Statistics of Reliability and Material Fatigue

Prof. Dr. Christine Müller
Technische Universität Dortmund

Wintersemester 2021/2022

Contents

Preface

We want to thank Professor B. M. Hillberry and Eric J. Tuegel for providing us the so-called Virkler data. Most of the other data were obtained in the DFG Collaborative Research Center SFB 823 *Statistical modeling of nonlinear dynamic processes*.

Special recommended books are:

1. Castillo, E. and Fernandez-Canteli, A. (2009). *A Unified Statistical Methodology for Modeling Fatigue Damage*. Springer Science + Business Media, Dordrecht.

2. Deshpande, J.V. and Purohit, S.G. (2015). *Lifetime Data: Statistical Models and Methods*. World Scientific Publishing Co Inc, New Jersey.

3. Kahle, W. and Liebscher, E. (2013). *Zuverlässigkeitsanalye und Qualitätssicherung*. Oldenbourg Verlag, München.

4. Kahle, W., Mercier, S., and Paroissin, C. (2016). *Degradation Processes in Reliability*. Wiley, New York.

5. Meeker, W.Q. and Escobar, L.A. (1998). *Statistical Methods for Reliability Data*. Wiley, New York.

6. Sobczyk, K. and Spencer, B.F. (1992). *Random Fatigue. From Data to Theory*. Academic Press, London.

Chapter 1

Introduction

1.0.1 Example (Prestressed concrete beam)
Uo to now, there are eleven experiments at prestressed concrete beams which different stress levels (TR01 with $\Delta\sigma_p = 200$ MPa, TR02 with $\Delta\sigma_p = 455$ MPa, TR03 with $\Delta\sigma_p = 200$ MPa, TR04 with $\Delta\sigma_p = 150$ MPa, TR05 with $\Delta\sigma_p = 98$ MPa, SB01 with $\Delta\sigma_p = 200$ MPa, SB02 with $\Delta\sigma_p = 100$ MPa, SB03 with $\Delta\sigma_p = 60$ MPa, SB04 with $\Delta\sigma_p = 80$ MPa, SB05 with $\Delta\sigma_p = 80$ MPa, SB06 with $\Delta\sigma_p = 50$ MPa). During the experiment the widening of an initial crack was observed. The left hand side of Figure 1.1 shows the growth curves of the crack width for TR01 and TR02. The jumps which can be seen in the growth curve are caused by the breakening of the tension wires, see the left hand side of Figure 1.1. There are 5 strands each with 7 wires in each beam so that at most 35 jumps could be observed. However much less jumps are observed before the failure of the beam. The data set `failure_times_tension_wires.RData` contains the times, given by the variable `t_jm`, between these jumps and the corresponding stress levels, given by the variable `s`. For the meaning of the third variable `stress`, see Section 6.3.

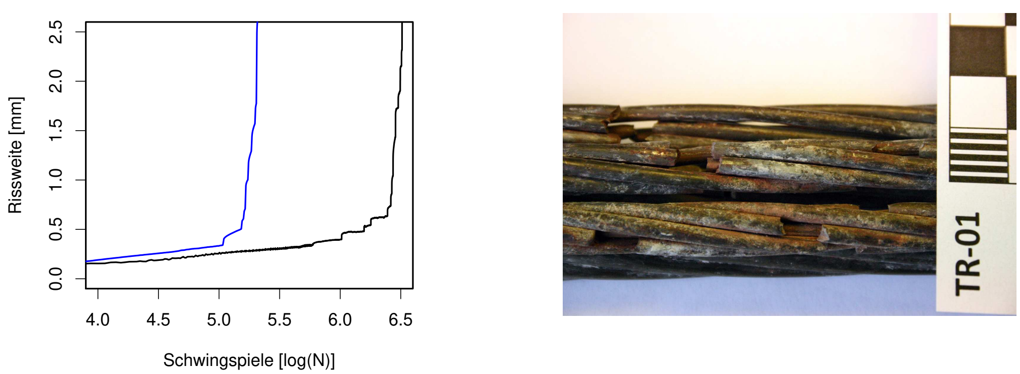

Figure 1.1: Left: crack growth curves for TR01 (black) and TR02 (blue), right: and broken tension wires

1.0.2 Example (Hudak crack grwoth data)
The data of Hudak et al. (1978) concern 21 steel specimen exposed to the same stress given by cyclic load. These data are contained in the file `Hudak_data.asc` which can be read with the R function `source` and was obtained from the R package `dhglm`. The observation of the crack

growth at each specimen started at an initial crack of 0.9 inches and the crack length of the crack was measured after specific numbers of load cycles. These load cycles are given in the data set `Hudak_data.asc` by the variable `cycle`. The crack length is given by the variable `crack0`, however it is the crack length at the predecessor time point. Figure 1.2 provides the crack growth curves for the untransformed data and the data with a logarithmic transformation of the crack length. The logarithmic transformation leads to an almost linear growth while this is clearly not the cause for the untransformed data.

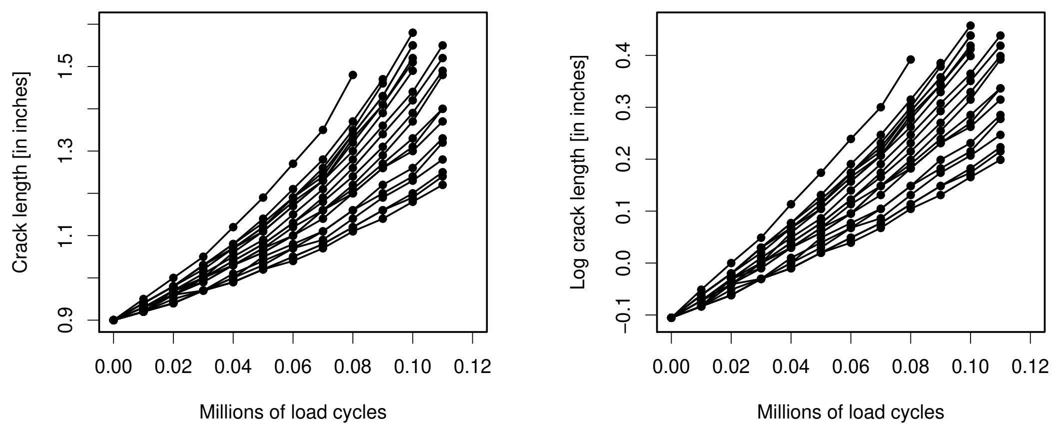

Figure 1.2: Crack growth curves for the Hudak data, left with untransformed crack lengths, right with the logarithm of the crack lengths

1.0.3 Example (Virkler data)
The data of Virkler et al. (1979) given by the data set `BasicVirklerdata.xls` and `Virkler_data.asc` concern 68 steel specimen. For each specimen, the time was measured when an initial crack of 9mm in the specimen reaches a given length value, i.e. the length is here the explanatory variable and the time the dependent variable. The measurements are taken at 164 length values which are the same for all specimens. The particular aim is to predict the time of a given length value in a new specimen as seen in Figure 1.3.

1.0.4 Example (Crack growth from photos)
To study the crack growth behavior of micro cracks in a steel specimen exposed to cyclic load, cracks were detected in photos of the steel surface by the crack detection package `crackrec` given by Gunkel et al. (2012) and the 112 longest cracks at the end were traced back. The resulting 112 crack growth values are given in the data set `top112.length.P10.Rdata` and the six longest cracks at the end are shown in Figure 1.4. The data set `top112.length.P10.Rdata` does not contain the time points of the photos. Hence it is important to know that these photos were done at the beginning (time 0), and in steps of thousand load cycles up to 20 000 load cycles and then at 25 000, 30 000, 35 000, 37 000, 39 000, 40 000, 42 000, 44 000 load cycles so that 29 time points are available. Figure 1.4 shows that there is no strict increasing growth. This is due to

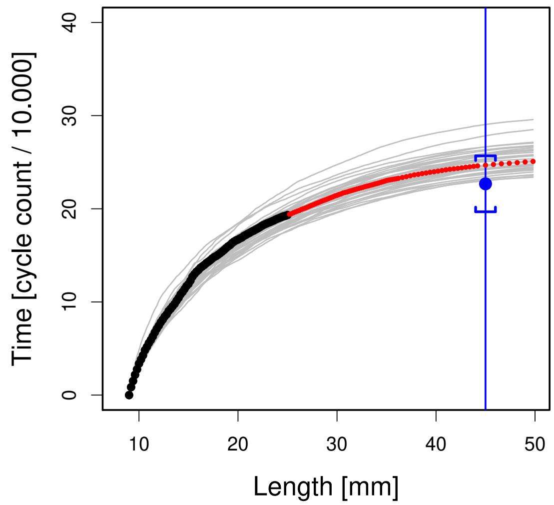

Figure 1.3: Crack growth curves for the Virkler data with a series (red) for which a future time shall be predicted

the varying quality of the photos. There is also a strange crack growth curve starting not near zero. This is caused by a contamination of the steel surface resulting in a big pit so that the automatic program detected this falsely as a crack.

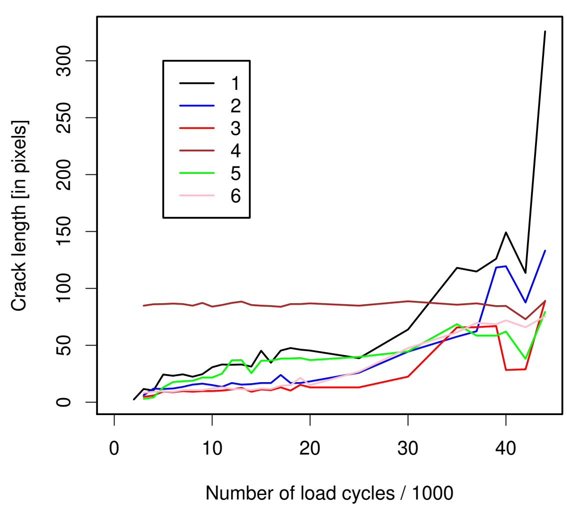

Figure 1.4: Growth curves of the six largest micro cracks obtained from photos by back tracing

1.0.5 Example (Experiments with isolated tension wires)
Figure 1.5 shows the life times in load cycles until the failure of 32 tension wires, given by the data set `alter_Spannstahl.txt`. There are 4 censored data since some tension wire did not break up to 10^7 load cycles so that the experiment was stopped before the failure could be observed.

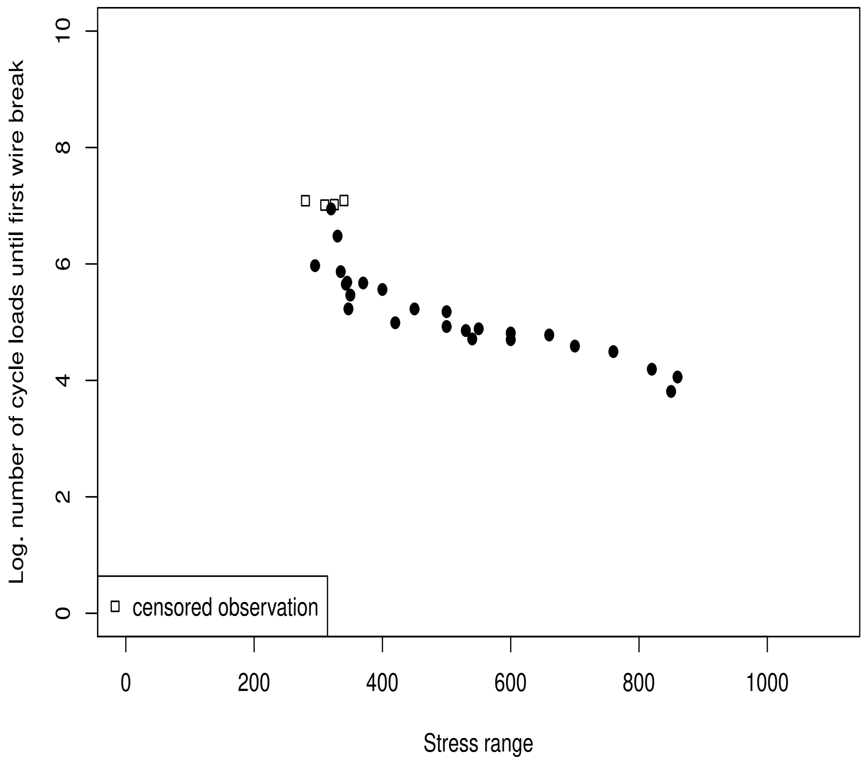

Figure 1.5: Lifetimes of tension wires

1.0.6 Example (Lifetimes of diamonds in a drilling tool)
Figure 1.6 shows the diamonds on the surface of a segment of a drilling tool after two and three minutes of drilling. Two diamonds are breaking out between the second and third minute (red cycles) and two diamonds appear during this time (blue cycles). The appearance of new diamonds is caused by the fact that hidden diamonds are contained in the steel matrix which appear only when some steel has been removed. In this experiment the drilling tool was used for 25 minutes and after every minute the diamonds which are visible are reported. The data set `Diamonds_B28_Matrix.xlsx` contains for each diamond when it is visible. Some diamonds which are visible in the beginning are still visible after 25 minutes when the experiment was stopped. Here we have censored observations.

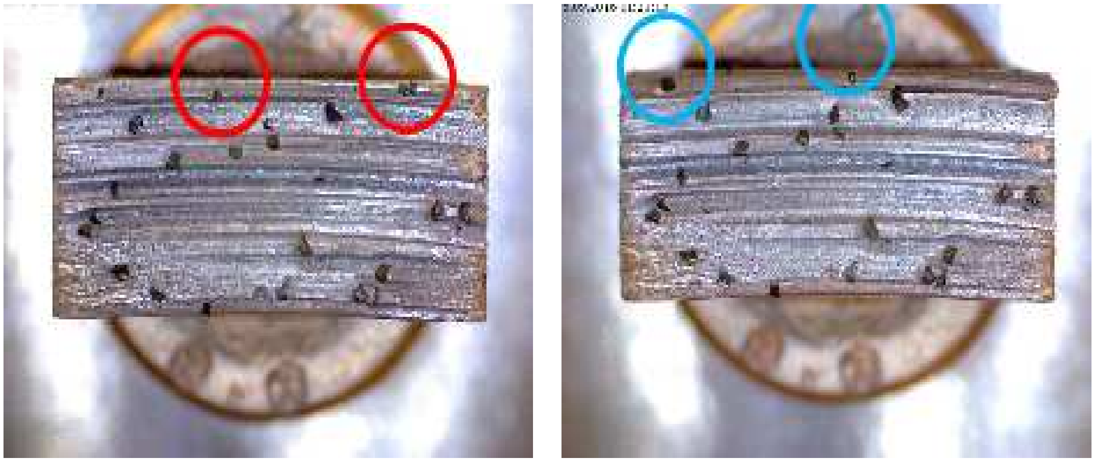

Figure 1.6: Visible diamonds after two minutes (left) and visible diamonds after three minutes (right)

1.0.7 Example (Failures of throttles)
The data set `throttle.csv` provides the distances in km to failure of 50 throttles of load-carrying vehicles. Some of them are shown in Table 1.1. In some cases, denoted by 1, no failure was observed up to the observed distance. Such observations are called censored observations and 1 stands for censoring while 0 denotes no censoring.

	Censored	Failure
1	0	478
2	1	484
3	0	583
4	1	626
5	0	753
6	0	753
⋮	⋮	⋮
47	0	11019
48	0	12986
49	1	13103
50	1	23245

Table 1.1: Distance in km to failure of 50 trottles from Jiang, and Murthy (1995), see also Blischke and Murthy (2000). The column "‘Censored"’ denotes whether a failure was observed (0) or the observation was stopped before failure (1).

1.0.8 Example (Lifetimes of electric lamps)
Table 1.2 provides the lifetime of 300 electric lamps given by the data set `LAMPS.DAT`.

Life time (in hours)	Absolute frequency
950-1000	2
1000-1050	2
1050-1100	3
1100-1150	6
1150-1200	7
1200-1250	12
1250-1300	16
1300-1350	20
1350-1400	24
1400-1450	27
1450-1500	29
1500-1550	29
1550-1600	28
1600-1650	25
1650-1700	21
1700-1750	16
1750-1800	12
1800-1850	8
1850-1900	6
1900-1950	3
1950-2000	2
2000-2050	1
2050-2100	1

Table 1.2: Lifetimes of 300 electric lamps from Gupta (1952) given in Hand et al. (1994), p. 108

Chapter 2

Experiments with one stress level

2.1 Lifetime distributions

Let be $t_1, \ldots, t_N$ observed lifetimes at the same stress level. Such observations are realizations of independent and identically distributed (i.i.d.) random variables $T_1, \ldots, T_N$. Let be $T : \Omega \to \mathbb{R}$ a random variable with the same distribution as that of $T_1, \ldots, T_N$, where $(\Omega, \mathcal{A}, P)$ is the underlying probability space. Important quantities of lifetime distributions are the cumulative distribution function, the reliability or survival function, the hazard function and the cumulative hazard function.

2.1.1 Definition (Cumulative distribution function, reliability or survival function)
a) The cumulative distribution function $F = F_T : \mathbb{R} \to [0, 1]$ is defined by

$$F(t) := P(T \leq t).$$

b) The reliability function or survival function $\overline{F} := S : \mathbb{R} \to [0, 1]$ is defined by

$$\overline{F}(t) := S(t) := P(T > t) = 1 - F(t).$$

2.1.2 Theorem
If T has continuous distribution with cumulative distribution function F and reliability function $\overline{F} = S$ then $F(T) \sim U(0, 1)$ and $S(T) \sim U(0, 1)$ where $U(0, 1)$ is the uniform distribution on $(0, 1)$.

Proof. For any $a \in (0, 1)$, we have

$$\begin{aligned} &P(F(T) \leq a) = P(F(T) < a) + P(F(T) = a) \\ &= \quad P(T < F^{-1}(a)) + P(T \in [F^{-1}(a), \sup\{t;\ F(t) \leq a\}]) = P(T \leq F^{-1}(a)) = F(F^{-1}(a)) = a \end{aligned}$$

which is the distribution function of the uniform distribution on $(0, 1)$. Note that we do not assume that F is strictly increasing. With $F(T)$ also $S(T) = 1 - F(T)$ has a uniform distribution on $(0, 1)$. $\square$

2.1.3 Definition (Hazard function for continous distributions)
The hazard function (hazard rate) $h : \mathbb{R}_+ \to \mathbb{R}$ for continuous distributions is defined by

$$h(t) := \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t}.$$

2.1.4 Remark
The hazard function for a discrete distributions is defined by

$$h(t) := P(T = t | T \ge t).$$

As a time the lifetime T has usually a continuous distribution with density $f = f_T$ which is the derivative of F, i.e. $f = F'$. Of course, the lifetime can be measured by days, weeks, years, load cycles, and then it will be a discrete variable. However, a continuous distribution is a good approximation in these cases as well.

2.1.5 Theorem
If T has a continuous distribution, then

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}.$$

Proof. Since $P(T \ge t) = P(T > t) = S(t)$ for continuous distributions, we obtain

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P(t < T \le t + \Delta t | T \ge t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(t < T \le t + \Delta t)}{P(T \ge t)\, \Delta t} \\ &= \frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{P(t < T \le t + \Delta t)}{\Delta t} = \frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{1}{S(t)}\, f(t). \ \square \end{aligned}$$

2.1.6 Definition (Cumulative hazard function for continuous distributions)
If T has a continuous distribution, then $H : \mathbb{R}_+ \to \mathbb{R}$ given by

$$H(t) := \int_0^t h(s)\, ds$$

is called cumulative hazard function.

2.1.7 Theorem

$$F(t) = 1 - \exp(-H(t)), \quad S(t) = \exp(-H(t)), \quad f(t) = h(t)\, \exp(-H(t)).$$

Proof. Since $f(t) = -S'(t)$, we have with Theorem 2.1.5

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{\partial}{\partial t} \ln(S(t))$$

so that with $S(0) = 1$

$$H(t) = \int_0^t h(s)\,ds = \int_0^t -\frac{S'(s)}{S(s)}\,ds = -\ln(S(s))\big|_0^t = -\ln(S(t)) + \ln(S(0)) = -\ln(S(t))$$

follows. This provides the first assertion. Differentation leads to the second assertion. □

The simplest continuous lifetime distribution is the exponential distribution.

2.1.8 Definition (Exponential distribution)
T has an exponential distribution, shortly $T \sim \mathcal{E}(\lambda)$, if

$$F(t) = 1 - e^{-\lambda t}.$$

2.1.9 Theorem
If $T \sim \mathcal{E}(\lambda)$ then

$$f(t) = \lambda e^{-\lambda t},\ S(t) = e^{-\lambda t},\ h(t) = \lambda,\ H(t) = \lambda\, t.$$

The exponential distribution is the only continuous distribution with constant hazard function. A generalization of the exponential distribution with nonconstant hazard function is the Weibull distribution.

2.1.10 Definition (Weibull distribution)
T has a Weibull distribution, shortly $T \sim \mathcal{W}(\alpha, \beta)$, if

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right].$$

The parameters α and β are called scale and shape parameter, respectively.

The exponential distribution $\mathcal{E}(\lambda)$ is obtained from the Weibull distribution with $\beta = 1$ and $\alpha = \frac{1}{\lambda}$.

2.1.11 Theorem
If $T \sim \mathcal{W}(\alpha, \beta)$ then

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right],$$

$$S(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right],\ \ h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1},\ \ H(t) = \frac{1}{\alpha^{\beta}}\, t^{\beta}.$$

Another generalization of the exponential distribution is the Gamma distribution.

2.1.12 Definition (Gamma distribution)
T has a Gamma distribution, shortly $T \sim \mathcal{G}(\lambda, \beta)$, if

$$F(t) = \frac{1}{\Gamma(\beta)} \int_0^{\lambda t} s^{\beta-1} \, e^{-s} \, ds,$$

where Γ denotes the Gamma function given by

$$\Gamma(x) = \int_0^{\infty} s^{x-1} \, e^{-s} \, ds.$$

The exponential distribution $\mathcal{E}(\lambda)$ is obtained from the Gamma distribution with $\beta = 1$.

2.1.13 Theorem
If $T \sim \mathcal{G}(\lambda, \beta)$ then

$$f(t) = \frac{\lambda^\beta \, t^{\beta-1}}{\Gamma(\beta)} \, e^{-\lambda t}, \quad h(t) = \frac{\frac{\lambda^\beta \, t^{\beta-1}}{\Gamma(\beta)} \, e^{-\lambda t}}{1 - \frac{1}{\Gamma(\beta)} \int_0^{\lambda t} s^{\beta-1} \, e^{-s} \, ds} = \frac{\lambda^\beta \, t^{\beta-1} \, e^{-\lambda t}}{\int_{\lambda t}^{\infty} s^{\beta-1} \, e^{-s} \, ds}.$$

2.1.14 Definition (Erlang distribution)
A Gamma distribution with $\beta = N \in \mathbb{N}$ is called an Erlang distribution.

2.1.15 Theorem
If $T \sim \mathcal{G}(\lambda, N)$ with $N \in \mathbb{N}$ then

$$f(t) = \frac{\lambda^N \, t^{N-1}}{(N-1)!} \, e^{-\lambda t}, \quad F(t) = 1 - \sum_{n=0}^{N-1} \frac{1}{n!} \, e^{-\lambda t} \, (\lambda \, t)^n.$$

Proof. For the form of F see `https://en.wikipedia.org/wiki/Erlang_distribution`. □

2.1.16 Definition (Lognormal distribution)
T has a lognormal distribution, shortly $T \sim \mathcal{LN}(\mu, \sigma^2)$, if $ln(T)$ has a normal distribution with expectation μ and variance σ^2, i.e. $ln(T) \sim \mathcal{N}(\mu, \sigma^2)$.

2.1.17 Theorem
If $T \sim \mathcal{LN}(\mu, \sigma^2)$ then

$$F(t) = \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right), \quad f(t) = \frac{1}{\sqrt{2\pi}\,\sigma\, t} \exp\left[-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right],$$

$$h(t) = \frac{\frac{1}{\sqrt{2\pi}\,\sigma\, t} \exp\left[-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right]}{1-\Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)},$$

where Φ is the cumulative distribution function of the standard normal distribution, i.e.

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}\, ds.$$

Proof. Since the logarithm function is a strictly increasing function we have

$$F(t) = P(T \le t) = P(\ln(T) \le \ln(t)) = \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right).$$

This implies

$$f(t) = F'(t) = \Phi'\left(\frac{\ln(t)-\mu}{\sigma}\right)\frac{1}{t\,\sigma} = \frac{1}{\sqrt{2\pi}\,\sigma\, t} \exp\left[-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right]$$

since $\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. □

2.1.18 Theorem (See e.g. Kahle and Liebscher 2013, S. 36)
If $T \sim \mathcal{LN}(\mu, \sigma^2)$ then

$$E(T) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad \mathrm{var}(T) = \exp(2\mu+\sigma^2)\,(\exp(\sigma^2)-1).$$

Proof. We obtain

$$\begin{aligned}
E(T) = \int_0^\infty t\, f(t)\, dt &= \int_0^\infty t\, \frac{1}{\sqrt{2\pi}\,\sigma\, t} \exp\left[-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right] dt \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right] dt \overset{(\star)}{=} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] e^y\, dt \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[\frac{-y^2+2y\mu-\mu^2+2\sigma^2 y}{2\sigma^2}\right] dt \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[\frac{-y^2+2y(\mu+\sigma^2)-\mu^2-2\mu\sigma^2-\sigma^4+2\mu\sigma^2+\sigma^4}{2\sigma^2}\right] dt \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y^2-[\mu+\sigma^2])^2}{2\sigma^2}\right] \exp\left[\mu+\frac{\sigma^2}{2}\right] dt \\
&= \exp\left[\mu+\frac{\sigma^2}{2}\right] \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y^2-[\mu+\sigma^2])^2}{2\sigma^2}\right] dt \overset{(\star\star)}{=} \exp\left[\mu+\frac{\sigma^2}{2}\right],
\end{aligned}$$

where $(\star\star)$ follows from

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y^2 - [\mu + \sigma^2])^2}{2\sigma^2}\right] \; dt = 1$$

and in $(\star)$ the substitution $y = \ln(t) \Leftrightarrow t = e^y = g(y)$ is used so that $g'(y) = e^y$. $\square$

2.2 Properties of the exponential distribution

The exponential distribution has several interesting properties. The first property means that the exponential distribution is memoryless and that it is the only continuous distribution with this property.

2.2.1 Theorem
a) If $T \sim \mathcal{E}(\lambda)$ then

$$P_\lambda(T > t | T > s) = P_\lambda(T > t - s)$$

for all $t \geq s \geq 0$.
b) If T has continuous distribution with

$$P(T > t | T > s) = P(T > t - s) \tag{2.1}$$

is satisfied for all $t \geq s \geq 0$, then $T \sim \mathcal{E}(\lambda)$ with $\lambda = -\ln(P(T > 1)) > 0$.

Proof.
a) If $T \sim \mathcal{E}(\lambda)$ then

$$P_\lambda(T > t | T > s) = \frac{P_\lambda(T > t)}{P_\lambda(T > s)} = \frac{\exp(-\lambda t)}{\exp(-\lambda s)} = \exp(-\lambda(t - s)) = P_\lambda(T > t - s).$$

b) Let $q = \frac{m}{n}$ an arbitrary rational number with $n, m \in I\!N$. The property (2.1) implies

$$\begin{aligned} P(T > q) &= P\left(T > \frac{m}{n}, T > \frac{1}{n}\right) \\ &= P\left(T > \frac{1}{n}\right) \cdot P\left(T > \frac{m}{n} \middle| T > \frac{1}{n}\right) = P\left(T > \frac{1}{n}\right) \cdot P\left(T > \frac{m}{n} - \frac{1}{n}\right). \end{aligned}$$

Per induction, we get

$$P\left(T > q\right) = P\left(T > \frac{1}{n}\right)^m.$$

In particular, we have

$$P\left(T > 1\right) = P\left(T > \frac{1}{n}\right)^n$$

so that the reliability function satisfies

$$\overline{F}(q) = S(q) = P\left(T > q\right) = P\left(T > 1\right)^{m/n} = P\left(T > 1\right)^q$$

for all rational $q > 0$. Since $\overline{F}$ is monotone decreasing, it holds

$$\overline{F}(t) = S(t) = P\,(T > 1)^t$$

for all $t \geq 0$. In particular, it follows $0 < P\,(T > 1) < 1$. Hence we have

$$\lambda := -\ln(P(T > 1)) > 0$$

and

$$\overline{F}(t) = S(t) = \exp(-\lambda t)$$

for all $t \geq 0$ so that $T \sim \mathcal{E}(\lambda)$. □

2.2.2 Theorem
If T has continuous distribution with cumulative hazard function H then $H(T) \sim \mathcal{E}(1)$.

Proof. Theorem 2.1.7 provides

$$\exp(-H(t)) = 1 - F(t) = \overline{F}(t) = S(t) \Leftrightarrow H(t) = -\ln(S(t))$$

so that

$$P(H(T) > t) = P(-\ln(S(T)) > t) = P(S(T) \leq e^{-t}) = e^{-t}$$

for all $t \geq 0$ since $S(T)$ has a uniform distribution on $(0, 1)$ according to Theorem 2.1.2. □

2.2.3 Theorem
If $T_1, \ldots, T_N \sim \mathcal{E}(\lambda)$ are independent, then

$$a)\ \sum_{n=1}^{N} T_n \sim \mathcal{G}(\lambda, N), \quad b)\ \lambda \sum_{n=1}^{N} T_n \sim \mathcal{G}(1, N), \quad c)\ 2\lambda \sum_{n=1}^{N} T_n \sim \chi^2_{2N}$$

where χ^2_n is the χ^2 distribution with n degrees of freedom.

Proof.
a) At first note that $T \sim \mathcal{G}(\lambda, N)$ implies $f_T(t) = \frac{\lambda^N t^{N-1}}{(N-1)!} e^{-\lambda t} \cdot 1_{[0,\infty)}(t)$. Then $S_N := \sum_{n=1}^{N} T_n \sim \mathcal{G}(\lambda, N)$ follows by induction:
For $N = 1$, we get $f_T(t) = \lambda e^{-\lambda t} = f_{T_1}(t)$.

If $f_{S_N}(t) = \frac{\lambda^N t^{N-1}}{(N-1)!} e^{-\lambda t}$ holds then the convolution theorem provides

$$
\begin{aligned}
&f_{S_{N+1}}(t) = f_{S_N + T_{N+1}}(t) = \int_{-\infty}^{\infty} f_{S_N}(x) f_{T_{N+1}}(t-x) dx \\
&= \int_{-\infty}^{\infty} \frac{\lambda^N x^{N-1}}{(N-1)!} e^{-\lambda x} 1_{[0,\infty)}(x) \lambda e^{-\lambda(t-x)} 1_{[0,\infty)}(t-x) dx \\
&= \frac{\lambda^{N+1}}{(N-1)!} e^{-\lambda t} \int_0^t x^{N-1} dx \, 1_{[0,\infty)}(t) \\
&= \frac{\lambda^{N+1}}{(N-1)!} e^{-\lambda t} \frac{1}{N} t^N \, 1_{[0,\infty)}(t) = \frac{\lambda^{N+1} t^N}{N!} e^{-\lambda t} \, 1_{[0,\infty)}(t).
\end{aligned}
$$

b) From $S_N := \sum_{n=1}^N T_n \sim \mathcal{G}(\lambda, N)$, we get for the distribution function with Definition 2.1.12

$$
\begin{aligned}
&F_{\lambda S_N}(t) = P(\lambda S_N \leq t) = P\left(S_N \leq \frac{t}{\lambda}\right) \\
&= F_{S_N}\left(\frac{t}{\lambda}\right) = \frac{1}{\Gamma(N)} \int_0^{\lambda \frac{t}{\lambda}} s^{N-1} e^{-s} ds = \frac{1}{\Gamma(N)} \int_0^t s^{N-1} e^{-s} ds
\end{aligned}
$$

which means $\lambda \sum_{n=1}^N T_n \sim \mathcal{G}(1, N)$.
c) Similar to b), we obtain

$$
F_{2\lambda S_N}(t) = \frac{1}{\Gamma(N)} \int_0^{\lambda \frac{t}{2\lambda}} s^{N-1} e^{-s} ds = \frac{1}{\Gamma(N)} \int_0^{\frac{t}{2}} s^{N-1} e^{-s} ds =: \frac{1}{\Gamma(N)} G\left(\frac{t}{2}\right)
$$

with $G'(t) = t^{N-1} e^{-t}$. Hence

$$
f_{2\lambda S_N}(t) = F'_{2\lambda S_N}(t) = \frac{1}{\Gamma(N)} \frac{1}{2} G'\left(\frac{t}{2}\right) = \frac{1}{\Gamma(N)} \frac{1}{2} \left(\frac{t}{2}\right)^{N-1} e^{-t/2}
$$

so that $2\lambda \sum_{n=1}^N T_n \sim \chi^2_{2N}$ since the density of the χ^2 distribution with n degrees of freedom which is given by

$$
f(t) = \frac{t^{n/2-1} e^{-t/2}}{2^{n/2} \Gamma\left(\frac{n}{2}\right)}.
$$

2.2.4 Theorem (see e.g. Deshpande, Purohit (2016))
Let $T_1, \ldots, T_N \sim \mathcal{E}(\lambda)$ be independent, $T_{(1)}, \ldots, T_{(N)}$ the corresponding order statistics,

$$Y_n := T_{(n)} - T_{(n-1)}, \;\; n = 2, \ldots, N, \;\; Y_1 := T_{(1)},$$

the consecutive sample spacings,

$$D_n := (N - n + 1) Y_n, \;\; n = 1, \ldots, N,$$

the normalised sample spacings, then
a) $Y_1, \ldots, Y_N$ are independent with $Y_n \sim \mathcal{E}((N - n + 1)\lambda)$ for $n = 1, \ldots, N$,
b) $D_1, \ldots, D_N \sim \mathcal{E}(\lambda)$ are independent.

