \text{no closed form} \Rightarrow \text{DMRL}$$

Reflected WEIBULL distribution

This distribution is also known as extreme value distribution of type III for the maximum and results from the WEIBULL distribution by reflection around $x = a$ whereby a turns into an upper threshold. For $c = 1$ the reflected WEIBULL distribution is equal to the reflected exponential distribution.

$$
\begin{aligned}
f(x) &= \frac{c}{b}\left(\frac{a-x}{b}\right)^{c-1} \exp\left[-\left(\frac{a-x}{b}\right)^{c}\right];\ x \le a;\ a \in \mathbb{R},\, b > 0 \\
S(x) &= 1 - \exp\left[-\left(\frac{a-x}{b}\right)^{c}\right] \\
h(x) &= \frac{c\left(\dfrac{a-x}{b}\right)^{c-1}}{b\left\{\exp\left[\left(\dfrac{x-a}{b}\right)^{c}\right] - 1\right\}} \Rightarrow \text{IHR} \\
\mu(x) &- \text{no closed form} \Rightarrow \text{DMRL}
\end{aligned}
$$

Semi–elliptical distribution

This distribution is also known as WIGNER's semi–circle distribution. For $b = \sqrt{2/\pi} \approx 0.7979$ the graph of $f(x)$ is a semi–circle, otherwise a semi–ellipse.

$$
\begin{aligned}
f(x) &= \frac{2}{b\,\pi}\sqrt{1-\left(\frac{x-a}{b}\right)^{2}};\ a-b \le x \le a+b;\ a \in \mathbb{R},\, b > 0 \\
S(x) &= \frac{1}{2} - \frac{1}{\pi}\left[\frac{x-a}{b}\sqrt{1-\left(\frac{x-a}{b}\right)^{2}} + \arcsin\left(\frac{x-a}{b}\right)\right] \\
h(x) &= \frac{4\sqrt{1-\left(\dfrac{x-a}{b}\right)^{2}}}{b\left\{\pi - 2\left[\dfrac{x-a}{b}\sqrt{1-\left(\dfrac{x-a}{b}\right)^{2}} + \arcsin\left(\dfrac{x-a}{b}\right)\right]\right\}} \Rightarrow \text{IHR} \\
\mu(x) &- \text{no closed form} \Rightarrow \text{DMRL}
\end{aligned}
$$

t distribution

This distribution is also known as STUDENT's distribution, the pseudonym of W. S. GOSSET, its discoverer.

$$
f(x) = \frac{\Gamma\left(\dfrac{\nu+1}{\nu}\right)}{\sqrt{\pi\,\nu}\,\Gamma\left(\dfrac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2};\ x \in \mathbb{R};\ \nu > 0
$$

There are no closed formulas for $S(x)$, $h(x)$ and $\mu(x)$. The latter does not exist for $\nu \le 1$. for smaller ν we have IDHR and DIMRL. With $\nu \to \infty$ the t distribution goes to a normal distribution, so then we have IHR and DMRL.

TEISSIER distribution

This distribution, suggested by TEISSIER (1934), is characterized by an exponentially declining mean residual life function.

$$
\begin{aligned}
f(x) &= \frac{1}{b}\left[\exp\left(\frac{x-a}{b}\right)-1\right]\exp\left[1+\frac{x-a}{b}-\exp\left(\frac{x-a}{b}\right)\right];\ x \geq a;\ a \in \mathbb{R},\, b>0 \\
S(x) &= \exp\left[1+\frac{x-a}{b}-\exp\left(\frac{x-a}{b}\right)\right] \\
h(x) &= \frac{1}{b}\left[\exp\left(\frac{x-a}{b}\right)-1\right] \Rightarrow \text{IHR} \\
\mu(x) &= \frac{1}{b}\exp\left(-\frac{x-a}{b}\right) \Rightarrow \text{DMRL}
\end{aligned}
$$

Triangular distribution, continuous

The triangular distribution includes the following special cases:

- $c = 0.5 \Rightarrow$ symmetric triangular distribution,
- $c = 0 \Rightarrow$ right–angled and positively skewed triangular distribution,
- $c = 1 \Rightarrow$ right–angled and negatively skewed triangular distribution.

A triangular distribution results from folding (= summing) two independently and uniformly distributed variates. The triangle is symmetric when both uniform distributions are identical, otherwise it is asymmetric.

$$
\begin{aligned}
f(x) &= \left\{\begin{array}{ll} \dfrac{2\,(x-a)}{c\,b^2} & \text{for } a \leq x \leq a+c\,b \\ \dfrac{2\,(a+b-x)}{(1-c)\,b^2} & \text{for } a+c\,b \leq x \leq a+b \end{array}\right\} a \in \mathbb{R},\ b>0,\ 0<c<1 \\
S(x) &= \left\{\begin{array}{ll} 1-\dfrac{(x-a)^2}{c\,b^2} & \text{for } a \leq x \leq a+c\,b \\ \dfrac{(a+b-x)^2}{(1-c)\,b^2} & \text{for } a+c\,b \leq x \leq a+b \end{array}\right. \\
h(x) &= \left\{\begin{array}{ll} \dfrac{2\,(x-a)}{c\,b^2-(x-a)^2} & \text{for } a \leq x \leq a+c\,b \\ \dfrac{2}{a+b-x} & \text{for } a+c\,b \leq x \leq a+b \end{array}\right\} \Rightarrow \text{IHR} \\
\mu(y) &= \left\{\begin{array}{ll} b\,\dfrac{c+c^2-3\,c\,y+y^3}{3\,(c-y^2)} & \text{for } 0 \leq y \leq c \\ b\,\dfrac{1-y}{3} & \text{for } c \leq y \leq 1 \end{array}\right\} \text{with } y=\frac{x-a}{b} \Rightarrow \text{DMRL}
\end{aligned}
$$

Uniform distribution, continuous

The uniform or rectangular distribution is the simplest continuous distribution having a constant PDF.

$$
\begin{aligned}
f(x) &= \frac{1}{b};\ a \leq x \leq a+b;\ a \in \mathbb{R},\, b>0 \\
S(x) &= 1-\frac{x-a}{b} \\
h)x) &= \frac{1}{a+b-x} \Rightarrow \text{IHR} \\
\mu(x) &= \frac{b-x-a}{2} \Rightarrow \text{DMRL}
\end{aligned}
$$

V–shaped distribution

We present the symmetric V–shaped distribution which may be regarded as a linear approximation to the U–shaped parabolic distribution.

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

$$S(x) = \left\{ \begin{array}{ll} 1-\dfrac{2\,(x-a)\,(a+b-x)}{b^2} & \text{for } a \le x \le a+b/2 \\ 0.5-\dfrac{(2\,x-2\,a-b)^2}{b^2} & \text{for } a+b/2 \le x \le a+b \end{array} \right.$$

$$h(x) = \left\{ \begin{array}{ll} \dfrac{2\,(2\,a+b-2\,x)}{b^2+2\,(a-x)\,(a+b-x)} & \text{for } a \le x \le a+b/2 \\ \dfrac{2\,a+b-2\,x}{(a-x)\,(a+b-x)} & \text{for } a+b/2 \le x \le a+b \end{array} \right\} \Rightarrow \text{DIHR}$$

$$\mu(y) = \left\{ \begin{array}{ll} b\,\dfrac{3-2\,y\,[3+y\,(2\,y-3)]}{6\,[1+2\,y\,(y-1)]} & \text{for } 0 \le y \le 0.5 \\ b\,\dfrac{(y-1)^2\,(y+0.5)}{3\,y\,(1-y)} & \text{for } 0.5 \le y \le 1 \end{array} \right\} \text{with } y=\frac{x-a}{b} \Rightarrow \text{DMRL}$$

The hazard rate has its minimum at $x^* = a + b/2$ with $h(x^*) = 0$.

WEIBULL distribution

This distribution is also known as extreme value distribution of type III for the minimum.

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

$$S(x) = \exp\left[-\left(\frac{x-a}{b}\right)^{c}\right]$$

$$h(x) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \Rightarrow \left\{ \begin{array}{l} \text{DHR for } 0<c\le 1 \\ \text{IHR for } c \ge 1 \end{array} \right.$$

$$\mu(y) = \frac{b}{c}\,\exp(y^c)\,\gamma\!\left(\frac{1}{c},y^c\right) \Rightarrow \left\{ \begin{array}{l} \text{IMRL for } 0<c\le 1 \\ \text{DMRL for } c \ge 1 \end{array} \right\} \text{with } y=\frac{x-a}{b}$$

We conclude this section on continuous distributions by showing a typical output of the programm *ContDist*. After choosing one of the 62 distributions implemented in that program — here the DHILLON–II distribution — the program shows — as a reminder — the PDF–formula of this distribution. Then the user is asked to input a value for each of the pertaining parameters. The program checks these values for admissibility. The chosen parameter values are displayed together with the graphs of $f(x),\ S(x),\ h(x)$ and $\mu(x)$.

Figure 3/1: PDF-formula display of the DHILLON–II distribution by the program *ContDist*

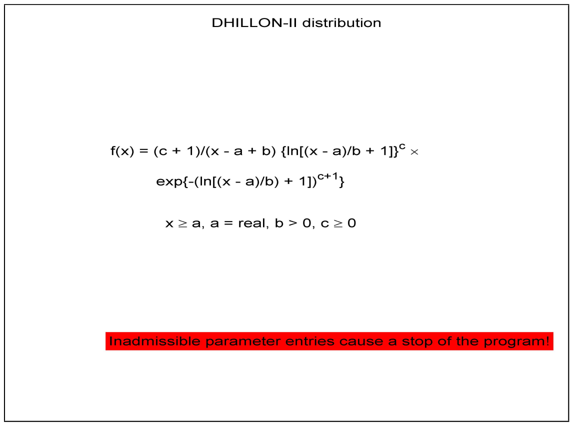
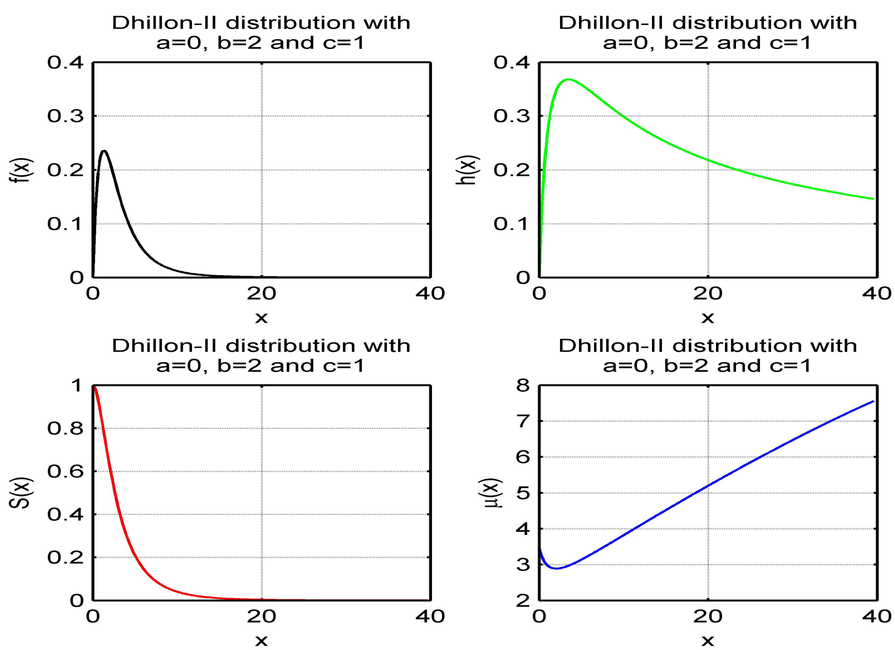


Figure 3/2: Display of the functions of a DHILLON–II distribution by the program *ContDist*



3.2 Discrete Distributions[5]

The most basic representative of a discrete distribution is its PMF

$$P_i = \Pr(X = i);\ i = 0,\ 1,\ 2 \ldots \text{ or } i = 1,\ 2,\ 3,\ \ldots$$

which can always be given in explicit and closed form. The survival function

$$S_i = \Pr(X \geq i) = \sum_{j \geq i} P_j$$

[5] Suggested reading for this section: LAI (2013), JOHNSON/KOTZ/KEMP (1992), PADGETT/SPURRIER (1985), RINNE (2009), SALVIA/BOLLINGER (1982), SHAKED et al. (1995), XEKALAKI (1983a, b). The interactive program *DiscDist*, which is written in MATLAB and which is included in the accompanying file 'Distributions.zip', displays for all distributions presented here a graph of the functions $P_i,\ S_i,\ h_i$ and — if existing — L_i for any set of parameter values.

seldom exists in closed form and must be found numerically by summing the P_i's. Consequently, the hazard rate

$$h_i = \frac{P_i}{S_i}$$

mostly has no closed form, too. This statement also holds with respect to the mean residual life function

$$L_i = \mathrm{E}(X - i | X \geq i)$$

which may be calculated as

$$L_i = \frac{\sum\limits_{j>i} S_j}{S_i} = \frac{\sum\limits_{j\geq i} P_j}{S_i} - i$$

when the support is finite. For an infinite support $i = i_{\min},\ i_{\min}{+}1,\ ...,\infty$ we avoid the evaluation of the sum $\sum\limits_{j=i+1}^{\infty} S_j$ with an infinite number of summands by remembering that — for existing $\mathrm{E}(X)$ — we have $L_{i_{\min}} = \sum\limits_{j>i_{\min}} S_j = \mathrm{E}(X) - i_{\min}$ as $S_{i_{\min}} = 1$. Thus, we can evaluate L_i for $i > i_{\min}$ as

$$L_i = \left\{ \Big[\mathrm{E}(X) - i_{\min}\Big] - \sum_{j=i_{\min}+1}^{i} S_j \right\} \Big/ S_i;\ i = i_{\min}+1,\ i_{\min}+2,\ ...$$

Together with P_i and — if existing in closed form — $S_i,\ h_i,\ L_i$ we will give the ratio

$$q_i = \frac{P_{i+1}}{P_i}$$

which comes up in the recursion formula $P_{i+1} = q_i\, P_i$. The quantity

$$\Delta\eta_i = \frac{P_{i+1}}{P_i} - \frac{P_{i+2}}{P_{i+1}} = q_i - q_{i+1},$$

which has been defined in (2/10b), will be given, too, as it serves in identifying the type of monotonicity of the hazard rate, i.e.:

$$\Delta\eta_i < 0 \ \Rightarrow\ \text{DHR},\ \ \Delta\eta_i > 0 \ \Rightarrow\ \text{IHR}.$$

Many univariate discrete distributions can be explained by the so–called **urn model**. An urn (= population) contains either a finite number N or an infinite number of balls from which either a finite number M or a fraction P is red. The color 'red' stands for any attribute. Sampling from this urn may be either with or without replacement of each ball drawn before drawing the next ball. Then the PMF gives the probability of having a certain number of red balls in the sample or of the number of balls to be drawn until the first, the second or so on red ball is found in the sample.

There is a caveat for the numerical evaluation of discrete distributions as severe rounding errors may distort the result for extreme values of the parameters or of the variable of the distribution.

Binomial distribution

The binomial PMF gives the probability of having i red balls in a sample of size n, drawn with replacement from an urn with fraction P of red balls.

$$\begin{aligned}
P_i &= \binom{n}{i} P^i (1-P)^{n-i};\ i = 0,\ 1,\ ...,\ n;\ n \geq 1,\ 0 < P < 1 \\
q_i &= \frac{n-i}{i+1}\,\frac{P}{1-P} \\
\Delta\eta_i &= \frac{n+1}{(i+1)\,(i+2)} > 0\,\forall\, i \ \Rightarrow\ \text{IHR} \\
L_i &- \ \text{no closed form} \ \Rightarrow\ \text{DMRL}
\end{aligned}$$

Binomial distribution, positive

The positive binomial distribution is a binomial distribution truncated on the left–hand side at $i = 0$.

$$\begin{aligned}
P_i &= \binom{n}{i} \frac{P^i(1-P)^{n-i}}{1-(1-P)^n};\ i = 1,\ 2,\ ...,\ n;\ n \geq 1,\ 0 < P < 1 \\
q_i,\ \Delta\eta_i &- \ \text{as for the ordinary binomial distribution above}
\end{aligned}$$

We have IHR and DMRL. For $i = 1,\ 2,\ ...,\ n$ the hazard rates of the common and the positive binomial distributions are identical.

Geometric distribution

The geometric distribution is a 'waiting time distribution' in the sense that its PMF gives the probability of drawing $i + 1$ balls (with replacement from a population having fraction P of red balls) until the first red ball is drawn, i.e., i is the *excess number* of sampled balls until the happening occurs. It is the discrete analogue to the exponential distribution, and it is a special case (with $m = 1$) of the negative binomial distribution.

$$\begin{aligned}
P_i &= P\,(1-P)^i;\ i = 0,\ 1,\ 2,\ ...;\ 0 < P < 1 \\
S_i &= (1-P)^i \\
q_i &= (1-P) \\
\Delta\eta_i &= 0 \\
h_i &= P \ \Rightarrow\ \text{IHR and DHR} \\
L_i &= \frac{1-P}{P} \ \Rightarrow\ \text{IMRL and DMRL}
\end{aligned}$$

Geometric distribution, positive

The positive geometric distribution results from truncating the ordinary geometric distribution on the left–hand side at $i = 0$. The resulting distribution gives the probability of the *total number* of balls to be drawn until the first red ball. It is nothing but a shifted ordinary geometric distribution, shifted one step to the right.

$$\begin{aligned}
P_i &= \frac{P\,(1-P)^i}{1-P} = P\,(1-P)^{i-1};\ i = 1,\ 2,\ 3,\ ...;\ 0 < P < 1 \\
S_i &= (1-P)^{i-1}
\end{aligned}$$

$$\begin{aligned}
q_i &= 1 - P \\
\Delta\eta_i &= 0 \\
h_i &= P \;\Rightarrow\; \text{IHR and DHR} \\
L_i &= \frac{1}{P} \;\Rightarrow\; \text{IMRL and DMRL}
\end{aligned}$$

Geometric distribution, zero–inflated

This modification of the ordinary geometric distribution has two parameters:

1. λ which corresponds to $1 - P$ of the ordinary geometric distribution,
2. α which is responsible for the inflation of Pr$(X = 0)$.

$$\left.\begin{aligned}
P_0 &= 1 - \alpha\,\lambda \\
P_i &= \alpha\,(1-\lambda)\,\lambda^i;\; i = 1,\,2,\,3,\,...
\end{aligned}\right\} 0 < \lambda < 1,\; 0 < \alpha < 1$$

$$\begin{aligned}
S_0 &= 1 \\
S_i &= \alpha\,\lambda^i;\; i = 1,\,2,\,3,\,...
\end{aligned}$$

$$\left.\begin{aligned}
h_0 &= 1 - \alpha\,\lambda \\
h_i &= 1 - \lambda;\; i = 1,\,2,\,3,\,...
\end{aligned}\right\} \Rightarrow \text{DHR}$$

$$\left.\begin{aligned}
L_0 &= \frac{\alpha\,\lambda}{1-\lambda} \\
L_i &= \frac{\lambda}{1-\lambda};\; i = 1,\,2,\,3,\,...
\end{aligned}\right\} \Rightarrow \text{IMRL}$$

Hypergeometric distribution

The hypergeometric PMF gives the probability of having i red balls in a sample of size n, drawn without replacement from a population of size N containing M red balls.

$$P_i = \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}};\; \begin{cases} \max(0,\,M+n-N) \le i \le \min(n,\,M) \\ n,\,N,\,M \in \mathbb{N}^+;\; n < N,\; M < N \end{cases}$$

For $M = 0$ or $M = N$ we would have a degenerate distribution with $P_0 = 1$ or $P_n = 1$, respectively. We would also have a degenerate distribution for $n = N$ with $P_M = 1$.

$$\begin{aligned}
q_i &= \frac{(M-i)\,(n-i)}{(i+1)\,(N-M+i+1)} \\
\Delta\eta_i &> 0 \;\Rightarrow\; \text{IHR} \\
L_i &- \text{no closed form} \;\Rightarrow\; \text{DMRL}
\end{aligned}$$

Hypergeometric distribution, positive

The positive hypergeometric distribution results from the ordinary hypergeometric distribution by truncation on the left–hand side at $i = 0$. For the truncation to be possible we must have $n \leq N - M$.

$$P_i = \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n} - \binom{N-M}{n}}; \begin{cases} 1 \leq i \leq \min(n,\, M) \\ n,\, N,\, M \in \mathbb{N}^+;\; n < N,\; M < N,\, n \leq N - M \end{cases}$$

$q_i, \Delta\eta_i$ are as with the ordinary hypergeometric distribution. We have IHR and DMRL. For $i = 1,\, 2,\, ...,\, \min(n, M)$ the hazard rates of the ordinary and the positive hypergeometric distributions are identical.

Logarithmic distribution

This distribution, also known as **logarithmic series distribution**, is derived from the MACLAURIN series expansion

$$-\ln(1 - P) = P + \frac{P^2}{2} + \frac{P^3}{3} + ...$$

It is the limit as $m \to 0$ of the zero–truncated negative binomial distribution.

$$P_i = \frac{a\,P^i}{i};\; i = 1,\, 2,\, 3,\, ...;\; 0 < P < 1,\; a = -\frac{1}{\ln(1-P)}$$

$$\mathrm{E}(X) = a\,\frac{P}{1-P}$$

$$q_i = P\,\frac{i}{i+1}$$

$$\Delta\eta_i = -\frac{P}{(i+1)\,(i+2)} < 0 \;\Rightarrow\; \text{DHR}$$

$$L_i - \text{no closed form} \;\Rightarrow\; \text{IMRL}$$

Logarithmic distribution, right–truncated

This truncated logarithmic distribution results from omitting all realizations of the ordinary logarithmic distributions greater than $r,\ r \geq 2$. For $r = 1$ we would have a degenerate distribution with $P_1 = 1$ and $P_i = 0\;\forall\, i > 1$.

$$P_i = \frac{P^i}{i}\,\frac{1}{\sum\limits_{j=1}^{i}\frac{P^j}{j}};\; i = 1,\, 2,\, ...,\, r;\; 0 < P < 1$$

$$h_i \Rightarrow \begin{cases} \text{IHR for } P \text{ large and } r \text{ small} \\ \text{DIHR otherwise} \end{cases}$$

$$L_i \Rightarrow \begin{cases} \text{DMRL for } P \text{ large and } r \text{ small} \\ \text{IDMRL otherwise} \end{cases}$$

Matching distribution

There is a population of size $N \in \mathbb{N}^+$ and the entities of this population are numbered 1, 2, ..., N. In the classical matching model the entities are arranged in a random order. Let X be the number of entities for which their position in the random order is the same as the number assigned to them.

$$P_i = \frac{1}{i!} \sum_{j=0}^{N-i} \frac{(-1)^j}{j!};\ i = 0,\ 1,\ ...,\ N;\ N \in \mathbb{N}^+,$$

where $P_N = 1\big/N!$ and $P_{N-1} = 0$. h_i and L_i are not monotone.

Negative binomial distribution

We look at successive random trials, each having a constant probability P of success (= drawing of a red ball). The number of *extra trials* to perform in order to observe a given number m of successes has a negative binomial distribution. For integer m it is called **PASCAL distribution** and for $m = 1$ we have the geometric distribution.

$$P_i = \binom{m+i-1}{i} P^m (1-P)^i;\ i = 0,\ 1,\ 2,\ ...;\ 0 < P < 1;\ m = 1,\ 2,\ ...$$

More generally, m may be any positive real number. Then we write:

$$P_i = \frac{\Gamma(m+i)}{\Gamma(i+1)\,\Gamma(m)}\, P^m (1-P)^i;\ i = 0,\ 1,\ 2,\ ...;\ 0 < P < 1;\ m > 0.$$

As $\Gamma(z+1) = z!$ for integer z the latter version of the negative binomial distribution is more general than the first one.

$$\begin{aligned}
\mathrm{E}(X) &= m\,\frac{1-P}{P} \\
q_i &= (1-P)\,\frac{m+i}{1+1} \\
\Delta\eta_i &= (1-P)\,\frac{m-1}{(i+1)\,(1+2)} \begin{cases} < 0 \text{ for } 0 < m < 1 \Rightarrow \text{ DHR with IMRL} \\ = 0 \text{ for } m = 1 \Rightarrow h_i = P \text{ with } L_i = 1/P \\ > 0 \text{ for } m > 1 \Rightarrow \text{ IHR with DMRL} \end{cases}
\end{aligned}$$

Negative hypergeometric distribution

The model of the negative binomial distribution above is modified in the following sense:

1. The drawing of balls is without replacement from a finite population of size N having $M,\ 1 \le M < N$, red balls.
2. The variate is the *total number* of trials until the occurrence of the m–th success (= red ball).

$$\begin{aligned}
P_i &= \frac{\binom{i-1}{m-1}\binom{N-i}{M-m}}{\binom{N}{M}};\ m \le i \le N-M+m;\ m,\ N,\ M \in \mathbb{N}^+;\ M < N,\ m \le M \\
\mathrm{E}(X) &= m\,\frac{N+1}{M+1} \\
q_i &= \frac{i\,(N-M+m-i)}{(i-m+1)\,(N-i)} \\
\Delta\eta_i &> 0 \Rightarrow \text{ IHR} \\
L_i &- \text{ no closed form } \Rightarrow \text{ DMRL}
\end{aligned}$$

Occupancy distribution

We have N distinct objects, e.g., balls, and m distinct boxes or cells. Now consider the placement of these objects into the m boxes. The number of ways to do this clearly m^N. Each of these ways is considered equiprobable. We are interested in the distribution of X, the number of empty boxes in a placement.

$$P_i = \binom{m}{i} \sum_{j=0}^{m-i} (-1)^j \binom{m-i}{j} \left(1 - \frac{i+j}{m}\right)^N ;\; i = 0,\, 1,\, ...,\, m-1;\; m,\, N \in \mathbb{N}^+$$

We have:

- IHR as $\Delta\eta_i > 0$ and
- DMRL.

POISSON distribution

This distribution is for the number of occurrences of an event in an interval of given length when the intensity of event–occurrence in this interval is λ.

$$\begin{aligned}
P_i &= \frac{\lambda^i}{i!}\exp(-\lambda);\; i = 0,\, 1,\, 2,\, ...;\; \lambda > 0 \\
\mathrm{E}(X) &= \lambda \\
q_i &= \frac{\lambda}{i+1} \\
\Delta\eta_i &= \frac{\lambda}{(i+1)\,(i+2)} > 0 \;\Rightarrow\; \text{IHR} \\
L_i &- \text{no closed form} \;\Rightarrow\; \text{DMRL}
\end{aligned}$$

POISSON distribution, positive

This distribution results from the ordinary POISSON distribution by truncation of the realization $i = 0$.

$$\begin{aligned}
P_i &= \frac{\lambda^i}{i!\,(e^\lambda - 1)}\exp(-\lambda);\; i = 1,\, 2,\, 3,\, ...;\; \lambda > 0 \\
\mathrm{E}(X) &= \frac{\lambda}{1 - e^{-\lambda}} \\
q_i &= \frac{\lambda}{i+1} \\
\Delta\eta_i &= \frac{\lambda}{(i+1)\,(i+2)} > 0 \;\Rightarrow\; \text{IHR} \\
L_i &- \text{no closed form} \;\Rightarrow\; \text{DMRL}
\end{aligned}$$

PÓLYA distribution

This distribution gives the probabilities of the number of successes (= red balls) in a rather general urn model. An urn contains N balls, M being red. A sample of size n is to be drawn from this urn. After each ball drawn this ball is given back to the urn together with K balls of the color just drawn, K may be any integer ..., $-1,\ -1,\ 0,\ 1,\ 2,\ \dots$. $K<0$ means that a number $|K|$ of balls of the color just drawn is eliminated from the urn, but the ball just drawn is laid back anyway. Three special values of K lead to special distributions:

- $K=-1$ gives the hypergeometric distribution as effectively no ball is replaced.
- $K=0$ gives the binomial distribution with $P=M/N$.
- $K=N/2$, when $M=N/2$ is the discrete uniform distribution.

$$\begin{aligned}
P_i &= \binom{n}{i}\frac{\prod\limits_{j=1}^{i}[M+(j-1)\,K]\prod\limits_{j=1}^{n-i}[N-M+(j-1)\,K]}{\prod\limits_{j=1}^{n}[N+(j-1)\,K]};\\
&\quad i=0,\ 1,\ ...,\ n \text{ for } K\geq 0;\ \max\Big[0,n+\frac{N-M}{K}\Big]\leq i\leq \min\Big[n,-\frac{M}{K}\Big] \text{ for } K<0\\
&\quad n,\ N,\ M\in\mathbb{N}^+;\ M<N;\ K=...,\ -2,\ -1,\ 0,\ 1,\ 2,\ ...;\ N+K\,(n-1)>0\\
q_i &= \frac{n-i}{i+1}\,\frac{M+i\,K}{(i+1)\,[N-M+(n-i+1)\,K]}\\
\Delta\eta_i &> 0\ \Rightarrow\ \text{IHR}\\
L_i &- \text{no closed form}\ \Rightarrow\ \text{DMRL}
\end{aligned}$$

Runs distribution

There are N_0 balls labeled 0 and N_1 balls labeled 1, arranged in random order. Let r_{0j} be the number of runs of j consecutive 0's and r_{1j} that of j consecutive 1's. Then we have

$$\sum_j j\,r_{0j}=N_0 \text{ and } \sum_j j\,r_{1j}=N_1.$$

The distribution of the total number of runs of 1's, $X=R_1$, is

$$\Pr(X=i)=P_i=\frac{\binom{N_0}{i-1}\binom{N_1}{N_1-i}}{\binom{N_0+N_1}{N_1-1}};\ i=1,\ 2,\ ...,\ \min(N_0+1,N_1);\ N_0,\ N_1\in\mathbb{N}^+$$

$i=1$ means that all the 1's are standing together with no 0's in between.

$$\begin{aligned}
\mathrm{E}(X) &= \frac{N_1\,(N_0+1)}{N_0+N_1}\\
q_i &= \frac{(N_0-i+1)\,(N_1-i)}{i\,(i+1)}\\
\Delta\eta_i &> 0\ \Rightarrow\ \text{IHR}\\
L_i &- \text{no closed form}\ \Rightarrow\ \text{DMRL}
\end{aligned}$$

SALVIA–BOLLINGER's DHR distribution This and the following distribution of SALVIA/BOLLINGER (1982) have their origin in looking for a rather simple form for the hazard rate, namely some modification of a harmonic series.

$$\begin{aligned}
h_i &= \frac{c}{i+1};\ i = 0,\,1,\,2,\,...;\ 0 < c < 1\\
P_0 &= h_0 = c\\
P_i &= \frac{c}{i+1}\prod_{j=0}^{i-1}\frac{j+1-c}{j+1};\ i = 1,\,2,\,...\\
S_0 &= 1\\
S_i &= \prod_{j=0}^{i-1}\frac{j+1-c}{j+1};\ i = 1,\,2,\,...\\
q_i &= \frac{i+1-c}{i+2}\\
\Delta\eta_i &= -\frac{c+1}{(i+2)\,(i+3)} < 0 \ \Rightarrow\ \text{DHR with } \lim_{i\to\infty} h_i = 0\\
L_i &- \text{no closed form} \ \Rightarrow\ \text{IMRL}
\end{aligned}$$

SALVIA–BOLLINGER's IHR distribution

$$\begin{aligned}
h_i &= 1-\frac{c}{i+1};\ i = 0,\,1,\,2,\,...;\ 0 < c < 1\\
P_i &= \frac{(i+1-c)\,c^i}{(i+1)!}\\
S_i &= \frac{c^i}{i!}\\
q_i &= \frac{c\,(i+2-c)}{(i+2)\,(i+1-c)}\\
\Delta\eta_i &> 0 \ \Rightarrow\ \text{IHR with } \lim_{i\to\infty} h_i = 1\\
L_i &- \text{no closed form} \ \Rightarrow\ \text{DMRL with } \lim_{i\to\infty} L_i = 1
\end{aligned}$$

SALVIA–BOLLINGER's generalized DHR distribution

This and the following generalization of SALVIA–BOLLINGER's distributions by PADGETT/SPURRIER (1985) have an extra parameter α causing a faster or a slower decline of the hazard rate. With $\alpha = 1$ we have the original distribution of SALVIA and BOLLINGER and with $\alpha = 0$ we have the geometric distribution with constant hazard rate.

$$\begin{aligned}
h_i &= \frac{c}{\alpha\,i+1};\ i = 0,\,1,\,2,\,...;\ 0 < c < 1;\ \alpha \ge 0\\
P_0 &= h_0 \;=\; c\\
P_i &= \frac{c}{\alpha\,i+1}\prod_{j=0}^{i-1}\frac{\alpha\,j+1-c}{\alpha\,j+1};\ i = 1,\,2,\,3,\,...\\
S_0 &= 1\\
S_i &= \prod_{j=0}^{i-1}\frac{\alpha\,j+1-c}{\alpha\,j+1};\ i = 1,\,2,\,3,\,...
\end{aligned}$$

$$\begin{aligned}
q_i &= \frac{\alpha\, i + 1 - c}{\alpha\,(i+1)+1} \\
\Delta\eta_i &= -\frac{\alpha\, c - \alpha^2}{[\alpha\,(i+1)+1]\,[\alpha\,(i+2)+1]} < 0 \;\Rightarrow\; \text{DHR with } \lim_{i\to\infty} h_i = 0 \\
L_i &- \text{ no closed form } \;\Rightarrow\; \text{IMRL}
\end{aligned}$$

SALVIA–BOLLINGER's generalized IHR distribution

For $\alpha = 1$ we have the original IHR distribution of SALVIA/BOLLINGER and for $\alpha = 0$ we have a geometric distribution.

$$\begin{aligned}
h_i &= 1 - \frac{c}{\alpha\, i + 1};\; i = 0,\, 1,\, 2,\, ...;\; 0 < c < 1;\; \alpha \geq 0 \\
P_0 &= h_0 \;=\; 1 - c \\
P_i &= \left(1 - \frac{c}{\alpha\, i + 1}\right) \prod_{j=0}^{i-1} \frac{c}{\alpha\, j + 1};\; i = 1,\, 2,\, 3,\, ... \\
S_0 &= 1 \\
S_i &= \prod_{j=0}^{i-1} \frac{c}{\alpha\, j + 1};\; i = 1,\, 2,\, 3,\, ... \\
q_i &= \frac{[\alpha\,(i+1) - c]\,(\alpha\, i + 1)\, c}{[\alpha\,(i+1)+1]\,(\alpha\, i + 1 - c)\,(\alpha\, i + 1)} \\
\Delta\eta_i &> 0 \;\Rightarrow\; \text{IHR with } \lim_{i\to\infty} h_i = 1 \\
L_i &- \text{ no closed form } \;\Rightarrow\; \text{DMRL with } \lim_{i\to\infty} L_i = 1
\end{aligned}$$

Triangular distribution, right–angled and negatively skew

The triangular and the uniform distributions are simple forms of a discrete distribution with linear PMF.

$$\begin{aligned}
P_i &= \frac{2\, i}{m\,(m+1)};\; i = 1,\, 2,\, ...,\, m;\; m \geq 2 \\
S_i &= 1 - \frac{i\,(i-1)}{m\,(m+1)} \\
h_i &= \frac{2\, i}{m\,(m+1) - i\,(i-1)} \\
q_i &= \frac{i+1}{i} \\
\Delta\eta_i &= \frac{1}{i\,(i+1)} > 0 \;\Rightarrow\; \text{IHR} \\
L_i &= \frac{(m-i)\,(2\, m + i - 1)}{3\,(m+i)} \;\Rightarrow\; \text{DMRL}
\end{aligned}$$

Triangular distribution, right–angled and positively skew

$$\begin{aligned}
P_i &= \frac{2\,(m-i+1)}{m\,(m+1)};\; i = 1,\, 2,\, ...,\, m;\; m \geq 2 \\
S_i &= \frac{(i-m-2)\,(i-m-1)}{m\,(m+1)}
\end{aligned}$$

$$
\begin{aligned}
h_i &= \frac{2}{m-i+2)}\\
q_i &= \frac{m-i}{m-i+1}\\
\Delta\eta_i &= \frac{1}{(m-i)\,(m-i+1)} > 0 \;\Rightarrow\; \text{IHR}\\
L_i &= \frac{m-i}{3} \;\Rightarrow\; \text{DMRL}
\end{aligned}
$$

Triangular distribution, symmetric

We have to distinguish two cases depending on m, the length of the support of the distribution:

- $m = 2\,k,\; k = 1,\,2,\,...$
- $m = 2\,k + 1;\; k = 1,\,2,\,...$

We require $m > 3$. For $m = 2$ we will have the uniform distribution over two points $i = 1$ and $i = 2$.

Case 1: $m = 2\,k,\; k = 1,\,2,\,...$

$$
\begin{aligned}
P_i &= \begin{cases} \frac{i}{k\,(k+1)} & \text{for } i = 1,\,2,\,...,\,k\\ \frac{m-i+k}{k\,(k+1)} & \text{for } i = k+1,\,k+2,\,...,\,2\,k\;(=m) \end{cases}\\
S_i &= \begin{cases} 1+\frac{i\,(1-i)}{2\,k\,(k+1)} & \text{for } i = 1,\,2,\,...,\,k\\ \frac{(i-2\,k-1)\,[i-2\,(k+1)]}{2\,k\,(k+1)} & \text{for } i = k+1,\,k+2,\,...,\,2\,k\;(=m) \end{cases}\\
h_i &= \begin{cases} \frac{2\,i}{i\,(1-i)\,2\,k\,(k+1)} & \text{for } i = 1,\,2,\,...,\,k\\ \frac{2}{2-i+2\,k} & \text{for } i = k+1,\,k+2,\,...,\,2\,k\;(=m) \end{cases}\\
q_i &= \begin{cases} \frac{i+1}{i} & \text{for } i = 1,\,2,\,...,\,k\\ \frac{2\,k-i}{2\,k-i+1} & \text{for } i = k+1,\,k+2,\,...,\,2\,k-1 \end{cases}\\
\Delta\eta_i &> 0 \;\Rightarrow\; \text{IHR}\\
L_i &- \text{complicated form} \;\Rightarrow\; \text{DMRL}
\end{aligned}
$$

Case 1: $m = 2\,k+1,\; k = 1,\,2,\,...$

$$
\begin{aligned}
P_i &= \begin{cases} \frac{i}{(k+1)^2} & \text{for } i = 1,\,2,\,...,\,k+1\\ \frac{m-i+1}{(k+1)^2} & \text{for } i = k+2,\,k+3,\,...,\,2\,k+1\;(=m) \end{cases}\\
S_i &= \begin{cases} 1+\frac{i\,(1-i)}{2\,(k+1)^2} & \text{for } i = 1,\,2,\,...,\,k+1\\ \frac{(i-2\,k-3)\,[i-2\,(k+1)]}{2\,(k+1)^2} & \text{for } i = k+2,\,k+3,\,...,\,2\,k+1\;(=m) \end{cases}\\
h_i &= \begin{cases} \frac{2\,i}{i\,(1-i)\,2\,(k+1)^2} & \text{for } i = 1,\,2,\,...,\,k+1\\ \frac{2}{3-i+2\,k} & \text{for } i = k+2,\,k+3,\,...,\,2\,k+1\;(=m) \end{cases}\\
q_i &= \begin{cases} \frac{i+1}{i} & \text{for } i = 1,\,2,\,...,\,k+1\\ \frac{2\,k-i}{2\,k-i+1} & \text{for } i = k+2,\,k+3,\,...,\,2\,k \end{cases}\\
\Delta\eta_i &> 0 \;\Rightarrow\; \text{IHR}\\
L_i &- \text{complicated form} \;\Rightarrow\; \text{DMRL}
\end{aligned}
$$

Uniform distribution

$$\begin{aligned}
P_i &= \frac{1}{m};\ i = 1,\, 2,\, ...,\, m;\ m \in \mathbb{N}^+ \vee m \geq 2 \\
S_i &= 1 - \frac{i-1}{m} \\
h_i &= \frac{1}{m-i+1} \\
q_i &= 1 \\
\Delta\eta_i &= 0 \ \Rightarrow\ \text{IHR} \\
L_i &= \frac{m-i}{2} \ \Rightarrow\ \text{DMRL}
\end{aligned}$$

WEIBULL distribution of type I

This and the following two discrete WEIBULL distributions mimic the continuous WEIBULL distribution with respect to the hazard rate behavior which may be IHR, DHR or constant depending on the value of a certain parameter β. For $\beta = 1$ this discrete WEIBULL distribution is equal to the geometric distribution with $P = 1 - q$.

$$\begin{aligned}
P_i &= q^{i^\beta} - q^{i+1)\,\beta};\ i = 0,\, 1,\, 2,\, ...;\ 0 < q < 1,\ \beta > 0 \\
S_i &= q^{i^\beta} \\
h_i &= 1 - q^{(i+1)^\beta - i^\beta} \ \Rightarrow \begin{cases} \text{DHR for } \ 0 < \beta < 1 \\ \text{constant (IHR and DHR) for } \ \beta = 1 \ \text{ with } \ h_i = 1 - q \\ \text{IHR for } \ \beta > 1 \end{cases} \\
L_i &- \ \text{no closed form} \ \Rightarrow \begin{cases} \text{IDMRL for } \ 0 < \beta < 1 \\ \text{constant for } \beta = 1 \ \text{ with } \ L_i = \frac{q}{1-q} \\ \text{DIMRL for } \ \beta > 1 \end{cases}
\end{aligned}$$

WEIBULL distribution of type II

For $\beta = 1$ this discrete WEIBULL distribution is equal to the geometric distribution with $P = \alpha$.

$$\begin{aligned}
P_1 &= \alpha \\
P_i &= \alpha\, i^{\beta-1} \prod_{j=1}^{i-1} \Big(1 - \alpha\, j^{\beta-1}\Big);\ i = 2,\, 3,\, ...,\, m;\ 0 < \alpha < 1,\ \beta > 0 \\
& \qquad m = \begin{cases} \infty & \text{for } \ 0 < \beta \leq 1 \\ \text{int}\Big[\alpha^{-1/(\beta-1)}\Big] & \text{for } \ \beta > 1 \end{cases} \\
S_1 &= 1 \\
S_i &= \prod_{j=1}^{i-1} \Big(1 - \alpha\, j^{\beta-1}\Big);\ i = 2,\, 3,\, ...,\, m
\end{aligned}$$

$$h_i = \alpha\, i^{\beta-1} \;\Rightarrow\; \left\{\begin{array}{l}\text{DHR for } 0<\beta<1\\ \text{constant (IHR and DHR) for } \beta=1 \text{ with } h_i=\alpha\\ \text{IHR for } \beta>1\end{array}\right\};\; i=1,\,2,\,...,\,m$$

$$L_i \;-\; \text{no closed form} \;\Rightarrow\; \left\{\begin{array}{l}\text{IMRL for } 0<\beta<1\\ \text{constant for } \beta=1 \text{ with } L_i=\frac{1-\alpha}{\alpha}\\ \text{DMRL for } \beta>1\end{array}\right\};\; i=1,\,2,\,...,\,m$$

WEIBULL distribution of type III

For $\beta = 0$ this discrete WEIBULL distribution is equal to the geometric distribution with $P = 1 - \exp(-d)$.

$$
\begin{aligned}
P_0 &= 1-\exp(-d)\\
P_i &= \Big\{1-\exp\big[-d\,(i+1)^{\beta}\big]\Big\}\exp\left[-d\sum_{j=1}^{i} j^{\beta}\right];\; i=1,\,2,\,...;\; \beta\in\mathbb{R},\; d>0\\
S_0 &= 1\\
S_i &= \exp\left[-d\sum_{j=1}^{i} j^{\beta}\right];\; i=1,\,2,\,...\\
h_i &= 1-\exp\big[-d\,(i+1)^{\beta}\big]\\
&\Rightarrow \left\{\begin{array}{l}\text{DHR for } \beta<0\\ \text{constant (IHR and DHR) for } \beta=0 \text{ with } h_i=1-\exp(-d)\\ \text{IHR for } \beta>0\end{array}\right\};\; i=0,\,1,\,...
\end{aligned}
$$

$$L_i \;-\; \text{no closed form} \;\Rightarrow\; \left\{\begin{array}{l}\text{IMRL for } \beta<0\\ \text{constant for } \beta=0 \text{ with } L_i=\frac{1}{\exp(d)-1}\\ \text{DMRL for } \beta>0\end{array}\right\};\; i=0,\,1,\,...$$

YULE distribution

The YULE distribution plays a role in biostatistics where it is the distribution for the number of species of biological organisms per family, see also XEKALAKI (1983a, b).

$$
\begin{aligned}
P_i &= \rho\, B(i,\rho+1) = \rho\,\frac{\Gamma(i)\,\Gamma(\rho+1)}{\Gamma(i+\rho+1)};\; i=1,\,2,\,...;\; \rho>0\\
S_1 &= 1\\
S_i &= (i-1)\,B(i-1,\rho+1);\; i=2,\,3,\,...\\
\mathrm{E}(X) &= \frac{\rho}{\rho-1} \quad \text{for } \rho>1\\
h_i &= \frac{\rho}{i+\rho}
\end{aligned}
$$

$$
\begin{aligned}
q_i &= \frac{i}{1+\rho+i} \\
\Delta\eta_i &= \frac{-\rho-1}{(1+\rho+i)\,(2+\rho+i)} < 0 \;\Rightarrow\; \text{DHR} \\
L_i &- \text{no closed form} \;\Rightarrow \begin{cases} \text{does not exist for } \rho \le 1 \\ \text{IMRL (linear) for } \rho > 1 \end{cases}
\end{aligned}
$$

Zeta distribution of ZIPF

This distribution is also called discrete PARETO distribution. The name zeta is justified by the appearance of RIEMANN's zeta function in the formula of the PMF. This distribution plays a role in linguistics giving the probability of the number of words appearing i times in long sequences of text.

$$
\begin{aligned}
P_i &= \frac{i^{-(\theta+1)}}{\zeta(\theta+1)};\; i = 1,\, 2,\, ...;\; \theta < 0 \\
& \quad \zeta(\theta+1) = \sum_{i=1}^{\infty} i^{-(\theta+1)} - \text{RIEMANN's zeta function} \\
\mathrm{E}(X) &= \begin{cases} \dfrac{\zeta(\theta)}{\zeta(\theta+1)} & \text{for } \theta > 1 \\ \infty & \text{else} \end{cases} \\
h_i &- \text{no closed form} \\
q_i &= \left(\frac{i}{i+1}\right)^{\theta+1} \\
\Delta\eta_i &= \left(\frac{i}{i+1}\right)^{\theta+1} - \left(\frac{i+1}{i+2}\right)^{\theta+1} < 0 \;\Rightarrow\; \text{DHR} \\
L_i &- \text{no closed form} \;\Rightarrow \begin{cases} \text{does not exist for } \theta \le 1 \\ \text{IMRL for } \theta > 1 \end{cases}
\end{aligned}
$$

Zeta distribution of HAIGHT

The name of the distribution has its reason in the fact that the first two raw moments can be expressed in terms of RIEMANN's zeta function.

$$
\begin{aligned}
P_i &= \frac{1}{(2\,i-1)^\alpha} - \frac{1}{(2\,i+1)^\alpha};\; i = 1,\, 2,\, ...;\; \alpha > 0 \\
\mathrm{E}(X) &= \begin{cases} \text{does not exist for } \alpha \le 1 \\ \left(1 - 2^{-\alpha}\right)\zeta(\alpha) \text{ for } \alpha > 1 \end{cases} \\
h_i &- \text{no closed form} \\
q_i &= \frac{[2\,i+1]^{-\alpha} - [2\,i+3]^{-\alpha}}{[2\,i-1]^{-\alpha} - [2\,i+1]^{-\alpha}} \\
\Delta\eta_i &= -\frac{(2\,i-1)^\alpha \left[1 - 2\left(1 - \frac{2}{3+2\,i}\right)^\alpha + \left(1 - \frac{4}{5+2\,i}\right)^\alpha\right]}{(2\,i-1)^\alpha - (2\,i+1)^\alpha} < 0 \;\Rightarrow\; \text{DHR} \\
L_i &- \text{no closed form} \begin{cases} \text{does not exist for } \alpha \le 1 \\ \text{IMRL for } \alpha > 1 \end{cases}
\end{aligned}
$$

We conclude this section on discrete distributions by showing a typical output of the programm *DiscDist*. After choosing one of the 31 distributions implemented in that program — here the discrete WEIBULL distribution of type I — the program shows — as a reminder — the PMF–formula of this distribution. Then the user is asked to input a value for each of the pertaining parameters. The program checks these values for admissibility. The chosen parameter values are displayed together with the graphs of P_i, S_i, h_i and L_i.

Figure 3/3: PMF-formula display of the WEIBULL type I distribution by the program *DiscDist*

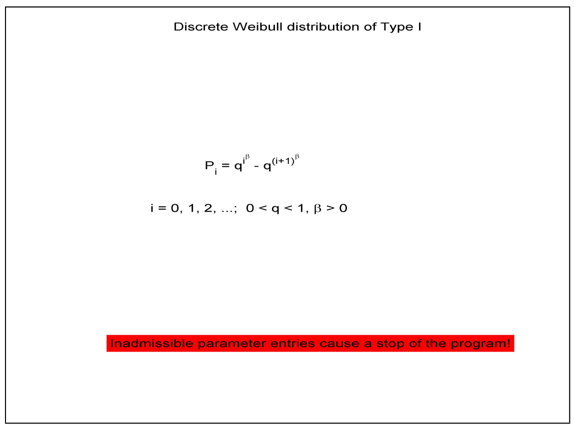
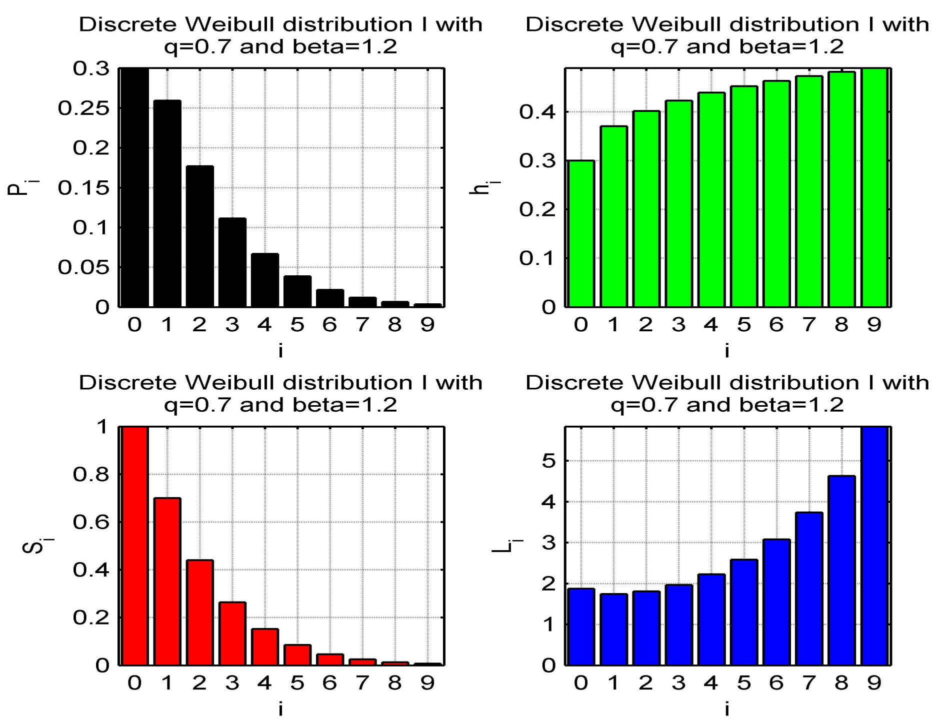


Figure 3/4: Display of the functions of a WEIBULL type I distribution by the program *DiscDist*



Part II

Inferential Aspects

4 Sampling Lifetime Data[1]

The estimation approach as well as the testing approach for the hazard rate and for any other lifetime representative have to take account of the type of the sampled data, i.e., on the way the data set has been generated and on the form in which the data are handed over to the statistician, either as individual observations (**non–grouped**) or as frequency counts per interval of time (**grouped**). We will revert to the latter aspect by the end of this chapter when commenting on Figures 4/1 and 4/2.

The true problem of sampling lifetime data is the fact that the characteristic to be measured is time itself. We might have long–lasting life–testing experiments unless we shorten them in one way or the other, e.g. by acceleration, see Sect. 1.1.2.4, or by censoring. Thus, in a lifetime data set we will find observations of complete lifetimes from birth to death (called **failure data** hereafter) and incomplete lifetimes ending before death or failure (called **censored data** hereafter). When the data set consists of the entire time spans ranging from birth or start to death or failure of each sampled unit, the data set is said to be **complete** or **uncensored**. In practice, uncensored lifetime data sets will be the rare exception. Clinical studies and biological trials or technical life testing will seldom lead to complete lifetimes of all sampled units for several reasons. A sample is called **incomplete** or **censored** when it consists of time spans covering the whole period of the unit's existence as well as of time spans with missing early and/or lifetime. The latter type of time spans are called censored times. In most cases the late lifetime is missing because the observation of an individual is not terminated by its death or its failure but by some other event. Thus, in clinical trials there may a loss to follow–up of a person or its death due to another risk other than that under study, and in technical life testing we meet planned withdrawal alive or stopping of the experiment before all units have failed.

We may distinguish between **random censoring** and non–random censoring. Most of the results presented in Part II are valid for random censoring only, but in practice the findings are also used when censoring is deterministic and planned. An assumption concerning random as well as non–random censoring is that **censoring is non–informative**. This means that the failure mechanism and the censoring mechanism are assumed to act independently. Stated otherwise, for each unit the censoring must not be predictive for the future and unobserved failure. Specifically, it must be true for each unit at each lifetime x that

$$\Pr\big\{[X \in [x,\, x+dx)\,|\, X \geq x\big\} = \Pr\big\{X \in [x,\, x+dx)\,|\, X \geq x,\, Z \geq x\big\},\ dx \text{ small}, \tag{4.1}$$

Z being the censoring variate. (4.1) means that the probability of failing shortly after x, given survival up to and including x, is unchanged by the added condition that censoring has not occurred up to and including time x. Unfortunately, the truth of (4.1) cannot be tested from the censored sample alone. In practice, a judgement about the truth of (4.1) should be sought on the best available understanding of the nature of censoring applied.

A very simple random censoring process that is often realistic is one in which each unit is assumed to be endowed with two random variables, a lifetime X and a censoring time Z, X and Z being independent and continuous variates, having CDFs $F(x)$ and $G(z)$, respectively. For example, Z may be the time associated to the happening of a competing risk. With n being the sample size let $(X_i\, Z_i);\ i = 1,\, 2,\, ...,\, n$; be independent and define

$$Y_i = \min(X_i,\, Z_i), \tag{4.2a}$$

[1] Suggested reading for this chapter: RINNE (2009, Chapter 8).

and the indicator

$$\delta_i = I_{(X_i \le Z_i)} = \begin{cases} 1, & \text{if } X_i \le Z_i \text{ (uncensored observation)}, \\ 0, & \text{if } X_i > Z_i \text{ (censored observation)}. \end{cases} \tag{4.2b}$$

The data from observing n units now consists of the pairs $(y_i,\, \delta_i)$, i.e., it is known which observation is a failure time and which is a censored time. The joint probability of $(y_i,\, \delta_i)$ is obtained using $f(x)$ and $g(z)$, the PDFs of X and Z, respectively. We have

$$\begin{aligned} \Pr(Y_i = y,\, \delta_i = 0) &= \Pr(X_i > y,\, Z_i = y) \\ &= \big[1 - F(y)\big]\, g(y) \end{aligned} \tag{4.2c}$$

and

$$\begin{aligned} \Pr(Y_i = y,\, \delta_i = 1) &= \Pr(Z_i > y,\, X_i = y) \\ &= \big[1 - G(y)\big]\, f(y). \end{aligned} \tag{4.2d}$$

These probabilities can be combined into the single expression

$$\Pr(Y_i = y,\, \delta_i) = \big\{f(y)\,[1 - G(y)]\big\}^{\delta_i}\, \big\{g(y)\,[1 - F(y)]\big\}^{1-\delta_i}. \tag{4.2e}$$

From (4.2e) the joint probability of the n pairs $(y_i,\, \delta_i)$ results as

$$\left.\begin{aligned} &\prod_{i=1}^{n} \big\{f(y_i)\,[1 - G(y_i)]\big\}^{\delta_i}\, \big\{g(y_i)\,[1 - F(y_i)]\big\}^{1-\delta_i} \\ &\qquad = \left(\prod_{i=1}^{n} \big[1 - G(y_i)\big]^{\delta_i} g(y_i)^{1-\delta_i}\right) \left(\prod_{i=1}^{n} f(y_i)^{\delta_i} \big[1 - F(y_i)\big]^{1-\delta_i}\right). \end{aligned}\right\} \tag{4.2f}$$

If $G(Y)$ and $g(y)$ do not involve any parameters of $F(y)$ and $f(y)$, then the first factor on the right–hand side of (4.2f) can be neglected and the resulting expression taken to be proportional to the **likelihood function** of the data:

$$L \propto \prod_{i=1}^{n} f(y_i)^{\delta_i} \big[1 - F(y_i)\big]^{1-\delta_i}, \tag{4.2g}$$

which constitutes the basis for maximum likelihood estimation, see Sections 7.1 and 7.2.

Censoring may be a random event and prevails in clinical studies, e.g., a person being member of a special cancer survival study dies of a stroke or has a fatal traffic accident. Non–random censoring prevails in life testing of technical units where the times of removing non–failed units are scheduled at the beginning of the experiment. Non–random censoring takes different forms.[2] According to what part of a lifetime is cut off and not reported by the sampling process, we distinguish between

- **censoring from above (on the right)** when we do not observe the failure of a unit, i.e., the last part of a lifetime is missing,
- **censoring from below (on the left)** when we do not know the 'date of birth' of a unit, i.e., observation starts at an unknown age of the unit and the first part of its lifetime is missing,
- **censoring on both side**, which is a combination of censoring from above and below.

[2] A detailed description of life test plans, their motivation and their economic as well as their statistical advantages and drawbacks is given in RINNE (2009, Chapter 8).

Censoring from above is most common with clinical studies and with technical life testing, and that is why we will assume this type of censoring throughout Part II unless stated otherwise.

We further distinguish between:

- **type–I censoring** (time–dependent censoring) when testing and observing of lifetimes are suspended when a **fixed time** x_{end} has been reached, (The maximum lifetime to be observed is x_{end}. The number of failures in x_{end} is random. Sometimes, depending on further censoring prescriptions, the number of censored lifetimes in x_{end} is random, too.),
- **type–II censoring** (failure–dependent censoring) when testing and observing are suspended by reaching a fixed number k of failures, $k < n,\ n$ being the sample size, (The maximum observable lifetime, which also is the length or the duration of the experiment, is the k-th order statistic $X_{k:n}$ which is random,[3])
- a **combination of type–I and type–II censoring**, meaning that testing and observation stop at $\min\big(x_{\text{end}}, X_{k:n}\big)$.

A third criterion in classifying non–random censoring is whether we have:

- **single censoring**, when all units that have not failed up to and including a certain time $\big[x_{\text{end}},\ X_{k:n}$ or $\min\big(x_{\text{end}}, X_{k:n}\big)\big]$ are withdrawn from the test so that all censored times are of equal length, or
- **multiple censoring** (**hypercensoring** or **progressive censoring**) when all the withdrawal of units still alive is performed through several stages so that the censored times are not of equal length as is the case with random censoring.

The following two figures show the quantities and data appearing in lifetime sampling. Fig. 4/1 is related to *non–grouped data* and depicts the most general case, i.e., multiple or random censoring with possibly tied observations. Other types of sampling with non–grouped data result from Fig. 4/1 when assigning special values to the quantities c_i and d_i which represent counts. Let $x_1 < x_2 < \ldots < x_k,\ k \leq n$, be the observed distinct times of a failure. We admit three possibilities of ties:

1) ties among censored observations,
2) ties among failure times, $d_i \geq 1$, being the **number of failures** happening at x_i,
3) ties among censored lifetimes and failure times.

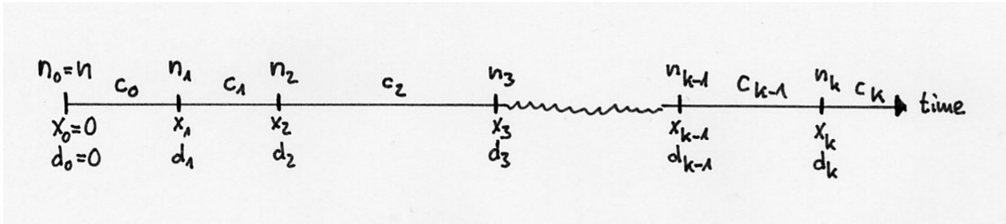
As time is a continuous variable ties of type 2) and 3) are theoretically impossible, but in practice time is nearly always counted in some unit, for instance in minutes, hours or so on, so that two or more events may happen 'simultaneously'. In order to avoid difficulties with case 2), **censored times tied with failure times**, we adopt the convention of moving such uncensored times in a tie a little amount to the right so that censoring is assumed to occur a little bit later than failure. This convention is sensible, since a unit observed alive at time x_i, certainly survives past x_i.

Beside d_i, which is attached to a certain point of time x_i, we have $c_i,\ c_i \geq 0$, the observed **number of censored lifetimes** between failure times x_i and x_{i+1}. Thus, c_i is attached to an interval of time, more precisely, to an interval of random length, which is right–opened: $[x_i,\ x_{i+1});\ i =$

[3] Commonly, $x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{n:n}$ denotes the ordered sample values. Sometimes, when there is no danger of confusion and in order to keep the mathematical notation as lean as possible we will refrain from using this special notation and x_i will stand for the i–th longest lifetime.