Proof. The joint density of $T_* := (T_{(1)}, \ldots, T_{(N)})$ is

$$\begin{aligned} f_{T_*}(t_*) &= f_{T_{(1)}, T_{(2)}, \ldots, T_{(N)}}(t_1, t_2, \ldots, t_N) \\ &= \begin{cases} N! \cdot \lambda e^{-\lambda t_1} \cdot \lambda e^{-\lambda t_2} \cdot \ldots \cdot \lambda e^{-\lambda t_N} & \text{for } 0 < t_1 < t_2 < \ldots < t_N < \infty, \\ 0 & \text{else.} \end{cases} \end{aligned}$$

Consider the transformation

$$g : \{(t_1, t_2, \ldots, t_N);\; 0 < t_1 < t_2 < \ldots < t_N < \infty\} \longrightarrow \{(y_1, y_2, \ldots, y_N);\; y_n > 0,\; n = 1, 2, \ldots, N\}$$

given by

$$g : \begin{cases} y_1 = t_1, \\ y_2 = t_2 - t_1, \\ \vdots \\ y_N = t_N - t_{N-1}, \end{cases} \qquad g^{-1} : \begin{cases} t_1 = y_1, \\ t_2 = y_1 + y_2, \\ \vdots \\ t_N = y_1 + \ldots + y_N. \end{cases}$$

g is a 1-1 transformation with the determinant of the Jacobian matrix given by

$$\det(g'(t_*)) = \det\left(\frac{\partial}{\partial(t_1, \ldots, t_N)} g(t_1, \ldots, t_N)\right) = \det \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ 0 & 0 & -1 & & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix} = 1.$$

Then $Y = (Y_1, \ldots, Y_N) = g(T_*)$ has the density

$$\begin{aligned}
& f_Y(y) = f_{(Y_1,\ldots,Y_N)}(y_1, \ldots, y_N) = f_{T_*}(g^{-1}(y))/|\det(g'(g^{-1}(y))| \\
= \; & N! \cdot \lambda e^{-\lambda y_1} \cdot \lambda e^{-\lambda(y_1+y_2)} \cdot \ldots \cdot \lambda e^{-\lambda(y_1+\ldots+y_N)} \\
= \; & N \cdot \lambda e^{-\lambda N\, y_1} \cdot (N-1)\lambda e^{-\lambda(N-1)y_2} \cdot \ldots \cdot \lambda e^{-\lambda y_N} \\
= \; & \prod_{n=1}^{N} (N-n+1)\lambda e^{-(N-n+1)\lambda\, y_n} = \prod_{n=1}^{N} f_{Y_n}(y_n).
\end{aligned}$$

Hence $Y_1, \ldots, Y_N$ are independent with $Y_n \sim \mathcal{E}((N-n+1)\lambda)$ for $n = 1, \ldots, N$ and $D_1, \ldots, D_N \sim \mathcal{E}(\lambda)$ are independent. $\square$

2.3 Classes of distributions with ageing properties

2.3.1 Definition (IFR distribution)
T with continuous distribution has a IFR distribution (Increasing Failure Rate) if

$$S_{t_1}(s) \geq S_{t_2}(s)$$

is satisfies for all $s > 0$, $t_1 \leq t_2$, where

$$S_t(s) := \frac{S(t+s)}{S(t)}.$$

2.3.2 Remark
If T has a continuous distribution then

$$\frac{S(t+s)}{S(t)} = \frac{P(T > t+s)}{P(T > t)} = \frac{P(T > t+s)}{P(T \geq t)} = P(T > t+s | T \geq t).$$

If T has not a continuous distribution, then S_t is defined as

$$S_t(s) := \frac{S(t+s)}{\lim_{h \downarrow 0} S(t-h)} = P(T > t+s | T \geq t),$$

so that a discrete IFR distribution is defined via this conditional distribution, see Kahle and Liebscher (2013), p. 43.

2.3.3 Theorem
T has a continuous IFR distribution if and only if its hazard function h is increasing.

Proof. Theorem 2.1.7 provides

$$S_t(s) := \frac{S(t+s)}{S(t)} = \frac{\exp\left(-\int_0^{t+s} h(u)du\right)}{\exp\left(-\int_0^{t} h(u)du\right)} = \exp\left(-\int_t^{t+s} h(u)du\right).$$

As soon as h is increasing, we get for all $t_1 \leq t_2$

$$S_{t_1}(s) = \exp\left(-\int_{t_1}^{t_1+s} h(u)du\right) \geq \exp\left(-\int_{t_2}^{t_2+s} h(u)du\right) = S_{t_2}(s)$$

for all $s > 0$. □

2.3.4 Definition (AIFR distribution)
T has a AIFR distribution (Increasing Failure Rate in Average) if g given by

$$g(t) := S(t)^{1/t}$$

is a decreasing function on $\mathbb{R}_+$.

2.3.5 Remark
If $S(t)^{1/t}$ is decreasing then also

$$\ln\left(S(t)^{1/t}\right) = \frac{\ln(S(t))}{t} = \frac{\ln(\exp(-H(t)))}{t} = \frac{-H(t)}{t} \tag{2.2}$$

is decreasing and $\frac{-H(t)}{t}$ can be interpreted as average cumulative failure rate.

2.3.6 Theorem
If a distribution is IFR then it is AIFR. The converse does not hold in general.

Proof. If a distribution is IFR then its hazard function h is increasing according to Theorem 2.3.3 so that

$$H(t) = \int_0^t h(u)\,du \le h(t)\,t.$$

Hence with (2.2)

$$\frac{\partial}{\partial t}\ln\left(S(t)^{1/t}\right) = \frac{\partial}{\partial t}\frac{-H(t)}{t} = \frac{-h(t)}{t} + \frac{H(t)}{t^2} \le \frac{-h(t)}{t} + \frac{h(t)\,t}{t^2} = 0$$

so that with $\ln\left(S(t)^{1/t}\right)$ also $S(t)^{1/t}$ is decreasing.
To show that the converse does not hold in general, consider the hazard function h of the counterexample given by Kahle and Liebscher 2013, p. 46, which is defined as

$$h(t) = \begin{cases} 0, & t < 0.5, \\ 2, & 0.5 \le t < 1, \\ 1, & t \ge 1. \end{cases}$$

Since h is not increasing, the corresponding distribution is not IFR. However, this distribution is AIFR since

$$H(t) = \begin{cases} 0, & t < 0.5, \\ 2t - 1, & 0.5 \le t < 1, \\ t, & t \ge 1, \end{cases}$$

and

$$\frac{H(t)}{t} = \begin{cases} 0, & t < 0.5, \\ 2 - \frac{1}{t}, & 0.5 \leq t < 1, \\ 1, & t \geq 1. \end{cases}$$

$\frac{H(t)}{t}$ is obviously increasing so that with $\frac{-H(t)}{t} = \ln\left(S(t)^{1/t}\right)$ also $S(t)^{1/t}$ is decreasing. □

2.3.7 Definition (NBU distribution)
T has a NBU distribution (New Better than Used) if

$$S(t+s) \leq S(t) \cdot S(s)$$

for all $t, s > 0$.

2.3.8 Remark
If T has a continuous NBU distribution and $S(t) > 0$ then

$$S_t(s) := \frac{S(t+s)}{S(t)} \leq S(s), \tag{2.3}$$

i.e. the conditional survival probability given a survival of an individual used up to t is not larger then the unconditional survival probability of a new individual without usage, i.e. new better than used.
Moreover, a distribution is NBU if and only if its cumulative hazard function H satisfies

$$-H(t+s) = \ln(S(t+s)) \leq \ln(S(t)) + \ln(S(s)) = -H(t) - H(s) \tag{2.4}$$

or $H(t+s) \geq H(t) + H(s)$, respectively.
Moreover, note that equality in (2.3) means that the distribution is memoryless, where the exponential distribution is the only continuous distribution which is memoryless.

2.3.9 Theorem
If a distribution is AIFR then it is NBU. The converse does not hold in general.

Proof. If a distribution is AIFR then

$$S(t+s)^{1/(t+s)} \leq S(t)^{1/t}, \quad S(t+s)^{1/(t+s)} \leq S(s)^{1/s}$$

implying

$$S(t+s)^{t/(t+s)} \leq S(t), \quad S(t+s)^{s/(t+s)} \leq S(s).$$

Multiplying both inequalities yields

$$S(t+s) = S(t+s)^{t/(t+s)} \cdot S(t+s)^{s/(t+s)} \leq S(t) \cdot S(s),$$

which means that the distribution is NBU.
To show that the converse does not hold in general, consider the hazard function h of the counterexample given by Kahle and Liebscher 2013, p. 47, which is defined as

$$h(t) = \begin{cases} 0, & t < 1, \\ 1, & 1 \leq t < 1.5, \\ 0, & 1.5 \leq t < 2, \\ 1, & t \geq 2. \end{cases} \tag{2.5}$$

That this hazard function indeed defines an NBU distribution which is not an AIFR distribution is an exercise. □

2.3.1 Example (Weibull distribution)
If $T \sim \mathcal{W}(\alpha, \beta)$ then the hazard function h is according to Theorem 2.1.11 given by

$$h(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}.$$

This function is strictly increasing for $\beta > 1$, constant for $\beta = 1$ and decreasing for $\beta < 1$. Hence the Weibull distribution is IFR and thus AIFR and NBU for $\beta \geq 1$. In particular, the exponential distribution is IFR, AIFR, and NBU. A Weibull distribution is not AIFR and thus not IFR for $\beta < 1$ since

$$S(t)^{1/t} = \left(\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]\right)^{1/t} = \exp\left[-\frac{t^{\beta-1}}{\alpha^{\beta}}\right]$$

is increasing in t for $\beta < 1$. For $\beta < 1$, the Weibull distribution is also not NBU, since the NBU property is according to (2.4) equivalent to

$$\frac{1}{\alpha^{\beta}}(t+s)^{\beta} = H(t+s) \geq H(t) + H(s) = \frac{1}{\alpha^{\beta}} t^{\beta} + \frac{1}{\alpha^{\beta}} s^{\beta}$$

for all $s, t \geq 0$. This would mean in particular for $s = t$

$$(2t)^{\beta} \geq 2\, t^{\beta} \quad \Leftrightarrow \quad 2^{\beta} \geq 2 \quad \Leftrightarrow \quad \beta \geq 1$$

so that the Weibull distribution with $\beta < 1$ cannot be NBU.

2.4 Censored lifetimes

Usually lifetime experiments in engineering are very time consuming and costly so that not every object can be observed until failure.

2.4.1 Definition (Type I censoring)
The objects are only observed until a predefined fixed time c, so that instead of the real lifetimes $t_1, \ldots, t_N$ only

$$z_n = \min(t_n, c), \quad d_n = \mathbb{1}\{t_n \le c\} = \begin{cases} 1, & \text{if failure of the } n\text{'th object is observed, i.e. } t_n \le c, \\ 0, & \text{if no failure of the } n\text{'th object is observed, i.e. } t_n > c, \end{cases}$$

is observed for $n = 1, \ldots, N$. Thereby, $\mathbb{1}\{t_n \le c\} := \mathbb{1}_{[0,c]}(t_n)$ denotes the indicator function.

2.4.2 Definition (Progressive Type I censoring)
There are predefined fixed progressive censoring times $c_1 < c_2 < \ldots < c_K$ and given samples sizes $0 = N_0 < N_1 < N_2 < \ldots < N_K = N$ so that objects with $n \in \{N_{k-1}+1, \ldots, N_k\}$ are only observed until c_k, for $k = 1, \ldots, K$. This means that instead of the real lifetimes $t_1, \ldots, t_N$ only

$$z_n = \min(t_n, c_k), \quad d_n = \mathbb{1}\{t_n \le c_k\} = \begin{cases} 1, & \text{if } t_n \le c_k, \\ 0, & \text{if } t_n > c_k, \end{cases}$$

is observed for $n \in \{N_{k-1}+1, \ldots, N_k\}$ for $k = 1, \ldots, K$.

If the lifetime experiments are running parallel then also Type II censoring can be used.

2.4.3 Definition (Type II censoring)
The objects are only observed until a predefined fixed number R of failures are observed. If $t_{(1)} \le t_{(2)} \le \ldots \le t_{(N)}$ are the ordered real lifetimes then only

$$z_n = \min(t_n, t_{(R)}), \quad d_n = \mathbb{1}\{t_n \le t_{(R)}\} = \begin{cases} 1, & \text{if } t_n \le t_{(R)}, \\ 0, & \text{if } t_n > t_{(R)}, \end{cases}$$

is observed for $n = 1, \ldots, N$. This means that $N - R$ observations are censored.

The corresponding random variables are always denoted by $Z_1, \ldots, Z_N$ and $D_1, \ldots, D_N$.

2.4.4 Definition (Progressive Type II censoring)
Here integers $R_1, \ldots, R_K$ and $N_1, \ldots, N_K$ are predefined such that $R_1 + \ldots + R_K + N_1 + \ldots + N_K \le N$. At first the objects are observed until R_1 failures are noted. From the remaining $N - R_1$ objects, N_1 objects are removed so that observations of these objects are censored by $t_{(R_1)}$. Then the remaining $N - R_1 - N_1$ objects are observed until they provide R_2 failures. From the remaining $N - R_1 - N_1 - R_2$ objects, N_2 objects are removed. This is repeated until $N - (R_1 + \ldots + R_K + N_1 + \ldots + N_K) \ge 0$ objects remain and provide censored observations as well. At all $R_1 + \ldots + R_K$ failures are observed and $N - (R_1 + \ldots + R_K)$ observations are censored.

2.4.5 Definition (Random censoring)
The objects are only observed until different censoring constants $c_1, \ldots, c_N$ which are realizations of i.i.d. random variables $C_1, \ldots, C_N$. This means that instead of the real lifetimes $t_1, \ldots, t_N$ only

$$z_n = \min(t_n, c_n), \quad d_n = 1\!\!1\{t_n \leq c_n\} = \begin{cases} 1, & \text{if } t_n \leq c_n, \\ 0, & \text{if } t_n > c_n, \end{cases}$$

is observed for $n = 1, \ldots, N$.

Note that Type I censoring is a special case of random censoring where the distribution of $C_1, \ldots, C_N$ is given by a one-point measure on c.

Sometimes the life times of N objects can be only observed at predefined fixed time points $0 = \tau_0 < \tau_1 < \ldots < \tau_I < \tau_{I+1} = \infty$. Then we have interval censored data.

2.4.6 Definition (Interval censoring)
The objects are only observed at predefined fixed time points $0 = \tau_0 < \tau_1 < \ldots < \tau_I < \tau_{I+1} = \infty$. This means that instead of the real life times $t_1, \ldots, t_N$ only

$$z_n = \begin{cases} i, & \text{if } t_n \in (\tau_{i-1}, \tau_i] \text{ for } i = 1, \ldots, I, \\ I+1, & \text{if } t_n \in (\tau_I, \infty), \end{cases}$$

is observed for $n \in 1, \ldots, N$.

Often objects are not entering at the same time in a study. We will denote the time point when the nth object has entered the study by b_n (b for beginning or birth time) and assume that $b_1, \ldots, b_N$ are realizations of random variables $B_1, \ldots, B_N$. If we can observe $b_1, \ldots, b_N$ then $t_n = \tilde{t}_n - b_n$ are the life times where $\tilde{t}_n$, $n = 1, \ldots, N$, are the observed failure times. If the study has a predefined fixed end point τ then we have randomly censored data with $c_n = \tau - b_n$. However, the situation becomes more complicated if the birth times and life times can be only observed at predefined fixed time points $0 = \tau_0 < \tau_1 < \ldots < \tau_I < \tau_{I+1} = \infty$.

2.4.7 Example (Diamonds pull outs)
The data set `Diamonds_B28_Matrix.xlsx` shows the appearance of 46 diamonds on a segment of a drilling tool. At the beginning of the experiments, only 22 diamonds are visible on the surface of the segment. During the drilling process, diamonds are pulled out but also new diamonds appear which are lying in deeper layers of the metal matrix. However, this can be only observed at predefined time points because the drilling process has to be interrupted for this. In this experiment, the process was interrupted every 60 seconds and this was done 25 times. If one would consider only the 22 diamonds which are visible at the beginning of the experiments then the data are interval censored data in the sense of Definition 2.4.6. To include the observations concerning the remaining 44 diamonds then Definition 2.4.6 must be extended. Since the time of the first appearance of a diamond can be interpreted as a birth of the diamond and the pull out time of a diamond as a death of a diamond, we call this type of censored data *interval censored births and deaths data*.

2.4.8 Definition (Doubly interval censoring)
The births (first appearance) and deaths (life time) are only observed at predefined fixed time points $0 = \tau_0 < \tau_1 < \ldots < \tau_I < \tau_{I+1} = \infty$. *This means that instead of the real lifetimes* $t_1, \ldots, t_N$ *only*

$$z_n = \begin{cases} (0,i) & \text{if } b_n = 0,\ t_n \in (\tau_{i-1}, \tau_i] \text{ for } i = 1, \ldots, I, \\ (h,i) & \text{if } b_n \in (\tau_{h-1}, \tau_h],\ t_n + b_n \in (\tau_{i-1}, \tau_i] \text{ for } h, i = 1, \ldots, I,\ h < i, \\ (h, I+1), & \text{if } b_n \in (\tau_{h-1}, \tau_h],\ t_n + b_n \in (\tau_I, \infty) \text{ for } h = 1, \ldots, I, \\ (0, I+1), & \text{if } b_n = 0,\ t_n \in (\tau_I, \infty), \end{cases}$$

is observed for $n \in 1, \ldots, N$.

2.5 Estimation for parametric lifetime distributions

Let be $\theta \in \Theta$ the unknown parameter of the lifetime distribution. Then P_θ, f_θ, F_θ, and S_θ denote the corresponding probability measure, the corresponding density function, the corresponding cumulative distribution function and the corresponding survival function, respectively.

2.5.1 Definition (Maximum likelihood estimate for uncensored observations)
$\widehat{\theta} = \widehat{\theta}(t_1, \ldots, t_N)$ *is called maximum likelihood estimate for* θ *if*

$$\widehat{\theta} \in \arg\max \prod_{n=1}^{N} f_\theta(t_n).$$

For censoring, we consider at first progressive Type I censoring which includes Type I censoring as a special case. Then we get the following likelihood function (see e.g. Klein and Moeschberger 2003)

$$l(\theta) := \prod_{n=1}^{N} f_\theta(z_n)^{d_n} S_\theta(z_n)^{1-d_n}.$$

This is also the likelihood function for random censoring if the censoring variable $C_1, \ldots, C_N$ and the life time variables $T_1, \ldots, T_N$ are independent and the distributions of $C_1, \ldots, C_N$ do not depend on θ.

2.5.2 Definition (Maximum likelihood estimate for progressive Type I and random censored observations)
$\widehat{\theta} = \widehat{\theta}((z_1, d_1), \ldots, (z_N, d_N))$ *is called maximum likelihood estimate for* θ *based on progressive Type I and random censored observations if*

$$\widehat{\theta} \in \arg\max \prod_{n=1}^{N} f_\theta(z_n)^{d_n} S_\theta(z_n)^{1-d_n}.$$

To derive the maximum likelihood estimate for this type of censored observations $(z_1, d_1), \ldots,$ (z_N, d_N) we can regard the loglikelihood function

$$\begin{aligned} L(\theta) &:= \ln\left(\prod_{n=1}^{N} f_\theta(z_n)^{d_n} S_\theta(z_n)^{1-d_n}\right) = \sum_{n=1}^{N} \ln(f_\theta(z_n)^{d_n}) + \sum_{n=1}^{N} \ln(S_\theta(z_n)^{1-d_n}) \\ &= \sum_{n=1}^{N} d_n \ln(f_\theta(z_n)) + \sum_{n=1}^{N} (1-d_n) \ln(S_\theta(z_n)). \end{aligned}$$

If $\frac{\partial}{\partial\theta} L(\theta)\Big|_{\theta=\tilde{\theta}} = 0$ and $\frac{\partial^2}{\partial^2\theta} L(\theta)\Big|_{\theta=\tilde{\theta}}$ is negative definite then $\tilde{\theta} = \widehat{\theta}$ is the maximum likelihood estimate.

2.5.3 Example (Exponential distribution with progressive Type I or random censoring)
The loglikelihood function is here with $z_n = \min(t_n, c_n)$ and $J := \sum_{n=1}^N d_n$

$$L(\lambda) = \sum_{n=1}^N d_n \ln(\lambda e^{-\lambda z_n}) + \sum_{n=1}^N (1-d_n)\ln(e^{-\lambda z_n})$$
$$= \quad J\,\ln(\lambda) - \lambda \sum_{n=1}^N z_n.$$

Then we have

$$L'(\lambda) = J\,\frac{1}{\lambda} - \sum_{n=1}^N z_n = 0$$

if and only if $\frac{1}{\lambda} = \frac{1}{J}\sum_{n=1}^N z_n$ so that the maximum likelihood estimator for λ is

$$\widehat{\lambda} = \frac{J}{\sum_{n=1}^N z_n}.$$

For Type I censoring with $c = c_n$, $n = 1, \dots, N$, this simplifies to

$$\widehat{\lambda} = \frac{J}{(N-J)\,c + \sum_{n=1}^N d_n t_n}.$$

If there are no censored observations then $J = N$ and

$$\widehat{\lambda} = \frac{N}{\sum_{n=1}^N t_n}.$$

The likelihood function for Type II censoring is given for the ordered sample $t_{(1)}, \dots, t_{(N)}$ by (see e.g. Deshpande, Purohit 2016)

$$l(\theta) = \binom{N}{R} R! \prod_{n=1}^R f_\theta(t_{(n)}) S_\theta(t_{(R)})^{N-R} = \binom{N}{R} R! \prod_{n=1}^N f_\theta(z_n)^{d_n} S_\theta(z_n)^{1-d_n}.$$

2.5.4 Definition (Maximum likelihood estimate for Type II censored observations)
$\widehat{\theta} = \widehat{\theta}((z_1, d_1), \dots, (z_N, d_N))$ is called maximum likelihood estimate for θ based on Type II censored observations if

$$\widehat{\theta} \in \arg\max \prod_{n=1}^N f_\theta(z_n)^{d_n} S_\theta(z_n)^{1-d_n}.$$

2.5.5 Example (Exponential distribution with Type II censoring)
Since the likelihood function is the same as for progressive Type I censoring with $R = \sum_{n=1}^N d_n$, the maximum likelihood estimator is given by

$$\widehat{\lambda} = \frac{R}{\sum_{n=1}^N z_n} = \frac{R}{(N-R)\, t_{(R)} + \sum_{n=1}^R t_{(n)}}.$$

2.5.6 Theorem (Likelihood function for interval censored data)
If $T_1, \ldots, T_N$ are independent with cumulative distribution function F_θ, then the likelihood function for interval censored data $z_1, \ldots, z_N$ is given by

$$l(\theta) = \prod_{n=1}^N \left(\prod_{i=1}^I \left(F_\theta(\tau_i) - F_\theta(\tau_{i-1})\right)^{\mathbb{1}\{z_n=i\}} \left(1 - F_\theta(\tau_I)\right)^{\mathbb{1}\{z_n=I+1\}} \right),$$

where $\mathbb{1}$ denotes the indicator function.

2.5.7 Theorem (Likelihood function for doubly interval censored birth and death data)
If $B_1, \ldots, B_N, T_1, \ldots, T_N$ are independent so that B_n has cumulative distribution function G_θ and T_n has cumulative distribution function F_θ for $n = 1, \ldots, N$, then the likelihood function for doubly interval censored birth and death data $z_1, \ldots, z_N$ satisfies

$$l(\theta) = \prod_{n=1}^N \prod_{i=1}^{I+1} \prod_{\substack{h=1\\h<i}}^{I+1} \left(\int_{\tau_{h-1}}^{\tau_h} \left(F_\theta(\tau_i - u) - F_\theta(\tau_{i-1} - u)\right) dG_\theta(u) \right)^{\mathbb{1}\{z_n=(h,i)\}}. \tag{2.6}$$

Proof. Set $D_n = B_n + T_n$ for the "death time". Then we have

$$\begin{aligned} l(\theta) &:= l(\theta; z_1, \ldots, z_N) = \prod_{n=1}^N P_\theta(Z_n = z_n) = \prod_{n=1}^N \prod_{\substack{h,i=0\\h<i}}^{I+1} P_\theta\big(Z_n = (h,i)\big) \\ &= \prod_{n=1}^N \prod_{i=1}^{I+1} \prod_{\substack{h=1\\h<i}}^{I+1} P_\theta\big(B_n \in (\tau_{h-1}, \tau_h],\, D_n \in (\tau_{i-1}, \tau_i]\big)^{\mathbb{1}\{z_n=(h,i)\}}. \end{aligned}$$

Since $T_n := D_n - B_n$, $n = 1, \ldots, N$, we can rewrite

$$\begin{aligned} &P_\theta\big(B_n \in (\tau_{h-1}, \tau_h],\, D_n \in (\tau_{i-1}, \tau_i]\big) = P_\theta\big(B_n \in (\tau_{h-1}, \tau_h],\, B_n + T_n \in (\tau_{i-1}, \tau_i]\big) \\ &\quad = \iint_{\mathbb{R}^2} \mathbb{1}_{(\tau_{h-1}, \tau_h]}(y_1)\, \mathbb{1}_{(\tau_{i-1}, \tau_i]}(y_2)\, dP_\theta^{(B_n, B_n+T_n)}(y_1, y_2). \end{aligned}$$

Using the elementary transformation theorem from the measure theory and then the independence of B_n and T_n, we obtain

$$
\begin{aligned}
&\iint_{\mathbb{R}^2} \mathbb{1}_{(\tau_{h-1},\tau_h]}(y_1)\ \mathbb{1}_{(\tau_{i-1},\tau_i]}(y_2)\ dP_\theta^{(B_n,B_n+T_n)}(y_1,y_2) \\
&\quad = \iint_{\mathbb{R}^2} \mathbb{1}_{(\tau_{h-1},\tau_h]}(u)\ \mathbb{1}_{(\tau_{i-1},\tau_i]}(v+u)\ dP_\theta^{(B_n,T_n)}(u,v) \\
&\quad = \iint_{\mathbb{R}^2} \mathbb{1}_{(\tau_{h-1},\tau_h]}(u)\ \mathbb{1}_{(\tau_{i-1},\tau_i]}(v+u)\ dP_\theta^{T_n}(v)\ dP_\theta^{B_n}(u) \\
&\quad = \int_{\tau_{h-1}}^{\tau_h} \int_{\tau_{i-1}-u}^{\tau_i-u} dF_\theta(v)\ dG_\theta(u) = \int_{\tau_{h-1}}^{\tau_h} \left(F_\theta(\tau_i-u) - F_\theta(\tau_{i-1}-u)\right) dG_\theta(u).
\end{aligned}
$$

This implies the assertion. □

2.6 Goodness-of-fit tests

Let be $\mathcal{F} = \{F_\theta; \theta \in \Theta\}$ a family of given cumulative distribution functions for $\theta \in \Theta \subset \mathbb{R}^r$ and F the true cumulative distribution function, then we want to test

$$H_0 : F \in \mathcal{F}, \quad H_1 : F \notin \mathcal{F}$$

If we have an estimate $\widehat{\theta}$ of θ, we can use $F_{\widehat{\theta}}$ for a test. There are several goodnes-of-fit tests based on $F_{\widehat{\theta}}$ to test the null hypothesis. The most flexible goodness-of-fit test is the χ^2 goodness-of-fit test.

2.6.1 Definition (χ^2 goodness-of-fit test, see e.g. Schervish 1997, Theorem 7.133)
Let be $I_1, I_2, \ldots, I_L$ disjoint intervals such that $\bigcup_{l=1}^L I_l = [0, \infty)$,

$$I_1 := [a_1, b_1] \text{ with } a_1 = 0,\ I_l := (a_l, b_l] \text{ for } l = 2, \ldots, L-1,\ I_L := (a_L, b_L) \text{ with } b_l = \infty,$$

$p_l := F_{\widehat{\theta}}(b_l) - F_{\widehat{\theta}}(a_l)$, $N_l := \sharp\{n;\ t_n \in I_l\}$,

$$T_{\chi^2} := T_{\chi^2}(t_1, \ldots, t_N) := \sum_{l=1}^{L} \frac{(N_l - N\,p_l)^2}{N\,p_l},$$

and $\chi^2_{L-r-1;1-\alpha}$ the $(1-\alpha)$-quantile of the χ^2 distribution with $L-r-1$ degrees of freedom, then the χ^2 goodness-of-fit test is given by the decision rule

$$\text{reject } H_0 : F \in \mathcal{F} \quad \text{if} \quad T_{\chi^2} > \chi^2_{L-r-1;1-\alpha}.$$

This is an approximate α-level test for $H_0 : F \in \mathcal{F}$ if $N\ p_l \geq 5$ holds for all $l = 1, \ldots, L$. Hence the a_l and b_l must be chosen so that this is satisfied.

2.6.2 Example
A natural choice for interval censored data is $I_1 = (0, \tau_1]$, $I_2 = (\tau_1, \tau_2], \ldots I_I = (\tau_{I-1}, \tau_I]$, $I_{I+1} = (\tau_I, \infty)$ so that $L = I + 1$. If $N\ p_l \geq 5$ is not satisfied then classes should be combined.

2.6.3 Lemma
If $F_{\widehat{\theta}}$ is continuous and strictly increasing and $k \in \mathbb{N}$ then

$$a_l = F_{\widehat{\theta}}^{-1}\left(\frac{(l-1)k}{N}\right),\quad b_l = F_{\widehat{\theta}}^{-1}\left(\frac{lk}{N}\right) \quad \text{for } l = 1, \ldots, \left\lfloor \frac{N}{k} \right\rfloor$$

satifies

$$N\,\left(F_{\widehat{\theta}}(b_l) - F_{\widehat{\theta}}(a_l)\right) = k \quad \text{for } l = 1, \ldots, \left\lfloor \frac{N}{k} \right\rfloor.$$

Proof. Since $F_{\widehat{\theta}}(F_{\widehat{\theta}}^{-1}(x)) = x$ for all $x \in \mathbb{R}$, it holds

$$F_{\widehat{\theta}}(b_l) - F_{\widehat{\theta}}(a_l) = \frac{lk}{N} - \frac{(l-1)k}{N} = \frac{k}{N}$$

for all $l = 1, \ldots, \lfloor \frac{N}{k} \rfloor$. □

2.6.4 Corollary

If $k \in I\!N$ *satisfies* $k \geq 5$, $L = \left\lfloor \frac{N}{k} \right\rfloor$, $N_1 = \sharp \left\{ n,\ F_{\widehat{\theta}}(t_n) \in \left[0, \frac{k}{N}\right] \right\}$, $N_l = \sharp \left\{ n,\ F_{\widehat{\theta}}(t_n) \in \left(\frac{(l-1)k}{N}, \frac{lk}{N} \right] \right\}$ *for* $l = 2, \ldots, L-1$, $N_L = \sharp \left\{ n,\ F_{\widehat{\theta}}(t_n) \in \left(\frac{(L-1)k}{N}, \infty \right) \right\}$,

$$T_{\chi^2} := T_{\chi^2}(t_1, \ldots, t_N) := \sum_{l=1}^{L} \frac{(N_l - k)^2}{k},$$

then the decision rule

$$\text{reject } H_0 : F \in \mathcal{F} \quad \text{if} \quad T_{\chi^2} > \chi^2_{L-r-1;1-\alpha}.$$

is an approximate α*-level test for* $H_0 : F \in \mathcal{F}$.

Proof. The assertion follows with

$$F_{\widehat{\theta}}(t_n) \in \left(\frac{(l-1)k}{N}, \frac{lk}{N} \right] \Longleftrightarrow t_n \in \left(F_{\widehat{\theta}}^{-1}\left(\frac{(l-1)k}{N} \right), F_{\widehat{\theta}}^{-1}\left(\frac{lk}{N} \right) \right]$$

and Lemma 2.6.3. $\square$

2.7 Confidence sets and prediction intervals

Let be $\mathcal{P}(\mathbb{R}^r)$ the set of all subsets of $\mathbb{R}^r$. The following definitions, lemmas and theorems are given for uncensored data $t_1, \ldots, t_N$. However, they hold also for censored data $z_1, \ldots, z_N$ and in more general situations so that we provide the general concepts for data of the form $z_1, \ldots, z_N$. We assume that $z_1, \ldots, z_N$ are realizations of $Z_1, \ldots, Z_N$.

2.7.1 Definition (Confidence set)
$\mathbb{C} : [0, \infty)^N \ni (z_1, \ldots, z_N) \to \mathbb{C}(z_1, \ldots, z_N) \in \mathcal{P}(\mathbb{R}^s)$ *is a* $(1-\alpha)$*-confidence set function for the aspect* $a(\theta) \in \mathbb{R}^s$ *of* θ *if*

$$P_\theta(a(\theta) \in \mathbb{C}(Z_1, \ldots, Z_N)) \geq 1 - \alpha$$

is satisfied for all $\theta \in \Theta$.

2.7.2 Lemma
If $\theta = (\theta_1, \ldots, \theta_r)^\top \in \mathbb{R}^r$ *and* $\mathbb{C}_i : [0, \infty)^N \ni (z_1, \ldots, z_N) \to \mathbb{C}_i(z_1, \ldots, z_N) \in \mathcal{P}(\mathbb{R})$ *are* $(1-\alpha)$*-confidence set functions for* θ_i *for* $i = 1, \ldots, r$ *then* $\mathbb{C} : [0, \infty)^N \ni (z_1, \ldots, z_N) \to \mathbb{C}(z_1, \ldots, z_N) \in \mathcal{P}(\mathbb{R}^r)$ *given by*

$$\mathbb{C}(z_1, \ldots, z_N) = \mathbb{C}_1(z_1, \ldots, z_N) \times \mathbb{C}_2(z_1, \ldots, z_N) \times \ldots \times \mathbb{C}_r(z_1, \ldots, z_N)$$

is a $(1 - r\alpha)$*-confidence set function for* θ.

Proof. Since $P_\theta(\theta_i \notin \mathbb{C}_i(Z_1, \ldots, Z_N)) = 1 - P_\theta(\theta_i \in \mathbb{C}_i(Z_1, \ldots, Z_N)) \leq 1 - (1-\alpha) = \alpha$, we obtain

$$\begin{aligned} & P_\theta(\theta \in \mathbb{C}(Z_1, \ldots, Z_N)) = P_\theta(\theta_1 \in \mathbb{C}_1(Z_1, \ldots, Z_N) \text{ and } \ldots \text{ and } \theta_r \in \mathbb{C}_r(Z_1, \ldots, Z_N)) \\ &= 1 - P_\theta(\theta_1 \notin \mathbb{C}_1(Z_1, \ldots, Z_N) \text{ or } \ldots \text{ or } \theta_r \notin \mathbb{C}_r(Z_1, \ldots, Z_N)) \\ &\geq 1 - \sum_{i=1}^r P_\theta(\theta_i \notin \mathbb{C}_i(Z_1, \ldots, Z_N)) = 1 - r\alpha. \square \end{aligned}$$

Let be Z_0 the random variable for a future observation. Typically we have here that $Z_1, \ldots, Z_N, Z_0$ are independent and identically distributed.

2.7.3 Definition (Prediction interval)
$\mathbb{P} : [0, \infty)^N \ni (z_1, \ldots, z_N) \to \mathbb{P}(z_1, \ldots, z_N) \in \mathcal{P}(\mathbb{R})$ *is a* $(1-\alpha)$*-prediction interval function for* Z_0 *if*

$$P_\theta(Z_0 \in \mathbb{P}(Z_1, \ldots, Z_N)) \geq 1 - \alpha$$

is satisfied for all $\theta \in \Theta$.

To compare several prediction intervals, one can consider the length of the prediction intervals and the coverage rate, i.e. add the relative number of future observations falling in the prediction interval. A combination of the length and the coverage rate of a prediction interval is the interval score of Gneiting und Raftery (2007).

2.7.4 Definition (Gneiting und Raftery (2007))
Is $[l,u] := [l(z_1,\ldots,z_N), u(z_1,\ldots,z_N)] := \mathbb{P}(z_1,\ldots,z_N)$ a $(1-\alpha)$-prediction interval for Z_0 and z_0 a future observation, then

$$\mathcal{S}(\mathbb{P}(z_1,\ldots,z_N), z_0) := (u-l) + \frac{2}{\alpha}(l-z_0)\mathbb{1}\{z_0 < l\} + \frac{2}{\alpha}(z_0-u)\mathbb{1}\{z_0 > u\}$$

is called an interval score of the $(1-\alpha)$-prediction interval $\mathbb{P}(z_1,\ldots,z_N)$ for Z_0 at z_0.

2.7.5 Remark (Naive or plug-in prediction interval)
If $Z_0, Z_1, \ldots, Z_N$ are i.i.d., each with cumulative distribution function F_θ and $\widehat{\theta} := \widehat{\theta}(z_1,\ldots,z_N)$ is an estimate for θ then a naive or plug-in $(1-\alpha)$-prediction interval $\mathbb{P}$ is given by

$$\mathbb{P}(z_1,\ldots,z_N) = \left[F_{\widehat{\theta}}^{-1}(\eta_1), F_{\widehat{\theta}}^{-1}(\eta_2)\right]$$

where $0 \leq \eta_1 < \eta_2 \leq 1$ with $\eta_2 - \eta_1 = 1-\alpha$. For finite samples N, these prediction intervals are usually too small so that $P_\theta(Z_0 \in \mathbb{P}(Z_1,\ldots,Z_N)) \geq 1-\alpha$ is not satisfied. However, if the estimator $\widehat{\theta}_N(Z_1,\ldots,Z_N)$ is a consistent estimator for θ then

$$\lim_{N\to\infty} F_{\widehat{\theta}_N}^{-1}(\eta_1) = F_\theta^{-1}(\eta_1),\ \lim_{N\to\infty} F_{\widehat{\theta}_N}^{-1}(\eta_2) = F_\theta^{-1}(\eta_2)$$

if $F_\theta^{-1}(\eta)$ is continuous in θ for $\eta = \eta_1, \eta_2$. Then

$$\begin{aligned}\lim_{N\to\infty} P_\theta(Z_0 \in \mathbb{P}(Z_1,\ldots,Z_N)) &= \lim_{N\to\infty}\Big(F_\theta(F_{\widehat{\theta}_N}^{-1}(\eta_2)) - F_\theta(F_{\widehat{\theta}_N}^{-1}(\eta_1))\Big)\\ &= F_\theta(F_\theta^{-1}(\eta_2)) - F_\theta(F_\theta^{-1}(\eta_1)) = \eta_2 - \eta_1 = 1-\alpha.\end{aligned}$$

For finite samples, the uncertainty of an estimator $\widehat{\theta}$ must be taken into account. A general simple approach for doing this is to base the prediction interval on a confidence set for θ.

2.7.6 Theorem (Prediction intervals based on confidence sets for indenpendet observations)
If $Z_1,\ldots,Z_N,Z_0$ are independent distributed, $0 \leq \eta_1 < \eta_2 \leq 1$ with $\eta_2 - \eta_1 = 1-\alpha_1$, and $\mathbb{C}$ is a $(1-\alpha_2)$-confidence set function, then $\mathbb{P}$ given by

$$\mathbb{P}(z_1,\ldots,z_N) = \bigcup_{\theta\in\mathbb{C}(z_1,\ldots,z_N)} \left[F_\theta^{-1}(\eta_1), F_\theta^{-1}(\eta_2)\right]$$

is a $(1-\alpha_1)(1-\alpha_2)$-prediction interval function for Z_0.

Proof. At first note that for any $\theta_* \in \Theta$

$$P_{\theta_*}\left(Z_0 \in \left[F_{\theta_*}^{-1}(\eta_1), F_{\theta_*}^{-1}(\eta_2)\right]\right) = F_{\theta_*}\left(F_{\theta_*}^{-1}(\eta_2)\right) - F_{\theta_*}\left(F_{\theta_*}^{-1}(\eta_1)\right) = \eta_2 - \eta_1 = 1-\alpha_1$$

is satisfied. Then we obtain for any $\theta_* \in \Theta$ with the independence of Z_0 and $Z_1, \ldots, Z_N$

$$\begin{aligned}
P_{\theta_*}(Z_0 \in \mathbb{P}(Z_1, \ldots, Z_N)) &= P_{\theta_*}\left(Z_0 \in \bigcup_{\theta \in \mathbb{C}(Z_1,\ldots,Z_N)} \left[F_\theta^{-1}(\eta_1), F_\theta^{-1}(\eta_2)\right]\right) \\
&\geq P_{\theta_*}\left(Z_0 \in \bigcup_{\theta \in \mathbb{C}(Z_1,\ldots,Z_N)} \left[F_\theta^{-1}(\eta_1), F_\theta^{-1}(\eta_2)\right],\ \theta_* \in \mathbb{C}(Z_1, \ldots, Z_N)\right) \\
&\geq P_{\theta_*}\left(Z_0 \in \left[F_{\theta_*}^{-1}(\eta_1), F_{\theta_*}^{-1}(\eta_2)\right],\ \theta_* \in \mathbb{C}(Z_1, \ldots, Z_N)\right) \\
&= P_{\theta_*}\left(Z_0 \in \left[F_{\theta_*}^{-1}(\eta_1), F_{\theta_*}^{-1}(\eta_2)\right]\right) \cdot P_{\theta_*}\left(\theta_* \in \mathbb{C}(Z_1, \ldots, Z_N)\right) \geq (1-\alpha_1)(1-\alpha_2) \ \square
\end{aligned}$$

2.7.7 Remark
Usually $\alpha_1 = \alpha_2$ with $(1-\alpha_1)^2 = 1-\alpha$ is used in Theorem 2.7.6. However, for large sample sizes N, it is better to use $\alpha_1 > \alpha_2$ with $(1-\alpha_1)(1-\alpha_2) = 1-\alpha$ to be competitive with the naive prediction interval. In particular one should use $\alpha_1(N) \uparrow \alpha$ and $\alpha_2(N) \downarrow 0$ for $N \to \infty$ to get the same asymptotic behavior as the naive prediction interval. One has only to ensure that $\alpha_2(N)$ is skrinking not too quickly to 0 so that the confidence set is not shrinking to one point which is the underlying parameter. But usually a rate $\alpha_2(N) \downarrow 0$ can be found so that the confidence set is shrinking to the true parameter.

Observations with exponential distribution

2.7.8 Lemma
If $T_1, \ldots, T_N$ are independent and $T_n \sim \mathcal{G}(\lambda, \beta_n)$ for $n = 1, \ldots, N$, then $\sum_{n=1}^N T_n \sim \mathcal{G}(\lambda, \sum_{n=1}^N \beta_n)$.

Proof. By induction as in the proof of Theorem 2.2.3.

2.7.9 Lemma
If $T_1, \ldots, T_N \sim \mathcal{E}(\lambda)$ are independent and $F_{\lambda,N}$ is the cumulative distribution function of the Gamma distribution with parameters λ and N, then $\mathbb{C}$ given by

$$\mathbb{C}(t_1, \ldots, t_N) = \left\{\lambda;\ \sum_{n=1}^N t_n \in \left[F_{\lambda,N}^{-1}\left(\frac{\alpha}{2}\right), F_{\lambda,N}^{-1}\left(1-\frac{\alpha}{2}\right)\right]\right\}$$

is a $(1-\alpha)$-confidence set function for λ.

2.7.10 Corollary
If $T_1, \ldots, T_N, T_0 \sim \mathcal{E}(\lambda)$ are independent, $0 \leq \eta_1 < \eta_2 \leq 1$ with $\eta_2 - \eta_1 = 1-\alpha$, and $\mathbb{C}$ is given by Lemma 2.7.9, then $\mathbb{P}$ given by

$$\mathbb{P}(t_1, \ldots, t_N) = \bigcup_{\lambda \in \mathbb{C}(t_1,\ldots,t_N)} \left[F_{\lambda,1}^{-1}(\eta_1), F_{\lambda,1}^{-1}(\eta_2)\right]$$

is a $(1-\alpha)^2$-prediction interval function for T_0.

2.7.11 Theorem (Meeker and Escobar (1998), p. 300)
If $T_1, \ldots, T_N, T_0 \sim \mathcal{E}(\lambda)$ are independent with ordering $T_{(1)} \leq \ldots \leq T_{(N)}$, $Z_n = min(T_n, T_{(R)})$, $n = 1, \ldots, N$, are the Typ II censored observations with $R \leq N$, $\widehat{\theta}(Z_1, \ldots, Z_N)$ is the maximum likelihood estimator for $\theta = \frac{1}{\lambda}$ based on $Z_1, \ldots, Z_N$ and $f_{n,m;\alpha}$ is the α-quantile of the F distribution with n and m degrees of freedom, then $\mathbb{P}$ given by

$$\mathbb{P}(z_1, \ldots, z_N) = \left[\widehat{\theta}(z_1, \ldots, z_N)\, f_{2,2R;\alpha/2}, \widehat{\theta}(z_1, \ldots, z_N)\, f_{2,2R;1-\alpha/2}\right],$$

is a $(1-\alpha)$-prediction interval function for T_0.

Proof. According to Example 2.5.5, the maximum likelihood estimator for λ is given by

$$\widehat{\lambda}(Z_1, \ldots, Z_N) = \frac{R}{(N-R)\, t_{(R)} + \sum_{n=1}^R t_{(n)}}$$

so that

$$\widehat{\theta}(Z_1, \ldots, Z_N) = \frac{(N-R)\, T_{(R)} + \sum_{n=1}^R T_{(n)}}{R} = \frac{(N-R+1)\, T_{(R)} + \sum_{n=1}^{R-1} T_{(n)}}{R}$$

is the maximum likelihood estimator for $\theta = \frac{1}{\lambda}$. Using the normalised sample spacings

$$D_n := (N-n+1)(T_{(n)} - T_{(n-1)}), \ \ n = 2, \ldots, N, \ \ D_1 := N\, T_{(1)},$$

we get

$$\widehat{\theta}(Z_1, \ldots, Z_N) = \frac{\sum_{n=1}^R D_n}{R},$$

where $D_1, \ldots, D_R \sim \mathcal{E}(\lambda)$ are independent according to Theorem 2.2.4. Theorem 2.2.3 provides

$$2\lambda \sum_{n=1}^R D_n \sim \chi^2_{2R}$$

and $2\lambda\, T_0 \sim \chi^2_2$. Hence

$$\frac{T_0}{\widehat{\theta}(Z_1, \ldots, Z_N)} = \frac{\frac{2\lambda\, T_0}{2}}{\frac{2\lambda \sum_{n=1}^R D_n}{2R}}$$

has a F distribution with 2 and $2R$ degrees of freedom. This implies

$$P_\lambda(T_0 \in \mathbb{P}(Z_1, \ldots, Z_N)) = P_\lambda\left(f_{2,2R;\alpha/2} \leq T_0/\widehat{\theta}(Z_1, \ldots, Z_N) \leq f_{2,2R;1-\alpha/2}\right) = 1-\alpha. \square$$

2.7.12 Remark
Theorem 2.7.11 provides in particular a prediction interval for uncensored data with exponential distribution. Since the exponential dsitribution is a special Gamma distribution, prediction intervals for Gamma distributions can be used as well.

Observations with Gamma distribution

2.7.13 Lemma
If X, Y are independent and $X \sim \mathcal{G}(\lambda, \beta_1)$, $Y \sim \mathcal{G}(\lambda, \beta_2)$, then

$$\frac{X}{X+Y} \sim \mathcal{B}(\beta_1, \beta_2),$$

where $\mathcal{B}(a, b)$ denotes the Beta distribution given by the density

$$f(x) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} x^a (1-x)^b \mathbb{1}_{[0,1]}(x).$$

Proof. See Kahle et al. (2016) Proposition 2.6.

2.7.14 Theorem
If $T_1, \ldots, T_N, T_0 \sim \mathcal{G}(\lambda, \beta)$ are independent, β is a known parameter, $S(t_1, \ldots, t_N) := \sum_{n=1}^N t_n$, $\eta_2 - \eta_1 = 1 - \alpha$ and $b_{a,b;\alpha}$ the α-quantile of the $\mathcal{B}(a, b)$ distribution, then $\mathbb{P}$ given by

$$\mathbb{P}(t_1, \ldots, t_N) = \left[\frac{S(t_1, \ldots, t_N)(1 - b_{N\beta,\beta;\eta_2})}{b_{N\beta,\beta;\eta_2}}, \frac{S(t_1, \ldots, t_N)(1 - b_{N\beta,\beta;\eta_1})}{b_{N\beta,\beta;\eta_1}}\right],$$

is a $(1-\alpha)$-prediction interval function for T_0.

Proof. We have $S(T_1, \ldots, T_N) \sim \mathcal{G}(\lambda, N\beta)$ according to 2.7.8 so that

$$\frac{S(T_1, \ldots, T_N)}{S(T_1, \ldots, T_N) + T_0} \sim \mathcal{B}(N\beta, \beta)$$

according to Lemma 2.7.13. This implies

$$\begin{aligned}
&P_{\lambda,\beta}(T_0 \in \mathbb{P}(T_1, \ldots, T_N)) \\
&= P_{\lambda,\beta}\left(\frac{S(T_1, \ldots, T_N)(1 - b_{N\beta,\beta;\eta_2})}{b_{N\beta,\beta;\eta_2}} \leq T_0 \leq \frac{S(T_1, \ldots, T_N)(1 - b_{N\beta,\beta;\eta_1})}{b_{N\beta,\beta;\eta_1}}\right) \\
&= P_{\lambda,\beta}\left(S(T_1, \ldots, T_N) \leq b_{N\beta,\beta;\eta_2}(S(T_1, \ldots, T_N) + T_0),\right. \\
&\qquad\qquad \left. b_{N\beta,\beta;\eta_1}(S(T_1, \ldots, T_N) + T_0) \leq S(T_1, \ldots, T_N)\right) \\
&= P_{\lambda,\beta}\left(b_{N\beta,\beta;\eta_1} \leq \frac{S(T_1, \ldots, T_N)}{S(T_1, \ldots, T_N) + T_0} \leq b_{N\beta,\beta;\eta_2}\right) \\
&= \eta_2 - \eta_1 = 1 - \alpha. \Box
\end{aligned}$$

Observations with (log)normal distribution

Assume now that $T_1, \dots, T_N$ are independent with a lognormal distribution. Then $Y_1 = \ln(T_n), \dots, Y_N = \ln(T_N)$ have a normal distribution. Set

$$y = (y_1, \dots, y_N)^\top, \quad \overline{y}. = \frac{1}{N}\sum_{n=1}^{N} y_n, \quad \widehat{\sigma}(y)^2 = \frac{1}{N-1}\sum_{n=1}^{N}(y_n - \overline{y}.)^2$$

and let be Y, $\overline{Y}.$, and $\widehat{\sigma}(Y)^2$ the corresponding random variables. If $Y_n \sim \mathcal{N}(\mu, \sigma^2)$, then $\overline{Y}. \sim \mathcal{N}(\mu, \frac{1}{N}\sigma^2)$, $(N-1)\frac{\widehat{\sigma}(Y)^2}{\sigma^2}$ has a χ^2-distribution with $N-1$ degrees of freedom and $\overline{Y}.$ and $\widehat{\sigma}(Y)^2$ are independent.

2.7.15 Lemma
If $Y_1, \dots, Y_N \sim \mathcal{N}(\mu, \sigma^2)$ are independent and $t_{N-1;\alpha}$ is the α-quantile of the t distribution with $N-1$ degrees of freedom, then $\mathbb{C}_\mu$ given by

$$\mathbb{C}_\mu(y_1, \dots, y_N) = \left[\overline{y}. - \frac{\widehat{\sigma}(y)}{\sqrt{N}}\, t_{N-1;1-\alpha/2}, \quad \overline{y}. + \frac{\widehat{\sigma}(y)}{\sqrt{N}}\, t_{N-1;1-\alpha/2}\right]$$

is a $(1-\alpha)$-confidence set (interval) function for μ.

Proof.

$$P_{\mu,\sigma^2}(\mu \in \mathbb{C}_\mu(Y_1, \dots, Y_N)) = P_{\mu,\sigma^2}\left(-\frac{\widehat{\sigma}(Y)}{\sqrt{N}}\, t_{N-1;1-\alpha/2} \le \mu - \overline{Y}. \le \frac{\widehat{\sigma}(Y)}{\sqrt{N}}\, t_{N-1;1-\alpha/2}\right)$$
$$= \quad P_{\mu,\sigma^2}\left(t_{N-1;\alpha/2} \le \sqrt{N}\,\frac{\overline{Y}. - \mu}{\widehat{\sigma}(Y)} \le t_{N-1;1-\alpha/2}\right) = 1 - \alpha$$

since $\sqrt{N}\,\frac{\overline{Y}.-\mu}{\widehat{\sigma}(Y)}$ has a t-distribution with $N-1$ degree of freedom. $\square$

2.7.16 Lemma
If $Y_1, \dots, Y_N \sim \mathcal{N}(\mu, \sigma^2)$ are independent and $\chi^2_{N-1;\alpha}$ is the α-quantile of the χ^2 distribution with $N-1$ degrees of freedom, then $\mathbb{C}_{\sigma^2}$ given by

$$\mathbb{C}_{\sigma^2}(y_1, \dots, y_N) = \left[\frac{(N-1)\widehat{\sigma}(Y)^2}{\chi^2_{N-1;1-\alpha/2}} \;,\; \frac{(N-1)\widehat{\sigma}(Y)^2}{\chi^2_{N-1;\alpha/2}}\right]$$

is a $(1-\alpha)$-confidence set (interval) function for σ^2.

Proof.

$$P_{\mu,\sigma^2}(\sigma^2 \in \mathbb{C}_{\sigma^2}(Y_1, \dots, Y_N)) = P_{\mu,\sigma^2}\left(\frac{(N-1)\widehat{\sigma}(Y)^2}{\chi^2_{N-1;1-\alpha/2}} \le \sigma^2 \le \frac{(N-1)\widehat{\sigma}(Y)^2}{\chi^2_{N-1;\alpha/2}}\right)$$
$$= \quad P_{\mu,\sigma^2}\left(\frac{1}{\chi^2_{N-1;1-\alpha/2}} \le \frac{\sigma^2}{(N-1)\widehat{\sigma}(Y)^2} \le \frac{1}{\chi^2_{N-1;\alpha/2}}\right)$$
$$= \quad P_{\mu,\sigma^2}\left(\chi^2_{N-1;\alpha/2} \le (N-1)\,\frac{\widehat{\sigma}(Y)^2}{\sigma^2} \le \chi^2_{N-1;1-\alpha/2}\right) = 1 - \alpha$$

since $(N-1)\,\frac{\widehat{\sigma}(Y)^2}{\sigma^2}$ has a χ^2-distribution with $N-1$ degree of freedom. $\square$

2.7.17 Corollary
Let $Y_1,\ldots,Y_N \sim \mathcal{N}(\mu,\sigma^2)$ be independent and $\mathbb{C}$ given by

$$\mathbb{C}(y_1,\ldots,y_N) = \mathbb{C}_\mu(y_1,\ldots,y_N) \times \mathbb{C}_{\sigma^2}(y_1,\ldots,y_N),$$

where $\mathbb{C}_\mu$ and $\mathbb{C}_{\sigma^2}$ are $(1-\alpha)$-confidence interval functions for μ and σ^2 given by Lemma 2.7.15 and Lemma 2.7.16, respectively. Then $\mathbb{C}$ is a $(1-2\alpha)$-confidence set function for (μ,σ^2).

Proof. The assertion follows from Lemma 2.7.2.

2.7.18 Remark
A first $(1-\alpha_1)(1-\alpha_2)$-prediction interval for Y_0 is given according to Theorem 2.7.6 by

$$\mathbb{P}(y_1,\ldots,y_N) = \bigcup_{(\mu,\sigma^2)\in\mathbb{C}(y_1,\ldots,y_N)} \left[F^{-1}_{(\mu,\sigma^2)}\left(\frac{\alpha_1}{2}\right), F^{-1}_{(\mu,\sigma^2)}\left(1-\frac{\alpha_1}{2}\right)\right]$$

where $\mathbb{C}$ is a $(1-\alpha_2)$-confidence set function for (μ,σ^2). Because of the symmetry of the normal distribution, the choice $\eta_1 = \frac{\alpha_1}{2}$ and $\eta_2 = 1-\frac{\alpha_1}{2}$ is the best choice. However, here a better prediction interval can be obtained by the following theorem.

2.7.19 Theorem
If $Y_1,\ldots,Y_N, Y_0 \sim \mathcal{N}(\mu,\sigma^2)$ are independent, then $\mathbb{P}$ given by

$$\mathbb{P}(y_1,\ldots,y_N) = \left[\overline{y}. - \widehat{\sigma}(y)\sqrt{1+\frac{1}{N}}\ t_{N-1;1-\alpha/2},\quad \overline{y}. + \widehat{\sigma}(y)\sqrt{1+\frac{1}{N}}\ t_{N-1;1-\alpha/2}\right]$$

is a $(1-\alpha)$-predition interval function for Y_0.

Proof. Since Y_0 and $\overline{Y}.$ are independent, it holds $Y_0 - \overline{Y}. \sim \mathcal{N}(0,\sigma^2_{\mathbb{P}})$ with

$$\sigma^2_{\mathbb{P}} = \mathrm{var}(Y_0-\overline{Y}.) = \mathrm{var}(Y_0) + \mathrm{var}(\overline{Y}.) = \sigma^2\left(1+\frac{1}{N}\right).$$

This implies

$$\frac{1}{\sqrt{1+\frac{1}{N}}}\,\frac{Y_0-\overline{Y}.}{\sigma} \sim \mathcal{N}(0,1).$$

Since $\overline{Y}.$ and $\widehat{\sigma}(Y)^2$ are independent, also $Y_0-\overline{Y}.$ and $\widehat{\sigma}(Y)^2$ are independent so that

$$\frac{1}{\sqrt{1+\frac{1}{N}}}\,\frac{Y_0-\overline{Y}.}{\widehat{\sigma}(Y)}$$

has a t-distribution with $N-1$ degrees of freedom. Hence we can conclude

$$\begin{aligned}
& P_{\mu,\sigma^2}(Y_0 \in \mathbb{P}(Y_1,\ldots,Y_N)) \\
= \;& P_{\mu,\sigma^2}\left(\overline{Y}. - \widehat{\sigma}(Y)\sqrt{1+\frac{1}{N}}\; t_{N-1;1-\alpha/2} \leq Y_0 \leq \overline{Y}. - \widehat{\sigma}(Y)\sqrt{1+\frac{1}{N}}\; t_{N-1;1-\alpha/2}\right) \\
= \;& P_{\mu,\sigma^2}\left(-t_{N-1;1-\alpha/2} \leq \frac{1}{\sqrt{1+\frac{1}{N}}}\,\frac{Y_0-\overline{Y}.}{\widehat{\sigma}(Y)} \leq t_{N-1;1-\alpha/2}\right) = 1-\alpha.\square
\end{aligned}$$

2.7.20 Corollary
If $T_1,\ldots,T_N,T_0 \sim \mathcal{LN}(\mu,\sigma^2)$ are independent, then $\mathbb{P}$ given by

$$\mathbb{P}(t_1,\ldots,t_N) = \left[\exp\left(\overline{y}. - \widehat{\sigma}(y)\sqrt{1+\frac{1}{N}}\; t_{N-1;1-\alpha/2}\right),\; \exp\left(\overline{y}. + \widehat{\sigma}(y)\sqrt{1+\frac{1}{N}}\; t_{N-1;1-\alpha/2}\right))\right]$$

with $y_n = \log(t_n)$, $n=1,\ldots,N$, is a $(1-\alpha)$-predition interval function for T_0.

2.8 Convergence in distribution

Asymptotic prediction intervals can be used in situations where exact predictions intervals cannot be calculated. To derive approximate prediction intervals, some results about convergence in distribution are necessary.

2.8.1 Definition (Convergence in probability and almost surely convergence)
Let be $(X_n)_{n\in\mathbb{N}}$, X random variables on $(\Omega,\ \mathcal{A},\ P)$ with values in $\mathbb{R}^r$.

(i) $(X_n)_{n\in\mathbb{N}}$ *converges almost surely to X (briefly $X_n \to X$ P-a.s.)*
$:\Leftrightarrow P(\{\omega \in \Omega;\ \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$

(ii) $(X_n)_{n\in\mathbb{N}}$ *converges in probability to X (briefly $X_n \xrightarrow{\mathcal{P}} X$)*
$:\Leftrightarrow \bigwedge\limits_{\varepsilon>0} \lim\limits_{n\to\infty} P(\{\omega \in \Omega;\ \|X_n(\omega) - X(\omega)\| > \varepsilon\}) = 0.$

A general definition of convergence in distribution (also called weak convergence) is the following.

2.8.2 Definition (Convergence in distribution)
Let be $(P_n)_{n\in\mathbb{N}}$, P probability measures on $(\mathbb{R}^r, \mathcal{B}_r)$ where $\mathcal{B}_r$ is the Borel-σ-algebra on $\mathbb{R}^r$, then $(P_n)_{n\in\mathbb{N}}$ converges in distribution (converges weakly) to P
$:\Leftrightarrow$
$\lim\limits_{n\to\infty} P_n(B) = P(B)$ *for all* $B \in \mathcal{B}_r$ *with* $P(\partial B) = 0$, *where* ∂B *is the border of* B.

The assumption $(\mathbb{R}^r, \mathcal{B}_r)$ can be weakened but here it will be enough.

2.8.3 Theorem (Theorem of Portmanteau, see e.g. Witting and Müller-Funk 1995, Satz 5.40)
$(P_n)_{n\in\mathbb{N}}$ *converges in distribution (converges weakly) to P*
$:\Leftrightarrow$
$\lim\limits_{n\to\infty} \int f\, dP_n = \int f\, dP$ *for all continuous and bounded* $f : \mathbb{R}^r \to \mathbb{R}$.

2.8.4 Definition (Convergence in distribution for random variables)
Let be $(X_n)_{n\in\mathbb{N}}$, X random variables on $(\Omega,\ \mathcal{A},\ P)$ with values in $\mathbb{R}^r$.
$(X_n)_{n\in\mathbb{N}}$ *converges in distribution to X (briefly $X_n \xrightarrow{\mathcal{D}} X$)*
$:\Leftrightarrow$
$\lim\limits_{n\to\infty} P^{X_n}(B) = P^X(B)$ *for all* $B \in \mathcal{B}_r$ *with* $P^X(\partial B)) = 0.$

The following Theorem is sometimes used also as the definition of convergence in distribution.

2.8.5 Theorem (Convergence in distribution to a continuous distribution, see e.g. Witting and Müller-Funk 1995, Satz 5.58)
Let $(X_n)_{n\in\mathbb{N}}$, X be random variables on $(\Omega,\ \mathcal{A},\ P)$ with values in $\mathbb{R}^r$ and cumulative distribution functions $(F_{X_n})_{n\in\mathbb{N}}$ and F_X. If F_X is continuous then
$(X_n)_{n\in\mathbb{N}}$ converges in distribution to X
$\Leftrightarrow$
$\lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x)$ *for all* $x \in \mathbb{R}^r$.

If F_X is not continuous, then $\lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x)$ must be satisfied for all $x \in \mathbb{R}^r$, at which F_X is continuous. However, here F_X will be always continuous so that the more general definition is not necessary.

2.8.6 Lemma (Lemma of Slutzky, see e.g. Witting and Müller-Funk 1995, Korollar 5.84)
If $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are sequences of random variables, X a random variable. If $(X_n)_{n\in\mathbb{N}}$ converges in distribution to X and $(\|X_n - Y_n\|)_{n\in\mathbb{N}}$ converges in probability to 0, then $(Y_n)_{n\in\mathbb{N}}$ converges to X. Shortly:

$$X_n \xrightarrow{\mathcal{D}} X, \quad \|X_n - Y_n\| \xrightarrow{\mathcal{P}} 0 \Longrightarrow Y_n \xrightarrow{\mathcal{D}} X\,.$$

2.8.7 Corollary

(i) $(X_n)_{n\in\mathbb{N}} \xrightarrow{\mathcal{P}} X \Longrightarrow (X_n)_{n\in\mathbb{N}} \xrightarrow{\mathcal{D}} X.$

(ii) *The reverse implication is not satisfied in general.*

(iii) $(X_n)_{n\in\mathbb{N}} \xrightarrow{\mathcal{D}} a \in \mathbb{R}^r \Longrightarrow (X_n)_{n\in\mathbb{N}} \xrightarrow{\mathcal{P}} a.$

(iv) $(\sqrt{n}X_n)_{n\in\mathbb{N}} \xrightarrow{\mathcal{D}} X \Longrightarrow (X_n)_{n\in\mathbb{N}} \xrightarrow{\mathcal{P}} 0.$

2.8.8 Corollary
Let be $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$, $(Z_n)_{n\in\mathbb{N}}$ sequences of random variables with $X_n \xrightarrow{\mathcal{P}} a \in \mathbb{R}^r$, $Y_n \xrightarrow{\mathcal{P}} A \in \mathbb{R}^{s\times r}$ and $Z_n \xrightarrow{\mathcal{D}} Z$. Then we have:

(i) $X_n + Z_n \xrightarrow{\mathcal{D}} a + Z,$

(ii) $Y_n\, Z_n \xrightarrow{\mathcal{D}} A\, Z.$

Both corollaries can be shown as in the one-dimensional case. For alternative proofs see also Witting and Müller-Funk (1995), Satz 5.83 and Korollar 5.84. To see in particular the assertions of Corollary 2.8.8 (iii) and (iv), the Cramér-Wold device is helpful.

2.8.9 Theorem (Cramér-Wold, see e.g. Witting and Müller-Funk 1995, Korollar 5.69)
If $(X_n)_{n\in\mathbb{N}}$, X are random varibales with values in $\mathbb{R}^r$, then

$$X_n \xrightarrow{\mathcal{D}} X \Longleftrightarrow u^\top X_n \xrightarrow{\mathcal{D}} u^\top X \ \text{ for all } \ u \in \mathbb{R}^r.$$

2.8.10 Lemma (See e.g. Witting and Müller-Funk 1995, Hilfssatz 5.80)
If $(X_n)_{n\in\mathbb{N}}$, X are random varibales with values in $\mathbb{R}^r$, $(Y_n)_{n\in\mathbb{N}}$, Y are random varibales with values in $\mathbb{R}^s$ and X_n and Y_n are independent for all $n \in \mathbb{N}$, then

$$X_n \xrightarrow{\mathcal{D}} X,\ \ Y_n \xrightarrow{\mathcal{D}} Y \quad\Longrightarrow\quad (X_n, Y_n) \xrightarrow{\mathcal{D}} (X, Y).$$

In particular, if $r = s$,

$$X_n \xrightarrow{\mathcal{D}} X,\ \ Y_n \xrightarrow{\mathcal{D}} Y \quad\Longrightarrow\quad X_n + Y_n \xrightarrow{\mathcal{D}} X + Y.$$

2.8.11 Theorem (Continuous mapping thoerem, see e.g. Witting and Müller-Funk 1995, Satz 5.43)
If $(X_n)_{n\in\mathbb{N}}$, X are random variables with values in $\mathbb{R}^r$ and $f : \mathbb{R}^r \to \mathbb{R}^s$ is continuous, then

$$\begin{aligned} (i) \qquad & X_n \xrightarrow{\mathcal{D}} X \Longrightarrow f(X_n) \xrightarrow{\mathcal{D}} f(X), \\ (ii) \qquad & X_n \xrightarrow{\mathcal{P}} X \Longrightarrow f(X_n) \xrightarrow{\mathcal{P}} f(X). \end{aligned}$$

2.8.12 Theorem (Delta method, see e.g. Witting and Müller-Funk 1995, Satz 5.78)
$(X_n)_{n\in\mathbb{N}}$, X are random variables with values in $\mathbb{R}^r$, $f : \mathbb{R}^r \to \mathbb{R}^s$ is differentiable at θ with derivative $\dot{f}(\theta) = \frac{\partial}{\partial x} f(x)\Big|_{x=\theta} \in \mathbb{R}^{r\times s}$, and $(c_n)_{n\in\mathbb{N}}$ is a sequence in $(0, \infty)$ with $c_n \longrightarrow \infty$, then

$$c_n(X_n - \theta) \xrightarrow{\mathcal{D}} X \Longrightarrow c_n(f(X_n) - f(\theta)) \xrightarrow{\mathcal{D}} \dot{f}(\theta)^\top X.$$

2.9 Asymptotic confidence sets and prediction intervals

We present here the results for Type I censored data so that $z_n = \min(t_n, c)$ for $n = 1, \ldots, N$ and $c \in (0, \infty)$. The special case of uncensored observations is included by $c = \infty$ which means that no censoring occurs.