$0, 1, ..., k$; $x_0 = 0$ and $x_{k+1} = \infty$. This interval is in accordance with the convention above and any censoring happening at x_{i+1} is counted in the number c_{i+1} of the following interval.

Figure 4/1: Illustration of the numbers c_i, d_i, n_i and the failure times x_i on the time axis (non–grouped data)



The numbers n_i; $i = 0, 1, ..., k$; are attached to a point just prior to x_i. n_i is called the **number of units at risk** at x_i, and it counts the number of units which are alive and not censored and which are exposed to the risk of failure at x_i. The quantities n_i, c_i, d_i are linked as

$$n_i = n_{i-1} - c_{i-1} - d_{i-1};\ i = 1, 2, ..., k; \tag{4.3a}$$

where

$$n_0 = n,\ d_0 = 0.$$

The sample size n can be expressed as

$$n = n_0 = (d_0 + ... + d_k) + (c_0 + ... + c_k), \tag{4.3b}$$

i.e., it is divided into the number of units with complete lifetimes and of incomplete lifetimes, respectively. The number of censored lifetimes in the right–opened interval $[x_i, x_{i+1})$ is

$$c_i = n_i - d_i - n_{i+1}. \tag{4.3c}$$

We now look at some special situations.

1. For $c_0 = c_1 = ... = c_k = 0$ there is no censoring and we have

   $$n_i = n - \sum_{j=1}^{i-1} d_j;\ i = 1, 2, ..., k;$$

   and especially

   $$d_k = n_k = n - \sum_{i=1}^{k-1} d_i.$$

2. When there are no tied failure times $(d_i = 1\ \forall\, i)$ and no censoring we have

   a) $k = n$ and

   b) $n_i = n - i + 1;\ i = 1, 2, ..., n.$

3. For single censoring of type–II with ℓ as given number of failures to be observed and possibly tied failure times we have k so that

   $$\sum_{i=1}^{k-1} d_i < \ell \le \sum_{i=1}^{k} d_i$$

   and $c_0 = c_1 = ... = c_{k-1} = 0$ and $c_k = n - \sum\limits_{i=1}^{k} d_i$.

4. For single censoring of type–I with the single censoring time x_{end} and possibly tied failure times we may have

   a) no failures before x_{end}, so that neither $d_1,\, d_2,\, ...,\, d_k$ nor $c_1,\, c_2,\, ...,\, c_k$ exist and $c_0 = n_0 = n$, or

   b) $k \geq 1$ failure times $x_1 < x_2 < ... < x_k < x_{\text{end}}$ so that $c_k = n_k - d_k \geq 0$ lifetimes will be censored at x_{end} somewhere behind the last failure time x_k. It might happen that $c_k = 0$ when all units have failed before x_{end}.

Figure 4/2: Illustration of the numbers $c_j,\; d_j,\; n_j$ for a divided time axis (grouped data)

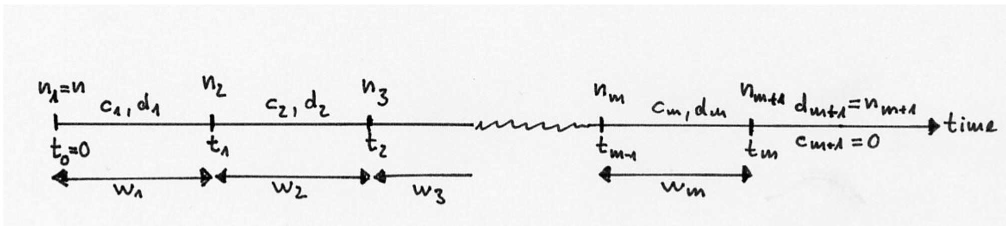


Fig. 4/2 depicts the situation for *grouped sampling data*. The time axis is divided into $m + 1$ intervals, not necessarily of equal length:

$$I_j = [t_{j-1},\, t_j);\; j = 1,\, 2,\, ...,\, m+1; \tag{4.4a}$$

where $t_0 = 0,\; t_m = t_{\text{end}}$ and $t_{m+1} = \infty$, and t_{end} is an upper limit on observation. These intervals are fix and have non–random width

$$w_j = t_j - t_{j-1};\; j = 1,\, 2,\, ...,\, m+1. \tag{4.4b}$$

The analysis of grouped lifetime data is done by actuarial methods within a so–called life table. In life tables for human populations, i.e., in demography, t_{end} generally is 100 years and the interval width has a constant length of one year for the first $m = 100$ intervals. But in general, the widths need not be constant. Each member in a sample of n units whose lifetime starts at t_0 either has a failure time or a censoring time. These times are counted per interval. We now define the following quantities:

d_j = number of lifetimes ending by a failure or death in I_j, (Remember that for non–grouped data d_i is attached to a point of time.)

c_j = number of lifetimes in I_j ending by censoring,

n_j = number of units entering I_j alive (= number of units at risk at t_{j-1}).

These numbers are linked as

$$n_j = n_{j-1} - d_{j-1} - c_{j-1};\; j = 2,\, 3,\, ...,\, m+1; \tag{4.4c}$$

where $n_1 = n$ is the sample size. In the last interval $I_{m+1} = [t_m,\, t_{m+1}) = [t_m,\, \infty)$ it can be considered that only uncensored lifetimes are in this interval since the n_{m+1} units not failed by $t_m = t_{\text{end}}$ must fail somewhere in I_{m+1}. Thus we have

$$n_{m+1} = d_{m+1} \;\text{ and }\; c_{m+1} = 0. \tag{4.4d}$$

5 Hazard Rate Estimation and the KAPLAN/MEIER and NELSON/AALEN Approaches

The estimating approach for the hazard rate in this chapter is especially apt for non-grouped data sets of small to medium size. The quantities $n_i,\ c_i$ and d_i coming up here have been defined in Chapter 4 and are explained in Fig. 4/1. The hazard rate estimator which is derived in Sect. 5.1 is a by–product when estimating the survival function and comes as a pointwise estimator which — for continuous distributed lifetime — may be smoothed by several methods, see Chapter 8. The hazard rate estimator of Sect. 5.1 is input for estimating the cumulative hazard rate by the NELSON/AALEN method in Sect. 5.2.[1]

5.1 Estimating the Hazard Rate and the Survival Function[2]

The search for an estimator of the survival function begins by assuming that the date have arisen from a discrete distribution with probability mass values at the ordered distinct failure times $x_1 < x_2 < \ldots < x_k,\ k \leq n$. For a discrete distribution the hazard rate h_i is a conditional probability with interpretation, see (1.60a):

$$h_i = \Pr(X = x_i \,|\, X \geq x_-i) = \frac{\Pr(X = x_i)}{\Pr(X \geq x_i)}.$$

As has been shown in (1.61c), the survival function can be written in terms of the hazard rate:

$$S(x) = \prod_{i\,:\,x_i \leq x} (1 - h_i). \tag{5.1a}$$

Thus, a reasonable estimator of $S(x)$ will be

$$\widehat{S}(x) = \prod_{i\,:\,x_i \leq x} (1 - \widehat{h}_i), \tag{5.1b}$$

which reduces the problem of estimating the survival function to that of estimating the hazard rate at each observed failure time x_i. Choosing the maximum likelihood procedure we have an appropriate element for the likelihood function as[3]

$$L_i = h_i^{d_i}\,(1 - h_i)^{n_i - d_i};\ i = 1,\,2,\,\ldots,\,k. \tag{5.2a}$$

This expression is correct since

1. $d_i \geq 1$ is the number of failures at x_i and h_i is the conditional probability of failure at x_i and

[1] The procedures of this chapter are programmed and assembled in the MATLAB–file Hazard$_{01}$ which can be downloaded from `htpp://geb.uni-giessen.de/volltexte/2013/?????.`

[2] Suggested reading for this chapter: COX/OAKES (1984), KAPLAN/MEIER (1958), KLEIN/MOESCHBERGER (1997), LAWLESS (1982), LEEMIS (1995), SMITH (2002).

[3] Remember that n_i, the number of units at risk at x_i, is given by

$$n_i = n_{i-1} - c_{i-1} - d_{i-1};\ i = 1,\,2,\,\ldots,\,k;$$

with $n_0 = n$ and $d_0 = 0$.

2. $n_i - d_i$ is the number of units on test not failing at x_i with $1 - h_i$ as the probability of failing after x_i, conditioned on survival to time x_i.

Thus, the likelihood function for all h_i is

$$L(h_1, \ldots, h_k) = \prod_{i=1}^{k} L_i = \prod_{i=1}^{k} h_i^{d_i} (1 - h_i)^{n_i - d_i}, \tag{5.2b}$$

with log–likelihood function

$$\begin{aligned} \mathcal{L}(h_1, \ldots, h_k) &= \ln L(h_1, \ldots, h_k) \\ &= \sum_{i=1}^{k} \big[d_i \ln h_i + (n_i - d_i) \ln(1 - h_i)\big]. \end{aligned} \tag{5.2c}$$

The i–th element of the so–called **score vector** is

$$\frac{\partial \mathcal{L}(h_1, \ldots, h_k)}{\partial h_i} = \frac{d_i}{h_i} - \frac{n_i - d_i}{1 - h_i}. \tag{5.2d}$$

Equating the score vector to zero and solving for h_i yields the **maximum likelihood estimator** (**MLE**) for h_i :

$$\widehat{h}_i = \frac{d_i}{n_i};\ i = 1, 2, \ldots, k. \tag{5.2e}$$

This estimator is sensible, since d_i of the n_i units still at risk at time x_i fail, so the ratio of d_i to n_i is an appropriate estimator of the conditional probability of failure at x_i. This derivation may strike a familiar chord since, at each time x_i, estimating h_i with d_i divided by n_i is equivalent to estimating the probability of 'success', i.e., failing at x_i, for each of of the n_i units still on test. Thus, this derivation is equivalent to finding the MLE for the probability of success for k binomial variables P_i.

Using the particular estimator (5.2e) for the hazard rate, the survival function estimator (5.1b) becomes

$$\widehat{S}(x) = \prod_{i\,:\,x_i \le x} \left(1 - \frac{d_i}{n_i}\right);\ i = 1, \ldots, k; \tag{5.3a}$$

which is a so–called **product—limit estimator** (**PLE**). This version is commonly known as the **KAPLAN/MEIER estimator** (**KME**), see KAPLAN/MEIER (1958).[4] This estimator is a maximum likelihood estimator, too, as it is a function of the maximum likelihood estimated h_i. In (5.3a) the censored observations are not forgotten, they have been allowed for in n_i, the number of units at risk just before x_i, and the effect of censored observations in the survival function estimator is a larger downward step compared to the step–size if the there had been no censoring.

One problem that arises with the KME is that it is not defined past the last observed failure time x_k. The usual way to handle this problem is to cut off the estimator at x_k. But there are other suggestions. Some authors define

- $\widehat{S}(x) = 0$ for $x > x_k$ when $d_k = n_k$, i.e., when no sample units survived past x_k,
- $\widehat{S}(x) = \prod\limits_{i=1}^{k} \big(1 - d_i/n_i\big)$ for $x > x_k$ when $d_k < x_k$, i.e., when there are sample units surviving past x_k.

[4] Another PLE has been suggested by HERD (1960) and JOHNSON (1964), see the Excursus further down.

The first suggestion means that the in the sampled population no individual would survive age x_k, whereas the second suggestion supposes ever–lasting lifetime for some individuals in the sampled population.

We now mention two *special variants of the KME–formula* (5.3a).

1. When there are *neither multiple failures nor censored observations before* x_k, the longest recorded failure time, we will have

   - $d_i = 1$ for $i = 1, \ldots, n$ (or k in the case of single censoring) and
   - $n_i = n - i + 1$ as the number of units at risk (not failed) just before x_i.

   In this case the KME of (5.3a) will turn into

   $$\widehat{S}(x) = \prod_{j=1}^{i} \frac{n-j}{n-j+1} = \frac{n-i}{n} \text{ for } \left\{ \begin{array}{l} x_i \leq x < x_{i+1}, \\ i = 1, \ldots, n-1 \text{ or } k-1 \end{array} \right\} \tag{5.3b}$$

   which is the familiar staircase empirical survival function with a downward step of size $1/n$ at each x_i. The hazard rate estimator (5.2e) turns into

   $$\widehat{h}_i = \widehat{h}(x_i) = \frac{1}{n-i+1};\ i = 1, \ldots, n \text{ or } k. \tag{5.3c}$$

2. When we have *records on all observed times* — failure times as well as censored times — which are given as pairs $(y_i,\ \delta_i)$, see (4.2a,b), the KME of (5.3a) may be written as

   $$\widehat{S}(x) = \prod_{i\,:\,y_i \leq x} \left(\frac{n-i}{n-i+1} \right)^{\delta_i}, \tag{5.3d}$$

   when there are *no tied failure times*. y_i is any observation, either censored or not. The hazard rate estimator for this case is

   $$\widehat{h}_i = \widehat{h}(y_i) = \left\{ \begin{array}{l} \dfrac{1}{n-i+1} \text{ for } \delta_i = 1, \\ \text{not defined for } \delta_i = 0. \end{array} \right\} \tag{5.3e}$$

Excursus: The KAPLAN/MEIER estimator and the HERD/JOHNSON estimator written in terms of reverse ranks

Besides the KME we have another PLE which has been proposed by HERD (1960) and JOHNSON (1964) and which will be called **HERD/JOHNSON estimator**, abbreviated **HJE**. When there are no tied observations both estimators can be defined recursively using the **reverse ranks** r_i of $y_1 < y_2 < \ldots < y_n$:

$$r_i = n - i + 1,\ i = 1, 2, \ldots, n. \tag{5.4}$$

Using r_i the KME of (5.3d) turns into

$$\widehat{S}(x) = {}_{\text{KME}}\widehat{P}_i = \left(\frac{r_i - 1}{r_i} \right)^{\delta_i} {}_{\text{KME}}\widehat{P}_{i-1};\ y_i \leq x < y_{i+1};\ i = 1, \ldots, n; \tag{5.5a}$$

with starting value

$${}_{\text{KME}}\widehat{P}_0 = 1. \tag{5.5b}$$

The HJE is defined as

$$\widehat{S}(x) = {}_{\text{HJE}}\widehat{P}_i = \left(\frac{r_i}{r_i + 1} \right)^{\delta_i} {}_{\text{HJE}}\widehat{P}_{i-1};\ y_i \leq x < y_{i+1};\ i = 1, \ldots, n; \tag{5.6a}$$

with starting value

$$_{\text{HJE}}\widehat{P}_0 = 1. \tag{5.6b}$$

The recursion factor $r_i/(r_i+1)$ of the HJE is greater than the factor $(r_i-1)/r_i$ of the KME resulting in an always larger HJE. When the sample is uncensored, i.e., $\delta_i = 1 \;\forall\; i$, the HJE will be

$$_{\text{HJE}}\widehat{P}_i = \frac{n+1-i}{n+1}$$

while the KME will be

$$_{\text{KME}}\widehat{P}_i = \frac{n-i}{n}.$$

We conclude this excursus by stating that the KME being a MLE is more popular than the HJE.

We now want to find estimators for the variances of $\widehat{h}_i$ and $\widehat{S}(x)$. If the the possible failure times x_i are fixed and the censoring mechanism allows the number of failures d_i at each x_i to increase at the same rate as the sample size n, then the standard large–sample theory for MLEs applies as COX/OAKES (1984) stated. Thus, asymptotically $\sqrt{n}\,(\widehat{h}_i - h_i)$ will have a multivariate normal distribution with mean vector equal to zero and a variance–covariance matrix which can be estimated by the observed **FISHER information matrix**. The elements of the latter matrix require the derivatives of the score vector (5.2d) and read

$$-\frac{\partial \mathcal{L}(h_1, \dots, h_k)}{\partial h_i\, \partial h_\ell} = \left\{ \begin{array}{cl} \dfrac{d_i}{h_i^2} + \dfrac{n_i - d_i}{(1-h_i)^2} & \text{for } i = \ell \\ 0 & \text{for } i \neq \ell \end{array} \right\};\; i,\, \ell = 1, \dots, k. \tag{5.7a}$$

Substituting h_i by its MLE the diagonal elements, which are not equal to zero, read

$$-\frac{\partial \mathcal{L}(h_1, \dots, h_k)}{\partial h_i\, \partial h_i}\bigg|_{h_i = d_i/n_i} = \frac{n_i^3}{d_i\,(n_i - d_i)}. \tag{5.7b}$$

Thus, the estimated variance of $\widehat{h}_i$ is

$$\widehat{\text{Var}}\big(\widehat{h}_i\big) = \frac{d_i\,(n_i - d_i)}{n_i^3} \tag{5.7c}$$

which is also the variance of $1 - \widehat{h}_i$:

$$\widehat{\text{Var}}\big(1 - \widehat{h}_i\big) = \widehat{\text{Var}}\big(\widehat{h}_i\big). \tag{5.7d}$$

When the data set is given as pairs $(y_i,\, \delta_i)$ the variance of $\widehat{h}_i$ in (5.3e) and of $1 - \widehat{h}_i$ is

$$\widehat{\text{Var}}\big(\widehat{h}_i\big) = \widehat{\text{Var}}\big(1 - \widehat{h}_i\big) = \left\{ \begin{array}{cl} \dfrac{n-i}{(n-i+1)^3} & \text{for } \delta_i = 1 \\ \text{not defined} & \text{for } \delta_i = 0 \end{array} \right\} \tag{5.7e}$$

Pointwise confidence intervals for h_i can now be obtained via the normal approximation. A two–sided $(1-\alpha)$–confidence interval such as

$$\widehat{h}_i - \tau_{1-\alpha/2}\sqrt{\widehat{\text{Var}}\big(\widehat{h}_i\big)} \le h_i \le \widehat{h}_i + \tau_{1-\alpha/2}\sqrt{\widehat{\text{Var}}\big(\widehat{h}_i\big)} \tag{5.7f}$$

refers to a single observation. A larger multiplier than the standard normal percentile $\tau_{1-\alpha/2}$ would be needed for a simultaneous confidence interval over more than one lifetime.

Turning to the variance of $\widehat{S}(x)$ we have from (5.1b)

$$\ln \widehat{S}(x) = \sum_{i\,:\,x_i \le x} \ln\big(1 - \widehat{h}_i\big). \tag{5.8a}$$

We have just seen that the $\widehat{h}_i$'s are asymptotically independent, so that the asymptotic variance of $\ln \widehat{S}(x)$ and hance of $\widehat{S}(x)$ can easily be found for any fixed x. First, using (1.30b) for the approximate variance of a transformed variate we find

$$\begin{aligned} \widehat{\text{Var}}\big[\ln\big(1-\widehat{h}_i\big)\big] &\approx \left(\frac{1}{1-\widehat{h}_i}\right)^2 \widehat{\text{Var}}\big(1-\widehat{h}_i\big) \\ &= \frac{d_i}{n_i\,(n_i - d_i)} \end{aligned} \tag{5.8b}$$

and then

$$\begin{aligned} \widehat{\text{Var}}\big[\ln \widehat{S}(x)\big] &\approx \sum_{i\,:\,x_i \le x} \widehat{\text{Var}}\big[\ln\big(1-\widehat{h}_i\big)\big] \\ &= \sum_{i\,:\,x_i \le x} \frac{d_i}{n_i\,(n_i - d_i)}. \end{aligned} \tag{5.8c}$$

Finally, applying (1.30b) once more, but now to (5.8c) we have

$$\widehat{\text{Var}}\big[\widehat{S}(x)\big] \approx \Big[\widehat{S}(x)\Big]^2 \sum_{i\,:\,x_i \le x} \frac{d_i}{n_i\,(n_i - d_i)}, \tag{5.8d}$$

which is known as **GREENWOOD's formula**, GREENWOOD (1926). When the data are given as pairs $(y_i,\ \delta_i)$ so that (5.3d,e) and (5.7e) hold, we have instead of (5.8b-d)

$$\widehat{\text{Var}}\big[\ln(1-\widehat{h}_i)\big] \approx \left\{ \begin{array}{ll} \dfrac{1}{(n-i)\,(n-i+1)} & \text{for } \ \delta_i = 1 \\ \text{not defined} & \text{for } \ \delta_i = 0 \end{array} \right\}, \tag{5.8e}$$

$$\widehat{\text{Var}}\big[\ln \widehat{S}(x)\big] \approx \sum_{i\,:\,y_i \le x} \frac{\delta_i}{(n-i)\,(n-i+1)}, \tag{5.8f}$$

$$\widehat{\text{Var}}\big[\widehat{S}(x)\big] \approx \Big[\widehat{S}(x)\Big]^2 \sum_{i\,:\,y_i \le x} \frac{\delta_i}{n-i)\,(n-i+1)}. \tag{5.8g}$$

A pointwise confidence interval for $S(x)$ can be obtained analogous to (5.7f). GREENWOOD's formula may be seen unstable in the right tail of the distribution,[5] so some authors have proposed an alternative and simpler estimator originating in the binomial distribution, namely

$$\widehat{\text{Var}}\big[\widehat{S}(x_i)\big] \approx \frac{\Big[\widehat{S}(x_i)\Big]^2 \Big[1 - \widehat{S}(x)\Big]}{n_i}. \tag{5.9}$$

A rationale for (5.9) is given in COX/OAKES (1984, p. 51) who also suggest likelihood based confidence intervals resting upon a χ^2–distribution.

[5] We see that (5.8d) grows with x_i because d_i comes closer to n_i. (5.8d) would even become ∞ when for the last observed failure time x_k we would have $d_k = n_k$.

Example 5/1: Computation of $\widehat{h}_i,\ \widehat{S}(x_i)$ and their variances

The following data have often been used in the literature on lifetime analysis, e.g., see COX/OAKES (1984) or LEEMIS (1986). The following observations $(y_i,\,\delta_i)$ with time measured in weeks

y_i	6	6	6	6	7	9	10	10	11	13	16	17	19	20	22	23	25	32	32	34	35
δ_i	1	1	1	0	1	0	1	0	0	1	1	0	0	0	1	1	0	0	0	0	0

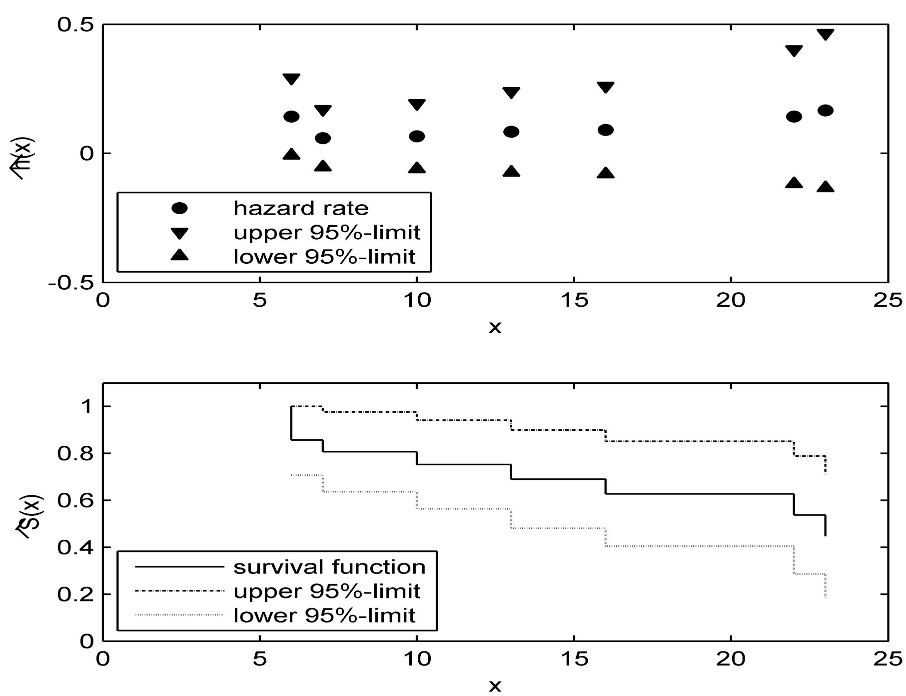
are the ordered times of remission, i.e., freedom of symptoms in a precisely defined medical sense, of 21 leukaemia patients who have been treated with the drug 6–MP (= 6-mercapto-purine). 12 patients have been lost to follow–up ($\delta_i = 0$). From the data above we have extracted and displayed in Tab. 5/1 the 'failure times' x_i, where — in this example — failure is a positive event.

Table 5/1: Estimates of h_i and $S(x_i)$ for the 21 leukaemia patients' data

i	x_i	n_i	d_i	$\widehat{h}_i$	$\widehat{\text{Var}}(\widehat{h}_i)$	$\widehat{S}(x_i)$	$\widehat{\text{Var}}\big[\widehat{S}(x_i)\big]$
1	6	21	3	0.1429	0.0058	0.8571	$5.8303 \cdot 10^{-3}$
2	7	17	1	0.0588	0.0033	0.8067	$7.5488 \cdot 10^{-3}$
3	10	15	1	0.0667	0.0041	0.7529	$9.2965 \cdot 10^{-3}$
4	13	12	1	0.0833	0.0064	0.6902	0.0114
5	16	11	1	0.0909	0.0075	0.6275	0.0130
6	22	7	1	0.1429	0.0175	0.5378	0.0165
7	23	6	1	0.1667	0.0231	0.4482	0.0181

$\widehat{S}(x)$ and $\widehat{\text{Var}}\big[\widehat{S}(x)\big]$ are not defined beyond $x = 23$.

Figure 5/1: Estimated hazard rate and survival function with pointwise 95%–confidence intervals for the 21 leukaemia–patients' data



Under the non–predictive censoring assumption the KME can be motivated in several ways. This estimator is

1. the **generalized MLE** in the same sense that the empirical distribution function is in the case of uncensored data, see MILLER (1981, pp. 57 ff.),
2. the **limit of the life–table estimators**, i.e., for data grouped in time intervals (see Sec. 8.2), as the intervals increase in number and decrease in length,
3. a **product od estimators** of conditional probabilities, see (5.1b),
4. a **self–consistent estimator**, see MILLER (1981, pp. 52 ff.),
5. the **redistribution–to–the–right estimator**, as proposed by EFRON (1967), see the following Example 5/2.

Example 5/2: The redistribution–to–the-right algorithm applied to the 21 leukemia patients' data

The algorithm starts with an empirical distribution that puts mass $1/n$ at each observation y_i; $i = 1, 2, ..., n$; and then eliminates and moves the mass of each censored observation $(y_i, 0)$ by distributing it equally to all observations to the right of it. After the last redistribution the estimated survival function at each $(y_i, 1)$ is unity minus the sum of the redistributed masses up to and including $(y_i, 1)$.

Mass at start ($*$ marks a censored observation)

6	6	6	6^*	7	9^*	10	10^*	11^*	13	16	$\dots$	34^*	35^*
$1/21$	$1/21$	$1/21$	$1/21$	$1/21$	$1/21$	$1/21$	$1/21$	$1/21$	$1/21$	$1/21$	$\dots$	$1/21$	$1/21$

Mass after the first redistribution

Combining the first three tied uncensored observations at $x = 6$ and first redistribution of $1/21$ at 6^* among the 17 subsequent observations we have

6	6^*	7	9^*	10	10^*	11^*	13	16	$\dots$	34^*	35^*
$1/7$	0	$6/119$	$6/119$	$6/119$	$6/119$	$6/119$	$6/119$	$6/119$	$\dots$	$6/119$	$6/119$

since $1/21 + (1/21)/17 = 6/119$.

Mass after the second redistribution

Redistribution of $6/119$ at 9^* to the 15 subsequent observations gives

6	6^*	7	9^*	10	10^*	11^*	13	16	$\dots$	34^*	35^*
$1/7$	0	$6/119$	0	$32/595$	$32/595$	$32/595$	$32/595$	$32/595$	$\dots$	$32/595$	$32/595$

since $6/119 + (6/119)/15 = 32/595$.

Mass after the third redistribution

Redistribution of $32/595$ at each of the two consecutive censored observations 10^* and 11^* to the 12 subsequent observations gives

6	6^*	7	9^*	10	10^*	11^*	13	16	$\dots$	34^*	35^*
$1/7$	0	$6/119$	0	$32/595$	0	0	$16/255$	$16/255$	$\dots$	$16/255$	$16/255$

since $32/595 + 2 \cdot (32/595)/12 = 16/255$.

When this process is continued through all the observed data and the resulting mass function is processed as described in the beginning of this example the KME results. To check this for one special value, e.g., for $x_5 = 16$, note that

$$\widehat{S}(16) = 1 - [1/7 + 6/119 + 32/595 + 16/255 + 16/255] = 0.6275$$

which matches the result given in Tab. 5/1.

5.2 Estimating the Cumulative Hazard rate[6]

We have two approaches to estimate $H(x)$ resulting from the two possibilities of defining $H(x)$ in the discrete case, see (1.62a,b):

- an indirect estimator built upon a previously estimated survival function and
- a direct estimator resting upon a previously estimated hazard rate.

Of course, both estimators have differing statistical properties and the generated estimates will not be identical to each other.

In Chapter 1, see (1.11f), we have seen that the cumulative hazard rate $H(x)$ and the survival function $S(x)$ are related as

$$H(x) = -\ln S(x).$$

So, the **indirect estimator** of $H(x)$, sometimes called the **natural estimator** of $H(x)$, is

$$\widetilde{H}(x) = -\ln \widehat{S}(x). \tag{5.10}$$

The KME of $S(x)$ is, see (5.3a) and (5.3d):[7]

$$\widehat{S}(x) = \left\{ \begin{array}{ll} \prod\limits_{i\,:\,x_i \le x} \left(1 - \dfrac{d_i}{n_i}\right); & i = 1,\, 2,\, \ldots,\, k \;\text{ for tied failure times} \\ \prod\limits_{i\,:\,y_i \le x} \left(\dfrac{n-i}{n-i+1}\right)^{\delta_i}; & i = 1,\, 2,\, \ldots,\, n \;\text{ for untied failure times.}^{8} \end{array} \right\} \tag{5.11a}$$

Thus, the natural estimator of $H(x)$ is

$$\widetilde{H}(x) = \left\{ \begin{array}{c} -\sum\limits_{i\,:\,x_i \le x} \ln \left(1 - \dfrac{d_i}{n_i}\right), \\ -\sum\limits_{i\,:\,y_i \le x} \delta_i \ln \left(\dfrac{n-i}{n-i+1}\right). \end{array} \right\} \tag{5.11b}$$

The estimated variance of $\ln \widehat{S}(X)$ is given in (5.8c) and (5.8f) so that for $\widetilde{H}(x) = -\ln \widehat{S}(x)$ the estimated variance reads

$$\widehat{\text{Var}}\big[\widetilde{H}(x)\big] = \left\{ \begin{array}{c} \sum\limits_{i\,:\,x_i \le x} \dfrac{d_i}{n_i\,(n_i - d_i)}, \\ \sum\limits_{i\,:\,y_i \le x} \dfrac{\delta_i}{(n-i)\,(n-i+1)}. \end{array} \right\} \tag{5.11c}$$

[6] Suggested reading for this section: AALEN (1978), ELANDT–JOHNSON/JOHNSON (1980, Chapter 2), GROSS/CLARK (1975, Sect. 4.7), LAWLESS (1982, Sect. 2.4), MILLER (1981, Chapter 3), NELSON (1969, 1970, 1972, 1982), SMITH (2002, Chapter 6).

[7] When we have grouped data we have to take the life table estimator of $S(x)$, see Sect. 6.2, which resembles the formula for tied failure times.

[8] For the definition of n_i and d_i see (4.31) and for that of y_i and δ_i see (4.2a,b).

The **direct estimator** of $H(x)$ comes with different names, either as **empirical accumulated hazard function**, see (1.62b), or as **NELSON/AALEN estimator**. It has first been suggested by NELSON (1972) in a reliability context and has been rediscovered by AALEN (1978) using modern counting process techniques. Using the MLEs of h_i in (5.2e) or (5.3e) the empirical accumulated hazard rate function results as[9]

$$\widehat{H}(x) = \left\{ \begin{array}{ll} \sum\limits_{i\,:\,x_i \le x} \dfrac{d_i}{n_i} & \text{for tied failure times,} \\ \sum\limits_{i\,:\,y_i \le x} \dfrac{\delta_i}{(n-i+1)} & \text{for untied failure times,} \end{array} \right\} \tag{5.12a}$$

with estimated variance

$$\widehat{\text{Var}}\big[\widehat{H}(x)\big] = \left\{ \begin{array}{l} \sum\limits_{i\,:\,x_i \le x} \dfrac{d_i\,(n_i - d_i)}{n_i^3}, \\ \sum\limits_{i\,:\,y_i \le x} \dfrac{\delta_i\,(n-i)}{(n-i+1)^3}. \end{array} \right\} \tag{5.12b}$$

AALEN (1978) gives an approximation: $\widehat{\text{Var}}\big[\widehat{H}(x_j)\big] \approx \sum\limits_{i=1}^{j} d_i/n_i^2$.

Since $\widetilde{H}(x) = -\ln \widehat{S}(x)$ we have from (5.11a)

$$\begin{aligned} \widetilde{H}(x) &= -\ln\left\{ \prod_{i\,:\,x_i \le x} \left(1 - \frac{d_i}{n_i}\right) \right\} \\ &= -\sum_{i\,:\,x_i \le x} \ln\left(1 - \frac{d_i}{n_i}\right) \\ &= \sum_{i\,:\,x_i \le x} \left(\frac{d_i}{n_i} + \frac{d_i^2}{2\,n_i^2} + \ldots \right). \end{aligned}$$

Thus, $\widehat{H}(x)$ in (5.12a) can be viewed as first–order approximation to $\widetilde{H}(x)$. Comparing the estimators $\widehat{H}(x)$ and $\widetilde{H}(x)$ we see that for a given set of failure data

- $\widehat{H}(x)$ produces smaller estimates than $\widetilde{H}(x)$ and
- the variance of $\widehat{H}(x)$ is smaller than that of $\widetilde{H}(x)$,

[9] We remark that based on the NELSON/AALEN estimator we may give another estimator of $S(x)$, namely

$$\widetilde{S}(x) = \exp\big[-\ln \widehat{H}(x_i)\big].$$

Furthermore, the estimator $\widehat{H}(x_i) = \sum\limits_{i\,:\,x_i \le x} d_i/n_i$ is correct only when we have non–grouped data with tied failure times where both d_i and n_i refer to the same points of time x_i, the failure time. With grouped data — see Sect. 6.2 — n_j refers to the start t_{j-1} of the j–th interval $I_j = [t_{j-1},\, t_j)$ and d_j is the number of failures in I_j. Thus, it seems reasonable in this case to use a slightly modified estimator

$$\widehat{H}(t_j^*) = \sum_{i=1}^{j} \frac{d_i}{n_i^*}$$

where

$$t_j^* = \frac{t_{j-1} + t_j}{2} \;\text{ and }\; n_j^* = \frac{n_i + n_{i+1}}{2}.$$

see Example 5/3 and Fig. 5/2. Under certain regulatory conditions one can show that both estimators are

- non–parametric MLEs,
- consistent,
- asymptotically equivalent and
- converge weakly to GAUSSIAN PROCESSES, see MILLER (1981), pp. 67 ff.), meaning that for fixed x, the estimators are approximately normally distributed.

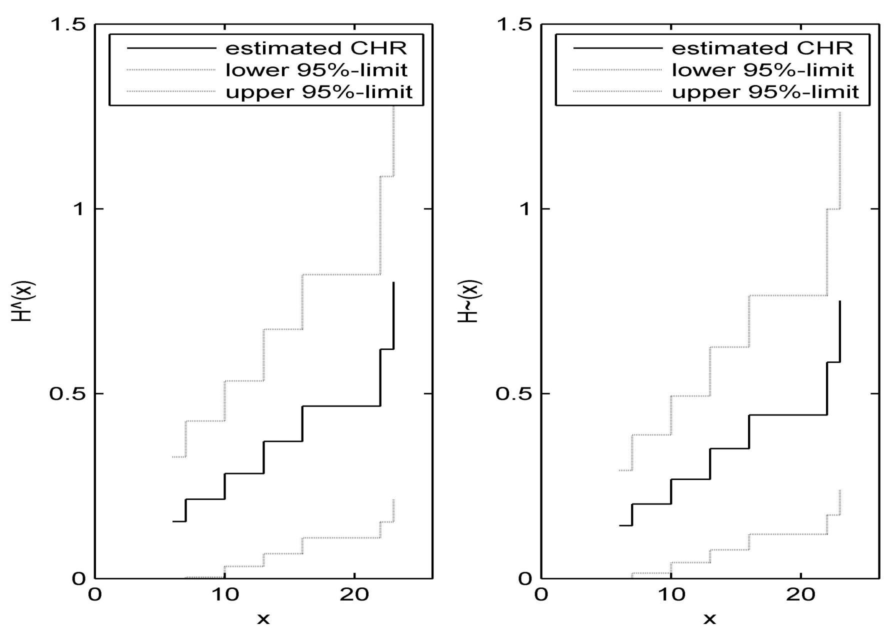
Example 5/3: Estimation of the CHR for the 21 leukaemia patients' data

We return to the data of Example 5/1 and Tab. 5/1.

Table 5/2: Estimates of $H(x_i)$ and its variances for the 21 leukaemia patients' data

i	x_i	n_i	d_i	$\widehat{S}(x_i)$	$\widetilde{H}(x_i)$	$\widehat{\text{Var}}\big[\widetilde{H}(x_i)\big]$	$\widehat{H}(x_i)$	$\widehat{\text{Var}}\big[\widehat{H}(x_i)\big]$
1	6	21	3	0.8571	0.1542	0.0079	0.1429	0.0058
2	7	17	1	0.8067	0.2148	0.0116	0.2017	0.0091
3	10	15	1	0.7529	0.2838	0.0164	0.2683	0.0132
4	13	12	1	0.6902	0.3708	0.0240	0.3517	0.0196
5	16	11	1	0.6275	0.4661	0.0330	0.4426	0.0271
6	22	7	1	0.5378	0.6202	0.0569	0.5854	0.0446
7	23	6	1	0.4482	0.8026	0.0902	0.7521	0.0678

Figure 5/2: Estimated CHR with pointwise 95%–confidence intervals
left part: indirect estimates; right part: direct (NELSON/AALEN) estimates



6 Estimating the Hazard Rate from Life Tables

One of the oldest tools in empirical statistics is the life table. Life tables have been used since the 17^{th} century[1] by demographers and actuaries to find the law of decrement for human populations and to calculate premiums for life insurances. But life tables apply equally well to reliability and to biostatistical situations for which grouped data rather than individual observations are available. The life table methods for estimating the survival function and other representatives of lifetime data are primarily designed for situations in which the sample size n is not too small and in which exact censored times and failure times shall not be processed or are not available at all. A drawback of hazard rates estimated by life table methods is their lack of smoothness and continuity.

This chapter is organized as follows:

- In Sect. 6.1 we will give basic definitions of life table functions and show how they are related.
- Sect. 6.2 is the core of this chapter where we see how life table functions including the hazard rate are to be estimated.
- In Sect.6.3 we review some further estimators of the hazard rate which are related to the life table approach.

6.1 Life Table Function[2]

A divided time axis is needed to set up a life table. We introduce $m+1$ preassigned time intervals[3]

$$I_i = [t_{j-1},\, t_j);\; j = 1,\, 2,\, ...,\, m+1;$$

where $t_0 = 0,\; t_m = t_{\text{end}}$ and $t_{m+1} = \infty$ with interval width

$$w_j = t_j - t_{j-1};\;\; j = 1,\, 2,\, ...,\, m+1.$$

t_{end} is an upper limit of observation, and for life tables we usually have $t_{\text{end}} = 100$ years.[4] For those life tables the interval width, generally, is constant and amounts to one year. In order to document the high infant mortality and its fast decline we sometimes have a finer division for the

[1] The first life tables have been compiled by J. GRAUNT in 1666 and E. HALLEY in 1693. Another pioneer is the Prussian statistician and demographer J. P. SÜSSMILCH (1707 – 1767). In the 19^{th} century B. GOMPERTZ (1779 – 1865) and W. MAKEHAM (1829 – 1871) were interested in the graduation of crude death rates for older people.

[2] Suggested reading for this section: ELANDT–JOHNSON/JOHNSON (1980), SHYROCK/SIEGEL (1976), SPIEGELMAN (1968).

[3] Another, but not so popular way of grouping is to construct the intervals so that they will contain fixed numbers of failures. By this way, the interval limits and their widths are random. Hazard rate estimation for this kind of grouping will be presented in Sect. 6.3

[4] As human life expectancy is growing we nowadays may find $t_{\text{end}} = 105$ years or even $t_{\text{end}} = 110$ years, especially for economically developed countries.

first and second years of life, i.e., days for the first week after birth, weeks and months thereafter. So–called **abridged life tables** on the other hand, have a width of five or even ten years.

A life table combines two types of quantities:

- **stocks,** measured or counted *at a* certain *point of time* or age, and
- **flows**, measured or counted *within a* certain *interval of time*.

Stocks and flows are linked by the general **updating formula**

stock at t_j = stock at t_{j-1} + incoming flows in $[t_{j-1},\, t_j)$ − outgoing flows in $[t_{j-1},\, t_j)$.

The updating formula shows that there are flows into the stock as well as out of the stock. Looking at a demographic life table we will only see flows out of the stock, where the stock is the number of persons ℓ_x alive at exact age x and the outgoing flow is the number ${}_1d_x$ of persons dying in $[x,\, x+1)$.

We will first look at the quantities and functions coming up in a demographic life table. Later on, when describing hazard rate estimation, we will change and simplify the notation and introduce some more quantities. The key quantities of a demographic life table are ${}_1q_x;\ x = 0,\, 1,\, \ldots, x_{\text{end}}$; the **conditional probabilities of dying** in $[x,\, x+1)$, given an individual is alive at exact age x. Roughly speaking, this ratio is obtained from dividing the number of persons dying in $[x,\, x+1)$, originating from a nation's mortality statistics by the number of persons entering the age of x, originating from a nation's population census. Then, a number ℓ_0 as starters at age $x = 0$ is taken to derive all the other quantities. ℓ_0 is called the **radix** of the life table, and it is often conventionally taken as 100,000 or 10,000. The ℓ_x, the **expected numbers of persons at exact age** x, are found recursively by applying

$$ {}_1p_x = 1 - {}_1q_x \tag{6.1a} $$

to ℓ_0 with ${}_1p_x$ as an individual's **conditional probability of not dying** in $[x,\, x+1)$, given being alive at exact age x. We thus find

$$ \begin{aligned} \ell_x &= (1 - {}_1q_{x-1})\, \ell_{x-1} \\ &= {}_1p_{x-1}\, \ell_{x-1} \\ &= \ell_0 \prod_{y=0}^{x-1} {}_1p_y. \end{aligned} \tag{6.1b} $$

The quantity

$$ {}_1d_x = {}_1q_x\, \ell_x = \ell_x - \ell_{x+1} \tag{6.1c} $$

is the **expected number of deaths** in $[x,\, x+1)$ from where we may write

$$ {}_1q_x = \frac{{}_1d_x}{\ell_x} = \frac{\ell_x - \ell_{x+1}}{\ell_x} \tag{6.1d} $$

and

$$ {}_1p_x = 1 - {}_1q_x = \frac{\ell_{x+1}}{\ell_x}. \tag{6.1e} $$

Whereas ℓ_x is the **expected number of survivors** out of ℓ_0 at exact age x, the **expected proportion of survivors** out of ℓ_0 is

$$ \Pi_x = \frac{\ell_x}{\ell_0}. \tag{6.1f} $$

This is in fact the survival function $S(x)$. With (6.1a,b) we may write Π_x as a telescope product:

$$\Pi_x = \prod_{y=0}^{x-1} {}_1p_y = \prod_{y=0}^{x-1}(1 - {}_1q_y), \tag{6.1g}$$

which parallels the product–limit estimator (PLE) of (5.3a). Also, the expected proportion surviving for k years, given alive at exact age x, is

$$\begin{aligned} {}_kp_x &= \prod_{y=x}^{x+k-1} {}_1p_y = \prod_{y=x}^{x+k-1} (1 - {}_1q_y) \\ &= \frac{\ell_{x+1}}{\ell_x}\,\frac{\ell_{x+2}}{\ell_{x+1}} \cdots \frac{\ell_{x+k}}{\ell_{x+k-1}} = \frac{\ell_{x+k}}{\ell_x}. \end{aligned} \tag{6.1h}$$

There are some more basic functions. The first one is L_x, **the expected total number of years lived in** $[x,\, x+1)$. L_x is nothing but the number of 'person $\times$ years' that ℓ_x persons, aged x exactly, are expected to live through $[x,\, x+1)$ and is recognized as a contribution of what is called total–time–on–test statistic in life testing. Each member in the group who survives the full year x to $x+1$ contributes exactly one year to L_x, whereas each member who dies in $[x,\, x+1)$ only contributes a fraction of a year to L_x. Formally, we have

$$L_x = \int_x^{x+1} \ell_y \, dy = \int_0^1 \ell_{x+u} \, du. \tag{6.2a}$$

This integral may be evaluated exactly when the age at death of each member is known. An approximation to L_x is

$$L_x \approx \frac{\ell_x + \ell_{x+1}}{2} = \ell_x - 0.5\, {}_1d_x, \tag{6.2b}$$

assuming the deaths to be equally distributed within $[x,\, x+1)$. If the interval width is other than one year, L_x of (6.2b) has to be multiplied by this width. The approximation (6.2b) tends to overestimate L_x for younger ages and to underestimate it for older ages.

T_x is the **expected total number of years live beyond age** x by the ℓ_x persons alive at that age:[5]

$$\begin{aligned} T_x &= L_x + L_{x+1} + \ldots + L_{x_{\text{end}}-1} \\ &= \sum_{u=0}^{x_{\text{end}}-x-1} L_{x+u}. \end{aligned} \tag{6.2c}$$

Of course,

$$T_x = T_{x-1} + L_x \tag{6.2d}$$

and using the approximation (6.2b)

$$\left.\begin{aligned} T_x &\approx \frac{\ell_x}{2} + \sum_{u=x+1}^{x_{\text{end}}} \ell_u \\ &= \sum_{x=0}^{x_{\text{end}}-1} (x+0.5)\, {}_1d_x \\ &= \frac{\ell_0}{2} + \sum_{x=0}^{x_{\text{end}}-1} x\, {}_1d_x. \end{aligned}\right\} \tag{6.2e}$$

[5] Remember that x_{end} is the oldest age reported, which is assumed not to be survived so that $\ell_{\text{end}} = 0$ and we have ${}_1d_{x_{\text{end}}-1} = \ell_{x_{\text{end}}-1} - \ell_{x_{\text{end}}} = \ell_{x_{\text{end}}-1}$. Furthermore, since all persons entering life at $x=0$ will die before x_{end} we have

$$\ell_0 = \sum_{x=0}^{x_{\text{end}}-1} {}_1d_x.$$

A last basic life table function is

$$\overset{o}{e}_x = \frac{T_x}{\ell_x}, \tag{6.3a}$$

the **expected future lifetime** of an individual alive at x. In Sect. 1.1.1.6 this quantity has been introduced by the name 'mean residual life' (MRL). Using (6.2e) $\overset{o}{e}_x$ may be written as

$$\overset{o}{e}_x = \frac{1}{2} + \frac{1}{\ell_x} \sum_{u=x+1}^{x_{\text{end}}} \ell_u. \tag{6.3b}$$

The **expected age at death** of a person surviving x is

$$\mathrm{E}(X \mid X \geq x) = x + \overset{o}{e}_x. \tag{6.3c}$$

The basic functions ${}_1q_x,\ \ell_x,\ {}_1d_x,\ L_x,\ T_x$ and $\overset{o}{e}_x$ are usually tabulated in a standard format as in Tab. 6/1, which is the life table 2000 – 2002 for German males. It has been constructed using updated results from the 1987 German population census and death statistics for the years 2000 through 2002. It is common practice to take the death statistics for more than one year to allow for years of over–mortality and of sub–mortality.

Table 6/1: Extraction from the German life table 2000 – 2002 for males

x	${}_1q_x$	ℓ_x	${}_1d_x$	L_x	T_x	$\overset{o}{e}_x$
0	0.00451281	100,000	451	99,605	7,537,995	75.38
1	0.00043340	99,549	43	99,527	7,438,390	74.72
2	0.00024513	99,506	24	99,493	7,338,863	73.75
3	0.00022031	99,481	22	99,479	7,239,370	72.77
4	0.00013878	99,459	14	99,452	7,139,899	71.79
⋮	⋮	⋮	⋮	⋮	⋮	⋮
96	0.30296858	2,242	679	1,902	5,541	2.47
97	0.32184621	1,563	503	1,311	3,639	2.33
98	0.34097780	1,060	361	879	2,328	2, 20
99	0.36031243	698	252	573	1,448	2.07
100	0.37979995	447	170	362	876	1.96

Source: Statistisches Bundesamt, ed., (2004)

6.2 Estimators for Life Table Functions Including the Hazard Rate[6]

In this section we assume strictly grouped data, i.e., we only know how many sample units either failed or had censored lifetimes in each interval. The case, where we have for each interval individually recorded failure times and censored times, will be treated in Sect. 6.3.

We now revert to the notation for grouped data as has been introduced in Fig. 4/2, i.e., we drop the two indices in ${}_nq_x,\ {}_np_x$ and ${}_nd_x$, which indicate flows out of the interval $[x,\ x+n)$, and we switch from x and y to i and j, conventionally used in counting. The demographic life table

[6] Suggested reading for this section: ELANDT–JOHNSON/JOHNSON (1980), GEHAN (1969), KIMBALL (1960), LONDON (1988), MILLER (1981), MÜLLER et al. (1997), SINGPURWALLA/WONG (1983), SMITH (2002).

does not know censored observations which usually might be encountered in life tables used in displaying and evaluating data from clinical and biological survival studies or from life testing. The type of life table used now is outlined in Tab. 6/2. We start with commenting upon the seven columns 2 – 8 which either reflect observations $(n_j,\ n'_j,\ c_j,\ d_j)$ or are quantities defined on the time axis $(t_j,\ t^*_j,\ w_j)$. The last five columns show estimates, and their estimators are the core of this section.

Table 6/2: Lay–out of a non–demographic life table

j	I_j	w_j	t^*_j	n_j	c_j	n'_j	d_j	$\widehat{q}_j$	$\widehat{p}_j$	$\widehat{\Pi}_j$	$\widehat{f}(t^*_j)$	$\widehat{h}(t^*_j)$
1	$[t_0,\ t_1)$	w_1	t^*_1	n_1	c_1	n'_1	d_1	$\widehat{q}_1$	$\widehat{p}_1$	$\widehat{\Pi}_1 = 1$	$\widehat{f}(t^*_1)$	$\widehat{h}(t^*_1)$
2	$[t_1,\ t_2)$	w_2	t^*_2	n_2	c_2	n'_2	d_2	$\widehat{q}_2$	$\widehat{p}_2$	$\widehat{\Pi}_2$	$\widehat{f}(t^*_2)$	$\widehat{h}(t^*_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	$[t_{m-1},\ t_m)$	w_m	t^*_m	n_m	c_m	n'_m	d_m	$\widehat{q}_m$	$\widehat{p}_m$	$\widehat{\Pi}_m$	$\widehat{f}(t^*_m)$	$\widehat{h}(t^*_m)$
$m+1$	$[t_m,\ \infty)$	–	–	n_{m+1}	c_{m+1}	n'_{m+1}	d_{m+1}	1	0	$\widehat{\Pi}_{m+1}$	–	–

1. $$I_j = [t_{j-1},\ t_j);\ j = 1,\ 2,\ ...,\ m+1;\ t_{m+1} = \infty,\ t_0 = 0; \tag{6.4a}$$

   is the *half–open time interval*. The last interval is infinite in length.

2. $$w_j = t_j - t_{j-1};\ j = 1,\ 2,\ ...,\ m; \tag{6.4b}$$

   is the *width of the interval* j. A constant width will be denoted w. The widths are required to estimate rates such as PDF and HR. Since the width of the last interval is infinite, no estimate of either PDF and HR can be given for this interval.

3. $$t^*_j = \frac{t_{j-1} + t_j}{2};\ j = 1,\ 2,\ ...,\ m; \tag{6.4c}$$

   is the *midpoint of interval j*. The midpoints are used as point of reference for the estimated PDF and HR which are assumed to be constant within I_j.

4. $c_j;\ j = 1,\ 2,\ ...,\ m+1;$ is the *total number of units* whose lifetime is **censored** in I_j without regard to the reason of censoring.[7]

5. $d_j;\ j = 1,\ 2,\ ...,\ m+1;$ is the *total number of units who die or fail* in the j–th interval.

6. $n_j;\ j = 1,\ 2,\ ...,\ m+1;$ is the *number of units entering* I_j. Especially we have

   $$n_1 = n, \tag{6.4d}$$

   n being the sample size. n_j is the number of units exposed to the risk of either dying (failing) or being censored in I_j. The updating formula linking the n_j's is

   $$n_j = n_{j-1} - d_{j-1} - c_{j-1};\ j = 2,\ 3,\ ...,\ m+1. \tag{6.4e}$$

[7] There are life tables with c_j split into several categories: the number of losses to follow–up, the number of withdrawals not failed or the number of deaths or failures due to another reason than that under investigation. These numbers are irrelevant for the estimation process and thus are not displayed here.

Furthermore, the sample size can be decomposed as

$$n = \sum_{j=1}^{m+1} d_j + \sum_{j+1}^{m} c_j. \tag{6.4f}$$

7. $n'_j;\ j = 1, 2, ..., m+1$; is the *number of units exposed to risk of dying or failing* in I_j for the special reason under study. If all the censoring would occur immediately at the start of I_j, then the number at risk (with the potential of dying or failing in the interval) is essentially $n_j - c_j$. On the other hand, if the censoring occurs just before the end of the interval, the censored units were at risk for the whole duration of I_j and the number of units at risk is obtained by interpreting $c_j = 0$. Averaging these extrem numbers of units at risk for estimation purposes we have

$$n'_j = \frac{(n_j - c_j) + n_j}{2} = n_j - \frac{c_j}{2}. \tag{6.4g}$$

This definition of n'_j essentially assumes that all censoring happens at t^*_j or occurs uniformly across I_j.

The estimated quantities $\widehat{q}_j$, $\widehat{p}_j$ and $\widehat{\Pi}_j$, which are needed in calculating $\widehat{f}(t^*_j)$ and $\widehat{h}(t^*_j)$, have counterparts in the sampled population. We will present these first. The *unconditional probability* for a member of the sampled population *to survive up to the start of* I_j is

$$\begin{aligned} \Pi_j &= S(t_{j-1});\ j = 1, 2, ..., m+1 \ \text{ with} \\ \Pi_1 &= S(t_0) = S(0) = 1. \end{aligned} \tag{6.5a}$$

The *unconditional probability of its failing* in I_j is

$$\begin{aligned} \pi_j &= \int\limits_{t_{j-1}}^{t_j} f(u)\,du = F(t_j) - F(t_{j-1}) = S(t_{j-1}) - S(t_j) \\ &= \Pi_j - \Pi_{j+1};\ j = 1, 2, ..., m+1;\ \text{ with } \Pi_{m+2} = 0. \end{aligned} \tag{6.5b}$$

Combining π_j and Π_j we have the following *conditional probability of failing in* I_j, given survival up to the start of I_j :

$$q_j = \frac{\pi_j}{\Pi_j} = 1 - \frac{\Pi_{j+1}}{\Pi_j}. \tag{6.5c}$$

The complement

$$p_j = 1 - q_j = \frac{\Pi_{j+1}}{\Pi_j} \tag{6.5d}$$

is the *conditional probability of surviving* I_j. We immediately see that

$$\Pi_j = p_1 \cdot p_2 \cdot ... \cdot p_{j-1};\ j = 1, 2, ..., m+1;\ \text{ with } p_0 = 1. \tag{6.5e}$$

Thus, Π_j is the cumulated product of the conditional survival probabilities for the first $j-1$ intervals. (6.5e) is the life table analogue of the product limit estimator (5.3a). Upon combining (6.5c,e) we may write the unconditional failure probability in I_j as[8]

$$\pi_j = \Pi_j\, q_j = q_j \prod_{i=1}^{j-1} p_i, \tag{6.5f}$$

[8] Remember $\prod\limits_{i=1}^{k} a_i = 1$ for $k < 1$.

and we have

$$\begin{aligned}\pi_1 &= q_1 \\ p_1 &= 1 - q_i = \Pi_2\end{aligned}$$

Estimating $\Pi_j,\ q_j$ and p_j when there is no censoring

When there is no censoring the construction of estimators for $q_j,\ p_j$ and Π_j is straightforward. In this case we have

$$n = \sum_{j=1}^{m+1} d_j, \tag{6.6a}$$

i.e., all members of the sample will be observed dying or failing. The set $\{d_1,\, d_2,\, ...,\, d_{m+1}\}$ has a **multinomial distribution**:

$$\Pr(d_1,\, d_2,\, ...,\, d_{m+1}) = n! \prod_{j=1}^{m+1} \frac{\pi^{d_j}}{d_j!} \tag{6.6b}$$

with

$$\mathrm{E}(d_j) = n\,\pi_j, \tag{6.6c}$$

$$\mathrm{Var}(d_j) = n\,\pi_j\,(1-\pi_j), \tag{6.6d}$$

$$\mathrm{Cov}(d_j,\, d_k) = -n\,\pi_j\,\pi_k,\ j \neq k. \tag{6.6e}$$

The total number of units entering I_j, i.e., surviving up to t_{j-1} is

$$n_j = n - (d_1 + d_2 + ... + d_{j-1});\ j = 2,\, 3,\, ...,\, m+1; \tag{6.7a}$$

with **binomial distribution**

$$\Pr(n_j) = \binom{n}{n_j} \Pi_j^{n_j}\,(1-\Pi_j)^{n-n_j};\ j = 2,\, 3,\, ...,\, m+1. \tag{6.7b}$$

The well–known MLE of the binomial parameter Π_j is

$$\widehat{\Pi}_j = \frac{n_j}{n};\ j = 2,\, 3,\, ...,\, m+1; \tag{6.7c}$$

which is also binomially distributed and has

$$\mathrm{E}(\widehat{\Pi}_j) = \Pi_j, \tag{6.7d}$$

$$\mathrm{Var}(\widehat{\Pi}_j) = \frac{\Pi_j\,(1-\Pi_j)}{n}. \tag{6.7e}$$

We mention that $\widehat{\Pi}_i$ and $\widehat{\Pi}_j$ with $i < j$ are positively correlated:

$$\mathrm{Cor}(\widehat{\Pi}_i,\, \widehat{\Pi}_j) = \sqrt{\frac{1-\Pi_i}{\Pi_i}\,\frac{1-\Pi_j}{\Pi_j}}\,. \tag{6.7f}$$

With $n \to \infty$ the distribution of $\widehat{\Pi}_j$ goes to a normal distribution, and its mean and variance are estimated by (6.7d,e) upon substituting Π_j by its estimator $\widehat{\Pi}_j$. Thus, we easily have approximate confidence intervals for Π_j.

In (6.5c) we have seen that the conditional probability of dying or failing in $I_j = [t_{j-1},\, t_j)$, conditional on survival up to t_{j-1}, is q_j. So, the proportion of deaths or failures in I_j,

$$\widehat{q}_j = \frac{d_j}{n_j} \tag{6.8a}$$

has a binomial distribution with parameters n_j and q_j and

$$\begin{aligned} \mathrm{E}(\widehat{q}_j \,|\, n_j) &= q_j, &(6.8b)\\ \mathrm{E}(\widehat{p}_j \,|\, n_j) &= p_j \text{ with } \widehat{p}_j = 1 - \widehat{q}_j = \frac{n_j - d_j}{n_j}\,, &(6.8c)\\ \mathrm{Var}(\widehat{q}_j \,|\, n_j) &= \mathrm{Var}(\widehat{p}_j \,|\, n_j) = \frac{p_j\, q_j}{n_j}\,. &(6.8d) \end{aligned}$$

Conditional on $(n_1 = n,\, n_2,\, ...,\, n_j)$ the random variables $\widehat{q}_j,\, ...,\, \widehat{q}_j$ are mutually independent.

Besides the estimator of Π_j in (6.7c) there is another estimator which rests upon (6.5e):

$$\widehat{\Pi}_j = \widehat{p}_1 \cdot \widehat{p}_2 \cdot ... \cdot \widehat{p}_{j-1};\; j = 1,\, 2,\, ...,\, m+1;\; \widehat{p}_0 = 1. \tag{6.9a}$$

Because of the mutual independence of the $\widehat{q}_j$'s and hance of the $\widehat{p}_j$'s the estimator (6.9a) is unbiased and has the approximate conditional variance[9]

$$\mathrm{Var}(\widehat{\Pi}_j \,|\, n_1,\, ...,\, n_j) = \Pi_j^2 \sum_{i=1}^{j-1} \frac{q_i}{n_i\, p_i};\; j = 2,\, 3,\, ...,\, m+1; \tag{6.9b}$$

which is **GREENWOOD's formula**. Substituting $\Pi_j,\; q_i$ and p_i by their estimates we obtain the estimated variance

$$\widehat{\mathrm{Var}}(\widehat{\Pi}_j \,|\, n_1,\, ...,\, n_j) = \widehat{\Pi}_j^2 \sum_{i=1}^{j-1} \frac{d_i}{n_i\,(n_i - d_i)};\; j = 2,\, 3,\, ...,\, m+1. \tag{6.9c}$$

Observing that $n_i - d_i = n_{i+1}$ we can rewrite a term of the sum in (6.9c) as

$$\frac{d_i}{n_i\,(n_i - d_i)} = \frac{n_i - n_{i+1}}{n_i\, n_{i+1}} = \frac{1}{n_{i+1}} - \frac{1}{n_i},$$

so that the sum turns into

$$\sum_{i=1}^{j-1} \left(\frac{1}{n_{i+1}} - \frac{1}{n_i}\right) = \frac{1}{n_j} - \frac{1}{n_1}.$$

As $n_1 = n$, we finally have

$$\begin{aligned} \widehat{\mathrm{Var}}(\widehat{\Pi}_j \,|\, n_1,\, ...,\, n_j) &= \widehat{\Pi}_j^2 \left(\frac{1}{n_j} - \frac{1}{n}\right) = \frac{\widehat{\Pi}_j^2\,(n - n_j)}{n\, n_j}\\ &= \frac{\widehat{\Pi}_j^2\,(1 - \widehat{\Pi}_j)}{n_j} = \frac{\widehat{\Pi}_j\,(1 - \widehat{\Pi}_j)}{n}, \end{aligned} \tag{6.9d}$$

since $\widehat{\Pi}_j = n_j/n$. So, (6.9c) also estimates the unconditional (exact) variance of $\widehat{\Pi}_j$ in (6.7e) after substituting Π_j with $\widehat{\Pi}_j$.

Estimating $\Pi_j,\; q_j$ and p_j when there is censoring

With censoring (6.7a) does no longer hold and the number n_j of units entering $I_j = [t_{j-1},\; t_j)$ is not the number of survivors up to t_{j-1} because some of the lifetimes censored before t_{j-1} may last longer than t_{j-1}. Consequently, Π_j cannot be estimated by (6.7c), and we have to revert to

[9] For a proof see ELANDT–JOHNSON/JOHNSON (1980, p. 140).

(6.5e) and its conditional survival probabilities in search for an estimator $\widehat{\Pi}_j$. Thus, we have to look for an estimator of $q_j = 1 - p_j$ under censoring. The estimator d_j/n_j in (6.8a) might be expected to underestimate q_j, since it is possible that some of the units censored in I_j might have failed or died before the end of I_j, hat they not been censored first. It is therefore desirable to make some adjustments for the censored units. The most commonly used procedure is to estimate q_j by the so–called **actuarial estimator**, also called **standard life table estimator**, which is

$$\widehat{q}_j = \frac{d_j}{n_i - c_j/2} = \frac{d_j}{n'_j}, \tag{6.10a}$$

i.e., we replace n_j in (6.8a) by n'_j, the average number of units exposed to risk of failing or dying, see (6.4g). This adjustment is arbitrary, but sensible in many situations. Its appropriateness depends on the failure and censoring process of course. Once estimates $\widehat{q}_j$ and $\widehat{p}_j = 1 - \widehat{q}_j$ have been calculated, Π_j can be estimated using (6.9a).

Conditioned on n_j and c_j and assuming that $\widehat{q}_j$ of (6.10a) is approximately a binomial proportion we have

$$\begin{aligned} \mathrm{Var}(\widehat{q}_j \,|\, n_j,\, c_j) &= \mathrm{Var}(\widehat{p}_j \,|\, n_j,\, c_j) \\ &\approx \frac{p_j\, q_j}{n'_j} = \frac{p_j\, q_j}{n_j - c_j/2}. \end{aligned} \tag{6.10b}$$

An estimator $\widehat{\mathrm{Var}}(\widehat{q}_j \,|\, n_j,\, c_j)$ is found by replacing p_j and q_j by their estimators. Conditional on the sets $\{n_j\} = (n_i,\, ...,\, n_j)$ and $\{c_j\} = (c_1,\, ...,\, c_j)$, GREENWOOD's formula for the conditional variance of $\widehat{\Pi}_j$, derived for the uncensored data in (6.9b,c), is approximately valid here when n_j is replaced with $n'_j = n_j - c_j/2$. So we have

$$\widehat{\mathrm{Var}}(\widehat{\Pi}_j \,|\, \{n_j\},\, \{c_j\}) \approx \widehat{\Pi}_j^2 \sum_{i=1}^{j-1} \frac{\widehat{q}_i}{n'_i\, \widehat{p}_i};\; j = 2,\, 3,\, ...,\, m+1. \tag{6.10c}$$

This estimator is reasonable provided $\mathrm{E}(n'_j)$ is not too small, though, when there is a lot of censoring, $m+1$, the number of intervals, should not be too small. (6.10c) sometimes tends to underestimate the variance of $\widehat{\Pi}_j$ for intervals in the right–hand tail of the lifetime distribution, essentially when $\mathrm{E}(n'_j)$ is quite small. However, in such instances the distribution of $\widehat{\Pi}_j$ is typically highly skewed, and its variance is not a particularly good indicator of estimator precision anyway.

Estimating $f(t_j^*)$

The PDF $f(x)$ of a lifetime distribution is also called **curve of deaths** in the context of life table analysis. With $t_j^* = (t_{j-1} + t_j)/2$ as midpoint and $w_j = t_j - t_{j-1}$ as width of the j–th interval[10] we can approximate the PDF at t_j^* by the well-known formula

$$\begin{aligned} f(t_j^*) &= -\frac{dS(t)}{dt}\bigg|_{t=t_j^*} \\ &\approx \frac{S(t_{j-1}) - S(t_j)}{w_j} \\ &= \frac{\Pi_j - \Pi_{j+1}}{w_j} \\ &= \frac{\Pi_j - p_j\,\Pi_j}{w_j} \end{aligned}$$

[10] As the last interval I_{m+1} has length ∞, in all the following formulas j ranges from 1 to m.

$$\begin{aligned} &= \frac{\Pi_j\, q_j}{w_j} \\ &= \frac{\pi_j}{\Pi_j};\ j = 1,\, 2,\, ...,\, m;\ \text{ see } \ (6.5\text{f}); \end{aligned} \tag{6.11}$$

i.e., $f(t_j^*)$ is the unconditional probability π_j of failing in I_j per unit width, the definition of the PDF.

To arrive at an estimator we insert estimators of Π_j and q_j. When the sample has *not* been *censored* we take $\widehat{q}_j = d_j/n_j$ and $\widehat{\Pi}_j = n_j/n$ and have the following estimator of the PDF at the midpoint t_j^* of I_j :

$$\widehat{f}(t_j^*) = \frac{\widehat{\Pi}_j\, \widehat{q}_j}{w_j} = \frac{d_j}{n\, w_j}. \tag{6.12a}$$

$\widehat{f}(t_j^*)$ is taken to construct the histogram–estimator of the PDF. From (6.6c,d) and (6.5c) we have

$$\begin{aligned} \mathrm{E}(d_j) &= n\, \pi_j \;=\; n\, q_j\, \Pi_j, & (6.12\text{b}) \\ \mathrm{Var}(d_j) &= n\, \pi_j\, (1 - \pi_j) \;=\; n\, q_j\, \Pi_j\, (1 - q_j\, \Pi_j), & (6.12\text{c}) \end{aligned}$$

so that *unconditional variance* of $\widehat{f}(t_j^*)$ is

$$\mathrm{Var}\Big[\widehat{f}(t_j^*)\Big] = \left(\frac{1}{n\, w_j}\right)^2 \mathrm{Var}(d_j) = \frac{q_j\, \Pi_j\, (1 - q_j\, \Pi_j)}{n\, w_j^2} = \frac{\pi_j\, (1 - \pi_j)}{n\, w_j^2}. \tag{6.12d}$$

The estimated version of (6.12d) results when q_j and Π_j are replaced by their estimators $\widehat{q}_j = d_j/n_j$ and $\widehat{\Pi}_j = n_j/n$, respectively:

$$\widehat{\mathrm{Var}}\Big[\widehat{f}(t_j^*)\Big] = \frac{d_j\, (n - d_j)}{n^3\, w_j^2}. \tag{6.12e}$$

When the sample is *censored* we have to use

$$\widehat{f}(t_j^*) = \frac{\widehat{\Pi}_j\, \widehat{q}_j}{w_j} = \frac{\widehat{p}_1\, \widehat{p}_2 ... \widehat{p}_{j-1}\, \widehat{q}_j}{w_j} =: g(\widehat{p}_1\, \widehat{p}_2 ... \widehat{p}_{j-1}\, \widehat{p}_j), \tag{6.13a}$$

where $\widehat{q}_j = d_j/n'_j$ and $\widehat{p}_j = 1 - \widehat{q}_j$. For this estimator we can only give a *conditional variance*.[11] The problem now is to find this variance as the function $g(\widehat{p}_1\, \widehat{p}_2 ... \widehat{p}_{j-1}\, \widehat{p}_j)$ of j variates. Based on the large–sample–approximation formula

$$\mathrm{Var}\big[g(\widehat{p}_1\, \widehat{p}_2 ... \widehat{p}_{j-1}\, \widehat{p}_j)\big] \approx \sum_{i,k=1}^{j} \left\{ \frac{\partial g}{\partial \widehat{p}_i} \frac{\partial g}{\partial \widehat{p}_k} \mathrm{Cov}(\widehat{p}_i, \widehat{p}_k) \right\},$$

GEHAN (1969, p. 644) gives the result

$$\widehat{\mathrm{Var}}\big[\widehat{f}(t_j^*) \,|\, \{n'_j\}\big] \approx \left(\frac{\widehat{\Pi}_j\, \widehat{q}_j}{w_j}\right)^2 \sum_{i=1}^{j-1} \left(\frac{\widehat{q}_i}{n'_i\, \widehat{p}_i} + \frac{\widehat{p}_i}{n'_i\, \widehat{q}_i} \right), \tag{6.13b}$$

Estimating $h(t_j^*)$

As is well known the hazard rate at any point of time is defined as $h(t) = f(t)/S(t)$. For $t = t_j^*$ it is not directly estimated as $\widehat{f}(t_j^*)/\widehat{\Pi}_j$, since $\widehat{\Pi}_j$ is the estimated probability of survival up to t_j

[11] Of course, when there is no censoring we may use (6.13a) with n_j instead of n'_j, and we have a conditional variance (6.13b) with n_j.

and not up to t_j^*. We thus must find $\widehat{\Pi}(t_j^*)$, the probability of of surviving up to t_j^*, the midpoint of I_j. Clearly, by linear interpolation

$$\widehat{\Pi}(t_j^*) = \frac{\widehat{\Pi}_j + \widehat{\Pi}_{j+1}}{2} = \frac{\widehat{\Pi}_j\,(1+\widehat{p}_j)}{2}. \tag{6.14a}$$

Upon combining (6.13a) and (6.14a) we have

$$\widehat{h}^{(1)}(t_j^*) = \frac{\widehat{f}(t_j^*)}{\widehat{\Pi}((t_j^*)} = \frac{2\,\widehat{q}_j}{w_j\,(1+\widehat{p}_j)}. \tag{6.14b}$$

$\widehat{h}^{(1)}(t_j^*)$ is the most popular estimator, sometimes called **classical** or **actuarial estimator** of $h(t_j^*)$. (6.14b) can be transformed by inserting $\widehat{q}_j = d_j/n'_j$ into

$$\widehat{h}^{(1)}(t_j^*) = \frac{d_j}{w_j\,(n'_j - d_j/2)} \tag{6.14c}$$

which is known by the name **central death rate**. The denominator $w_j\,(n'_j - d_j/2)$ estimates the total–time–on–test spent in I_j, and so $\widehat{h}^{(1)}(t_j^*)$ is related to the estimators of Chapter 7, and in fact, GRENANDER (1956) has shown that this a MLE when the hazard rate is non–decreasing. An approximate variance estimator, conditioned on $\{n_j\}$, has bee derived by GEHAN (1969):

$$\widehat{\mathrm{Var}}\big[\,\widehat{h}^{(1)}(t_j^*)\,|\,\{n'_j\}\big] \approx \frac{\big[\,\widehat{h}^{(1)}(t_j^*)\big]^2}{n'_j\,\widehat{q}_j}\left\{1 - \left[\frac{w_j\,\widehat{h}^{(1)}(t_j^*)}{2}\right]^2\right\}. \tag{6.14d}$$

When n'_j is small, (6.14d) looses accuracy. This implies that if there are very few survivors in the later intervals, the computation of $\widehat{\mathrm{Var}}\big[\,\widehat{h}^{(1)}(t_j^*)\,|\,\{n_j\}\big]$ is not worthwhile for these later stages.

Another estimator for $h(t_j^*)$ is found by converting the conditional failure probability $\widehat{q}_j = d_j/n'_j$ into a rate:

$$\widehat{h}^{(2)}(t_j^*) = \frac{\widehat{q}_j}{w_j} = \frac{d_j}{w_j\,n'_j}, \tag{6.15a}$$

called **death rate**. Its estimated approximate variance, conditioned on $\{n'_j\}$, follows from (6.10b) as

$$\widehat{\mathrm{Var}}\big[\,\widehat{h}^{(2)}(t_j^*)\,|\,\{n'_j\}\big] \approx \frac{1}{w_j^2}\,\frac{\widehat{p}_j\,\widehat{q}_j}{n'_j}. \tag{6.15b}$$

The statistics $\widehat{h}^{(1)}(t_j^*)$ and $\widehat{h}^{(2)}(t_j^*)$ converge to different functions as $n \to \infty$, assuming for now that we have a constant width w, assumed fixed:[12]

$$\lim_{n\to\infty} \widehat{h}^{(1)}(t_j^*) = \frac{S(t_j^* - w/2) - S(t_j^* + w/2)}{\frac{w}{2}\big[S(t_j^* - w/2) - S(t_j^* + w/2)\big]} = \widetilde{h}^{(1)}(t_j^*), \tag{6.16a}$$

$$\lim_{n\to\infty} \widehat{h}^{(2)}(t_j^*) = \frac{S(t_j^* - w/2) - S(t_j^* + w/2)}{w\,S(t_j^* - w/2)} = \widetilde{h}^{(2)}(t_j^*). \tag{6.16b}$$

We see, that

$$\widetilde{h}^{(1)}(t_j^*) \le \frac{2}{w} \text{ and } \widetilde{h}^{(2)}(t_j^*) \le \frac{1}{w},$$

[12] See MÜLLER et al. (1997).

so that neither statistic can approximate $h(t_j^*)$ whenever $h(t_j^*) > 2/w$. This is the underlying reason for the **discretisation biases** inherent in $\widehat{h}^{(1)}(t_j^*)$ and $\widehat{h}^{(2)}(t_j^*)$ when viewed as estimators of $h(t_j^*)$.

The estimators in (6.14b) and (6.15a) have been derived for the hazard rate at the midpoint of $I_j = [t_{j-1}, t_j)$, but they are often taken as estimators for all t in I_j, i.e., by assuming the hazard rate is constant within I_j. There exists, however, a special estimator for the assumption that $h(t)$ is constant within an interval, but varies among the intervals. This estimator has been proposed by SACHER (1956). Let h_j be this constant hazard rate in I_j, then, given survival up to the start t_{j-1} of I_j, the chance of failure or death in I_j is $\big[1-\exp(-h_j\,w_j)\big]$ and the chance of surviving is $\big[\exp(-h_j\,w_j)\big]$. Assuming that there is *no censoring* the number of failures in I_j has a binomial distribution with parameters n_j (number of survivors up to t_{j-1}) and binomial proportion $1-\exp(-h_j\,w_j)$, i.e.,

$$\Pr(d_j) = \binom{n_j}{d_j}\big[1-\exp(-h_j\,w_j)\big]^{d_j}\,\big[\exp(-h_j\,w_j)\big]^{n_j-d_j}. \tag{6.17a}$$

The usual MLE of the binomial parameter $\big[1-\exp(-h_j\,w_j)\big]$ is

$$1-\widehat{\exp(-h}_j\,w_j) = \frac{d_j}{n_j}. \tag{6.17b}$$

Observing $d_j = n_j - n_{j+1}$ and $\widehat{p}_j = n_{j+1}/n_j$ we find, after some manipulations, **SACHER's estimator**[13]

$$\widehat{h}_j^{(3)} = -\frac{\ln\widehat{p}_j}{w_j}. \tag{6.17c}$$

This is a MLE too, and as it is a natural log–function of the binomial variate d_j its variance can be approximated by the method of statistical differentials using (1.30b). The result for the estimated version of the conditional variance then is

$$\widehat{\mathrm{Var}}\big[\widehat{h}_j^{(3)}\,|\,\{n_j\}\big] \approx \frac{1}{w_j^2}\,\frac{\widehat{q}_j}{n_j\,\widehat{p}_j}. \tag{6.17d}$$

Often, (6.17c,d) are applied when there is censoring, so that n_j is replaced with n_j'. This procedure is satisfying as long as the numbers of censored lifetimes per interval do not differ much. As GEHAN (1969) reports, Monte Carlo Studies have shown that $\widehat{h}^{(1)}(t_j^*)$ is less biased than $\widehat{h}_j^{(3)}$, and SINGPURWALLA/WONG (1983) report that $\widehat{h}_j^{(3)}$ is almost positively biased whereas $\widehat{h}^{(1)}(t_j^*)$ tends to be negatively biased as t_j^* increases.

Example 6/1: Life testing of 200 pieces of equipment

$n = 200$ pieces of equipment had been put on a life test for a certain type of failure. The test was scheduled to last at most $t_{\text{end}} = 240$ hours. By the end of the test 5 pieces had not failed. The reason for censoring in this life test is failure due to another reason than that under study and due to withdrawal of non–failed units for special further investigations. Table 6/3 gives the data and Table 6/4 displays estimates together with their estimated variances.

[13] An estimator similar to (6.17c) is

$$\widehat{h}_j^{(4)} = -\frac{1}{2}\big(\ln\widehat{p}_{j-1} + \ln\widehat{p}_j\big),$$

and from the series presentation of $\ln\widehat{p}_j = \ln(1-\widehat{q}_j) = -\widehat{q}_j + 0.5\,\widehat{q}_j^2 - \ldots$ we are — for h_j and q_j very small — back to (6.15a).

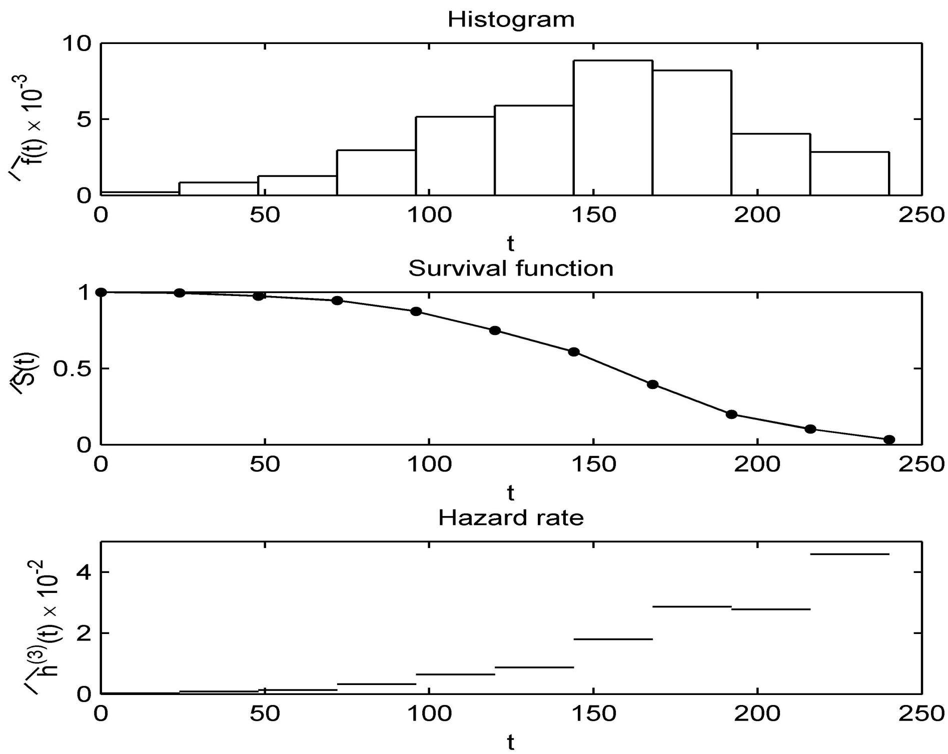
Table 6/3: Data for life table estimation

j	I_j	w_j	t_j^*	d_j	c_j	n_j	n_j'
1	$0 - 24$	24	12	1	0	200	200
2	$24 - 48$	24	36	4	1	199	198.5
3	$48 - 72$	24	60	6	1	194	193.5
4	$72 - 96$	24	84	14	2	187	186
5	$96 - 120$	24	108	24	3	171	169.5
6	$120 - 144$	24	132	27	1	144	143.5
7	$144 - 168$	24	156	40	3	116	114.5
8	$168 - 192$	24	180	36	1	73	72.5
9	$192 - 216$	24	204	17	2	36	35
10	$216 - 240$	24	228	11	1	17	16.5
11	$240 - \infty$	–	–	5	0	5	5

Table 6/4: Estimates (with variances) of life table quantities

j	$\widehat{q}_j$	$\widehat{p}_j$	$\widehat{\text{Var}}(\widehat{q}_j \mid n_j')$	$\widehat{\Pi}_j$	$\widehat{\text{Var}}(\widehat{\Pi}_j \mid \{n_j'\})$	$\widehat{f}(t_j^*)$	$\widehat{\text{Var}}[\widehat{f}(t_j^*) \mid \{n_j'\}]$	$\widehat{h}^{(1)}(t_j^*)$	$\widehat{\text{Var}}[\widehat{h}^{(1)}(t_j^*) \mid \{n_j'\}]$	$\widehat{h}^{(3)}(t_j^*)$	$\widehat{\text{Var}}[\widehat{h}^{(3)}(t_j^*) \mid \{n_j'\}]$
	$(7.10a)$		$(7.10b)$	$(7.9a)$	$(7.10c)$	$(7.13a)$	$(7.13b)$	$(7.14c)$	$(7.14d)$	$(7.17c)$	$(7.17d)$
			$\times 10^{-4}$		$\times 10^{-4}$	$\times 10^{-3}$	$\times 10^{-4}$	$\times 10^{-2}$	$\times 10^{-6}$	$\times 10^{-2}$	$\times 10^{-6}$
1	0.0050	0.9950	0.249	1	0	0.208	0.010	0.021	0.044	0.021	0.044
2	0.0202	0.9798	0.995	0.9950	0.249	0.835	0.208	0.085	0.180	0.085	0.180
3	0.0310	0.9690	1.553	0.9749	1.224	1.260	0.534	0.131	0.287	0.131	0.287
4	0.0753	0.9247	3.742	0.9447	2.625	2.963	3.093	0.326	0.757	0.326	0.760
5	0.1416	0.8584	7.171	0.8736	5.584	5.154	9.595	0.635	1.670	0.636	1.689
6	0.1882	0.8118	10.645	0.7499	9.588	5.879	12.747	0.865	2.744	0.869	2.804
7	0.3493	0.6507	19.852	0.6088	12.305	8.861	29.358	1.764	7.428	1.791	8.141
8	0.4966	0.5034	34.481	0.3961	12.568	8.196	25.555	2.752	18.747	2.859	23.618
9	0.4857	0.5143	71.370	0.1994	8.596	4.036	6.421	2.673	37.704	2.771	46.847
10	0.6667	0.3333	134.680	0.1026	5.112	2.849	3.495	4.167	118.371	4.578	210.438
11	1	0	0	0.0342	1.985	–	–	–	–	–	–

Figure 6/1: Plot of life table quantities



6.3 Related Hazard Rate estimators

Even if the data are grouped in fixed interval, we may still have *records of the exact failure and censoring times*. Let t^*_{ji} denote the recorded time of failure or censoring for the i–th unit in interval $I_j = [t_{j-1},\, t_j);\ j = 1,\, 2,\, ...,\, m+1;\ t_{m+1} = \infty$. For t^*_{ji} we have

$$t^*_{ji} = \left\{ \begin{array}{ll} t^+_{ji} & \text{if unit } i \text{ failed in } I_j, \\ t^-_{ji} & \text{if unit } i \text{ has been censored in } I_j, \\ t_j & \text{if unit } i \text{ survived } I_j. \end{array} \right\} \tag{6.18a}$$

The t^*_{ji}–s can be processed to give the amount of 'item $\times$ time units' spent in I_j :

$$\begin{aligned} L_j &= \sum_{i=1}^{n_j} (t^*_{ji} - t_{j-1}) \\ &= \sum_i (t^+_{ji} - t_{j-1}) + \sum_i (t^-_{ji} - t_{j-1}) + n_{j+1}\, w_j. \end{aligned} \tag{6.18b}$$

L_j is the sample equivalent of L_x in (6.2a), called the **amount exposed to risk**. The central death rate for I_j, which is taken as an estimator of the hazard rate in I_j, results as

$$\widehat{h}^{(5)}(t^*_j) = \frac{d_j}{L_j}, \tag{6.18c}$$

expressing the number of failures per item per unit of time. We note that when we assume that over I_j

- the time at failure has mean $(t_{j-1} + t_j)/2$ and
- the time of censoring has mean $(t_{j-1} + t_j)/2$, too,

then the average amount of item $\times$ time units is

$$\begin{aligned} L_j^* &= w_j \left(n_j - \frac{c_j}{2} - \frac{d_j}{2} \right) \\ &= w_j \left(n_j' - \frac{d_j}{2} \right), \ \text{ see (6.4g).} \end{aligned} \tag{6.18d}$$

With L_j^* inserted in (6.18c) instead of L_j we are back to $\widehat{h}^{(1)}(t_j^*)$ of (6.14c).

The sampling procedure or 'stop rule' for another hazard rate estimator is based on the number of failures observed rather than on the attainment of predetermined time limits, i.e., we now have *fixed numbers of failures per interval* and the boundaries of the intervals are random variables. Based on this sampling procedure SEAL (1954) has found an unbiased estimator of $h(t)$ when there is *no censoring*. The boundaries t_j and the width w_j of the j–th interval now are random and result from the cumulated numbers of failures n_j such that $n_{j+1} - n_j = d_j$ is equal to the preassigned number of failures in I_j. If w_j is small enough so that one may assume

$$h(t_{j-1} + \tau) = h_j \text{ for } \ 0 \le \tau < w_j \tag{6.19a}$$

the estimator given by SEAL reads

$$\widehat{h}_j^{(6)} = \frac{d_j - 1}{\sum\limits_{i=0}^{d_j-1} (n_j - i)\,(t_{j,i+1}^+ - t_{ji}^+)}, \ t_{j-1} \le t < t_j, \tag{6.19b}$$

where t_{ji}^+ is the time of failure of item i in I_j. The denominator in (6.19b) is nothing but the number of time units lived in I_j by those items that failed in I_j. The estimated variance of $\widehat{h}_j^{(6)}$ is

$$\widehat{\text{Var}}\big(\widehat{h}_j^{(6)}\big) \approx \frac{\big[\widehat{h}_j^{(6)}\big]^2}{d_j - 2}. \tag{6.19c}$$

Assumption (6.19a), although suitable for some purposes, is not consistent with most data on mortality and failure intensities which indicate that $h(t)$ usually is a non–decreasing function of t. KIMBALL (1960) has given hazard rate estimators when (6.19a) does not hold, but instead it is assumed that $h(t)$ increases linearly with τ over the interval for which $h(t)$ is being estimated. The estimates and their variances are obtained recursively.

7 Maximum Likelihood Estimation of Monotone Hazard Rates[1]

The direct non–parametric ML estimation of the hazard rate started as early as 1956 by papers of GRENANDER (1956) and KIEFER/WOLFOWITZ (1956). They assumed a continuous distribution with increasing hazard rate and a sample with uncensored data. Later papers generalized their ideas to decreasing as well as to U-shaped hazard rates, to discrete distributions, to censored samples and to the discovery of the statistical properties of the resulting estimators. A general drawback of the ML–estimated hazard rate is its non–smoothness, i.e., the hazard rate is assumed constant between observed failure times. Another drawback is that for U–shaped hazard rates the change point has to be known and that for monotone hazard rates one first has to test whether the distribution is IHR or DHR.[2] We will only discuss the estimation of monotone hazard rates — for U–shaped hazard rates see MYKYTYN/SANTNER (1981) — and we will first present results for complete sample data (Sect. 7.1) and then give results for censored samples (Sect. 7.2).

ML methods are especially apt for processing non–grouped data of smaller sample sizes. When we have a continuous lifetime distribution, the estimates found here subsequently should be smoothed in one way or the other, whereas for a discrete lifetime distribution the estimates found here are definitive.

7.1 The Case of Complete Samples

Since the hazard rate is given by

$$h(x) = \frac{f(x)}{S(x)}$$

and the survival function by

$$S(x) = \exp\left[-\int\limits_0^x h(u)\,du\right]$$

the likelihood function for n independent uncensored lifetimes $x_1 \le x_2 \le \ldots \le x_n$ follows as

$$\begin{aligned} L &= \prod_{i=1}^n f(x_i) = \prod_{i=1}^n h(x_i)\,S(x_i) \\ &= \prod_{i=1}^n h(x_i)\,\exp\left[-\int\limits_0^{x_i} h(u)\,du\right], \end{aligned} \tag{7.1a}$$

so that the log–likelihood is

$$\mathcal{L} = \ln L = \sum_{i=1}^n \ln h(x_i) - \sum_{i=1}^n \int\limits_0^{x_i} h(u)\,du. \tag{7.1b}$$

[1] Suggested reading for this chapter: BARLOW et al. (1972), GRENANDER (1956), KIEFER/WOLFOWITZ (1956), MARSHALL/PROSCHAN (1965), MYKYTYN/SANTNER (1981), PADGETT (1988), PADGETT/WEI (1980), PRAKASA RAO (1970)

[2] Such test will be presented in Sect.10.2.2.

We first consider the case of a *continuous IHR distribution*. It is not possible to obtain a MLE directly by maximizing either L or $\mathcal{L}$, since $(h(x)$ can be arbitrarily large. It follows from argumentation of MARSHALL/PROSCHAN (1965) that

$$\mathcal{L} \leq \sum_{i=1}^{h} \ln h(x_i) - \sum_{i=1}^{n-1} (n-i)\,(x_{i+1}-x_i)\,h(x_i) =: \mathcal{L}^*. \tag{7.1c}$$

The maximization of $\mathcal{L}^*$ subject to $h(x_1) \leq h(x_2) \leq ... \leq h(x_n)$ is performed in GRENANDER (1956). This yields for $h(x_i)$:

$$\widehat{h}(x_i) = \min_{\nu \geq i+1} \max_{\kappa \leq i} \left\{ \frac{\nu - \kappa}{\sum\limits_{j=\kappa}^{\nu-1} (n-j)\,(x_{j+1}-x_j)} \right\};\; i = 1,\, 2,\, ...,\, n-1 \tag{7.1d}$$

and

$$\widehat{h}(x_n) = \infty. \tag{7.1e}$$

For the remaining values of x, $\widehat{h}(x)$ is determined as

$$\widehat{h}(x) = \left\{ \begin{array}{cl} 0 & \text{for } x < x_1 \\ \widehat{h}(x_i) & \text{for } x_i \leq x < x_{i+1};\; i = 1,\, 2,\, ..., n-1 \\ \infty & \text{for } x \geq x_n \end{array} \right\}, \tag{7.1f}$$

so that $\widehat{h}(x)$ is a monotone increasing step–function, see Fig. 7/1. The estimator (7.1d) is consistent and its — not simple looking — asymptotic distribution has been found by PRAKASA RAO (1970). The corresponding estimators of $S(x)$ and $f(x)$ are obtained using $\widehat{h}(x)$:

$$\begin{aligned} \widehat{S}(x) &= \exp\left[-\int_0^x \widehat{h}(u)\,du\right] \\ &= \exp\Big\{-\sum \widehat{h}(x)\big[\min(x,\, x_{i-1}) - x_i\big]\Big\} \end{aligned} \tag{7.1g}$$

where the sum is over i such that $x_i \leq x$, and

$$\widehat{f}(x) = \widehat{h}(x)\,\widehat{S}(x). \tag{7.1h}$$

The resulting curve for $\widehat{S}(x)$ has knees (break points) at the distinct failure times and that for $\widehat{f}(x)$ looks rather strange, see Fig. 7/1.

The estimator in (7.1d) is built on quantities

$$T_i = (n-i)\,(x_{i+1}-x_i);\; i = 0,\, 1,\, ...,\, n-1; \tag{7.2a}$$

where $x_0 = 0$. T_i is nothing but the total–time–on–test spent by the x_i–survivors in the interval $[x_i,\, x_{i+1})$, i.e., between the i–th and the $(i+1)$–st failure times. Another estimator of $h(x)$, called **naive estimator** by some authors, is the reciprocal of T_i :

$$\widetilde{h}(x_i) = \left\{ \begin{array}{cl} \dfrac{1}{(n-i)\,(x_{i+1}-x_i)} & \text{for } x_i \leq x < x_{i+1};\; i = 0,\, 1,\, ...,\, n-1 \\ 0 & \text{for } x \geq x_n. \end{array} \right\} \tag{7.2b}$$

The naive estimator is asymptotically unbiased, but it is not consistent since it has a limiting non-degenerate distribution. Furthermore, since for any n distinct time points x_1, ..., x_n the estimators $\widetilde{h}(x_i)$, ..., $\widetilde{h}(x_n)$ are asymptotically independent, the graph of $\widetilde{h}(x),\ x \geq 0$, will exhibit wild fluctuations, prohibiting its use as an estimator for a monotone hazard rate. For this reason SINGPURWALLA/WONG (1983) have proposed a smoothed version of $\widetilde{h}(x)$, smoothed by kernel methods.

The estimator $\widehat{h}(x_i)$ in (7.1d) can be interpreted as the result of averaging naive estimators until an increasing sequence $\widehat{h}(x_i) \leq \widehat{h}(x_2) \leq ... \leq \widehat{h}(x_n)$ has been found. First, the maximum of $\mathcal{L}^*$ in (7.1c) is found, giving $\widetilde{h}(x_i)$ in (7.2b). If there is a reversal, say $\widetilde{h}(x_i) > \widetilde{h}(x_{i+1})$, then set $h(x_i) = h(x_{i+1})$ in (7.1c) and repeat the procedure. After, at most n steps of this kind, a monotone estimator is obtained. The maximum of $\mathcal{L}^*$ derived with $h(x_i) = h(x_{i+1})$ can be directly obtained by replacing $\widetilde{h}(x_i)$ and $\widetilde{h}(x_{i+1})$ by their **harmonic mean** $\Big\{\big[\widetilde{h}(x_i)^{-1}+\widetilde{h}(x_{i+1})^{-1}\big]/2\Big\}^{-1}$. Succeeding steps amount to further such harmonic averaging which is extended just to the point necessary to eliminate all reversals.

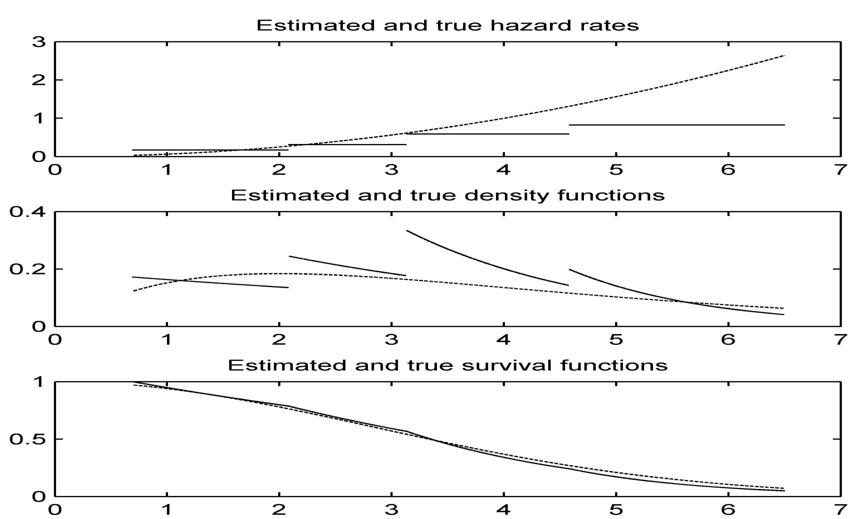
Example 7/1: ML–estimation of an increasing hazard rate of a continuous distribution

The following $n = 10$ observations x_i in Tab. 7/1 have been simulated from a WEIBULL distribution (see Sect. 3.1) with parameters $a = 0,\ b = 4$ and $c = 2$ and thus come from an IHR distribution. Tab. 7/1 gives the ML–estimated hazard rate values according to (7.1d) together with the naive estimates according to (7.2b).

Table 7/1: ML–estimates and naive estimates of an increasing hazard rate

i	$x_{i-1} \leq x$	$< x_i$	$\widehat{h}(x)$	$\widetilde{h}(x)$
1	0	0.69	0	0.1449
2	0.69	1.20	0.1720	0.2179
3	1.20	2.08	0.1720	0.1420
4	2.08	2.21	0.3110	1.0989
5	2.21	3.13	0.3110	0.1811
6	3.13	3.28	0.5894	1.3333
7	3.28	3.72	0.5894	0.5681
8	3.72	4.58	0.5894	0.3876
9	4.58	5.09	0.8230	0.9804
10	5.09	6.50	0.8230	0.7092

Figure 7/1: ML–estimates for a continuous distribution with increasing hazard rate



We now show how to find $\widehat{h}(x)$ from $\widetilde{h}(x)$. There is a first reversal of $\widetilde{h}(x)$ between 0.2179 and 0.1420 ($i = 2$ and $i = 3$), so we replace both values by $\big[(0.2179^{-1} + 0.1420^{-1})/2\big]^{-1} = 0.1720$. The next reversal is between 1.0989 and 0.1811 and both values are replaced by $\big[(1.0981^{-1} + 01811^{-1})/2\big]^{-1} = 0.3130$. The next reversal is between $1.3333,\ 0.5681$ and 0.3876 which are replaced by $\big[(1.3333^{-1} + 0.5681^{-1} + 0.3876^{-1})/3\big]^{-1} = 0.5894$. The last reversal between 0.9804 and 0.7092 is replaced by $\big[(0.9804^{-1} + 0.7092^{-1})/2\big]^{-1} = 0.8230$. Fig. 7/1 depicts the estimated hazard rate, the estimated density function and the estimated survival function (solid lines), each supplemented by their true curves (dashed lines).

We now turn to a *continuous DHR distribution*. Estimation in the DHR case parallels that of the preceding IHR case with some obvious modifications:

1. As the hazard rate is assumed decreasing there is no trivial estimate for $x < x_1$.
2. For the same reason the estimator is defined only for $x \leq x_n$, but it may be extended beyond x_n in any manner that preserves the DHR property.

Thus the estimator in the DHR case reads

$$\widehat{h}(x) = \widehat{h}(x_i) \ \text{ for } \ x_{i-1} < x \leq x_i;\ i = 2,\,3,\,\ldots,\,n; \tag{7.3a}$$

where

$$\widehat{h}(x_i) = \max_{\nu \geq i}\ \min_{\kappa \leq i-1} \left\{ \frac{\nu - \kappa}{\sum\limits_{j=\kappa}^{\nu-1} (n-j)\,(x_{j+1} - x_j)} \right\};\ i = 2,\,3,\,\ldots,\,n. \tag{7.3b}$$

As in the IHR case the estimator (7.3b) results from harmonic averaging the naive estimators until a decreasing sequence has been reached.

Example 7/2: ML–estimation of a decreasing hazard rate of a continuous distribution

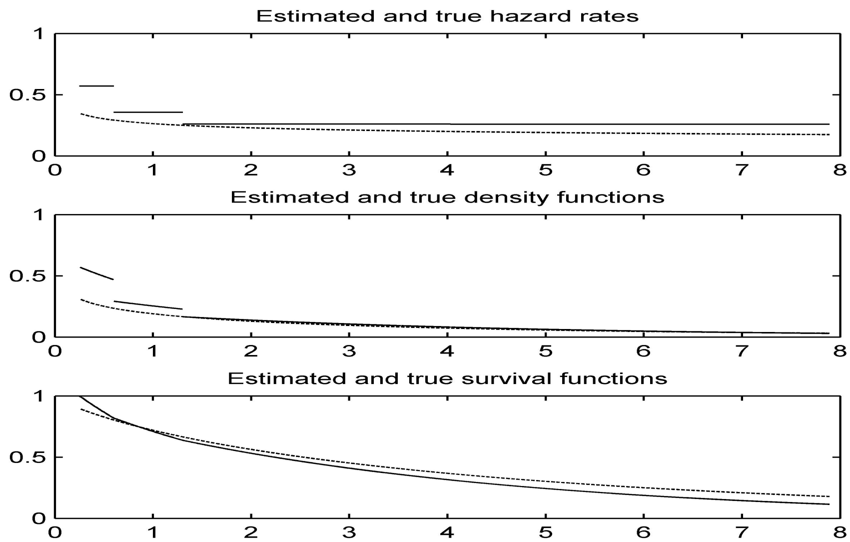
The $n = 6$ observations x_i in Tab. 7/2 come from a WEIBULL distribution with parameters $a = 0,\ b = 4$ and $c = 0.8$, so the sampled population has a decreasing hazard rate. Tab. 7/2 shows the estimated hazard rate $\widehat{h}(x)$ according to (7.3a,b) and the naive estimates. There is only one reversal for $i = 4$ and $i = 5$. Replacing the values 0.1494 and 1.0000 by their harmonic mean yields $\big[(0.1494^{-1} + 1^{-1})/2\big]^{-1} = 0.2600$.

Table 7/1: ML–estimates and naive estimates of a decreasing hazard rate

i	$x_{i-1} \leq x < x_i$		$\widehat{h}(x)$	$\widetilde{h}(x)$
2	0.25	0.60	0.5714	0.5714
3	0.60	1.30	0.3571	0.3571
4	1.30	3.53	0.2600	0.1494
5	3.53	4.03	0.2600	1.0000
6	4.03	7.89	0.2591	0.2591

Fig. 7/2 shows the estimated hazard rate, density function and survival function (solid lines) together with the true curves (dashed lines).

Figure 7/2: ML–estimates for a continuous distribution with decreasing hazard rate



A related problem of interest occurs in the case of a *discrete distribution*. Let $F(\cdot)$ be a discrete distribution with probability mass P_i at x_i, the x_i ordered increasingly. Then, for convenience we encode $x_i =: i;\ i = 1,\, 2,\, \ldots$. The ratio

$$h_i = \frac{P_i}{S_i};\ i = 1,\, 2,\, \ldots \tag{7.4a}$$

with survival function

$$S_i = \Pr(X \geq x_i) = \sum_{j=i}^{\infty} P_j \tag{7.4b}$$

is called discrete hazard rate.[3] Given h_i we find P_i as[4]

$$P_i = h_i \prod_{j=1}^{i-1} (1 - h_j);\ i = 1,\, 2,\, \ldots \tag{7.4c}$$

Let a sample of n independent observations from $F(\cdot)$ consist of k_i occurrences of $x_i = i;\ i = 1,\, 2,\, \ldots,\, m$ where

$$n = \sum_{i=1}^{m} k_i.$$

The log–likelihood function now reads

$$\mathcal{L} = \sum_{i=1}^{m} k_i\, \ln P_i. \tag{7.5a}$$

Inserting (7.4c) we find after some rearrangements[5]

$$\mathcal{L} = \sum_{i=1}^{m} \left[k_i\, \ln h_i + \ln(1 - h_i) \sum_{j=i+1}^{m} k_j \right]. \tag{7.5b}$$

[3] For more on discrete hazard rates see Sect. 1.2.1.

[4] Remember: $\prod\limits_{i=j}^{k} a_i = 1$ for $k < j$.

[5] Remember: $\sum\limits_{i=j}^{k} a_i = 0$ for $k < j$.

The maximization of $\mathcal{L}$ with respect to h_i yields the naive estimators

$$\widetilde{h}_i = \left\{ \begin{array}{cl} 0 & \text{for } i < 1 \\ \dfrac{k_i}{k_i + \ldots + k_m} & \text{for } i = 1, 2, \ldots, m \\ 1 & \text{for } i > m. \end{array} \right\} \tag{7.5c}$$

$\widetilde{h}_i = k_i \big/ (k_i + \ldots + k_m)$ is nothing but the estimator of the conditional probability $P_i \big/ S_i = \Pr(X = x_i \mid X \geq x_i)$ where

$$\widehat{P}_i = \frac{k_i}{n} \text{ and } \widehat{S}_i = \sum_{j=i}^{m} k_j \bigg/ n.$$

In Sect. 1.2.1 we have shown that for a discrete distribution the hazard rate corresponds to this conditional probability, see (1.60c).

An *increasing hazard rate* of a *discrete distribution* is found by maximizing (7.5b) subject to $h_1 \leq h_2 \leq \ldots \leq h_m$. The result is

$$\widehat{h}_i = \min_{i \leq \nu \leq m} \max_{\kappa \leq i} \left\{ \frac{k_\kappa + k_{\kappa+1} + \ldots + k_\nu}{\sum\limits_{j=\nu}^{\kappa} (k_j + \ldots + k_m)} \right\};\ i = 1, 2, \ldots, m. \tag{7.6}$$

Likewise, the $\widehat{h}_i$ of (7.6) are found by averaging through adding numerators and denominators of the naive estimators involved in any reversal.

A *decreasing discrete hazard rate* is estimated as

$$\widehat{h}_i = \max_{i \leq \nu \leq m} \min_{\kappa \leq 1} \left\{ \frac{k_\kappa + k_{\kappa+1} + \ldots + k_\nu}{\sum\limits_{j=\nu}^{\kappa} (k_j + \ldots + k_m)} \right\};\ i = 1, 2, \ldots, m. \tag{7.7}$$

Estimates of the probability mass function and the survival function can be found by evaluating (7.4c) and (7.4b), respectively, with the $\widehat{h}_i$ of (7.6) or (7.7).

Example 7/3: ML–estimation of an increasing hazard rate of a discrete distribution

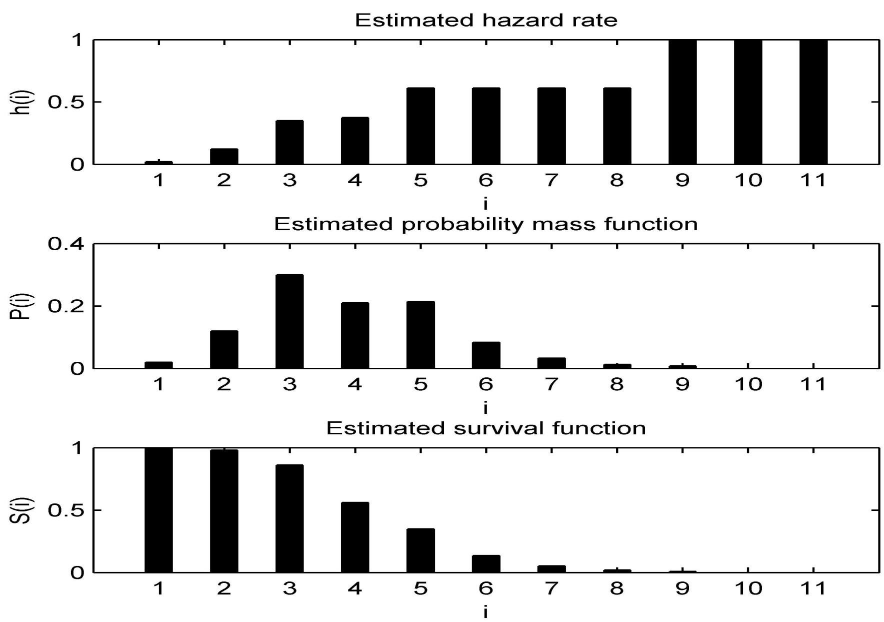
Tab. 7/3 contains the counts k_i of $n = 100$ replicates of a binomial distribution with parameters $N = 10$ and $P = 0.3$. The binomial distribution always has an increasing hazard rate, see Sect. 3.2. In Tab. 7/3 we also give — besides the $\widehat{h}_i$ according to (7.6) — the naive estimates $\widetilde{h}_i$ according to (7.5c).

We have a reversal between $\widetilde{h}_6$ and $\widetilde{h}_7$. The result of averaging through adding the pertinent numerators and denominators is $\widetilde{h}_6^{\text{new}} = \widetilde{h}_7^{\text{new}} = (8 + 1)\big/(12 + 4) = 0.5625$. Now, we have a reversal between $\widetilde{h}_5$, $\widetilde{h}_6^{\text{new}}$, $\widetilde{h}_7^{\text{new}}$ and $\widetilde{h}_8$. Thus, we replace these estimates by $(23+8+1+1)\big/(35+12+4+3) = 0.6111$. Fig. 7/3 displays the estimated hazard rate together with the estimates PMF and survival function.

Table 7/3: ML–estimates and naive estimates of the increasing hazard rate of a discrete distribution

i	k_i	$\widehat{h}_i$	$\widetilde{h}_i$
1	2	0.0200	0.0200
2	12	0.1224	0.1224
3	30	0.3488	0.3488
4	21	0.3750	0.3750
5	23	0.6111	0.6571
6	8	0.6111	0.6667
7	1	0.6111	0.2500
8	1	0.6111	0.3333
9	2	1.0000	1.0000
10	0	1.0000	1.0000
11	0	1.0000	1.0000

Figure 7/3: ML–estimates of an increasing discrete hazard rate



7.2 The case of Randomly Censored Samples

When the observations are singly censored on the right the formulas of the preceding section hold, but they can be evaluated only for the $k < n$ failure times observed, and we will only have estimates $\widehat{h}(x)$ for $x < x_k$.

The methods of this section request samples that are randomly censored on the right. The data set for such samples is presented by the pairs $(y_i,\, \delta_i);\ i = 1,\, 2,\, ...,\, n$. δ_i indicates whether y_i is an uncensored observation ($\delta_i = 1$) or not ($\delta_i = 0$) and $y_i = \min(x_i,\, z_i)$, where x_i is a realization of the interesting lifetime variate X and z_i a realization of some censoring variate Z, independent of X. Using the hazard rate $h(x) = f(x)\big/S(x)$ of X, the likelihood function, see (4.2g),

$$L = \prod_{i=1}^{n} f(y_i)^{\delta_i}\, S(y_i)^{1-\delta_i} \tag{7.8a}$$

turns into

$$L = \prod_{i=1}^{n} h(y_i)^{\delta_i}\, S(y_i). \tag{7.8b}$$

From $S(y) = \exp\Big[-\int\limits_0^y h(u)\,du\Big]$ we finally have the log–likelihood function

$$\mathcal{L} = \sum_{i=1}^{n} \delta_i\, \ln h(y_i) - \sum_{i=1}^{n} \int\limits_0^{y_i} h(u)\,du. \tag{7.8c}$$

Suppose, that $h(x)$ is *increasing* and — without loss of generality — further assume that $y_1 \leq y_2 \leq \ldots \leq y_n$. It follows from (7.1b) that

$$\mathcal{L} \leq \sum_{i=1}^{n} \delta_i\, \ln h(y_i) - \sum_{i=1}^{n-1} (n-i)\,(y_{i+1}-y_i)\,h(y_i) =: \mathcal{L}^*, \tag{7.8d}$$

and the problem of maximizing $\mathcal{L}$ is equivalent to that of maximizing $\mathcal{L}^*$. The following results are due to PADGETT/WEI (1980).

We denote the *distinct* uncensored failure times by $x_1 < x_2 < \ldots < x_k$ and let d_j be the number of uncensored failure times exactly at x_j; $j = 1, 2, \ldots, k$. Also, let c_j denote the number of losses (due to censoring) which occur in the interval $[x_j, x_{j+1})$ for $j = 0, 1, \ldots, k$, where $x_0 = 0$ and $x_{k+1} = \infty$. Furthermore, let the times of the c_j losses be denoted by $\ell_\iota^{(j)}$; $\iota = 1, 2, \ldots, \lambda_j$. The quantities just defined — without $\ell_\iota^{(j)}$ — are illustrated in Fig. 4/1.

Now, for any given increasing $h(x)$, we can define

$$h^*(x) = \begin{cases} 0 & \text{for } x < x_1 \\ h(x_j) & \text{for } x_j \leq x < x_{j+1};\ j = 1, 2, \ldots, k-1 \\ h(x_k) & \text{for } x \geq x_k. \end{cases}$$

Then, for each j we have that

$$\left.\begin{aligned} &\sum_{i=a_j+1}^{b_j} (n-i)(y_{i+1}-y_i)h(y_i) \\ &\qquad \geq \Big[(n-a_j)\big(\ell_1^{(j)}-x_j\big) + (n-a_j-1)\big(\ell_2^{(j)}-\ell_1^{(j)}\big) + \ldots + \\ &\qquad\quad + (n-b_j)\,\big(x_{j+1}-\ell_{c_j}^{(j)}\big)\Big]\,h^*(x_j) \\ &\qquad = \Big[\sum_{\iota=1}^{\lambda_j} \ell_\iota^{(j)} + (n-b_j)\,x_{j+1} - (n-a_j)\,x_j\Big]\,h(x_j), \end{aligned}\right\} \tag{7.9a}$$

where

$$a_j = \sum_{i=0}^{j-1} c_i + \sum_{i=1}^{j} d_i, \tag{7.9b}$$

$$b_j = \sum_{i=0}^{j} c_i + \sum_{i=1}^{j} d_i. \tag{7.9c}$$

Replacing $h\big(\ell_\iota^{(0)}\big)$ by zero for $\iota = 1, 2, \ldots, \lambda_0$, we have that

$$\mathcal{L}^* \leq \sum_{j=1}^{k} d_j \ln h(x_j) - \sum_{j=1}^{k} \alpha_j\, h(x_j) =: \mathcal{L}^{**}, \tag{7.10a}$$

say, where

$$\alpha_j = \left\{ \begin{array}{l} \sum\limits_{\iota=1}^{c_j} \ell_\iota^{(j)} + (n - b_j)\, x_{j+1} - (n - a_j)\, x_j;\ j = 1, 2, \ldots, k-1 \\ \sum\limits_{\iota=1}^{c_k} \ell_\iota^{(k)} - c_k\, x_k;\ j = k. \end{array} \right\} \tag{7.10b}$$

Since $h^*(x)$ is increasing, it follows that the maximization of $\mathcal{L}$ is equivalent to that of $\mathcal{L}^{**}$. Note that only α_k can be zero and this happens when there are no censored observations strictly larger than x_k, the largest uncensored lifetime observed. The problem of obtaining an estimator of $h(x)$ subject to its increasing is reduced to that of maximizing $\mathcal{L}^{**}$ subject to the constraint $h(x_1) \le h(x_2) \le \ldots \le h(x_k)$.

In maximizing $\mathcal{L}^{**}$ we have to distinguish two cases.

1. The last observation y_n is uncensored so that $\alpha_k = 0$. In this case $\mathcal{L}^{**}$ is unbounded, and it is not possible to find MLEs of $h(x)$ directly from $\mathcal{L}^{**}$. Following the argumentation of MARSHALL/PROSCHAN (1965) we estimate $h(x)$ by

$$\widehat{h}(x) = \left\{ \begin{array}{ll} 0 & \text{for } x < x_1, \\ \widehat{h}(x_j) & \text{for } x_j \le x < x_{j+1};\ j = 1, 2, \ldots, k-1; \\ \widehat{h}(x_k) & \text{for } x \ge x_k, \end{array} \right\} \tag{7.11a}$$

   where

$$\left. \begin{array}{rcl} \widehat{h}(x_j) & = & \min\limits_{j \le \nu \le k-1}\ \max\limits_{1 \le \kappa \le j} \left[\dfrac{\sum\limits_{\mu=\kappa}^{\nu} d_\mu}{\sum\limits_{\mu=\kappa}^{\nu} \alpha_\mu} \right];\ j = 1, \ldots, k-1; \\ \widehat{h}(x_k) & = & \infty. \end{array} \right\} \tag{7.11b}$$

   $\widehat{h}(x)$ truly is not the MLE of $h(x)$, but can be considered as the limit of a sequence of MLEs in the sense of MARSHALL/PROSCHAN (1965).

2. The last observation y_n is uncensored so that $\alpha_k \ne 0$. In this case the MLE of $h(x)$ is given by (7.11a) with

$$\widehat{h}(x_j) = \min\limits_{j \le \nu \le k}\ \max\limits_{1 \le \kappa \le j} \left[\dfrac{\sum\limits_{\mu=\kappa}^{\nu} d_\mu}{\sum\limits_{\mu=\kappa}^{\nu} \alpha_\mu} \right];\ j = 1, \ldots, k. \tag{7.11c}$$

The ML–estimation of a *decreasing* hazard rate with sample data randomly censored on the right follows along the same lines as above. But there are some evident minor modifications which we already encountered in Sect. 7.1 for the change–over from the IHR case to the DHR case.

- As $h(x)$ is now assumed decreasing there is no trivial estimate of $h(x)$ for $x < x_1$.
- For the same reason the estimate is defined only for $x < x_k$, but it may be extended beyond x_k in any manner that preserves the DHR property.

So we have

$$\widehat{h}(x) = \widehat{h}(x_j) \ \text{ for } \ x_j \le x < x_{j+1};\ j = 1, 2, \ldots, k-1;$$

with

$$\widehat{h}(x_j) = \max_{j \le \nu \le k-1} \min_{1 \le \le j} \left[\frac{\sum\limits_{\mu=\kappa}^{\nu} d_\mu}{\sum\limits_{\mu=\kappa}^{\nu} \alpha_\mu} \right]; \; j = 1, ..., k-1. \tag{7.12}$$

Formulas (7.11a-c) and (7.12) do not only process multiply censored data sets, where the censored and uncensored observations are mixed, but they also cope with complete data sets as well as with data sets which are singly censored whether of type–I or of type–II. For these situations the input has to be organized properly. First of all the observed times y_i together with their indicators δ_i have to be in ascending order with respect to y.

- When there are no censored observations within the sample of size n the input has to look like:

$$\left\{ \begin{array}{ccccc} y_1 \le & y_2 \le & ... \le & y_{n-1} \le & y_n \\ 1 & 1 & ... & 1 & 1 \end{array} \right\}.$$

- When the sample of size n is singly censored of type–I with censoring time $y_\ell,\ \ell \le n$, the input has to be

$$\left\{ \begin{array}{cccccccc} y_1 \le & y_2 \le & ... \le & y_{\ell-1} \le & y_\ell = & y_{\ell+1} = & ... = & y_n \\ 1 & 1 & ... & 1 & 0 & 0 & ... & 0 \end{array} \right\}.$$

- When the sample of size n is singly censored of type–II with censoring at the k–th failure the input should read

$$\left\{ \begin{array}{cccccccc} y_1 \le & y_2 \le & ... \le & y_{k-1} \le & y_k = & x_{k+1} = & ... = & y_n \\ 1 & 1 & ... & 1 & 1 & 0 & ... & 0 \end{array} \right\}.$$

Example 7/4: ML–estimation of an increasing hazard rate with multiply censored data

The following observations in Tab. 7/4 have been taken from KAPLAN/MEIER (1958, p. 464). The hazard rate has been found by (7.11a,c) because the last observation y_8 is censored.

Table 7/4: Hazard rate estimate for KAPLAN/MEIER's data set

i	y_i	δ_i	j	$x_{j-1} \le x <$	x_j	$\widehat{h}(x)$
1	0.8	1	1	$-\infty$	0.8	0.0000
2	1.0	0				
3	2.7	0				
4	3.1	1	2	0.8	3.1	0.0375
5	5.4	1	3	3.1	5.4	0.1087
6	7.0	0				
7	9.2	1	4	5.4	9.2	0.1087
8	12.1	0	5	9.2	∞	0.3448

8 Smooth Hazard Rate Estimators

In the preceding chapters we have presented several estimation approaches that only led to pointwise or to non–smooth estimates of the hazard rate. When lifetime is a continuous variate with a continuous hazard rate we, of course, like to have a continuous estimate for its whole course or at least for the course between the shortest and longest lifetime observed. So, we have to look for an appropriate method to smooth or to graduate discrete hazard rate estimates. A lot of smoothing or graduating techniques have been developed in mathematics and in statistics, e.g., moving averages, least squares, splines, orthogonal series, wavelets, or kernels, however an accepted standard solution to hazard rate smoothing does not exist. It seems that the kernel technique prevails because it rests upon an intuitive idea, has mostly known mathematical and statistical properties and is relatively easy to implement. The kernel smoothing approach has been most thoroughly developed and has an extensive literature. For these reasons the focus of this chapter is on kernel smoothing, but we also present other techniques in the final sections of this chapter.

8.1 Kernel Smoothing[1]

A kernel estimator is a convolution of a smooth function and a rough empirical function estimator chosen in such a way as to produce a smooth functional estimator. The underlying idea is to take advantage of the fact that this linear functional transfers continuity properties from the smooth function, the so–called **kernel**,[2] to the final estimator. Although potentially useful in a variety of settings, kernel methods have been principally exploited in four settings:

- probability density estimation,
- hazard rate estimation, which is very closely related to PDF estimation,
- spectral density estimation, and
- non–parametric regression.

In this Section 8.1 we will only present PDF and HR kernel smoothing for non–grouped data.

8.1.1 Motivation and Basic Concepts

Kernel estimators for the PDF and the HR have the same structure, similar formulas, and share the same set of problems and nearly the same set of approaches to solve these problems. We start this section by looking at PDF smoothers which have been the first field of kernel estimation and which started by papers of ROSENBLATT (1956) and PARZEN (1962), whereas HR estimation by kernels started a little bit later with papers of WATSON/LEADBETTER (1964a,b).

[1] Suggested reading for this section: IZENMAN (1991), PRAKASA RAO (1983), WAND/JONES (1995).

[2] CACOULOS (1966) appears to be the first to call this smoothing function a kernel function. Previously it was referred to as a **weight function** or as a **window**.