2.9.1 Definition (Asymptotic confidence set)
$\mathbb{C}_N : [0, \infty)^N \ni (z_1, \ldots, z_N) \to \mathbb{C}_N(z_1, \ldots, z_N) \in \mathcal{P}(\mathbb{R}^r)$ *is an asymptotic* $(1 - \alpha)$*-confidence set function for* $a(\theta)$ *if*

$$\lim_{N\to\infty} P_\theta(a(\theta) \in \mathbb{C}_N(Z_1, \ldots, Z_N)) \geq 1 - \alpha$$

is satisfied for all $\theta \in \Theta$.

2.9.2 Definition (Asymptotic prediction interval)
$\mathbb{P}_N : [0, \infty)^N \ni (z_1, \ldots, z_N) \to \mathbb{P}_N(z_1, \ldots, z_N) \in \mathcal{P}(\mathbb{R})$ *is an asymptotic* $(1 - \alpha)$*-prediction interval function for* T_0 *if*

$$\lim_{N\to\infty} P_\theta(T_0 \in \mathbb{P}_N(Z_1, \ldots, Z_N)) \geq 1 - \alpha$$

is satisfied for all $\theta \in \Theta$.

2.9.3 Theorem (Naive / plug-in prediction interval)
Let $\widehat{\theta}_N := \widehat{\theta}_N(Z_1, \ldots, Z_N)$ *be a weak consistent estimator for* θ *and* F_θ *the cumulative distribution of* T_0 *so that* $F_\theta(t)$ *is continuous in* t *and* $F_\theta^{-1}(\eta)$ *is continuous in* θ *for* $\eta = \eta_1, \eta_2$*. If* $0 \leq \eta_1 < \eta_2 \leq 1$ *and* $\eta_2 - \eta_1 = 1 - \alpha$ *then the naive or plug-in prediction interval function* $\mathbb{P}_N$ *given by*

$$\mathbb{P}_N(z_1, \ldots, z_N) = \left[F_{\widehat{\theta}_N}^{-1}(\eta_1) \ , \ F_{\widehat{\theta}_N}^{-1}(\eta_2)\right]$$

is an asymptotic $1 - \alpha$*-prediction interval function for* T_0.

Proof. Since $F_\theta^{-1}(\eta)$ is continuous in θ for $\eta = \eta_1, \eta_2$, the weak consistency of $\widehat{\theta}_N$ implies for any $\theta \in \Theta$ and any $\epsilon > 0$

$$\begin{aligned}
&\lim_{N\to\infty} P_\theta(F_{\widehat{\theta}_N}^{-1}(\eta_1) > F_\theta^{-1}(\eta_1) + \epsilon) \\
&\leq \lim_{N\to\infty} P_\theta(|F_{\widehat{\theta}_N}^{-1}(\eta_1) - F_\theta^{-1}(\eta_1)| > \epsilon) \leq \lim_{N\to\infty} P_\theta(\|\widehat{\theta} - \theta\| > \tilde{\epsilon}) = 0
\end{aligned}$$

so that

$$\lim_{N\to\infty} P_\theta(F_{\widehat{\theta}_N}^{-1}(\eta_1) > F_\theta^{-1}(\eta_1) + \epsilon) = 0.$$

Analogously, it follows

$$\lim_{N\to\infty} P_\theta(F_{\widehat{\theta}_N}^{-1}(\eta_2) < F_\theta^{-1}(\eta_2) - \epsilon) = 0.$$

Hence

$$\begin{aligned}
&\lim_{N\to\infty} P_\theta(T_0 \notin \mathbb{P}_N(Z_1,\ldots,Z_N)) \le \lim_{N\to\infty}\left(P_\theta\left(T_0 < F^{-1}_{\widehat{\theta}_N}(\eta_1)\right) + P_\theta\left(T_0 > F^{-1}_{\widehat{\theta}_N}(\eta_2)\right)\right)\\
\le\; & \lim_{N\to\infty}\Big(P_\theta\left(T_0 < F_\theta^{-1}(\eta_1)+\epsilon\right) + P_\theta(F^{-1}_{\widehat{\theta}_N}(\eta_1) > F_\theta^{-1}(\eta_1)+\epsilon)\\
&\qquad + \; P_\theta\left(T_0 > F_\theta^{-1}(\eta_2)-\epsilon\right) + P_\theta(F^{-1}_{\widehat{\theta}_N}(\eta_2) < F_\theta^{-1}(\eta_2)-\epsilon)\Big)\\
=\; & F_\theta\left(F_\theta^{-1}(\eta_1)+\epsilon\right) + 1 - F_\theta\left(F_\theta^{-1}(\eta_2)-\epsilon\right).
\end{aligned}$$

This holds for any $\epsilon > 0$. Since $F_\theta(t)$ is continuous in t, there exists for all $\epsilon^* > 0$ a $\epsilon > 0$ with

$$\left|F_\theta\left(F_\theta^{-1}(\eta)\pm\epsilon\right) - F_\theta\left(F_\theta^{-1}(\eta)\right)\right| \le \epsilon^*$$

for $\eta = \eta_1,\ \eta_2$. This implies for any $\epsilon^* > 0$

$$\begin{aligned}
&\lim_{N\to\infty} P_\theta(T_0 \notin \mathbb{P}_N(Z_1,\ldots,Z_N))\\
&\le\; F_\theta\left(F_\theta^{-1}(\eta_1)\right) + 1 - F_\theta\left(F_\theta^{-1}(\eta_2)\right) + 2\epsilon^* = \eta_1 + 1 - \eta_2 = 1-(\eta_2-\eta_1)+2\epsilon^* = \alpha + 2\epsilon^*
\end{aligned}$$

and thus $\lim_{N\to\infty} P_\theta(T_0 \notin \mathbb{P}_N(Z_1,\ldots,Z_N)) \le \alpha$. $\square$

Recall that the likelihood function for Type I censored observations is given by

$$l(\theta, z_n) := l(\theta, z_n, d_n) := f_\theta(z_n)^{d_n}\ S_\theta(z_n)^{1-d_n}$$

where $d_n := \mathbb{1}_{[0,c]}(t_n)$ and f is the density of T_n. In particular, we have then

$$\ln l(\theta, z_n) = \ln\ f_\theta(z_n)\,\mathbb{1}_{[0,c]}(t_n) + \ln\left(1 - F_\theta(z_n)\right)\mathbb{1}_{(c,\infty)}(t_n).$$

2.9.4 Definition
Is T a random variable with density f_θ and $Z = \min(T, c)$ then

$$\begin{aligned}
I_\theta(Z) &:= E_\theta\left(\left[\frac{\partial}{\partial\theta}\ln l(\theta, Z)\right]^2\right) \quad \text{if } \theta\in\mathbb{R},\\
I_\theta(Z) &:= E_\theta\left(\left[\frac{\partial}{\partial\theta}\ln l(\theta, Z)\frac{\partial}{\partial\theta}\ln l(\theta, Z)^\top\right]\right) \quad \text{if } \theta\in\mathbb{R}^r,
\end{aligned}$$

is called the Fisher information of Z at θ.

2.9.5 Theorem
a) The Fisher information satisfies under regularity conditions

$$I_\theta(Z) = -E_\theta\left[\frac{\partial^2}{\partial^2\theta}\ln l(\theta, Z)\right]$$

for all $\theta \in \Theta$.

b) If $Z_1, \ldots, Z_N$ are i.i.d., then with $Z_* = (Z_1, \ldots, Z_N)$

$$I_\theta(Z_*) = N\ I_\theta(Z_1)$$

for all $\theta \in \Theta$.

Proof.
a) At first note that it holds

$$\frac{\partial^2}{\partial^2\theta} \ln l(\theta, z) = \frac{\partial}{\partial\theta} \frac{\frac{\partial}{\partial\theta} l(\theta, z)}{l(\theta, z)} = \frac{\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta} l(\theta, z)}{l(\theta, z)} - \frac{\frac{\partial}{\partial\theta} l(\theta, z) \frac{\partial}{\partial\theta} l(\theta, z)^\top}{l(\theta, z)^2}.$$

If the regularity conditions ensure that the differentiation can be exchanged with the expectation then

$$\begin{aligned} E_\theta\left[\frac{\partial^2}{\partial^2\theta} \ln l(\theta, Z)\right] &= E_\theta\left[\frac{\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta} l(\theta, Z)}{l(\theta, Z)}\right] - E_\theta\left[\frac{\frac{\partial}{\partial\theta} l(\theta, Z)}{l(\theta, Z)} \frac{\frac{\partial}{\partial\theta} l(\theta, Z)^\top}{l(\theta, Z)}\right] \\ &= \frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta} E_\theta\left[\frac{l(\theta, Z)}{l(\theta, Z)}\right] - I_\theta(Z) = \frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\, 1 - I_\theta(Z) = -I_\theta(Z). \end{aligned}$$

b) The independence of $Z_1, \ldots, Z_N$ implies that the likelihood function of Z_* is given by

$$l(\theta, z_*) := \prod_{n=1}^N f_\theta(z_n)^{d_n}\ S_\theta(z_n)^{1-d_n} = \prod_{n=1}^N l(\theta, z_n)$$

so that the assertion follows from a). □

2.9.6 Lemma

$$\begin{aligned}
&a) \quad \int_0^c \lambda e^{-\lambda y} dy = -e^{-\lambda y}\Big|_0^c = 1 - e^{-\lambda c}. \\
&b) \quad \int_0^c y\, \lambda e^{-\lambda y} dy = -c\, e^{-\lambda c} + \frac{1}{\lambda}\left(1 - e^{-\lambda c}\right). \\
&c) \quad \int_0^c y^2\, \lambda e^{-\lambda y} dy = e^{-\lambda c}\left[-c^2 - \frac{2c}{\lambda} - \frac{2}{\lambda^2}\right] + \frac{2}{\lambda^2}. \\
&d) \quad \int_0^c \left(\frac{1}{\lambda} - y\right) \lambda e^{-\lambda y} dy - c\int_c^\infty \lambda e^{-\lambda y} dy = 0. \\
&e) \quad \int_0^c \left(\frac{1}{\lambda} - y\right)^2 \lambda e^{-\lambda y} dy + c^2 \int_c^\infty \lambda e^{-\lambda y} dy = \frac{1}{\lambda^2}\left(1 - e^{-\lambda c}\right).
\end{aligned}$$

Proof.
b) Using partial integration and a) we get

$$\int_0^c y\,\lambda e^{-\lambda y} dy = -y\,e^{-\lambda y}\Big|_0^c - \int_0^c -e^{-\lambda y} dy$$
$$= \; -c\,e^{-\lambda c} + \frac{1}{\lambda}\int_0^c \lambda e^{-\lambda y} dy = -c\,e^{-\lambda c} + \frac{1}{\lambda}\left(1 - e^{-\lambda c}\right).$$

c) Using partial integration and b) we get

$$\int_0^c y^2\,\lambda e^{-\lambda y} dy = -y^2\,e^{-\lambda y}\Big|_0^c - \int_0^c -2y\,e^{-\lambda y} dy$$
$$= \; -c^2\,e^{-\lambda c} + \frac{2}{\lambda}\int_0^c y\,\lambda e^{-\lambda y} dy$$
$$= \; -c^2\,e^{-\lambda c} + \frac{2}{\lambda}\left[-c\,e^{-\lambda c} + \frac{1}{\lambda}\left(1 - e^{-\lambda c}\right)\right]$$
$$= \; e^{-\lambda c}\left[-c^2 - \frac{2c}{\lambda} - \frac{2}{\lambda^2}\right] + \frac{2}{\lambda^2}.$$

d) Using a) and b) we obtain

$$\int_0^c \left(\frac{1}{\lambda} - y\right)\lambda e^{-\lambda y} dy - c\int_c^\infty \lambda e^{-\lambda y} dy$$
$$= \; \frac{1}{\lambda}\left(1 - e^{-\lambda c}\right) - \left[-c\,e^{-\lambda c} + \frac{1}{\lambda}\left(1 - e^{-\lambda c}\right)\right] - c\,e^{-\lambda c} = 0$$

e) Finally, we get with a), b), and c)

$$\int_0^c \left(\frac{1}{\lambda} - y\right)^2 \lambda e^{-\lambda y} dy + c^2\int_c^\infty \lambda e^{-\lambda y} dy$$
$$= \; \int_0^c \left(\frac{1}{\lambda^2} - \frac{2}{\lambda}y + y^2\right)\lambda e^{-\lambda y} dy + c^2\,e^{-\lambda c}$$
$$= \; \frac{1}{\lambda^2}\left(1 - e^{-\lambda c}\right) - \frac{2}{\lambda}\left[-c\,e^{-\lambda c} + \frac{1}{\lambda}\left(1 - e^{-\lambda c}\right)\right] + e^{-\lambda c}\left[-c^2 - \frac{2c}{\lambda} - \frac{2}{\lambda^2}\right] + \frac{2}{\lambda^2} + c^2\,e^{-\lambda c}$$
$$= \; \frac{1}{\lambda^2} - \frac{1}{\lambda^2}e^{-\lambda c} + \frac{2c}{\lambda}\,e^{-\lambda c} - \frac{2}{\lambda^2} + \frac{2}{\lambda^2}e^{-\lambda c} - c^2\,e^{-\lambda c} - \frac{2c}{\lambda}e^{-\lambda c} - \frac{2}{\lambda^2}e^{-\lambda c} + \frac{2}{\lambda^2} + c^2\,e^{-\lambda c}$$
$$= \; \frac{1}{\lambda^2}\left(1 - e^{-\lambda c}\right).\ \square$$

2.9.7 Lemma
Let be $T \sim \mathcal{E}(\lambda)$ and $Z = \min\{T, c\}$ (Typ I censoring). Then

a) $I_\lambda(T) = \frac{1}{\lambda^2}$,

b) $I_\lambda(Z) = \frac{1}{\lambda^2}(1 - e^{\lambda c})$.

Proof. a) We have

$$\frac{\partial}{\partial \lambda} f_\lambda(t) = \frac{\partial}{\partial \lambda} \lambda e^{-\lambda t} = e^{-\lambda t}(1-\lambda t) = f_\lambda(t)\left(\frac{1}{\lambda} - t\right)$$

so that

$$I_\lambda(T) = E_\lambda\left(\left[\frac{\partial}{\partial \lambda} \ln f_\lambda(T)\right]^2\right) = E_\lambda\left(\left[\frac{\frac{\partial}{\partial \lambda} f_\lambda(T)}{f_\lambda(T)}\right]^2\right) = E_\lambda\left(\left[\frac{1}{\lambda} - T\right]^2\right) = \frac{1}{\lambda^2},$$

since $E_\lambda(T) = \frac{1}{\lambda}$ and $\mathrm{var}_\lambda(T) = \frac{1}{\lambda^2}$.

b) In the censored case, we have

$$\ln l(\lambda, z) = \ln f_\lambda(z) 1\!\!1_{[0,c]}(t) + \ln S_\lambda(z) 1\!\!1_{(c,\infty)}(t) = (\ln \lambda - \lambda t) 1\!\!1_{[0,c]}(t) - \lambda c 1\!\!1_{(c,\infty)}(t)$$

so that

$$\frac{\partial}{\partial \lambda} \ln l(\lambda, z) = \left(\frac{1}{\lambda} - t\right) 1\!\!1_{[0,c]}(t) - c 1\!\!1_{(c,\infty)}(t).$$

Hence, Lemma 2.9.6 e) provides

$$I_\lambda(Z) = E_\lambda\left(\left[\frac{\partial}{\partial \lambda} \ln l(\lambda, Z)\right]^2\right) = \int_0^c \left(\frac{1}{\lambda} - t\right)^2 \lambda e^{-\lambda t} dt + \int_c^\infty c^2 \lambda e^{-\lambda t} dt = \frac{1}{\lambda^2}\left(1 - e^{-\lambda c}\right). \square$$

2.9.8 Lemma
If $Y = \log(Z) = \log(T) \sim \mathcal{N}(\mu, \sigma^2)$, *then*

$$I_{(\mu,\sigma^2)}(Y) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

Proof. We have

$$\frac{\partial}{\partial \mu} f_{(\mu,\sigma^2)}(y) = \frac{\partial}{\partial \mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} = f_{(\mu,\sigma^2)}(y)\left(\frac{y-\mu}{\sigma^2}\right)$$

and

$$\frac{\partial}{\partial \sigma^2} f_{(\mu,\sigma^2)}(y) = \frac{\partial}{\partial \sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} = f_{(\mu,\sigma^2)}(y)\left(-\frac{1}{2\sigma^2} + \frac{(y-\mu)^2}{2\sigma^4}\right)$$

so that

$$\begin{aligned} & E_{(\mu,\sigma^2)}\left(\left[\frac{\partial}{\partial \mu} \ln f_{(\mu,\sigma^2)}(Y)\right]^2\right) \\ = \; & E_{(\mu,\sigma^2)}\left(\left[\frac{\frac{\partial}{\partial \mu} f_{(\mu,\sigma^2)}(Y)}{f_{(\mu,\sigma^2)}(Y)}\right]^2\right) = E_{(\mu,\sigma^2)}\left(\left[\frac{y-\mu}{\sigma^2}\right]^2\right) = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}, \end{aligned}$$

$$\begin{aligned}
&E_{(\mu,\sigma^2)}\left(\left[\frac{\partial}{\partial\mu}\ln f_{(\mu,\sigma^2)}(Y)\right]\,\left[\frac{\partial}{\partial\sigma}\ln f_{(\mu,\sigma^2)}(Y)\right]\right)\\
&= E_{(\mu,\sigma^2)}\left(\left[\frac{\frac{\partial}{\partial\mu}f_{(\mu,\sigma^2)}(Y)}{f_{(\mu,\sigma^2)}(Y)}\right]\,\left[\frac{\frac{\partial}{\partial\sigma}f_{(\mu,\sigma^2)}(Y)}{f_{(\mu,\sigma^2)}(Y)}\right]\right)\\
&= E_{(\mu,\sigma^2)}\left(-\frac{(y-\mu)}{2\sigma^4}+\frac{(y-\mu)^3}{2\sigma^6}\right)=0,
\end{aligned}$$

$$\begin{aligned}
&E_{(\mu,\sigma^2)}\left(\left[\frac{\partial}{\partial\sigma^2}\ln f_{(\mu,\sigma^2)}(Y)\right]^2\right)\\
&= E_{(\mu,\sigma^2)}\left(\left[\frac{\frac{\partial}{\partial\sigma^2}f_{(\mu,\sigma^2)}(Y)}{f_{(\mu,\sigma^2)}(Y)}\right]^2\right)=E_{(\mu,\sigma^2)}\left(\left[-\frac{1}{2\sigma^2}+\frac{(y-\mu)^2}{2\sigma^4}\right]^2\right)\\
&= E_{(\mu,\sigma^2)}\left(\frac{1}{4\sigma^4}-\frac{(y-\mu)^2}{2\sigma^6}+\frac{(y-\mu)^4}{4\sigma^8}\right)=\frac{1}{4\sigma^4}-\frac{\sigma^2}{2\sigma^6}+\frac{3\sigma^4}{4\sigma^8}=\frac{1}{2\sigma^4}.
\end{aligned}$$

Hence the assertion follows. □

2.9.9 Theorem (Schervish 1997, Theorem 7.57 or Theorem 7.63)
If $Z_1,\dots,Z_N$ are i.i.d., $\widehat{\theta}_N := \widehat{\theta}_N(Z_1,\dots,Z_N)$ is the maximum likelihood estimator for θ, then under regularity conditions (see Schervish 1997, Theorem 7.57 or Theorem 7.63),

$$\begin{aligned}
&a)\quad \widehat{\theta}_N \longrightarrow \theta_*\ P_{\theta^*}\text{-almost surely},\\
&b)\quad \sqrt{N}(\widehat{\theta}_N-\theta_*)\xrightarrow{\mathcal{D}}\mathcal{N}(0_r, I_{\theta_*}(Z_1)^{-1}),
\end{aligned}\tag{2.7}$$

if θ_* is the true parameter, i.e. the maximum likelihood estimator converges in distribution to a normal distribution.

Proof.
a) At first we show $\widehat{\theta}_N \longrightarrow \theta_*$ P_{θ_*}-almost surely.
The strong law of large numbers provides

$$g_N(\theta,\omega) := \frac{1}{N}\sum_{n=1}^N \ln l(\theta, Z_n(\omega)) \longrightarrow E_{\theta_*}(\ln l(\theta, Z_1)) =: g(\theta)\ \text{ for all }\ \theta\in\Theta \tag{2.8}$$

for P_{θ_*}-almost all ω. Consider any ω satisfying (2.8). If the regularity conditions ensure that g has a unique maximum at $\theta_*\in\Theta$ and Θ is compact then we obtain the following properties. The compactness of Θ implies uniform convergence which means that for any $\delta>0$, there exists N_δ such that

$$|g_N(\theta,\omega)-g(\theta)|<\delta\ \text{ for all }\ N\ge N_\delta,\ \theta\in\Theta. \tag{2.9}$$

Now let $B_\epsilon(\theta_*)$ be an ϵ-ball around θ_*. The uniqueness of the maximum of g at θ_* provides the existence of an $\epsilon_0 > 0$ so that for all $0 < \epsilon < \epsilon_0$ a $\delta > 0$ exists with

$$\max_{\theta \in \Theta \setminus B_\epsilon(\theta_*)} g(\theta) < g(\theta_*) - 2\delta.$$

The uniform convergence in (2.9) implies then for $N \geq N_\delta$

$$\begin{array}{rcl} g_N(\theta,\omega) & \leq & |g_N(\theta,\omega) - g(\theta)| + g(\theta) < \delta + g(\theta_*) - 2\delta = g(\theta_*) - \delta \ \text{ for } \ \theta \in \Theta \setminus B_\epsilon(\theta_*) \\ g_N(\theta_*,\omega) & \geq & -|g_N(\theta_*,\omega) - g(\theta_*)| + g(\theta_*) > g(\theta_*) - \delta. \end{array}$$

This means

$$\widehat{\theta}_N = \arg\max \ln \prod_{n=1}^N l(\theta, Z_n(\omega)) = \arg\max g_N(\theta,\omega) \in B_\epsilon(\theta_*)$$

for all $N \geq N_\delta$ so that $\widehat{\theta}_N \longrightarrow \theta_*$ P_{θ_*}-almost surely.
b) To show $\sqrt{N}(\widehat{\theta}_N - \theta_*) \xrightarrow{\mathcal{D}} \mathcal{N}(0_r, I_{\theta_*}(Z_1)^{-1})$ we use a necessary condition for

$$\widehat{\theta}_N = \arg\max \sum_{n=1}^N \ln l(\theta, Z_n)$$

and the mean value theorem so that we get

$$0_r = \sum_{n=1}^N \frac{\partial}{\partial\theta} \ln l(\theta, Z_n)\Big|_{\theta=\widehat{\theta}_N} = \sum_{n=1}^N \frac{\partial}{\partial\theta} \ln l(\theta, Z_n)\Big|_{\theta=\theta_*} + \sum_{n=1}^N \frac{\partial^2}{\partial^2\theta} \ln l(\theta, Z_n)\Big|_{\theta=\widetilde{\theta}_N} (\widehat{\theta}_N - \theta_*)$$

where $\widetilde{\theta}_N = (1-\alpha_N)\theta_* + \alpha_N\widehat{\theta}_N$ with $\alpha_N \in (0,1)$. Hence we get

$$V_n := \sqrt{N}\frac{1}{N} \sum_{n=1}^N \frac{\partial}{\partial\theta} \ln l(\theta, Z_n)\Big|_{\theta=\theta_*} = -\frac{1}{N} \sum_{n=1}^N \frac{\partial^2}{\partial^2\theta} \ln l(\theta, Z_n)\Big|_{\theta=\widetilde{\theta}_N} \sqrt{N}(\widehat{\theta}_N - \theta_*)$$

so that

$$\sqrt{N}(\widehat{\theta}_N - \theta_*) = W_N^{-1} V_N$$

with

$$W_N := -\frac{1}{N} \sum_{n=1}^N \frac{\partial^2}{\partial^2\theta} \ln l(\theta, Z_n)\Big|_{\theta=\widetilde{\theta}_N}.$$

With $\widehat{\theta}_N \longrightarrow \theta_*$ P_{θ_*}-almost surely, we have also $\widetilde{\theta}_N \longrightarrow \theta_*$ P_{θ_*}-almost surely. The weak law of large numbers provides with Theorem 2.9.5

$$W_N^* := -\frac{1}{N}\sum_{n=1}^N \frac{\partial^2}{\partial^2\theta}\ln l(\theta,Z_n)\Big|_{\theta=\theta_*} \xrightarrow{\mathcal{P}} -E_\theta\left[\frac{\partial^2}{\partial^2\theta}\ln l(\theta,Z_1)\Big|_{\theta=\theta_*}\right] = I_{\theta_*}(Z_1)$$

so that with a regularity condition (for example $\frac{\partial^2}{\partial^2\theta}\ln l(\theta,z)$ is continuous in θ uniformly in z)

$$W_N = W_N - W_N^* + W_N^* \xrightarrow{\mathcal{P}} I_{\theta_*}(Z_1). \tag{2.10}$$

Again with the regularity assumption that differentiation and expectation can be exchanged, we get

$$E_{\theta_*}\left[\frac{\partial}{\partial\theta}\ln l(\theta,Z_n)\Big|_{\theta=\theta_*}\right] = E_{\theta_*}\left[\frac{\frac{\partial}{\partial\theta}l(\theta,Z_n)\big|_{\theta=\theta_*}}{l(\theta,Z_n)}\right] = \frac{\partial}{\partial\theta}\,E_{\theta_*}\left[\frac{l(\theta,Z_n}{l(\theta,Z_n}\right]\Big|_{\theta=\theta_*} = 0_r$$

and for any $u \in \mathbb{R}^r$

$$\mathsf{Var}_{\theta_*}\left[u^\top\frac{\partial}{\partial\theta}\ln l(\theta,Z_n)\Big|_{\theta=\theta_*}\right] = E_{\theta_*}\left[\left(u^\top\frac{\partial}{\partial\theta}\ln l(\theta,Z_n)\Big|_{\theta=\theta_*}\right)^2\right] = u^\top I_{\theta_*}(Z_1)\,u.$$

Hence the central limit theorem provides

$$\sqrt{N}\frac{1}{N}\sum_{n=1}^N \frac{u^\top\frac{\partial}{\partial\theta}\ln l(\theta,Z_n)\big|_{\theta=\theta_*}}{\sqrt{u^\top I_{\theta_*}(Z_1)\,u}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

and thus

$$\sqrt{N}\frac{1}{N}\sum_{n=1}^N u^\top\frac{\partial}{\partial\theta}\ln l(\theta,Z_n)\Big|_{\theta=\theta_*} \xrightarrow{\mathcal{D}} \mathcal{N}(0,u^\top I_{\theta_*}(Z_1)\,u)$$

or, respectively, with the Cramér-Wold device of Theorem 2.8.9

$$V_n = \sqrt{N}\frac{1}{N}\sum_{n=1}^N \frac{\partial}{\partial\theta}\ln l(\theta,Z_n)\Big|_{\theta=\theta_*} \xrightarrow{\mathcal{D}} \mathcal{N}_r(0_r,I_{\theta_*}(Z_1))$$

With Corollary 2.8.8 we get finally with (2.10)

$$\sqrt{N}(\widehat{\theta}_N - \theta_*) = W_N^{-1}V_N \xrightarrow{\mathcal{D}} \mathcal{N}_r(0_r,I_{\theta_*}(Z_1)^{-1}).\square$$

If $a:\Theta\to\mathcal{A}$ is an aspect function, then $a(\widehat{\theta})$ is a maximum likelihood estimator for $a(\theta)$ if $\widehat{\theta}$ is the maximum likelihood estimator for θ. The delta method, Theorem 2.8.12, provides at once the asymptotic distribution of $\sqrt{N}(a(\widehat{\theta})-a(\theta_*))$.

2.9.10 Corollary
If $Z_1, \ldots, Z_N$ are i.i.d., $\widehat{\theta}_N := \widehat{\theta}_N(Z_1, \ldots, Z_N)$ is the maximum likelihood estimator for $\theta \in \Theta \subset \mathbb{R}^r$, $a : \Theta \to \mathbb{R}^s$ is a differentiable aspect function with derivative $\dot{a}(\theta_*) = \frac{\partial}{\partial \theta} a(x)\Big|_{\theta=\theta_*} \in \mathbb{R}^{r \times s}$, then under regularity conditions

$$\sqrt{N}(a(\widehat{\theta}_N) - a(\theta_*)) \xrightarrow{\mathcal{D}} \mathcal{N}(0_s, \dot{a}(\theta_*)^\top I_{\theta_*}(Z_1)^{-1} \dot{a}(\theta_*)), \tag{2.11}$$

if θ_* is the true parameter.

Proof. Since

$$\sqrt{N}(\widehat{\theta}_N - \theta_*) \xrightarrow{\mathcal{D}} Y \sim \mathcal{N}_r(0_r, I_{\theta_*}(Z_1)^{-1}),$$

the δ method, Theorem 2.8.12, provides

$$\sqrt{N}(a(\widehat{\theta}_N) - a(\theta_*)) \xrightarrow{\mathcal{D}} \dot{a}(\theta_*)^\top Y \sim \mathcal{N}(0_s, \dot{a}(\theta_*)^\top I_{\theta_*}(Z_1)^{-1} \dot{a}(\theta_*)).\Box$$

2.9.11 Definition

$$I_{a(\theta)}(Z) := \left[\dot{a}(\theta)^\top I_\theta(Z)^{-1} \dot{a}(\theta)\right]^{-1}$$

is called the Fisher information of Z for $a(\theta)$.

2.9.12 Theorem (Wald-type confidence set)
If $Z_1, \ldots, Z_N$ are i.i.d., $\widehat{\theta}_N := \widehat{\theta}_N(z_1, \ldots, z_N)$ is the maximum likelihood estimator for $\theta \in \mathbb{R}^r$, $a : \mathbb{R}^r \to \mathbb{R}^s$ is a differentiable aspect function, then under regularity conditions, $\mathbb{C}_N$ given by

$$\mathbb{C}_N(z_1, \ldots, z_N) = \left\{a(\theta);\ N\left(a(\widehat{\theta}_N) - a(\theta)\right)^\top I_{a(\widehat{\theta}_N)}(Z_1)\left(a(\widehat{\theta}_N) - a(\theta)\right) \le \chi^2_{s;1-\alpha}\right\}$$

is an asymptotic $(1-\alpha)$-confidence set function for $a(\theta)$.

Proof. Let be $\theta_* \in \Theta$ arbitrary. Since $I_{a(\theta_*)}(Z_1)$ is a symmetric matrix, $I_{a(\theta_*)}(Z_1)^{1/2}$ exists with $I_{a(\theta_*)}(Z_1) = \left(I_{a(\theta_*)}(Z_1)^{1/2}\right)^\top I_{a(\theta_*)}(Z_1)^{1/2}$ and $I_{a(\theta_*)}(Z_1)^{1/2}\ I_{a(\theta_*)0}(Z_1)^{-1}\ \left(I_{a(\theta_*)}(Z_1)^{1/2}\right)^\top = I_{s\times s}$, where $I_{s\times s}$ is the $s \times s$ identity matrix. The convergence in (2.11) implies with Corollary 2.8.7 (iv) that $a(\widehat{\theta}_N)$ converges to $a(\theta_*)$ if θ_* is the true parameter. If the regularity conditions mean that $I_{a(\theta_*)}(Z_1)$ and thus $I_{a(\theta_*)}(Z_1)^{1/2}$ is continuous in $a(\theta)$, then also $I_{a(\widehat{\theta}_N)}(Z_1)^{1/2}$ converges in probability to $I_{a(\theta_*)}(Z_1)^{1/2}$ if θ_* is the true parameter. Hence

$$\sqrt{N}(a(\widehat{\theta}_N) - a(\theta_*)) \xrightarrow{\mathcal{D}} X \sim \mathcal{N}(0_s, I_{a(\theta_*)}(Z_1)^{-1})$$

implies with Corollary 2.8.8 (ii)

$$\begin{aligned}
&\sqrt{N}\, I_{a(\widehat{\theta}_N)}(Z_1)^{1/2}(a(\widehat{\theta}_N) - a(\theta_*)) \xrightarrow{\mathcal{D}} V := I_{a(\theta_*)}(Z_1)^{1/2}\, X \\
&\quad \sim \mathcal{N}\left(0_s, I_{a(\theta_*)}(Z_1)^{1/2}\ I_{a(\theta_*)}(Z_1)^{-1}\ \left(I_{a(\theta_*)}(Z_1)^{1/2}\right)^\top\right) = \mathcal{N}(0_s, I_{s\times s}).
\end{aligned}$$

Sine $f : \mathbb{R}^r \ni u \to f(u) = u^\top u \in \mathbb{R}$ is a continuous function, the Continuous Mapping Theorem (Theorem 2.8.11) provides then

$$N(a(\widehat{\theta}_N) - a(\theta_*))^\top I_{a(\widehat{\theta}_N)}(Z_1)\,(a(\widehat{\theta}_N) - a(\theta_*)) \xrightarrow{\mathcal{D}} V^\top V \sim \chi^2_s,$$

where χ^2_r is the χ^2-distribution with r degrees of freedom. Hence we obtain

$$\begin{aligned}
&\lim_{N\to\infty} P_{\theta_*}(a(\theta_*) \in \mathbb{C}_N(t_1,\dots,t_N))\\
=& \lim_{N\to\infty} P_{\theta_*}\left(N(a(\widehat{\theta}_N) - a(\theta_*))^\top I_{a(\widehat{\theta}_N)}(Z_1)\,(a(\widehat{\theta}_N) - a(\theta_*)) \le \chi^2_{s;1-\alpha}\right)\\
=& P_{\theta_*}\left(V^\top V \le \chi^2_{s;1-\alpha}\right) = 1-\alpha. \square.
\end{aligned}$$

2.9.13 Corollary (One dimensional aspect)
If $Z_1,\dots,Z_N$ are i.i.d., $\widehat{\theta}_N := \widehat{\theta}_N(z_1,\dots,z_N)$ is the maximum likelihood estimator for $\theta \in \mathbb{R}^r$, $a : \mathbb{R}^r \to \mathbb{R}$ is a differentiable aspect function, then under regularity conditions, $\mathbb{C}_N$ given by

$$\mathbb{C}_N(z_1,\dots,z_N) = \left[\, a(\widehat{\theta}_N) - \sqrt{\frac{1}{N}\, I_{a(\widehat{\theta}_N)}(Z_1)^{-1}\chi^2_{1;1-\alpha}}\;,\; a(\widehat{\theta}_N) + \sqrt{\frac{1}{N}\, I_{a(\widehat{\theta}_N)}(Z_1)^{-1}\chi^2_{1;1-\alpha}}\,\right]$$

is an asymptotic $(1-\alpha)$-confidence interval function for $a(\theta)$.