8.1.1.1 The Convolution Formula

The most simple **kernel estimator of a PDF** for an uncensored set of distinct and ordered observations $x_1 < x_2 < ... < x_n$ is given by

$$\widehat{f}_n(x) = \frac{1}{n\,b}\sum_{i=1}^{n} K\left(\frac{x-x_i}{b}\right). \tag{8.1}$$

The idea of this estimator is the following: The empirical distribution function $\widehat{F}_n(x) = i/n$, where i is the number of sample observations less or equal to x, or the empirical survival function $\widehat{S}_n(x) = (n-i)/n$ are discrete discrete functions each placing mass $1/n$ at each of the observations x_i, giving the rough empirical function. By formula (8.1) this probability mass is smeared out continuously, smearing according to the choice of the kernel. The kernel is a smooth function[3] that determines the pattern of how the mass $1/n$ is redistributed around the the observation x_i, and b, the **bandwidth** or **window width**,[4] is responsible for 'how far' the kernel stretches out to either side of x_i when the kernel is symmetric. Stated in another way, one may say that all observations x_i that are within a distance b on either side of a given point x contribute to the density estimate at this point x.

(8.1) can be motivated by generalizing the **sliding histogram**:

$$\frac{\widehat{F}_n(x+b) - \widehat{F}_n(x-b)}{2\,b} = \int\limits_{x-b}^{x+b} \frac{1}{2\,b}\, d\widehat{F}_n(u) \tag{8.2a}$$

where $\mathrm{d}\widehat{F}_n(\cdot)$ is the empirical measure. (8.2a) is a special case of (8.1) when $\mathrm{d}\widehat{F}_n(\cdot) = 1/n$ and

$$K\left(\frac{x-x_i}{b}\right) = \left\{\begin{array}{cl} 1/2 & \text{for } |x-x_i| \le b \\ 0 & \text{else.} \end{array}\right\} \tag{8.2b}$$

The kernel (8.2b) is known as **uniform kernel** or **rectangular kernel**.

Generally, a kernel estimator of the PDF smoothes the increments of the empirical distribution function $\widehat{F}_n(\cdot)$ which are also the decrements of the empirical survival function $\widehat{S}_n(\cdot)$. It is more common to write the kernel PDF estimator by means of the empirical survival function because — using the KAPLAN/MEIER estimator of Sect. 5.1 — we have a compact notation which covers both, censored as well as uncensored data sets.

When there are *no tied observations*[5] we denote any observed lifetime — whether uncensored or not — by y_i which are given in ascending order. $\delta_i = 1$ indicates a failure time (uncensored lifetime) and $\delta_i = 0$ a censored observation. The data set thus consists of n pairs $(y_i,\, \delta_i)$. The KME of $S(x)$ reads, see (5.3d):

$$\widehat{S}_n(x) = \left\{\begin{array}{cl} 1 & \text{for } x < x_1 \\ \prod\limits_{i\,:\,y_i \le x}\left(\dfrac{n-i}{n-i-1}\right)^{\delta_i} & \text{else,} \end{array}\right\} \tag{8.3a}$$

[3] More on the properties and types of kernels is found in Sect. 8.1.1.3.

[4] More on bandwidths is found in Sect. 8.1.1.4.

[5] Censored observations may be tied among themselves or with uncensored data. In the latter case censored lifetimes are moved a little amount to the right of the uncensored lifetime so that censoring is assumed to happen later.

where $x_1 = \min_i(y_i,\, 1)$, i.e., x_1 is the shortest uncensored lifetime observed. Let

$$ {}_n\Delta_i = \widehat{S}_n(y_{i-1}) - \widehat{S}_n(y_i);\; i = 1,\, 2,\, ...,\, n;\; \widehat{S}_n(y_0) = 1; \tag{8.3b} $$

be the magnitude of the discontinuity of $\widehat{S}_n(\cdot)$ at y_i with

$$ {}_n\Delta_i \left\{ \begin{array}{ccl} = & 0 & \text{for } (y_i,\, 0), \\ > & 0 & \text{for } (y_i,\, 1). \end{array} \right\} \tag{8.3c} $$

The PDF kernel estimator is then given by

$$ \widehat{f}_n(x) = \frac{1}{b} \sum_{i=1}^{n} {}_n\Delta_i\, K\left(\frac{x - y_i}{b}\right). \tag{8.3d} $$

In the special case where all observations are uncensored ($\delta_i = 1 \;\forall\; i$), $\widehat{S}_n(x)$ is a staircase with n steps and $1/n$ as constant step height and (8.3d) reduces to (8.1). When there are censored observations we have fewer steps and the step heights are greater than $1/n$.

In the case of *tied observations*, $d_i \geq 1$ being the size of the tie at x_i with at least one $d_i > 1$, we have $k < n$ distinct uncensored lifetimes x_i. In the interval $[x_i,\, x_{i+1})$; $i = 0,\, 1,\, ...,\, k$; between two uncensored lifetimes, where $x_0 = 0$ and $x_{k+1} = \infty$, there may be $c_i,\; c_i \geq 0$, censored lifetimes. The number of sample units at risk just before x_i is

$$ n_i = n_{i-1} - c_{i-1} - d_{i-1};\; i = 1,\, 2,\, ...,\, k; $$

where

$$ n_0 = n \text{ and } d_0 = 0. $$

For an illustration of these quantities see Fig. 4/1. The KME of the survival function in this case of tied uncensored lifetimes, see (5.3a), reads

$$ \widehat{S}_n(x) = \prod_{i\,:\,x_i \leq x} \left(1 - \frac{d_i}{n_i}\right);\; i = 1,\, 2,\, ...,\, k; \tag{8.4a} $$

with step height

$$ {}_n\Delta_i = \widehat{S}_n(x_{i-1}) - \widehat{S}_n(x_i);\; i = 1,\, 2,\, ...,\, k;\; \widehat{S}_n(x_0) = 1. \tag{8.4b} $$

With respect to the step height we can state

$$ {}_n\Delta_i \left\{ \begin{array}{ccl} = & d_i/n & \text{when the sample has no censored observations at all} \\ \geq & d_i/n & \text{when there are censored observations somewhere in the sample..} \end{array} \right\} \tag{8.4c} $$

The formula of the PDF kernel estimator again is (8.3d), but the summation now goes from $i = 1$ to $i = k$ and y_i is replaced by x_i :

$$ \widehat{f}_n(x) = \frac{1}{b} \sum_{i=1}^{k} {}_n\Delta_i\, K\left(\frac{x - x_i}{b}\right). \tag{8.5} $$

We know turn to the **kernel estimator of the hazard rate**. The rough empirical estimator to be convoluted with a smooth kernel is given by the increments of an estimated CHR. In Sect. 5.2 we have found two such estimators, the indirect or natural estimator, see (5.10), and the direct estimator, known as NELSON/AALEN estimator, see (5.12a), which is nothing but the cumulation of the empirical hazard rate values. In kernel estimation the NELSON/AALEN estimator is preferred over the indirect estimator because it avoids taking logarithms and has a smaller variance, compare (5.11c) to (5.12b). So we have when there are *no tied observations*

$$\widehat{H}(x) = \left\{ \begin{array}{cl} 0 & \text{for } x < x_1 = \min_i(y_i,\, 1) \\ \sum\limits_{i\,:\,y_i \le x} \dfrac{\delta_i}{n-i-1} & \text{else,} \end{array} \right\} \tag{8.6a}$$

with step height

$$_nD_i = \widehat{H}_n(y_i) - \widehat{H}_n(y_{i-1});\ i = 1,\, 2,\, ...,\, n;\ \widehat{H}_n(y_0) = 0. \tag{8.6b}$$

and

$$_nD_i = \left\{ \begin{array}{cl} 0 & \text{for } (y_i,\, 0) \\ \dfrac{1}{n-i+1} & \text{for } (y_i,\, 1). \end{array} \right\} \tag{8.6c}$$

For *tied observations* we have

$$\widehat{H}(x) = \left\{ \begin{array}{cl} 0 & \text{for } x < x_1 = \min_i(y_i,\, 1) \\ \sum\limits_{i\,:\,x_i \le x} \dfrac{d_i}{n_i} & \text{else,} \end{array} \right\} \tag{8.7a}$$

with step height

$$_nD_i = \widehat{H}_n(x_i) - \widehat{H}_n(x_{i-1});\ i = 1,\, 2,\, ...,\, k;\ \widehat{H}_n(x_0) = 0 \tag{8.7b}$$

and

$$_nD_i \left\{ \begin{array}{cll} = & d_i/n & \text{when there is no censoring in the sample} \\ \ge & d_i/n & \text{with censoring somewhere in the sample.} \end{array} \right\} \tag{8.7c}$$

The HR kernel estimator is then given by

$$\widehat{h}_n(x) = \left\{ \begin{array}{ll} \dfrac{1}{b} \sum\limits_{i=1}^{n} {}_nD_i\, K\!\left(\dfrac{x-y_i}{b}\right) & \text{for a sample having no ties} \\ \dfrac{1}{b} \sum\limits_{i=1}^{k} {}_nD_i\, K\!\left(\dfrac{x-x_i}{b}\right) & \text{for a sample with tied observations.} \end{array} \right\} \tag{8.8}$$

8.1.1.2 Performance of Kernel Smoothing

Like any statistical procedure, kernel estimators are recommended only if they possess desirable properties. These properties depend on — besides the sample size — the chosen kernel and the chosen size of the bandwidth, but the greatest influence comes from the bandwidth. Finite–sample properties are available for special situations, but, in general, research emphasis is and has been on large–sample properties.

Let $\psi(x)$ denote the continuous curve to be estimated by kernel techniques, e.g.,

- the PDF $f(x)$ in density estimation,
- the HR $h(x)$ in hazard rate estimation,

- the regression function in regression estimation, or
- the spectral density in spectral estimation,

and let $\widehat{\psi}(x)$ denote its kernel estimator.

Consider, for example, **unbiasedness**. The estimator $\widehat{\psi}(x)$ is unbiased for $\psi(x)$ if, for all $x \in \mathbb{R}$, which — without loss of general validity — is assumed to be the domain of $\psi(x)$, we have $\mathrm{E}\big[\widehat{\psi}(x)\big] = \psi(x)$. Unbiasedness seldom exists for kernel estimators, hence attention has focused on sequences $\big\{\widehat{\psi}_n(x)\big\}$ of kernel estimators that are **asymptotically unbiased** for $\psi(x)$, that is, for all $x \in \mathbb{R}$, $\mathrm{E}\big[\widehat{\psi}_n(x)\big] \;\to\; \psi(x)$ as $n \;\to\; \infty$.

A more important property is consistency. As $n \;\to\; \infty$, $\widehat{\psi}(x)$ is **weakly pointwise consistent** for $\psi(x)$ if $\widehat{\psi}(x) \;\to\; \psi(x)$ in probability for every $x \in \mathbb{R}$, and it is **strongly consistent** if convergence holds almost surely. Other types of consistency depend on the error criterion chosen, i.e., on the distance function (= metric) used, either the L_1–norm or the L_2–norm, but L_2–approaches are dominant as being more tractable than L_1–approaches.

We first look at L_2–approaches, also known as squared–error criteria. If $\psi(x)$ is assumed square–integrable, then the performance of $\widehat{\psi}(x)$ at $x \in \mathbb{R}$ is measured by the **mean squared error**

$$\begin{aligned}
\mathrm{MSE}\big[\widehat{\psi}(x)\big] &= \mathrm{E}\big[\widehat{\psi}(x) - \psi(x)\big]^2 \\
&= \mathrm{Var}\big[\widehat{\psi}(x)\big] + \Big\{\mathrm{Bias}\big[\widehat{\psi}(x)\big]\Big\}^2 \qquad (8.9a)
\end{aligned}$$

where the expectation is with respect to random sampling and

$$\mathrm{Var}\big[\widehat{\psi}(x)\big] = \mathrm{E}\Big\{\widehat{\psi}(x) - \mathrm{E}\big[\psi(x)\big]\Big\}^2 \qquad (8.9b)$$

$$\mathrm{Bias}\big[\widehat{\psi}(x)\big] = \mathrm{E}\big[\widehat{\psi}(x)\big] - \psi(x). \qquad (8.9c)$$

If $\mathrm{MSE}\big[\widehat{\psi}(x)\big] \;\to\; 0$ for all $x \in \mathbb{R}$ as $n \;\to\; \infty$, then $\widehat{\psi}(x)$ is said to be **pointwise consistent in quadratic mean**. More important is to measure how well the entire curve $\widehat{\psi}(x)$ estimates $\psi(x)$. One such measure of goodness of fit is found by integrating $\mathrm{MSE}\big[\widehat{\psi}(x)\big]$ over all values of x, yielding the **integrated mean squared error**[6]

$$\mathrm{IMSE}\big[\widehat{\psi}(\cdot)\big] = \int \mathrm{E}\big[\widehat{\psi}(x) - \psi(x)\big]^2\, dx. \qquad (8.10)$$

Another measure is the **integrated squared error**

$$\mathrm{ISE}\big[\widehat{\psi}(\cdot)\big] = \int \big[\widehat{\psi}(x) - \psi(x)\big]^2\, dx. \qquad (8.11a)$$

Taking expectation over $\widehat{\psi}(\cdot)$ in (8.11a) gives the **mean integrated squared error**

$$\mathrm{MISE}\big[\widehat{\psi}(\cdot)\big] = \mathrm{E}\big\{\mathrm{ISE}\big[\widehat{\psi}(\cdot)\big]\big\}. \qquad (8.11b)$$

Note that $\mathrm{MISE}\big[\widehat{\psi}(\cdot)\big] = \mathrm{IMSE}\big[\widehat{\psi}(\cdot)\big]$. $\mathrm{ISE}\big[\widehat{\psi}(\cdot)\big]$ is often preferred as a criterion rather than its mean MISE, since ISE determines how closely $\widehat{\psi}(\cdot)$ approximates $\psi(\cdot)$ for a given data set, whereas MISE is concerned with the average over all possible data sets. On the other hand asymptotic MSE and MISE are more often used to find the optimal kernel and the optimal bandwidth as will be shown below.

[6] At all times, an unqualified integral sign $\int$ will be taken to mean integration over $\mathbb{R}$.

When L_2–approaches are used in PDF kernel estimation the tail behavior of the density becomes less important, possibly resulting in peculiarities in the tails of the density estimates. For this and other reasons some authors prefer L_1–approaches like the **integrated absolute error**

$$\text{IAE}\big[\widehat{\psi}(\cdot)\big] = \int \big|\widehat{\psi}(x) - \psi(x)\big|\, dx \tag{8.12a}$$

which is invariant under monotone transformations with $0 \leq \text{IAE} \leq 2$ for $\psi(x) = f(x)$. The expectation of (8.12a) over all $\widehat{\psi}(\cdot)$ yields the **mean integrated absolute error**

$$\text{MIAE}\big[\widehat{\psi}(\cdot)\big] = \text{E}\big\{\text{IAE}\big[\widehat{\psi}(\cdot)\big]\big\}. \tag{8.12b}$$

The labor needed to get L_1–results is more difficult than that needed to obtain analogous L_2–results. It should be realized that the MIAE and the MISE do not necessarily conform to the human perception of closeness of a curve estimate to its target.

We now take a closer look at the L_2–criteria MSE and MISE when a *PDF is to be estimated*. These results for $f(x)$ are needed in Sect. 8.1.2 on indirect hazard rate smoothing which is based on $\widehat{f}_n(x)$. Furthermore, these results can — more or less easily — be transferred to and generalised for direct hazard rate smoothing, i.e., smoothing the increments of the empirical cumulative hazard rate as given in (8.8). To keep things easy we assume an uncensored sample with untied observations so that the estimator to be studied reads

$$\widehat{f}_n(x) = \frac{1}{n\, b_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b_n}\right). \tag{8.13}$$

where X_i is the i-th ordered lifetime to be observed in the sample.

We will make the following assumptions concerning the PDF, the kernel and the bandwidth:

(i) $f(x)$ is such that its second derivative $f''(x)$, which measures the curvature of $f(x)$, is continuous, square integrable and ultimately monotone.[7]

(ii) The bandwidth $b := b_n$ is a non–random sequence of positive numbers, where the dependence on the sample size n will be suppressed in the following formulas in order to keep the notation as lean as possible, but we assume

$$\lim_{n\to\infty} b_n \;\to\; 0 \;\text{ and }\; \lim_{n\to\infty} (n\, b_n) \;\to\; \infty,$$

i.e., b_n approaches zero, but at a rate slower than n^{-1}.

(iii) The kernel is a bounded PDF which is symmetric about X_i and has a finite second moment about the origin.

We first look at the *expectation of (8.13)* at a given $x \in \mathbb{R}$. The X_i's in (8.13) are iid (= independently and identically distributed) variables with the same PDF as given by the target function $f(\cdot)$. So we have

$$\begin{aligned}
\text{E}\big[\widehat{f}_n(x;b)\big] &= \text{E}\left[\frac{1}{n\,b}\sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right)\right] \\
&= \frac{1}{n\,b}\int \sum_{i=1}^{n} K\left(\frac{x - v}{b}\right) f(v)\, dv \\
&= \frac{1}{b}\int K\left(\frac{x - v}{b}\right) f(v)\, dv.
\end{aligned}$$

[7] An ultimately monotone function is one that is monotone over both $(-\infty,\, x^*)$ and $(x^*,\, \infty)$ for some $x^* > 0$.

Introducing a new variable

$$z = \frac{x - v}{b}$$

and observing the symmetry of $K(\cdot)$ around $z = 0$ we arrive at

$$\mathrm{E}\big[\widehat{f}_n(x;b)\big] = \int K(z)\, f(x - b\, z)\, dz. \tag{8.14a}$$

Expanding $f(x - b\, z)$ in a TAYLOR series around x gives

$$f(x - b\, z) = f(x) - b\, z\, f'(x) + \frac{1}{2}\, b^2\, z^2\, f''(x) + o(b^2) \tag{8.14b}$$

uniformly in z, and a remainder, which approaches zero more rapidly than b^2 goes to zero with $n \to \infty$. With (8.14b) we have for (8.14a):

$$\mathrm{E}\big[\widehat{f}_n(x;b)\big] = f(x) + \frac{1}{2}\, b^2\, f''(x) \int z^2\, K(z)\, dz + o(b^2) \tag{8.14c}$$

because of assumption (iii) above

$$\int K(z)\, dz = 1, \quad \int z\, K(z) dz = 0, \quad \int z^2\, K(z)\, dz < \infty.$$

For the second moment about zero of the kernel we write

$$\mu_2(K) := \int z^2\, K(z)\, dz \tag{8.15}$$

and thus find the bias of $\widehat{f}_n(x, b)$ as

$$\begin{aligned}\mathrm{Bias}\big[\widehat{f}_n(x;b)\big] &= \mathrm{E}\big[\widehat{f}_n(x,b)\big] - f(x)\\ &= \frac{1}{2}\, b^2\, f''(x)\, \mu_2(K) + o(b^2).\end{aligned} \tag{8.16}$$

A closer look at (8.16) reveals that the bias

- is of order $o(b^2)$ implying that $\widehat{f}_n(x, b)$ is asymptotically unbiased,
- is proportional to the squared given bandwidth b,
- is proportional the variance of the kernel,[8]
- depends on the second derivative of the target function $f(x)$ and is zero when there is no curvature at x and is the higher the greater $|f''(x)|$.

We now look at the *variance of (8.13)* which — after some manipulations like those for finding the expectation — reads

$$\begin{aligned}\mathrm{Var}\big[\widehat{f}_n(x;b)\big] &= \frac{1}{n\, b} \int K(z)^2\, f(x - b\, z)\, dz - \frac{1}{n} \Big\{\mathrm{E}\big[\widehat{f}_n(x,b)\big]\Big\}^2\\ &= \frac{1}{n\, b} \int K(z)^2\, \big[f(x) + o(1)\big]\, dz - \frac{1}{n}\big[f(x) + o(1)\big]^2\\ &= \frac{1}{n\, b}\, f(x) \int K(z)^2\, dz + \big[o(n\, b)^{-1}\big]\\ &= \frac{1}{n\, b}\, f(x)\, R(K) + \big[o(n\, b)^{-1}\big]\end{aligned} \tag{8.17a}$$

[8] Remember that — because of $\mu_1(K) = \int z\, K(z)\, dz = 0$ — $\mu_2(K)$ is equal to the variance. Tab. 8/1 shows this variance for different kernels.

where we have introduced[9]

$$R(K) := \int K(z)^2\, dz, \tag{8.17b}$$

and in general

$$R(\phi) := \int \phi(u)^2 du \tag{8.18}$$

for any square–integrable function $\phi(\cdot)$. Since the variance is of order $(n\,b)^{-1}$ assumption (ii) above assures that $\operatorname{Var}\big[\widehat{f}_n(x;b)\big]$ converges to zero.

Adding (8.17a) and the square of (8.16) gives the mean squared error of $\widehat{f}_n(x;b)$:

$$\operatorname{MSE}\big[\widehat{f}_n(x;b)\big] = \frac{1}{n\,b}\, f(x)\, R(K) + \frac{1}{4}\, b^4\, f''(x)^2\, \mu_2(K)^2 + o\big[(n\,b)^{-1} + b^4\big]. \tag{8.19}$$

Integrating (8.19) under the integrability assumptions in (i) above we obtain

$$\operatorname{MISE}\big[\widehat{f}_n(\cdot;b)\big] = \frac{1}{n\,b}\, R(K) + \frac{1}{4}\, b^4\, \mu_2(K)^2\, R(f'') + o\big[(n\,b)^{-1} + b^4\big]. \tag{8.20}$$

The first two terms on the right–hand side constitute AMISE, the **asymptotic mean integrated squared error**:

$$\operatorname{AMISE}\big[\widehat{f}_n(\cdot;b)\big] = \frac{1}{n\,b}\, R(K) + \frac{1}{4}\, b^4\, \mu_2(K)^2\, R(f''), \tag{8.21}$$

which is a large–sample–approximation to the MISE. AMISE is — besides n — influenced by

- the bandwidth b,
- the kernel K and
- the target function $f(\cdot)$ via its curvature $f''(\cdot)$.

We see that the second term of AMISE, the integrated squared bias, is proportional to b^4, so for this term to decrease one needs to take b to be small. However, taking b small means an increase in the leading factor of the first term, the integrated variance, which is proportional to $(n\,b)^{-1}$. Therefore, as n increases b should vary in such a way that each of the two terms of AMISE becomes smaller. This is known as the **variance–bias trade–off** and is in accordance with the intuitive role of b demonstrated in Fig. 8/1 below. For very small b, $\widehat{f}_n(.;b)$ is very spiky and hence very variable in the sense that, over repeated sampling from $f(\cdot)$ the spikes wold appear in different places. There is, however, very little bias.

(8.21) lends to find the **optimal bandwidth** with respect to this criterion. The bandwidth minimizing AMISE can be given in closed form as

$$b_{\text{AMISE}} = \left[\frac{R(K)}{n\, R(f'')\, \mu_2(K)^2}\right]^{1/5}. \tag{8.22}$$

Aside from its dependence on the known $R(K)$, $\mu_2(K)$ and n, (8.22) shows that b_{AMISE} is inversely proportional to the unknown $R(f'')^{1/5}$. The functional $R(f'')^{1/5} = \int f''(x)^2\, dx$ measures the total curvature of $f(\cdot)$. Thus, for a PDF with little curvature, $R(f'')$ will be small and a large bandwidth is called for, on the other hand, when $R(f'')$ is large, little smoothing with a smaller bandwidth will be optimal. Unfortunately, direct use of (8.22) to choose a good bandwidth in practice is impossible since $R(f'')$ is not known. Some proposals for estimating $R(f'')$ and then selecting b will be presented in Sect. 8.1.1.4.

[9] Tab. 8/1 shows $R(K)$ for different kernels.

Inserting (8.22) into (8.21) leads to the smallest possible AMISE of $\widehat{f}_n(\cdot;b)$ using a given kernel K :

$$\inf_{b>0} AMISE\big[\widehat{f}_n(\cdot;b)\big] = \frac{5}{4}\Big[\mu_2(K)^2\,R(K)^4\,R(f'')\Big]^{1/5} n^{-4/5}. \tag{8.23}$$

This expression gives the **rate of convergence** of the minimum AMISE to zero as $n \to \infty$. Under the stated assumptions, the best obtainable rate is of order $n^{-4/5}$. This rate is slower than the typical parametric rate of order n^{-1}, e.g., $\widehat{E(X)} = \overline{X}$ with $\text{Var}\big(\overline{X}\big) = \text{Var}(X)/n$. To arrive at a higher order of convergence one has to choose special kernels, the so–called higher–order kernels, see Sect. 8.1.1.3.

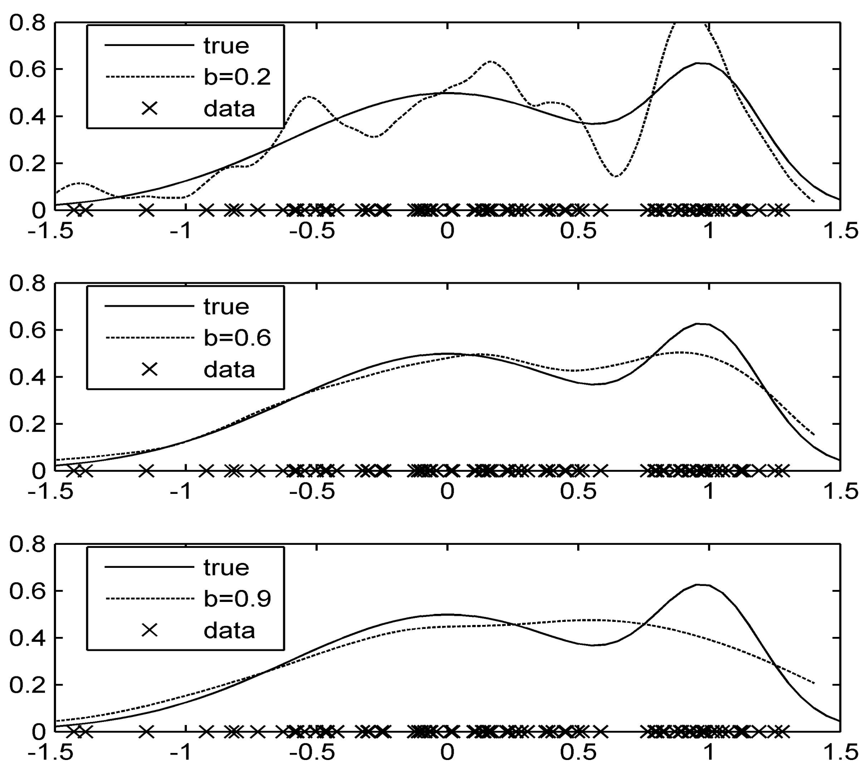
Example 8/1: Effect of varying bandwidth on smoothed PDF estimates

We have randomly generated $n = 80$ realizations of the following mixture of two normal distributions:

$$0.75 \cdot No(\mu = 0,\, \sigma = 0.6) + 0.25 \cdot No(\mu = 1,\, \sigma = 0.2).$$

Fig. 8/1 displays the true density as a solid line. Using the biweight kernel, see Tab. 8/1, with three different bandwidths we have smoothed the data. Smoothing with $b = 0.2$ gives a very rugged curve because the kernel is very narrow and the averaging process only covers relatively few observation. This estimate pays too much attention to the particular data set at hand and does not allow for the variation across the sample and thus is **undersmoothed**. Using $b = 0.9$ results in a much smoother estimate of $f(\cdot)$ which is really too smooth since the true bimodality has been smoothed away, so this an **oversmoothed** estimate. The graph in the middle of Fig. 8/1 is a compromise with $b = 0.6$. This kernel estimate is not overly noisy and the structure of the true density, i.e., its bimodality, has been recovered.

Figure 8/1: Biweight–kernel smoothing with different bandwidths



8.1.1.3 Kernel Selection

The simplest class of kernels consists of symmetric[10] PDFs satisfying

(1) $K(u) \geq 0 \; \forall \, u \in \mathbb{R}$,

(2) $K(u) = K(-u) \implies \int u \, K(u) \, du = 0$,

(3) $\int K(u) \, du = 1$,

(4) $\int u^2 \, K(u) \, du =: \mu_2(k) \neq 0$.

Because of (4) such a kernel is called of order 2 or **second order kernel**. The argument of the kernel is the scaled variable

$$u = \frac{x - x_i}{b}.$$

Second order kernels with an **infinite support** are, e.g.:

- the GAUSS kernel,
- the CAUCHY kernel,
- the LAPLACE kernel,
- the logistic kernel.

For more details of these kernels see Tab. 8/1. More popular, especially in HR and PDF estimation of lifetime data, are kernels with **finite support** which mostly are polynomial functions related to the beta distribution, more precisely, they are symmetric beta distributions on the interval $[-1, \, 1]$. Their generating formula is

$$K(u) = \kappa_{r,s} \left(1 - |u|^r\right)^s I_{|u| \leq 1} \tag{8.24a}$$

with

$$\kappa_{r,s} = \frac{r}{2 \, B(s+1, \, 1/r)}, \; r > 0, \; s \geq 0, \tag{8.24b}$$

and the beta function

$$B(p, \, q) = \int\limits_0^1 v^{p-1} \, (1-v)^{q-1} dv = \frac{\Gamma(p) \, \Gamma(q)}{\Gamma(p+q)}. \tag{8.24c}$$

Among these kernels — some of which come with different names — the most popular are:

- the **uniform kernel** or **rectangular kernel** with

$$s = 0, \; r = 1 \implies \kappa_{1,0} = 1/2;$$

- the **triangular kernel** with

$$s = 1, \; r = 1 \implies \kappa_{1,1} = 1;$$

- **EPANECHNIKOV kernel** or **quadratic kernel** with

$$s = 1, \; r = 2 \implies \kappa_{2,1} = 3/4;$$

[10] Asymmetric kernels will be needed when estimating near the boundaries, see further down.

- **biweight kernel** or **quartic kernel** or **biquadratic kernel** with

$$s = 2,\ r = 2 \implies \kappa_{2,2} = 15/16;$$

- **triweight kernel** or **triquadratic kernel** with

$$s = 3,\ r = 2 \implies \kappa_{2,3} = 35/32;$$

- **tricube kernel** with

$$s = 3,\ r = 3 \implies \kappa_{3,3} = 70/81.$$

After a suitable rescaling the GAUSS kernel is seen to be of the above type with $r = 2,\ s = \infty$. Two other kernels with finite support but not of polynomial type are the **cosine kernel**

$$K(u) = \frac{\pi}{4}\,\cos\Bigl(\frac{\pi}{2}\,u\Bigr)\; I_{|u|\le 1} \tag{8.25}$$

and the **semi–elliptical kernel**

$$K(u) = \frac{2}{\pi}\sqrt{1-u^2}\, I_{|u|\le 1}. \tag{8.26}$$

Tab. 8/1 summarizes the kernels mentioned above and displays them together with the pertaining $\mu_2(K) = \int u^2\, K(u)\, du$ and $R(K) = \int K(u)^2\, du$. $I_{|u|\le 1}$ is the indicator function:

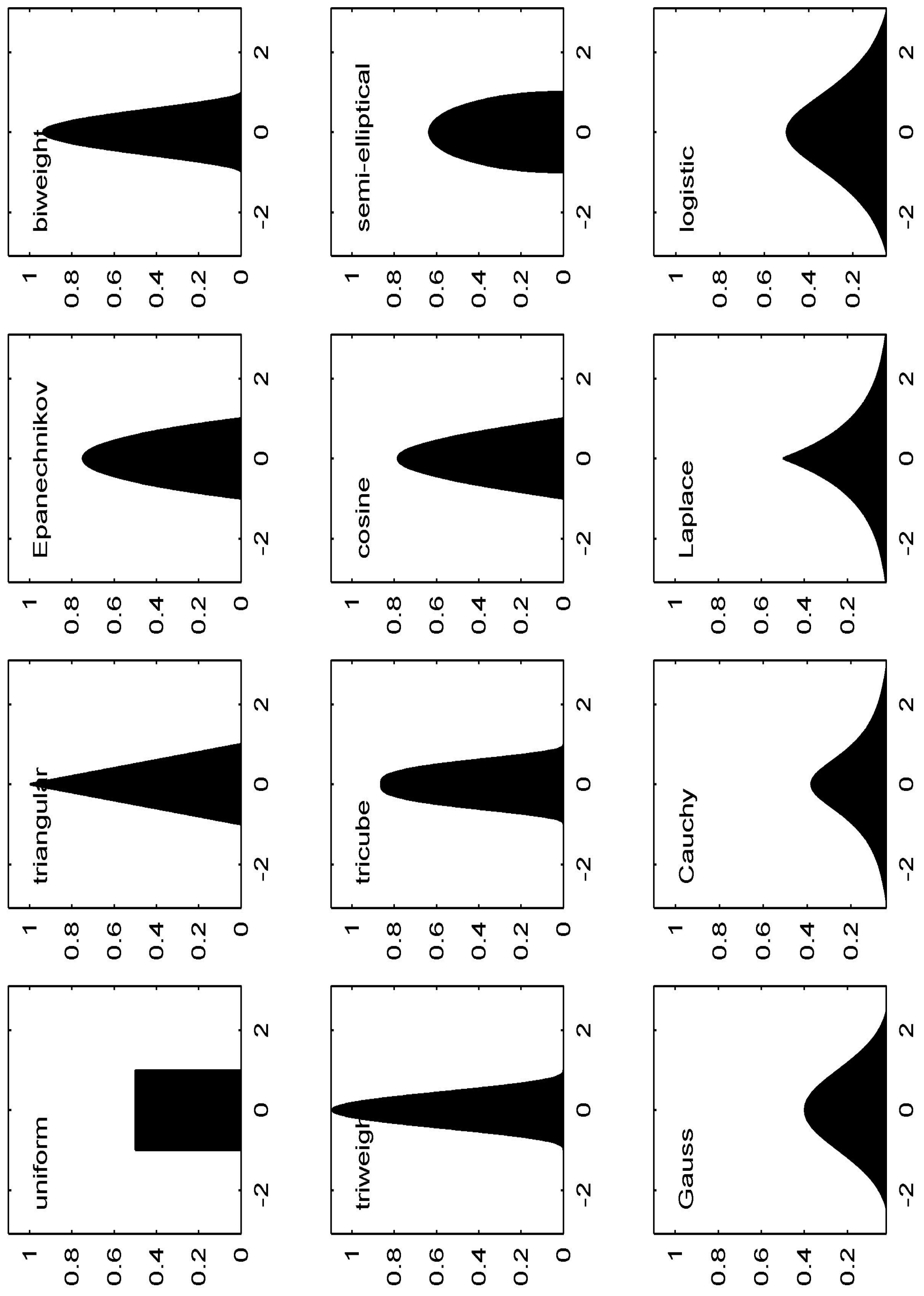
$$I_{|u|\le 1} = \left\{ \begin{array}{ll} 1 & \text{for } u \in [-1,\,1] \\ 0 & \text{for } u \text{ else.} \end{array} \right\} \tag{8.27}$$

Fig. 8/2 displays all the kernels mentioned above.

Table 8/1: Common kernels

Name	Formula	$\mu_2(K)$	$R(K)$
uniform (rectangular)	$K(u) = \frac{1}{2}\, I_{\lvert u\rvert\le 1}$	$\frac{1}{3} \approx 0.3333$	$\frac{1}{2} = 0.5$
triangular	$K(u) = (1-\lvert u\rvert)\, I_{\lvert u\rvert\le 1}$	$\frac{1}{6} \approx 0.1667$	$\frac{2}{3} \approx 0.667$
EPANECHNIKOV (quadratic)	$K(u) = \frac{3}{4}\,\bigl(1-u^2\bigr)\, I_{\lvert u\rvert\le 1}$	$\frac{1}{5} = 0.2$	$\frac{3}{5} = 0.6$
biweight (quartic, biquadratic)	$K(u) = \frac{15}{16}\,\bigl(1-u^2\bigr)^2\, I_{\lvert u\rvert\le 1}$	$\frac{1}{7} \approx 0.1429$	$\frac{5}{7} \approx 0.7143$
triweight (triquadratic)	$K(u) = \frac{35}{32}\,\bigl(1-u^2\bigr)^3\, I_{\lvert u\rvert\le 1}$	$\frac{1}{9} \approx 0.1111$	$\frac{350}{429} \approx 0.8159$
tricube	$K(u) = \frac{70}{81}\,\bigl(1-\lvert u\rvert^3\bigr)^3\, I_{\lvert u\rvert\le 1}$	$\frac{35}{243} \approx 0.1440$	$\frac{175}{247} \approx 0.7086$
cosine	$K(u) = \frac{\pi}{4}\,\cos\bigl(\frac{\pi}{2}\,u\bigr)\; I_{\lvert u\rvert\le 1}$	$1 - \frac{8}{\pi^2} \approx 0.1884$	$\frac{\pi^2}{16} \approx 0.6169$
semi–elliptical	$K(u) = \frac{2}{\pi}\,\sqrt{1-u^2}\, I_{\lvert u\rvert\le 1}$	$\frac{1}{4} = 0.25$	$\frac{16}{3\,\pi^2} \approx 0.5404$
GAUSS	$K(u) = \frac{1}{\sqrt{2\,\pi}}\,\exp\bigl[-u^2/2\bigr],\ u\in\mathbb{R}$	1	$\frac{1}{2\,\sqrt{\pi}} \approx 0.2821$
CAUCHY	$K(u) = \bigl[\pi\,(1+u^2)\bigr]^{-1},\ u\in\mathbb{R}$	non–existent	$\frac{1}{2\,\pi} \approx 0.1592$
LAPLACE	$K(u) = \frac{1}{2}\,\exp\bigl[-\lvert u\rvert\bigr],\ u\in\mathbb{R}$	2	$\frac{1}{4} = 0.25$
logistic	$K(u) = \frac{\exp(-u)}{\bigl[1+\exp(-u)\bigr]^2}\ u\in\mathbb{R}$	$\frac{\pi^2}{3} \approx 3.2899$	$\frac{1}{6} \approx 0.1667$

Figure 8/2: Common kernels



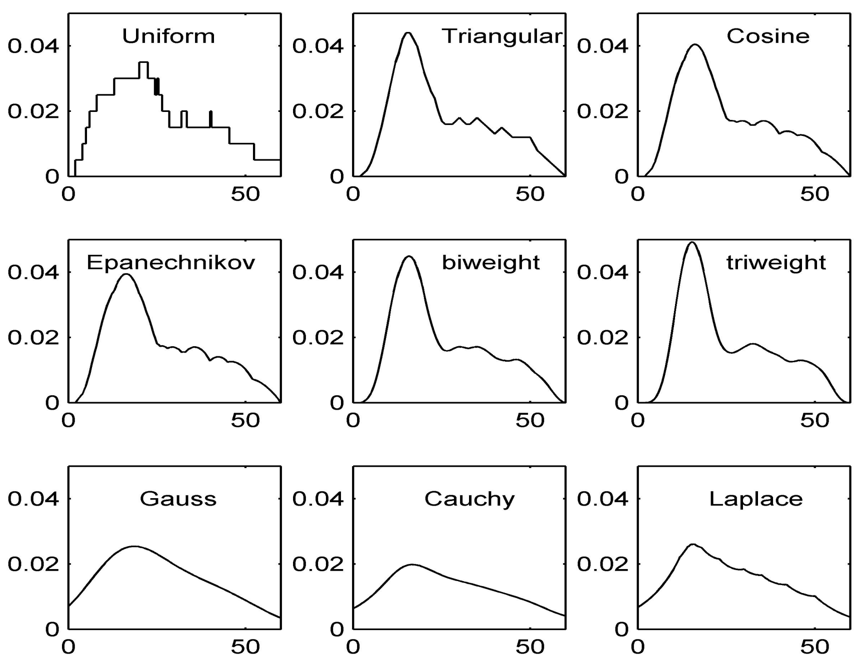
Optimizing AMISE (8.21) *with respect to the kernel* is not an easy problem since the scaling of K is coupled with the bandwidth b. EPANECHNIKOV (1969) has found the AMISE–optimizing kernel to be

$$K(u) = \frac{3}{4}\left(1 - u^2\right) I_{|u| \leq 1}. \tag{8.28}$$

Investigations have revealed that using another 'suboptimal' kernel does not cause great loss of efficiency, i.e., seldom more than 5%. Indeed, these results suggest that most unimodal densities perform about the same as each other when used as a kernel. Thus, the choice between kernels can be made on other grounds such as computational efficiency. The kernel effects the local smoothness whereas the bandwidth is responsible for the global smoothness of the estimate. The smaller the sample size the greater the effect of the kernel and when $n = 1$, then the kernel wholly determines the graph of the estimate.

Fig. 8/3 shows the effect of the kernel. The data used in this figure are: 12, 14, 15, 16, 18, 23, 35, 42, 50, and the bandwidth has been set to $b = 10$. The uniform kernel gives an estimate of the density that is piecewise konstant, the triangular kernel and the LAPLACE kernel have a kink which is reflected in the estimated densities. Even the EPANECHNIKOV kernel gives an estimate having a discontinuous first derivative which sometimes can be unattractive because of its kinks. Very smooth estimates are produced by the GAUSS and the CAUCHY kernels, respectively. The use of the triweight kernel seems to be a good compromise.

Figure 8/3: Effect of kernel choice



We have seen in (8.23) that the best obtainable rate of convergence of the kernel estimator considered there is of order $n^{-4/5}$. But it is possible to obtain a better rate of convergence at the price of relaxing the restriction that the kernel be a density and $K(u) \geq 0 \; \forall \; u$. Such kernels are of higher order than two. We say that $K(u)$ is an **ℓ-th order kernel** if

$$\begin{aligned}
\mu_0(K) &= \textstyle\int u^0 \, K(u) \, du = 1, \\
\mu_j(K) &= \textstyle\int u^j \, K(u) \, du = 0 \;\text{ for }\; j = 1, 2, ..., \ell - 1 \;\text{ and} \\
\mu_\ell(K) &= \textstyle\int u^\ell \, K(u) \, du \neq 0.
\end{aligned}$$

Still requiring that $K(u)$ be symmetric we see that ℓ must be even. With $\ell \to \infty$ the convergence rate can be made arbitrarily close to n^{-1}, the parametric convergence rate.

There are several rules to construct higher–order kernels. Let $K_\ell(u)$ denote the ℓ-th order kernel which is assumed to be differentiable, then formula

$$K_{\ell+2}(u) = \frac{3}{2}\,K_\ell(u) + \frac{1}{2}\,u\,K'_\ell(u) \tag{8.29}$$

can be use to generate higher–order kernels. Taking the GAUSS kernel

$$K_2(u) = \frac{1}{\sqrt{2\,\pi}}\,\exp(-u^2/2)$$

we find from (8.29)

$$K_4(u) = 0.5\,(3 - u^2)\,\frac{1}{\sqrt{2\,\pi}}\,\exp(-u^2/2),$$

and taking the triweight kernel

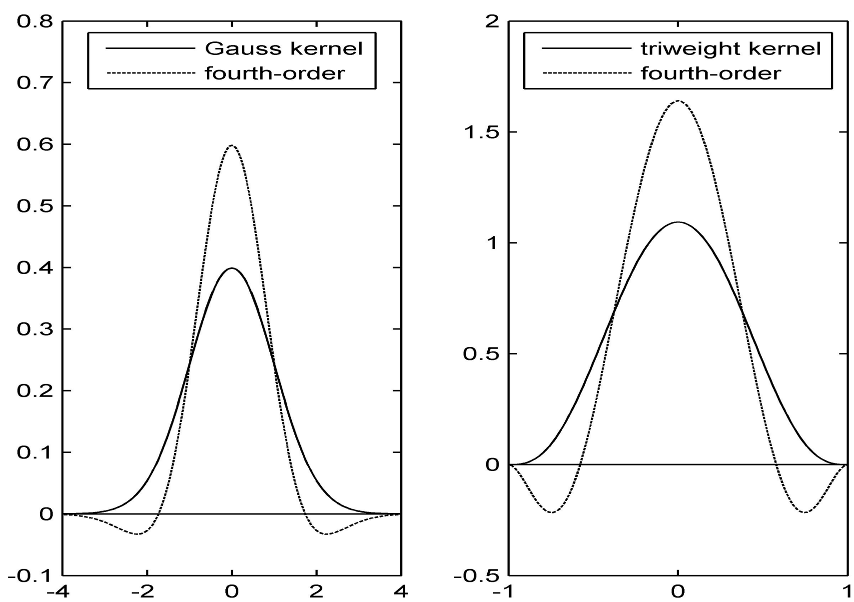
$$K_2(u) = \frac{35}{32}\,\big(1 - u^2\big)^3\,I_{|u|\le 1}$$

we have

$$K_4(u) = \frac{105}{65}\,\big(1 - u^2\big)^2\,\big(1 - 3\,u^2\big)\,I_{|u|\le 1}.$$

Fig. 8/4 shows these kernels. Notice the negative lobes of $K_4(u)$ which entail that the resulting smoothed density will not be a density itself. Higher–order kernels necessarily take on negative values, so there is a price to be paid in interpretability and plausibility.

Figure 8/4: Fourth–order kernels based on GAUSS and triweight kernels, respectively



There are some situations where there is scope for improvement of the basic kernel presented up to here. Some of the modifications will be needed in Sect. 8.1.3, and we will shortly present their ideas here. A first modification is the **local kernel estimator**. Given that the optimal amount of smoothing varies across the real line, an obvious extension of

$$\widehat{f}_n(x;b) = \frac{1}{n\,b}\sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right)$$

is to that having different bandwidth $b(x)$, say, for each x where $f(\cdot)$ is to be estimated. This leads to the local kernel estimator

$$\widehat{f}_n[x; b(x)] = \frac{1}{n\, b(x)} \sum_{i=1}^{n} K\left(\frac{x - x_i}{b(x)}\right), \tag{8.30}$$

where a different basic kernel estimator is employed at each point. A popular method which fits into the framework of (8.30) is the **nearest neighbor kernel estimator** that uses distances from x to the data point being the k-th nearest to x.

A quite different idea from local kernel estimation is that of **variable kernel estimation** where the single b is replaced by n values $b(x_i);\ i = 1,\, 2,\, \ldots,\, n$; rather than by $b(x)$. The estimator has the form

$$\widehat{f}_n[x; b(x_i)] = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{b(x_i)}\, K\left(\frac{x - x_i}{b(x_i)}\right), \tag{8.31}$$

so that the kernel centered on x_i has associated with it its own scale parameter $b(x_i)$ allowing different degrees of smoothing depending on where x_i is in relation to other data points. The aim is to smooth out the mass associated with data values that are in sparse regions much more than those situated in the main body of the data. Variable kernel estimation can also be realized by a nearest neighbor approach, but using the distance from x_i to the data point being the k-th nearest to x_i.

A special situation arises when the function to be estimated has a bounded support. For the lifetime variable we have a naturally lower bound equal to zero. When the point x where to estimate $f(\cdot)$ or $h(\cdot)$ is smaller than the bandwidth a symmetric kernel is not appropriate because no lifetimes less than zero are observable. In the region $0 \le x < b$ the use of an asymmetric kernel is suggested. This is also true when there is a right endpoint x_{end} of the support where we have to consider an asymmetric kernel for $x_{\text{end}} - b < x \le x_{\text{end}}$. The right endpoint is often taken to be the greatest observed lifetime in the sample. The asymmetric kernels needed near the boundaries and which are different for each x within a distance of b to the boundary are called **boundary kernels**. In Sect. 8.1.3.1 we will present different types of boundary kernels.

8.1.1.4 Bandwidth Selection

The implementation of a kernel estimator requires the specification of a bandwidth b. One possibility is to choose the bandwidth **subjectively by eye** and on aesthetic grounds. This would involve looking at several PDF or HR estimates over a range of bandwidths and selecting the estimate that is the 'most pleasing' in some sense. One such strategy is to begin with a large bandwidth and to decrease the amount of smoothing until fluctuations that are more random than structural start to appear. However, there are also many circumstances where it is very beneficial to have the bandwidth automatically selected from the data.

A method that uses the data $x_1,\, x_2,\, \ldots,\, x_n$ to produce a bandwidth b is called a **bandwidth selector**. A look into the professional journals reveals that work on constructing bandwidth selectors is still going on. Available selectors can be roughly divided into two classes:

- quick and simple selectors, sometimes called 'quick and dirty methods', and
- sophisticated selectors aiming to minimize AMISE or some other criterion.

The first class consists of formulas which are easy to evaluate, but without any mathematical guarantee of being close to the optimal bandwidth. These selectors often provide a starting point for the subjective choice of the smoothing parameter. The two methods falling into the category 'quick and dirty' are the **rule of thumb**, sometimes called normal scale bandwidth selector, and the **maximal smoothing** or **oversmoothing principle**. Both are based on the optimal bandwidth minimizing the AMISE:

$$b_{\text{AMISE}} = \left[\frac{R(K)}{n\,R(f'')\,\mu_2(K)^2}\right]^{1/5}.$$

Note that only the term $R(f'')$ is unknown in this expression.

The rule of thumb replaces the unknown PDF $f(\cdot)$ in this functional by a reference distribution function. The reference distribution is rescaled to have variance equal to the sample variance. If, e.g., we take K as the GAUSS kernel and the standard normal distribution as reference distribution the rule of thumb yields the bandwidth

$$b_{\text{RoT}} = 1.0592\,\widehat{\sigma}\,n^{-1/5} \tag{8.32a}$$

where $\widehat{\sigma}^2 = \sum(x_i - \overline{x}/(n-1)$ is the sample variance. A version which is more robust against outliers in the sample uses the interquartile range rq as a measure of spread instead of the variance giving the modified estimator

$$b_{\text{RoT}} = 1.0592\,\left[\min\left(\widehat{\sigma},\,\frac{rq}{1.34}\right)\right]\,n^{-1/5}. \tag{8.32b}$$

The maximal smoothing principle is due to TERRELL (1990). He showed that there is a lower bound for the functional $R(f'')$ for all densities having standard deviation σ, and this bound is attained by the triweight density, see Tab. 8/1. Thus we have an upper bound for b_{AMISE} leading to the oversmoothed bandwidth

$$b_{\text{os}} = \left[\frac{243\,R(K)}{35\,\mu_2(K)^2\,n}\right]^{1/5}\,\widehat{\sigma}. \tag{8.33}$$

While b_{os} will give a too large bandwidth for optimal estimation of a general density $f(\cdot)$ it provides an excellent starting point for subjective choice of the bandwidth. A graphical strategy is to plot an estimate with the bandwidth b_{os} and then successively look at plots based on fractions of b_{os} to see what features are present in the data.

There are two fundamental approaches in the class of sophisticated selectors:

- the **cross–validation method** and
- the **plug–in method**,

each coming with several versions.

The most popular and best studied cross–validation method is **least–squares cross–validation**, originally proposed by RUDEMO (1982) and BOWMAN (1984). Its motivation comes from expanding the MISE of $\widehat{f}_n(\cdot;b)$ to obtain

$$\text{MISE}\big[\widehat{f}_n(\cdot;b)\big] = \text{E}\Big[\int \widehat{f}_n(x;b)^2\,dx\Big] - 2\,\text{E}\Big[\int \widehat{f}_n(x;b)\,f(x)dx\Big] + \int f(x)^2\,dx. \tag{8.34a}$$

As $\int f(x)^2\,dx$ does not depend on b the minimization of MISE is equivalent to minimization of

$$\text{MISE}\big[\widehat{f}_n(\cdot;b)\big] - \int f(x)^2\,dx = \text{E}\Big[\int \widehat{f}_n(x;b)^2\,dx - 2\int \widehat{f}_n(x;b)\,f(x)\,dx\Big]. \tag{8.34b}$$

The right–hand side of (8.34b) is unknown since it depends on $f(x)$. Using a method of moments to estimate this term results in the least–squares cross–validation function

$$LSCV(b) = \int \widehat{f}_n(x;b)^2\,dx - \frac{2}{n}\sum_{i=1}^{n}\widehat{f}_{-i}(x_i;b) \tag{8.35a}$$

where

$$\widehat{f}_{-i}(x_i; b) = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{b} K\left(\frac{x - x_j}{b}\right) \tag{8.35b}$$

is the density estimate based on the sample with x_i deleted, often called the **leave–one–out density estimator**. This is the reason for the name 'cross–validation' which refers to the use of part of a sample to obtain information about another part. It therefore seems reasonable to chose b to minimize $LSCV(b)$; the bandwidth chosen by this way is denoted b_{LSCV}. This estimate of b suffers a lot under sample variation, i.e., for different samples from the same distribution the estimated bandwidths have a big variance. Another drawback of *LSCV* is that it often has several minima. Simulation studies have shown that this problem can be fixed by selecting the largest value of b for which a local minimum occurs.

The idea of plug–in methods goes back to WOODROOFE (1970). These methods are based on the asymptotically best choice of b given by b_{AMISE} in (8.22). The only unknown quantity in b_{AMISE} is the functional $R(f'')$. WOODROOFE proposed to use a first bandwidth b_1 to calculate $\widehat{f}_n(x; b_1)$, take this estimate to calculate $\widehat{R}(f'') = R\big[\widehat{f}_n(x; b_1)\big]$ and to plug $\widehat{R}(f'')$ into (8.22) to obtain b_2, the final bandwidth. This **direct plug–in rule** may be generalized by iterating the process, i.e., calculating $\widehat{R}(f'') = R\big[\widehat{f}_n(x; b_2)\big]$, plug $\widehat{R}(f'')$ into (8.22) to obtain b_3 etc., until b_i converges.

8.1.2 Indirect Smoothing — The Ratio–type Estimator[11]

The hazard rate is defined as

$$h(x) = \frac{f(x)}{S(x)}.$$

A smoothed hazard rate estimator based on this definition and on suitably chosen estimators of the nominator and the denominator is called a **ratio–type estimator** or an **indirectly smoothed estimator**. The latter name is justified by the fact that we do not smooth rough estimates of the hazard rate itself. Direct smoothing as presented in Sect. 8.1.3 prevails owing to its theoretical tractability (exact mean square errors are available) and aesthetic superiority over the ratio–type estimator even though – as has been shown by RICE/ROSENBLATT (1976) — the direct and the indirect estimators have the same asymptotic variance but different asymptotic biases.

The ratio–type estimator comes in two variants, resulting from the use of different estimators of the survival function $S(x)$. The first variant, called **simple indirect estimator**, takes the KAPLAN/MEIER estimator, see (8.3a) and (8.4a):[12]

$$\widehat{S}_n(x) = \left\{ \begin{array}{ll} \displaystyle\prod_{i\,:\,y_i \leq x} \left(\frac{n-i}{n-i+1}\right)^{\delta_i} & \text{for untied observations} \\ \displaystyle\prod_{i\,;\,x_i \leq x} \left(1 - \frac{d_i}{n_i}\right) & \text{for tied observations.} \end{array} \right\} \tag{8.36a}$$

The second variant, called **smoothed indirect estimator**, is based on the integrated smoothed density estimator

$$\widehat{S}_n(x) = 1 - \int_0^x \widehat{f}_n(u)\, du \tag{8.36b}$$

where $\widehat{f}_n(\cdot)$ is a kernel estimator of $f(\cdot)$ with a kernel that has to be PDF itself. The results from using (8.36a) or (8.36b) do not differ much, but as (8.36a) is a stair–case function the simple

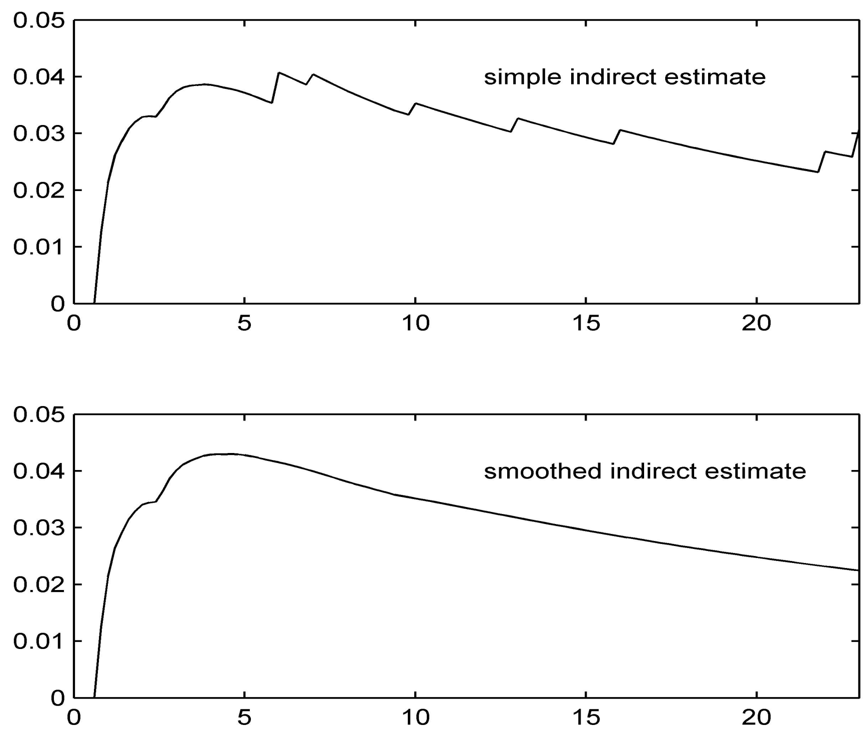
[11] Suggested reading for this section: LO et al. (1989), RICE/ROSENBLATT (1976), WATSON/LEADBETTER (1964a,b).

[12] SARDA/VIEU (1990) show how to find the ISE –minimizing bandwidth by cross–validation.

indirect estimator will generate hazard rate courses which are less smooth than those coming from the smoothed indirect estimator.

To show this effect we have plotted in Fig. 8/5 both versions of the ratio–type estimator. This figure rests upon the data of Example 5/1 (leukaemia patients' data). We have used the function '`ksdensity`' of MATLAB with a positive support, the EPANECHNIKOV kernel and bandwidth 10.

Figure 8/5: Ratio–type estimates of the hazard rate for the leukaemia patients' data



The smoothed indirect estimator has first been studied by WATSON/LEADBETTER (1964a,b) for uncensored samples. The paper of LO et al. (1989) investigates this estimator for samples with censored observations. We shortly repeat the results of WATSON/LEADBETTER, but will not go further into the details of indirect smoothing as it is of no great importance in practice. WATSON/LEADBETTER use a sequence $\{\delta_n(x)\}$ of smoothing functions tending, as $n \to \infty$, to a **DIRAC delta–function**.[13] This delta–sequence method is quite general and covers several types of smoothing methods, including the kernel method with $\delta_n(u) = (1/b)\, K(u/b)$. The smoothed indirect estimator of WATSON/LEADBETTER reads

$$\widehat{h}_n(x) = \frac{\widehat{f}_n(x)}{1 - \widehat{F}_n(x)} \tag{8.37a}$$

[13] The δ–function is (informally) a generalized function on $\mathbb{R}$ that is zero anywhere except at zero, with an integral of unity over $\mathbb{R}$. This unit impulse function may be written as

$$\delta(x) = \left\{ \begin{array}{cl} +\infty & \text{for } x = 0 \\ 0 & \text{for } x \neq 0 \end{array} \right\}$$

with $\int\limits_{-\infty}^{+\infty} \delta(x)\, dx = 1$. Rigorously defined the δ–function is a distribution or a measure.

with

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\delta_n(x - x_i), \tag{8.37b}$$

$$\widehat{F}_n(x) = \int_0^x \widehat{f}_n(u)\,du. \tag{8.37c}$$

If the sequence $\{\delta_n(x)\}$ is suitably chosen, i.e., if

$$\alpha_n = \int \delta_n^2(x)\,dx < \infty \tag{8.37d}$$

at every point of continuity x of $h(\cdot)$ at which $F(x) < 1$, then $\widehat{h}_n(x)$ is shown to be asymptotically unbiased with an asymptotic variance

$$\text{Var}\Big[\widehat{h}_n(x)\Big] \approx \frac{\alpha_n}{n}\,\frac{h(x)}{1 - F(x)}. \tag{8.37e}$$

If in addition to (8.37d) $\alpha_n = o(n)$, then the asymptotic variance converges to zero in order of α_n/n, i.e., $\widehat{h}_n(x)$ is consistent. Under some further slightly more restrictive conditions on $\big\{\delta_n(x)\big\}$ the random variable

$$\big[1 - F(x)\big]\,\left[\frac{n}{\alpha_n\,f(x)}\right]^{1/2}\,\big[\widehat{h}_n(x) - h(x)\big]$$

is asymptotically standard normally distributed at every continuity point of $h(\cdot)$.

SALHA (w.y.) has generalized the smoothed indirect estimator to incorporate a bandwidth depending on x_i as $b_i = c_i\,b$ with

$$c_i = \left[\frac{\widetilde{f}(x_i)}{g}\right]^{-\alpha} \tag{8.38a}$$

where $\widetilde{f}(x_i)$ is a pilot estimate and

$$g = \left[\prod_{i=1}^{n}\widetilde{f}(x_i)\right]^{1/n} \tag{8.38b}$$

and $0 < \alpha < 1$. He suggests to take $\alpha = 0.5$.

8.1.3 Direct Smoothing

In this section we — generally — will look at samples with untied observations and possible censoring. The general form of a direct kernel estimator of the hazard rate in such a situation can be written as

$$\widehat{h}_n(x) = \sum_{i=1}^{n}\frac{\delta_i}{n - i - 1}\,\frac{1}{b_n(i,x)}K_x\left(\frac{x - y_i}{b_n(i,x)}\right);\; b_n(i,x) > 0,\; 0 \le x \le x_{\text{end}}. \tag{8.39}$$

δ_i is the censoring indicator corresponding to the i-th ordered observation y_i. $\delta_i = 1$ stands for an uncensored observation and $\delta_i = 0$ for a censored observation. $b_n(i,x)$ represents the bandwidth function. The bandwidth will depend inversely on the sample size n, but we can also make it dependent on the point x for which $h(\cdot)$ is to be estimated and/or on the data point y_i processed by the kernel. The numerous ways of specifying the bandwidth have been developed in order to optimize the estimator and here they serve as guideline for the organization of this section.