Note that we have $\sqrt{\chi^2_{1;1-\alpha}} = q_{1-\alpha/2}$ where $q_{1-\alpha/2}$ is the $1-\alpha/2$-quantile of the standard normal distribution.

2.9.14 Corollary (One-sided confidence intervals)
If $Z_1,\dots,Z_N$ are i.i.d., $\widehat{\theta}_N := \widehat{\theta}_N(z_1,\dots,z_N)$ is the maximum likelihood estimator for $\theta \in \mathbb{R}^r$, $a : \mathbb{R}^r \to \mathbb{R}$ is a differentiable aspect function, then under regularity conditions, $\mathbb{C}^u_N$ and $\mathbb{C}^l_N$ given by

$$\mathbb{C}^l_N(z_1,\dots,z_N) = \left[\, a(\widehat{\theta}_N) - \sqrt{\frac{1}{N}\, I_{a(\widehat{\theta}_N)}(Z_1)^{-1}}\;\; q_{1-\alpha}\;,\;\infty\right]$$

$$\mathbb{C}^u_N(z_1,\dots,z_N) = \left[-\infty\;,\; a(\widehat{\theta}_N) + \sqrt{\frac{1}{N}\, I_{a(\widehat{\theta}_N)}(Z_1)^{-1}}\; q_{1-\alpha}\right]$$

are asymptotic one-sided $(1-\alpha)$-confidence interval function for $a(\theta)$.

2.9.15 Corollary
If $Z_1,\dots,Z_N$ are i.i.d., $\widehat{\theta}_N := \widehat{\theta}_N(z_1,\dots,z_N)$ is the maximum likelihood estimator for $\theta \in \mathbb{R}^r$, then under regularity conditions, $\mathbb{C}_N$ given by

$$\mathbb{C}_N(z_1,\dots,z_N) = \left\{\theta;\; N\,\left(\widehat{\theta}_N - \theta\right)^\top I_{\widehat{\theta}_N}(Z_1)\left(\widehat{\theta}_N - \theta\right) \le \chi^2_{r;1-\alpha}\right\}$$

is an asymptotic $(1-\alpha)$-confidence set function for θ.

2.9.16 Example ((Log)normal distribution: Confidence set for (μ, σ^2))
At first, we check that $g(\mu,\sigma^2) := E_{(\mu_*,\sigma_*^2)}\left(\ln(f_{(\mu,\sigma^2)}(Y)\right)$ has a unique maximum at (μ_*, σ_*^2). To see this, note that

$$\begin{aligned}
g(\mu,\sigma^2) &= E_{(\mu_*,\sigma_*^2)}\left(\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}(Y-\mu)^2\right)\\
&= E_{(\mu_*,\sigma_*^2)}\left(\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}(Y-\mu_*+\mu_*-\mu)^2\right)\\
&= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}E_{(\mu_*,\sigma_*^2)}\left((Y-\mu_*)^2+2(Y-\mu_*)(\mu_*-\mu)+(\mu_*-\mu)^2\right)\\
&= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\sigma_*^2+(\mu_*-\mu)^2)\\
&\overset{(*)}{\leq} -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\frac{\sigma_*^2}{\sigma^2} = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\left(\ln(\sigma^2)+\frac{\sigma_*^2}{\sigma^2}\right),
\end{aligned}$$

where "$<$" is satisfied in $(*)$ if and only if $\mu_* \neq \mu$. The function f given by $f(\sigma^2) := \ln(\sigma^2) + \frac{\sigma_*^2}{\sigma^2}$ has a unique minimum at σ_*^2 because of

$$f'(\sigma^2) = \frac{1}{\sigma^2} - \frac{\sigma_*^2}{\sigma^4} = \frac{1}{\sigma^2}\left(1-\frac{\sigma_*^2}{\sigma^2}\right) = 0 \;\Leftrightarrow \sigma^2 = \sigma_*^2$$

and

$$f''(\sigma^2) = -\frac{1}{\sigma^4} + 2\frac{\sigma_*^2}{\sigma^6} \overset{\sigma^2=\sigma_*^2}{=} \frac{1}{\sigma_*^4} > 0.$$

Hence g has a unique maximum at (μ_*, σ_*^2).

According to Corollary 2.9.15, an asymptotic $(1-\alpha)$-confidence interval function for (μ,σ^2) is given by

$$\mathbb{C}_N(y_1,\ldots,y_N) = \left\{(\mu,\sigma^2);\; N\left(\frac{(\widehat{\mu}-\mu)^2}{\widehat{\sigma}^2} + \frac{(\widehat{\sigma}^2-\sigma^2)^2}{2\widehat{\sigma}^4}\right) \leq \chi^2_{2;1-\alpha}\right\}$$

where $(\widehat{\mu}, \widehat{\sigma}^2)$ is the maximum likelihood estimator for (μ,σ^2). This follows with Lemma 2.9.8 from

$$\left(\widehat{\theta}-\theta\right)^\top I_{\widehat{\theta}}(Y)\left(\widehat{\theta}-\theta\right) = \begin{pmatrix} \widehat{\mu}-\mu \\ \widehat{\sigma}^2-\sigma^2 \end{pmatrix}^\top \begin{pmatrix} \frac{1}{\widehat{\sigma}^2} & 0 \\ 0 & \frac{1}{2\widehat{\sigma}^4} \end{pmatrix} \begin{pmatrix} \widehat{\mu}-\mu \\ \widehat{\sigma}^2-\sigma^2 \end{pmatrix}.$$

2.9.17 Example ((Log)normal distribution: Confidence set for μ)
We have $a(\theta) = a(\mu,\sigma^2) = \mu$ so that $\dot{a}(\theta) = \binom{1}{0}$ which implies with Lemma 2.9.8

$$I_{a(\theta)}(Y)^{-1} = \dot{a}(\theta)^\top I_\theta(Y)^{-1}\dot{a}(\theta) = (1,0)I_{(\mu,\sigma^2)}(Y)^{-1}\begin{pmatrix}1\\0\end{pmatrix} = (1,0)\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}^{-1}\begin{pmatrix}1\\0\end{pmatrix} = \sigma^2.$$

Hence Corollary 2.9.13 provides that

$$\mathbb{C}_N(y_1,\ldots,y_N) = \left[\,\widehat{\mu} - \sqrt{\frac{1}{N}\,\widehat{\sigma}^2\chi^2_{1;1-\alpha}}\;,\;\widehat{\mu} + \sqrt{\frac{1}{N}\,\widehat{\sigma}^2\chi^2_{1;1-\alpha}}\,\right]$$

$$= \left[\,\widehat{\mu} - \frac{\widehat{\sigma}}{\sqrt{N}}\,q_{1-\alpha/2}\;,\;\widehat{\mu} + \frac{\widehat{\sigma}}{\sqrt{N}}\,q_{1-\alpha/2}\,\right]$$

is an asymptotic $(1-\alpha)$-confidence interval function for μ if $(\widehat{\mu},\widehat{\sigma}^2)$ is the maximum likelihood estimator for (μ,σ^2).

The statistic $N(\widehat{\theta}_N-\theta_*)^\top I_{\widehat{\theta}_N}(Z_1)\,(\widehat{\theta}_N-\theta_*)$ can be used to test $H_0:\theta=\theta_*$ against $H_1:\theta\neq\theta_*$. This test known as Wald test. An alternative test for testing this hypotheses is given by the likelihood ratio test based on the test statistic

$$L_N(\theta_*) := L_N(z_1,\ldots,z_N) := \frac{l(\theta_*,(z_1,\ldots,z_N))}{l(\widehat{\theta}_N,(z_1,\ldots,z_N))},$$

where $\widehat{\theta}_N := \widehat{\theta}_N(z_1,\ldots,z_N)$ is again the maximum likelihood estimator for θ and $l(\theta,(z_1,\ldots,z_N)) := \prod_{n=1}^N l(\theta,z_n)$.

2.9.18 Theorem (Likelihood-ratio confidence set)
If $Z_1,\ldots,Z_N$ are i.i.d., $\widehat{\theta}_N := \widehat{\theta}_N(z_1,\ldots,z_N)$ is the maximum likelihood estimator for $\theta\in\mathbb{R}^r$, then under regularity conditions, $\mathbb{C}_N$ given by

$$\mathbb{C}_N(z_1,\ldots,z_N) = \left\{\theta; -2\,\ln(L_N(\theta)) \le \chi^2_{r;1-\alpha}\right\}$$

is an asymptotic $(1-\alpha)$-confidence set function for θ.

Proof. See for example Schervish (1997), Section 7.5.1. Set $l_N(\theta) := \ln l(\theta,(z_1,\ldots,z_N)) = \sum_{n=1}^N \ln l(\theta,z_n)$. Then Taylor expansion provides for arbitrary $\theta_*\in\Theta$

$$l_N(\theta_*) = l_N(\widehat{\theta}_N) + (\theta_*-\widehat{\theta}_N)^\top \frac{\partial}{\partial\theta} l_N(\theta)\big|_{\theta=\widehat{\theta}_N} + \frac{1}{2}(\theta_*-\widehat{\theta}_N)^\top \frac{\partial^2}{\partial^2\theta} l_N(\theta)\big|_{\theta=\widetilde{\theta}_N}(\theta_*-\widehat{\theta}_N)$$

where $\widetilde{\theta}_N = \theta_* + \alpha_N(\widehat{\theta}_N-\theta_*)$ for some $\alpha_N\in(0,1)$. The definition of a maximum likelihood estimator means that $\frac{\partial}{\partial\theta} l_N(\theta)\big|_{\theta=\widehat{\theta}_N} = 0_r$ so that

$$-2\ln(L_N(\theta_*)) = -2\left(l_N(\theta_*) - l_N(\widehat{\theta}_N)\right) = -(\theta_*-\widehat{\theta}_N)^\top \frac{\partial^2}{\partial^2\theta} l_N(\theta)\big|_{\theta=\widetilde{\theta}_N}(\theta_*-\widehat{\theta}_N).$$

The law of large numbers and Theorem 2.9.5 imply

$$\frac{1}{N}\frac{\partial^2}{\partial^2\theta} l_N(\theta)\big|_{\theta=\theta_*} = \frac{1}{N}\sum_{n=1}^N \frac{\partial^2}{\partial^2\theta}\ln l(\theta_*,Z_n)\big|_{\theta=\theta_*} \xrightarrow{\mathcal{P}} E_{\theta_*}\left[\frac{\partial^2}{\partial^2\theta}\ln l(\theta,Z_1)\big|_{\theta=\theta_*}\right] = -I_{\theta_*}(Z_1).$$

According to Theorem 2.9.9 a), $\widehat{\theta}_N$ converges to θ_* in probability if θ_* is the true parameter. With $\widehat{\theta}_N$ also $\widetilde{\theta}_N$ converges to θ_* in probability. Completely analog to the proof of Theorem 2.9.12 we obtain

$$\frac{1}{N}\frac{\partial^2}{\partial^2\theta} l_N(\theta)\big|_{\theta=\widetilde{\theta}_N} \xrightarrow{\mathcal{P}} -I_{\theta_*}(Z_1).$$

and

$$(\theta_* - \widehat{\theta})^\top \left(-\frac{\partial^2}{\partial^2 \theta} l_N(\theta)\big|_{\theta=\widetilde{\theta}_N} \right) (\theta_* - \widehat{\theta}) \xrightarrow{\mathcal{D}} \chi^2_r . \square$$

2.9.19 Theorem (Prediction intervals based on confidence sets)
If $T_0, T_1, \ldots, T_N$ are independent distributed with cumulative distribution function F_θ, $0 \le \eta_1 < \eta_2 \le 1$ with $\eta_2 - \eta_1 = 1 - \alpha_1$, and $\mathbb{C}_N$ is an asymptotic $(1-\alpha_2)$-confidence set function based on $z_n = \min(t_n, c)$, $n = 1, \ldots, N$, then $\mathbb{P}_N$ given by

$$\mathbb{P}_N(z_1, \ldots, z_N) = \bigcup_{\theta \in \mathbb{C}_N(z_1,\ldots,z_N)} \left[F_\theta^{-1}(\eta_1), F_\theta^{-1}(\eta_2)\right]$$

is an asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval function for T_0.

Proof. The proof is the same as for Theorem 2.7.6.

Another type of prediction intervals can be constructed by one-sided asymptotic confidence intervals for $a_1(\theta) = F_\theta^{-1}(\eta_1)$ and $a_1(\theta) = F_\theta^{-1}(\eta_1)$ obtained with the δ-method.

2.9.20 Theorem (Prediction intervals based on the δ-method)
Let $T_0, T_1, \ldots, T_N$ be i.i.d. with distribution function F_θ, $Z_n = \min(T_n, c)$ for $n = 1, \ldots, N$, $\widehat{\theta}_N := \widehat{\theta}_N(Z_1, \ldots, Z_N)$ is the maximum likelihood estimator for $\theta \in \mathbb{R}^r$, $0 \le \eta_1 < \eta_2 \le 1$ with $\eta_2 - \eta_1 = 1 - \alpha_1$, $a_1, a_2 : \mathbb{R}^r \to \mathbb{R}$ with $a_1(\theta) = F_\theta^{-1}(\eta_1)$ and $a_2(\theta) = F_\theta^{-1}(\eta_2)$ are differentiable aspect functions, and

$$\begin{aligned} v_1 &:= \sqrt{\frac{1}{N} I_{a_1(\widehat{\theta}_N)}(Z_1)^{-1}}\, q_{1-\alpha_2/2}, \\ v_2 &:= \sqrt{\frac{1}{N} I_{a_2(\widehat{\theta}_N)}(Z_1)^{-1}}\, q_{1-\alpha_2/2}. \end{aligned}$$

Then $\mathbb{P}_N$ given by

$$\mathbb{P}_N(z_1, \ldots, z_N) = \left[F_{\widehat{\theta}_N}^{-1}(\eta_1) - v_1 \;\;,\;\; F_{\widehat{\theta}_N}^{-1}(\eta_2) + v_2\right]$$

is an asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval function for T_0.

Proof. At first note

$$\begin{aligned}
&P_\theta\left(T_0 \in \mathbb{P}_N(Z_1, \ldots, Z_N)\right) = P_\theta\left(F_{\widehat{\theta}_N}^{-1}(\eta_1) - v_1 \le T_0 \le F_{\widehat{\theta}_N}^{-1}(\eta_2) + v_2\right) \\
&\quad\ge\;\; P_\theta\left(T_0 \ge F_\theta^{-1}(\eta_1) \wedge F_\theta^{-1}(\eta_1) \ge F_{\widehat{\theta}_N}^{-1}(\eta_1) - v_1 \wedge T_0 \le F_\theta^{-1}(\eta_2) \wedge F_\theta^{-1}(\eta_2) \le F_{\widehat{\theta}_N}^{-1}(\eta_2) + v_2\right) \\
&\quad\overset{(*)}{=}\;\; P_\theta\left(F_\theta^{-1}(\eta_1) \le T_0 \le F_\theta^{-1}(\eta_2)\right)\; P_\theta\left(F_\theta^{-1}(\eta_1) \ge F_{\widehat{\theta}_N}^{-1}(\eta_1) - v_1 \wedge F_\theta^{-1}(\eta_2) \le F_{\widehat{\theta}_N}^{-1}(\eta_2) + v_2\right) \\
&\quad=\;\; (\eta_2 - \eta_1)\;\left(1 - P_\theta\left(F_\theta^{-1}(\eta_1) < F_{\widehat{\theta}_N}^{-1}(\eta_1) - v_1 \vee F_\theta^{-1}(\eta_2) > F_{\widehat{\theta}_N}^{-1}(\eta_2) + v_2\right)\right) \\
&\quad\ge\;\; (1 - \alpha_1)\;\left(1 - P_\theta\left(F_\theta^{-1}(\eta_1) < F_{\widehat{\theta}_N}^{-1}(\eta_1) - v_1\right) - P_\theta\left(F_\theta^{-1}(\eta_2) > F_{\widehat{\theta}_N}^{-1}(\eta_2) + v_2\right)\right),
\end{aligned}$$

where $(*)$ holds because of the independence of T_0 and $\widehat{\theta}_N(Z_1,\ldots,Z_N)$. Corollary 2.9.14 provides

$$\lim_{N\to\infty} P_\theta\left(F_\theta^{-1}(\eta_1) \geq F_{\widehat{\theta}_N}^{-1}(\eta_1) - v_1\right) \geq 1 - \frac{\alpha_2}{2},$$
$$\lim_{N\to\infty} P_\theta\left(F_\theta^{-1}(\eta_2) \leq F_{\widehat{\theta}_N}^{-1}(\eta_2) + v_2\right) \geq 1 - \frac{\alpha_2}{2},$$

so that

$$\begin{aligned}&\lim_{N\to\infty} P_\theta\left(T_0 \in \mathbb{P}_N(Z_1,\ldots,Z_N)\right)\\ \geq\ & (1-\alpha_1)\left(1 - \frac{\alpha_2}{2} - \frac{\alpha_2}{2}\right) = (1-\alpha_1)(1-\alpha_2).\Box\end{aligned}$$

2.9.21 Example (Asymptotic prediction intervals for the exponential distribution)
Here we have $T_n \sim \mathcal{E}(\lambda)$.
1. Asymptotic prediction interval based on the δ-method
Since

$$\begin{aligned}F_\lambda^{-1}(\eta) &= \inf\{t;\ F_\lambda(t) \geq \eta\} = \inf\{t;\ 1 - e^{-\lambda t} \geq \eta\} = \inf\{t;\ 1-\eta \geq e^{-\lambda t}\}\\ &= \inf\{t;\ 1-\eta \geq e^{-\lambda t}\} = \inf\{t;\ -\lambda t \leq \ln(1-\eta)\} = \frac{-\ln(1-\eta)}{\lambda} = a(\lambda)\end{aligned}$$

and $I_\lambda(T_1) = \frac{1}{\lambda^2}$ according to Lemma 2.9.7 a), we obtain

$$I_{a(\lambda)}(T_1)^{-1} = \dot{a}(\lambda)^\top I_\lambda(T_1)^{-1}\dot{a}(\lambda) = \frac{\ln(1-\eta)^2}{\lambda^4}\,\lambda^2 = \frac{\ln(1-\eta)^2}{\lambda^2}.$$

Setting for example $\eta_1 = \frac{\alpha_1}{2}$ and $\eta_2 = 1 - \frac{\alpha_1}{2}$ so that $a_1(\lambda) = F_\lambda^{-1}(\eta_1)$, $a_2(\lambda) = F_\lambda^{-1}(\eta_2)$, then the asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval for T_0 based on the δ-method is given by

$$\begin{aligned}&\mathbb{P}_N(t_1,\ldots,t_N)\\ =\ & \left[F_{\widehat{\lambda}_N}^{-1}(\eta_1) - \sqrt{\frac{1}{N}\,I_{a_1(\widehat{\lambda}_N)}(T_1)^{-1}}\,q_{1-\alpha_2/2}\ ,\ F_{\widehat{\lambda}_N}^{-1}(\eta_2) + \sqrt{\frac{1}{N}\,I_{a_2(\widehat{\lambda}_N)}(T_1)^{-1}}\,q_{1-\alpha_2/2}\right]\\ =\ & \left[\frac{-\ln(1-\eta_1)}{\widehat{\lambda}_N} - \frac{|\ln(1-\eta_1)|}{\sqrt{N}\widehat{\lambda}_N}\,q_{1-\alpha_2/2}\ ,\ \frac{-\ln(1-\eta_1)}{\widehat{\lambda}_N} + \frac{|\ln(1-\eta_2)|}{\sqrt{N}\widehat{\lambda}_N}\,q_{1-\alpha_2/2}\right]\\ =\ & \left[\frac{-\ln(1-\eta_1)}{\widehat{\lambda}_N}\left(1 - \frac{q_{1-\alpha_2/2}}{\sqrt{N}}\right)\ ,\ \frac{-\ln(1-\eta_2)}{\widehat{\lambda}_N}\left(1 + \frac{q_{1-\alpha_2/2}}{\sqrt{N}}\right)\right]\end{aligned}$$

where $\widehat{\lambda}_N$ is the ML estimate. If $\alpha_2 = \alpha_{2N}$ converges to zero so that

$$\frac{q_{1-\alpha_2/2}}{\sqrt{N}} \overset{N\to\infty}{\longrightarrow} 0 \tag{2.12}$$

then the constistency of the ML estimate leads to

$$\mathbb{P}_N(t_1,\ldots,t_N) \overset{N\to\infty}{\longrightarrow} \left[F_{\lambda^*}^{-1}(\eta_1)\ ,\ F_{\lambda^*}^{-1}(\eta_2)\right] \tag{2.13}$$

which would be used if the underlying parameter λ_* is known.

2. Asymptotic prediction interval based on the Wald-type confidence set
Since $I_\lambda(T_1) = \frac{1}{\lambda^2}$ according to Lemma 2.9.7, we get with Corollary 2.9.15 that

$$\begin{aligned}
\mathbb{C}_N(t_1,\ldots,t_N) &= \left\{\lambda;\ N(\widehat{\lambda}_N-\lambda)I_{\widehat{\lambda}_N}(T_1)(\widehat{\lambda}_N-\lambda)\le \chi^2_{1;1-\alpha_2}\right\}\\
&= \left\{\lambda;\ \frac{(\widehat{\lambda}_N-\lambda)^2}{\widehat{\lambda}_N^2}\le\frac{\chi^2_{1;1-\alpha_2}}{N}\right\} = \left\{\lambda;\ -\widehat{\lambda}_N\sqrt{\frac{\chi^2_{1;1-\alpha}}{N}}\le\widehat{\lambda}_N-\lambda\le\widehat{\lambda}_N\sqrt{\frac{\chi^2_{1;1-\alpha_2}}{N}}\right\}\\
&= \left\{\lambda;\ \widehat{\lambda}_N-\frac{\widehat{\lambda}_N}{\sqrt{N}}q_{1-\alpha_2/2}\le\lambda\le\widehat{\lambda}_N+\frac{\widehat{\lambda}_N}{\sqrt{N}}q_{1-\alpha_2/2}\right\} = \left[\widehat{\lambda}_l\ ,\ \widehat{\lambda}_u\right]
\end{aligned}$$

with $\widehat{\lambda}_l := \widehat{\lambda}_N - \frac{\widehat{\lambda}_N}{\sqrt{N}}q_{1-\alpha_2/2}$ and $\widehat{\lambda}_u := \widehat{\lambda}_N + \frac{\widehat{\lambda}_N}{\sqrt{N}}q_{1-\alpha_2/2}$ is an asymptotic Wald-type $(1-\alpha_2)$-confidence interval for λ. Since the cumulative distribution F_λ of the exponential distribution is an increasing function of λ and thus F_λ^{-1} is an decreasing function of λ, we get that

$$\mathbb{P}_N(t_1,\ldots,t_N) = \left[\ F^{-1}_{\widehat{\lambda}_u}\left(\frac{\alpha_1}{2}\right)\ ,\ F^{-1}_{\widehat{\lambda}_l}\left(1-\frac{\alpha_1}{2}\right)\ \right],$$

is an asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval for T_0. Here again, the convergence (2.12) implies the convergence (2.13) of the prediction interval.

3. Asymptotic prediction interval based on the likelihood-ratio confidence set
The asymptotic prediction interval can be also created with Theorem 2.9.18 via the likelihood ratio statistics. Since $\widehat{\lambda}_N = \frac{N}{\sum_{n=1}^N t_n}$ and thus

$$l(\widehat{\lambda}_N,(t_1,\ldots,t_N)) = \prod_{n=1}^N \widehat{\lambda}_N\ e^{-\widehat{\lambda}_N t_n} = \widehat{\lambda}_N^N e^{-\widehat{\lambda}_N\sum_{n=1}^N t_n} = \widehat{\lambda}_N^N e^{-N}$$

we get

$$\begin{aligned}
\ln(L_N(\lambda)) &= \ln\left(\frac{\lambda^N\exp(-\lambda\sum_{n=1}^N t_n)}{\widehat{\lambda}_N^N\exp(-N)}\right)\\
&= \ln\left(\lambda^N\left(\frac{\sum_{n=1}^N t_n}{N}\right)^N\exp\left(-\lambda\sum_{n=1}^N t_n+N\right)\right) = N\ln\left(\lambda\frac{\sum_{n=1}^N t_n}{N}\right)-\lambda\sum_{n=1}^N t_n+N.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{C}_N(t_1,\ldots,t_N) &= \left\{\lambda;\ 2\lambda\sum_{n=1}^N t_n - 2N\left(1+\ln\left(\lambda\frac{\sum_{n=1}^N t_n}{N}\right)\right)\le\chi^2_{1;1-\alpha_2}\right\}\\
&= \left\{\lambda;\ 2N\left[\frac{\lambda}{\widehat{\lambda}_N}-1-\ln\left(\frac{\lambda}{\widehat{\lambda}_N}\right)\right]\le\chi^2_{1;1-\alpha_2}\right\} = \left\{\lambda;\ \frac{\lambda}{\widehat{\lambda}_N}-1-\ln\left(\frac{\lambda}{\widehat{\lambda}_N}\right)\le\frac{\chi^2_{1;1-\alpha_2}}{2N}\right\}
\end{aligned}$$

is an asymptotic $(1-\alpha_2)$-confidence set for λ. The asymptotic prediction interval is given by Theorem 2.9.19. However, in this case, a more explicit version is not possible. However, since $\frac{\chi^2_{1;1-\alpha_2}}{N} = \left(\frac{q_{1-\alpha_2/2}}{\sqrt{N}}\right)^2$, the convergence (2.12) implies again that the confidence set shrinks to the underlying parameter λ_* and hence the convergence (2.13) of the corresponding prediction interval follows as well.

2.9.22 Example (Asymptotic prediction interval for the (log)normal distribution)
If $Y_n = \ln(T_n) \sim \mathcal{N}(\mu,\sigma^2)$ then $F_{(\mu,\sigma^2)}(y) = F_{(0,1)}\left(\frac{y-\mu}{\sigma}\right)$ so that

$$\begin{aligned}
F^{-1}_{(\mu,\sigma^2)}(\eta) &= \inf\{z;\ F_{(\mu,\sigma^2)}(z) \geq \eta\} = \inf\left\{z;\ F_{(0,1)}\left(\frac{z-\mu}{\sigma}\right) \geq \eta\right\} \\
&\overset{z=\sigma u+\mu}{=} \inf\left\{\sigma u + \mu;\ F_{(0,1)}(u) \geq \eta\right\} = \sigma F^{-1}_{(0,1)}(\eta) + \mu = \sqrt{\sigma^2}\, q_\eta + \mu = a((\mu,\sigma^2)) = a(\theta)
\end{aligned}$$

where $q_\eta := F^{-1}_{(0,1)}(\eta)$. With

$$\dot{a}(\theta) = \begin{pmatrix} 1 \\ \frac{1}{2}(\sigma^2)^{-\frac{1}{2}} q_\eta \end{pmatrix}, \quad I_{(\mu,\sigma^2)}(Y_1) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix},$$

we get

$$I_{a(\theta)}(Y_1)^{-1} = \dot{a}(\theta)^\top\, I_{(\mu,\sigma^2)}(Y_1)^{-1}\, \dot{a}(\theta) = \left(1, \frac{q_\eta}{2\sigma}\right) \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{q_\eta}{2\sigma} \end{pmatrix} = \sigma^2 + \frac{q_\eta^2\,\sigma^2}{2}.$$

Because of the symmetry of the normal distribution, we can set $\eta_1 = \frac{\alpha_1}{2}$ and $\eta_2 = 1 - \frac{\alpha_1}{2}$. Then the asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval for $Y_0 = \ln(T_0)$ based on the δ-method is given by

$$\begin{aligned}
&\mathbb{P}_N(y_1,\ldots,y_N) \\
&= \left[\widehat{\mu} + \widehat{\sigma} q_{\eta_1} - \sqrt{\frac{1}{N}\left(\widehat{\sigma}^2 + \frac{q^2_{\eta_1}\,\widehat{\sigma}^2}{2}\right)}\, q_{1-\frac{\alpha_2}{2}}\ ,\ \widehat{\mu} + \widehat{\sigma} q_{\eta_2} + \sqrt{\frac{1}{N}\left(\widehat{\sigma}^2 + \frac{q^2_{\eta_2}\,\widehat{\sigma}^2}{2}\right)}\, q_{1-\frac{\alpha_2}{2}}\right],
\end{aligned}$$

where $(\widehat{\mu}, \widehat{\sigma}^2)$ is the maximum likelihood estimator for $\theta = (\mu,\sigma^2)$.

2.10 Nonparametric methods

If the distribution class of the lifetime distribution is unknown then nonparametric methods can be used. However, prediction intervals cannot be derived.

If there are no censored observations then the empirical distribution function

$$F_N(t) := \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{(\infty,t]}(t_n)$$

is a nonparametric estimator of the underlying distribution function $F(t)$ and $1 - F_N(t)$ is a nonparametric estimator of the survival function $S(t) := 1 - F(t)$.

2.10.1 Theorem
a) An asymptotic $(1-\alpha)$-confidence interval for $F(t)$ for noncensored observations is given by

$$\left[F_N(t) - q_{1-\alpha/2}\,\frac{\sqrt{F_N(t)\,(1-F_N(t))}}{\sqrt{N}}, F_N(t) + q_{1-\alpha/2}\,\frac{\sqrt{F_N(t)\,(1-F_N(t))}}{\sqrt{N}}\right].$$

b) An asymptotic naive $(1-\alpha)$-prediction interval for T_0 for noncensored observations is given by

$$[F_N^{-1}(\eta_1), F_{N+}^{-1}(\eta_2)]$$

where $\eta_2 - \eta_1 = 1 - \alpha$ and $F_{N+}^{-1}(\eta) := \sup\{z; F_N(z) \leq \eta\}$.

Proof. By central limit theorem and law of large numbers for Bernoulli variables.□

However, this does not work anymore as soon as there are censored observations.

Since an empirical distribution function provides always a discrete distribution we shall consider at first discrete lifetime distributions.

2.10.2 Definition (Hazard function for discrete distribution)
The hazard function (hazard rate) $h : \mathbb{R}_+ \to \mathbb{R}$ for a discrete random variable T is defined by

$$h(t) := P(T = t | T \geq t).$$

2.10.3 Theorem
The hazard function for a discrete distribution with support $0 \leq \tau_1 < \tau_2 < \tau_3 < \ldots$ satisfies:

$$a) \qquad h(t) = 1 - \frac{S(t)}{S_-(t)} \quad \text{with} \quad S_-(t) := \lim_{s \uparrow t} S(s),$$

$$b) \qquad S(\tau_k) = \prod_{i=1}^{k} (1 - h(\tau_i)) \;\; \text{for} \;\; k \in \{1, 2, 3, \ldots\}.$$

Proof. a) The definition provides

$$\begin{aligned} h(t) &:= P(T = t | T \geq t) = \frac{P(T = t)}{P(T \geq t)} \\ &= \frac{F(t) - \lim_{s \uparrow t} F(s)}{1 - \lim_{s \uparrow t} F(s)} = \frac{1 - S(t) - \lim_{s \uparrow t}(1 - S(s))}{\lim_{s \uparrow t} S(s)} = \frac{S_-(t) - S(t)}{S_-(t)} = 1 - \frac{S(t)}{S_-(t)}. \end{aligned}$$

b) The assertion a) implies

$$1 - h(\tau_i) = \frac{S(\tau_i)}{S_-(\tau_i)} = \frac{S(\tau_i)}{S(\tau_{i-1})} \;\; \text{for} \;\; i \in \{2, 3, \ldots\} \quad \text{and} \quad 1 - h(\tau_1) = \frac{S(\tau_1)}{S_-(\tau_1)} = S(\tau_1)$$

so that

$$\prod_{i=1}^{k} (1 - h(\tau_i)) = S(\tau_1) \cdot \frac{S(\tau_2)}{S(\tau_1)} \cdot \frac{S(\tau_3)}{S(\tau_2)} \cdot \ldots \cdot \frac{S(\tau_k)}{S(\tau_{k-1})} = S(\tau_k). \square$$

Let $z_1, \ldots, z_N$ with $z_n = \min\{t_n, c_n\}$ be right censored observations, $d_n = \mathbb{1}_{[0,c_n]}(t_n)$, and $\tau_1 < \tau_2 < \tau_3 < \ldots < \tau_I$ be ordered distinct time points of observed failures (deaths) of the noncensored observations z_n with $z_n = t_n$ and $d_n = 1$. Define for $i = 1, \ldots, I$

$$\begin{aligned} b_i &:= \sharp\{n \in \{1, \ldots, N\}; z_n = \tau_i \;\; \text{and} \;\; d_n = 1\}, \\ y_i &:= \sharp\{n \in \{1, \ldots, N\}; z_n \geq \tau_i\}, \end{aligned}$$

where $\sharp$ stands for the number of elements of a set. Thereby y_i is the number of individuals at risk at time point τ_i and b_i is the number of individuals which fails (dies) at time point τ_i. An estimator of the hazard rate at τ_i for $i = 1, \ldots, I$ is given by

$$\widehat{h}(\tau_i) := \frac{b_i}{y_i}.$$

Theorem 2.10.3 provides then an estimate for the survival function based on $z_1, \ldots, z_N$.