The bandwidth function leads to different properties of the resulting hazard rate estimator whereas the influence of the particular kernel function $K_x(\cdot)$ is only marginal except for the behavior near the boundaries. We thus may have different kernels near the boundaries, i.e., for $x \in \big[0;\, b_n(i,x)\big)$ and $x \in \big(x_{\text{end}} - b_n(i,x);\, x_{\text{end}}\big]$, and some other kernel in the interior of the data body, i.e., for $x \in \big[b_n(i,x);\, x_{\text{end}} - b_n(i,x)\big]$.

We will first present boundary kernels (Sect. 8.1.3.1) before we turn to different possibilities of choosing a bandwidth (Sect. 8.1.3.2 – 8.1.3.3).

8.1.3.1 Boundary Kernels[14]

Let $\big[0,\, x_{\text{end}}\big]$ be the support of the hazard rate to be estimated, where x_{end} is the greatest uncensored observation in the sample. When we have a kernel with bandwidth b, i.e., $K\big(\frac{x-x_i}{b}\big)$, a part of those kernels having either $0 \le x_i < b$ or $x_{\text{end}} - b < x_i \le x_{\text{end}}$ is outside the support. It has been noted that bias problems occur when estimating near an endpoint of the data. The application of unmodified kernel estimators leads to meaningless estimates in the boundary regions near the endpoints. Therefore some authors suggest to estimate $h(x)$ — or $f(x)$ in case of density estimation — only for the interior region $[b,\, x_{\text{end}} - b]$. Another suggestion is the so–called 'cut–and–normalize' modification whereby that part of the kernel lying outside the boundary is omitted and the remaining part is normalized to be a proper density. This solution achieves consistency near the boundary but results in a large bias there. A variety of further modifications is possible to achieve a smaller bias, see KARUNAMUNI/ALBERTS (2005). One can think of these boundary modifications in terms of special boundary kernels which are different for each x within a distance of b to the boundary.

For the following formulas we define

$$q = \frac{x}{b} \;\text{ for }\; 0 \le x < b$$

and

$$q = \frac{x_{\text{end}} - x}{b} \;\text{ for }\; x_{\text{end}} - b < x \le x_{\text{end}},$$

so we have $0 \le q < 1$. We will look at boundary kernels corresponding to some basic second order kernels.

One simple family of boundary kernels for the left–hand side (lower boundary side) is the following **linear multiple of a given kernel** $K(u)$

$$K_q^L(u) = \frac{\mu_{2,q}(K) - \mu_{1,q}(K)\,u}{\mu_{0,q}(K)\,\mu_{2,q}(K) - \mu_{1,q}(K)^2}\,K(u)\,I_{\{-1\le u\le q\}} \tag{8.40a}$$

with

$$\mu_{\ell,q}(K) = \int_{-1}^{q} u^\ell\,K(u)\,du. \tag{8.40b}$$

This family goes back to GASSER/MÜLLER (1979). For the left boundary we have

Uniform kernel

$$\left.\begin{aligned} {}_UK(u) &= 0.5 \;\text{ for for }\; -1 \le u \le 1 \\ {}_UK_q^L(u) &= \frac{4\big[2\,(1-q+q^2) + 3\,(1-q)\,u\big]}{(1+q)^3}\;{}_UK(u) \text{ for } -1 \le u \le q \end{aligned}\right\} \tag{8.41a}$$

[14] Suggested reading for this section: BOUEZMARNI/ROMBOUTS (2008), GASSER/MÜLLER (1979), GASSER/MÜLLER/MAMMITZSCH (1985), KARUNAMUNI/ALBERTS (2005), MESSER/GOLDSTEIN (1993), MÜLLER (1991, 1993), MÜLLER/WANG (1994).

EPANECHNIKOV kernel

$$\left.\begin{aligned}{}_E K(u) &= 0.75\,(1-u^2) \text{ for } -1 \le u \le 1 \\ {}_E K_q^L(u) &= \frac{64\,(2-4q+6q^2-3q^3)+240\,(1-q)^2\,u}{(1+q)^4\,(19-18\,q+3\,q^2)}\;{}_E K(u) \text{ for } -1 \le u \le q\end{aligned}\right\} \quad (8.41b)$$

Biweight kernel

$$\left.\begin{aligned}{}_B K(u) &= \frac{15}{16}\,(1-u^2)^2 \text{ for } -1 \le u \le 1 \\ {}_B K_q^L(u) &= \frac{\begin{array}{c}64\,(8-24\,q+48\,q^2-45\,q^3+15\,q^4)\\ +1120\,(1-q)^3\,u\end{array}}{(1-q)^5(81-168q+126q^2-40q^3+5q^4)}\;{}_B K(u) \text{ for } -1 \le u \le q\end{aligned}\right\} \quad (8.41c)$$

Triweight kernel

$$\left.\begin{aligned}{}_T K(u) &= \frac{35}{32}\,(1-u^2)^3 \text{ for } -1 \le u \le 1 \\ {}_T K_q^L(u) &= \frac{\begin{array}{c}256(-8(-16+q(64+5q(-32+q(43\\ +7(-4+q)q)))))+80640\,(1-q^4)u\end{array}}{\begin{array}{c}(1+q)^6(5359+5q(-3550+q\,(4909\\ +q\,(-3620+q(1517+35(-10+q)q)))))\end{array}}\;{}_T K(u) \text{ for } -1 \le u \le q\end{aligned}\right\} \quad (8.41d)$$

Two other classes have been suggested by MÜLLER (1991) and MÜLLER/WANG (1994). Both classes result as solutions of a variational problem under asymmetric support and lead to classes of compactly supported polynomial kernel functions. The class proposed in MÜLLER/WANG (1994) gives rise to smaller leading constants of the asymptotic MSE than the previously suggested class. When the basic kernel is the uniform kernel the corresponding boundary kernels in both classes are the same as (8.41a), but for other types of basic kernels the formulas are different from formulas(8.41b-d).

The left boundary versions of the **1991–class** are

EPANECHNIKOV kernel

$${}_E K_q^{91}(u) = \frac{6(1+u)(q-u)}{(1+q)^3}\left\{1+5\left(\frac{1-q}{1+q}\right)^2+10\frac{1-q}{(1+q)^2}u\right\} \text{ for } -1 \le u \le q, \quad (8.42a)$$

Biweight kernel

$${}_B K_q^{91}(u) = \frac{30(1+u)^2(q-u)}{(1+q)^5}\left\{1+7\left(\frac{1-q}{1+q}\right)^2+14\frac{1-q}{(1+q)^2}u\right\} \text{ for } -1 \le u \le q, \quad (8.42b)$$

Triweight kernel

$${}_T K_q^{91}(u) = \frac{140(1+u)^3(q-u)^3}{(1+q)^7}\left\{1+9\left(\frac{1-q}{1+q}\right)^2+18\frac{1-q}{(1+q)^2}u\right\} \text{ for } -1 \le u \le q, \quad (8.42c)$$

and for the **1994–class**

EPANECHNIKOV kernel

$${}_E K_q^{94}(u) = \frac{12(u+1)}{(1+q)^4}\left\{\frac{3q^2-2q+1}{2}+(1-2q)u\right\} \text{ for } -1 \le u \le q, \quad (8.43a)$$

Biweight kernel

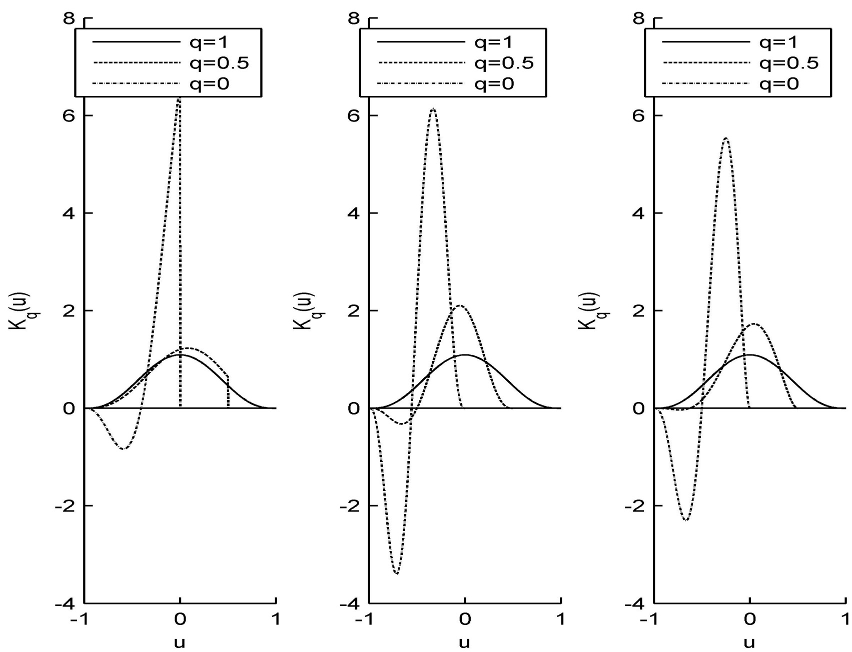
$$ {}_B K_q^{94}(u) = \frac{15(u+1)^2(q-u)}{(1+q)^5}\left\{2u\left(5\frac{1-q}{1+q}-1\right)+(3q-1)+\frac{5(1-q)^2}{1+q}\right\} \text{ for } -1 \le u \le q, \tag{8.43b} $$

Triweight kernel

$$ {}_T K_q^{94}(u) = \frac{70(u+1)^3(q-u)^2}{(1+q)^7}\left\{2u\left(7\frac{1-q}{1+q}-1\right)+(3q-1)+\frac{7(1-q)^2}{1+q}\right\} \text{ for } -1 \le u \le q, \tag{8.43c} $$

Fig. 8/6 displays all three variants of the triweight boundary kernel.

Figure 8/6: Triweight boundary kernels
(left: linear multiple, center: MÜLLER–91, right: MÜLLER/WANG–94)



For all three classes of boundary kernels we can state the following results:

1. All kernels conform to the moments conditions of second order kernels:
$\int_{-1}^{q} K_q(u)\,du = 1,\ \int_{-1}^{q} u\,K_q(u)\,du = 0,\ \int_{-1}^{q} u^2\,K_q(u)\,du \neq 0.$
2. For the right boundary the formulas are the same, but with $-u$ instead of u.
3. For $q \rightarrow 1$ the boundary kernel approaches the basic kernel.
4. For $q \rightarrow 0$ the boundary kernel takes on negative values. This might lead to negative hazard rate estimates which should be replaced by zero values.

8.1.3.2 Kernel Estimators with Globally Constant (Fixed) Bandwidth[15]

In (8.39) we have introduced the general form of the direct kernel estimator with a bandwidth depending on the sample size n, the data to be processed $\{y_i\}$, and the point x for which the

[15] Suggested reading for this section: DIEHL/STUTE (1988), LO et al. (1989), RAMLAU–HANSEN (1983), RICE/ROSENBLATT (1976), TANNER/WONG (1983), UZUNOGULLARI/WANG (1992), WATSON/LEADBETTER (1964a,b), YANDELL (1983).

hazard rate is to be estimated. The simplest case of kernel estimation is given by a bandwidth which is independent of $\{y_i\}$ as well as of x and is fixed for a given sample size:

$$b_n = b_n(i, x) \; \forall \, i \text{ and } \forall \, x.$$

Sometimes this bandwidth is called **global** and is determined by one of the methods presented in Sect. 8.1.1.4. Of course, the optimal, i.e., the MISE or AMISE minimizing global bandwidth depends on the kernel chosen and on the distribution of the lifetime variable X and on that of the censoring variable Z, respectively, see (8.52b). The fixed–bandwidth kernel estimator to be presented in this section reads

$$\widehat{h}_n(x) = \frac{1}{b_n} \sum_{i=1}^{n} \frac{\delta_i}{n-i+1} \, K\left(\frac{x-y_i}{b_n}\right). \tag{8.44}$$

This estimator has been discussed extensively and analyzed with different techniques.

We first look at the properties of the estimator when there is *no censoring*, i.e., we have $\delta_i = 1 \; \forall \, i$ and each observation y_i is a failure time x_i. This case has already been studies by WATSON/LEDBETTER (1064a,b), and subsequently by RICE/ROSENBLATT (1976) and RAMLAU–HANSEN (1983). The estimator in this case of no censoring reads

$$\widehat{h}_n(x) = \frac{1}{b_n} \sum_{i=1}^{n} \frac{1}{n-i+1} \, K\left(\frac{x-x_i}{b_n}\right). \tag{8.45a}$$

Sometimes we use a shortened notation for the kernel function

$$K_b(x-x_i) := \frac{1}{b_n} \, K\left(\frac{x-x_i}{b_n}\right) \tag{8.45b}$$

resulting in

$$\widehat{h}_n(x) = \sum_{i=1}^{n} \frac{1}{n-i+1} \, K_b(x-x_i). \tag{8.45c}$$

The failure times are assumed to be increasing order $x_1 < x_2 < \ldots < x_n$. Thus, the mean or expectation of $\widehat{h}_n(x)$ is

$$\mathrm{E}\Big[\widehat{h}_n(x)\Big] = \sum_{i=1}^{n} \int\limits_0^\infty \frac{1}{n-i+1} \, K_b(x-u) \, f_{i:n}(u) \, du \tag{8.46a}$$

where

$$f_{i:n}(u) = \frac{n!}{(i-1)! \, (n-i)!} \, F(u)^{i-1} \big[1-F(u)\big]^{n-i} \, f(u) \tag{8.46b}$$

is the PDF of the i-th order statistic. Upon inserting (8.46b) into (8.46a) and changing the order of summation and integration we have

$$\mathrm{E}\Big[\widehat{h}_n(x)\Big] = \int\limits_0^\infty \sum_{i=1}^{n} \binom{n}{i-1} F(u)^{i-1} \big[1-F(u)\big]^{n-i} \, K_b(x-u) \, f(u) \, du. \tag{8.46c}$$

As

$$\sum_{i=1}^{n} \binom{n}{i-1} F(u)^{i-1} \big[1-F(u)\big]^{n-i} = \frac{1-F(u)^n}{1-F(u)}$$

and observing $h(x) = f(x)/(1 - F(x)$ we finally arrive at

$$\begin{aligned}
\mathrm{E}\Big[\widehat{h}_n(x)\Big] &= \int\limits_0^\infty \big[1 - F(u)^n\big]\, h(u)\, K_b(x-u)\, du \\
&= \int\limits_0^\infty K_b(x-u)\, h(u)\, du - \int\limits_0^\infty F(u)^n\, h(u)\, K_b(x-u)\, du. \qquad (8.46d)
\end{aligned}$$

If x is such that $F(x) < 1$, the second term tends to zero geometrically as $n \to \infty$. If $f(x)$ is twice continuously differentiable the first term of (8.46d) can be expanded in a TAYLOR series analogous to (8.14a,b):

$$\begin{aligned}
\int\limits_0^\infty \frac{1}{b_n} K\Big(\frac{x-u}{b_n}\Big) \frac{f(u)}{1-F(u)} &= \int\limits_0^\infty K(v) \frac{f(x-b_n\, v)}{1-F(x-b_n\, v)}\, dv \\
&= \frac{f(x)}{1-F(x)} + \frac{b_n^2}{2}\left[\frac{f(x)}{1-F(x)}\right]'' \int\limits_0^\infty K(v)\, v\, dv + o\big(b_n^2\big). \qquad (8.46e)
\end{aligned}$$

In (8.46e) we see that the bias of $\widehat{h}_n(x)$ depends on the second derivative of $h(x) = f(x)/[1 - F(x)]$, but $\widehat{h}_n(x)$ is asymptotically unbiased.

WATSON/LEADBETTER (1964a,b) give the exact variance of $\widehat{h}_n(x)$ as

$$\begin{aligned}
\mathrm{Var}\Big[\widehat{h}_n(x)\Big] &= \int\limits_0^\infty \frac{K^2(x-u)}{1-F(u)}\, I_n\big[F(u)\big]\, du + \\
&2 \iint\limits_{0\le u_1\le u_2<\infty} \frac{K(x-u_1)K(x-u_2)}{1-F(u_2)} \left\{\frac{1-F(u_1)^n}{1-F(u_1)} F(u_2)^n - \frac{F(u_2)^n - F(u_1)^n}{F(u_2)-F(u_1)}\right\} dF(u_1)dF(u_2) \qquad (8.46f)
\end{aligned}$$

where

$$I_n(F) = \int\limits_0^{1-F} \frac{(F-B)^n - F^n}{B}\, dB.$$

Formula (8.46f) is difficult to appraise for finite n, but as $n \to \infty$ only the first term needs to be considered with the result that

$$\mathrm{Var}\Big[\widehat{h}_n(x)\Big] \approx \frac{1}{n\, b_n} \frac{h(x)}{1-F(x)} \int K^2(u)\, du. \qquad (8.46g)$$

This variance formula is similar in construction to that of the PDF kernel estimator given in (8.17a). Furthermore, at each continuity point of $h(x)$ the estimator is asymptotically normally distributed.

We now look at the properties of $\widehat{h}_n(x)$ when there is *censoring*. We assume a random censorship model and observe $Y_i = \min(X_i,\, Z_i)$ together with $\delta_i = I_{(X_i \le Z_i)}$. The $X_i;\ i = 1, 2, \ldots, n$; are i.i.d. lifetimes with CDF $F_X(\cdot)$ and PDF $f(\cdot)$. The X_i are independent of the censoring variables $Z_1, \ldots, Z_n$ each having CDF $F_Z(\cdot)$ and PDF $f_Z(\cdot)$. The CDF of Y_i will be denoted $F_Y(\cdot)$ and is given via

$$1 - F_Y(t) = \big[1 - F_X(t)\big]\, \big[1 - F_Z(t)\big] \qquad (8.47a)$$

and the PDF of the Y_i follows as

$$f_Y(t) = f_X(t)\, \big[1 - F_Z(t)\big] + f_Z(t)\, \big[1 - F_X(t)\big]. \qquad (8.47b)$$

The observations y_i are assumed increasingly ordered together with the corresponding indicator δ_i.

Let

$$m(t) = \frac{f_X(t)\,\big[1 - F_Z(t)\big]}{f_Y(t)} \;\text{ if }\; f_Y(T) > 0, \tag{8.48a}$$

then, see TANNER/WONG (1983),

$$\begin{aligned} \mathrm{E}(\delta_i \big| Y_i = t) &= m(t) && (8.48b)\\ \mathrm{E}(\delta_i\,\delta_j \big| Y_i = t,\, Y_j = s) &= m(t)\,m(s)\;\forall\; i < j,\; t < s. && (8.48c) \end{aligned}$$

Using (8.48b,c) the derivation of the formulas for $\mathrm{E}\Big[\widehat{h}_n(x)\Big]$ and for $\mathrm{Var}\Big[\widehat{h}_n(x)\Big]$ proceeds in essentially the same way as in WATSON/LEADBETTER (1964a,b) for the uncensored case. To illustrate the idea,

$$\begin{aligned} \mathrm{E}\Big[\widehat{h}_n(x)\Big] &= \sum_{i=1}^{n} \int\limits_0^\infty \frac{\mathrm{E}(\delta_i \big| Y_i = u)}{n - i + 1}\, K_b(x-u)\; {}_Y f_{i:n}(u)\,du \\ &= \int\limits_0^\infty \left[\sum_{i=1}^{n} \binom{n}{i-1} F_Y(u)^{i-1}\big[1 - F_Y(u)\big]^{n-i}\right] f_Y(u)\,m(u)\,K_b(x-u)\,du \\ &= \int\limits_0^\infty \frac{1 - F_Y(u)^n}{1 - F_Y(u)}\, f_X(u)\big[1 - F_Z(u)\big]K_b(x-u)\,du \\ &= \int\limits_0^\infty \big[1 - F_Y(u)^n\big]\, h_X(u)\,K_b(x-u)\,du, \end{aligned} \tag{8.49}$$

observing (8.47a), (8.48a) and $h_X(u) = f_X(u)\big/[1 - F_X(u)]$. The only difference between (8.49) and (8.46d) is that we have to use the CDF of Y instead of the CDF of X in the censoring case. The asymptotic variance ($n \to \infty$) in the censoring case is similar to (8.46g):

$$\mathrm{Var}\Big[\widehat{h}_n(x)\Big] \approx \frac{1}{n\,b_n}\, \frac{h_X(x)}{[1 - F_X(x)]\,[1 - F_Z(x)]} \int K^2(u)\,du. \tag{8.50}$$

The rate of convergence of the kernel hazard rate estimator

$$\widehat{h}_n(x) = \frac{1}{b_n} \sum_{i=1}^{n} \frac{\delta_i}{n - i + 1}\, K\bigg(\frac{x - y_i}{b_n}\bigg)$$

depends on the order of the kernel, the bandwidth and the differentiability of the hazard rate. Typically, the order of the kernel is chosen to be an even number with $k = 2$ being the standard choice.The resulting bias and variance are

$$\begin{aligned} \mathrm{Bias}\Big[\widehat{h}_n(x)\Big] &= b_n^k \left[h^{(k)}(x)\,\frac{(-1)^k}{k!} \int u^k\,K(u)\,du + o(1)\right] && (8.51a)\\ \mathrm{Var}\Big[\widehat{h}_n(x)\Big] &= \frac{1}{n\,b_n} \left\{\frac{h(x)}{[1 - F_X(x)]\,[1 - F_Z(X)]} \int K^2(u)\,du + o(1)\right\} && (8.51b) \end{aligned}$$

The influence of the bandwidth b_n and the trade–off between the bias and the variance is seen from (8.51) and (8.51b). The optimal rate for the MSE of $\widehat{h}_(x)$ is attained when the squared bias and the variance are of the same order. This results in an optimal MSE rate of convergence of

$n^{2k/(2k+1)}$, which is $n^{4/5}$ for the standard choice of $k = 2$. This rate is slower than the usual rate of n regardless of the order k. For the asymptotic distribution we further assume that

$$d = \lim_{n \to \infty} n\, b_n^{2k+1}$$

exists for some $0 \le d < \infty$. Then

$$\frac{\widehat{h}_n(x) - h(x)}{\sqrt{n\, b_n}} \xrightarrow{\text{D}} No\left[h^k(x)\,\frac{(-1)^k}{k!};\frac{h(x)}{[1-F_X(x)]\,[1-F_Z(x)]}\int K^2(u)\,du\right]. \tag{8.51c}$$

Extensions to the estimation of derivatives $h^{(k)}(x)$ of the hazard function can be found in MÜLLER/WANG (1990b). These essentially involve a change in the kernel. Derivatives are of interest to detect rapid changes in the hazard rate or for data based bandwidth choice.

To find the **optimal global bandwidth**, we have to restrict the range of x to a compact interval $[0, \tau]$ with $F_X(\tau) < 1$ and $F_Z(\tau) < 1$. The global optimal bandwidth which minimizes the leading term of

$$\text{MISE}\Big[\widehat{h}_n(x)\Big] = \text{E}\bigg\{\int_0^\tau \Big[\widehat{h}_n(u) - h(u)\Big]^2 du\bigg\} \tag{8.52a}$$

is

$$b_{opt} = \left\{\frac{1}{n\,k}\int_0^\tau \frac{h(u)}{[1-F_X(u)][1-F_Z(u)]}du\,\frac{\int K^2(u)\,du}{\Big[\frac{(-1)^k}{k!}\int_0^\tau u^k K(u)du\Big]\int_0^\tau\big[h^k(u)\big]^2du}\right\}^{1/(2k+1}. \tag{8.52b}$$

This optimal global bandwidth involves unknown quantities, so that in practice one has to find alternatives. There is an extensive literature on bandwidth selection, see Sect. 8.1.1.4.

One way to pick a good bandwidth is to use a cross–validation technique for determining the bandwidth that minimizes some measure of how well the estimate performs. One such measure is the mean integrated squared error (MISE) of $\widehat{h}_n(x)$ over the range 0 to τ, see (8.52a), defined by

$$\begin{aligned}\text{MISE}\Big[\widehat{h}_n(x)\Big] &= \text{E}\bigg\{\int_0^\tau \Big[\widehat{h}_n(u) - h(u)\Big]^2 du\bigg\}\\ &= \text{E}\bigg\{\int_0^\tau \widehat{h}_n^2(u)\,du\bigg\} - 2\,\text{E}\bigg\{\int_0^\tau \widehat{h}_n(u)\,h(u)\,du\bigg\} + \text{E}\bigg\{\int_0^\tau h^2(u)\,du\bigg\}.\end{aligned} \tag{8.53a}$$

This function depends both on the kernel used to estimate $h(\cdot)$ and on the bandwidth b_n. Note that, although the last term in (8.53a) depends on the unknown hazard rate, it is independent of the choice of the kernel and of the bandwidth and can be ignored when finding the best value of b_n. The first term of (8.53a) can be estimated by $\int\limits_0^\tau \widehat{h}_n^2(u)\;du$. If we evaluate $\widehat{h}_n(\cdot)$ at a grid of points $0 < u_1 < \ldots < u_\ell = \tau$, then, we find an approximation to this integral by some formula of numerical integration over $\widehat{h}_n(\cdot)$, e.g., by the trapezoid rule as

$$\int_0^\tau \widehat{h}_n^2(u)\,du \approx \sum_{i=1}^{\ell-1}\frac{u_{i+1}-u_i}{2}\Big[\widehat{h}_n^2(u_i) + \widehat{h}_n^2(u_{i+1})\Big]. \tag{8.53b}$$

The second term of (8.53a) can be estimated by a cross–validation estimate suggested by RAMLAU–HANSEN (1983). This estimate is

$$\widehat{\text{E}}\bigg\{\int_0^\tau \widehat{h}_n(u)\,h(u)\,du\bigg\} = \frac{1}{b_n}\sum_{i\neq j}K\left(\frac{y_i-y_j}{b_n}\right)\frac{\delta_i}{n-i+1}\,\frac{\delta_j}{n-j+1} \tag{8.53c}$$

where the sum is over all observed times between 0 and τ. Thus, to find the best value of b_n which minimizes the MISE for a fixed kernel, we find b_n which minimizes the function

$$g(b_n) = \sum_{i=1}^{\ell-1} \frac{u_{i+1}-u_i}{2} \left[\widehat{h}_n^2(u_i)+\widehat{h}_n^2(u_{i+1})\right] - \frac{2}{b_n} \sum_{i \neq j} K\left(\frac{y_i-y_j}{b_n}\right) \frac{\delta_i}{n-i+1} \frac{\delta_j}{n-j+1}. \tag{8.53d}$$

One also has to find alternatives to calculate the variance (8.51b) which contains unknown quantities. One possibility is to use $\widehat{h}_n(x)$ for $h(x)$, $\exp\Big\{-\int\limits_0^x \widehat{h}_n(u)\,du\Big\}$ for $1 - F_X(x)$, where the integral may be evaluated by the trapezoid rule, and to neglect $1 - F_Z(x)$ what will lead to an underestimation of $\text{Var}\big[\,\widehat{h}_n(x)\big]$ for censored samples. Another possibility has been suggested by KLEIN/MOESCHBERGER (1997, p. 153). Since the kernel–smoothed estimator is a linear combination of the increments of the cumulated hazard rate

$$\widehat{h}_n(x) = \frac{1}{b_n} \sum_{i=1}^{n} K\left(\frac{x-y_i}{b_n}\right) \Delta\widehat{H}_n(y_i)$$

with

$$\Delta\widehat{H}_n(y_i) = \frac{\delta_i}{n-i+1}$$

a crude estimator of $\text{Var}\big[\,\widehat{h}_n(x)\big]$ follows as

$$\widehat{\text{Var}}\big[\,\widehat{h}_n(x)\big] = \frac{1}{b_n^2} \sum_{i=1}^{n} K\left(\frac{x-y_i}{b_n}\right) \Delta\widehat{\text{Var}}\big[\widehat{H}_n(y_i)\big] \tag{8.54a}$$

with $\widehat{\text{Var}}\big[\widehat{H}_n(x)\big]$ given in (5.12b) so that

$$\Delta\widehat{\text{Var}}\big[\widehat{H}_n(y_i)\big] = \frac{\delta_i\,(n-i)}{(n-i+1)^3}. \tag{8.54b}$$

Example 8/2: Fixed–bandwidth kernel estimation of the hazard rate

The following data are from BYSON/SIDDIQUI (1961) and represent the ordered times (in days) at death of $n = 43$ patients suffering from chronic granulocytic leukemia with $x = 0$ taken as the patient's date of diagnosis. This is an uncensored sample.

7	47	58	74	177	232	273	285	317	429
440	445	455	468	495	497	532	571	579	581
650	702	715	779	881	900	930	968	1077	1109
1314	1334	1367	1534	1712	1784	1877	1886	2045	2056
2260	2429	2509							

Fig. 8/7 shows the hazard rate estimated with four different kernels. The grid has 100 evenly spaced points between 0 and 2509 in each case and the bandwidth is $b_n = 250$ for all cases. The resulting hazard rate is nearly constant for $0 < x < 1600$ and is increasing for $x \geq 1600$. With the exception of the uniform kernel the other three kernels produce nearly the same hazard rates.

In Fig. 8/8 we find four estimated hazard rates, each using the EPANECHNIKOV kernel with a grid of 100 evenly spaced points between 0 and 2509, and bandwidths 600, 300, 200, 100, respectively. $b_n = 600$ gives an oversmoothed estimate whereas $b_n = 100$ produces a rather erratic course. $b_n = 300$ and $b_n = 200$ seem to be good compromise. The computations for this example have been done by using the MATLAB program *Hazard_04* of the appended software package Inference.zip.

Figure 8/7: Hazard rate estimates for the survival time of 43 patients having granulocytic leukemia — Different kernels and common bandwidth $b_n = 250$

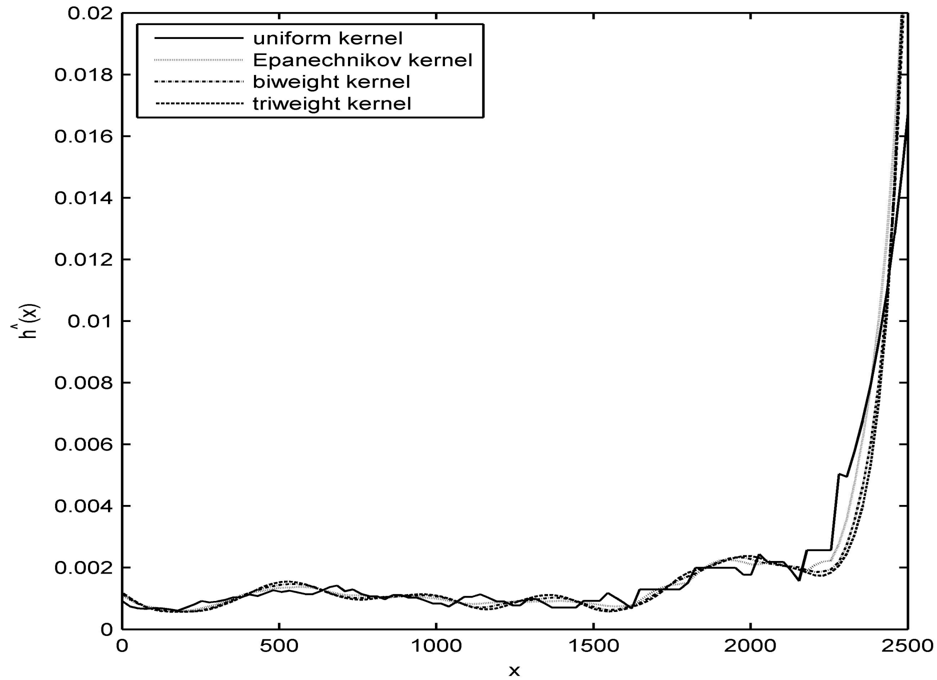
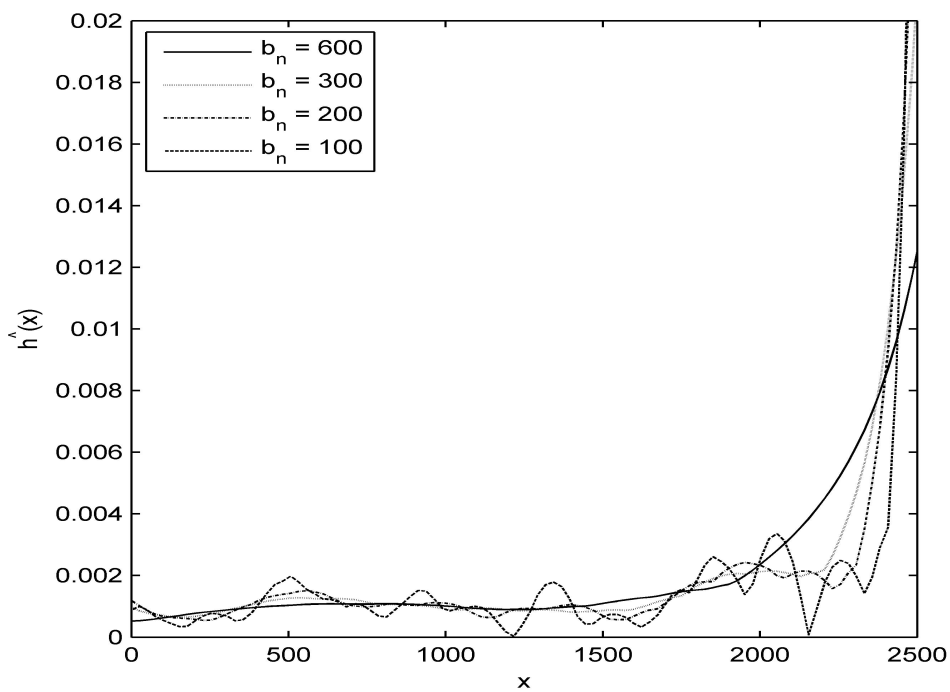


Figure 8/8: Hazard rate estimates for the survival time of 43 patients having granulocytic leukemia — Different bandwidths and common EPANECHNIKOV kernel



Up to now we have presented results and formulas for samples with untied observations. The key formulas have to be modified properly when the sample has *tied observations* where the ties may occur among the uncensored as well as among the censored observations. We have the following notation, also see Fig. 4/1:

$x_j;\ j = 1, 2, \ldots, k$ — distinct failure times (= uncensored observations),

$d_j \geq 1$ — number of failures at x_j,

$c_j \geq 0;\ j = 0, 1, \ldots, k-1$ — number of censored observations in the right–opened interval $[x_j, x_{j+1})$ where $x_0 = 0$,

$n_j = n_{j-1} - c_{j-1} - d_{j-1};\ j = 1, 2, \ldots, k$ — number of units at risk at x_j where $n_0 = n$ and $d_0 = 0$,

$\widehat{h}_j = \dfrac{d_j}{n_j}$ — MLE for $h_j = \dfrac{f(x_j)}{S(x_j)}$, see (5.2e),

$\widehat{\text{Var}}\big(\widehat{h}_j\big) = \dfrac{d_j\,(n_j - d_j)}{n_j^3}$, see (5.7c)

The fixed–bandwidth kernel estimator of (8.44) turns into

$$\widehat{h}_n(x) = \frac{1}{b_n} \sum_{j=1}^{k} \frac{d_j}{n_j}\, K\left(\frac{x - x_j}{b_n}\right) \tag{8.55a}$$

and the crude estimator of $\text{Var}\big[\widehat{h}_n(x)\big]$ in (8.54a) now reads

$$\widehat{\text{Var}}\big[\widehat{h}_n(x)\big] = \frac{1}{b_n^2} \sum_{j=1}^{k} K^2\left(\frac{x - x_j}{b_n}\right) \frac{d_j\,(n_j - d_j)}{n_j^3} \tag{8.55b}$$

and the goal function (8.53d) is

$$g(b_n) = \sum_{i=1}^{\ell-1} \frac{u_{i+1} - u_i}{2} \left[\widehat{h}_n^2(u_i) + \widehat{h}_n^2(u_{i+1})\right] - \frac{2}{b_n} \sum_{i \neq j} K\left(\frac{x_i - x_j}{b_n}\right) \frac{d_i}{n_i}\,\frac{d_j}{n_j}. \tag{8.55c}$$

8.1.3.3 Kernel Estimators with Varying Bandwidth[16]

The fixed–bandwidth kernel estimator of the hazard rate has many good properties, e.g., asymptotic unbiasedness, mean square error consistency, asymptotic normality, but in practical application it has been observed that a globally constant bandwidth leads to undesirable effects whenever the data are not evenly distributed over the whole range of interest. The fixed–bandwidth kernel estimator cannot adopt to unevenness in the distribution of the data and thus tends to oversmooth in regions with many observations and to undersmooth in regions with few observations revealing many misleading peaks. One approach to overcome these problems of the fixed–bandwidth estimator is to incorporate the idea of **nearest neighbor** into the definition of the bandwidth. The resulting estimator has a bandwidth which adapts to the configuration of the data. There are two estimators emerging from this idea depending on what is the point from where to look for the k-th nearest neighbor.

[16] Suggested reading for this section: BAGKAVOS/PATIL (2009), CHENG (1987), DETTE/GEFELLER (1995), GEFELLER/DETTE (1991, 1992), GEFELLER/MICHELS (1992), HESS et al. (1999), MÜLLER/WANG (1990b, 1994), NIELSEN (2003), SCHÄFER (1985, 1986), TANNER (1983, 1984), TANNER/WANG (1984).

1. The **local kernel estimator** takes the distance from x, the point of interest where to estimate $h(x)$, to its k-th nearest neighbor among the observations as bandwidth.
2. The **variable kernel estimator** defines the bandwidth as distance from X_i, the i-th uncensored observation in ascending order to its k-th nearest neighbor among the observations.

There are questions to be answered for each of these two approaches:

- How to choose k? — FAILING (1984), based on published experience, suggests to specify k as
$$k \approx c\sqrt{n} \;\text{ with }\; 1 \leq c \leq 2.5. \tag{8.56}$$
TANNER (1984) finds k by a time–consuming cross–validation method.[17]
- How to find the k-th nearest neighbor when there are censored observations? — This question has been answered in different ways as will be shown further down.

We first turn to the *local kernel estimator for* a sample of n *uncensored and distinct observations* $X_1,\, X_2,\, \ldots,\, X_n$. In this case, the kernel estimator of $h(x)$ is defined as

$$\widehat{h}(x) = \sum_{i=1}^{n} \frac{1}{n-i+1}\,\frac{1}{D(k_n,x)}\,K\left(\frac{x-X_i}{D(k_n,x)}\right), \tag{8.57}$$

where $D(k_n,x)$ is the the distance of x to its k-th nearest neighbor among $X_1,\, X_2,\, \ldots,\, X_n$. DETTE/GEFELLER (1995) have derived the asymptotic MISE of this estimator. TANNER (1983) has shown that the estimator (8.57) converges (almost surely) to $h(\cdot)$ at each point of continuity of $h(\cdot)$, provided that the sequence k_n fulfills the condition $k_n = n^{\alpha},\ \alpha \in (0.5, 1)$. Other versions of the estimator $\widehat{h}(x)$ have been considered by LIU/VAN RYZIN (1985) and CHENG (1987). These authors have proved asymptotic normality and strong consistency. Mathematical disadvantages of (8.57) are due to the fact that the bandwidth function by the k-th nearest neighbor distance is not differentiable at all $x > 0$ and that the integral from 0 to ∞ over (8.57) is not bounded in general.

The problem of transferring appropriately the definition of $D(k_n,x)$ from the uncensored to the *censored setting* has been solved in different ways.

1. A straightforward and simple solution — as suggested in TANNER (1983), TANNER/WONG (1984), LIU/VAN RYZIN (1985) and CHENG (1987) — ignores all censored observations when looking for the k-th neighbor and defines $D_1(k_n,x)$ as the distance of x to its k-th neighbor among the uncensored data $(Y_i, \delta_i = 1);\ i = 1,\, 2,\, \ldots,\, n$:
$$\widehat{h}(x) = \sum_{i=1}^{n} \frac{\delta_i}{n-i+1}\,\frac{1}{D_1(k_n,x)}\,K\left(\frac{x-Y_i}{D_1(k_n,x)}\right). \tag{8.58}$$
SCHÄFER (1985) points out that these distances are biased by the censoring distribution in the sense that they adapt to the conditional density of X_i under the condition $X_i \leq Z_i$ of being uncensored rather than to the density function $f(\cdot)$ or the hazard rate to be estimated.
2. Therefore, SCHÄFER (1985) proposes an alternative definition of $D(k_n,x)$:
$$D_2(k_n,x) = \sup\left\{d > 0 \,\middle|\, \widehat{H}_n(x+d) - \widehat{H}_n(x-d) \leq \frac{k_n-1}{n}\right\} \tag{8.59a}$$

[17] The paper of TANNER gives a FORTRAN–code for the variable kernel estimator of the hazard rate.

resulting in

$$\widehat{h}(x) = \sum_{i=1}^{n} \frac{\delta_i}{n-i+1}\,\frac{1}{D_2(k_n,x)}\,K\left(\frac{x-Y_i}{D_2(k_n,x)}\right). \tag{8.59b}$$

This definition of $D(k_n,x)$ incorporates the information of the censored observations by using the NELSON/AALEN estimator (5.12a) of the cumulated hazard rate, and the nearest neighbor distances are no longer biased by the censoring distribution. It suffers, however, as GEFELLER/DETTE (1992) and DETTE/GEFELLER (1995) criticize, from other serious conceptual drawbacks:

a) "Even if no censored data were observed in the sample, $D_2(k_n,x)$ is not identical to $D(k_n,x)$ of (8.57), i.e., $D_2(k_n,x)$ does not reveal the natural definition of nearest neighbor distances in the uncensored setting."

b) "In addition, one inherent property of $\widehat{H}_n(\cdot)$ has an awkward effect on the definition of nearest neighbor distances: the heights of the steps in $\widehat{H}_n(\cdot)$ increase automatically by definition as $x \to Y_n$. Consequently, in the right tail of the lifetime distribution this effect dominates the value $(k_n-1)/n$ used in the definition of $D_2(\cdot,\cdot)$."

3. To avoid the problems mentioned in a) and b) above, the authors DETTE and GEFELLER propose a modification of SCHÄFER's idea defining

$$D_3(k_n,x) = \sup\left\{d>0 \,\middle|\, \widehat{S}_n(x-d) - \widehat{S}_n(x+d-0) \le \frac{k_n-1}{n}\right\} \tag{8.60}$$

where $\widehat{S}_n(x+d-0)$ denotes the limit from the left of the KAPLAN/MEIER estimator of $S(\cdot)$ — see (5.3d) — at the point $x+d$. Using $\widehat{S}_n(\cdot)$ instead of $\widehat{H}_n(\cdot)$ resolves the drawbacks of $D_2(\cdot,\cdot)$.

In an intensive simulation study GEFELLER/DETTE (1992) found that the kernel hazard rate estimator based an $D_3(\cdot,\cdot)$ always had a smaller MISE than that based on $D_1(\cdot,\cdot)$ or $D_2(\cdot,\cdot)$.

We now turn to the *variable kernel estimator of the hazard rate* which has been proposed by TANNER/WONG (1984) using a bandwidth which is the distance $D(k_n,X_i)$ between the observation X_i and its k-th nearest neighbor among the remaining uncensored observations:

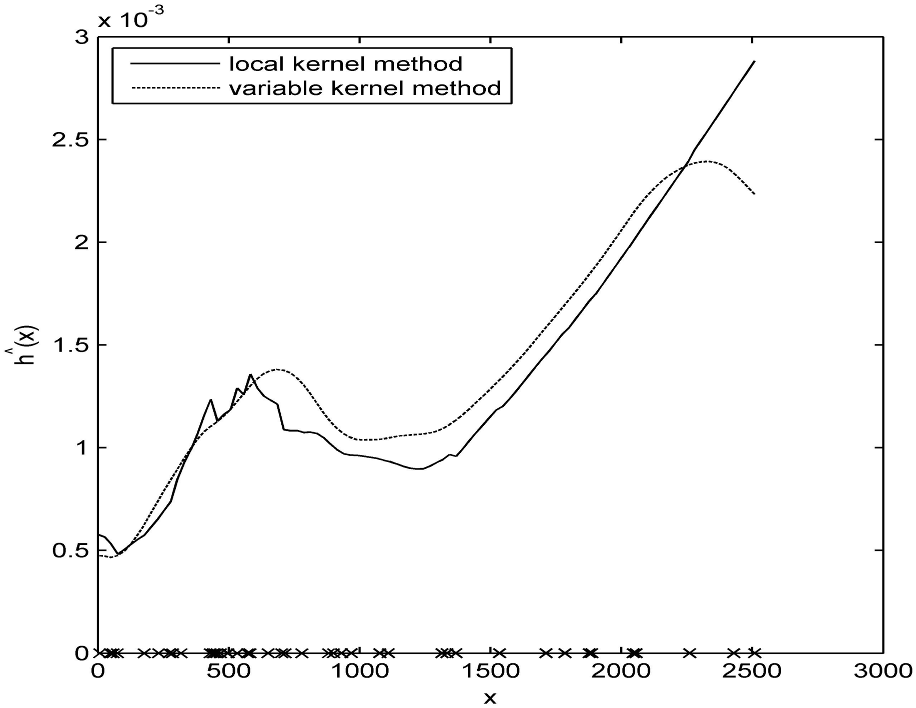
$$\widehat{h}_n(x) = \sum_{i=1}^{n} \frac{1}{n-i+1}\,\frac{1}{D(k_n,X_i)}\,K\left(\frac{x-X_i}{D(k_n,X_i)}\right), \tag{8.61}$$

assuming a sample with uncensored data. This definition of the bandwidth is independent of the points of interest x and adapts only to the configuration of the data. The kernel estimator (8.61) is differentiable for appropriate kernel functions $K(\cdot)$ at all $x>0$ and its integral is bounded. We mention that contrary to the hazard rate estimator with fixed bandwidth, the bandwidth here is not globally constant and contrary to the local kernel estimators (8.57) – (8.59) the number of observations influencing $\widehat{h}(x)$ is not fixed either. Statistical properties of (8.61), e.g., uniform convergence to $h(x)$, have been derived by SCHÄFER (1985).

Example 8/3: Local and variable kernel estimators of the hazard rate

Using the data of the preceding Example 8/2 Fig. 8/9 shows the smoothed hazard estimated by the local and the variable kernel method, respectively, taking the biweight kernel with $k=13$ and 100 gridpoints. Evidently and ceteris paribus the variable kernel method produces a smoother curve. The computations have been done with the help of the MATLAB programs *Hazard_05* and *Hazard_06* of the appended software package Inference.zip.

Figure 8/9: Hazard rate estimates for the survival time of 43 patients having granulocytic leukemia using a local and a variable biweight kernel with $k = 13$ and 100 gridpoints



A kernel–based estimator with varying bandwidths can be found by optimizing the bandwidth for a given point of interest x. For kernel estimators, the bandwidth regulates the trade–off between local bias and local variance. Thus, another way to overcome the non–adaptive behavior of the fixed–bandwidth estimator is to vary the local bandwidth in order to balance local variance and local bias. A natural objective function is the local MSE expressing the estimation error as a function of local bias and local variance. An **optimal local bandwidth kernel estimator** can be found by minimizing an estimate of the local MSE with respect to the bandwidth. This problem has been solved by MÜLLER/WANG (1990b, 1994) who also give a MATLAB program named HADES for doing this job.[18]

8.2 Further Smoothing Techniques[19]

There are two serious competitors to the kernel approach when looking for smooth hazard rate estimators, the spline approach and the wavelet approach, both being not very popular in practice. We will shortly comment on these two techniques without going into details, starting with the **spline estimator**.

In mathematics a spline[20] is a sufficiently smooth polynomial function that is piecewise defined and possesses a high degree of smoothness at the places where the polynomial pieces connect.

[18] See www.stat.ucdavis.edu/~ntyang/hades/. Other programs — written in FORTRAN or S–plus — performing different approaches in kernel estimation of the hazard rate can be found on the website htpp://odin.mdacc.tmc.edu of HESS et al. (1999).

[19] Suggested reading for this section: ANDERSON/SENTHILSELVAN (1980), ANTONIADIS et al. (1994, 1999), BÉZANDRY et al. (2005), DUPUY/GNEYOU (2011), GU (1996), JARJOURA (1988), LI (2002), O'SULLIVAN (1988a,b), PATIL (1997), ROSENBERG (1995), WU/WELLS (2003).

[20] The term 'spline' is adapted from the name of a flexible strip of metal used by draftsmen to assist in drawing curved lines. They are very popular among naval architects in designing a ship's hull.

These places are called knots and in hazard rate estimation they — generally — are the observed lifetimes X_i. The most commonly used splines are cubic splines, i.e., they consist of polynomials of order 3 and their first derivatives coincide at the knots so that there are no sharp edges in the curvature.

There are several types of spline methods, the most widely investigated spline method for hazard smoothing is the penalized likelihood approach. Let $\eta(x) = \ln h(x)$ be the log hazard rate. Then the log likelihood function for the censored data is

$$\mathcal{L}(\eta) = \sum_{i=1}^{n} \left\{ \delta_i\, \eta(Y_i) - \int\limits_0^{Y_i} e^{\eta} \right\}, \tag{8.62a}$$

which is unbounded if no shape restriction on $\eta(\cdot)$ is imposed. A penalty $P(\eta)$, measuring the roughness of $\eta(\cdot)$, is therefore introduced in (8.62a). The penalized estimator $\widehat{\eta}(\cdot)$ of $\eta(\cdot)$ is the maximum of the penalized log likelihood

$$\mathcal{L}(\eta) = \frac{1}{n} \sum_{[i=1}^{n} \left\{ \delta_i\, \eta(Y_i) - \int\limits_0^{Y_i} e^{\eta} \right\} - \frac{\alpha}{2}\, P(\eta) \tag{8.62b}$$

among all $\eta(\cdot)$ in a HILBERT space. α is a smoothing parameter playing the same role as the bandwidth b in kernel estimation. A smaller α yields a better fit bit a more rough curve. JARJOURA (1988) describes how to determine the smoothing parameter α by a cross–validation likelihood approach.

The penalty function $P(\eta)$ determines the kind of spline resulting from (8.62b). With

$$P(\eta) = \int \left[\eta^{(2)}(x)\right]^2 dx \tag{8.62c}$$

we find an estimator which is twice continuously differentiable and a piecewise cubic polynomial between two consecutive X_i's. For computing details see the papers of O'SULLIVAN (1988a,b) and for asymptotic results see GU (1996). ANDERSON/SENTHILSELVAN (1980) take

$$P(\eta) = \int \left[h'(x)\right]^2 dx \tag{8.62d}$$

which leads to $\widehat{h}(x)$ as a piecewise quadratic spline that may result in negative values under heavy censoring. Another type of spline method is regression splines or B–splines which adopt a fixed number of knots and basis functions. ROSENBERG (1995) describes how to select the number and the location of the knots.

Wavelet–based hazard rate estimation has been treated by many authors for a long time. A wavelet is a wave–like oscillation with an amplitude that begins at zero, increases, and then decreases back to zero. For instance, it can be visualized as a brief oscillation like one might see recorded by a seismograph or heart monitor.

Let ϕ be the so–called 'father wavelet' and ψ the 'mother wavelet'.[21] We assume that the functions

$$\phi_{j,k}(x) := 2^{j/2}\, \phi\big(2^j\, x - k\big) \ \text{ and } \ \psi_{j,k}(x) := 2^{j/2}\, \psi\big(2^j\, x - k\big)$$

satisfy the following conditions:

1. $\big\{\phi_{j,k},\, \psi_{j,k} \,:\, k \in \mathbb{Z},\, j \ge j_0\big\}$ is an orthogonal basis of $L^2(\mathbb{R})$,

[21] The term 'father' (respectively 'mother') comes from the fact that the functions $\phi_{j,k}$ (respectively $\psi_{j,k}$) are derived by dilations and translations from the original function ϕ (respectively ψ).

2. ϕ and ψ are bounded and compactly supported,

3. $\int y^k\,\psi(y)\,dy = 0$ for $0 \le k \le r-1$ and $\kappa = (r!)^{-1}\int y^r \psi(y)\,dy = 0$.

This implies that an arbitrary square–integrable function may be expanded in a generalized FOURIER series.

Readers who are interested in the rather complicated theory and computation of wavelet–based hazard rate estimation should consult one of the original papers on this subject: ANTONIADIS et al. (1994, 1999), BÉZANDRY et al. (2005), DUPUY/GNEYOU (2011), LI (2002), PATIL (1997), WU/WELLS (2003). An introduction in wavelets and their statistical applications is HÄRDLE et al. (1998).

9 Hazard Plotting

Hazard plotting allows estimation as well as informal testing of hypotheses. So this chapter marks the transition from one topic of statistical inference to another one.

9.1 Introduction and Motivation

Data plotting, especially probability plotting, has been applied for a long time by engineers to display and interpret failure data because of its simplicity and effectiveness. Data plotting is often used in place of or in addition to standard numerical methods of data analysis because it serves a lot of purposes:

- A plot provides a complete and easy–to–grasp picture of data according to an old Chinese proverb saying that one picture is worth a thousand words, or in the present context, a thousand numbers. A plot is particular useful in presenting data, since it aids in convincing others of conclusions drawn from data by numerical methods.
- A plot provides a convenient means of fitting a theoretical distribution to data. This can can be done by drawing a straight line by eye through the plotted data points on specialized graph paper. This line is used to smooth, interpolate and extrapolate data. Estimates of distribution parameters, percentiles, and predictions of number of failed and unfailed units in specified periods of time are easily obtained from this straight line.
- A plot allows one to assess whether a chosen theoretical distribution provides an adequate fit to the data or not. The data points will tend to plot as a straight line on the plotting paper for a satisfactory distribution. Non–random departures of the plotted data from a straight line can provide useful information to the statistician. Such departures may indicate that the chosen distribution is incorrect, that there is more than one failure mode, or that certain data points are outliers that do not fit in with the rest of the data.

The object of plotting in this chapter is the cumulative hazard function $H(x)$. For this purpose we first have to find estimates of $H(x)$, see Sect. 5.2., which then will be displayed either on normal (= naturally or linearly scaled) graph paper or on specialized graph paper as presented in Sect. 9.2. The plot on normal graph paper serves to make non–parametric inference whereas the plot on special hazard paper aims at making inference on a hypothetical parametric lifetime distribution. We close this chapter by a section on how to find the appropriate hazard–scale for distributions belonging to the location–scale family.

9.2 Hazard Plots and Hazard Paper[1]

Only little knowledge can be gained from line plots of the direct estimator $\widehat{H}(x)$ or of the indirect estimator $\widetilde{H}(x)$ as given in Fig. 5/2. When the left–hand edges of the stair–case plot form a nearly convex (concave) line we may guess that the sampled distribution is IHR (DHR) and when they scatter around a straight line we may have a sample from an exponential distribution. More insight can be found when we make a point–plot on hazard paper. The basic idea is to make plots that should be roughly linear if the proposed family of distributions seems to have generated the sample at hand, since departures from linearity can readily be appreciated by eye.

[1] Suggested reading for this section: NELSON (1969, 1970, 1972, 1982), RINNE (2010).

Hazard plotting — like probability plotting — can be successfully applied to location–scale distributions[2] and to those distributions that after suitable transformation can be converted into a location–scale type. Hazard plots and probability plots are closely related to one another, the main difference is the scaling of the ordinate where to lay down the CHR–values instead of the CDF–values and the choice of the plotting position, i.e., the ordinate–value to be plotted against the ordered sample values x_i on the abscissa. Each member of the location–scale family has an ordinate–scaling of its own, distorted in such a way that, when the sample comes from the pertinent distribution, the plotted points on the graph paper will randomly scatter around a straight line, thus giving a graphical and informal goodness–of–fit test. The fitted line — either fitted by eye or by regression — enables the statistician to read off estimates of the location parameter and of the scale parameter, respectively, either as point or as an interval on the abscissa.

We will first demonstrate how to construct a hazard paper and how to find estimates of the location and scale parameters for a genuine location–scale distribution and for distributions transformed to location–scale type. Then we comment on the choice of the plotting position.

A random variable X is said to belong to the **location–scale family** when its CDF

$$F_X(x\,|\,a,b) = \Pr(X \le x\,|\,a,b) \tag{9.1a}$$

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

$$F_X(x\,|\,a,b) = F\left(\frac{x-a}{b}\right);\ a \in \mathbb{R},\ b > 0; \tag{9.1b}$$

where $F(\cdot)$ denotes a distribution having no other parameters. Different $F(\cdot)$'s correspond to different members of the family. The random variable

$$Y = \frac{X-a}{b} \tag{9.1c}$$

is called the **reduced variable**.[3] The location parameter a is either a measure of central tendency (mean, median, mode) or a threshold parameter. The scale parameter b can be either the standard deviation or the length of the distribution's support or the length of a central $(1-\alpha)$–interval for X.

We will write the **reduced CDF** as

$$F_Y(y) := F\left(\frac{x-a}{b}\right),\ y = \frac{x-a}{b}, \tag{9.1d}$$

and the CCDF belonging to (9.1a,b) read

$$\begin{aligned} S_X(x\,|\,a,b) &= 1 - F_X(x\,|\,a,b) && (9.1e)\\ S_Y(y) &= 1 - F_Y(y). && (9.1f)\end{aligned}$$

We now turn to the CHR and find

$$\begin{aligned} H_X(x\,|\,a,b) &= -\ln\big[1 - F_X(x\,|\,a,b)\big]\\ &= -\ln\big[1 - F_Y(y)\big]\\ &= H_Y(y),\ y = (x-a)/b. && (9.1g)\end{aligned}$$

[2] See RINNE (2010) for a detailed representation of probability plotting and linear estimation techniques. This monograph is supplemented by a MATLAB program that — among other features — produces probability papers.

[3] 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 $a = \mathrm{E}(X)$ and $b^2 = \mathrm{Var}(X)$. So, in this case reducing and standardizing are the same.

Let $\Lambda,\ \Lambda \geq 0$, be a value of the CHR, then the **hazard quantile** of order Λ in the reduced case is

$$y_\Lambda = H_Y^{-1}(\Lambda) \tag{9.1h}$$

and consequently

$$x_\Lambda = a + b\, y_\Lambda. \tag{9.1i}$$

The hazard paper for a location–scale distribution is now constructed by taking the horizontal axis (abscissa) for x or x_Λ and the vertical axis (ordinate) for y or y_Λ[4] where the labeling of this axis is according to the corresponding CHR–value Λ. This procedure gives a scaling with respect to Λ which is non–linear, the only exception is the **exponential distribution**,[5] where Λ and Y_Λ coincide:

$$\begin{aligned} F_X(x\,|\,a,b) &= 1 - \exp\left(-\frac{x-a}{b}\right) = F_Y(y) \\ H_X(x\,|\,a,b) &= \frac{x-a}{b} = y = H_Y(y) = \Lambda \\ y_\Lambda &= H_Y^{-1}(\Lambda) = \Lambda. \end{aligned}$$

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

$$\Lambda = -\ln(1-P) \tag{9.1j}$$

$$P = 1 - \exp(-\Lambda), \tag{9.1k}$$

where P is a given value of the CDF. Thus, a probability grid may be used for hazard plotting when the P–scaling on the ordinate is supplemented by a Λ–scaling, see Fig. 9/1. Conversely, a hazard paper may be used for probability plotting.

The extreme value distribution of type I for the minimum, the **log–WEIBULL distribution**, is a genuine location–scale distribution:

$$F_X(x\,|\,a,b) = 1 - \exp\left[-\exp\left(\frac{x-a}{b}\right)\right];\ a \in \mathbb{R},\ b > 0,\ x \in \mathbb{R}, \tag{9.2a}$$

$$F_Y(y) = 1 - \exp\big[-\exp(y)\big],\ y \in \mathbb{R}. \tag{9.2b}$$

So, the reduced CHR reads

$$\begin{aligned} H_Y(y) &= -\ln\big[1 - F_Y(y)\big] \\ &= -\ln\big\{\exp\big[-\exp(y)\big]\big\} \\ &= \exp(y). \end{aligned} \tag{9.2c}$$

The reduced hazard quantile of order Λ is

$$y_\Lambda = \ln \Lambda \tag{9.2d}$$

[4] Some authors construct hazard paper by interchanging the axes. This approach has some justification when looking at (9.1i), where the dependent variable is x_Λ which normally is laid down on the ordinate, and y_Λ is the independent variable, to be displayed on the abscissa.