2.10.4 Definition (Kaplan-Meier estimator / Product-Limit estimator, see Klein and Moeschberger 2003, p. 92)
The Kaplan-Meier estimate (Product-Limit estimate) of the survival function S of the underlying distribution based on censored observations $z_1, \dots, z_N$ is given by

$$\widehat{S}_{KM}(t) := \begin{cases} 1 & \text{for } t < \tau_1, \\ \prod_{i;\, \tau_i \le t} \left(1 - \frac{b_i}{y_i}\right) & \text{for } \tau_1 \le t \le \tau_I. \end{cases}$$

If $b_I = y_I$ then $\widehat{S}(\tau_I) = 0$ and $\widehat{S}$ is decreasing from 1 to 0.

If the observations are coming from a continuous distribution then the uncensored observations are usually pairwise different. If no observation is censored then the Kaplan-Meier estimate provides the empirical distribution function.

2.10.5 Theorem
If there are no censored observations and the observations are pairwise different then

$$F_N = 1 - \widehat{S}_{KM},$$

i.e. the Kaplan-Meier estimate provides the empirical distribution function.

Proof. The assumption means $\tau_1 = t_{(1)} < \tau_2 = t_{(2)} < \ldots < \tau_N = t_{(N)}$, where $t_{(1)}, \dots, t_{(N)}$ are the ordered noncensored observations. This implies $b_i = 1$ and $y_i = N - i + 1$ for $i = 1, \dots, N$ so that for any $t \in [\tau_k, \tau_{k+1})$ with $k = 1, \dots, N-1$

$$\begin{aligned} 1 - \widehat{S}_{KM}(t) &= 1 - \prod_{i;\, \tau_i \le t} \left(1 - \frac{b_i}{y_i}\right) = 1 - \prod_{i;\, \tau_i \le t} \left(1 - \frac{1}{N-i+1}\right) \\ &= 1 - \prod_{i;\, \tau_i \le t} \frac{N-i}{N-i+1} = 1 - \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \frac{N-3}{N-2} \cdot \ldots \frac{N-k}{N-k+1} \\ &= 1 - \frac{N-k}{N} = \frac{k}{N} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{(\infty,t]}(t_n) = F_N(t). \end{aligned}$$

Obviously, we have $1 - \widehat{S}_{KM}(t) = 1 - 1 = F_N(t)$ for $t < \tau_1$ and $1 - \widehat{S}_{KM}(t) = 1 - 0 = F_N(t)$ for $t \ge \tau_N$. $\square$

2.10.6 Theorem (See Kahle and Liebscher 2013, p. 99)
The Kaplan-Meier estimator is a generalized maximum likelihood estimator for S in the class of step functions with jumps at $\tau_1, \dots, \tau_I$.

Proof. The observations consist of the event times $\tau_1, \dots, \tau_I$, the number of failures/deaths b_i at event time τ_i, and the number y_i of individuals under risk at event time τ_i, $i = 1, \dots, I$.

Unknown parameters are $h_i = h(\tau_i)$ with $i = 1, \dots, I$. Set

$$j := \arg\max\{i \in \{1, \dots, I\};\ \tau_j \leq t\}$$

and

$$\theta := (h_1, \dots, h_j).$$

Then we have

$$S(t) = \prod_{i;\ \tau_i \leq t} (1 - h_i) = g((h_1, \dots, h_j)) = g(\theta)$$

so that an ML estimate $\widehat{\theta}$ for θ provides an ML estimate $g(\widehat{\theta})$ for $S(t)$. To derive the ML estimate for θ, note that b_i can be interpreted as realization of $B_i \sim \mathcal{B}(y_i, h_i)$, i.e. B_i has binomial distribution with parameters y_i and h_i. Then the likelihood function of the observations is given by

$$l(\theta; (b_1, \dots, b_j)) := l(\theta; (b_1, \dots, b_j), (y_1, \dots, y_j), (\tau_1, \dots, \tau_j)) = \prod_{i=1}^{j} \binom{y_i}{b_i} h_i^{b_i} (1 - h_i)^{y_i - b_i}$$

so that

$$\ln(l(\theta; (b_1, \dots, b_j))) = \sum_{i=1}^{j} \left(\ln\left(\binom{y_i}{b_i} \right) + b_i \ln(h_i) + (y_i - b_i) \ln(1 - h_i) \right)$$

and

$$\begin{aligned} &\frac{\partial}{\partial h_i} \ln(l(\theta; (b_1, \dots, b_j))) = \frac{b_i}{h_i} - \frac{y_i - b_i}{1 - h_i} = 0 \\ &\iff b_i - h_i\, b_i = h_i\, y_i - h_i\, b_i \iff h_i = \frac{b_i}{y_i}. \end{aligned}$$

Hence we obtain $\widehat{\theta} = \left(\frac{b_1}{y_1}, \dots, \frac{b_j}{y_j} \right)$. $\square$

Since the Kaplan-Meier estimator is a generalized maximum likelihood estimator for S, it has an asymptotic normal distribution and confidence sets for $S(t)$ can be derived.

2.10.7 Theorem (See Kahle and Liebscher 2013, p. 100, Klein and Moeschberger 2003, p. 92) *An asymptotic $(1-\alpha)$-confidence interval for $S(t)$ is given by*

$$[\widehat{S}_{KM}(t) - q_{1-\alpha/2}\,\widehat{\sigma}_{KM}(t), \widehat{S}_{KM}(t) + q_{1-\alpha/2}\,\widehat{\sigma}_{KM}(t)],$$

where q_α is the α-quantile of the standard normal distribution and $\widehat{\sigma}_{KM}(t)$ is given by Greenwood's formula

$$\widehat{\sigma}_{KM}(t)^2 := \widehat{S}_{KM}(t)^2 \sum_{i;\,\tau_i \le t} \frac{b_i}{y_i(y_i - b_i)}.$$

Proof. The number of failures/deaths b_i at event time τ_i can be interpreted as realization of $B_i \sim \mathcal{B}(y_i, h_i)$, i.e. B_i has binomial distribution with parameters y_i and h_i. In particular we have $\mathsf{E}(B_i) = y_i\,h_i$ or

$$\mathsf{E}\left(\frac{B_i}{y_i}\right) = h_i \text{ and } \operatorname{var}\left(\frac{B_i}{y_i}\right) = \frac{h_i\,(1-h_i)}{y_i}$$

so that with the central limit theorem

$$\frac{\frac{B_i}{y_i} - h_i}{\sqrt{\frac{h_i\,(1-h_i)}{y_i}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

which means with $\widehat{h}_i := \frac{B_i}{y_i}$

$$\widehat{h}_i - h_i \xrightarrow{\mathcal{D}} V_i \sim \mathcal{N}\left(0, \frac{h_i\,(1-h_i)}{y_i}\right).$$

With the δ-method, we get

$$\ln(1-\widehat{h}_i) - \ln(1-h_i) \xrightarrow{\mathcal{D}} W_i := \frac{-1}{1-h_i} V_i \sim \mathcal{N}\left(0, \frac{1}{(1-h_i)^2}\frac{h_i\,(1-h_i)}{y_i}\right) = \mathcal{N}\left(0, \frac{h_i}{y_i\,(1-h_i)}\right).$$

Then we obtain with the martingale central limit theorem

$$\ln(\widehat{S}_{KM}(t)) - \ln(S(t)) = \sum_{i=1}^{j}\Big(\ln(1-\widehat{h}_i) - \ln(1-h_i)\Big) \xrightarrow{\mathcal{D}} \sum_{i=1}^{j} W_i \sim \mathcal{N}\left(0, \sum_{i=1}^{j}\frac{h_i}{y_i\,(1-h_i)}\right).$$

Again applying the δ-method leads to

$$\begin{aligned}\widehat{S}_{KM}(t) - S(t) &= \exp(\ln(\widehat{S}_{KM}(t))) - \exp(\ln(S(t)))\\ &\xrightarrow{\mathcal{D}} \exp(\ln(S(t)))\sum_{i=1}^{j} W_i = S(t)\sum_{i=1}^{j} W_i \sim \mathcal{N}\left(0, S(t)^2\sum_{i=1}^{j}\frac{h_i}{y_i\,(1-h_i)}\right).\end{aligned}$$

A consistent estimate of the asymptotic variance $S(t)^2 \sum_{i=1}^{j} \frac{h_i}{y_i\,(1-h_i)}$ is

$$\widehat{S}_{KM}(t)^2 \sum_{i=1}^{j} \frac{\widehat{h}_i}{y_i\,(1-\widehat{h}_i)} = \widehat{S}_{KM}(t)^2 \sum_{i=1}^{j} \frac{\frac{B_i}{y_i}}{y_i\,(1-\frac{B_i}{y_i})} = \widehat{S}_{KM}(t)^2 \sum_{i;\,\tau_i \le t} \frac{B_i}{y_i(y_i - B_i)} = \widehat{\sigma}_{KM}(t)^2$$

so that with Lemma 2.8.8

$$\frac{\widehat{S}_{KM}(t) - S(t)}{\widehat{\sigma}_{KM}(t)} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).\square$$

An alternative estimator for the survival function can be obtained via the cumulative hazard function for discrete distributions.

2.10.8 Definition (Cumulative hazard function for discrete distributions)
If T has a discrete distribution with support $0 \le \tau_1 < \tau_2 < \tau_3 < \ldots$, then $H : \mathbb{R}_+ \to \mathbb{R}$ given by

$$H(t) := \sum_{i;\,\tau_i \le t} h(\tau_i)$$

is called cumulative hazard function.

An estimator for $H(t)$ is given via the estimates $\widehat{h}(\tau_i) = \frac{b_i}{y_i}$ for $h(\tau_i)$.

2.10.9 Definition (Nelson-Aalen estimator, see Klein and Moeschberger 2003, p. 94)
The Nelson-Aalen estimate of the cumulative hazard function H of the underlying distribution based on censored observations $z_1, \ldots, z_N$ is given by

$$\widehat{H}(t) := \sum_{i;\,\tau_i \le t} \frac{b_i}{y_i}.$$

2.10.10 Theorem
If T has a continuous distribution and $\widehat{H}(t)$ is the Nelson-Aalen estimate of the cumulative hazard function H then

$$\widehat{S}_{NA}(t) := \exp(-\widehat{H}(t))$$

is an estimate for the survival function $S(t)$.

Proof. This follows at once from Theorem 2.1.7.□

Note that for small $x > 0$ we have $1 - x \approx \exp(-x)$ so that

$$\widehat{S}_{NA}(t) := \exp(-\widehat{H}(t)) = \prod_{i;\,\tau_i \le t} \exp\left(-\frac{b_i}{y_i}\right) \approx \prod_{i;\,\tau_i \le t} \left(1 - \frac{b_i}{y_i}\right) = \widehat{S}_{KM}(t).$$

An estimated variance of the Nelson-Aalen estimate $\widehat{H}(t)$ is given by

$$\widehat{\sigma}_{NA}(t)^2 := \sum_{i;\, \tau_i \leq t} \frac{b_i}{y_i^2},$$

see Klein and Moeschberger 2003, p. 94.

Chapter 3

Experiments with different stress levels

Usually lifetime experiments under realistic stress are lasting very long so that a failure often cannot be observed. To overcome this problem are step-stress experiments or accelerated lifetime experiments.

3.1 Step-stress experiments

Before the experiment, stress levels $0 < s_1 < s_2 < \ldots < s_K$ and censoring times $c_1 < c_2 < \ldots < c_K$ are fixed. Then each experiment starts with a the first stress level s_1. If no failure is observed until time c_1 then the stress is increased to s_2. If no failure is observed until time c_2 then the stress is increased to s_3. And so on until time c_K is reached. The question is how to model the influence of the increased stress. Balakrishnan (2009) proposed to make the following assumption:

$$F_{(\theta_1,\ldots,\theta_K)}(t) := \begin{cases} F_{\theta_1}(t) & \text{for } 0 < t \le c_1, \\ F_{\theta_k}(t + a_{k-1} - c_{k-1}) & \text{for } c_{k-1} < t \le c_k,\ k = 2,\ldots,K-1, \\ F_{\theta_K}(t + a_{K-1} - c_{K-1}) & \text{for } c_{K-1} < t < \infty, \end{cases} \tag{3.1}$$

where

$$a_{k-1} := \theta_k \sum_{i=1}^{k-1} \left(\frac{c_i - c_{i-1}}{\theta_i} \right) \quad \text{for} \quad k = 2, \ldots, K,$$

$c_0 := 0$, $a_0 := 0$ and F_{θ_k}, $k = 1, \ldots, K$, is the cumulative distribution function of a scale family of distributions, i.e. $F_{\theta_k}(t) = F\left(\frac{t}{\theta_k}\right)$ for some cumulative distribution function F. Note that $a_{k-1} - c_{k-1}$ is a location shift.

3.1.1 Lemma
$F_{(\theta_1,\ldots,\theta_K)}$ given by (3.1) is continuous and the density is given by

$$f_{(\theta_1,\ldots,\theta_K)}(t) = \begin{cases} f_{\theta_1}(t) & \text{for } 0 < t \le c_1, \\ f_{\theta_k}(t + a_{k-1} - c_{k-1}) & \text{for } c_{k-1} < t \le c_k,\ k = 2,\ldots,K-1, \\ f_{\theta_K}(t + a_{K-1} - c_{K-1}) & \text{for } c_{K-1} < t < \infty, \end{cases}$$

where $f_{\theta_k}(t) = F'_{\theta_k}(t)$ for $k = 1,\ldots,K$.

Proof. For $k = 1,\ldots,K$ we have

$$\begin{aligned} F_{\theta_k}(c_k + a_{k-1} - c_{k-1}) &= F\left(\frac{a_{k-1} + c_k - c_{k-1}}{\theta_k}\right) \\ &= F\left(\sum_{i=1}^{k}\left(\frac{c_i - c_{i-1}}{\theta_i}\right)\right) = F\left(\theta_{k+1}\sum_{i=1}^{k}\left(\frac{c_i - c_{i-1}}{\theta_i}\right)\frac{1}{\theta_{k+1}}\right) \\ &= F_{\theta_{k+1}}(a_k) = F_{\theta_{k+1}}(c_k + a_k - c_k). \end{aligned}$$

Similarly

$$f_{\theta_k}(c_k + a_{k-1} - c_{k-1}) = f_{\theta_{k+1}}(c_k + a_k - c_k). \square$$

For the family of exponential distributions with scale parameter $\theta = \frac{1}{\lambda}$ we obtain

$$F_{(\theta_1,\ldots,\theta_K)}(t) := \begin{cases} 1 - e^{-t/\theta_1} & \text{for } 0 < t \le c_1, \\ 1 - e^{-(t + \frac{\theta_2}{\theta_1}c_1 - c_1)/\theta_2} & \text{for } c_1 < t \le c_2, \\ 1 - e^{-(t + a_{k-1} - c_{k-1})/\theta_k} & \text{for } c_{k-1} < t \le c_k,\ k = 2,\ldots,K-1, \\ 1 - e^{-(t + a_{K-1} - c_{K-1})/\theta_K} & \text{for } c_{K-1} < t < \infty. \end{cases}$$

The parameter $(\theta_1,\ldots,\theta_K)$ can be estimated with the maximum likelihood method if at least one observations is obtained in each of the intervals $(0, c_1]$, $(c_1, c_2]$, $\ldots$, $(c_{K-1}, c_K]$. Note that the observations are given by $z_1,\ldots,z_N$ with $z_n = \min(t_n, c_K)$ and $d_n = 1\!\!1\{t_n \le c_K\}$ for $n = 1,\ldots,N$. Then the likelihood function is given by

$$f_{(\theta_1,\ldots,\theta_K)}(z_1,\ldots,z_N) = \prod_{n=1}^{N} f_{(\theta_1,\ldots,\theta_K)}(t_n)^{d_n}\left(1 - F_{(\theta_1,\ldots,\theta_K)}(c_K)\right)^{1-d_n}.$$

The problem is that the maximum likelihood estimator and corresponding confidence sets and prediction intervals cannot be determined if there is no observation in $[0, c_1]$.

3.2 Accelerated lifetime experiments

In accelerated lifetime experiments, different experiments are run under different stress levels, usual at stress level larger then the stress level which is of interest. Then the expected lifetime $E(T(s))$ depends on the stress level s and the dependence is given by via a given link function. Here it is assumed that this function is known up to a parametervector θ. We assume that N life time experiments at different stress levels $s_n \in \mathcal{S}$ for $n = 1, \ldots, N$ are executed and that the lifetime T_n of the product shall be observed in each lifetime experiment. However, if the stress is too low then often the lifetime cannot be observed since the time up to the event, the "death", is too long. Therefore usually a time c is fixed at which the lifetime experiment is stopped. Then the only information is that the product has survived the time c. Such observations are socalled censored observations. It is clear that the censored observations should also be used in an analysis of lifetime data. Therefore define

$$Z_n := \begin{cases} T_n, & \text{if } T_n \leq c, \\ c, & \text{if } T_n > c, \end{cases} \quad \text{and} \quad D_n := \begin{cases} 1, & \text{if } T_n \leq c, \\ 0, & \text{if } T_n > c. \end{cases}$$

Then $(Z_1, D_1, s_1), \ldots, (Z_N, D_N, s_N)$ are the available informations where D_n is the censoring variable. Let be t_n, z_n, and d_n the realizations of T_n, Z_n and D_n respectively and $z_* = (z_1, \ldots, z_N)^\top$, $d_* = (d_1, \ldots, d_N)^\top$, $s_* = (s_1, \ldots, s_N)^\top$. The likelihood function is then given by (see e.g. Klein and Moeschberger 2003, p.75)

$$L_\theta(z_*, d_*, s_*) := \prod_{n=1}^{N} f_{\theta,s_n}(z_n)^{d_n} S_{\theta,s_n}(z_n)^{1-d_n}$$

if f_{θ,s_n} is the lifetime distribution density of T_n at stress s_n and

$$S_{\theta,s_n}(t) := \int_t^\infty f_{\theta,s_n}(u)\, du$$

the survival function of T_n at time t and stress s_n. Assume $T_n \sim T(s_n)$.

Typical link functions (see e.g. Haibach 2006, S.25)

1870	Wöhler:	$\log(E(T(s))) = \theta_0 - \theta_1\, s$
1910	Basquin:	$\log(E(T(s))) = \theta_0 - \theta_1\, \log(s)$
1914	Stromeyer:	$\log(E(T(s))) = \theta_0 - \theta_1\, \log(s - s_L)$
1963	Bastenaire:	$\log(E(T(s))) = \theta_0 - \log(s - s_L) - \theta_1(s - s_L)^{\theta_3}$

Thereby, s_L is the **fatigue limit** (in German Dauerfestigkeit), i.e. the stress level at which no failure can be observed. This is an unknown parameter like the other parameter θ_0, θ_1, θ_2 and must be estimated as well. However, a fatigue limit greater zero cause problems in estimation. Moreover, it is not clear what $\log(E(T(s)))$ should be when $s < s_L$ is used in this case.

S-N curves (in German Wöhlerlinien)

Since lifetime is often measures in number of load cycles when the stress is given by cyclic load, engineers use N instead of t for the dependent variable. They plot this variable on the horizontal axis of a diagram. And the vertical axis provides then the levels of the stress, given by the amplitude of the cyclic load. The estimated link functions are then the S-N curves. Since the dependent variable is plotted on the horizontal axis, the S-N curves are often not correctly estimated when the method of least squares is used in direction of the vertical axis. This is the reason that the fatigue limit s_L can be estimated then. However this is not correct since then the lifetime, the number of load cycles, is not anymore the dependent variable.

Hence it makes more sense to regard the following link function:

$$\ln(E(T(s))) = g_\theta(s)$$

with

$$g_\theta(s) = \theta_0 - \theta_1\, s, \tag{3.2}$$
$$g_\theta(s) = \theta_0 - \theta_1\, \log(s), \tag{3.3}$$
$$g_\theta(s) = \theta_0 + \theta_1\, \frac{1}{s}, \tag{3.4}$$
$$g_\theta(s) = \theta_0 + \theta_1\, \left(\frac{1}{s}\right)^{\theta_2}, \tag{3.5}$$
$$g_\theta(s) = \theta_0 - \theta_1\, s - \theta_2\, \log(s), \tag{3.6}$$
$$g_\theta(s) = \theta_0 - \theta_1\, \log(s) + \theta_2\, \left(\frac{1}{s}\right)^{\theta_3}. \tag{3.7}$$

with $\theta_0, \theta_1, \theta_2, \theta_3 > 0$.

A better definition of the fatigue limit would be the largest stress level s with $g_\theta(s) \geq L$ where L is a large number so that experiments having this lifetime can be considered as experiments without failure in a relevant time period.

3.2.1 Definition (Alternative definition of fatigue limit)
Is L a large value, then the stress level s_L with $g_\theta(s_L) = L$, i.e. $s_L = g_\theta^{-1}(L)$ is called fatigue limit.

Of interest are especially the estimators and confidence sets for two one-dimensional aspects. Namely the fatigue limit $a_L(\theta) = g_\theta^{-1}(L)$ and the lifetime time at a given stress level s_0, i.e. $a_0(\theta) = \exp(g_\theta(s_0))$.

The aspects $a_L(\theta) = g_\theta^{-1}(L)$ and $a_0(\theta) = \exp(g_\theta(s_0))$ can be estimated by the maximum likelihood method by $a_L(\widehat{\theta})$ and $a_0(\widehat{\theta})$ if $\widehat{\theta}$ is the maximum likelihood estimator for θ. The maximum likelihood estimator for θ is given by

$$\begin{aligned}\widehat{\theta} := \widehat{\theta}(z_*, d_*, s_*) &:= \arg\max_\theta L_\theta(z_*, d_*, s_*)\\ &= \prod_{n=1}^N f_{\theta,s_n}(z_n)^{d_n} S_{\theta,s_n}(z_n)^{1-d_n} = \prod_{n=1}^N f_{\theta,s_n}(t_n)^{\mathbb{1}_{[0,c]}(t_n)} \left(1 - F_{\theta,s_n}(c)\right)^{\mathbb{1}_{(c,\infty)}(t_n)}.\end{aligned}$$

Set

$$\begin{aligned} l(\theta,t,s) &:= \ln\left(f_{\theta,s}(t)^{\mathbb{1}_{[0,c]}(t)}\,(1-F_{\theta,s}(c))^{\mathbb{1}_{(c,\infty)}(t)}\right)\\ &= \ln(f_{\theta,s}(t))\mathbb{1}_{[0,c]}(t)+\ln\left(1-F_{\theta,s}(c)\right)\mathbb{1}_{(c,\infty)}(t). \end{aligned}$$

and

$$\dot{l}(\theta,t,s) := \frac{\partial}{\partial\theta}l(\theta,t,s).$$

Further we assume that $(t_1,s_1),\ldots,(t_N,s_N)$ are realizations of i.i.d. random variables (T_1,S_1), $\ldots,(T_N,S_N)$ where $P^{T|S=s}([0,t])=F_{\theta,s}$ and $P^S=\delta$. This means that the stress levels $s_1,\ldots,s_N$ are given by the *design measure* δ, i.e. they follow a *random design*. If $(t_1,s_1),\ldots,(t_N,s_N)$ are realizations of i.i.d. random variables $(T_1,S_1),\ldots,(T_N,S_N)$ then also $(z_1,d_1,s_1),\ldots,(z_N,d_N,s_N)$ are realizations of i.i.d. random variables $(Z_1,D_1,S_1),\ldots,(Z_N,D_N,S_N)$. Assuming that $z_*=(z_1,\ldots,z_N)^\top$, $d_*=(d_1,\ldots,d_N)^\top$, $s_*=(s_1,\ldots,s_N)^\top$ are realizations of $Z_*=(Z_1,\ldots,Z_N)^\top$, $D_*=(D_1,\ldots,D_N)^\top$, $S_*=(S_1,\ldots,S_N)^\top$ then a modification of Theorem 2.9.9 holds.

3.2.2 Theorem (Schervish 1997, Theorem 7.57 or Theorem 7.63)
If $(T_1,S_1),\ldots,(T_N,S_N)$ are i.i.d., $\widehat{\theta}_N:=\widehat{\theta}_N(Z_*,D_*,S_*)$ is the maximum likelihood estimator for θ, then under regularity conditions (see Schervish 1997, Theorem 7.57 or Theorem 7.63),

$$\sqrt{N}(\widehat{\theta}_N-\theta_*)\xrightarrow{\mathcal{D}}\mathcal{N}(0_r,I_{\theta_*}(\delta)^{-1}),\tag{3.8}$$

if θ_* is the true parameter, where

$$I_{\theta_*}(\delta):=E_{\theta_*}\left(\dot{l}(\theta_*,T_1,S_1)\,\dot{l}(\theta_*,T_1,S_1)^\top\right)=\int E_{\theta_*}\left(\dot{l}(\theta_*,T_1,s)\,\dot{l}(\theta_*,T_1,s)^\top\right)\delta(ds).$$

If the design measure δ is a discrete probability measure, i.e. $\delta=\sum_{k=1}^K a_k e_{\widetilde{s}_k}$ with $\sum_{k=1}^K a_k=1$, then we have

$$I_{\theta_*}(\delta)=\sum_{k=1}^K a_k\;E_{\theta_*}\left(\dot{l}(\theta_*,T_1,\widetilde{s}_k)\,\dot{l}(\theta_*,T_1,\widetilde{s}_k)^\top\right).$$

If δ is a continuous probability measure with probability density g then

$$I_{\theta_*}(\delta)=\int E_{\theta_*}\left(\dot{l}(\theta_*,T_1,s)\,\dot{l}(\theta_*,T_1,s)^\top\right)\;g(s)\;ds.$$

Usually one would use concrete designs where $s_1,\ldots,s_N$ are given by the experimenter and not by random. However, each concrete design $d_N=(s_1,\ldots,s_N)$ can be associated with the probability measure given by

$$\delta_N:=\frac{1}{N}\sum_{n=1}^N e_{s_n}$$

where e_s denotes the one-point (Dirac) measure on the point s, i.e. $e_s(A) = 1$ if $s \in A$ and $e_s(A) = 0$ if $s \notin A$ for any A of the Borel-σ-algebra. The probability measure δ_N is also called *generalized design*.

Often the generalized design δ_N converges weakly to a probability measure δ which is then considered as asymptotic design measure.

3.2.3 Example (Example of convergent generalized designs)
a) A concrete design with

$$d_N = (s_1, s_2, s_1, s_2, \ldots, s_1, s_2^\top),$$

where s_1, s_2 is repeated $\frac{N}{2}$ times, has the generalized design given by

$$\delta_N = \frac{1}{N}\left(\frac{N}{2}e_{s_1} + \frac{N}{2}e_{s_1}\right) = \frac{1}{2}e_{s_1} + \frac{1}{2}e_{s_2}$$

so that the convergence to $\delta = \frac{1}{2}e_{s_1} + \frac{1}{2}e_{s_2}$ is obvious.
b) A concrete design with

$$d_N = \left(\frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\right)$$

has a generalized design measure δ_N which converges weakly to the uniform measure on $[0, 1]$, i.e. to a continuous design measure δ with density $g(s) = 1\!\!1_{[0,1]}(s)$.

3.2.4 Lemma
If $s_1, \ldots, s_N$ are realizations of independent $S_1, \ldots, S_N \sim \delta$, δ has finite support, and $\delta_N = \frac{1}{N}\sum_{n=1}^N e_{s_n}$, then $\delta_N \to \delta$ weakly almost surely.

Proof. Let $\{\tilde{s}_1, \ldots, \tilde{s}_I\}$ be the finite support of δ. The strong law of large numbers provides

$$\lim_{N\to\infty} \delta_N(\{\tilde{s}_i\}) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N e_{S_n(\omega)}(\{\tilde{s}_i\}) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N 1\!\!1_{\{\tilde{s}_i\}}(S_n(\omega)) = E(1\!\!1_{\{\tilde{s}_i\}}(S_1)) = \delta(\{\tilde{s}_i\})$$

for all $\omega \in \Omega_0$ with $P(\Omega_0) = 1$. Each $\omega \in \Omega_0$ satisfies

$$\lim_{N\to\infty} \int f(s)\,\delta_N(ds) = \lim_{N\to\infty} \sum_{i=1}^I f(\tilde{s}_i)\,\delta_N(\{\tilde{s}_i\}) = \sum_{i=1}^I f(\tilde{s}_i)\,\delta(\{\tilde{s}_i\}) = \int f(s)\,\delta(ds)$$

for all all continuous and bounded functions $f : [0, \infty) \to \mathbb{R}$. Hence δ_N converges weakly to δ according to the Theorem of Portmanteau (Theorem 2.8.3) for all $\omega \in \Omega_0$. This means $\delta_N \to \delta$ weakly almost surely. $\square$

3.2.5 Corollary
If $T_1,\ldots,T_N$ are independent and δ_N converges weakly to δ almost surely then the maximum likelihood estimator $\widehat{\theta}_N$ for θ satisfies

$$\sqrt{N}(\widehat{\theta}_N - \theta_*) \xrightarrow{\mathcal{D}} \mathcal{N}(0_r, I_{\theta_*}(\delta)^{-1}),$$

if θ_* is the true parameter and $I_{\theta_*}(\delta)$ is given by Theorem 3.2.2.

3.2.6 Corollary
If $T_1,\ldots,T_N$ are independent and δ_N converges weakly to δ almost surely then the maximum likelihood estimator $a(\widehat{\theta}_N)$ for $a(\theta)\in\mathbb{R}^s$ satisfies

$$\sqrt{N}(a(\widehat{\theta}_N) - a(\theta_*)) \xrightarrow{\mathcal{D}} \mathcal{N}(0_s, \dot{a}_{\theta_*}^\top I_{\theta_*}(\delta)^{-1}\dot{a}_{\theta_*}),$$

if θ_* is the true parameter, $I_{\theta_*}(\delta)$ is given by Theorem 3.2.2 and $\dot{a}_\theta = \frac{\partial}{\partial\theta}a(\theta)$.

3.2.7 Lemma
If $\widehat{\theta}_N$ is the maximum likelihood estimator for θ, then under regularity conditions,

$$I_{\widehat{\theta}_N}(\delta_N) := \frac{1}{N}\sum_{n=1}^N E_{\widehat{\theta}_N}\left(\dot{l}(\widehat{\theta}_N, T_1, s_n)\,\dot{l}(\widehat{\theta}_N, T_1, s_n)^\top\right) \xrightarrow{\mathcal{P}} I_{\theta_*}(\delta).$$

Proof. If δ_N has support included in $[s_{min}, s_{max}]$ and

$$\sup_{s\in[s_{min},s_{max}]} \|E_{\widehat{\theta}_N}\left(\dot{l}(\widehat{\theta}_N, T_1, s_n)\,\dot{l}(\widehat{\theta}_N, T_1, s_n)^\top\right) - E_{\theta_*}\left(\dot{l}(\theta_*, T_1, s_n)\,\dot{l}(\theta_*, T_1, s_n)^\top\right)\| \xrightarrow{\mathcal{P}} 0$$

then

$$I_{\widehat{\theta}_N}(\delta_N) - I_{\theta_*}(\delta_N) \xrightarrow{\mathcal{P}} 0.$$

The Theorem of Portmanteau (Theorem 2.8.3) provides

$$I_{\theta_*}(\delta_N) \xrightarrow{\mathcal{P}} I_{\theta_*}(\delta)$$

so that the assertion follows. □

If $\dot{a}_\theta$ is continuous in θ then also

$$\dot{a}_{\widehat{\theta}_N}^\top\, I_{\widehat{\theta}_N}(\delta_N)^{-1}\,\dot{a}_{\widehat{\theta}_N}$$

is a consistent estimator of $\dot{a}_\theta^\top I_{\theta_*}(\delta)^{-1}\dot{a}_\theta$. If additional $a(\theta) \in \mathbb{R}$, i.e. $a(\theta)$ is one-dimensional, then Corollary 3.2.6 implies

$$\sqrt{N}\frac{a(\widehat{\theta}_N) - a(\theta_*)}{\sqrt{\dot{a}_{\widehat{\theta}_N}^\top \, I_{\widehat{\theta}_N}(\delta_N)^{-1}\, \dot{a}_{\widehat{\theta}_N}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

so that we get the following theorem.

3.2.8 Theorem
$\mathbb{C}_N$ given by

$$\begin{aligned}&\mathbb{C}_N(z_*, d_*, s_*)\\ =& \left[a(\widehat{\theta}_N) - \sqrt{\frac{1}{N}\,\dot{a}_{\widehat{\theta}_N}^\top I_{\widehat{\theta}_N}(\delta_N)^{-1}\dot{a}_{\widehat{\theta}_N}}\;\; q_{1-\alpha/2} \;\;,\;\; a(\widehat{\theta}_N) + \sqrt{\frac{1}{N}\,\dot{a}_{\widehat{\theta}_N}^\top I_{\widehat{\theta}_N}(\delta_N)^{-1}\dot{a}_{\widehat{\theta}_N}}\;\; q_{1-\alpha/2}\right]\end{aligned}$$

is an asymptotic $(1-\alpha)$-confidence interval function for $a(\theta) \in \mathbb{R}$.

Hence Theorem 3.2.8 provides confidence intervals for the fatigue limit $a_L(\theta) = g_\theta^{-1}(L)$ and the expected lifetime time at a given stress level s_0, i.e. $a_0(\theta) = \exp(g_\theta(s_0))$. Since g_θ is a decreasing function, we can conclude that on

$$\left[0 \;\;,\;\; a_L(\widehat{\theta}_N) - \sqrt{\frac{1}{N}\,\dot{a}_{L\widehat{\theta}_N}^\top I_{\widehat{\theta}_N}(\delta_N)^{-1}\dot{a}_{L\widehat{\theta}_N}}\;\; q_{1-\alpha/2}\right]$$

the fatigue limit given by L is satisfied in $(1-\alpha)\cdot 100\%$ of cases. Prediction intervals for the lifetime T_0 at a stress level s_0 are given by the following theorems.