[5] For a discrete distribution the exception is the geometric distribution with

$$\Lambda = H_Y(y) = P\,y;\ y = 0,\,1,\,2\,\ldots,\ 0 \leq P < 1$$

and

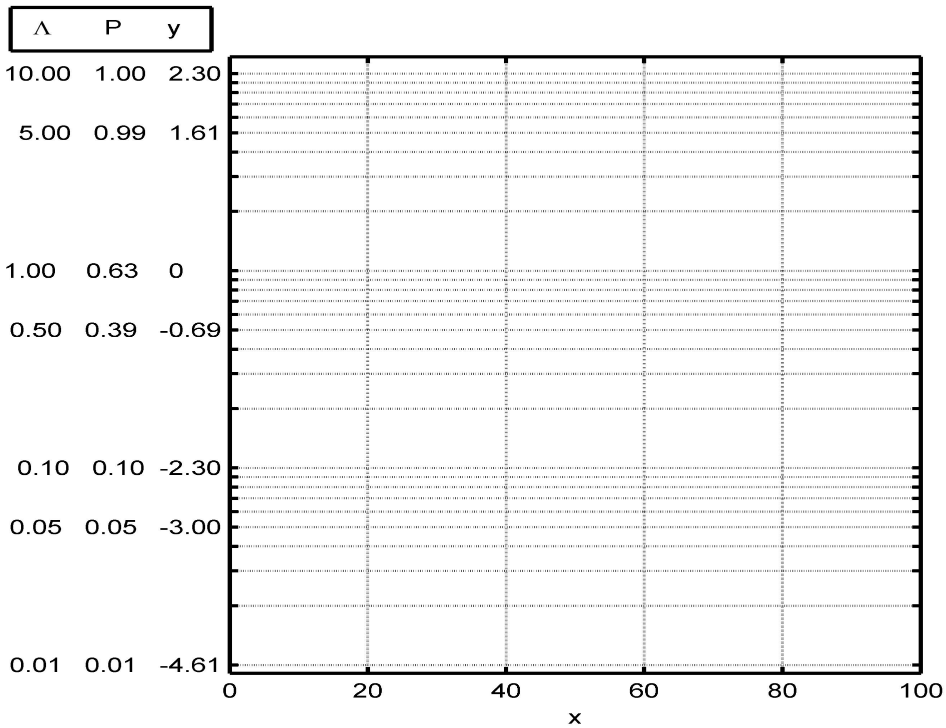
$$y_\Lambda = H_Y^{-1}(\Lambda) = \frac{\Lambda}{P}.$$

and the X–hazard quantile is

$$x_\Lambda = a + b\, y_\Lambda = a + b\, \ln \Lambda. \tag{9.2e}$$

The hazard paper of the log–WEIBULL distribution has a logarithmic scale on the ordinate and a linear scale on the abscissa. For didactic reasons we have given three scalings in Fig. 9/1, the Λ–, the P– and the y–scaling, and only the last one is linear.

Figure 9/1: Hazard paper of the log–WEIBULL distribution



When the straight line $x_\Lambda = a + b\, y_\Lambda$ is given, we may find a and b by suitably chosen values of Λ. For $\Lambda = 1$ we have

$$x_1 = a + b\, \ln 1 = a \tag{9.2f}$$

and the distance between the x–hazard quantiles of order $\Lambda = 1$ and $\Lambda = \exp(1) = e$ leads to b :

$$x_e - x_1 = (a + b\, \ln e) - (a + b\, \ln 1) = b. \tag{9.2g}$$

A distribution transformable to location–scale type is the **WEIBULL distribution** with

$$F_X(x\,|\,a, b, c) = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^c\right];\; a \in \mathbb{R},\; b,\; c > 0,\; x \geq a. \tag{9.3a}$$

$a,\; b,\; c$ are the original location, scale and shape parameters. When a, the lower threshold of X, is known (mostly $a = 0$) or has been estimated in some way or the other, the transformed variable

$$X^* = \ln(X - a) \tag{9.3b}$$

has the log–WEIBULL distribution:

$$F_{X^*}(x^*\,|\,a^*, b^*) = 1 - \exp\left[-\exp\left(\frac{x^* - a^*}{b^*}\right)\right];\; a^* \in \mathbb{R},\; b^* > 0,\; x^* \in \mathbb{R} \tag{9.3c}$$

where

$$a^* = \ln b, \tag{9.3d}$$
$$b^* = 1/c. \tag{9.3e}$$

So, the hazard paper of the WEIBULL distribution has the same ordinate as the log–WEIBULL distribution, but a logarithmic scale on the abscissa.

We now turn to how to find an estimate of $H(x) = \Lambda$, i.e., how to find a **plotting position** on the hazard grid. Hazard plots are not based on the PLE of $S(x)$, which would give $\widetilde{H}(x_i) = -\ln \widehat{S}(x_i)$, but they rest upon the empirical cumulative hazard function (5.12a). When the data set is singly type–II censored, the observed distinct lifetimes $x_1 < x_2 < \ldots < x_k$ are the first k lifetimes in a sample of size n,[6] and the number of units at risk at x_i is $n_i = n - i + 1$. This gives

$$\widehat{H}(x_j) = \widehat{\Lambda}_j = \sum_{i=1}^{j} \frac{1}{n-i+1};\; j = 1,\, 2,\, \ldots,\, k. \tag{9.4}$$

The quantity $n_i = n - i + 1$ in (9.4) is nothing but the **reverse rank** which results in the case of random censoring when all observations — censored as well as uncensored — would be ordered, but then the summation is only over those reciprocal reverse ranks belonging to uncensored observations, see Example 9/1.

One argument in support of (9.4) is that with singly type–II censoring it can be shown that the estimator (9.4) is unbiased:

$$\mathrm{E}\big[\widehat{H}(x_j)\big] = H(x_j). \tag{9.5}$$

To prove this suppose X has survivor function $S(x)$. As is well known, the random variable $U = S(X)$ has an uniform distribution on $[0,\, 1]$, and hence $W = -\ln U = H(X)$ has the reduced exponential distribution with PDF $f(w) = \exp(-w),\; w \geq 0$. Therefore, if $X_1 < X_2 < \ldots < X_n$ are the ordered random observations in a sample of size n, the random variable $W_j = H(X_j)$ is the j-th ordered observation in a random sample of size n from the reduced exponential distribution. As is known too, see RINNE (2010, p. 121), the mean of W_j is

$$\mathrm{E}(W_j) = \sum_{i=1}^{j} \frac{1}{n-i+1}, \tag{9.6}$$

and thus the stated result follows from (9.4). It is more tedious, see NELSON (1972), to show that if the data are progressively (= multiply) type–II censored, the result (9.5) still holds, where X_j represents the j-th smallest uncensored observation. NELSON (1982) suggests modified plotting positions obtained by averaging the hazard step function at the jumps x_j. The modified position of the earliest failure x_1 is half its regular hazard value $\Lambda_1 = 1/n$. The modified positions agree better with a distribution fitted by maximum likelihood.

How to find estimates for the location–scale parameters a and b? — A first possibility is **least–squares estimation** of (9.1i), interpreted as a regression of Y_Λ on X_Λ, where the regressand is taken as uncensored observation of order j and the regressor is

$$Y_{\widehat{\Lambda}_j} = H_Y^{-1}\big(\widehat{\Lambda}_j\big), \tag{9.7a}$$

the reduced hazard quantile of order $\widehat{\Lambda}_j$ estimated by (9.4). Introducing the following vectors and

[6] The reasoning will be the same for an uncensored sample where $k = n$.

matrices

$$\mathbf{x} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}; \ \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}; \ \widehat{\mathbf{y}} = \begin{pmatrix} Y_{\widehat{\Lambda}_1} \\ Y_{\widehat{\Lambda}_2} \\ \vdots \\ Y_{\widehat{\Lambda}_k} \end{pmatrix}; \ \theta = \begin{pmatrix} a \\ b \end{pmatrix}; \ \mathbf{A} = \big(\mathbf{1}\ \widehat{\mathbf{y}}\big) \tag{9.7b}$$

the **ordinary least–squares (OLS) estimator** of θ is

$$\widehat{\theta} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \big(\mathbf{A}'\,\mathbf{A}\big)^{-1}\,\mathbf{A}'\,\mathbf{x}. \tag{9.7c}$$

This estimator is not statistically optimal as the regressor variables $Y_{\widehat{\Lambda}_j}$ are neither homoscedastic nor free of autocorrelation.[7]

A second possibility is an **eye–fitted straight line** to the point–cluster on the hazard paper whereby the user should keep in mind that the *horizontal* distances between the x_j's and the straight line have to be rendered as small as possible. If we look at (9.1i) we see that

$$- \quad a \quad = \quad x_{\Lambda_0}, \text{ where } \Lambda_0 \text{ is such that } y_{\Lambda_0} = 0 \text{ and} \tag{9.8a}$$

$$- \quad b \quad = \quad x_{\Lambda_1} - x_{\Lambda_0}, \text{ where } \Lambda_1 \text{ is such that } y_{\Lambda_1} = 1. \tag{9.8b}$$

So we find so–called **hazard–quantile estimates** of a and b by processing — according to (9.8a,b) — $\widehat{x}_{\Lambda_0}$ and $\widehat{x}_{\Lambda_1}$ belonging to Λ_0 and Λ_1 on the eye–fitted straight line.[8] When the straight line has been fitted by OLS the estimates read–off according to (9.8a,b) will be identical — apart from errors due to rounding and reading–off — to the OLS estimates.

Example 9/1: Hazard plotting and parameter estimation for a logistic distribution

The following table gives a simulated data set ($n = 15$) from a logistic distribution having $a = 20$ and $b = 2$. The sample has been randomly censored. The logistic distribution has

$$S_X(x) = \left[1 + \exp\left(\frac{x-a}{b}\right)\right]^{-1}, \ a \in \mathbb{R},\ b > 0,\ x \in \mathbb{R}, \tag{9.9a}$$

giving the cumulated hazard rate

$$H_X(x) = -\ln S_X(x) = \ln\left[1 + \exp\left(\frac{x-a}{b}\right)\right]. \tag{9.9b}$$

From the reduced distribution we find

$$\Lambda \quad = \quad H_Y(y) = \ln\big[1 + \exp(y)\big],\ y \in \mathbb{R}, \tag{9.9c}$$

$$y_\Lambda \quad = \quad H_Y^{-1}(\Lambda) = \ln\big[\exp(\Lambda) - 1\big],\ \Lambda \geq 0. \tag{9.9d}$$

The hazard paper in Fig. 9/2 has an ordinate scaled according to (9.9c). The figure displays the plotted data and the OLS–fitted straight line with parameter estimates

$$\widehat{a} = 21.1580, \quad \widehat{b} = 2.1428$$

[7] See RINNE (2010, Chapter 4) for the alternative general least–squares estimator, which is implemented in the MATLAB program LEPP appended to the monograph.

[8] We have location–scale distributions where we have to choose other reduced quantiles than $y = 0$ and/or $y = 1$ to find hazard–quantile estimators for a and b, see, e.g., the arc–sine distribution in Sect. 9.3.

which do not differ much from the input parameters $a = 20$ and $b = 2$. In order to find the hazard–quantile estimates we need

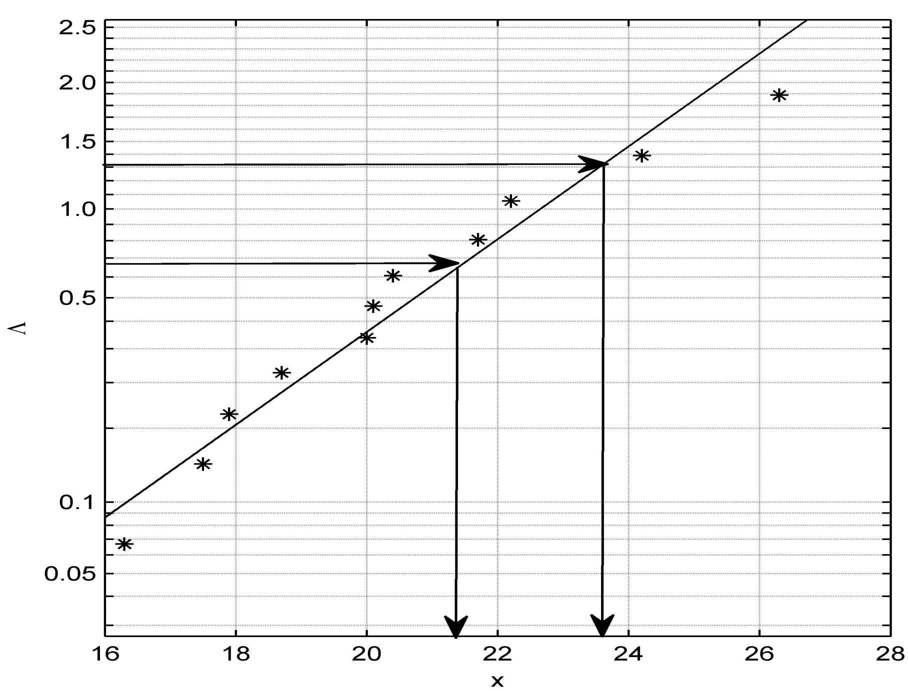
$$\Lambda_0 = \ln\big[1 + \exp(0)\big] = \ln 2 \approx 0.6931,$$

$$\Lambda_1 = \ln\big[1 + \exp(1)\big] \approx \ln 3.7183 \approx 1.3133.$$

Fig. 9/2 then shows the way to read–off the hazard–quantile estimates which — apart from random errors — agree with the OLS–estimates.

j	x_j	δ_j	n_j	δ_j/n_j	$\widehat{\Lambda_j} = \sum_{i=1}^{j} 1/n_i$
1	16.3	1	15	0.0667	0.0667
2	16.5	0	14	0	–
3	17.5	1	13	0.0769	0.1436
4	17.9	1	12	0.0833	0.2269
5	17.9	0	11	0	–
6	18.7	1	10	0.1000	0.3269
7	20.0	1	9	0.1111	0.4380
8	20.1	1	8	0.1250	0.5630
9	20.4	1	7	0.1429	0.7059
10	20.7	0	6	0	–
11	21.7	1	5	0.2000	0.9059
12	22.2	1	4	0.2500	1.1559
13	24.2	1	3	0.3333	1.4892
14	26.3	1	2	0.5000	1.9892
15	26.3	0	1	0	–

Figure 9/2: Hazard plot for the logistic distribution



9.3 Hazard Papers for Location–scale Distributions

This section informs on how to scale the cumulative hazard rate when the random variable either has a genuine location–scale distribution or a distribution transformable to location–scale type. We will also give those CHR–values and their hazard quantiles that allow to comfortably read–off estimates of the location and scale parameters.

We start with genuine location–scale distributions.

Arcsine distribution

$$\begin{aligned}
S(x) &= \frac{\pi - 2\,\arcsin\!\left(\frac{x-a}{b}\right)}{2\,\pi};\; a \in \mathbb{R},\, b > 0,\; a - b \le x \le a + b \\
\Lambda &= H_Y(y) \;=\; \ln(2\,\pi) - \ln\big[\pi - 2\,\arcsin(y)\big];\; -1 \le y \le 1 \\
y_\Lambda &= \cos\big[\pi\,\exp(-\Lambda)\big];\; \Lambda \ge 0 \\
a &= x_{(-\ln[1/2])} \;\approx\; x_{0.6931} \\
b &= x_{(-\ln[1/3])} - x_{(-\ln[2/3])} \;\approx\; x_{1.0986} - x_{0.4055}
\end{aligned}$$

CAUCHY distribution

$$\begin{aligned}
S(X) &= \tfrac{1}{2} - \tfrac{1}{\pi}\,\arctan\!\left(\tfrac{x-a}{b}\right);\; a \in \mathbb{R},\; b > 0,\; x \in \mathbb{R} \\
\Lambda &= H_Y(y) \;=\; \ln(2\,\pi) - \ln\big[\pi - 2\,\arctan(y)\big];\; y \in \mathbb{R} \\
y_\Lambda &= \tan\big[\pi\,\exp(-\Lambda)\big];\; \Lambda \ge 0 \\
a &= x_{(-\ln[1/2])} \;\approx\; x_{0.6931} \\
b &= x_{(-\ln[1/4])} - x_{(-\ln[3/4])} \;\approx\; x_{1.3863} - x_{0.2877}
\end{aligned}$$

Cosine distribution, ordinary

$$\begin{aligned}
S(x) &= 0.5\,\big[1 - \sin\!\left(\tfrac{x-a}{b}\right)\big]\,; \\
&\qquad a \in \mathbb{R},\, b > 0,\; a - \tfrac{\pi}{2}\,b \le x \le a + \tfrac{\pi}{2}\,b \\
\Lambda &= H_Y(y) \;=\; \ln 2 - \ln\big[1 - \sin(y)\big];\; -\tfrac{\pi}{2} \le y \le \tfrac{\pi}{2} \\
y_\Lambda &= \arcsin\big[1 - 2\,\exp(-\Lambda)\big];\; \Lambda \ge 0 \\
a &= x_{(-\ln[1/2])} \;\approx\; x_{0.6931} \\
b &= x_{(\ln\{2/[1-\sin(0.5)]\})} - x_{(\ln\{2/[1-\sin(-0.5)]\})} \;\approx\; x_{1.3460} - x_{0.3015}
\end{aligned}$$

Cosine distribution, raised

$$\begin{aligned}
S(x) &= \tfrac{1}{2}\,\big[1 - \tfrac{x-a}{b} - \tfrac{1}{\pi}\,\sin\!\big(\pi\,\tfrac{x-a}{b}\big)\big]\,;\; a \in \mathbb{R},\, b > 0,\; a - b \le x \le a + b \\
\Lambda &= H_Y(y) \;=\; \ln(2\,\pi) - \ln\big[\pi\,(1-y) - \sin(\pi\,y)\big];\; -1 \le y \le 1 \\
y_\Lambda &- \;\;\text{cannot be given in closed form} \\
a &= x_{(-\ln[1/2])} \;\approx\; x_{0.6931} \\
b &\approx x_{(-\ln 0.0908)} - x_{(-\ln 0.9092)} \;\approx\; x_{2.3391} - x_{0.0952}
\end{aligned}$$

Exponential distribution

$$\begin{aligned} S(x) &= \exp\left(-\tfrac{x-a}{b}\right);\ a \in \mathbb{R},\ b > 0,\ x \ge a \\ \Lambda &= H_Y(y) \ = \ y;\ y \ge 0 \\ y_\Lambda &= \Lambda;\ \Lambda \ge 0 \\ a &= x_0 \\ b &= x_1 - x_0 \end{aligned}$$

Exponential distribution, reflected

$$\begin{aligned} S(x) &= 1 - \exp\left(\tfrac{x-a}{b}\right);\ a \in \mathbb{R},\ b > 0,\ x \le a \\ \Lambda &= H_Y(y) \ = \ -\ln\big[1 - \exp(y)\big];\ y \le 0 \\ y_\Lambda &= \ln\big[1 - \exp(-\Lambda)\big];\ \Lambda \ge 0 \end{aligned}$$

a is the upper threshold of the reflected exponential distribution where $\Lambda = H_X(a) = \infty$. Thus, a cannot be read off as a point on the abscissa. Therefore, we propose the following procedure. First, from the difference of $x(y=-1) = a - 1\,b$ and $x(y=-2) = a - 2\,b$ we find

$$b = x_{(\Lambda[-1])} - x_{(\Lambda[-2])} \approx x_{0.4587} - x_{0.1482},$$

and then from

$$x(y=-1) + b = x(y=-1) + \big[x(y=-1) - x(y=-2)\big] = 2\,x(y=-1) - x(y=-2)$$

we have

$$a = 2\,x_{(\Lambda[-1])} - x_{(\Lambda[-2])} \approx 2\,x_{0.4587} - x_{0.1482}.$$

Extreme value distribution of type I for the maximum (GUMBEL distribution)

$$\begin{aligned} S(x) &= 1 - \exp\big[-\exp\big(-\tfrac{x-a}{b}\big)\big];\ a \in \mathbb{R},\ b > 0,\ x \in \mathbb{R} \\ \Lambda &= H_Y(y) \ = \ -\ln\big\{1 - \exp\big[-\exp(-y)\big]\big\};\ y \in \mathbb{R} \\ y_\Lambda &= \ln\big\{-\ln\big[1 - \exp(-\Lambda)\big]\big\};\ \Lambda \ge 0 \\ a &= x_{(-\ln\{1-\exp[-\exp(0)]\})} \ \approx \ x_{0.4587} \\ b &= x_{(-\ln\{1-\exp[-\exp(0)]\})} - x_{(-\ln\{1-\exp[-\exp(-1)]\})} \ \approx \ x_{0.4587} - x_{0.0683} \end{aligned}$$

Extreme value distribution of type I for the minimum (Log–WEIBULL distribution)

$$\begin{aligned} S(x) &= \exp\big[-\exp\big(\tfrac{x-a}{b}\big)\big];\ a \in \mathbb{R},\ b > 0,\ x \in \mathbb{R} \\ \Lambda &= H_Y(y) \ = \ \exp(y);\ y \in \mathbb{R} \\ y_\Lambda &= \ln(\Lambda);\ \Lambda \ge 0 \\ a &= x_{(\exp[0])} \ = \ x_1 \\ b &= x_{(\exp[1])} - x_{(\exp[0])} \ \approx \ x_{2.7181} - x_1 \end{aligned}$$

Half–CAUCHY distribution

$$
\begin{aligned}
S(x) &= 1 - \tfrac{2}{\pi}\,\arctan\bigl(\tfrac{x-a}{b}\bigr)\,;\; a \in \mathbb{R},\, b > 0,\; x \ge a \\
\Lambda &= H_Y(y) \;=\; -\ln\bigl[1 - \tfrac{2}{\pi}\,\arctan(y)\bigr];\; y \ge 0 \\
y_\Lambda &= \tan\bigl\{\tfrac{\pi}{2}\,\bigl[1 - \exp(-\Lambda)\bigr]\bigr\}\,;\; \Lambda \ge 0 \\
a &= x_0 \\
b &= x_{(-\ln[1-(2/\pi)\,\arctan(1)])} - x_0 \;\approx\; x_{0.6931} - x_0
\end{aligned}
$$

Half–logistic distribution

$$
\begin{aligned}
S(x) &= \frac{2\,\exp\bigl(-\frac{x-a}{b}\bigr)}{1 + \exp\bigl(-\frac{x-a}{b}\bigr)};\; a \in \mathbb{R},\, b > 0,\; x \ge a \\
\Lambda &= H_Y(y) \;=\; -\ln\Bigl[\tfrac{2}{1+\exp(y)}\Bigr]\,;\; y \ge 0 \\
y_\Lambda &= \ln\bigl[2\,\exp(\Lambda) - 1\bigr];\; \Lambda \ge 0 \\
a &= x_0 \\
b &= x_{(-\ln\{2/[1+\exp(1)]\})} - x_0 \;\approx\; x_{0.6201} - x_0
\end{aligned}
$$

Half–normal distribution

$$
\begin{aligned}
S(x) &= 2\,\bigl[1 - \Phi\bigl(\tfrac{x-a}{b}\bigr)\bigr]\,;\; a \in \mathbb{R},\, b > 0,\; x \ge a \\
\Lambda &= H_Y(y) \;=\; -\ln\bigl\{2\,\bigl[1 - \Phi(y)\bigr]\bigr\};\; y \ge 0 \\
y_\Lambda &= \Phi^{-1}\Bigl[1 - \tfrac{\exp(-\Lambda)}{2}\Bigr]\,;\; \Lambda \ge 0 \\
a &= x_0 \\
b &= x_{(-\ln\{2\,[1-\Phi(1)]\})} - x_0 \;\approx\; x_{1.1479} - x_0
\end{aligned}
$$

$\Phi(\cdot)$ is the CDF of the standardized normal distribution and $\Phi^{-1}(\cdot)$ is its percentile function.

Hyperbolic secant distribution

$$
\begin{aligned}
S(x) &= 1 - \tfrac{2}{\pi}\,\arctan\bigl[\exp\bigl(\tfrac{x-a}{b}\bigr)\bigr]\,;\; a \in \mathbb{R},\, b > 0,\; x \in \mathbb{R} \\
\Lambda &= H_Y(y) \;=\; -\ln\bigl\{1 - \tfrac{2}{\pi}\,\arctan\bigl[\exp(y)\bigr]\bigr\}\,;\; y \in \mathbb{R} \\
y_\Lambda &= \ln\bigl\{\tan\bigl[\tfrac{\pi}{2}\,\bigl(1 - \exp(-\Lambda)\bigr)\bigr]\bigr\}\,;\; \Lambda \ge 0 \\
a &= x_{(-\ln\{1-(2/\pi)\,\arctan[\exp(0)]\})} \;\approx\; x_{0.6931} \\
b &= x_{(-\ln\{1-(2/\pi)\,\arctan[\exp(1)]\})} - x_{(-\ln\{1-(2/\pi)\,\arctan[\exp(0)]\})} \;\approx\; x_{1.14941} - x_{0.6931}
\end{aligned}
$$

LAPLACE distribution

$$
\begin{aligned}
S(x) &= \left\{\begin{array}{ll} 1-0.5\,\exp\!\left(-\frac{a-x}{b}\right) & \text{for } x \le a \\ 0.5\,\exp\!\left(-\frac{x-a}{b}\right) & \text{for } x \ge a \end{array}\right\};\ a\in\mathbb{R},\ b>0 \\
\Lambda &= H_Y(y) \;=\; \left\{\begin{array}{ll} -\ln\big[1-0.5\,\exp(-y)\big] & \text{for } y\le 0 \\ -\ln\big[0.5\,\exp(-y)\big] & \text{for } y \ge 0 \end{array}\right. \\
y_\Lambda &= \left\{\begin{array}{ll} -\ln\big\{2\,\big[1-\exp(-\Lambda)\big]\big\} & \text{for } \Lambda \le -\ln 0.5 \approx 0.6931 \\ -\ln\big[2\,\exp(-\Lambda)\big] & \text{for } \Lambda \ge -\ln 0.5 \approx 0.6931 \end{array}\right. \\
a &= x_{(-\ln[1-0.5])} \;\approx\; x_{0.6931} \\
b &= x_{(-\ln[0.5\,\exp(-1)]} - x_{(-\ln[1-0.5])} \;\approx\; x_{1.6931} - x_{0.6931}
\end{aligned}
$$

Logistic distribution

$$
\begin{aligned}
S(x) &= \big[1+\exp\!\left(\tfrac{x-a}{b}\right)\big]^{-1};\ a\in\mathbb{R},\ b>0,\ x\in\mathbb{R} \\
\Lambda &= H_Y(y) \;=\; \ln\big[1+\exp(y)\big];\ y\in\mathbb{R} \\
y_\Lambda &= \ln\big[\exp(\Lambda)-1\big];\ \Lambda \ge 0 \\
a &= x_{(\ln[1+\exp(0)])} \;\approx\; x_{0.6931} \\
b &= x_{(\ln[1+\exp(1)])} - x_{(\ln[1+\exp(0)])} \;\approx\; x_{1.3133} - x_{0.6931}
\end{aligned}
$$

MAXWELL–BOLTZMANN distribution

$$
\begin{aligned}
S(x) &= 1 - F_{\chi^2_3}\Big[\left(\tfrac{x-a}{b}\right)^2\Big];\ a\in\mathbb{R},\ b>0,\ x\ge a \\
\Lambda &= H_Y(y) \;=\; -\ln\Big[1-F_{\chi^2_3}(y)\Big];\ y \ge 0 \\
y_\Lambda &- \text{cannot be given in closed form} \\
a &= x_{(-\ln[1-F_{\chi^2_3}(0)])} \;=\; x_0 \\
b &= x_{(-\ln[1-F_{\chi^2_3}(1)])} - x_{(-\ln[-F_{\chi^2_3}(0)])} \;\approx\; x_{0.2215} - x_0
\end{aligned}
$$

$F_{\chi^2_3}(\cdot)$ is the CDF of the χ^2–distribution with 3 degrees of freedom.

Normal distribution

$$
\begin{aligned}
S(x) &= 1-\Phi\!\left(\tfrac{x-a}{b}\right);\ a\in\mathbb{R},\ b>0,\ x\in\mathbb{R} \\
\Lambda &= H_Y(y) \;=\; -\ln\big[1-\Phi(y)\big];\ y\in\mathbb{R} \\
y_\Lambda &= \Phi^{-1}\big[1-\exp(-\Lambda)\big];\ \Lambda\ge 0 \\
a &= x_{(-[1-\Phi(0)])} \;\approx\; x_{0.6931} \\
b &= x_{(-\ln[1-\Phi(1)])} - x_{(-[1-\Phi(0)])} \;\approx\; x_{1.8407} - x_{0.6931}
\end{aligned}
$$

$\Phi(\cdot)$ is the CDF of the standardized normal distribution and $\Phi^{-1}(\cdot)$ is its percentile function.

Parabolic U–shaped distribution

$$
\begin{aligned}
S(x) &= \tfrac{1}{2}\Big[1-\big(\tfrac{x-a}{b}\big)^3\Big];\ a\in\mathbb{R},\ b>0,\ a-b\le x\le a+b\\
\Lambda &= H_Y(y) \ = \ \ln 2-\ln\big[1-y^3\big];\ -1\le y\le 1\\
y_\Lambda &= \big[1-\exp(\ln 2-\Lambda\big]^{1/3};\ \Lambda\ge 0\\
a &= x_{(\ln 2-\ln[1-0^3])} \ \approx \ x_{0.6931}\\
b &= x_{(\ln 2-\ln[1-0.5^3])}-x_{(\ln 2-\ln[1-(-0.5)^3])} \ \approx \ x_{0.8267}-x_{0.5754}
\end{aligned}
$$

Parabolic inverted U–shaped distribution

$$
\begin{aligned}
S(x) &= \tfrac{1}{2}-\tfrac{1}{4}\Big[\tfrac{3\,(x-a)}{b}-\big(\tfrac{x-a}{b}\big)^3\Big];\ a\in\mathbb{R},\ b>0,\ a-b\le x\le a+b\\
\Lambda &= H_Y(y) \ = \ \ln 4-\ln\big[2-3\,y+y^3\big];\ -1\le y\le 1\\
y_\Lambda &- \text{admissible solution, i.e., } \ -1\le y_\Lambda\le 1, \ \text{ of } \ 3\,y_\Lambda-y_\Lambda^3=4\,\exp(-\Lambda)-2\\
a &= x_{(\ln 4-\ln 2)} \ \approx \ x_{0.6931}\\
b &= x_{(\ln 4-\ln[2-3\cdot 0.5+0.5^3])}-x_{(\ln 4-\ln[2-3\,(-0.5)+(-0.5)^3])} \ \approx \ x_{18563}-x_{0.1699}
\end{aligned}
$$

RAYLEIGH distribution

$$
\begin{aligned}
S(x) &= \exp\Big[-\tfrac{1}{2}\,\big(\tfrac{x-a}{b}\big)^2\Big];\ a\in\mathbb{R},\ b>0,\ x\ge a\\
\Lambda &= H_Y(y) \ = \ y^2/2;\ y\ge 0\\
y_\Lambda &= \sqrt{2\,\Lambda};\ \Lambda\ge 0\\
a &= x_0\\
b &= x_{0.5}-x_0
\end{aligned}
$$

RAYLEIGH distribution, inverse

$$
\begin{aligned}
S(x) &= 1-\exp\bigg[-\Big(\frac{b}{x-a}\Big)^2\bigg];\ a\in\mathbb{R},\ b>0,\ x\ge a\\
\Lambda &= H_Y(y) \ = \ -\ln\Big[1-\exp\Big(-\frac{1}{y^2}\Big)\Big];\ y\ge 0\\
y_\Lambda &= \sqrt{\frac{1}{-\ln\big[1-\exp(-\Lambda)\big]}};\ \Lambda\ge 0\\
a &= x_0\\
b &= x_{(-\ln[1-\exp(-1)])}-x_0 \ \approx \ x_{0.4587}-x_0
\end{aligned}
$$

Semi–elliptical distribution

$$
\begin{aligned}
S(x) &= \tfrac{1}{2}-\tfrac{1}{\pi}\bigg[\tfrac{x-a}{b}\sqrt{1-\big(\tfrac{x-a}{b}\big)^2}+\arcsin\big(\tfrac{x-a}{b}\big)\bigg];\ a\in\mathbb{R},\ b>0,\ a-b\le x\le a+b\\
\Lambda &= H_Y(y) \ = \ -\ln\Big\{\tfrac{1}{2}-\tfrac{1}{\pi}\Big[y\sqrt{1-y^2}+\arcsin(y)\Big]\Big\};\ -1\le y\le 1\\
y_\Lambda &- \text{admissible solution, i.e., } \ -1\le y_\Lambda\le 1, \ \text{ of}\\
&\quad\ y_\Lambda\sqrt{1-y_\Lambda^2}+\arcsin(y_\Lambda)=\pi\,[0.5-\exp(-\Lambda)]\\
a &= x_{(-\ln 0.5)} \ \approx \ x_{0.6931}\\
b &= x_{(-\ln[(1/2)-(1/\pi)\{0.5\,\sqrt{1-0.5^2}+\arcsin(0.5)\}])}-\\
&\quad\ x_{(-\ln[(1/2)-(1/\pi)\{0.5\,\sqrt{1-(-0.5)^2}+\arcsin(-0.5)\}])} \ \approx \ x_{1.6322}-x_{0.2175}
\end{aligned}
$$

TEISSIER distribution

$$\begin{aligned}
S(x) &= \exp\left[1+\tfrac{x-a}{b}-\exp\left(\tfrac{x-a}{b}\right)\right];\ a\in\mathbb{R},\ b>0,\ x\ge a\\
\Lambda &= H_Y(y)\ = \exp(y)-y-1;\ y\ge 0\\
y_\Lambda &- \text{admissible solution, i.e., } y_\Lambda\ge 0,\ \text{ of } \exp(y_\Lambda)-y_\Lambda=1+\Lambda\\
a &= x_0\\
b &= x_{(\exp[1]-2)}-x_0\ \approx\ x_{0.7183}-x_0
\end{aligned}$$

Triangular distribution, right–angled and negatively skew

$$\begin{aligned}
S(x) &= \frac{a-x}{b}\left(\frac{x-a}{b}+2\right);\ a\in\mathbb{R},\ b>0,\ a-b\le x\le a\\
\Lambda &= H_Y(y)\ = -\ln(-y^2-2\,y);\ -1\le y\le 0\\
y_\Lambda &- \text{admissible solution, i.e., } -1\le y_\Lambda\le 0,\ \text{ of } y_\Lambda^2+2\,y_\Lambda=-\exp(-\Lambda)
\end{aligned}$$

a is the upper threshold where $H_X(a)=\infty$, and thus cannot be read off. From the difference of $x(y=-0.5)=a-0.5\,b$ and $x(y=-1)=a-b$ we find

$$\frac{b}{2}=x_{\Lambda(-0.5)}-x_{\Lambda(-1)}\ \approx\ x_{0.2877}-x_0$$

and then

$$a=x_0+2\,\frac{b}{2}\ \approx\ 2\,x_{0.2877}-x_0.$$

Triangular distribution, right–angled and positively skew

$$\begin{aligned}
S(x) &= \left(\frac{a+b-x}{b}\right)^2;\ a\in\mathbb{R},\ b>0,\ a\le x\le a+b\\
\Lambda &= H_Y(y)\ = -2\,\ln(1-y);\ 0\le y\le 1\\
y_\Lambda &= 1-\exp\left(-\tfrac{\Lambda}{2}\right);\ \Lambda\ge 0\\
a &= x_0\\
b &= 2\,(x_{1.3863}-x_0)
\end{aligned}$$

Triangular distribution, symmetric

$$\begin{aligned}
S(x) &= \left\{\begin{array}{ll} 1-\tfrac{1}{2}\left(1+\dfrac{x-a}{b}\right)^2 & \text{for } a-b\le x\le a\\[2ex] \tfrac{1}{2}\left(1-\dfrac{x-a}{b}\right)^2 & \text{for } a\le x\le a+b \end{array}\right\};\ a\in\mathbb{R},\ b>0\\
\Lambda &= H_Y(y)\ = \left\{\begin{array}{ll} -\ln\left[1-\tfrac{1}{2}\,(1+y)^2\right] & \text{for } -1\le y\le 0\\ \ln 2-2\,\ln(1-y) & \text{for } 0\le y\le 1 \end{array}\right.\\
y_\Lambda &= \left\{\begin{array}{ll} 2\sqrt{\left[1-\exp(-\Lambda)\right]}-1 & \text{for } \Lambda\le\ln 2\approx 0.6931\\ 1-\sqrt{2\,\exp(-\Lambda)} & \text{for } \Lambda\ge\ln 2\approx 0.6931 \end{array}\right.\\
a &= x_{(\ln 2)}\ \approx\ x_{0.6931}\\
b &= x_{(\ln 2-\ln[1-0.5])}-x_{(-\ln[1-0.5\,(1-0.5)^2])}\ \approx\ x_{2.0794}-x_{0.1335}
\end{aligned}$$

Uniform distribution

$$\begin{aligned}
S(x) &= 1-\frac{x-a}{b};\ a\in\mathbb{R},\ b>0,\ a\le x\le a+b\\
\Lambda &= H_Y(y)\ = -\ln(1-y);\ 0\le y\le 1\\
y_\Lambda &= 1-\exp(-\Lambda);\ \Lambda\ge 0\\
a &= x_0\\
b &= 2\left(x_{(-\ln[0.5])}-x_0\right)\ \approx\ 2\left(x_{0.6931}-x_0\right)
\end{aligned}$$

V–shaped distribution

$$\begin{aligned}
S(x) &= \left\{\begin{array}{ll}\frac12\left[1+\left(\frac{a-x}{b}\right)^2\right] & \text{for } a-b\le x\le a\\ \frac12\left[1-\left(\frac{a-x}{b}\right)^2\right] & \text{for } a\le x\le a+b\end{array}\right\};\ a\in\mathbb{R},\ b>0\\
\Lambda &= H_Y(y)\ = \begin{cases}\ln 2-\ln(1+y^2) & \text{for } -1\le y\le 0\\ \ln 2-\ln(1-y^2) & \text{for } 0\le y\le 1\end{cases}\\
y_\Lambda &= \begin{cases}\sqrt{2\,\exp(-\Lambda)-1} & \text{for } \Lambda\le\ln 2\approx 0.6931\\ \sqrt{1-2\,\exp(-\Lambda)} & \text{for } \Lambda\ge\ln 2\approx 0.6931\end{cases}\\
a &= x_{(\ln 2)}\approx x_{0.6931}\\
b &= x_{(\ln 2-\ln[1-0.5^2])}-x_{(\ln 2-\ln[1-(-0.5)^2])}\ \approx\ x_{0.9808}-x_{0.4700}
\end{aligned}$$

We now turn to those distributions that by a ln–transformation can be converted into a location–scale type. Generally, the original form of these distributions has three parameters, a location (shift) parameter a, a scale parameter b and a shape parameter c. The location parameter a must be known to make the ln–transformation. One way to find an estimate of a is by trial–and–error, i.e., a is chosen so that the hazard plot of $\widehat{\Lambda}_i$ over $\ln(x_i-a)$ is sufficiently linear.

Extreme value distribution of type II for the maximum (Inverse WEIBULL distribution)

$$S(x)=1-\exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right];\ a\in\mathbb{R},\ b>0,\ c>0,\ x\ge a$$

$X^*=\ln(X-a)$ has an extreme value distribution of type I for the maximum (GUMBEL distribution) with parameters $a^*=\ln b$ and $b^*=1/c$. Thus we find

$$\begin{aligned}
a^* &\approx x^*_{0.4587}\\
b^* &\approx x^*_{0.4587}-x^*_{0.0683}
\end{aligned}$$

on the $\ln(x-a)$–scaled abscissa of a hazard paper with an ordinate scaled as $\Lambda=-\ln\big\{1-\exp\big[-\exp(y)\big]\big\},\ y\in\mathbb{R}$. b and c follow as re–transforms of a^* and b^*.

Extreme value distribution of type II for the minimum (FRÉCHET distribution)

$$S(x)=\exp\left[-\left(\frac{a-x}{b}\right)^{-c}\right];\ a\in\mathbb{R},\ b>0,\ c>0,\ x\le a$$

$X^*=-\ln(a-X)$ has an extreme value distribution if type I for the minimum (Log–WEIBULL distribution) with parameters $a^*=-\ln b$ and $b^*=1/c$. Thus we find

$$\begin{aligned}
a^* &\approx x^*_1\\
b^* &\approx x^*_{2.7181}-x^*_1
\end{aligned}$$

on the $-\ln(a-x)$–scaled abscissa of a hazard paper with an ordinate scaled as $\Lambda = \exp(y)$, $y \in \mathbb{R}$. b and c follow as re–transforms of a^* and b^*.

Extreme value distribution of type III for the maximum (Reflected WEIBULL distribution)

$$S(x) = 1 - \exp\left[-\left(\frac{a-x}{b}\right)^c\right];\ a \in \mathbb{R},\ b > 0,\ c > 0,\ x \le a$$

$X^* = -\ln(a-X)$ has an extreme value distribution of type I for the maximum (GUMBEL distribution) with parameters $a^* = -\ln b$ and $b^* = 1/c$. Thus we find

$$\begin{aligned} a^* &\approx x^*_{0.4587} \\ b^* &\approx x^*_{0.4587} - x^*_{0.0683} \end{aligned}$$

on the $-\ln(x-a)$–scaled abscissa of a hazard paper with an ordinate scaled as $\Lambda = -\ln\big\{1 - \exp\big[-\exp(y)\big]\big\},\ y \in \mathbb{R}$. b and c follow as re–transforms of a^* and b^*.

Extreme value distribution of type III for the minimum (WEIBULL distribution)

$$S(x) = \exp\left[-\left(\frac{x-a}{b}\right)^c\right];\ a \in \mathbb{R},\ b > 0,\ c > 0,\ x \ge a$$

$X^* = \ln(X-a)$ has an extreme value distribution if type I for the minimum (log–GUMBEL distribution) with parameters $a^* = \ln b$ and $b^* = 1/c$. Thus we find

$$\begin{aligned} a^* &\approx x^*_1 \\ b^* &\approx x^*_{2.7181} - x^*_1 \end{aligned}$$

on the $\ln(x-a)$–scaled abscissa of a hazard paper with an ordinate scaled as $\Lambda = \exp(y),\ y \in \mathbb{R}$. b and c follow as re–transforms of a^* and b^*.

Log–normal distribution with lower threshold

$$S(x) = \Phi\left(-\frac{\ln(x-a) - \widetilde{a}}{\widetilde{b}}\right);\ a \in \mathbb{R},\ \widetilde{a} \in \mathbb{R},\ \widetilde{b} > 0,\ x > a$$

$X^* = \ln(X-a)$ has a normal distribution with parameters $\widetilde{a} = \mathrm{E}(X^*)$ and $\widetilde{b} = \sqrt{\mathrm{Var}(X^*)}$. Thus we find

$$\begin{aligned} \widetilde{a} &\approx x^*_{0.6931} \\ \widetilde{b} &\approx x^*_{1.8407} - x^*_{0.6931} \end{aligned}$$

on the $\ln(x-a)$–scaled abscissa of a hazard paper with an ordinate scaled as $\Lambda = -\ln\big[1 - \Phi(y)\big],\ y \in \mathbb{R}$.

Log–normal distribution with upper threshold

$$S(x) = \Phi\left(-\frac{\ln(a-x) - \widetilde{a}}{\widetilde{b}}\right);\ a \in \mathbb{R},\ \widetilde{a} \in \mathbb{R},\ \widetilde{b} > 0,\ x < a$$

$X^* = \ln(a-x)$ has a normal distribution with parameters $\widetilde{a} = \mathrm{E}(X^*)$ and $\widetilde{b} = \sqrt{\mathrm{Var}(X^*)}$. $\widetilde{a}$ and $\widetilde{b}$ are found in the same way as above.

PARETO distribution of the first kind

$$S(x) = \left(\frac{x-a}{b}\right)^{-c};\ a \in \mathbb{R},\ b > 0, c > 0,\ x \geq a + b$$

$X^* = \ln(X - a)$ has an exponential distribution with parameters $a^* = \ln b$ and $b^* = 1/c$. Thus we find

$$\begin{aligned} a^* &= x_0^* \\ b^* &= x_1^* - x_0^* \end{aligned}$$

on the $\ln(x - a)$–scaled abscissa of a hazard paper with an ordinate scaled as $\Lambda = y,\ y \geq 0$. b and c follow as re–transforms of a^* and b^*.

Power function distribution

$$S(x) = 1 - \left(\frac{x-a}{b}\right)^{c};\ a \in \mathbb{R},\ b > 0, c > 0,\ a \leq x \leq a + b$$

$X^* = \ln(X - a)$ has a reflected exponential distribution with parameters $a^* = \ln b$ and $b^* = 1/c$. Thus we find

$$\begin{aligned} a^* &= x_0^* \\ b^* &\approx x_{0.9808}^* - x_{0.4700}^* \end{aligned}$$

on the $\ln(x - a)$–scaled abscissa of a hazard paper with an ordinate scaled as $\Lambda = -\ln(1 - y)$, $0 \leq y \leq 1$. b and c follow as re–transforms of a^* and b^*.

10 Testing Hypotheses on Life Distributions

This chapter is mainly devoted to testing hypotheses concerning the hazard rate (Sect. 10.2), but in Sect. 10.3 we will also look for tests deciding whether a life distribution has one or the other aging property, which have been introduced in Sect. 2.4. The approaches of this chapter are non–parametric, the exception being Sect. 10.2.1 where we test the constancy of the hazard rate. As the only continuous distribution with constant hazard rate is the exponential distribution, the tests of Sect. 10.2.1 will be test for an exponential distribution.

This chapter is organized as follows:

- Sect. 10.1 present those statistical concepts which directly or indirectly give the test statistics for most of the subsequent tests. These concepts are order statistics which lead to spacings and the TTT–statistics which are a function of spacings.
- Sect. 10.2 examines the behavior of the hazard rate, i.e., its course and curvature.
- Finally, the topic of Sect. 10.3 are tests for several classes of aging.

A topic non treated here is that of comparing hazard rates, see QIU/SHENG (2008), and of testing the equality of two hazard rates, see CHENG (1985).

10.1 Prerequisites: Order Statistics, Spacings, TTT–statistics[1]

In life testing without replacement of failed units the observed life times naturally occur in ascending order. Thus, it is obvious to investigate order statistics or function thereof like spacings and TTT–statistics. These concepts will furnish us with test statistics for a lot of hypotheses on life distributions.

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ denote a sample of n independent random variables, each with CDF $F(\cdot)$ and PDF $f(\cdot)$, and let $\boldsymbol{X_n} = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ denote the vector of the associated **order statistics**: $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$. The joint PDF of any subset of the order statistics $X_{(k_1)}, X_{(k_2)}, \ldots, X_{(k_m)}$ with $1 \leq k_1 < k_2 < \ldots < k_m \leq n$ is given by PYKE (1965) as

$$f_{\boldsymbol{X_k}}(x_1, \ldots, x_m) = \left\{ \begin{array}{ll} n! \prod\limits_{j=1}^{m+1} \Big[F(x_j) - F(x_{j-1})\Big]^{k_j - k_{j-1} - 1} \times & \\ \qquad \times \dfrac{f(x_j)}{(k_j - k_{j-1} - 1)!} & \text{if } x_1 < x_2 < \ldots < x_m \\ \qquad 0 & \text{otherwise.} \end{array} \right\} \tag{10.1}$$

where $x_0 = -\infty,\ x_{m+1} = \infty,\ F(x_0) = 0,\ F(x_{m+1}) = 1,\ k_{m+1} = n+1,\ k_0 = 0$ and $f(x_{m+1}) = 1$.

---

[1] Suggested reading for this section: BALAKRISHNAN/COHEN (1991), BARLOW (1979), BARLOW/CAMPO (1975), BARLOW et al. (1972), BERGMAN (1979), BERGMNA/KLEFSJÖ (1984), DAVID/NAGARAJA (2003), KLEFSJÖ (1982b), PYKE (1965).

We look at special subsets of $\boldsymbol{X_n}$.

1. $m = 1$ — a single order statistic $X_\ell;\ \ell = 1,\, 2,\, \ldots,\, n$

$$f_{X_{(\ell)}}(x) = \frac{n!}{(\ell-1)!\,(n-\ell)!}\, f(x)\big[F(x)\big]^{\ell-1}\big[1-F(x)\big]^{n-\ell},\ x \in \mathbb{R} \quad (10.2a)$$

$$F_{X_{(\ell)}}(x) = \sum_{i=\ell}^{n}\binom{n}{i}\big[F(x)\big]^{i}\big[1-F(x)\big]^{n-i},\ x \in \mathbb{R} \quad (10.2b)$$

$F_{X_\ell}(x)$ is the CCDF of a binomial distribution with parameters n and $P = F(x)$ and may be expressed by an incomplete beta function ratio (= CDF of a beta distribution with parameters ℓ and $n-\ell+1$.)

Especially, we have:

- $\ell = 1$ — **sample minimum**

$$f_{X_{(1)}}(x) = n\, f(x)\big[1-F(x)\big]^{n-1},\ x \in \mathbb{R} \quad (10.2c)$$

$$F_{X_{(1)}}(x) = 1-\big[1-F(x)\big]^{n},\ x \in \mathbb{R} \quad (10.2d)$$

- $\ell = n$ — **sample maximum**

$$f_{X_{(n)}}(x) = n\, f(x)\big[F(x)\big]^{n-1},\ x \in \mathbb{R} \quad (10.2e)$$

$$F_{X_{(n)}}(x) = \big[F(x)\big]^{n},\ x \in \mathbb{R}. \quad (10.2f)$$

2. $m = 2$ — a pair of order statistics $(X_{(k)},\, X_{(\ell)});\ k,\, \ell = 1,\, 2,\, \ldots\, n;\ k \neq \ell$

$$f_{X_{(k)},X_{(\ell)}}(x,y) = \left\{\begin{array}{l} \dfrac{n!}{(k-1)!(\ell-k-1)!(n-\ell)!}\, f(x)\, f(y)\big[F(x)\big]^{k-1}\times \\ \quad \times\big[F(y)-F(x)\big]^{\ell-k-1}\big[1-F(y)\big]^{n-\ell} \\ \quad \text{for } k<\ell,\ x<y \text{ and } x,\, y \in \mathbb{R} \\ 0 \quad \text{otherwise} \end{array}\right\} \quad (10.3a)$$

$$F_{X_{(k)},X_{(\ell)}}(x,y) = \left\{\begin{array}{l} \dfrac{n!}{(k-1)!(\ell-k-1)!(n-\ell)!}\displaystyle\int\limits_{0}^{F(x)}\int\limits_{u}^{F(y)} u^{k-1}\times \\ \quad \times (v-u)^{\ell-k-1}(1-v)^{n-\ell}\, dv\, du \\ \quad \text{for } k<\ell,\ x<y \text{ and } x,\, y \in \mathbb{R} \end{array}\right\} \quad (10.3b)$$

Especially, we have:

- $k = i,\ \ell = i+1,\ i = 1,\, 2,\, \ldots,\, n-1$ (two adjacent order statistics)

$$f_{X_{(i)},X_{(i+1)}}(x,y) = \left\{\begin{array}{l} \dfrac{n!}{(i-1)!(n-i-1)!}\, f(x)\, f(y)\big[F(x)\big]^{i-1}\times \\ \quad \times\big[1-F(y)\big]^{n-i-1} \\ \quad \text{for } i = 1,\, 2,\, \ldots,\, n-1,\ x<y \text{ and } x,\, y \in \mathbb{R} \\ 0 \quad \text{otherwise} \end{array}\right\} \quad (10.3c)$$

- $k = 1,\ \ell = n$ (sample minimum and sample maximum)

$$f_{X_{(1)},X_{(n)}}(x,y) = \left\{\begin{array}{l} n\,(n-1)\,f(x)\,f(y)\,\big[F(y)-F(x)\big]^{n-1} \\ \quad \text{for } x<y \text{ and } x,\,y\in\mathbb{R} \\ 0 \quad \text{otherwise} \end{array}\right\} \tag{10.3d}$$

3. $m = n$ — the complete vector of order statistics $\boldsymbol{X_n} = (X_{(1)},\,\ldots,\,X_{(n)})$

$$f_{\boldsymbol{X_n}}(x_1,\,\ldots,\,x_n) = \left\{\begin{array}{cc} n!\,f(x_1)\cdot\,\ldots\,\cdot f(x_n) & \text{for } x_1<\ldots<x_n \text{ and } x_i\in\mathbb{R}\,\forall\,i \\ 0 & \text{otherwise} \end{array}\right\} \tag{10.4}$$

Order statistics form a MARKOV process, more precisely: $\big\{X_{(i)},\ 1\le i\le n\big\}$ is a non-homogeneous, discrete-parameter, real-valued MARKOV process whose initial measure is given by (10.2d):

$$F_{X_{(1)}}(x) = 1 - \big[1-F(x)\big]^n,$$

and whose transition-CDF $\Pr(X_{(i+1)}\le x\,|\,X_{(i)}=y)$ is the CDF of the minimum of $n-i$ independent observations of the CDF $F(\cdot)$ truncated at y, namely:

$$\Pr(X_{(i+1)}\le x\,|\,X_{(i)}=y) = 1-\big[1-F(x)\big]^{n-i}\big[1-F(y)\big]^{-n+i} \ \text{ for } x>y. \tag{10.5}$$

The PDF and CDF of order statistics from most parametric distributions have no nicely looking and closed formulas. Two exceptions are the uniform and the exponential distributions which play a dominant role in testing hypotheses on life distributions.

1. Order statistics from the reduced uniform distribution

$$f(x) = \left\{\begin{array}{ll} 1 & \text{for } 0\le x\le 1 \\ 0 & \text{otherwise} \end{array}\right\} \tag{10.6a}$$

$$F(x) = \left\{\begin{array}{ll} 0 & \text{for } x<0 \\ x & \text{for } 0\le x\le 1 \\ 1 & \text{for } x>1 \end{array}\right\} \tag{10.6b}$$

$$f_{X_\ell}(x) = \frac{n!}{(\ell-1)!(n-\ell)!}\,x^{\ell-1}\,(1-x)^{n-\ell} \quad \text{for } \ell=1,\,\ldots,\,n;\ 0\le x\le 1 \tag{10.6c}$$

$$f_{X_k,X_\ell}(x,y) = \frac{n!}{(k-1)!(\ell-k-1)!(n-\ell)!}\,x^{k-1}(y-x)^{\ell-k-1}\times \times(1-y)^{n-\ell} \ \text{ for } k<\ell,\ x<y \text{ and } x,y\in[0,1] \tag{10.6d}$$

$$f_{\boldsymbol{X_n}}(x_1,\,\ldots,\,x_n) = n! \ \text{ for } 0\le x_1<x_2<\ldots<x_n\le 1 \tag{10.6e}$$

$$\mathrm{E}(X_{(\ell)}) = \frac{\ell}{n+1};\ \ell=1,\,2,\,\ldots,\,n \tag{10.6f}$$

$$\mathrm{E}\big(X_{(\ell)}^2\big) = \frac{\ell+1}{n+2}\,\frac{\ell}{n+1};\ \ell=1,\,2,\,\ldots,\,n \tag{10.6g}$$

$$\operatorname{Var}(X_{(\ell)}) = \frac{\ell\,(n-\ell+1)}{(n+1)^2\,(n+2)};\; ell = 1,\,2,\,\ldots,\,n \tag{10.6h}$$

$$\operatorname{Cov}(X_{(k)},X_{(\ell)}) = \frac{k\,(n-\ell+1)}{(n+1)^2\,(n+2)};\; k<\ell;\; k,\ell = 1,\,\ldots,\,n \tag{10.6i}$$

$$\operatorname{Cor}(X_{(k)},X_{(\ell)}) = \sqrt{\frac{k\,(n-\ell+1)}{\ell\,(n-k+1)}};\; k<\ell;\; k,\ell = 1,\,\ldots,\,n \tag{10.6j}$$

2. Order statistics from the exponential distribution

$$f(x) = \begin{Bmatrix} \frac{1}{b}\exp\Big(-\frac{x}{b}\Big) & \text{for } x\ge 0,\; b>0 \\ 0 & \text{otherwise} \end{Bmatrix} \tag{10.7a}$$

$$F(x) = \begin{Bmatrix} 0 & \text{for } x<0 \\ 1-\exp\Big(-\frac{x}{b}\Big) & \text{for } x\ge 0 \end{Bmatrix} \tag{10.7b}$$

$$\begin{aligned} f_{X_{(\ell)}}(x) &= \frac{n!}{b\,(\ell-1)!(n-\ell)!}\,\exp\Big[-\frac{(n-\ell+1)\,x}{b}\Big]\times \\ &\quad\times\Big[1-\exp\Big(-\frac{x}{b}\Big)\Big]^{\ell-1};\; x\ge 0 \end{aligned} \tag{10.7c}$$

$$\begin{aligned} f_{X_{(k)},X_{(\ell)}}(x,y) &= \frac{n!}{b^2\,(k-1)!(\ell-k-1)!(n-\ell)!}\,\exp\Big[-\frac{x+y\,(n-\ell+1)\,x}{b}\Big]\times \\ &\quad\times\Big[1-\exp\Big(-\frac{x}{b}\Big)\Big]^{k-1}\Big[\exp\Big(-\frac{x}{b}\Big)-\exp\Big(-\frac{y}{b}\Big)\Big]^{\ell-k-1}; \\ &\quad k<\ell,\; x<y \text{ and } x,y\ge 0 \end{aligned} \tag{10.7d}$$

$$f_{\boldsymbol{X_n}}(x_1,\,\ldots,\,x_n) = \frac{n!}{b^n}\,\exp\Bigg(-\frac{1}{b}\sum_{i=1}^{n}x_i\Bigg);\; 0\le x_1<x_2<\ldots<x_n<\infty \tag{10.7e}$$

$$\operatorname{E}(X_{(\ell)}) = b\sum_{i=1}^{\ell}\frac{1}{n-i+1};\; \ell = 1,\,2,\,\ldots,\,n \tag{10.7f}$$

$$\operatorname{E}\big(X_{(\ell)}^2\big) = b^2\Bigg[\sum_{i=1}^{\ell}\frac{1}{(n-i+1)^2}+\Bigg(\sum_{i=1}^{\ell}\frac{1}{n-i+1}\Bigg)^2\Bigg];\ell = 1,\,2,\,\ldots,\,n \tag{10.7g}$$

$$\operatorname{Var}(X_{(\ell)}) = b^2\sum_{i=1}^{\ell}\frac{1}{(n-i+1)^2};\; \ell = 1,\,2,\,\ldots,\,n \tag{10.7h}$$

$$\operatorname{Cov}(X_{(k)},X_{(\ell)}) = \operatorname{Var}(X_{(k)}) = b^2\sum_{i=1}^{k}\frac{1}{(n-i+1)^2};\; k<\ell;\; k,\ell = 1,\,\ldots,\,n \tag{10.7i}$$

$$\operatorname{Cor}(X_{(k)},X_{(\ell)}) = \sqrt{\frac{\sum_{i=1}^{k}\frac{1}{(n-i+1)^2}}{\sum_{i=1}^{\ell}\frac{1}{(n-i+1)^2}}};\; k<\ell;\; k,\ell = 1,\,\ldots,\,n \tag{10.7j}$$

The difference S_i of two adjacent order statistics $X_{(i-1)},\ X_{(i)}$ is called **spacing**:

$$S_i = X_{(i)} - X_{(i-1)};\ i = 2,\,3,\,\ldots,\,n. \tag{10.8a}$$

For a lifetime variable the spacing is nothing but the waiting time (time elapsed) between the $(i+1)$-st and the i-th failure. Let

$$\boldsymbol{S} = (S_2,\,\ldots,\,S_n) \tag{10.8b}$$

be the vector of all spacings in a sample of size n, then by means of a linear transformation of (10.4) we find

$$f_{\boldsymbol{S}}(s_2,\,\ldots,\,s_n) = \left\{\begin{array}{cc} n!\int\limits_{-\infty}^{\infty}\prod\limits_{i=2}^{n} f(x+s_2+\ldots+s_i)\,dx & \text{for } s_i > 0,\ 2 \le i \le n \\ 0 & \text{otherwise.}\end{array}\right\} \tag{10.8c}$$

For a single spacing we have from (10.3c)

$$\begin{aligned} f_{S_i}(x) &= \int\limits_{-\infty}^{\infty} f_{X_{(i-1)},X_{(i)}}(y,\,y+x)\,dy \\ &= \frac{n!}{(i-2)!(n-i)!}\int\limits_{-\infty}^{\infty}\big[F(y)\big]^{i-2}\big[1-F(y+x)\big]^{n-i} f(y)f(y+x)dy; \\ &\quad 2 \le i \le n,\ x > 0 \end{aligned} \tag{10.8d}$$

As this expression indicates, the formulas for the PDFs of sets of arbitrary spacings are not particularly simple, although they are derived straightforwardly. PYKE (1965, p. 400), based on the MARKOV property of order statistics, gives the following PDF of $(S_i,\,S_j),\ i \ne j$,

$$\begin{aligned} f_{S_i,S_j}(u,v) &= \frac{n!}{(i-2)!(j-i-2)!(n-j)!}\int\limits_{-\infty}^{\infty}\int\limits_{x+u}^{\infty}[F(x)]^{i-2}[F(y)-F(x+u)]^{j-i-2}\ \times \\ &\quad \times [1-F(y+v)]^{n-j} f(x)\,f(x+u)\,f(y)\,f(y+v)\,dy\,dx; \\ &\quad i \ne j;\ u,v > 0. \end{aligned} \tag{10.8e}$$

For two adjacent spacings S_i and S_{i+1} we have

$$\begin{aligned} f_{S_i,S_{i+1}}(u,v) &= \frac{n!}{(i-2)!(n-i-1)!}\int\limits_{-\infty}^{\infty}[F(x)]^{i-2}\,[1-F(x+u+v)]^{n-i-1}\ \times \\ &\quad \times f(x)\,f(x+u)\,f(x+u+v)\,dx; i = 2,\,\ldots,\,n-1;\ u,v > 0. \end{aligned} \tag{10.8f}$$

We now look at spacings from the uniform and the exponential distributions. These spacings are of special interest in lifetime analysis.