3.2.9 Theorem (Naive / plug-in prediction interval)
Let $\widehat{\theta}_N := \widehat{\theta}_N(z_*, d_*, s_*)$ be a weakly consistent estimator for θ and $F_{\theta,s_0}(t)$ is continuous in t and $F_{\theta,s_0}^{-1}(\eta)$ continuous in θ at $\eta = \eta_1, \eta_2$. If $0 \le \eta_1 < \eta_2 \le 1$ and $\eta_2 - \eta_1 = 1-\alpha$ then the naive or plug-in prediction interval function $\mathbb{P}_N$ given by

$$\mathbb{P}_N(z_*, d_*, s_*) = \left[F_{\widehat{\theta}_N,s_0}^{-1}(\eta_1) \;\;,\;\; F_{\widehat{\theta}_N,s_0}^{-1}(\eta_2)\right]$$

is an asymptotic $1-\alpha$-prediction interval function for T_0.

3.2.10 Theorem
Let be $\widehat{\theta}_N := \widehat{\theta}_N(Z_*, D_*, S_*)$ be the maximum likelihood estimator for θ. Then $\mathbb{C}_N$ given by

$$\mathbb{C}_N(z_*, d_*, s_*) = \left\{\theta \in \mathbb{R}^r;\ N\ \left(\widehat{\theta}_N - \theta\right)^\top I_{\widehat{\theta}_N}(\delta_N)\left(\widehat{\theta}_N - \theta\right) \le \chi^2_{r;1-\alpha_2}\right\}$$

or

$$\mathbb{C}_N(z_*, d_*, s_*) = \left\{\theta \in \mathbb{R}^r; -2\ \ln\left(\frac{\prod_{n=1}^N f_{\theta,s_n}(z_n)^{d_n} S_{\theta,s_n}(z_n)^{1-d_n}}{\prod_{n=1}^N f_{\widehat{\theta}_N,s_n}(z_n)^{d_n} S_{\widehat{\theta}_N,s_n}(z_n)^{1-d_n}}\right) \le \chi^2_{r;1-\alpha_2}\right\}$$

is an asymptotic $(1-\alpha_2)$-confidence set function for θ and $\mathbb{P}_N$ given by

$$\mathbb{P}_N(z_*, d_*, s_*) = \bigcup_{\theta \in \mathbb{C}_N(z_*, d_*, s_*)} \left[F^{-1}_{\theta,s_0}(\eta_1), F^{-1}_{\theta,s_0}(\eta_2)\right]$$

with $0 \le \eta_1 < \eta_2 \le 1$ and $\eta_2 - \eta_1 = 1 - \alpha_1$ is an asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval function for T_0 at s_0.

3.2.11 Theorem (Prediction intervals based on on the δ-method)
Let $T_0, T_1, \ldots, T_N$ be independent, $\widehat{\theta}_N := \widehat{\theta}_N(Z_*, D_*, S_*)$ is the maximum likelihood estimator for $\theta \in \mathbb{R}^r$, $0 \le \eta_1 < \eta_2 \le 1$ with $\eta_2 - \eta_1 = 1 - \alpha_1$, $a_1, a_2 : \mathbb{R}^r \to \mathbb{R}$ with $a_1(\theta) = F^{-1}_{\theta,s_0}(\eta_1)$ and $a_2(\theta) = F^{-1}_{\theta,s_0}(\eta_2)$ are differentiable aspect functions, and

$$\begin{aligned} v_{1N} &:= \sqrt{\frac{1}{N}\,\dot{a}^\top_{1\widehat{\theta}_N}\, I_{\widehat{\theta}_N}(\delta_N)^{-1}\,\dot{a}_{1\widehat{\theta}_N}}\ \ q_{1-\alpha_2/2}, \\ v_{2N} &:= \sqrt{\frac{1}{N}\,\dot{a}^\top_{2\widehat{\theta}_N}\, I_{\widehat{\theta}_N}(\delta_N)^{-1}\,\dot{a}_{2\widehat{\theta}_N}}\ \ q_{1-\alpha_2/2}. \end{aligned}$$

Then $\mathbb{P}_N$ given by

$$\mathbb{P}_N(z_*, d_*, s_*) = \left[F^{-1}_{\widehat{\theta}_N,s_0}(\eta_1) - v_{1N}\ \ ,\ \ F^{-1}_{\widehat{\theta}_N,s_0}(\eta_2) + v_{2N}\right]$$

is an asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval function for T_0 at s_0.

Proof. The proof is the same as for Theorem 2.9.20.

3.3 Accelerated lifetime experiments with exponential distribution

At first, we consider also the exponential distribution which is the simplest lifetime distribution so that

$$f_{\theta,s}(t) = \lambda_\theta(s) \exp(-\lambda_\theta(s)t)$$

is the lifetime density. Here the link function $\lambda_\theta : \mathcal{S} \to (0,\infty)$ is known up to the parameter vector $\theta \in \mathbb{R}^r$. The expected lifetime is then

$$E_\theta(T_n) = \frac{1}{\lambda_\theta(s_n)}.$$

Simple reasonable functions for λ_θ are the following:

$$\begin{aligned}
\lambda_\theta(s) &= \theta s, \quad \theta \in (0,\infty), && (3.9)\\
\lambda_\theta(s) &= \theta_0 + \theta_1 s, \quad \theta = (\theta_0,\theta_1)^\top \in (0,\infty)^2, && (3.10)\\
\lambda_\theta(s) &= \exp(\theta_0 + \theta_1 s), \quad \theta = (\theta_0,\theta_1)^\top \in \Re \times (0,\infty), && (3.11)\\
\lambda_\theta(s) &= \frac{1}{\exp(g_\theta(s))}, && (3.12)
\end{aligned}$$

where g_θ is given by (3.2) to ((3.7). All these functions ensure that the expected life time is decreasing with increasing stress s. The function given by (3.9) provides an infinite life time if there is no stress while function (3.10) is more flexible allowing a finite expected life time for no stress. A similar flexibility is achieved by function (3.11), however without allowing infinte expected life time at $s = 0$. Function (3.11) is function proposed by Wöhler.

We derive now the maximum likelihood estimator of θ given by

$$\widehat{\theta} := \arg\max_\theta L_\theta(z_*, d_*, s_*)$$

and the information matrix.

Since the survival function for the exponential distribution satisfies $S_{\theta,s}(t) = \exp(-\lambda_\theta(s)t)$, the loglikelihood function has the form

$$\log L_\theta(z_*, d_*, s_*) = \sum_{d_n=1} (\log \lambda_\theta(s_n) - \lambda_\theta(s_n) z_n) + \sum_{d_n=0} (-\lambda_\theta(s_n)c) = \sum_{n=1}^N l(\theta, t_n, s_n)$$

with

$$l(\theta, t, s) := (\log \lambda_\theta(s) - \lambda_\theta(s)t)\, \mathbb{1}_{[0,c]}(t) - \lambda_\theta(s)c\, \mathbb{1}_{(c,\infty)}(t).$$

The maximum likelihood estimator $\widehat{\theta}$ is a solution of

$$\sum_{n=1}^N \dot{l}(\widehat{\theta}, t_n, s_n) = 0,$$

where

$$\dot{l}(\theta,t,s) := \frac{\partial}{\partial\theta} l(\theta,t,s) = \frac{\partial}{\partial\theta}\lambda_\theta(s)\left[\left(\frac{1}{\lambda_\theta(s)} - t\right) 1_{[0,c]}(t) - c\, 1_{(c,\infty)}(t)\right].$$

Set also

$$\begin{aligned}\ddot{l}(\theta,t,s) := \frac{\partial}{\partial\theta}\dot{l}(\theta,t,s) = \frac{\partial^2}{\partial^2\theta}\lambda_\theta(s)&\left[\left(\frac{1}{\lambda_\theta(s)} - t\right) 1_{[0,c]}(t) - c\, 1_{(c,\infty)}(t)\right]\\ &+\frac{\partial}{\partial\theta}\lambda_\theta(s)\frac{\partial}{\partial\theta}\lambda_\theta(s)^\top\left(-\frac{1}{\lambda_\theta(s)^2}\right) 1_{[0,c]}(t).\end{aligned}$$

3.3.1 Lemma (See for the proof Lemma 2.9.6)

a) $\displaystyle\int_0^c \lambda e^{-\lambda y} dy = -e^{-\lambda y}\Big|_0^c = 1 - e^{-\lambda c}.$

b) $\displaystyle\int_0^c y\,\lambda e^{-\lambda y} dy = -c\,e^{-\lambda c} + \frac{1}{\lambda}\left(1 - e^{-\lambda c}\right).$

c) $\displaystyle\int_0^c y^2\,\lambda e^{-\lambda y} dy = e^{-\lambda c}\left[-c^2 - \frac{2c}{\lambda} - \frac{2}{\lambda^2}\right] + \frac{2}{\lambda^2}.$

d) $\displaystyle\int_0^c \left(\frac{1}{\lambda} - y\right)\lambda e^{-\lambda y} dy - c\int_c^\infty \lambda e^{-\lambda y} dy = 0.$

e) $\displaystyle\int_0^c \left(\frac{1}{\lambda} - y\right)^2\lambda e^{-\lambda y} dy + c^2\int_c^\infty \lambda e^{-\lambda y} dy = \frac{1}{\lambda^2}\left(1 - e^{-\lambda c}\right).$

3.3.2 Lemma
Let be S a random variable giving the stress levels s by the distribution δ and T the life time which has the exponential distribution with parameter $\lambda_\theta(s)$ as conditional distribution given $S = s$. Then we have

$$\begin{aligned}I_\theta(\delta) &:= E_\theta\left(\dot{l}(\theta,T,S)\,\dot{l}(\theta,T,S)^\top\right)\\ &= \int \frac{1}{\lambda_\theta(s)^2}\left(1 - e^{-\lambda_\theta(s)c}\right)\frac{\partial}{\partial\theta}\lambda_\theta(s)\frac{\partial}{\partial\theta}\lambda_\theta(s)^\top\delta(ds) = -\,E_\theta\left(\ddot{l}(\theta,T,S)\right).\end{aligned}$$

Proof Since $1\!\!1_{[0,c]}(t)1\!\!1_{(c,\infty)}(t)=0$, we have with Lemma 3.3.1 e)

$$\begin{aligned}
&E_\theta\left(\dot l(\theta,T,S)\,\dot l(\theta,T,S)^\top\right)\\
=&\int\int \frac{\partial}{\partial\theta}\lambda_\theta(s)\frac{\partial}{\partial\theta}\lambda_\theta(s)^\top\left[\left(\frac{1}{\lambda_\theta(s)}-t\right)1\!\!1_{[0,c]}(t)-c\,1\!\!1_{(c,\infty)}(t)\right]^2\\
&\qquad\qquad\qquad\qquad\qquad\cdot\,\lambda_\theta(s)e^{-\lambda_\theta(s)t}\,dt\,\delta(ds)\\
=&\int \frac{\partial}{\partial\theta}\lambda_\theta(s)\frac{\partial}{\partial\theta}\lambda_\theta(s)^\top\left[\int_0^c\left(\frac{1}{\lambda_\theta(s)}-t\right)^2\lambda_\theta(s)e^{-\lambda_\theta(s)t}dt\right.\\
&\left.+c^2\int_c^\infty\lambda_\theta(s)e^{-\lambda_\theta(s)t}dt\right]\delta(ds)\\
=&\int\frac{1}{\lambda_\theta(s)^2}\left(1-e^{-\lambda_\theta(s)c}\right)\frac{\partial}{\partial\theta}\lambda_\theta(s)\frac{\partial}{\partial\theta}\lambda_\theta(s)^\top\delta(ds).
\end{aligned}$$

Lemma 3.3.1 a) and d) implies

$$\begin{aligned}
&E_\theta\left(\ddot l(\theta,T,S)\right)\\
=&\int\int\left\{\frac{\partial^2}{\partial^2\theta}\lambda_\theta(s)\left[\left(\frac{1}{\lambda_\theta(s)}-t\right)1_{[0,c]}(t)-c\,1_{(c,\infty)}(t)\right]\right.\\
&\qquad\left.+\frac{\partial}{\partial\theta}\lambda_\theta(s)\frac{\partial}{\partial\theta}\lambda_\theta(s)^\top\left(-\frac{1}{\lambda_\theta(s)^2}\right)1_{[0,c]}(t)\right\}\lambda_\theta(s)e^{-\lambda_\theta(s)t}dt\,\delta(ds)\\
=&\int\left\{\frac{\partial^2}{\partial^2\theta}\lambda_\theta(s)\left[\int_0^c\left(\frac{1}{\lambda_\theta(s)}-t\right)\lambda_\theta(s)e^{-\lambda_\theta(s)t}dt-c\int_c^\infty\lambda_\theta(s)e^{-\lambda_\theta(s)t}dt\right]\right.\\
&\qquad\left.+\frac{\partial}{\partial\theta}\lambda_\theta(s)\frac{\partial}{\partial\theta}\lambda_\theta(s)^\top\left(-\frac{1}{\lambda_\theta(s)^2}\right)\int_0^c\lambda_\theta(s)e^{-\lambda_\theta(s)t}dt\right\}\delta(ds)\\
=&\ 0-\int\frac{1}{\lambda_\theta(s)^2}\left(1-e^{-\lambda_\theta(s)c}\right)\frac{\partial}{\partial\theta}\lambda_\theta(s)\frac{\partial}{\partial\theta}\lambda_\theta(s)^\top\delta(ds).\ \square
\end{aligned}$$

3.4 Accelerated lifetime experiments with lognormal distribution

Here we will consider only uncensored independent observations.

If $T \sim \mathcal{LN}(\mu_\theta(s), \sigma^2)$ then $\ln(T) \sim \mathcal{N}(\mu(s), \sigma)$ and $\ln E(T) = \ln\left(\exp(\mu_\theta(s) + \frac{1}{2}\sigma^2)\right) = \mu_\theta(s) + \frac{1}{2}\sigma^2$. Hence we can work with normally distributed random variables $Y_0 = \ln(T_0) \sim \mathcal{N}(\mu_\theta(s_0), \sigma^2), Y_1 = \ln(T_1) \sim \mathcal{N}(\mu_\theta(s_1), \sigma^2), \ldots, Y_N = \ln(T_N) \sim \mathcal{N}(\mu_\theta(s_N), \sigma^2)$ with

$$\mu_\theta(s) = g_\theta(s),$$

where g_θ is given by (3.2) to (3.7), for example. We should have only in mind that the unknown parameter are θ and σ^2 here.

Classical linear model
If

$$\mu_\theta(s) = x(s)^\top \theta$$

then we have a classical linear model.

3.4.1 Example
If $Y_n = \ln(T_n) \sim \mathcal{N}(\mu_\theta(s_n), \sigma^2)$ then $\ln(E_\theta(T_n)) = \mu_\theta(s_n) + \frac{1}{2}\sigma^2$ according to Theorem 2.1.18. Then we get in particular the Basquin model $\ln(E_\theta(T_n)) = \theta_0 - \theta_1 \ln(s_n)$ if $\theta_0 = \frac{1}{2}\sigma^2$ and $\mu_\theta(s_n) = -\theta_1 \ln(s_n)$.

Let be $\theta \in \mathbb{R}^r$, $y_* = (y_1, \ldots, y_N)^\top$ the realization of $Y_* = (Y_1, \ldots, Y_N)^\top$ and set $s_* = (s_1, \ldots, s_N)^\top$, $X = (x(s_1), \ldots, x(s_N))^\top$. Then the maximum likelihood estimator $\widehat{\theta}$ for θ is the least squares estimator, i.e. $\widehat{\theta} = (X^\top X)^{-1} X^\top y_*$. Moreover,

$$\widehat{\sigma}^2 = \frac{1}{N-r} \sum_{n=1}^{N} (y_n - x(s_n)^\top \widehat{\theta})^2$$

is an unbiased estimator of σ^2. Then $\mu_{\widehat{\theta}}(s_0) = x(s_0)^\top \widehat{\theta}$ is an unbiased estimator for $\mu_\theta(s_0) = x(s_0)^\top \theta$ and $\mu_{\widehat{\theta}}(s_0) + \frac{1}{N}\sum_{n=1}^{N} (y_n - x(s_n)^\top \widehat{\theta})^2$ is the maximum likelihood estimator for $\ln(E(T_0))$.

3.4.2 Theorem
If $\theta \in \mathbb{R}^r$, then $\mathbb{C}$ given by

$$\mathbb{C}(y_*, s_*) = \left[\; x(s_0)^\top \widehat{\theta} \pm \sqrt{x(s_0)^\top (X^\top X)^{-1} x(s_0)\, \widehat{\sigma}^2}\;\; t_{N-r;1-\alpha/2} \;\right]$$

is a $(1-\alpha)$-confidence interval function for $\mu_\theta(s_0) = x(s_0)^\top \theta$.

Proof. Since $Y_* \sim \mathcal{N}(X\theta, \sigma^2 I_{N \times N})$ we have

$$E_\theta(x(s_0)^\top \widehat{\theta}) = x(s_0)^\top\, (X^\top X)^{-1} X^\top X\, \theta = x(s_0)^\top \theta$$

and

$$\mathrm{var}_\theta(x(s_0)^\top \widehat{\theta}) = \sigma^2\, x(s_0)^\top\, (X^\top X)^{-1}\, x(s_0),$$

so that

$$\frac{x(s_0)^\top \widehat{\theta} - x(s_0)^\top \theta}{\sqrt{\sigma^2\, x(s_0)^\top\,(X^\top X)^{-1}\, x(s_0)}} \sim \mathcal{N}(0,1).$$

Since $\widehat{\theta}$ and $\widehat{\sigma}^2$ are independent and $(N-r)\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-r}$, we obtain

$$\frac{x(s_0)^\top \widehat{\theta} - x(s_0)^\top \theta}{\sqrt{\widehat{\sigma}^2\, x(s_0)^\top\,(X^\top X)^{-1}\, x(s_0)}} \sim t_{N-r}. \ \square$$

A confidence set for $\ln(E_\theta(T_0)) = \mu_\theta(s_0) + \frac{1}{2}\sigma^2 = g_\theta(s_0)$ or $E_\theta(T_0) = \exp(\mu_\theta(s_0) + \frac{1}{2}\sigma^2)$ is not that easy to obtain. However for prediction intervals, the transfer from Y_0 to T_0 is more easy.

3.4.3 Theorem
If $\theta \in \mathbb{R}^r$, then $\mathbb{P}$ given by

$$\mathbb{P}(y_*, s_*) = \left[\ x(s_0)^\top \widehat{\theta} \pm \sqrt{(1 + x(s_0)^\top (X^\top X)^{-1} x(s_0))\ \widehat{\sigma}^2}\ \ t_{N-r;1-\alpha/2}\ \right]$$

is a $(1-\alpha)$-predition interval function for Y_0 at s_0.

Proof. Since $E_\theta(Y_0) = x(s_0)^\top \theta$, we have

$$E_\theta(Y_0 - x(s_0)^\top \widehat{\theta}) = 0$$

and

$$\mathrm{var}_\theta(Y_0 - x(s_0)^\top \widehat{\theta}) = \sigma^2 + \sigma^2\, x(s_0)^\top\,(X^\top X)^{-1}\, x(s_0)$$

since Y_0 and $Y_* = (Y_1, \ldots, Y_N)^\top$ are independent. Hence we obtain

$$\frac{Y_0 - x(s_0)^\top \widehat{\theta}}{\sqrt{\sigma^2\,(1 +\, x(s_0)^\top\,(X^\top X)^{-1}\, x(s_0))}} \sim \mathcal{N}(0,1).$$

Since $Y_0 - x(s_0)^\top \widehat{\theta}$ and $\widehat{\sigma}^2$ are independent and $(N-r)\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-r}$, we obtain

$$\frac{Y_0 - x(s_0)^\top \widehat{\theta}}{\sqrt{\widehat{\sigma}^2\,(1 +\, x(s_0)^\top\,(X^\top X)^{-1}\, x(s_0))}} \sim t_{N-r}. \ \square$$

3.4.4 Corollary
If $\theta \in \mathbb{R}^r$and $\mathbb{P}$ given by

$$\mathbb{P}(y_*, s_*) = [L(y_*, s_*)\ ,\ U(y_*, s_*)] = \left[\ x(s_0)^\top \widehat{\theta} \pm \sqrt{(1 + x(s_0)^\top (X^\top X)^{-1} x(s_0))\ \widehat{\sigma}^2}\ \ t_{N-r;1-\alpha/2}\ \right]$$

is a $(1-\alpha)$-prediction interval for Y_0 based on y_* and $y_* = (\ln(t_1), \ldots, \ln(t_N))^\top$, then

$$\mathbb{P}_0(t_*, s_*) = [\ \exp(L(y_*, s_*))\ ,\ \exp(U(y_*, s_*))\ ]$$

is a $(1-\alpha)$-predition interval function for T_0 at s_0 based on $t_* = (t_1, \ldots, t_N)$.

Proof. Since $T_0 = \exp(Y_0)$, we get

$$P_\theta(T_0 \in \mathbb{P}_0(T_*, s_*)) = P_\theta\left(\exp\left(L(Y_*, s_*)\right) \leq T_0 \leq \exp\left(U(Y_*, s_*)\right)\right)$$
$$= \quad P_\theta\left(L(Y_*, s_*) \leq Y_0 \leq U(Y_*, s_*)\right) = 1 - \alpha. \Box$$

Nonlinear model
If $\mu_\theta(s)$ is not linear in $\theta \in \mathbb{R}^r$ then we have a classical nonlinear model given by

$$Y_n = \mu_\theta(s_n) + E_n$$

with $E_n \sim \mathcal{N}(0, \sigma^2)$ for $n = 1, \ldots, N$.

3.4.5 Example
If $Y_n = \ln(T_n) \sim \mathcal{N}(\mu_\theta(s_n), \sigma^2)$ with $\ln(E_\theta(T_n)) = \mu_\theta(s_n) + \frac{1}{2}\sigma^2$ according to Theorem 2.1.18, then we get in particular the model $\ln(E_\theta(T_n)) = \theta_0 + \theta_1 \left(\frac{1}{s_n}\right)^{\theta_2}$ if $\theta_0 = \frac{1}{2}\sigma^2$ and $\mu_\theta(s_n) = \theta_1 \left(\frac{1}{s_n}\right)^{\theta_2}$.

The Taylor expansion provides

$$Y_n = \mu_\theta(s_n) + E_n \approx \mu_{\theta_*}(s_n) + \dot{\mu}_{\theta_*}(s_n)^\top(\theta - \theta_*) + E_n$$

so that

$$Z_n := Y_n - \mu_{\theta_*}(s_n) = \dot{\mu}_{\theta_*}(s_n)^\top \tilde{\theta} + E_n.$$

Hence Z_n follows approximately a linear model. Using this approximation, approximate confidence intervals and prediction intervals can be derived. Again let $\widehat{\theta}_N$ be the least squares estimator, which is also the maximum likelihood estimator of the nonlinear model, and

$$\widehat{\sigma}_N^2 := \frac{1}{N - r} \sum_{n=1}^{N} (y_n - \mu_{\widehat{\theta}_N}(s_n))^2$$

the estimator of σ^2. Thereby $\widehat{\sigma}_N^2$ is a consistent estimator of σ^2. If the concrete design measure $\delta_N = \sum_{n=1}^N e_{s_n}$ is again converging to the a design δ, then the least squares estimator has an asymptotic normal distribution.

3.4.6 Theorem (Jennrich 1969)
If δ_n converges weakly to δ almost surely and $\widehat{\theta}_N$ is the least squares estimator in the nonlinear model, then

$$\sqrt{N}\left(\widehat{\theta}_N - \theta_*\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0_r, \sigma^2 \tilde{I}_{\theta_*}(\delta)^{-1}\right) \tag{3.13}$$

with $\tilde{I}_{\theta_*}(\delta) := \int \dot{\mu}_{\theta_*}(s)\, \dot{\mu}_{\theta_*}(s)^\top\, \delta(ds)$, $\dot{\mu}_{\theta_*}(s) := \frac{\partial}{\partial\theta}\mu_\theta(s)\big|_{\theta=\theta_*}$ if θ_* is the true parameter.

Proof. The proof follows as for any maximum likelihood estimator. Hence we have only to calculate the information matrix. At first note

$$\begin{aligned}\frac{\partial}{\partial\theta}\ln(f_{\theta,s}(y)) &= \frac{\partial}{\partial\theta}\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(y-\mu_\theta(s))^2\right)\right)\\ &= \frac{\partial}{\partial\theta}\left(-\frac{1}{2\sigma^2}(y-\mu_\theta(s))^2\right) = \frac{1}{\sigma^2}(y-\mu_\theta(s))\frac{\partial}{\partial\theta}\mu_\theta(s)\end{aligned}$$

so that

$$\begin{aligned}I_\theta(\delta) &= \int E\left(\frac{\partial}{\partial\theta}\ln(f_{\theta,s}(Y))\frac{\partial}{\partial\theta}\ln(f_{\theta,s}(Y))^\top\right)\delta(ds)\\ &= \frac{1}{\sigma^2}\int E\left(\left(\frac{Y-\mu_\theta(s)}{\sigma}\right)^2\right)\dot\mu_\theta(s)\,\dot\mu_\theta(s)^\top\delta(ds) = \frac{1}{\sigma^2}\int 1\cdot\dot\mu_\theta(s)\,\dot\mu_\theta(s)^\top\delta(ds). \quad\square\end{aligned}$$

Set $\dot M_{\theta,N} := (\dot\mu_\theta(s_1),\ldots,\dot\mu_\theta(s_N))^\top$ and let be $q_{1-\alpha/2}$ the $(1-\alpha/2)$-quantile of the standard normal distribution.

3.4.7 Theorem
If the support of δ is included in $[s_{min}, s_{max}]$, δ_n converges weakly to δ almost surely, $\dot\mu_\bullet(s)$ as function of θ is equicontinuous for $s\in[s_{min},s_{max}]$, $\widehat\theta_N$ is the least squares estimator in the nonlinear model, then $\mathbb{C}_N$ given by

$$\mathbb{C}_N(y_*,s_*) = \left[\mu_{\widehat\theta_N}(s_0) \pm \sqrt{\widehat\sigma^2\,\dot\mu_{\widehat\theta_N}(s_0)^\top\left(\dot M_{\widehat\theta_N,N}^\top\dot M_{\widehat\theta_N,N}\right)^{-1}\dot\mu_{\widehat\theta_N}(s_0)}\;q_{1-\alpha/2}\right]$$

is an asymptotic $(1-\alpha)$-confidence interval function for $\mu_\theta(s_0)$.

Proof. The asymptotic normality in (3.13) and Theorem 2.8.12 (δ-method) provide

$$\sqrt{N}\left(\mu_{\widehat\theta_N}(s_0)-\mu_{\theta_*}(s_0)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\sigma^2\,\dot\mu_{\theta_*}(s_0)^\top\,\tilde I_{\theta_*}(\delta)^{-1}\,\dot\mu_{\theta_*}(s_0)\right). \tag{3.14}$$

The asymptotic normality in (3.13) implies with Corollary 2.8.7 (iv) also $\widehat\theta_N\xrightarrow{\mathcal{P}}\theta_*$ so that with the equicontinuity of $\dot\mu_\bullet(s)$

$$\frac{1}{N}\dot M_{\widehat\theta_N,N}^\top\dot M_{\widehat\theta_N,N} - \frac{1}{N}\dot M_{\theta_*}^\top\dot M_{\theta_*} = \frac{1}{N}\sum_{n=1}^N\left(\dot\mu_{\widehat\theta_N}(s_n)\,\dot\mu_{\widehat\theta_N}(s_n)^\top - \dot\mu_{\theta_*}(s_n)\,\dot\mu_{\theta_*}(s_n)^\top\right)\xrightarrow{\mathcal{P}}0_{r\times r}$$

and

$$\dot\mu_{\widehat\theta_N}(s_0)\xrightarrow{\mathcal{P}}\dot\mu_{\theta_*}(s_0).$$

Since $\delta_N=\frac{1}{N}\sum_{n=1}^N e_{s_n}$ converges weakly (in distribution) to δ almost surely, we obtain with the Theorem of Portmanteau (Theorem 2.8.3) additionally

$$\frac{1}{N}\dot M_{\theta_*}^\top\dot M_{\theta_*} = \frac{1}{N}\sum_{n=1}^N\dot\mu_{\theta_*}(s_n)\,\dot\mu_{\theta_*}(s_n)^\top = \tilde I_{\theta_*}(\delta_N)\xrightarrow{\mathcal{P}}\tilde I_{\theta_*}(\delta)$$

so that

$$\frac{1}{N}\dot{M}_{\widehat{\theta}_N,N}^{\top}\dot{M}_{\widehat{\theta}_N,N} \xrightarrow{\mathcal{P}} \tilde{I}_{\theta_*}(\delta).$$

This implies

$$N\left(\dot{M}_{\widehat{\theta}_N,N}^{\top}\dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \xrightarrow{\mathcal{P}} \tilde{I}_{\theta_*}(\delta)^{-1}$$

and with $\dot{\mu}_{\widehat{\theta}_N}(s_0) \xrightarrow{\mathcal{P}} \dot{\mu}_{\theta_*}(s_0)$ and $\widehat{\sigma}_N^2 \xrightarrow{\mathcal{P}} \sigma^2$

$$\widehat{\sigma}_N^2\, N\, \dot{\mu}_{\widehat{\theta}_N}(s_0)^{\top}\left(\dot{M}_{\widehat{\theta}_N,N}^{\top}\,\dot{M}_{\widehat{\theta}_N,N}\right)^{-1}\,\dot{\mu}_{\widehat{\theta}_N}(s_0) \xrightarrow{\mathcal{P}} \sigma^2\,\dot{\mu}_{\theta_*}(s_0)^{\top}\,\tilde{I}_{\theta_*}(\delta)^{-1}\,\dot{\mu}_{\theta_*}(s_0). \tag{3.15}$$

Hence we obtain with (3.14) and Corollary 2.8.8

$$\begin{aligned}&\frac{\mu_{\widehat{\theta}_N}(s_0)-\mu_{\theta_*}(s_0)}{\sqrt{\widehat{\sigma}_N^2\,\dot{\mu}_{\widehat{\theta}_N}(s_0)^{\top}\left(\dot{M}_{\widehat{\theta}_N,N}^{\top}\,\dot{M}_{\widehat{\theta}_N,N}\right)^{-1}\,\dot{\mu}_{\widehat{\theta}_N}(s_0)}}\\ =\;& \frac{\sqrt{N}\left(\mu_{\widehat{\theta}_N}(s_0)-\mu_{\theta_*}(s_0)\right)}{\sqrt{\widehat{\sigma}_N^2\,N\,\dot{\mu}_{\widehat{\theta}_N}(s_0)^{\top}\left(\dot{M}_{\widehat{\theta}_N,N}^{\top}\,\dot{M}_{\widehat{\theta}_N,N}\right)^{-1}\,\dot{\mu}_{\widehat{\theta}_N}(s_0)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).\ \square\end{aligned}$$

3.4.8 Theorem (Simple prediction interval)
If $\widehat{\theta}_N$ is the least squares estimator in the nonlinear model, then $\mathbb{P}_N$ given by

$$\mathbb{P}_N(y_*, s_*) = \left[\mu_{\widehat{\theta}_N}(s_0) - \widehat{\sigma}_N\, q_{1-\alpha/2}, \mu_{\widehat{\theta}_N}(s_0) + \widehat{\sigma}_N\, q_{1-\alpha/2}\right]$$

is an asymptotic $(1-\alpha)$-predition interval function for Y_0 at s_0.

Proof. Since

$$\widehat{\sigma}_N^2 \xrightarrow{\mathcal{P}} \sigma^2,\ \ \mu_{\theta_*}(s_0) - \mu_{\widehat{\theta}_N}(s_0) \xrightarrow{\mathcal{P}} 0,\ \ \frac{Y_0 - \mu_{\theta_*}(s_0)}{\sigma} = \frac{E_0}{\sigma} \sim \mathcal{N}(0,1)$$

we get

$$\frac{Y_0 - \mu_{\widehat{\theta}_N}(s_0)}{\widehat{\sigma}_N} = \frac{E_0}{\widehat{\sigma}_N} + \frac{\mu_{\theta_*}(s_0) - \mu_{\widehat{\theta}_N}(s_0)}{\widehat{\sigma}_N} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).\ \square$$

A wider asymptotic prediction interval is obtained by the following approximations.