Spacings of the reduced uniform distribution

Let $\boldsymbol{X_n} = (X_{(1)},\,X_{(2)},\,\ldots,\,X_{(n)})$ denote the order statistics in a sample of size n from the reduced uniform distribution (10.6a,b). Set $X_{(0)} = 0$ and $X_{(n+1)} = 1$. The spacings are defined by

$$S_i = X_{(i)} - X_{(i-1)}\ \text{ for }\ 1 \le i \le n+1. \tag{10.9a}$$

Since $S_1 + S_2 + \ldots + S_n = 1$, the random vector $\boldsymbol{S} = (S_1,\,S_2,\,\ldots,\,S_{n+1})$ has a singular distribution, but when restricted to this hyperplane has the PDF

$$f_{\boldsymbol{S}}(s_1,\,\ldots,\,s_{n+1}) = n!\ \text{ if }\ s_i \ge 0 \text{ and } s_1 + \ldots + s_{n+1} = 1. \tag{10.9b}$$

It follows from (10.9b) that the PDF of $\boldsymbol{S}$ remains unchanged under any permutation of its coordinates, i.e., uniform spacings are interchangeable variates. This implies in particular, that the PDF of any S_i is equal to that of $S_1 = X_{(1)}$ and the joint PDF of any pair $(S_i,\, S_j),\ i \neq j$, is the same as that of $(S_1,\, S_2)$. So we have

$$F_{S_i}(x) \;=\; F_{S_1}(x) \;=\; F_{X_{(1)}}(x) \;=\; 1-(1-x)^n,\ 0 \le x \le 1, \tag{10.9c}$$

$$f_{S_i}(x) \;=\; f_{S_1}(x) \;=\; f_{X_{(1)}}(x) \;=\; n\,(1-x)^{n-1},\ 0 \le x \le 1, \tag{10.9d}$$

and for $x,\, y \ge 0$ with $x + y \le 1$

$$\begin{aligned} F_{S_i,S_j}(x,y) &= \Pr(X_{(1)} \le x,\, X_{(2)} - X_{(1)} \le y) \\ &= n\int\limits_0^1 \left[1-\left(1-\frac{y}{1-u}\right)^{n-1}\right](1-u)^{n-1}du \\ &= 1-\big[(1-x)^n + (1-y)^n - (1-x-y)^n, \qquad (10.9e) \\ f_{S_i,S_j}(x,y) &= n\,(n-1)\,(1-x-y)^{n-2}. \qquad (10.9f) \end{aligned}$$

For the single moments of S_i we find

$$\mathrm{E}(S_i) \;=\; \mathrm{E}(S_1) \;=\; \mathrm{E}(X_{(1)}) \;=\; \frac{1}{n+1}, \tag{10.9g}$$

$$\mathrm{E}(S_i^2) \;=\; \mathrm{E}(S_1^2) \;=\; \mathrm{E}(X_{(1)}^2) \;=\; \frac{2}{(n+2)\,(n+1)}, \tag{10.9h}$$

$$\mathrm{Var}(S_i) \;=\; \mathrm{Var}(S_1) \;=\; \mathrm{Var}(X_{(1)}) \;=\; \frac{n}{(n+2)\,(n+1)^2} \tag{10.9i}$$

and, since

$$\mathrm{E}(S_2 \,|\, S_1 = u) = \frac{1-u}{n} \ \text{ for } \ 0 < u < 1,$$

the product moments are

$$\mathrm{E}(S_1\,S_2) \;=\; \frac{\mathrm{E}[S_1(1{-}S_1)]}{n} = \frac{1}{(n{+}1)(n{+}2)} = \mathrm{E}(S_i\,S_j);\ i \neq j;\ i,j = 1,\ldots,n{+}1, \tag{10.9j}$$

$$\mathrm{Cov}(S_i,S_j) \;=\; -\frac{1}{(n+1)^2\,(n+2)};\ i \neq j;\ i,j = 1,\,\ldots,\,n+1, \tag{10.9k}$$

$$\mathrm{Cor}(S_i,S_i) \;=\; -\frac{1}{n};\ i\ \neq j;\ i,j = 1,\,\ldots,\,n+1. \tag{10.9l}$$

Spacings of the exponential distribution

Let $\boldsymbol{X_n} = (X_{(1)},\, X_{(2)},\, \ldots,\, X_{(n)})$ denote the set of order statistics from an exponential distribution with $f(x) = (1/b)\,\exp[-(x/b)],\ x \ge 0$. Set $X_{(0)} = 0$ and $S_i = X_{(i)} - X_{(i-1)},\ 1 \le i \le n$. Then for the vector

$$\boldsymbol{S} = (S_1,\, S_2,\, \ldots,\, S_n)$$

the joint PDF follows from (10.4) as

$$\begin{aligned} f_{\boldsymbol{S}}(s_1,\,\ldots,\,s_n) &= \frac{n!}{b^n}\prod_{i=1}^n \exp\left[-\frac{s_1+\ldots+s_i}{b}\right] \\ &= \frac{n!}{b^n}\,\exp\left[-\frac{1}{b}\sum_{i=1}^n (n-i+1)\,s_i\right] \\ &= \prod_{i=1}^n \frac{n-i+1}{b}\,\exp\left[-\frac{(n-i+1)\,s_i}{b}\right];\ s_i > 0;\ 1 \le i \le n. \qquad (10.10a) \end{aligned}$$

Hence, the joint PDF of exponential spacings is the product of n marginal exponential densities

$$f_{S_i}(s) = \frac{n-i+1}{b}\,\exp\left[-\frac{(n-i+1)\,s}{b}\right];\ s>0;\ i=1,\,\ldots,\,n; \tag{10.10b}$$

the parameters being $n/b,\ (n-1)/b,\,\ldots,\,b$ and

$$\mathrm{E}(S_i) = \frac{b}{n-i+1}, \tag{10.10c}$$

$$\mathrm{Var}(S_i) = \frac{b^2}{(n-i+1)^2}. \tag{10.10d}$$

Equivalently, one may say that

$$D_i^* = \frac{n-i+1}{b}\,S_i;\ 1\le i\le n; \tag{10.10e}$$

the **normalized exponential spacings**, are iid exponential variates with parameter $b=1$ and $\mathrm{E}(D_i^*) = \mathrm{Var}(D_i^*) = 1;\ i=1,\,\ldots,\,n.$

For a general variate X the quantity

$$D_i = (n-i+1)\,(X_{(i)} - X_{(i-1)}) \tag{10.11}$$

is simply called a **normalized spacing**. It plays a dominant role in testing hypotheses on the hazard rate.

Let $0 = X_{(0)} < X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ denote an ordered sample of size n from a life distribution with CDF $F(\cdot),\ F(0^-)=0$, survival function $S(\cdot) = 1-F(\cdot)$ and finite mean $\mu = \int_o^\infty S(x)\,dx$. T_n, the total time spent on test by the n sample units until the failure of the longest living unit, may be expressed in two different ways:

1. as the sum of all observed life spans

$$T_n = \sum_{i=1}^{n} X_{(i)}, \tag{10.12a}$$

   i.e., as the area given by n horizontal beams of length $X_{(i)}$, each having height equal to 1, see the upper display in Fig. 10/1, or

2. as the sum of all normalized spacings

$$T_n = \sum_{i=1}^{n}(n-i+1)\,(X_{(i)} - X_{(i-1)}) = \sum_{i=1}^{n} D_i, \tag{10.12b}$$

   i.e., as the area given by n vertical beams having length $D_i = X_{(i)} - X_{(i-1)}$ and corresponding height $n-i+1$. Such a vertical beam gives the time spent on test of those $n-i+1$ units having lived in the interval $(X_{(i)} - X_{(i-1)})$, see the lower display in Fig. 10/1.

T_n is known as **TTT–statistic** (**total–time–on–test statistic**). The **successive TTT–statistics** are defined

- according to (10.12a) as

$$T_i = \sum_{j=1}^{i} X_{(j)} + (n-i)\,X_{(i)}, \tag{10.12c}$$

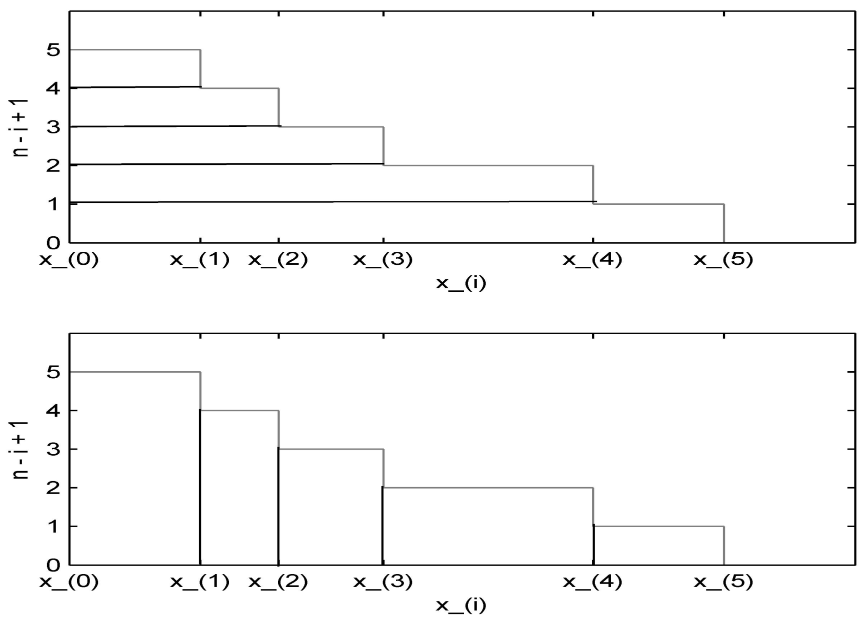
- according to (10.12b) as

$$T_i = \sum_{j=1}^{i}(n-j+1)\,(X_{(j)}-X_{(j-1)}) = \sum_{j=1}^{i} D_j. \tag{10.12d}$$

The **scaled TTT–statistics**, defined on $[0,1]$, are

$$T_i^* = \frac{T_i}{T_n};\; i = 1,\,2,\,\ldots,\,n. \tag{10.12e}$$

Figure 10/1: Two ways of expressing the total time spent on test



By plotting T_i^* on the ordinate against the empirical CDF $F_n(x_{(i)}) = i/n$ on the abscissa and then connecting these points by straight lines we obtain a curve within the unit square of the (x,y)–plane, called **TTT–plot**. The message of the TTT–plot is easy to understand: The shortest living $100\,(i/n)\%$ of the sample units contribute $100\,(T_i^*)\%$ of the total time lived by all sample units.[2]

TTT–statistics were first used by EPSTEIN/SOBEL (1953) to make inference about the exponential distribution. Starting with a paper by BARLOW/CAMPO (1975), researchers have studied generalizations of the original TTT–concept that have proven to be very useful in a great number of applications, e.g., for model identifications, as a basis for the characterization of life distribution classes, for hypothesis testing, and to determine optimal age replacement intervals.

The empirical quantities defined in (10.12a-e) have theoretical counterparts.

$$H_F^{-1}(P) = \int\limits_0^{F^{-1}(P)} S(x)\,dx,\; 0 \le P \le 1, \tag{10.13a}$$

[2] The TTT–plot resembles the LORENZ–curve which is used in economics to describe the inequality in income distributions. Contrary to the TTT–plot the LORENZ–curve is never situated above the $45°$–line.

the counterpart of T_i, is called the **TTT–transform** of $F(\cdot)$. $F^{-1}(P)$ is the percentile x_P of order P. For $P = 1$ we have

$$\int\limits_0^{F^{-1}(P)} S(x)\,dx = \mathrm{E}(X) = \mu. \tag{10.13b}$$

The counterpart of T_i^* in (10.12e) is

$$\phi_F(P) = \frac{H_F^{-1}(P)}{H_F^{-1}(1)} = \frac{H_F^{-1}(P)}{\mu}, \tag{10.13c}$$

the **scaled TTT–transform** of $F(\cdot)$; for examples see Fig. 10/4 and 10/5.

We realize — e.g. Fig 10/2 — that the TTT–plot of a sample from a population with $F(\cdot)$ will approach the graph of the scaled TTT–transform $\phi_F(P)$ of $F(\cdot)$ as n, the sample size, increases. This is so, because

$$T_i = \sum_{j=1}^{i}(n-j+1)\,(X_{(j)} - X_{(j-1)}) = \int\limits_0^{F_n^{-1}(i/n)} \big[1 - F_n(x)\big]\,dx,$$

$F_n(\cdot)$ being the empirical CDF and, because by the GLIVENKO—CANTELLI theorem and the strong law of large numbers with probability one:

$$T_i \;\longrightarrow\; \int\limits_0^{F^{-1}(P)} \big[1 - F(x)\big]\,dx \;\text{ uniformly in }\; [0,1]$$

when $n \to \infty$ and $i/n \to F(\cdot)$.

It is easy to verify that for an exponential distribution with $F(x) = 1 - \exp[-(x/b)]$, $F^{-1}(P) = -b\,\ln(1-P)$ and $\mu = b$ we have

$$H_F^{-1}(P) = \int\limits_0^{-b\,\ln(1-P)} \exp[-(u/b)]\,dx = b\,P \tag{10.14a}$$

and

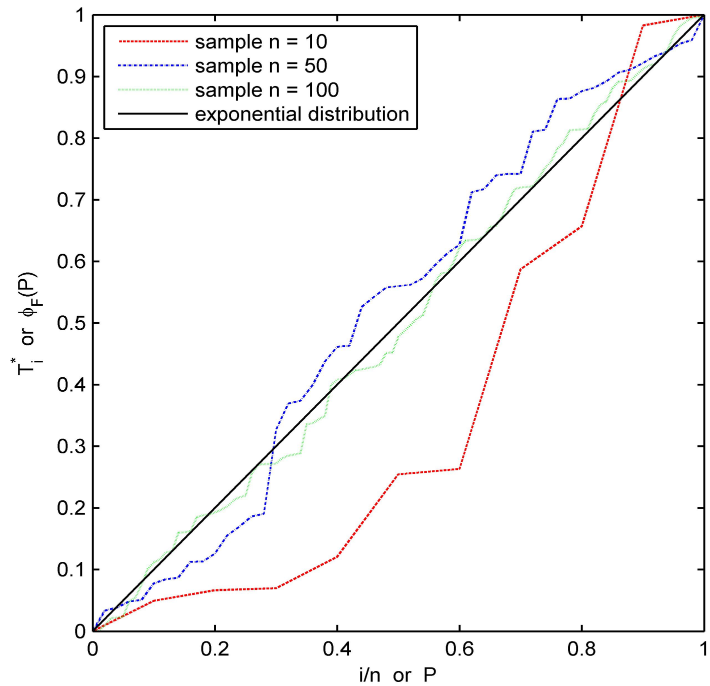
$$\phi_F(P) = \frac{b\,P}{P} = P,\; 0 \le P \le 1. \tag{10.14b}$$

So, the TTT–plot will be a $45°$–line running from $(0,0)$ to $(1,1)$, see Fig. 10/2. The notation of the TTT–transform as an inverse CDF indicates that the inverse of $H_F^{-1}(P)$ will be a CDF of some variate Y with support $[0,\mu]$. In case of (10.14a) the corresponding CDF is

$$H_F(y) = P = y/b,\; 0 \le y \le b, \tag{10.14c}$$

i.e., a uniform distribution on $[0,b]$.

Figure 10/2: TTT–plots based on simulated exponential data ($b = 5$; $n = 10, 50, 100$) and the scaled TTT–transform of the exponential distribution



We state the following properties of the TTT–transform $H_F^{-1}(P)$:[3]

1. There is a one–to–one correspondence between life distributions and their TTT–transforms.
2. If $F(\cdot)$ is strictly increasing or, equivalently, if $F^{-1}(\cdot)$ is continuous, then $H_F^{-1}(\cdot)$ is continuous.
3. If $F(\cdot)$ is absolutely continuous and strictly increasing the derivative of $H_F^{-1}(P)$ is found to be
$$\frac{dH_F^{-1}(P)}{dP} = \frac{1 - F(x)}{f(x)} = \frac{1}{h(x)}, \tag{10.15}$$
for almost all $P \in [0, 1]$, where $h(\cdot)$ is the hazard rate. The property, that the derivative of the TTT–transform $\phi_F(P) = H_F^{-1}(P)/\mu$ is proportional to the reciprocal of the hazard rate, is of utmost importance in finding test statistics for hypotheses on the hazard rate in later sections.

10.2 Testing Hazard Rate Properties

We are interested in the behavior of the hazard rate and want to know whether it is constant (Sect. 10.2.1), monotonely increasing or decreasing (Sect. 10.2.2) or has a bathtub shape or an upside–down bathtub shape (Sect. 10.2.3)

10.2.1 Constancy of the Hazard Rate[4]

The exponential distribution is the only continuous distribution with constant hazard rate. Thus, testing

[3] For more properties see BARLOW et al. (1972), BARLOW/CAMPO (1975) and BARLOW (1979).

[4] Suggested reading for this section: DOKSUM/YANDELL (1984), EPSTEIN (1960), EPSTEIN/SOBEL (1953).

$$\text{H}_0: \text{`}h(x) \text{ is constant.'} \quad \text{against} \quad \text{H}_\text{A}: \text{`}h(x) \text{ is not constant.'}$$

amounts to decide whether the sample comes from an exponential distribution or not.[5] For this purpose we may revert to one of the numerous existing goodness–of–fit tests, e.g., the tests of KOLMOGOROV–SMIRNOV, ANDERSON–DARLING, CRAMÉR–VON MISES or WATSON.

Here, we will only recommend informal procedures which are based on graphs to be judged personally. These approaches have neither a test statistic nor a calculable error probability. The accompanying MATLAB–program *HAZARD_09* produces the following graphs.

1) Plot of the transformed empirical CDF

The CDF of the general exponential distribution reads

$$F(x) = 1 - \exp\left(-\frac{x-a}{b}\right);\ a \in \mathbb{R},\ b > 0,\ x \geq a. \tag{10.16a}$$

It follows

$$y = \ln\left[\frac{1}{1-F(x)}\right] = \frac{x-a}{b}. \tag{10.16b}$$

So, if we plot $y = \ln\{1/[1-F(x)]\}$ against x — what is nothing but the probability plot of the exponential distribution, see RINNE (2010, pp. 118 ff.) — we get a straight line with slope $1/b$ cutting the x–axis at the point a. This suggests the following procedure for testing departures from exponentiality:

1.1) If we have an *uncensored sample* of n items, $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ being the ordered times of failure, we estimate $F(x)$ by

$$\widehat{F}_n(x_{(i)}) = \frac{i}{n};\ i = 1, 2, \ldots, n; \tag{10.16c}$$

and plot

$$\widehat{y}_i = \ln\left[\frac{1}{1-\widehat{F}_n(x_{(i)})}\right] \tag{10.16d}$$

against $x_{(i)}$. The points $(x_{(i)}, \widehat{y}_i)$ will fluctuate at random around a straight line for an exponential population.

1.2) If we have a *singly censored* type–I or type–II *sample* with altogether k failures observed, k being fixed for type–II and random for type–I, we evaluate (10.16c,d) for $i = 1, 2, \ldots, k$ and find a good linear fit up to $\big(x_{(k)}, \ln\left[1/(1-k/n)\right]\big)$ under exponentiality.

1.3) If we have a *multiply censored* or *randomly censored sample* with observations $(y_{(i)}, \delta_i)$, $\delta_i = 1$ for uncensored data and $\delta_i = 0$ for censored data, we take the KAPLAN/MEIER estimator

$$\widehat{S}_n(x_{(i)}) = \prod_{j=1}^{i}\left(\frac{n-j}{n-j+1}\right)^{\delta_i};\ 1 = 1, 2, \ldots, n. \tag{10.16e}$$

Then we look for linearity in the plot of

$$\widehat{y}_i = \ln\left[\frac{1}{\widehat{S}_n(x_{(i)})}\right];\ \delta_i = 1, \tag{10.16f}$$

against the uncensored observations $(x_{(i)}, 1)$.

[5] Testing constancy of the hazard rate for a discrete distribution based on the score statistics of the log–likelihood function is described in SCHIFFMAN (1986).

2) TTT–plot

The scaled TTT–transform of an exponential distribution is $\phi_F(P) = P$, i.e., a $45°$–line running from $(0,0)$ to $(1,1)$ and the scaled TTT–statistic T_i^* from an exponential sample will deviate randomly from this $45°$–line.

2.1) For an *uncensored sample* of size n we plot T_i^* on the ordinate against i/n on the abscissa.

2.2) For a *singly censored* type–I or type–II *sample* we plot

$$T_i^* = \frac{T_i^{**}}{T_k^*};\ i = 1,\,2,\,\ldots,\,k; \tag{10.17a}$$

against i/k with k as the total number of failed items observed and

$$T_i^{**} = \sum_{j=1}^{i}(k-j+1)\,(x_{(j)} - x_{(j-1)});\ i = 1,\,2,\,\ldots,\,k. \tag{10.17b}$$

2.3) For a *multiply censored* or *randomly censored sample* we take the KAPLAN/MEIER estimator of (10.16e) and plot

$$T_i^* = \frac{T_i^{***}}{T_k^{***}};\ i = 1,\,2,\,\ldots,,\,k; \tag{10.18a}$$

against $\widehat{F}_n(x_{(i)})$ for the uncensored observations $(x_{(i)},1)$. k is the total number of failed items observed and

$$T_i^{***} = \sum_{j=1}^{i}\frac{\widehat{S}_n(x_{(i)})+\widehat{S}_n(x_{(i-1)})}{2\,(x_{(i)}-x_{(i-1)})};\ i = 1,\,\ldots,\,k;\ x_{(0)} = 0;\ \widehat{S}_n(x_{(0)}) = 1. \tag{10.18b}$$

We mention that the TTT–graphs resulting from (10.17a,b) and (10.18a,b) would lie below those that would result had the sample been uncensored.

3) We mention another graphical approach suggested by EPSTEIN (1960). This approach needs a greater sample size. A property of the exponential distribution is that its conditional probability of failing in $(x,\,x+\Delta x)$, given survival up to x, is independent of x. This conditional probability is

$$\frac{f(x)\,\Delta x}{1-F(x)} = \frac{\frac{1}{b}\,\exp\big(-x/b\big)\,\Delta x}{\exp(-x/b)} = \frac{\Delta x}{b}. \tag{10.19}$$

So, if we start with a large number n of items, we divide the x–axis into intervals $(0,\Delta),\ (\Delta,2\,\Delta),\ (2\,\Delta,3\,\Delta),\ldots$ where Δ is suitably chosen, and if $n_1,\ n_2,\ n_3,\ldots$ are the numbers of items failing in these intervals, then

$$\frac{n_1}{n},\ \frac{n_2}{n-n_1},\ \frac{n_3}{n-n_1-n_2},\ \ldots$$

should fluctuate within reasonable limits about a constant, namely the hazard rate $1/b$.

Example 10/1: Checking for exponentiality

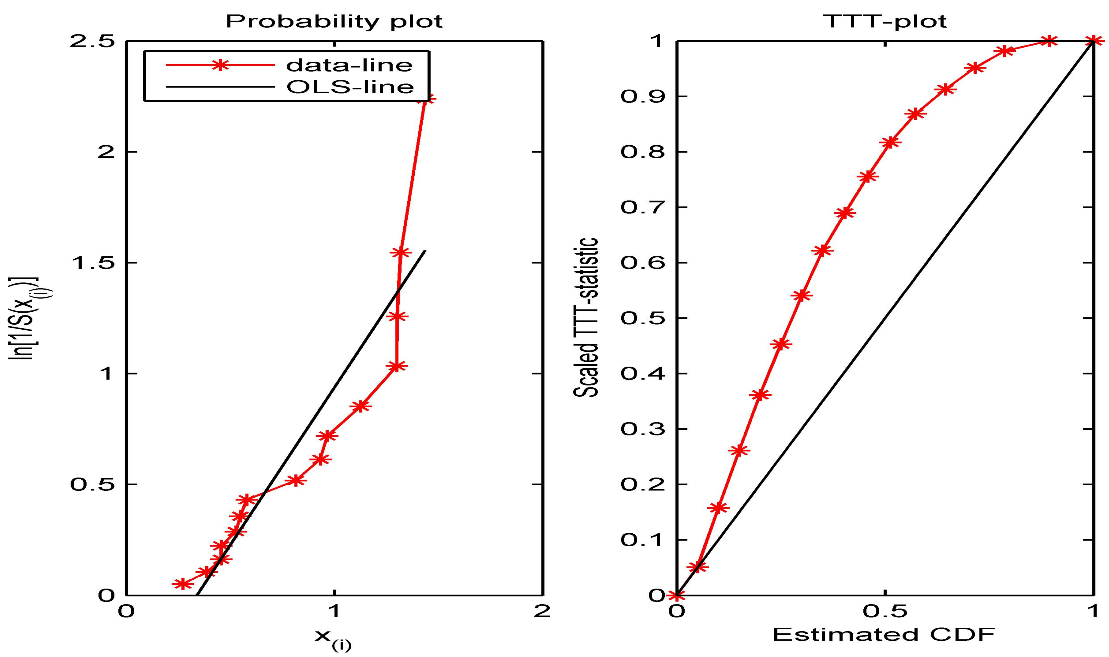
The following $n = 20$ observations are randomly censored. They have been generated from a WEIBULL distribution with scale parameter $b = 1$ and shape parameter $c = 2.5$, so the hazard rate is increasing more than linear.

$x_{(i)}$	0.2727	0.3877	0.4556	0.4565	0.5271	0.5487	0.5789	0.7846	0.8142	0.9329
δ_i	1	1	1	1	1	1	1	0	1	1

$x_{(i)}$	0.9659	1.1217	1.1267	1.2378	1.3001	1.3012	1.3181	1.3803	1.4362	1.9594
δ_i	1	0	1	0	1	1	1	0	1	1

Fig. 10/3 shows the probability plot on the left where the points significantly deviate from the straight OLS–fitted line. The same is true for the TTT–plot on the right where the points lie on a concave curve above the 45°–line what is typical for IHR distributions.

Figure 10/3: Graphs for judging exponentiality



10.2.2 Monotonicity of the Hazard Rate[6]

The topic of this section is to decide whether a distribution is IHR (DHR) or not. A lot of tests have been designed to test

H_0 : '$F(x) = 1 - \exp(-x/b), b$ unspecified.'
against
H_A : '$F(x)$ is IHR and not exponential.'
or
$\mathrm{H}_\mathrm{A}^\star$: '$F(x)$ is DHR and not exponential.'

[6] Suggested reading for this section: BARLOW (1968), BARLOW/PROSCHAN (1965, 1969), BICKEL (1964), BICKEL/DOKSUM (1969), DOKSUM/YANDELL, HALL/VAN KEILEGOM (2005), HOLLANDER/PROSCHAN (1984), KLEFSJÖ (1983a), PROSCHAN/PYKE (1967).

We will only present a few of these tests that can be found in the suggested reading for this section. We first mention a graphical approach that starts from (10.13c) in connection with (10.15):

$$\frac{d\,\phi_F(P)}{dP} = \frac{d}{dP}\left(\frac{1}{\mu}\,H_F^{-1}(P)\right) = \frac{1}{\mu}\,\frac{1}{h(x)}. \tag{10.20}$$

From (10.20) we see that the graph of the scaled TTT–transform $\phi_F(P)$ will be

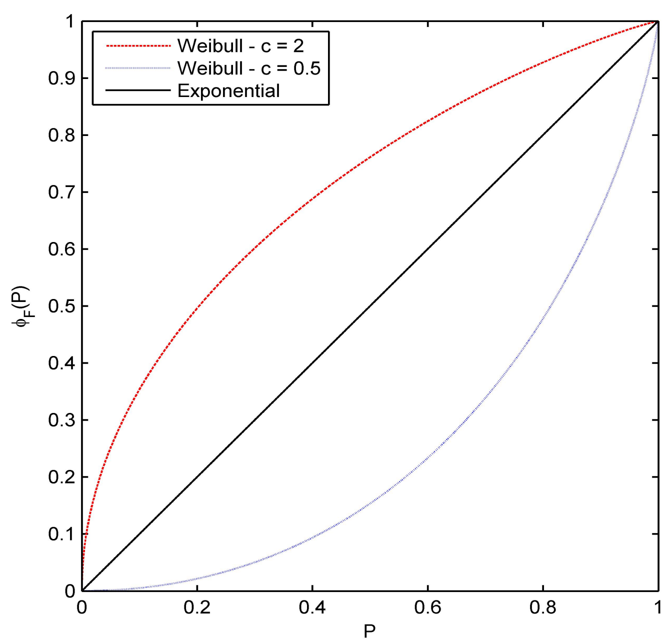
- a straight line for an exponential distribution,
- concave for an IHR distribution and
- convex for a DHR distribution.

As the graph of $\phi_F(P)$ runs from $(0,0)$ to $(1,1)$ it is clear that this graph will have

- a decreasing slope and lie above the $45°$–line for an IHR distribution,
- an increasing slope and lie below the $45°$–line for a DHR distribution,

see Fig. 10/4 which shows $\phi_F(P)$ for three WEIBULL distributions having $h(x) = c\,x^{c-1}$: $c = 0.5$ gives DHR, $c = 1$ gives the exponential distribution and $c = 2$ gives IHR. The graph of the scaled TTT–statistic T_i^* approaches that of $\phi_F(P)$ as $n \to \infty$ and we can reject H_0 in favor of H_A ($\mathrm{H}_\mathrm{A}^\star$) when the T_i^*–graph wholly lies above (below) the $45°$–line of the exponential distribution and is concave (convex). In case of an uncensored sample of size n the level of significance will be $\alpha = 1/n$.

Figure 10/4: Scaled TTT–transforms of three WEIBULL distributions



A numerical testing approach attached to the curvature of the TTT–transform goes back to KLFSJÖ (1983a). This tests needs an uncensored sample! Suppose that the $\phi_F(P)$–graph is concave (convex). Since the graph of the scaled TTT–statistic T_i^* converges to that of $\phi_F(P)$, it is reasonable to expect the TTT–plot based on a sample from an IHR (DHR) distribution to behave concavely (convexly), too, i.e.:

$$2\,T_i^* - T_{i-1}^* - T_{i+1}^* \underset{(<)}{>} 0;\; i = 1,\,\ldots,\,n-1;\; T_0^* = 0,\; T_n^* = 1. \tag{10.21a}$$

A possible test statistic against the IHR (DHR) alternative therefore is

$$A_1 = \sum_{i=1}^{n-1} \big(2\,T_i^* - T_{i-1}^* - T_{i+1}^*\big), \tag{10.21b}$$

and we expect a positive (negative) value of A_1 if $F(\cdot)$ is IHR (DHR), but not exponential. We immediately see that — using the normalized spacings D_i of (10.11) — A_1 can be written as

$$\begin{aligned} A_1 &= T_1^* + T_{n-1} - 1 \\ &= \frac{D_1 - D_n}{T_n}. \end{aligned} \tag{10.21c}$$

KLEFSJÖ gives the asymptotic distribution of A_1 under H_0 as a LAPLACE distribution and remarks that — because the numerator $(D_1 - D_n)$ of A_1 is independent of $D_2,\ D_3,\ \ldots,\ D_{n-1}$ — a test based on A_1 is not consistent against the whole IHR (DHR) class. For this reason KLEFSJÖ suggests a second test based on the idea that, when $\phi_F(P)$ is concave (convex) we would not only expect (10.21a) to hold, but we also expect, for $i = 1,\,2,\,\ldots,\,n-2$ and $k = 2,\,3,\,\ldots,\,n-j$, that

$$T_j^* + \frac{T_{j+k}^* - T_j^*}{k/n}\,\frac{\nu}{n} \underset{(<)}{>} T_{j+\nu}^* \ \text{ for } \ \nu = 1,\,2,\,\ldots,\,k-1 \tag{10.22a}$$

or

$$k\,\big(T_{j+\nu}^* - T_j^*\big) \underset{(>)}{<} \nu\,\big(T_{j+k}^* - T_j^*\big). \tag{10.22b}$$

From (10.22b) we can construct the test statistic

$$A_2 = \sum_{j=0}^{n-2}\sum_{k=2}^{n-j}\sum_{\nu=1}^{k-1} \Big[k\,\big(T_{j+\nu}^* - T_j^*\big) - \nu\,\big(T_{j+k}^* - T_j^*\big)\Big]. \tag{10.22c}$$

For $F(\cdot)$ to be IHR (DHR), but not exponential, we expect A_2 to be positive (negative). A_2 can be written more comfortably as

$$A_2 = \sum_{j=1}^{n} \alpha_j\,\frac{D_j}{T_n} \tag{10.22d}$$

with

$$\alpha_j = \frac{1}{6}\,\big[(n+1)^3\,j - 3\,(n+1)^2\,j^2 + 2\,(n+1)\,j^3\big]. \tag{10.22e}$$

KLEFSJÖ then considers the slightly modified test statistic

$$A = A_2\,\sqrt{\frac{7560}{n^7}} \tag{10.22f}$$

which is asymptotically $No(0,\,1)$. So, the asymptotic critical values for A are the percentiles τ_γ of $No(0,\,1)$:

	lower tail			upper tail		
γ	0.01	0.05	0.10	0.90	0.95	0.99
τ_γ	-2.3263	-1.6449	-1.2816	1.2816	1.6449	2.3263

H_0 is rejected in favor of H_A ($\mathrm{H}_\mathrm{A}^\star$) at level α when A is greater (smaller) than the critical value. Exact critical values have also been calculated by KLEFSJÖ, and an extract of his table is given in the following Tab. 10/1.

Table 10/1: Critical values of $A_2\sqrt{7560/n^7}$

	upper tail†		
n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
5	2.739	3.396	4.402
10	1.912	2.412	3.270
15	1.680	2.133	2.936
20	1.573	2.002	2.777
25	1.511	1.927	2.683
30	1.470	1.878	2.622
35	1.442	1.843	2.578
40	1.421	1.817	2.546
45	1.405	1.797	2.521
50	1.392	1.782	2.501
55	1.382	1.769	2.485
60	1.373	1.758	2.472
65	1.366	1.749	2.460
70	1.361	1.742	2.450
75	1.358	1.736	2.442
∞	1.282	1.645	2.326

Source: KLEFSJÖ (1983a, p. 922)

† For lower tail change the sign!

Many test of $\mathrm{H_0}$ versus $\mathrm{H_A}$ or $\mathrm{H_A^\star}$ are based on the normalized spacings $D_i = (n-i+1)\big(X_{(i)} - X_{(i-1)}\big)$ while other tests only utilize the ranks of the normalized spacings. One of the oldest tests based on normalized spacings is the **cttot–test** (cumulative total–time–on–test) **of EPSTEIN** (1960). see also BARLOW (1968). This test can be used for uncensored samples as well as for singly censored type–I and type–II sample. The test rests upon the fact that under $\mathrm{H_0}$ the total lifes (= successive TTT–statistics) $T_i = \sum_{j=1}^{i} D_j = \sum_{j=1}^{i}(n-j+1)\big(X_{(j)} - X_{(j-1)}\big)$ are uniformly distributed over $[0,\,T_r]$ where $1 \le r \le n$ is the number of failures in a sample of size n. The test statistic considered is

$$K_r = \frac{\sum\limits_{j=1}^{r-1}\sum\limits_{i=1}^{j} D_i}{\sum\limits_{i=1}^{r} D_i} = \frac{\sum\limits_{j=1}^{r-1}(r-j)\,D_j}{\sum\limits_{i=1}^{r} D_i}. \tag{10.23a}$$

BARLOW (1968) has given the exact percentage points $k_{r,\,\gamma}$ (critical values) for this test statistic, see Tab. 10/2. $\mathrm{H_0}$ is rejected in favor of $\mathrm{H_A}$ ($\mathrm{H_A^\star}$) at level α when

$$K_r \ge k_{r,\,1-\alpha} \quad (K_r \le k_{r,\,\alpha}). \tag{10.23b}$$

Even for small r we can use a normal approximation of K_r under $\mathrm{H_0}$. The approximate critical values are

$$k_{r,\,\gamma} \approx \frac{r-1}{2} + \tau_\gamma\sqrt{\frac{r-1}{12}}, \tag{10.23c}$$

with τ_γ as the γ–percentile of $No(0,\,1)$.

Table 10/2: Percentage points $k_{r,\gamma}$ of EPSTEINS's cttot–test

$r \backslash \gamma$	0.01	0.05	0.10	0.90	0.95	0.99
2	0.01	0.05	0.10	0.90	0.95	0.99
3	0.14	0.32	0.45	1.55	1.68	1.85
4	0.39	0.68	0.84	2.15	2.33	2.61
5	0.69	1.04	1.25	2.75	2.95	3.30
6	1.02	1.43	1.65	3.34	3.57	4.00
7	1.41	1.83	2.08	3.90	4.15	4.60
8	1.77	2.24	2.52	4.49	4.75	5.24
9	2.12	2.65	2.94	5.06	5.35	5.88
10	2.52	3.06	3.38	5.62	5.92	6.45

Source: BARLOW (1968, p. 558)

BICKEL/DOKSUM (1969) extensively studied tests of H_0 versus H_A ($\mathrm{H}_\mathrm{A}^\star$) based on the ranks of the normalized spacings. Their test statistics (see below) are partially motivated by the **test of PROSCHAN/PYKE** (1967) which is as follows. Let

$$V_{ij} = \left\{ \begin{array}{ll} 1 & \text{if } D_i \geq D_j \text{ for } i, j = 1, 2, \ldots, n \\ 0 & \text{otherwise.} \end{array} \right\} \tag{10.24a}$$

So, this test requires uncensored samples. The test statistic is

$$V_n = \sum_{\substack{i,j=1 \\ i<j}}^{n} V_{ij}. \tag{10.24b}$$

H_0 is rejected in favor of H_A ($\mathrm{H}_\mathrm{A}^\star$) at level α if

$$V_n \geq v_{n,1-\alpha} \quad (V_n \leq v_{n,\alpha}) \tag{10.24c}$$

where the critical value $v_{n,\gamma}$ is determined so that

$$\Pr(V_n < v_{n,\gamma} \,|\, \mathrm{H}_0) = \gamma. \tag{10.24d}$$

The exact value of $v_{n,\gamma}$ is calculated from

$$\Pr(V_n = k \,|\, \mathrm{H}_0) = \frac{P_n(k)}{n!}, \tag{10.24e}$$

where $P_n(k)$ is the number of orderings of $D_1, D_2, \ldots, D_n$ with exactly k inversions of indices. An inversion of indices $i < j$ occurs when $D_i > D_j$. V_n is asymptotic normal with

$$\mathrm{E}(V_n) = \frac{n\,(n-1)}{4} \quad \text{and} \quad \mathrm{Var}(V_n) = \frac{(2\,n+5)\,(n-1)\,n}{72}. \tag{10.24f}$$

So, we have

$$v_{n,\gamma} \approx \frac{n\,(n-1)}{4} + \tau_\gamma \sqrt{\frac{(2\,n+5)\,(n-1)\,n}{72}}. \tag{10.24g}$$

This test is justified as follows: Under H_0 the normalized spacings $D_1, D_2, \ldots, D_n$ are iid, each with PDF $\exp(-x/b)\big/b$, so that $\Pr(V_{ij} = 1) = 0.5$ for $i, j = 1, 2, \ldots, n;\ i \neq j$. However, under H_A we have $\Pr(V_{ij} = 1) > 0.5$ for $i, j = 1, 2, \ldots, n;\ i < j$. Thus, each V_{ij} and consequently

V_n, tend to be large under $\mathrm{H_A}$, so that rejection of $\mathrm{H_0}$ in favor of $\mathrm{H_A}$ occurs for large values of V_n. We finally mention that the asymptotic relative efficiency of the cttot–test is higher than that of the PROSCHAN/PYKE–test.

Let R_i be the rank of D_i. Based on these ranks BICKEL/DOKSUM (1969) suggested a great number of test statistics, e.g.:

$$\begin{aligned}
W_0 &= \sum_{i=1}^{n} \frac{i}{n+1}\,\frac{R_i}{n+1},\\
W_1 &= \sum_{i=1}^{n} -\frac{i}{n+1}\,\ln\Big(1-\frac{R_i}{n+1}\Big),\\
W_2 &= \sum_{i=1}^{n} \ln\Big(1-\frac{i}{n+1}\Big)\,\ln\Big(1-\frac{R_i}{n+1}\Big),\\
W_3 &= \sum_{i=1}^{n} \ln\Big[-\ln\Big(1-\frac{i}{n+1}\Big)\Big]\,\ln\Big(1-\frac{R_i}{n+1)}\Big).
\end{aligned}$$

Large (small) values of these test statistics are significant for $\mathrm{H_A}$ ($\mathrm{H_A^\star}$).

The tests of cttot–test and the tests of KELFSJÖ and PROSCHAN/PYKE have been implemented in the accompanying MATLAB–program *HAZARD_10*.

Example 10/2: Testing for IHR and DHR

A first uncensored data set of $n = r = 15$ from a WEIBUILL distribution with scale parameter $b = 10$ and shape parameter $c = 0.7$ is:

0.0309	0.3641	0.5317	0.9545	1.0119
1.9145	3.5331	3.9321	4.1219	10.9776
13.5405	14.9801	15.2600	15.8278	31.4019

So, this sample comes from a DHR–distribution. The cttot–test gives $K_r = 5.58$, so that $\mathrm{H_0}$ is rejected in favor of $\mathrm{H_A^\star}$ (DHR), the level of significance is about 0.10. The KLEFSJÖ–test gives $A = -1.766$, so that $\mathrm{H_0}$ is rejected in favor of $\mathrm{H_A^\star}$ with a level of significance of about 0.10. The PROSCHAN/PYKE–test gives $V_n = 41$, so that $\mathrm{H_0}$ is rejected in favor of $\mathrm{H_A^\star}$ with a level of significance of approximately 0.12. A second uncensored sample of $n = r = 20$ from a WEIBULL distribution with $b = 20,\ c = 2$ is:

2.5349	2.6149	4.4532	4.8567	5.7627
11.0273	15.5141	16.8996	18.3318	18.5556
18.6857	19.8772	20.7875	22.1906	22.9138
25.4471	26.2949	29.9485	34.3137	40.8301

This sample comes from an IHR–distribution. The cttot–test gives $K_r = 12.57$ and $\mathrm{H_0}$ ist rejected in favor of $\mathrm{H_A}$ (IHR) with $\alpha < 0.01$. The KLEFJÖ–test with $A = 3.260$ rejects $\mathrm{H_0}$ in favor of $\mathrm{H_A}$ with $\alpha \ll 0.01$. The PROSCHAN/PYKE–test with $V_n = 122$ rejects $\mathrm{H_0}$ in favor of $\mathrm{H_A}$ with $\alpha \approx 0.04$.

10.2.3 Bathtub Shape of the Hazard Rate[7]

There are two non–monotone courses of a hazard rate which are of special interest:

- the bathtub shape (DIHR = decreasing–increasing hazard rate) where the hazard rate initially is decreasing during the so–called 'infant mortality' phase, then constant during the 'useful life' phase, and finally increasing during the 'wear–out' phase, and

[7] Suggested reading for the section: AARSET (1985, 1987), BERGMAN (1979), KUNITZ (1989).

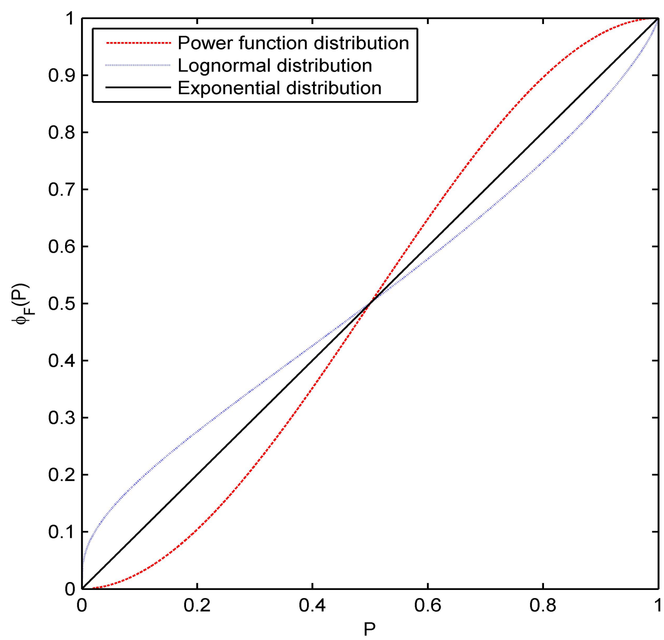
- the inverted bathtub shape (IDHR = increasing–decreasing hazard rate) where the three phases mentioned above are changed in order.

From the behavior of the scaled TTT–transform $\phi_F(P)$ of $F(\cdot)$ with respect to the hazard rate as given in (10.20) we expect for the DIHR case that

- $\phi_F(P)$ is convex and lies below the $45°$–line for P being small, i.e., in the leftmost part of the plot, and
- $\phi_F(P)$ is concave and lies above the $45°$–line for P being large, i.e., in the rightmost part of the plot.

In the IDHR case the order of these phases of $\phi_F(P)$ is reverted. Fig. 10/5 shows the scaled TTT–transforms of a lognormal distribution, which is an IDHR distribution, and of a power function distribution, which is DIHR.

Figure 10/5: Scaled TTT–transforms of lognormal and power function distributions



BERGMAN (1979) suggests the following procedure for testing

H_0 : '$F(\cdot)$ is the exponential distribution.'
against
H_A : '$F(\cdot)$ is DIHR (bathtub–shaped).'

based on the TTT-plot with i/n on the abscissa and T_i^* on the ordinate. This test asks for uncensored samples! We introduce

$$V_n^* = \left\{ \begin{array}{l} \min\left\{ i \geq 1 \,:\, T_i^* \geq \dfrac{i}{n} \right\} \\ n \;\text{ if }\; T_i^* < \dfrac{i}{n} \;\text{ for }\; i = 1, 2, \ldots, n-1; \end{array} \right\} \tag{10.25a}$$

$$K_n^* = \left\{ \begin{array}{l} \max\left\{ i \leq n-1 \,:\, T_i^* \leq \dfrac{i}{n} \right\} \\ 0 \;\text{ if }\; T_i^* > \dfrac{i}{n} \;\text{ for }\; i = 1, 2, \ldots, n-1; \end{array} \right\} \tag{10.25b}$$

$$G_n^* = V_n^* + n - K_n^*, \tag{10.25c}$$

and reject H_0 in favor of H_A when G_n^* is large. The motivation for this test is that when the distribution is DIHR, then we may expect V_n^* as well as $(n - K_n^*)$ to be large. G_n^* obviously takes on integer values in $[2, n+1]$.

The graph of $\phi_F(P)$ for an IDHR distribution (inverted bathtub shape) behaves like the reflection of $\phi_F(P)$ for a DIHR distribution, see Fig. 10/5, the line of reflection being the $45°$–line of the exponential distribution. Thus, we can modify (10.25a-c) to test for IDHR in the following way:

$$V_n^{**} = \left\{ \begin{array}{l} \min\left\{ i \geq 1 \,:\, T_i^* \leq \dfrac{i}{n} \right\} \\ n \text{ if } T_i^* > \dfrac{i}{n} \text{ for } i = 1,\, 2,\, \ldots,\, n-1; \end{array} \right\} \tag{10.26a}$$

$$K_n^{**} = \left\{ \begin{array}{l} \max\left\{ i \leq n-1 \,:\, T_i^* \geq \dfrac{i}{n} \right\} \\ 0 \text{ if } T_i^* < \dfrac{i}{n} \text{ for } i = 1,\, 2,\, \ldots,\, n-1; \end{array} \right\} \tag{10.26b}$$

$$G_n^{**} = V_n^{**} + n - K_n^{**}. \tag{10.26c}$$

H_0 is rejected in favor of $H_A^\star$: '$F(\cdot)$ is IDHR' when G_n^{**} is large.

AARSET (1985) has derived the following distribution of G_n^* under exponentiality, the so–called null distribution:

$$\Pr(G_n^* = i) = \left\{ \begin{array}{ll} \displaystyle\sum_{\ell=1}^{i-1} \frac{(n-1)!}{(\ell-1)!(n-i+1)!(i-\ell-1)!} \left(\frac{\ell}{n}\right)^{\ell-1} \times & \\ \quad \times \left(\dfrac{n-i}{n}\right)^{n-i+1} \left(\dfrac{i-\ell}{n}\right)^{i-\ell-1} \dfrac{1}{\ell}\, \dfrac{1}{i-\ell} & \text{for } i = 2,\, \ldots,\, n-1 \\ 0 & \text{for } i = n \\ \dfrac{2}{n} \left(\dfrac{n+1}{n}\right)^{n-2} & \text{for } i = n+1. \end{array} \right\} \tag{10.27a}$$

He also tabulated the CCDF of G_n^*, see Tab. 10/3, which also allows to give the level of significance of this test. We see from Tab. 10/3 that for smaller sample sizes we cannot achieve a small probability of the first–kind–error. An asymptotic result for $\Pr(G_n^* = n - k)$ is also given by AARSET:

$$\lim_{n\to\infty} n \Pr(G_n^* = n - k) = 2\, \frac{k^{k+1}}{(k+1)!} \exp(-k). \tag{10.27b}$$

Table 10/3: $\Pr(G_n^* \geq n - k)$

k \ n	10	50	75	100	125	150	175	200	250
-1	0.42872	0.10348	0.07013	0.05303	0.04263	0.03565	0.03063	0.02685	0.02153
1	0.46698	0.11091	0.07506	0.05673	0.04559	0.03811	0.03273	0.02869	0.02300
2	0.51102	0.11843	0.08001	0.06041	0.04852	0.04055	0.03482	0.03051	0.02446
3	0.56003	0.12564	0.08470	0.06389	0.05129	0.04284	0.03678	0.03222	0.02582
4	0.61577	0.13257	0.08917	0.06718	0.05389	0.04499	0.03862	0.03383	0.02710
5	0.68139	0.13927	0.09343	0.07030	0.05636	0.04703	0.04035	0.03534	0.02830
6	0.76202	0.14579	0.09753	0.07329	0.05871	0.04897	0.04200	0.03677	0.02944
7	0.86578	0.15217	0.10149	0.07617	0.06097	0.05082	0.04358	0.03814	0.03052
8	1.00000	0.15846	0.10535	0.07895	0.06314	0.05261	0.04509	0.03945	0.03156

Source: AARSET (1985, p. 59)

AARSET (1987) proposed another test — for uncensored samples — of H_0 against H_A which is an adaption to the well–known CRAMÉR–VON–MISES goodness–of–fit test. The latter test has the test statistic

$$W_n^2 = n \int\limits_{-\infty}^{\infty} \left[F_n(x) - F(x)\right]^2 dF(x), \tag{10.28}$$

i.e., it rests upon the squared distances between the empirical CDF $F_n(x)$ and the hypothetical $F(x)$. AARSET now suggests the test statistic

$$R_n = \int\limits_0^1 \Delta_n^2(u)\, du \tag{10.29a}$$

where

$$\Delta_n(u) = \left\{ \begin{array}{cl} \sqrt{n}\left[T_i^* - \phi_F(u)\right] & \text{for } \dfrac{i-1}{n} < u \le \dfrac{i}{n},\ 1 \le i \le n \\ 0 & \text{for } u = 0. \end{array} \right\} \tag{10.29b}$$

R_n rests upon the squared distances between the scaled TTT–statistics T_i^* and the TTT–transform of $F(\cdot)$. Under H_0 $F(\cdot)$ is given by the exponential distribution and the test statistic turns into

$$R_n = \sum_{i=1}^{n} T_i^* \left(T_i^* - \frac{2\,i-1}{n}\right) + \frac{n}{3}. \tag{10.29c}$$

According to the invariance principle R_n has the same asymptotic distribution as the CRAMÉR–VON–MISES statistic W_n^2 with the following percentage points:

$\lim_{n\to\infty}$ Pr$(W_n^2 \le w)$	0.90	0.95	0.99	0.999
w	0.34730	0.46136	0.74346	1.16786

Source: ANDERSON/DARLING (1952, p. 203)

The reason for a large value of R_n is any great discrepancy between T_i^* and $\phi_F(P)$ of the exponential distribution. Thus, we cannot rely on only R_n to decide for DIHR (IDHR). We have to take into account other evidences like the TTT–plot and the statistics G_n^* or G_n^{**}.

The tests of BERGMAN and AARSET are implemented in the accompanying MATLAB–program *HAZARD_11*.

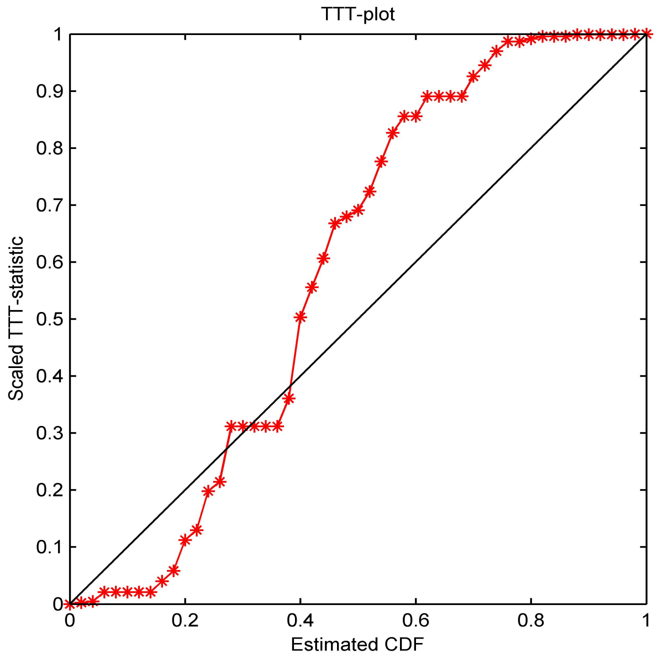
Example 10/3: Testing for DIHR (bathtub–shaped hazard rate)

We want to know whether the following data set $(n = 50)$ comes from a DIHR distribution or not:

0.2	0.4	2	2	2	2	2	4	6	12
14	22	24	36	36	36	36	36	43	64
72	80	90	92	94	100	110	120	126	126
134	134	134	134	144	150	158	164	164	166
168	168	168	170	170	170	170	170	172	172

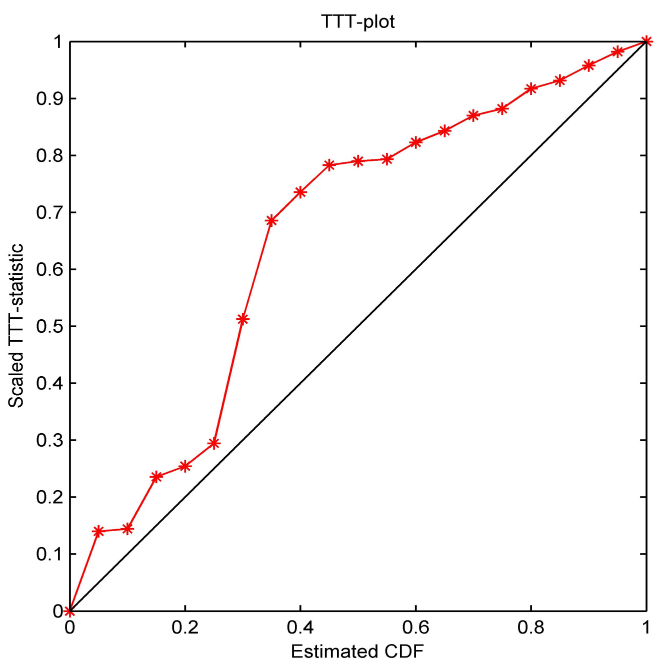
Fig. 10/6 clearly indicates DIHR. The BERGMAN–test gives $G_{50}^* = 45$. As — according to Tab. 10/3 — Pr$(G_{50}^* \ge 45) = 0.13927$ this test has a level of significance for rejecting exponentiality against DIHR. The AARSET–test gives $R_{50} = 1.2922$ and we can reject exponentiality in favor of DIHR at $\alpha \ll 0.001$.

Figure 10/6: TTT–plot for a data set coming from a DIHR distribution



When we submit the second data set $(n = 20)$ of Example 10/2 to the BERGMAN–test and to the AARSET–test we find $G^*_{20} = 21$, which is insignificant for DIHR as well as for IDHR, and $R_{20} = 0..8750$. The latter statistic alone would be significant $(\alpha < 0.01)$ for deviation from exponentiality, but the TTT–plot in Fig. 10/7 clearly indicates IHR and neither DIHR nor IDHR.

Figure 10/7: TTT–plot of a data set coming from an IHR distribution



When there is evidence for an IDHR or a DIHR distribution we often want to known that special lifetime where the hazard rate changes from increasing to decreasing or vice versa. This special lifetime is called **change point**.[8] There exists an extensive literature on turning points of the hazard rate. Most papers deal with the estimation of the change point, see ANTONIADIS et al. (2000), BEBBINGTON et al. (2008), JOSHI/MCEACHERN (1997), LAI et al. (2001), LOADER (1991), MÜLLER/WANG (1990a), NGUYEN et al. (1984) or PATRA/DEY (2002). The paper of HENDERSON (1990) tests for the existence of a change point.

10.3 Testing for Aging Classes[9]

In Chapter 2 we have introduced several classes of aging. A detailed description of the classes IHRA, NBU, NBUE, DMRL and HNBUE along with their duals DHRA, NWU, NWUE, IMRL and HNWUE is given in Sect. 2.4. Here, we look for testing procedures to decide whether a sample comes from a distribution belonging to one of these classes. Most of these tests rest upon the relationships between the TTT–transform and the aging properties as described in KLEFSJÖ (1982b). The exponential distribution is the border case of each of these classes and the test is of H_0 : '$F(\cdot)$ is exponential' against H_A : '$F(\cdot)$ is member of this class without the exponential distribution'.

10.3.1 IHRA (DHRA) Tests

A distribution $F(\cdot)$ is IHRA (DHRA) — increasing (decreasing) hazard rate average — if

$$\frac{H(x)}{x} = \frac{1}{x}\int\limits_0^x h(u)\,du$$

is increasing (decreasing). BARLOW/CAMPO (1975) and BARLOW (1979) give the following *theorem* relating IHRA (DHRA) to the TTT–transform $\phi_F(P)$:

> 'If $F(\cdot)$ is a life distribution which is IHRA (DHRA) then $\phi_F(P)\big/P$ is decreasing (increasing) for $0 < P < 1$.'

Thus, since $\phi_F(P)\big/P$ being decreasing is a necessary (but not sufficient) condition for $F(\cdot)$ to be IHRA KLEFSJÖ (1983a) proposes a statistic which investigates whether the analogous property holds for the TTT–plot. If $\phi_F(P)\big/P$ is decreasing we expect the corresponding to hold for the TTT–plot. This means that

$$\frac{T_i^*}{i/n} > \frac{T_j^*}{j/n} \ \text{ for } j > i,\ i = 1, 2, \ldots, n-1. \tag{10.30a}$$

Multiplication by $(i\,j)\big/n$ and summing over i and j gives the test statistic

$$B = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Big(j\,T_i^* - i\,T_j^*\Big). \tag{10.30b}$$

Significantly large (small) values of B lead to the rejection of

H_0 : '$F(\cdot)$ is exponential.'
in favor of
H_A : '$F(\cdot)$ is IHRA, but not exponential.'
(in favor of
$\mathrm{H}_\mathrm{A}^\star$: '$F(\cdot)$ is DHRA, but not exponential'.)

---

[8] In statistics the notion 'change point', generally has another meaning: it is that realization of a variate where there is a change of the distribution function, see CSÖRGÖ/HORVÁTH (1968) or KRISHNAIAH/MIAO (1988).

[9] Suggested reading for this section: HOLLANDER/PROSCHAN (1984), KLEFSJÖ (1982b, 1983a).

The test statistic B can be simplified by using the normalized spacings $D_i = (n-i+1)\,(X_{(i)} - X_{(i-1)})$ and the total–time–on–test statistic $T_n = \sum_{i=1}^{n} D_i$:

$$B = \frac{1}{T_n}\sum_{j=1}^{n} b_j\, D_j \tag{10.30c}$$

where

$$b_j = \frac{1}{6}\left[2\,j^3 - 3\,j^2 + j\,(1-3\,n-3\,n^2) + 2\,n + 3\,n^2 + n^3\right]. \tag{10.30d}$$

KLEFSJÖ (1983a) has found the exact null distribution of the slightly modified test statistic

$$B^* = B\sqrt{\frac{210}{n^5}}. \tag{10.30e}$$

The upper $0.01,\ 0.05,\ 0.10$ percentiles are in Tab. 10/4. The asymptotic distribution of B^* under H_0 is $No(0,\,1)$. Example 10/4 further down shows the working of this test which has been implemented in the accompanying MATLAB–program *HAZARD_12*.

Table 10/4: Critical values of $B^* = B\sqrt{210/n^5}$

n	upper tail†		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
5	1.703	2.257	3.227
10	1.508	2.003	2.951
15	1.441	1.909	2.815
20	1.406	1.858	2.736
25	1.385	1.827	2.684
30	1.371	1.804	2.646
35	1.360	1.788	2.618
40	1.352	1.755	2.595
45	1.346	1.765	2.577
50	1.341	1.757	2.562
55	1.337	1.750	2.550
60	1.333	1.744	2.538
65	1.330	1.739	2.528
70	1.328	1.734	2.520
75	1.325	1.730	2.512
∞	1.282	1.645	2.326

Source: KLEFSJÖ (1983a, p. 923)

† For lower tail change the sign!

There are other tests for IHRA (DHRA). BARLOW (1968) derives a likelihood ratio test statistic, lower percentiles of which are intended for testing IHRA against DHRA. He also suggests to take the cttot–test of EPSTEIN (1960) — see (10.23a-c) — for testing H_0 versus H_A or $\mathrm{H}_\mathrm{A}^\star$. BARLOW/CAMPO (1975) proposed to take the number of crossings between the the TTT–graph and the $45°$–line as a test statistic and to reject H_A when this number is small. BERGMAN (1977) gives the exact and the asymptotic distribution of this test statistic under H_0 together with a table of its CCDF.

10.3.2 NBU (NWU) Tests

A life distribution $F(\cdot)$ is NBU (new better than used) if

$$S(x+y) \le S(x)\,S(y)\;\forall\; x,\,y \ge 0,$$

and NWU (new worse than used) if the reversed inequality holds. The following NBU/NWU test of HOLLANDER/PROSCHAN (1972) is motivated by considering the parameter

$$\begin{aligned}\gamma &= \int\limits_0^\infty\int\limits_0^\infty \big[S(x)\,S(y) - S(x+y)\big]\,dF(x)\,dF(y)\\ &= \frac{1}{4} - \int\limits_0^\infty\int\limits_0^\infty S(x+y)\,dF(y)\\ &= \frac{1}{4} - \Delta(F) \end{aligned}\tag{10.31a}$$

with

$$\Delta(F) = \int\limits_0^\infty\int\limits_0^\infty S(x+y)\,dF(y) = \Pr(X_1 > X_2 + X_3) \tag{10.31b}$$

where $X_1,\,X_2,\,X_3$ are iid according to $F(\cdot)$. γ is a measure of deviation of $F(\cdot)$ from exponentiality towards NBU (NWU) alternatives. $\Delta(F) = 1\big/4$ when $F(\cdot)$ is exponential.

The classical non–parametric approach of replacing $F(\cdot)$ by the empirical CDF $F_n(\cdot)$ suggests rejecting

H$_0$: '$F(\cdot)$ is exponential.'
in favor of
H$_\text{A}$: '$F(\cdot)$ is NBU, but not exponential.'
(in favor of
H$^\star_\text{A}$: '$F(\cdot)$ is NWU, but not exponential.')

if $\int\int\big[1 - F_n(x+y)\big]\,dF_n(x)\,dF_n(y)$ is too small (large). HOLLANDER/PROSCHAN (1972) found it more convenient to reject H$_0$ for small (large) values of the asymptotically equivalent statistic

$$J = \frac{2}{n\,(n-1)\,(n-2)}\sum\nolimits^*\Psi\big(X_{a_1}, X_{a_2} + X_{a_3}\big) \tag{10.31c}$$

where

$$\Psi(a,\,b) = \left\{\begin{array}{ll} 1 & \text{for } a > b\\ 0 & \text{for } a \le b\end{array}\right\} \tag{10.31d}$$

and $\sum^*$ as the sum over all $n\,(n-1)\,(n-2)$ triples $(a_1,\,a_2,\,a_3)$ of three integers such that $1 \le a_1 \le n,\; a_1 \ne a_2,\; a_2 \ne a_3$ and $a_2 < a_3$. Defining

$$M_n = \frac{n\,(n-1)\,(n-2)}{2}\,J = \sum\nolimits^*\Psi\big(X_{a_1}, X_{a_2} + X_{a_3}\big) \tag{10.31e}$$

and denoting $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ as the ordered X's and since $i \le \max(j,\,k)$ implies $\Psi\big(X_{(i)}, X_{(j)} + X_{(k)}\big) = 0$ we can rewrite M_n as

$$M_n = \sum_{i>j>k}\Psi\big(X_{(i)}, X_{(j)} + X_{(k)}\big) \tag{10.31f}$$

with $M_n = 0, 1, 2, \ldots, n\,(n-1)\,(n-2)/6$. The following Tab. 10/5, based on the critical values of M_n to be found as table 4.1 in HOLLANDER/PROSCHAN (1972) — gives critical values $j_{n,\gamma}$ of the test statistic J in (10.31c). H_0 has to be rejected in favor of H_A ($\mathrm{H}_\mathrm{A}^\star$) at level $\alpha = 0.05$ or 0.10 if $J \le j_{n,\alpha}$ ($J \ge j_{n,1-\alpha}$). The normal approximation treats

$$J^* = \left(J - \frac{1}{4}\right)\sqrt{\frac{432\,n}{5}} \tag{10.31g}$$

as $No(0,\,1)$–distributed.

Table 10/5: Critical values $j_{n,\gamma}$ of J

$n \backslash \gamma$	0.05	0.10	0.90	0.95
8	0.1488	0.1786	0.3095	0.3155
9	0.1587	0.1865	0.3016	0.3095
10	0.1667	0.1917	0.2972	0.3056
11	0.1737	0.1960	0.2949	0.3030
12	0.1772	0.1984	0.2924	0.3000
13	0.1830	0.2028	0.2902	0.2972
14	0.1864	0.2042	0.2885	0.2949
15	0.1897	0.2066	0.2872	0.2938
16	0.1923	0.2083	0.2857	0.2923
17	0.1946	0.2103	0.2843	0.2907
18	0.1961	0.2116	0.2831	0.2896
19	0.1985	0.2129	0.2820	0.2883
20	0.2003	0.2140	0.2810	0.2871
25	0.2068	0.2184	0.2780	0.2835
30	0.2113	0.2220	0.2750	0.2800
35	0.2147	0.2238	0.2729	0.2783
40	0.2171	0.2260	0.2716	0.2763
45	0.2194	0.2275	0.2702	0.2751
50	0.2214	0.2286	0.2691	0.2736

An application of this test — worked out with the MATLAB–program *HAZARD_12* — is found further down in Example 10/4.

10.3.3 DMRL (IMRL) Tests

A continuous life distribution $F(\cdot)$ with $S(\cdot) = 1 - F(\cdot)$ is DMRL (IMRL) if its mean residual life function

$$\mu(x) = \frac{1}{S(x)}\int_x^\infty S(u)\,du$$

is decreasing (increasing). KLEFSJÖ (1982b) proved the following *theorem* connecting this aging property with the TTT–transform $\phi_F(P)$:

> 'A life distribution is DMRL (IMRL) if and only if
>
> $$Q(P) = \frac{1-\phi_F(P)}{1-P}$$
>
> is decreasing (increasing) for $0 \le P < 1$.'

Based on this theorem and on the same idea as in Sect. 10.3.1 we expect that — if $F(\cdot)$ is DMRL (IMRL) — the following holds:

$$\frac{1-T_j^*}{1-j/n} \underset{(>)}{<} \frac{1-T_i^*}{1-i/n} \quad \text{for } j>i \text{ and } i=0,\,1,\,2,\,\ldots,\,n-1. \tag{10.32a}$$

After multiplication by $(n-i)\,(n-j)/n$ and summation we get the test statistic

$$K=\sum_{i=0}^{n-1}\sum_{j=i+1}^{n}\big[(n-j)\,(1-T_i^*)-(n-i)\,(1-T_j^*)\big]. \tag{10.32b}$$

If $F(\cdot)$ is DMRL (IMRL), but not exponential, we expect K to be positive (negative). It is to be noted that the test statistic K is proportional to the test statistic proposed by HOLLANDER/PROSCHEN (1975) for the same test. Their test statistic is

$$W_n^*=\frac{1}{\overline{X}}\left[\frac{1}{n^4}\sum_{i=1}^{n}c_i\,X_{(i)}\right] \tag{10.33a}$$

where $\overline{X}=\sum X_i/n$ and

$$c_i=(4/3)\,i^3-4\,n\,i^2+3\,n^2\,i-0.5\,n^3+0.5\,n^2-0.5\,i^2+i/6. \tag{10.33b}$$

Significantly large (small) values of W_n^* suggest DMRL (IMRL) alternatives to exponentiality. Under H_0 : '$F(\cdot)$ is exponential' the statistic

$$W=W_n^*\sqrt{210\,n}$$

is asymptotically $No(0,\,1)$. For smaller sample sizes HOLLANDER/PROSCHAN (1975) give upper and lower critical values found by Monte Carlo simulation, see Tab. 10/6. The simulation evidently accounts for the slight disturbances in the monotonicity of the tabulated percentiles.

Table 10/6: Critical values $w_{n,\gamma}$ for W

$n \backslash \gamma$	0.01	0.05	0.10	0.90	0.95	0.99
8	−3.029	−2.095	−1.565	1.415	1.703	2.162
9	−2.956	−2.017	−1.532	1.385	1.659	2.131
10	−2.946	−2.038	−1.511	1.365	1.651	2.155
11	−2.887	−1.983	−1.506	1.355	1.657	2.125
12	−2.838	−1.959	−1.496	1.339	1.642	2.145
13	−2.828	−1.950	−1.482	1.349	1.639	2.143
14	−2.826	−1.918	−1.474	1.347	1.641	2.145
15	−2.822	−1.910	−1.466	1.317	1.615	2.143
16	−2.736	−1.900	−1.458	1.323	1.630	2.131
17	−2.788	−1.905	−1.447	1.314	1.625	2.153
18	−2.752	−1.881	−1.444	1.303	1.623	2.127
19	−2.669	−1.889	−1.429	1.309	1.620	2.141
20	−2.730	−1.865	−1.434	1.299	1.617	2.137
25	−2.694	−1.834	−1.398	1.315	1.627	2.141
30	−2.678	−1.805	−1.382	1.288	1.609	2.129
35	−2.570	−1.767	−1.365	1.286	1.606	2.153
40	−2.606	−1.775	−1.358	1.282	1.605	2.171
45	−2.564	−1.745	−1.361	1.273	1.606	2.154
50	−2.529	−1.762	−1.351	1.293	1.609	2.159
∞	−2.326	−1.645	−1.282	1.282	1.645	2.325

Source: HOLLANDER/PROSCHAN (1975, p. 589)

An application of this test — worked out with the MATLAB–program *HAZARD_12* — is found further down in Example 10/4.

10.3.4 NBUE (NWUE) Tests

A continuous life distribution $F(\cdot)$ with $S(\cdot) = 1 - F(\cdot)$ is called NBUE (new better than used in expectation) [NWUE (new worse than used in expectation)] if

$$\mu(0) \geq \mu(x) \quad [\mu(0) \leq \mu(x)],\ \forall\, x > 0,$$

or equivalently

$$\int_x^\infty S(u)\,du < \mu(0)\,S(x), \quad \Big[\int_x^\infty S(u)\,du > \mu(0)S(x)\Big],\ \forall\, x > 0.$$

The following *theorem* of KLEFSJÖ (1983a) relates these properties to the TTT–transform $\phi_F(P)$:

> 'A life distribution $F(\cdot)$ is NBUE (NWUE) if and only if $\phi_F(P) \underset{(\leq)}{\geq} P$ for $0 \leq P \leq 1$.'

This theorem leads to the following test statistic based on the differences between the scaled TTT–statistics $T_j^* = T_j/T_n$ and the empirical CDF j/n :

$$\begin{aligned} C &= \sum_{j=1}^{n}\left(T_j^* - \frac{j}{n}\right) \\ &= \sum_{j=1}^{n-1} T_j^* - \frac{n-1}{2}, \text{ as } T_n^* = 1. \end{aligned} \tag{10.34a}$$

If $F(\cdot)$ is NBUE (NWUE), but not exponential, C is expected to be positive (negative). We remark that

$$V = \sum_{j=1}^{n-1} T_j^* = \frac{\sum\limits_{j=1}^{n-1} T_j}{T_n} = \frac{\sum\limits_{j=1}^{n-1} T_j}{\sum\limits_{j=1}^{n} D_j} \tag{10.34b}$$

is the cumulative TTT–statistic used in (10.23a) for IHR(DHR)–testing. Thus, we have

$$C = V - \frac{n-1}{2}. \tag{10.34c}$$

HOLLANDER/PROSCHAN (1975) proposed another test statistic for testing

H_0 : '$F(\cdot)$ is exponential.'
in favor of
H_A : '$F(\cdot)$ is NBUE, but not exponential.'
(in favor of
$H_A^\star$: '$F(\cdot)$ is NWUE, but not exponential.')

Their statistic rests upon the weighted difference between $\mu(0)$ and $\mu(x)$ and reads

$$K^* = \frac{\sum\limits_{j=1}^{n}\left(\frac{3\,n+1}{2} - 2\,j\right) X_{(j)}}{n\sum\limits_{j=1}^{n} D_j}. \tag{10.34d}$$

We note that

$$C = V - \frac{n-1}{2} = n\,K^*. \qquad (10.34e)$$

Hence, $C,\ V$ and K^* are equivalent test statistics which can be traced back to the test statistic

$$K_n = \frac{\sum\limits_{j=1}^{n-1} T_j}{T_n}$$

of (10.23a). We have

$$K_n = n\,K^* + \frac{n-1}{2}. \qquad (10.34f)$$

Significantly large (small) values of K^* suggest NBUE (NWUE). We do not need to furnish critical values of K^* because — based on (10.34f) — we can use the percentage points $k_{n,\gamma}$ of Tab. 10/2 in the following way: "Reject $\mathrm{H_0}$ in favor of $\mathrm{H_A}$ ($\mathrm{H_A^\star}$) if

$$K^* \geq \frac{1}{n}\,k_{n,\,1-\alpha} - \frac{n-1}{2\,n} \quad \left(K^* \leq \frac{1}{n}\,k_{n,\,\alpha} - \frac{n-1}{2\,n}\right)."$$

Under $\mathrm{H_0}$ we have $K^*\sqrt{n} \;\to\; No(0,\,1/12)$ so that for large n we reject $\mathrm{H_0}$ in favor of $\mathrm{H_A}$ ($\mathrm{H_A^\star}$) with level α if

$$K^* \geq \tau_{1-\alpha}\sqrt{\frac{1}{12\,n}} \quad \left(K^* \leq \tau_\alpha\sqrt{\frac{1}{12\,n}}\right).$$

This test has been implemented in the MATLAB–program *HAZARD_12*. For an application see Example 10/4.

10.3.5 HNBUE (HNWUE) Tests

A continuous distribution with $F(\cdot)$ and $S(\cdot) = 1 - F(\cdot)$ is said to be HNBUE (harmonic new better than used in expectation) [HNWUE (harmonic new worse than used in expectation)] if

$$\int\limits_x^\infty S(u)\,du \underset{(\geq)}{\leq} \mu\,\exp\!\left(-\frac{x}{\mu}\right) \quad \text{for } x \geq 0.$$

KLEFSJÖ (1983b) proposed several test statistics for testing

$\mathrm{H_0}$: '$F(\cdot)$ is exponential.'
in favor of
$\mathrm{H_A}$: '$F(\cdot)$ is HNBUE, but not exponential.'
(in favor of
$\mathrm{H_A^\star}$: '$F(\cdot)$ is HNWUE, but not exponential.')

He recommends

$$Q_1 = \sum_{j=1}^{n}\left[3\left(1-\frac{j}{n}\right)^2 - \frac{1}{3}\right] X_{(j)}\Big/T_n \qquad (10.35a)$$

as it has hight PITMAN efficiency and good power.

He derived the following exact null distribution of Q_1 :

$$\Pr(Q_1 > x) = \sum_{j=1}^{n}\prod_{\substack{i=1\\ i\neq j}}^{n}\frac{\alpha_j - x}{\alpha_j - \alpha_i}\,\vartheta_j \qquad (10.35b)$$

with

$$\alpha_j = -\frac{1}{3} + \left(1 - \frac{j}{n}\right)^2 + \frac{1 - j/n}{2\,n} \tag{10.35c}$$

and

$$\vartheta_j = \left\{ \begin{array}{ll} 1 & \text{for } \alpha < x \\ 0 & \text{otherwise.} \end{array} \right\} \tag{10.35d}$$

The limiting distribution of

$$Q = Q_1 \sqrt{\frac{45\,n}{4}} \tag{10.35e}$$

is $No(0,1)$. Exact critical values $q_{n,\gamma}$ resulting from (10.35b) are in the following table:

Table 10/7: Critical values $q_{n,\gamma}$ of $Q_1\sqrt{45\,n/4}$

n \ γ	0.01	0.05	0.10	0.90	0.95	0.99
10	−2.087	−1.656	−1.397	0.929	1.309	2.032
20	−2.191	−1.678	−1.383	1.060	1.449	2.195
∞	−2.326	−1.645	−1.282	1.282	1.645	2.326

Source: KLEFSJÖ (1983b, p. 71)

H_0 is rejected in favor of H_A $\left(\mathrm{H}_\mathrm{A}^\star\right)$ with level α if

$$Q_1\sqrt{45\,n/4} > q_{n,1-\alpha} \quad \left(Q_1\sqrt{45\,n/4} < q_{n,\alpha}\right).$$

This test is implemented in the MATLAB–program *HAZARD_12*.

Example 10/4: Testing for aging classes

We take the data set of Example 8/8: survival times (in days) of the $n = 43$ patients suffering from granulocytic leukemia. $x = 0$ is taken as the patient's date of diagnosis and begin of treatment. We want to test of this sample comes from a distribution belonging to one or the other of the aging classes discussed in Sect. 10.3.

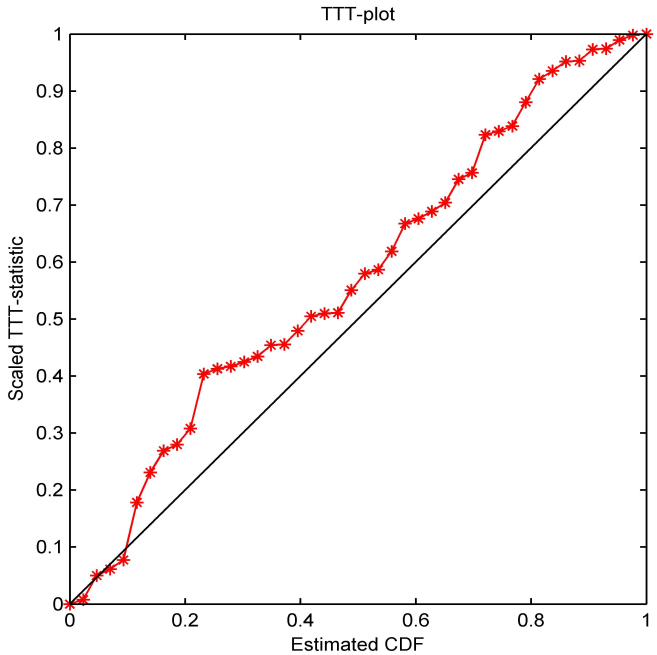
The TTT–plot of Fig. 10/8 clearly indicates an IHR distribution. This is also confirmed by the three IHR tests of Sect. 10.2.2 with EPSTEINS's cttot–test statistic $K_r = 24.11$, KLEFSJÖ's test statistic $A = 1.079$ and PROSCHAN/PYKE's test statistic $V_n = 512$.

For the test of the aging classes we get the following results:

- $B^* = 8.1546$ → Exponentiality is rejected in favor of IHRA with $\alpha \ll 0.01$.
- $J = 0.2248$ → Exponentiality is rejected in favor of NBU with $\alpha \approx 0.10$.
- $W = 1.3881$ → Exponentiality is rejected in favor of DMRL with $\alpha < 0.10$.
- $K^* = 0.0723$ → Exponentiality is rejected in favor of NBUE with $\alpha \approx 0.05$.
- $Q = 1.4356$ → Exponentiality is rejected in favor of HNBUE with $\alpha \approx 0.07$.

These results are in accordance with the chain of implications in Fig. 2/7. As we have evidence for IHR we expect to have evidence for IHRA, NBU, DMRL, NBUE and HNBUE.

Figure 10/8: TTT–plot for the 43 granulocytic leukemia patients



Part III

Appendices

Bibliography

A-A-A-A-A

AALEN, O. O. (1978): Nonparametric inference for a family of counting processes; *Ann. Statist.* **6**, 701–726

AARSET, M. V. (1985): The null distribution for a test of constant versus bathtub–failure rate; *Scand. Jour. Statist.* **12**, 55–62

AARSET, M. V. (1987): How to identify a bathtub hazard rate; *IEEE Trans. Rel.* **36**, 106–108

AL–HUSSAINI, E. K. / SULTAN, K. S. (2001): Reliability and hazard based on finite mixture models; in BALAKRISHNAN/RAO (eds.): *Handbook of Statistics,* Vol. 20 (Advances in Reliability), 139–183, North–Holland, Amsterdam etc.

ANDERSON, J. / SENTHILSELVAN, A. (1980): Smooth estimates for the hazard function; *Jour. Roy. Statist. Soc.* **B 42**, 322–327

ANDERSON, T. W. / DARLING, D. A. (1952): Asymptotic theory of certain goodness–of–fit criteria based on stochastic processes; *Ann. Math. Statist.* **23**, 193–213

ANTONIADIS, A. / GIJBELS, I. / MCGIBBON, B. (2000): Non–parametric estimation of the location of a change point in an otherwise smooth hazard rate function under random censoring; *Scand. Jour. Statist.* **27**, 501–519

ANTONIADIS, A. / GRÉGOIRE, G. / MCKEAGUE, I. W. (1994): Wavelet methods for curve estimation; *Jour. Amer. Statist. Ass.* **89**, 1340–1353

ANTONIADIS, A. / GRÉGOIRE, G. / NASON, G. (1999): Density and hazard rate estimation for right–censored data using wavelet methods; *Jour. Roy. Statist. Soc.* **B 61**, 63–84

ARNOLD, B. C. / ZAHEDI, H. (1988): On multivariate mean remaining life functions; *Jour. Multivar. Analysis* **25**, 1–9

ASADIA, M. (1999): Multivariate distributions characterized by a relationship between mean residual life and hazard rate; *Metrika* **49**, 121–126

B-B-B-B-B

BAGKAVOS, D. / PATIL, P. (2009): Variable bandwidth for nonparametric hazard rate estimation; *Comm. Statist. — Theory & Methods* **38**, 1055–1078

BAIN, L. J. / ENGELHARDT, M. (1991): *Statistical Analysis of Reliability and Life–Testing Models — Theory and Methods,* 2^{nd} ed.; Marcel Dekker, New York etc.

BALAKRISHNAN, N. / COHEN, A. C. (1991): *Order Statistics and Inference: Estimation Methods;* Academic Press, San Diego

BARLOW, R. E. (1968): Likelihood ratio tests for restricted families of probability distributions; *Ann. Math. Statist.* **39**, 547–566

BARLOW, R. E. (1979): Geometry of the total time on test transforms; *Naval Res. Log. Quart.* **26** 393–402

BARLOW, R. E. / BARTHOLOMEW, D. J. / BREMNER, J. M. / BRUNK, H. D. (1972): *Statistical Inference under Order Restrictions;* Wiley, New York etc.

BARLOW, R. E. / CAMPO, R. (1975): Total time on test processes and applications to failure analysis; in BARLOW/FUSSEL/SINGPURWALLA (eds.): *Reliability and Fault Tree Analysis*, SIAM, Philadelphia

BARLOW, R. E. / MARSHALL, A. W. (1964): Bounds for distributions with monotone hazard rate I, II; *Ann. Math. Statist.* **35**, 1234–1257, 1258–1274

BARLOW, R. E. / MARSHALL, A. W. / PROSCHAN, F. (1963): Properties of probability distributions with monotone hazard rate; *Ann. Math. Statist.* **34**, 375–389

BARLOW, R.E. / PROSCHAN, F. (1965): *Mathematical Theory of Reliability;* Wiley, New York etc.

BARLOW, R.E. / PROSCHAN, F. (1969): A note on tests for monotone failure rate based on incomplete data; *Ann. Math. Statist.* **40**, 595–600

BARLOW, R. E. / PROSCHAN, F. (1975): *Statistical Theory of Reliability and Life Testing*; Holt, Rinehart and Winston, New York etc.

BASU, A. P. (1971): Bivariate failure rate; *Jour. Amer. Statist. Ass.* **66**, 103–104

BEBBINGTON, M. / LAI, C. D. / ZITIKIS, R. (2008): Estimating the turning point of a bathtub–shaped failure distribution; *Jour. Statist. Plan. Inf.* **138**, 1157–1166

BERGMAN, B. (1977): Crossings in the total–time–on–test plot; *Scand. Jour. Statist.* **4**, 171–177

BERGMAN, B. (1979): On age replacement and the total time on test concept; *Scand. Jour. Statist.* **6**, 161–168

BERGMAN, B. / KLEFSJÖ, B. (1984): The total time on test concept and its use in reliability theory; *Operations Research* **32**, 596–606

BÉZANDRY, D. H. / BONNEY, G. E. / GANNOUN, A. (2005): Consistent estimation of the density and hazard rate functions for censored data via the wavelet method; *Statistics and Probability Letters* **74**, 366–372

BICKEL, P. J. (1969): Tests for monotone failure rate II; *Ann. Math. Statist.* **40**, 1250–1260

BICKEL, P. J. / DOKSUM, K. A. (1969): Tests for monotone failure rate based on normalized spacings; *Ann. Math. Statist.* **40**, 1216–1235

BIRNBAUM, Z. W. / SAUNDERS, S. C. (1968): A probabilistic interpretation of miner's rule; *SIAM — Jour. Appl. Math.* **16**, 637–652

BIRNBAUM, Z. W. / SAUNDERS, S. C. (1969): A new formula of life distribution; *SIAM — Jour. Appl. Prob.* **6**, 319–317

BLOCK, H. W. / SAVITS, T. H. (1982): The class of MIFRA lifetimes and its relation to other classes; *Nav. Res. Log. Quart.* **29**, 55–61

BLOCK, H. W. / SAVITS, T. H. (1984): Multivariate nonparametric classes in reliability; in KRISHNAIAH/RAO (eds.): *Handbook of Statistics*, Vol. 7 (Quality Control and Reliability), 121–129, North–Holland, Amsterdam etc.

BOUEZMARNI, T. / ROMBOUTS, J. V. K. (2008): Density and hazard rate estimation for censored and α–mixing date using gamma kernels; *Jour. Nonpar. Statist.* **20**, 627–643

BOWMAN, A. (1984): An alternative method of cross–validation for the smoothing of density estimates; *Biometrika* **71**, 353–366

BRINDLEY, E. C. JR. / THOMPSON, W. A. JR (1972): Dependence and aging aspects of multivariate survival; *Jour. Amer. Statist. Ass.* **67**, 822–830

BRYSON, M. C. / SIDDIQUI, M. M. (1969): Some criteria for ageing; *Jour. Amer. Statist. Ass.* **64**, 1472–1483

BURR, I. W. (1942): Cumulative frequency functions; *Ann. Math. Statist.* **13**, 215–232

C-C-C-C-C

CACOULOS, T. (1966): Estimation of a multivariate density; *Ann. Inst. Statist. Math.* **18**, 178–189

CHANDRA, N. K. / ROY, D. (2001): Some results on reversed hazard rate; *Probability in the Engineering and Informational Sciences* **15**, 95–102

CHANDRA, N. K. / ROY, D. (2005): Classification of distribution based on reversed hazard rate; *Calcutta Statist. Ass. Bull.* **56**, 231–249

CHENG, K. F. (1985): Tests for the equality of failure rates; *Biometrika* **72**, 211–215

CHENG, P. E. (1987): A nearest neighbour hazard rate estimator for randomly censored data; *Comm. Statist. — Theory & Methods* **16**, 613–625

COHEN, A. C. (1991): *Truncated and Censored Samples*; Marcel Dekker, New York etc.

COX, D. R. (1972): Regression models and life tables; *Jour. Roy. Statist. Soc.* **B 34**, 187–220

COX, D. R. / OAKES, D. (1984): *Analysis of Survival Data*; Chapman & Hall, London

CROWDER, M. J. / KIMBER, A. C. / SMITH, R. L. / SWEETING, T. J. (1991): *Statistical Analysis of Reliability Data*, Chapman & Hall, London etc.

CSÖRGÖ, M. / HORVÁTH. L. (1988): Nonparametric methods for change point problems; in KRISHHAIH/RAO (eds.): *Handbook of Statistics*, Vol. 7 (Quality Control and Reliability), 403–425, North–Holland, Amsterdam etc.

D-D-D-D-D

DAVID, H. A. / NAGARAJA, H. N. (2003): *Order Statistics;* Wiley, New York etc.

DETTE, H. / GEFELLER, O. (1995): The impact of different definitions of nearest neighbour distances for censored data on nearest neighbour kernel estimators of the hazard rate; *Jour. Nonpar. Statistics* **4**, 271–282

DHILLON, B. S. (1979): A hazard rate model; *IEEE Trans. Rel.* **28**, 150

DHILLON, B. S. (1981): Life distributions; *IEEE Trans. Rel.* **30**, 457–459

DIEHL, S. / STUTE, W. (1988): Kernel density and hazard function estimator in the presence of censoring; *Jour. Multivar. Analysis* **25**, 293–310

DOKSUM, K. A. / YANDELL, B. S. (1984): Testing for exponentiality; in KRISHNAIAH/SEN (eds.): *Handbook of Statistics*, Vol. 4 (Nonparametric Methods), 579–611, North–Holland, Amsterdam etc.

DUPUY, J.–F. / GNEYOU, K. E. (2011): A wavelet estimator of the intensity function with censored data; *Quality Technology and Quantitative Management* **8**, 401–410

E-E-E-E-E

EBRAHIMI, N. (1986): Classes of discrete decreasing and increasing mean residual life distributions; *IEEE Trans. Rel.* **35**, 403–405

EFRON, B. (1967): The two sample problem with censored data; *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. IV, Univ. of California Press, Berkeley, 831–853

ELANDT–JOHNSON, R. C. / JOHNSON, N. L. (1980): *Survival Models and Data Analysis;* Wiley, New York etc.

EPANECHNIKOV, V. A. (1969): Non–parametric estimation of a multivariate probability density; *Theory Probab. Appl.* **14**, 153–158

EPSTEIN, B. (1960): Testing for the validity of the assumption that the underlying distribution of life is exponential; *Technometrics* **2**, 83–101, 167–183

EPSTEIN, B. / SOBEL, M. (1953): Life testing; *Jour. Amer. Statist. Ass.* **48**, 486–502

F-F-F-F-F

FAILING, K. (1984): Neue Methoden zur nicht–parametrischen Schätzung von Dichte- und Hazardfunktionen bei zensierten Daten mit Anwendungen in klinischen Studien; in KÖHLER/TAUTU/WAGNER (eds.): *Der Beitrag der Informationsverarbeitung zum Fortschritt der Medizin*, 92–99, Springer, Berlin etc.

G-G-G-G-G

GASSER, T. / MÜLLER, H.–G. (1979): Kernel estimation for regression functions; in GASSER/ROSENBLATT (eds.): *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics, 23–68, Springer, Berlin/Heidelberg

GASSER, T. / MÜLLER, H.–G. / MAMMITZSCH, V. (1985): Kernels for nonparametric curve estimation; *Jour. Roy. Statist. Soc.* **B 47**, 238–252

GEFELLER, O. / DETTE, H. (1991): A comparative study on hazard function estimators employing nearest neighbour distances as bandwidths; in ADLASSNIG/GRABNER/BENGTSON/HANSEN (eds.): *Medical Information Europe 1991*, 963–987, Springer, Berlin etc.

GEFELLER, O. / DETTE, H. (1992): Nearest neighbour kernel estimation of the hazard function from censored data; *Jour. Statist. Comp. Simul.* **43**, 93–101

GEFELLER, O. / MICHELS, P. (1992): A review on smoothing methods for the estimation of the hazard rate based on kernel functions; in DODGE/WHITTAKER (eds.): *Proceedings of the* 10^{th} *Symposium on Computational Statistics*, 459–464

GEHAN, E. A. (1969): Estimating survival functions from the life table; *Jour. Chronic Disease* **21**, 629–644

GLASER, R. E. (1980): Bathtub and related failure rate characterizations; *Jour. Amer. Statist. Ass.* **75**, 667–672

GREENWOOD, M. (1926): The natural duration of cancer; *Report on Public Health and Medical Subjects*; Her Majesty's Stationary Office, London, Vol. 33, 1–26

GRENANDER; U. (1956): On the theory of mortality measurement, Part II; *Skand. Aktuarietidskr.* **39**, 125–153

GRIFFITH, W. S. (1982): Representations of distributions having monotone or bathtub–shaped failure rate; *IEEE Trans. Rel.* **31**, 95–96

GROSS, A. J. / CLARK, V. A. (1975): *Survival Distributions: Reliability Applications in the Biomedical Sciences;* Wiley, New York etc.

GU, C. (1996): Penalized likelihood hazard estimation: A general procedure; *Statistica Sinica* **6**, 861–876

GUESS, F. M. / PARK, D. H. (1988): Modeling discrete bathtub and upside–down bathtub mean residual life functions; *IEEE Trans. Rel.* **37**, 545–549

GUESS, F. M. / PROSCHAN, F. (1988): Mean residual life: Theory and application; in KRISHNAIAH/RAO (eds.): *Handbook of Statistics*, Vol. 7 (Quality Control and Reliability), 215–224, North–Holland, Amsterdam etc.

GUPTA, P. L. / GUPTA, R. C. (1997): On the multivariate normal hazard; *Jour. Multivar. Analysis* **62**, 64–73

GUPTA, P. L. / GUPTA, R. C. / TRIPATHI, R. C. (1997): On the monotonic properties of discrete failure rates; *Jour. Statist. Plan. Inf.* **65**, 255–268

GUPTA, R. C. (1981): Moments in terms of the mean residual life function; *IEEE Trans. Rel.* **30**, 450–451

GUPTA, R. D. / NANDA, A. K. (2001): Some results on reversed hazard rate ordering; *Comm. Statist. — Theory & Methods* **30**, 2447–2457

H-H-H-H-H

HÄRDLE, W. / KERKYACHARIAN, G. / PICARD, D. / TSYBAKOV, A. (1998): *Wavelets, Approximations and Statistical Applications;* Springer, Berlin etc.

HALL, P. / VAN KEILEGOM, I. (2005): Testing for monotone increasing hazard rate; *Ann. Statist.* **33**, 1109–1137

HARRIS, R. (1970): A multivariate definition for increasing hazard rate distribution functions; *Ann. Math. Statist.* **41**, 713–717

HENDERSON, R. (1990): A problem with the likelihood ratio test for a change–point hazard rate model; *Biometrika* **77**, 835–843

HERD, G. R. (1960): Estimation of reliability from incomplete data; *Proc. of the Sixth National Symposium on Reliability and Quality Control*, 202–217

HESS, K. R. / SERACHITOPOL, D. M. / BROWN, B. W. (1999): Hazard function estimators: A comparative simulation study; *Statistics in Medicine* **18**, 3075–3088

HJORTH, U. (1980): A reliability distribution with increasing, decreasing, constant, and bath–tub–shaped failure rates; *Technometrics* **22**, 99–107

HOLLANDER, M. / PARK, D. H. / PROSCHAN, F. (1986): A class of life distributions for aging; *Jour. Amer. Statist. Ass.* **81**, 91–95

HOLLANDER, M. / PROSCHAN, F. (1972): Testing whether new is better than used; *Ann. Math. Statist.* **43**, 1136–1146

HOLLANDER, M. / PROSCHAN, F. (1975): Tests for mean residual life; *Biometrika* **62**, 585–593

HOLLANDER, M. / PROSCHAN, F. (1984): Nonparametric concepts and methods in reliability; in KRISHNAIAH/SEN (eds.): *Handbook of Statistics*, Vol. 4 (Nonparametric Methods), 613–655, Elsevier, Amsterdam etc.

I-I-I-I-I

IZENMAN, A. I. (1991): Recent developments in nonparametric curve estimation; *Jour. Amer. Statist. Ass.* **86**, 205–224

J-J-J-J-J

JARJOURA, D. (1988): Smoothing hazard rates with cubic splines; *Comm. Statist. — Simul. & Comp.* **17**, 377–392

JOHNSON, L. G. (1964): *The Statistical Treatment of Fatigue Experiments;* Amsterdam

JOHNSON, N. L. / KOTZ, S. (1975): A vector multivariate hazard rate; *Jour. Multivar. Analysis* **5**, 53–66, 498

JOHNSON, N. L. / KOTZ, S. / BALAKRISHNAN, N. (1994): *Continuous Univariate Distributions, Vol. I,* 2^{nd} ed.; Wiley, New York etc.

JOHNSON, N. L. / KOTZ, S. / BALAKRISHNAN, N. (1995): *Continuous Univariate Distributions, Vol. II,* 2^{nd} ed.; Wiley, New York etc.

JOHNSON, N. L. / KOTZ, S. / KEMP, A. W. (1992): *Univariate Discrete Distributions*; Wiley, New York etc.

JOSHI, S. N. / MCEACHERN, S. N. (1997): Isotonic maximum likelihood estimation for the change point of a hazard rate; *Sankhya* **A 59**, 392–407

K-K-K-K-K

KALBFLEISCH, J. D. / PRENTICE, R. L. (1980): *The Statistical Analysis of Failure Time Data;* Wiley, New York etc.

KAPLAN, E. L. / MEIER, P. (1958): Non–parametric estimation from incomplete data; *Jour. Amer. Statist. Ass.* **53**, 457–481

KARUNAMUNI, R. J. / ALBERTS, T. (2005): On boundary correction in kernel density estimation; `www.cims.nyu.edu alberts/pub/SM2005.pdf`

KEMP, A. W. (2004): Classes of discrete lifetime distributions; *Comm. Statist. — Theory & Methods* **33**, 3069–3093

KIEFER, J. / WOLFOWITZ, J. (1956): Consistency of the maximum likelihood estimator in the presence of infinitely many identical parameters; *Ann. Math. Statist.* **27**, 897–906

KIMBALL, A. W. (1960): Estimation of mortality intensities in animal experiments; *Biometrics* **16**, 505–521

KLEFSJÖ, B. (1982a): The HNBUE and HNWUE classes of life distributions; *Nav. Res. Log. Quart.* **2**, 331–344

KLEFSJÖ, B. (1982b): On ageing properties and total time on test transforms; *Scand. Jour. Statist.* **9**, 37–41

KLEFSJÖ, B. (1983a): Some tests against ageing based on the total time on test transform; *Comm. Statist. — Theory & Methods* **12**, 907–927

KLEFSJÖ, B. (1983b): Testing exponentiality against HNBUE; *Scand. Jour. Statist.* **10**, 65–75

KLEIN, J. P. / MOESCHBERGER, M. L. (1997): *Survival Analysis;* Springer, New York etc.

KOTZ, S. / BALAKRISHNAN, N. / JOHNSON, N. L. (2000): *Continuous Multivariate Distributions,* Vol. I; Wiley, New York etc.

KRISHNAIAH, P. R. / MIAO, B. Q. (1988): Review about estimation of change points; in KRISHNAIAH/RAO (eds.): *Handbook of Statistics*, Vol. 7 (Quality Control and Reliability), 375–402, North–Holland, Amsterdam etc.

KUNDU, C. / NANDA, A. K. / HU, T. (2009): A note on the reversed hazard rate of order statistics and record values; *Jour. Statist. Plan. Inf.* **139**, 1257–1265

KUNITZ, H. (1989): A new class of bathtub–shaped hazard rates and its application in a comparison of two test–statistics; *IEEE Trans. Rel.* **38**, 351–354

L-L-L-L-L

LAGAKOS, S. W. / BARRAJ, L. M. / DE GRUTTOLA, V. (1988): Nonparametric analysis of truncated survival data with application to AIDS; *Biometrika* **75**, 515–523

LAI, C. D. (2013): Issues concerning constructions of discrete lifetime models; *Quality Technology and Quality Management* **10**, 251—262

LAI, C. D. / XIE, M. / MURTHY, D.N. P. (2001): Bathtub shaped failure rate life distributions; in BALAKRISHNAN/RAO (eds.): *Handbook of Statistics*, Vol. 20, (Advances in Reliability), 69–104, Elsevier, Amsterdam

LANGBERG, N. A. / LEON, R. V. / LYNCH, J. / PROSCHAN, F. (1980): Extreme points of the class of discrete decreasing failure rate life distributions; *Math. Oper. Res.* **5**, 35–42

LANGBERG, N. A. / LEON, R. V. / LYNCH, J. / PROSCHAN, F. (1982): Extreme points of the class of discrete decreasing failure rate average life distributions; *TIMS — Studies in the Management Sciences* **19**, 297–304

LAWLESS, J. F. (1982): *Statistical Models and Methods for Lifetime Data;* Wiley, New York etc.

LEEMIS, L. M. (1986): Lifetime distribution identities; *IEEE Trans. Rel.* **35**, 170–174

LEEMIS, L. M. (1995): *Reliability: Probabilistic Models and Statistical Methods;* Prentice–Hall, New Jersey

LI, L. (2002): Hazard rate estimation for censored data by wavelet methods; *Comm. Statist. — Theory & Methods* **31**, 943–960

LIU, Y. C. / VAN RYZIN, J. (1985): A histogram estimator of the hazard rate with censored data; *Ann. Statist.* **13**, 592—605

LO, S. H. / MACK, Y. P. / WANG, J.–L. (1989): Density and hazard rate estimation for censored data via strong representation of Kaplan/Meier estimate; *Prob. Theo. Rel. Fields* **80**, 461–473

LOADER, C. R. (1991): Inference for a hazard rate change point; *Biometrika* **78**, 749–757

LONDON, D. (1988): *Survival Models and Their Estimation*, 2nd ed.; ACTEX Publications, Winstedt & Avon, Conn.

M-M-M-M-M

MA, C. (2000): A note on the multivariate normal hazard; *Jour. Multivar. Analysis* **73**, 282–283

MARSHALL, A. W. / OLKIN, I. (2007): *Life Distributions — Structure of Nonparametric, Semiparametric and Parametric Families*; Springer, New York

MARSHALL, A. W. / PROSCHAN, F. (1965): Maximum likelihood estimation for distributions with monotone failure rate; *Ann. Math. Statist.* **36**, 69–77

MARSHALL, A. W. / PROSCHAN, F. (1972): Classes of distributions applicable in replacement with renewal implications; in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3, 395–415; Univ. of California Press, Berkeley

MCGILL, J. I. (1992): The multivariate hazard gradient and moments of the truncated multinormal distribution; *Comm. Statist. — Theory & Methods* **21**, 3053–3060

MEEKER, W. Q. / ESCOBAR, L. A. (1998): *Statistical Methods for Reliability Data*; Wiley, New York etc.

MEILIJSON, I. 1972): Limiting properties of the mean residual lifetime function; *Ann. Math. Statist.* **43**, 354–357

MESSER, K. / GOLDSTEIN, L. (1993): A new class of kernels for nonparametric curve estimation; *Ann. Statist.* **21**, 179–195

MI, J. (1995): Bathtub failure rates and upside–down bathtub mean residual life; *IEEE Trans. Rel.* **44**, 388–391

MILLER, R. G. JR. (1981): *Survival Analysis;* Wiley, New York etc.

MILLS, J. P. (1926): Tables of the ratio area to bounding ordinate, for any portion of normal curve; *Biometrika* **18**, 395–400

MORRISON, D. G. (1978): On linearly increasing mean residual lifetimes; *Jour. Appl. Prob.* **15**, 617–620

MÜLLER, H.–G. (1991): Smooth optimum kernel estimators near endpoints; *Biometrika* **78**, 521–530

MÜLLER, H.–G. (1993): On the boundary kernel method for non–parametric curve estimation near endpoints; *Scand. Jour. Statist.* **20**, 313–328

MÜLLER, H.–G. / WANG, J.–L. (1990a): Nonparametric analysis of changes in hazard rates for censored data: An alternative to change–point models; *Biometrika* **77**, 305–314

MÜLLER, H.–G. / WANG, J.–L. (1990b): Locally adaptive hazard smoothing; *Probability Theory and Related Fields* **85**, 823–838

MÜLLER, H.–G. / WANG, J.–L. (1994): Hazard rate estimation under random censoring with varying kernels and bandwidths; *Biometrics* **50**, 61–76

MÜLLER, H.–G. / WANG, J.–L. / CAPRA, W. B. (1997): From life tables to hazard rates: the transformation approach; *Biometrika* **84**, 881–892

MUTH, E. J. (1974): Moments expressed in terms of the hazard function and applications; *Microelectronics and Reliability* **13**, 469–471

MUTH, E. J. (1980): Memory as a property of probability functions; *IEEE Trans. Rel.* **29**, 160–164

MYKYTYN, S. W. / SANTNER, T. J. (1981): Maximum likelihood estimation of the survival function based on censored data under hazard rate assumptions; *Comm. Statist. — Theory & Methods* **10**, 1369–1387

N-N-N-N-N

NAKAGAWA, T. / OSAKI, S. (1975): The discrete Weibull distribution; *IEEE Trans. Rel.* **24**, 300–301

NANDA, A. K. / GUPTA, R. D. (2001): Some properties of reversed hazard rate function; *Statist. Methods* **3**, 108–124

NANDA, A. K. / GUPTA, R. D. (2004): Some properties of reversed hazard rate function — Corrections; *Statist. Methods* **6**, 90–91

NAVARRO, J. / RUIZ, J. M. (2004): A characterization of the multivariate normal distribution by using the hazard gradient; *Ann. Inst. Statist. Math.* **56**, 361–367

NELSON, W. (1969): Hazard plotting for incomplete data; *Journal of Quality Technology* **1**, 27–52

NELSON, W. (1970): Hazard plotting methods for analysis of life data with different failure modes; *Journal of Quality Technology* **2**, 126–149

NELSON, W. (1972): Theory and applications of hazard plotting for censored failure data; *Technometrics* **14**, 945–966

NELSON, W. (1982): *Applied Life Data Analysis;* Wiley, New York etc.

NELSON, W. (1984): *Accelerated Testing — Statistical Models, Plans, and Data Analysis*; Wiley, New York etc.

NGUYEN, H. T. / ROGERS, G. S. / WALKER, E. A. (1984): Estimation in change–point hazard rate models; *Biometrika* **71**, 299–304

NIELSEN, J. P. (2003): Variable bandwidth kernel hazard estimators; *Jour. Nonpar. Statist.* **15**, 355–376

O-O-O-O-O

OAKES, D. / DASU, T. (1980): A note on residual life; *Biometrika* **71**, 409–410

O'SULLIVAN, F. (1988a): Nonparametric estimation of relative risk using splines and cross–validation; *SIAM — Jour. Sci. Statist. Comp.* **9**, 531–542

O'SULLIVAN, F. (1988b): Fast computation of fully automated log–density and log–hazard estimators; *SIAM — Jour. Sci. Statist. Comp.* **9**, 363–379

P-P-P-P-P

PADGETT, W. J. (1988): Nonparametric estimation of density and hazard rate function when samples are censored; in KRISHNAIAH/RAO (eds.): *Handbook of Statistics*, Vol. 7 (Quality Control and Reliability), 313–333, North–Holland, Amsterdam etc.

PADGETT, W. J. / SPURRIER, J. D. (1985): On discrete failure models; *IEEE Trans. Rel.* **34**, 253–256

PADGETT, W. J. / WEI, L. J. (1980): Maximum likelihood estimation of a distribution with increasing failure rate based on censored observations; *Biometrika* **67**, 470–474

PARZEN, E. (1962): On estimation of a probability density function and mode; *Ann. Math. Statist.* **33**, 1065–1076

PATEL, J. (1973): A catalogue of failure distributions; *Comm. Statist.* **1**, 281–284

PATIL, P. (1997): Nonparametric hazard rate estimation by orthogonal wavelet methods; *Jour. Statist. Plan. Inf.* **60**, 153–168

PATRA, K. / DEY, D. K. (2002): A general class of change point and change curve models for lifetime data; *Ann. Inst. Statist. Math.* **54**, 517–530

PRAKASA RAO, B. L. S. (1970): Estimation for distributions with monotone failure rate; *Ann. Math. Statist.* **41**, 507–529

PRAKASA RAO, B. L. S. (1983): *Nonparametric Functional Estimation*; Academic Press, New York

PROSCHAN, F. / PYKE, R. (1967): Tests for monotone failure rate; *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.* Vol. 3, 293–312

PURI, P. S. / RUBIN, H. (1974): On a characterization of the family of distributions with constant multivariate failure rates; *The Annals of Probability* **2**, 738–740

PYKE, R. (1965): Spacings; Jour. Roy. Statist. Soc. **B 27**, 395–449

Q-Q-Q-Q-Q

QIU, P. / SHENG, J. (2008): A two–stage procedure for comparing hazard rate functions; *Jour. Roy. Statist. Soc.* **B 70**, 191–208

R-R-R-R-R

RAMLAU–HANSEN, H. (1983): Smoothing counting process intensities by means of kernel functions; *Ann. Statist.* **11**, 453–466

RICE, J. / ROSENBLATT, M. (1976): Estimation of log survivor function and hazard rate; *Sankhya* **A 38**, 60–78

RINNE, H. (2009): *The Weibull Distribution — A Handbook;* CRC Press, Boca Raton

RINNE, H. (2010): *Location–scale Distributions — Linear Estimation and Probability Plotting Using MATLAB;* `httpp://geb.uni-giessen.de/volltexte/2010/7607`

ROLSKI, T. (1975): Mean residual life; *Bull. Int. Statist. Inst.* **46**, 266–270

ROSENBERG, P. S. (1995): Hazard function estimation using B–splines; *Biometrics* **51**, 874–887

ROSENBLATT, M. (1956): Remarks on some nonparametric estimates of a density function; *Ann. Math. Statist.* **27**, 832–837

ROY, D. / GUPTA, R. P. (1992): Classifications of discrete lives; *Microelectronics and Reliability* **32**, 1459–1473

RUDEMO, M. (1982): Empirical choice of histograms and kernel density estimators; *Scand. Jour. Statist.* **9**, 65–78

S-S-S-S-S

SACHER, G. A. (1956): On the statistical nature of mortality with special reference to chronic radiation mortality; *Radiology* **67**, 250–257

SALHA, R. (w.y.): Adaptive kernel estimation of the hazard rate function; *Research Magazine* **17**, 71–81; Islamic University of Gaza, Department of Mathematics

SALVIA, A. A. (1985): Reliability application of the alpha distribution; *IEEE Trans. Rel.* **34**, 251–252

SALVIA, A. A. / BOLLINGER, R. C. (1982): On discrete hazard rates; *IEEE Trans. Rel.* **31**, 458–459

SANKARAN, P. G. / GLEEJA, V. L. / JACOB, T. M. (2007): Nonparametric estimation of reversed hazard rate; *Calcutta Statist. Ass. Bull.* **59**, 55–68

SARDA, P. / VIEU, P. (1990): Smoothing parameter selections in hazard rate estimation; *Statist. Prob. Letters* **11**, 429–434

SCHÄFER, H. (1985): A note on data–adaptive kernel estimation of the hazard rate and density function in the random censorship situation; *Ann. Statist.* **13**, 818–826

SCHÄFER, H. (1986): Local convergence of empirical measures in the random censorship situation with application to density and rate estimation; *Ann. Statist.* **14**, 1240–1245

SCHIFFMAN, D. A. (1986): The score statistic in constancy testing for a discrete hazard rate; *IEEE Trans. Rel.* **35**, 590–594

SEAL, H. L. (1954): The estimation of mortality and other decrement probabilities; *Skand. Aktuarietidskr.* **37**, 137–162

SHAKED, M. / SHANTHIKUMAR, J. G. (1987): Multivariate hazard rates and stochastic ordering, *Adv. Appl. Probab.* **19**, 123–137

SHAKED, M. / SHANTHIKUMAR, J. G. (1989): Multivariate conditional hazard rate and the MIFRA and MIFR properties, *Jour. Appl. Probab.* **25**, 150–168

SHAKED, M. / SHANTHIKUMAR, J. G. / VALDEZ–TORRES, J. B. (1995): Discrete hazard rate functions; *Comput. Oper. Res.* **22**, 391–402

SHYROCK, H. S. / SIEGEL, J. S. (1976): *The Methods and Materials of Demography*; Academic Press, San Diego

SILVA, R. B. / BARRETO–SOUZA, W. / CORSEIRA, G. M. (2010): A new distribuiton with decreasing, increasing and upside–down bathtub failure rate; *Comp. Statist. & Data Analysis* **54**, 935-944

SINGPURWALLA, N. D. (2006): The hazard potential: Introduction and overview; *Jour. Amer. Statist. Ass.* **101**, 1705–1717

SINGPURWALLA, N. D. / WONG, M. Y. (1983): Estimation of the failure rate — A survey of nonparametric methods (Part I – Non–Bayesian methods); *Comm. Statist. — Theory & Methods* **12**, 559–588

SMITH, P. J. (2002): *Analysis of Failure and Survival Data;* Chapman & Hall / CRC Press, Boca Raton

SPIEGELMAN, M. (1968): *Introduction to Demography*; rev. ed., Harvard University Press, Cambridge/Mass.

STACY, E. W. (1962): A generalization of the gamma distribution; *Ann. Math. Statist.* **33**, 1187–1192

STATISTISCHES BUNDESAMT (ed.) 2004): *Perioden–Sterbetafeln für Deutschland (1871/1881 bis 2001/2003)*; Wiesbaden

STEIN, W. E. / DATTERO, R. (1984): A new discrete Weibull distribution; *IEEE Trans. Rel.* **33**, 196–197

SWARTZ, G. B. (1973): The mean residual life function; *IEEE Trans. Rel.* **22**, 108–109

SWEET, A. L. (1990): On the hazard rate of the lognormal distribution; *IEEE Trans. Rel.* **39**, 325–328

T-T-T-T-T

TANNER, M. A. (1983): A note on the variabel kernel estimator of the hazard function from randomly censored data; *Ann. Statist.* **11**, 994–998

TANNER, M. A. (1984): Data–based nonparametric hazard estimation (Algorithm AS 202); *Jour. Roy. Statist. Soc. (Applied Statistics)* **C 33**, 248–258

TANNER, M. A. / WONG, W. H. (1983): The estimation of the hazard rate function from randomly censored data by the kernel method; *Ann. Statist.* **11**, 989–993

TANNER, M. A. / WONG, W. H. (1984): Data–based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis; *Jour. Amer. Statist. Ass.* **79**, 174–182

TEISSIER, G. (1934): Recherches sur le viellissement et sur les lois de la mortalité; *Annales de Physiologie et de Physico–chemie Biologique*, **X**, 237–284

TERRELL, G. R. (1990): Maximal smoothing principle in density estimation; *Jour. Amer. Statist. Ass.* **85**, 470–477

U-U-U-U-U

UZUNOGULLARI, Ü. / WANG, J.–L. (1992): A comparison of hazard rate estimators for left truncated and right censored data; *Biometrika* **79**, 297–310

W-W-W-W-W

WAND, M. P. / JONES, M. C. (1995): *Kernel Smoothing*; Chapman & Hall, London

WATSON, G. S. / LEADBETTER, M. R. (1964a): Hazard analysis I; *Biometrika* **51**, 175–184

WATSON, G. S. / LEADBETTER, M. R. (1964b): Hazard analysis II; *Sankhya* **A 26**, 101–116

WATSON, G. S. / WELLS, W. T. (1961): On the possibility of improving the mean useful life of items by eliminating those with short lives; *Technometrics* **3**, 281–298

WOODROOFE, M. (1970): On choosing a delta–sequence; *Ann. Math. Statist.* **41**, 1665–1671

WU, S. / WELLS, M. T. (2003): Nonparametric estimation of hazard functions by wavelet methods; *Jour. Nonpar. Statist.* **15**, 187–203

X-X-X-X-X

XEKALAKI, E. (1983a): A property of the Yule distribution and its applications; *Comm. Statist. — Theory & Methods* **12**, 1181–1189

XEKALAKI, E. (1983b): Hazard functions and life distributions in discrete time; *Comm. Statist. — Theory & Methods* **12**, 2503–2509

Y-Y-Y-Y-Y

YANDELL, B. S. (1983): Nonparametric inference for rates with censored survival data; *Ann. Statist.* **11**, 1119–1135

Author Index

Subject Index

Included MATLAB-programs

The PDF–file of the monograph "The Hazard Rate — Theory and Inference" is supplemented with two ZIP–files containing MATLAB–programs. You should have version 7.4.0 (R 2007a) or higher of MATLAB to run these programs successfully.

First ZIP–file

Distributions.zip, which should be extracted into a new directory — perhaps named 'Distributions' — contains the programs creating plots of the density function (or the probability mass function), the survival function, the hazard rate and the mean residual life function of 62 continuous and 32 discrete distributions. The programs are menu–driven. To invoke continuous (discrete) distributions type ContDist (DiscDist) into the Command Window after you have switched to the directory mentioned above. After having chosen a distribution you will see a picture with the formula of the PDF (PMF) together with the domain of its parameters. After your parameter–input has been checked you will see a plot of the four functions mentioned above. You can repeat with another set of parameter values for the same distribution or you can go to another distribution.

Second ZIP–file

Inference.zip, which should be extracted into a new directory — perhaps named 'Inference' — contains 12 programs HAZARD_xx intended to do estimation and testing as described in Part II of this monograph. Here is information on the Hazard–programs.

HAZARD_01 — This program computes the pointwise hazard rate by maximum–likelihood, the survival function according to KAPLAN/MEIER and the cumulative hazard rate according to NELSON/AALEN, all functions with 95%–confidence limits. The relevant formulas are in Chapter 5 of the monograph.
Input: A sample of non–grouped data stored in the Workspace as a $(n \times 2)$–matrix named y. The first column is for the observations in arbitrary order, the second column is for the corresponding censoring indicator: 1 for an uncensored observation, 0 for an censored observation.
Output: A table with the numerical results and a figure showing the three estimated functions with their 95%–confidence limits.

HAZARD_02 — This program estimates a life table and all its functions according to the formulas in Chapter 6 of the monograph.
Input: A $(3 \times k)$–matrix named y has to be stored in the Workspace. The first column is for lower class limits in ascending order, the second column for the number of censored lifetimes in corresponding class, the third column for the uncensored lifetimes in the corresponding class. The program asks you for the sample size n and the number k of classes.
Output: A life table with 12 columns and k rows and a figure displaying the histogram, the survival function and the hazard rate.

HAZARD_03 — Maximum likelihood estimation of an *increasing* hazard rate for a *continuous* distribution according to Chapter 7.
Input: A $(n \times 2)$–matrix named y has to be stored in the Workspace. The first column is for the observations in ascending order, the second column for the corresponding censoring indicator: 1 for an uncensored observation, 0 for a censored observation.
Output: A table showing the uncensored observations and the hazard rate which is constant between any two uncensored observation, a figure displaying the graphs of the estimated hazard rate, the density function and the survival function.

HAZARD_04 — Maximum likelihood estimation of a *decreasing* hazard rate for a *continuous* distribution according to Chapter 7.
Input: same as in HAZARD_03
Output: same as in HAZARD_03

HAZARD_05 — Maximum likelihood estimation of the hazard rate for a *discrete* distribution with realizations $x = 1, 2, 3, \ldots$ according to Chapter 7.
Input: You are asked whether the hazard has to be increasing or decreasing. A vector y to be stored in the Workspace with the counts for each realization.
Output: A table showing the estimated hazard rate, the probability mass function and the survival function and a figure display the graph of the three functions.

HAZARD_06 — *User–supplied fixed–bandwidth kernel estimation* of the hazard rate with one out of four kernels and corresponding boundary kernel
Input: A $(n \times 2)$–matrix named y to be stored in the Workspace. The first column is for the — not necessarily ordered — observations, the second column for the corresponding censoring indicator: 1 for an uncensored observation, 0 for a censored observation. The program asks for the number of gridpoints in the plot of the smoothed hazard rate, for the kernel to be used (uniform, EPANECHNIKOV, biweight or triweight) and for a bandwidth. (The first bandwidth is set automatically.)
Output: List of all bandwidths chosen (maximum number: 20), plot of the pointwise cumulative hazard rate and plot of each smoothed hazard rate with 95%–confidence limits.

HAZARD_07 — *Local kernel estimation* of the hazard rate with one out of four kernels and corresponding boundary kernel
Input: A $(n \times 2)$–matrix named y to be stored in the Workspace. The first column is for the observations in ascending order, the second column for the corresponding censoring indicator: 1 for an uncensored observation, 0 for a censored observation. There have to be no ties neither among the uncensored nor among the censored observations, but ties between uncensored and censored observations are allowed. In this case the uncensored observation precedes the censored observation The program asks for the number of gridpoints in the plot of the smoothed hazard rate, for the kernel to be used (uniform, EPANECHNIKOV, biweight or triweight) and for a parameter specifying the k–nearest neighbor.
Output: Plot of the pointwise cumulative hazard rate and plot of the smoothed hazard rate with 95%–confidence limits.

HAZARD_08 — *Variable kernel estimation* of the hazard rate with one out of four kernels and corresponding boundary kernel
Input: A $(n \times 2)$–matrix named y to be stored in the Workspace. The first column is for the observations in ascending order, the second column for the corresponding censoring indicator: 1 for an uncensored observation, 0 for a censored observation. There have to be no ties neither among the uncensored nor among the censored observations, but ties between uncensored and censored observations are allowed. In this case the uncensored observation precedes the censored observation The program asks for the number of gridpoints in the plot of the smoothed hazard rate, for the kernel to be used (uniform, EPANECHNIKOV, biweight or triweight) and for a parameter specifying the neighborhood.
Output: Plot of the pointwise cumulative hazard rate and plot of the smoothed hazard rate with 95%–confidence limits.

HAZARD_09 — Graphical check for constancy of the hazard rate
Input: A $(n \times 1)$–matrix named y to be stored in the Workspace or
a $(k \times 1)$–matrix named y to be stored in the Workspace of singly censored and ascendingly ordered observations, where the sample size n is asked by the program or
a $(n \times 2)$–matrix named y to be stored in the Workspace of multiply or randomly censored observations in ascending order, first column for the observations, second column for the corresponding censoring indicator: 1 for uncensored observation, 0 for censored observation.
Output: A figure with the probability plot and the TTT–plot

HAZARD_10 — This program tests for IHR or DHR using the procedures of KLEFSJÖ, EPSTEIN and PROSCHAN/PYKE depending on whether the sample is singly censored or uncensored.
Input: A column vector y to be stored in the Workspace with ascendingly ordered observations. The program asks to enter r the number of observations and n the sample size. For $r = n$ we have an uncensored sample, for $r < n$ a singly censored sample.
Output: TTT–plot and the test statistics with critical values.

HAZARD_11 — This program tests for bathtub–shape or inverted bathtub–shape of the hazard rate using the procedures of BERGMAN and AARSET.
Input: A column vector y to be stored in the Workspace with n ascendingly ordered and uncensored observations. The program asks whether you want to test for bathtub-shape or for inverted bathtub–shape.
Output: TTT–plot and test statistics of BERGMAN and AARSET together with critical values.

HAZARD_12 — This program tests for aging classes other than IHR and DHR
Input: A column vector y to be stored in the Workspace with n ascendingly ordered and uncensored observations. The program asks you what aging class you want to test for.
Output: TTT–plot and test statistic for the chosen class together with critical values.