Convergence (3.14) implies

$$\mu_{\widehat{\theta}_N}(s_0) - \mu_{\theta_*}(s_0) \approx \mathcal{N}\left(0, \sigma^2\,\frac{1}{N}\,\dot{\mu}_{\theta_*}(s_0)^{\top}\,I_{\theta_*}(\delta)^{-1}\,\dot{\mu}_{\theta_*}(s_0)\right).$$

Since $Y_0 = \mu_{\theta_*}(s_0) + E_0$ with $E_0 \sim \mathcal{N}(0, \sigma^2)$ and E_0 is independent of $\widehat{\theta}_N$, we have

$$\mu_{\widehat{\theta}_N}(s_0) - Y_0 = \mu_{\widehat{\theta}_N}(s_0) - \mu_{\theta_*}(s_0) - E_0 \approx \mathcal{N}\left(0, \sigma^2 \frac{1}{N}\, \dot{\mu}_{\theta_*}(s_0)^\top\, I_{\theta_*}(\delta)^{-1}\, \dot{\mu}_{\theta_*}(s_0) + \sigma^2\right).$$

As shown in the proof of Theorem 3.4.7, $\dot{\mu}_{\theta_*}(s_0)^\top\, I_{\theta_*}(\delta)^{-1}\, \dot{\mu}_{\theta_*}(s_0)$ can be approximated by $N\, \dot{\mu}_{\widehat{\theta}_N}(s_0)^\top \left(\dot{M}_{\widehat{\theta}_N,N}^\top\, \dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \dot{\mu}_{\widehat{\theta}_N}(s_0)$ and σ^2 by $\widehat{\sigma}_N^2$ so that

$$\mu_{\widehat{\theta}_N}(s_0) - Y_0 \approx \mathcal{N}\left(0, \widehat{\sigma}_N^2 \left(1 + \dot{\mu}_{\widehat{\theta}_N}(s_0)^\top \left(\dot{M}_{\widehat{\theta}_N,N}^\top \dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \dot{\mu}_{\widehat{\theta}_N}(s_0)\right)\right),$$

or, respectively,

$$\frac{\mu_{\widehat{\theta}_N}(s_0) - Y_0}{\sqrt{\widehat{\sigma}_N^2 \left(1 + \dot{\mu}_{\widehat{\theta}_N}(s_0)^\top \left(\dot{M}_{\widehat{\theta}_N,N}^\top\, \dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \dot{\mu}_{\widehat{\theta}_N}(s_0)\right)}} \approx \mathcal{N}(0,1).$$

3.4.9 Theorem
If $\widehat{\theta}_N$ is the least squares estimator in the nonlinear model, then $\mathbb{P}_N$ given by

$$\mathbb{P}_N(y_*, s_*) = \left[\mu_{\widehat{\theta}_N}(s_0) \pm \sqrt{\widehat{\sigma}_N^2 \left(1 + \dot{\mu}_{\widehat{\theta}_N}(s_0)^\top \left(\dot{M}_{\widehat{\theta}_N,N}^\top \dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \dot{\mu}_{\widehat{\theta}_N}(s_0)\right)}\; q_{1-\alpha/2}\right]$$

is an asymptotic $(1-\alpha)$-predition interval function for Y_0 at s_0.

Proof. Property (3.15) implies

$$\widehat{\sigma}_N^2\, \dot{\mu}_{\widehat{\theta}_N}(s_0)^\top \left(\dot{M}_{\widehat{\theta}_N,N}^\top\, \dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \dot{\mu}_{\widehat{\theta}_N}(s_0) \xrightarrow{\mathcal{P}} 0$$

so that the result follows from Theorem 3.4.8. □

For estimating the stress level s_L such that $\mu_\theta(s_L) = L$, i.e. for estimating $a_L(\theta) := \mu_\theta^{-1}(L)$, we can use again $a_L(\widehat{\theta}_N) = \mu_{\widehat{\theta}_N}^{-1}(L)$ as estimator. Set $\dot{a}_L(\theta_*) = \frac{\partial}{\partial\theta} a_L(\theta)\Big|_{\theta=\theta_*}$.

3.4.10 Theorem
If $\widehat{\theta}_N$ is the least squares estimator in the nonlinear model, then $\mathbb{C}_N$ given by

$$\mathbb{C}_N(y_*, s_*) = \left[a_L(\widehat{\theta}_N) \pm \sqrt{\widehat{\sigma}_N^2\, \dot{a}_L(\widehat{\theta}_N)^\top \left(\dot{M}_{\widehat{\theta}_N,N}^\top \dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \dot{a}_L(\widehat{\theta}_N)}\; q_{1-\alpha/2}\right]$$

is an asymptotic $(1-\alpha)$-confidence interval function for $a_L(\theta) := \mu_\theta^{-1}(L)$.

Proof. The asymptotic normality in (3.13) and Theorem 2.8.12 (δ-method) provide

$$\sqrt{N}\left(a_L(\widehat{\theta}_N) - a_L(\theta_*)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\, \dot{a}_L(\theta_*)^\top\, I_{\theta_*}(\delta)^{-1}\, \dot{a}_L(\theta_*)\right).$$

and thus, like in the proof of Theorem 3.4.7,

$$\begin{aligned}
&\frac{a_L(\widehat{\theta}_N) - a_L(\theta_*)}{\sqrt{\widehat{\sigma}_N^2\, \dot{a}_L(\theta_*)^\top \left(\dot{M}_{\widehat{\theta}_N,N}^\top\, \dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \dot{a}_L(\theta_*)}} \\
= \quad & \frac{\sqrt{N}\left(a_L(\widehat{\theta}_N) - a_L(\theta_*)\right)}{\sqrt{\widehat{\sigma}_N^2\, N\, \dot{a}_L(\theta_*)^\top \left(\dot{M}_{\widehat{\theta}_N,N}^\top\, \dot{M}_{\widehat{\theta}_N,N}\right)^{-1} \dot{a}_L(\theta_*)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \;\Box
\end{aligned}$$

If the stress level s_L with $\ln(E_\theta(T)) = L$ should be estimated, then we should know σ^2 since $E_\theta(T) = \exp(\mu_\theta(s) + \frac{1}{2}\sigma^2)$ and thus $\ln(E_\theta(T)) = \mu_\theta(s) + \frac{1}{2}\sigma^2 = L$ or $\mu_\theta(s) = L - \sigma^2$. There are two possibilities: neglecting $\sigma^2 > 0$ so that the lifetime at s_L satisfies $\ln(E_\theta(T)) > L$ or to use $\tilde{L} = L - \widehat{\sigma}^2$.

3.5 Nonparametric methods

If the class of distribution is not known then a nonparametric approach can be used. However, prediction is then not possible as in the case of one stress level. A very successful approach is available with the Cox model of proportional hazard rate. It can be used also if not only the stress level is the covariate but also if other and several covariates shall be included in the model. Let $x \in \mathbb{R}^r$ be a r-dimensional covariate.

3.5.1 Definition (Cox model of proportional hazard rate)
In a Cox model of proportional hazard rate it is assumed that $T_1, \ldots, T_N$ are independent with hazard functions $h_\theta(\cdot, x_1), \ldots h_\theta(\cdot, x_N)$ satisfying

$$h_\theta(t, x) = h_0(t)\ \exp(x^\top \theta),$$

for all $t \in \mathbb{R}_+$, $x \in \mathbb{R}^r$, where $\theta \in \mathbb{R}^r$ and $h_0 : \mathbb{R}_+ \to \mathbb{R}$ is the so called baseline hazard rate.

In a Cox model of proportional hazard rate, we have

$$\frac{h\theta(t,x)}{h_\theta(t, \tilde{x}} = \exp((x - \tilde{x})^\top \theta)$$

so that the quotient is independent of t and depends only on the covariates. This is of course a special assumption which is not always satisfied so that it must be verified.

We assume that we have randomly censored and uncensored observations $(z_1, d_1, x_1), \ldots, (z_N, d_N, x_N)$ where $z_n = \min(t_n, c_n)$, $d_n = \mathbb{1}_{(-\infty, c_n]}(t_n)$ for $n = 1, \ldots, N$. Let $0 \le \tau_1 < \tau_2 < \tau_3 < \ldots < \tau_I$ be the observed event times (failures/deaths) and set

$$R_i := \{n \in \{1, \ldots, N\};\ z_n \ge \tau_i\}.$$

If the event times are pairwise different then the relation between event time and covariate is unique so that $x_{(1)}, \ldots, x_{(I)}$ can be defined as the covariates corresponding to $\tau_1, \ldots, \tau_I$.

3.5.2 Definition (Partial maximum likelihood estimator in the Cox model)
If the event times are pairwise different, then the (partial) maximum likelihood estimator in a Cox model of proportional hazard rate is given as the parameter $\widehat{\theta}$ which maximizes the partial likelihood function

$$l_P(\theta) := \prod_{i=1}^{I} \frac{\exp(x_{(i)}^\top \theta)}{\sum_{n \in R_i} \exp(x_n^\top \theta)}.$$

For the motivation of the partial likelihood we need the following lemma.

3.5.3 Lemma
The function

$$g(h) = h\,\exp(-hk)$$

with $k > 0$ *has a unique maximum at* $h = \frac{1}{k}$.

Proof. We have

$$\begin{aligned} g'(h) &= \exp(-hk) + h\,\exp(-hk)\,(-k) = \exp(-hk)\,(1 - hk) = 0 \Leftrightarrow h = \frac{1}{k}, \\ g''(h) &= \exp(-hk)\,(-k)\,(1 - hk) - \exp(-hk)\,k = \exp(-hk)\,(-2k + hk^2), \end{aligned}$$

so that

$$g''\left(\frac{1}{k}\right) = \exp(-hk)\,(-2k + k) < 0. \square$$

3.5.4 Remark (Motivation of the partial likelihood function, see Klein and Moeschberger 2003, p. 258)
Define the cumulative baseline hazard function as

$$H_0(t) := \int_0^\infty h_0(t)\,dt.$$

Then the cumulative hazard function $H(\cdot, x)$ at x is given by

$$H_\theta(t, x) = H_0(t)\exp(x^\top\theta).$$

Theorem 2.1.7 provides for the survival function $S(\cdot, x)$ at x

$$S_\theta(t, x) = \exp(-H_\theta(t)) = \exp\left(-H_0(t)\,\exp(x^\top\theta)\right).$$

and for the density $f_\theta(\cdot, x)$ at x

$$f_\theta(t, x) = h_\theta(t, x) S_\theta(t, x).$$

If we assume that the censoring variables $C_1, \ldots, C_n$ are independent of $T_1, \ldots, T_N$ and not depending on θ then the likelihood function for θ is given by, see Section 2.5

$$\begin{aligned} l(\theta) = \prod_{n=1}^N f_\theta(z_n, x_n)^{d_n}\, S_\theta(z_n, x_n)^{1-d_n} &= \prod_{n=1}^N h_\theta(z_n, x_n)^{d_n}\, S_\theta(z_n, x_n) \\ = \left(\prod_{i=1}^I h_0(\tau_i)\exp(x_{(i)}^\top\theta)\right) & \left(\prod_{n=1}^N \exp\left(-H_0(z_n)\,\exp(x_n^\top\theta)\right)\right) \\ = \left(\prod_{i=1}^I h_0(\tau_i)\exp(x_{(i)}^\top\theta)\right) & \exp\left(-\sum_{n=1}^N H_0(z_n)\,\exp(x_n^\top\theta)\right). \end{aligned}$$

This expression is maximized with respect to the function h_0 if $h_0(t) = 0$ for all $t \notin \{\tau_1, \dots, \tau_I\}$ and $h(\tau_i) = h_{0i}$ for $i = 1, \dots, I$ so that

$$H_0(t) = \sum_{i;\,\tau_i \le t} h_{0i}. \tag{3.16}$$

Hence we get for the maximized likelihood function

$$\begin{aligned}
\max_{h:\mathbb{R}_+\to\mathbb{R}} l(\theta) &= \left(\prod_{i=1}^{I} h_{0i} \exp(x_{(i)}^\top\theta)\right) \exp\left(-\sum_{n=1}^{N} \sum_{i;\,\tau_i\le z_n} h_{0i}\, \exp(x_n^\top\theta)\right)\\
&= \left(\prod_{i=1}^{I} h_{0i} \exp(x_{(i)}^\top\theta)\right) \exp\left(-\sum_{i=1}^{I} \sum_{n;\,\tau_i\le z_n} h_{0i}\, \exp(x_n^\top\theta)\right)\\
&= \left(\prod_{i=1}^{I} h_{0i} \exp(x_{(i)}^\top\theta)\right) \exp\left(-\sum_{i=1}^{I} h_{0i} \sum_{n\in R_i} \exp(x_n^\top\theta)\right)\\
&= \left(\prod_{i=1}^{I} \exp(x_{(i)}^\top\theta)\right) \left(\prod_{i=1}^{I} h_{0i}\, \exp\left(-h_{0i} \sum_{n\in R_i} \exp(x_n^\top\theta)\right)\right)\\
&= \left(\prod_{i=1}^{I} \exp(x_{(i)}^\top\theta)\right) \left(\prod_{i=1}^{I} g_i(h_{0i})\right),
\end{aligned}$$

where

$$g_i(h) := h\, \exp(-h\, c_i), \quad c_i := \sum_{n\in R_i} \exp(x_n^\top\theta)$$

for $i = 1, \dots, I$. Lemma 3.5.3 implies

$$\max_{h\in\mathbb{R}} g_i(h) = g_i\left(\frac{1}{c_i}\right) = \frac{1}{c_i}\, \exp(1) \tag{3.17}$$

so that maximizing the likelihood function with respect to $h_{01}, \dots, h_{0I}$ leads to

$$\max_{h_{01},\dots,h_{0I}} \max_{h:\mathbb{R}_+\to\mathbb{R}} l(\theta) = \left(\prod_{i=1}^{I} \exp(x_{(i)}^\top\theta)\right) \left(\prod_{i=1}^{I} \frac{1}{c_i}\, \exp(1)\right) = e^I \prod_{i=1}^{I} \frac{\exp(x_{(i)}^\top\theta)}{\sum_{n\in R_i} \exp(x_n^\top\theta)}$$

which is proportional to the partial likelihood $l_P(\theta)$.

3.5.5 Remark (Estimation of the baseline cumulative hazard function and the survival function) If $\widehat{\theta}$ is the partial maximum likelihood estimator, then (3.16) and (3.17) imply that $\widehat{H}_0$ given by

$$\widehat{H}_0(t) := \sum_{i;\,\tau_i\le t} \frac{1}{\sum_{n\in R_i} \exp(x_n^\top\widehat{\theta})}$$

is an estimate of the baseline hazard function. This estimator is also called Breslow estimator for the baseline hazard function, see Klein and Moeschberger 2003, p. 258 and 283. Then $\widehat{S}(\cdot, x)$ given by

$$\widehat{S}(t, x) := \exp\left(-\widehat{H}_0(t)\, \exp(x^\top \widehat{\theta})\right)$$

is the estimate of the survival function $S_\theta(\cdot, x)$ at x.

If the lifetime follows a continuous distribution then the time points of events should be different. However, this is often not the case because of rounding. Then so called *ties* appear. If ties are present then the partial likelihood must be modified. Define for $i = 1, \ldots, I$

$$\begin{aligned} B_i &:= \{n \in \{1, \ldots, N\}; z_n = \tau_i \text{ and } d_n = 1\}, \\ b_i &:= \sharp B_i, \\ \xi_i &:= \sum_{n \in B_i} x_n, \\ y_i &:= \sharp\{n \in \{1, \ldots, N\}; z_n \geq \tau_i\} = \sharp R_i. \end{aligned}$$

We have $b_i = 1$ for $i = 1, \ldots, I$ if no ties appear.

3.5.6 Definition (Partial maximum likelihood estimator / Breslow estimator in the Cox model with ties, see see Klein and Moeschberger 2003, p. 259)
The general (partial) maximum likelihood estimator (Breslow estimator) in a Cox model of proportional hazard rate is given as the parameter $\widehat{\theta}_B$ which maximizes the partial likelihood function of Breslow given by

$$l_B(\theta) := \prod_{i=1}^{I} \frac{\exp(\xi_i^\top \theta)}{\left(\sum_{n \in R_i} \exp(x_n^\top \theta)\right)^{b_i}}.$$

3.5.7 Remark
It is clear that $l_P(\theta) = l_B(\theta)$ holds in the case of no ties. Moreover, we have for the general case

$$l_B(\theta) = \prod_{i=1}^{I} \prod_{n \in B_i} \frac{\exp(x_n^\top \theta)}{\sum_{n \in R_i} \exp(x_n^\top \theta)}.$$

There are also other proposals for the partial likelihood function for the case with ties which generalize the partial likelihood function for the case without ties, see Klein and Moeschberger 2003, p. 259.

3.5.8 Definition (Breslow estimator of the survival function, see see Klein and Moeschberger 2003, p. 283)
The Breslow estimator of the survival function $S_\theta(\cdot, x)$ at x is $\widehat{S}_B(\cdot, x)$ given by

$$\widehat{S}_B(t, x) := \exp\left(-\widehat{H}_{0B}(t)\,\exp(x^\top \widehat{\theta}_B)\right)$$

where

$$\widehat{H}_{0B}(t) := \sum_{i;\ \tau_i \le t} \frac{b_i}{\sum_{n \in R_i} \exp(x_n^\top \widehat{\theta})},$$

and $\widehat{\theta}_B$ is partial maximum likelihood estimator of Breslow.

3.5.9 Remark
Again we have $\widehat{H}_{0B} = \widehat{H}_0$ and $\widehat{S}_B(\cdot, x) = \widehat{S}(\cdot, x)$ in the case of no ties. Moreover, we get

$$\widehat{H}_{0B}(t) = \sum_{i;\ \tau_i \le t} \frac{b_i}{y_i},$$

i.e. the Nelson-Aalen estimator of the cumulative hazard function, in the case where no influence of covariates exists, i.e. $\widehat{\theta}_B = 0$.

To check the model, the so called *Cox-Snell residuals* can be used.

3.5.10 Definition (Cox-Snell residuals, see Klein and Moeschberger 2003, p. 355)
The Cox-Snell residuals are defined as

$$\widehat{r}_n := \widehat{H}_{0B}(t_n)\,\exp(x_n^\top \widehat{\theta}_B)$$

for $n = 1, \ldots, N$, where $\widehat{H}_{0B}$ and $\widehat{\theta}_B$ are the Breslow estimators for the cumulative hazard function and θ, respectively.

3.5.11 Lemma
If T is a continuous lifetime distribution with strictly increasing cumulative distribution function F on $\mathbb{R}_+$ and cumulative hazard function H, then
a) $F(T)$ has a uniform distribution on $[0,1]$,
b) $H(T) \sim \mathcal{E}(1)$, i.e. $H(T)$ has an exponential distribution with $\lambda = 1$.

Proof. a) With F also F^{-1} is continuous and strictly increasing and F^{-1} is the inverse of F on $\mathbb{R}_+$ so that $F(F^{-1}(u)) = u$ for all $u \in [0,1]$. Hence we have for any $u \in [0,1]$

$$P(F(T) \le u) = P(F^{-1}(F(T)) \le F^{-1}(u)) = F(F^{-1}(u)) = u.$$

b) Since g given by $g(z) = 1 - \exp(-z)$ is strictly increasing on $\mathbb{R}_+$, the assertion in a) together with Theorem 2.1.7 provides for any $z \in \mathbb{R}_+$

$$\begin{aligned} & P(H(T) \le z) = P(g(H(T)) \le g(z)) \\ = \quad & P(1 - \exp(-H(T)) \le 1 - \exp(-z)) = P(F(T) \le 1 - \exp(-z)) = 1 - \exp(-z). \ \square \end{aligned}$$

3.5.12 Theorem
If $T_1, \ldots, T_n$ are independent observations from the Cox model given by Definition 3.5.1, $x_1, \ldots, x_N$ are fixed covariates then the theoretical Cox-Snell residuals

$$R_n := H_0(T_n) \exp(x_n^\top \theta)$$

satisfy $R_n \sim \mathcal{E}(1)$ for $n = 1, \ldots, N$ and are independent.

Proof. Since

$$h_\theta(\cdot, x_n) = h_0(\cdot) \exp(x_n^\top \theta),$$

is the hazard function of T_n at x_n, the cumulative hazard function T_n at x_n is

$$H_\theta(\cdot, x_n) = H_0(\cdot) \exp(x_n^\top \theta).$$

Hence Lemma 3.5.11 provides $R_n = H_\theta(T_n, x_n) \sim \mathcal{E}(1)$. The independence of $R_1, \ldots, R_N$ follows from the independence of $T_1, \ldots, T_N$. $\square$

3.5.13 Remark (Model check)
If the model is correct then the Cox-Snell residuals $\widehat{r}_1, \ldots, \widehat{r}_N$ should follow approximately an $\mathcal{E}(1)$ distribution. This can be checked by classical goodness-of-fit tests as presented in Section 2.6. Moreover, since the cumulative hazard function $H_{\mathcal{E}(1)}$ of a $\mathcal{E}(1)$ distribution satisfies $H_{\mathcal{E}(1)}(t) = t$, also the points $(\widehat{r}_n, H_{\mathcal{E}(1)}(\widehat{r}_n))$ for $n = 1, \ldots, N$ can be plotted. If the model is a good approximation then these points should follow a straight line through the origin with a slope of 1.
If there are only few different values of the covariates x_n, say $x_{(1)}, \ldots, x_{(J)}$ then the survival function for each $x_{(j)}$ can be estimated with the methods of Section 2.10 and plotted together. If the proportional hazard model is correct then these survival function should be more or less parallel since they are only shifted by $\exp(x_{(j)}^\top \theta)$. In particular, if some survival functions are crossing then the model of proportional hazard function is doubtful.

3.5.14 Remark (Analysis with `R`)
The `R` package `survival` includes several methods for survival analysis as the Kaplan-Meier and Nelson-Aalen estimator, analysis of Cox models, and parametric accelerated failure time models. In particular, the `R` function `coxph` provides the analysis of the cox model.

Chapter 4

Crack initiation

4.1 Crack detection using surface photos

Figure 4.1 shows a round specimen of steel which was exposed to cyclic load. Photos were obtained from a small inner part during the fatigue experiment. Figure 4.2 shows how micro cracks were initiated and how micro cracks grow during the experiment.

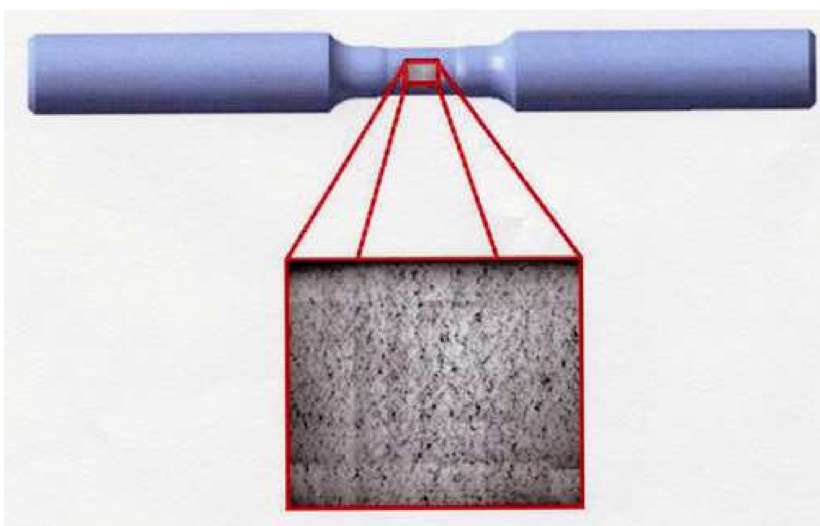

Figure 4.1: Round specimen of steel

With the R package `crackrec` decribed in Gunkel et al. (2012) micro cracks can be detected. Figure 4.3 shows the detected cracks after 18.000 load cycles. With the package `crackrec`, crack cluster can determined. Crack clusters are connected sets of pixels below a given threshhold. Such a crack cluster is shown in Figure 4.4. The longest shortest path through the crack cluster is then a crack path, which is shown on the righthand side of Figure 4.4.

Here is of special interest the micro cracks which are initiated in the beginning. An example of detected micro cracks after 1.000 load cycles is shown in Figure 4.5. In this image 131 cracks were found, where the maximum length of a detected crack consisted of 27.72 pixels. The data set

At the beginning | After 5.000 load cycles | After 18.000 load cycles

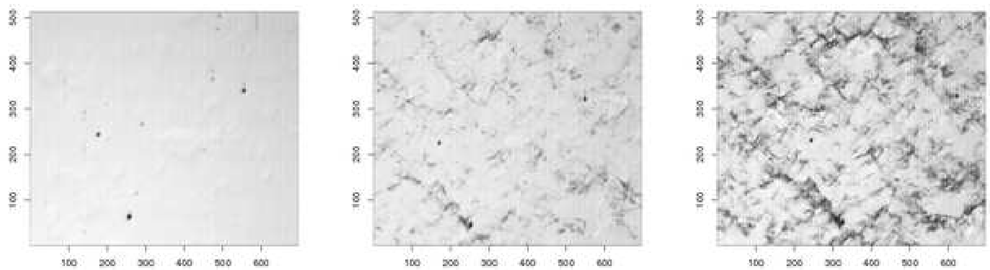

Figure 4.2: Development of micro cracks

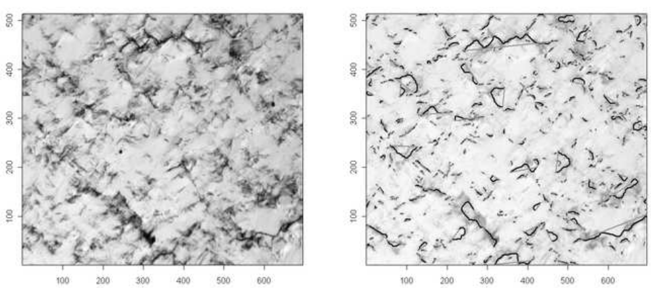

Figure 4.3: Image after 18.000 load cycles without and with detected micro cracks

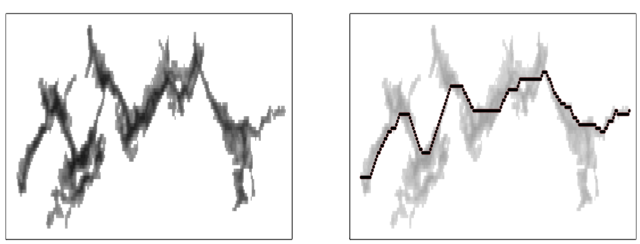

Figure 4.4: The longest shortest path (righthand side) through a crack cluster (lefthand side)

`S01_B30_X_62.425_Y_212.250_erg_cracks.asc` provides the size of the detected crack clusters, the length of the detected crack paths and the coordinates of the start and end points of the crack paths.

The question is, how to model the crack initiation. One possibility is to model the initiated micro cracks by spatial Poisson point process. A Poisson process is a stochastic process.

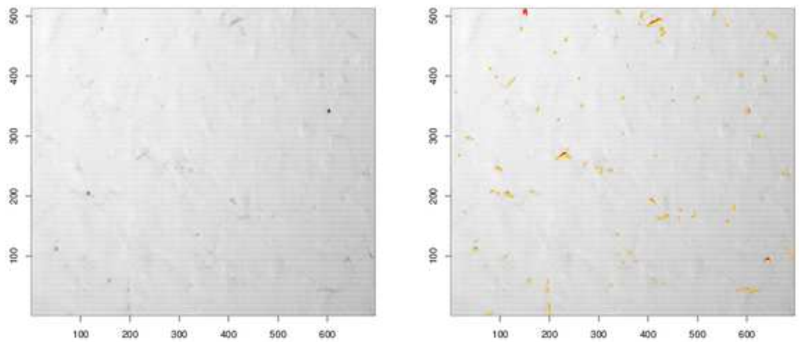

Figure 4.5: Image after 1.000 load cycles without and with detected micro cracks

4.2 Point processes for crack initiation

4.2.1 Definition (Stochastic process, see e.g. Iacus 2008, p. 14)
Let $(\Omega, \mathcal{A}, P)$ be a probability space. A stochastic process on $\mathcal{X}$ with state space $\mathcal{S}$ is a family of random variables $Y = \{Y(x);\ x \in \mathcal{X}\} = \{Y_x;\ x \in \mathcal{X}\}$ defined on Ω and with values in $\mathcal{S}$, i.e.

$$Y = \{Y(x);\ x \in \mathcal{X}\} = Y_{x;\, x\in\mathcal{X}} : \Omega \ni \omega \longrightarrow Y(\omega) = Y_{x;\, x\in\mathcal{X}}(\omega) = \{Y(x)(\omega);\ x \in \mathcal{X}\} \in S^{\mathcal{X}}.$$

a) If $\mathcal{X} = I\!N$, then Y a discrete-time process.
b) If $\mathcal{X} = [a, b] \subset \mathbb{R}$ with $a < b$, then Y is a continuous-time process.
c) If $\mathcal{X} \subset \mathbb{R}^d$ with $d > 1$, then Y is a spatial process, also called random field.
d) If $\mathcal{S} = \mathbb{R}$, then Y is real valued stochastic process.

Note that $\mathcal{S}^{\mathcal{X}}$ is the space of all functions $f : \mathcal{X} \to \mathcal{S}$.

If $\mathcal{X} = \{1, \ldots, n\}$ and $\mathcal{S} = \mathbb{R}$ then $\{Y(x);\ x \in \mathcal{X}\} = (Y(1), \ldots, Y(n)) = (Y_1, \ldots, Y_n)$ is a n-dimensional random vector.

4.2.2 Definition (Simple point process, see Jacobsen 2006, p. 10)
A simple point process is a sequence $\mathcal{T} = (T_n)_{n\geq 1}$ of $[0, \infty]$-valued random variables (events) defined on $(\Omega, \mathcal{A}, P)$ such that
a) $P(0 < T_1 \leq T_2 \leq \ldots) = 1$,
b) $P(T_n < T_{n+1}, T_n < \infty) = P(T_n < \infty)$ for all $n \geq 1$,
c) $P(\lim_{n\to\infty} T_n = \infty) = 1$.
If $\mathcal{T} = (T_n)_{n\geq 1}$ does not satisfy Condition c), then $\mathcal{T} = (T_n)_{n\geq 1}$ is called simple point process with explosion.

4.2.3 Definition (Counting process, Jacobsen 2006, p. 11/12)
Let $\mathcal{T} = (T_n)_{n\geq 1}$ be a simple point process. The associated counting process $N = (N(t))_{t\geq 0} = (N_t)_{t\geq 0}$ is a continuous time process with

$$N_t = \sum_{n=1}^{\infty} \mathbb{I}_{[0,t]}(T_n).$$

Hence N_t counts the number of events in the time interval $[0,t]$ and $N_0 \equiv 0$. A counting process is a stochastic process with state space $\mathcal{S} = \mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For P-almost all $\omega \in \Omega$, the samples path $N.(\omega) : \mathbb{R}^+ \ni t \to N_t(\omega) \in \mathbb{N}_0$ belongs to the space W of counting process paths,

$$W = \{w \in \mathbb{N}_0^{\mathbb{R}^+};\ w(0) = 0,\ w \text{ is right-continuous, increasing, } \Delta w(t) = 0 \text{ or } 1 \text{ for all } t \geq 0\}$$

with $\mathbb{R}^+ := [0,\infty)$ and $\Delta w(t) := w(t) - w(t-)$ where $w(t-) := \lim_{\tau\uparrow t} w(\tau)$. The functions in W are also called *cadlag* functions (cadlag = continue à droite, limite à gauche).

Having a random variable N with values in W, then the associated simple point process $\mathcal{T} = (T_n)_{n\geq 1}$ is given by

$$T_n = \inf\{t \geq 0;\ N_t = n\}.$$

In particular, we have

$$T_n \leq t \Longleftrightarrow N_t \geq n.$$

Since any $w \in W$ is a right-continuous, increasing step, it defines a discrete measure on $\mathbb{R}^+$ given by

$$\sum_{t;\Delta w(t)=1} \epsilon_t$$

where ϵ_t is the one-point (Dirac) measure on t, i.e. $\epsilon_t(A) = \mathbb{I}_A(t)$ for all $A \in \mathcal{B}$, if $\mathcal{B}$ denotes the Borel-σ-algebra on $\mathbb{R}$. Hence each path $N(\omega)$ defines a discrete measure on $\mathbb{R}^+$ given by

$$\mu_\omega := M(\omega) := \sum_{t;\Delta N_t(\omega)=1} \epsilon_t = \sum_{n\in\mathbb{N};\ T_n(\omega)<\infty} \epsilon_{T_n(\omega)}. \tag{4.1}$$

Denoting with $\mathcal{P}(\mathbb{R}^+)$ all discrete measures on $\mathbb{R}^+$, then

$$M : \Omega \ni \omega \to M(\omega) \in \mathcal{P}(\mathbb{R}^+)$$

is a random measure, the *counting measure*.

4.2.4 Definition (Point process on $\mathbb{R}^d$, see e.g. `https://en.wikipedia.org/wiki/Point_process`)
Let $(\Omega, \mathcal{A}, P)$ be a probability space. If $\mathcal{P}(\mathbb{R}^d)$ is the set of all discrete measures on $\mathbb{R}^d$ equipped with a appropriate σ-algebra, then the random variable (random measure)

$$M : \Omega \ni \omega \to M(\omega) \in \mathcal{P}(\mathbb{R}^d)$$

is called point process on $\mathbb{R}^d$. It is called simple point process if $M(\omega)(\{x\}) = \mu_\omega(\{x\}) \in \{0,1\}$ holds for all $x \in \mathbb{R}^d$ for P-almost all ω.

The random discrete measure M can be also represented via the points (events) as in the one-dimensional case given in (4.1), i.e.

$$M = \sum_{n=1}^{\infty} \epsilon_{X_n}, \text{ or } M(\omega) = \sum_{n=1}^{\infty} \epsilon_{X_n(\omega)}, \text{ respectively,}$$

where $X_n : \Omega \to \mathbb{R}^d$, $n = 1, 2, \ldots$, are random vectors providing the event point (see Baddeley et al. 2006, p. 3).

4.2.5 Definition (Cox point process, Poisson point process)
(see e.g. `https://en.wikipedia.org/wiki/Point_process`*)*
Let $\lambda : \mathbb{R}^d \to \mathbb{R}$ be a integrable function and set $\Lambda(B) = \int_B \lambda(x)dx$ for any $B \in \mathcal{B}^d$, the Borel-σ-algebra on $\mathbb{R}^d$.
a) A point process M is called Cox point process if

(i) for any $B \in \mathcal{B}^d$, $M(B)$ has Poisson distribution with parameter $\Lambda(B)$, i.e. the density of $M(B)$ is given by $p(k) = \frac{\Lambda(B)^k}{k!}\, e^{-\Lambda(B)}$,

(ii) $M(B_1), \ldots, M(B_I)$ are independent for any finite collection of disjoint subsets $B_1, \ldots, B_I \in \mathcal{B}^d$.

b) A Cox point process is called a Poisson point process if λ is the constant function so that $\Lambda(B) = \lambda \|B\|$, where $\|B\|$ is the Lebesgue measure of B.
c) The function λ is called the intensity function of the process. If it is constant, then it is called intensity parameter.

Testing a Poisson point process
If it should be tested whether given points $x_1, \ldots, x_N \in \mathbb{R}^d$ are realizations of a Poisson point process, then the following simple test can be applied: Divide the space in L d-dimensional disjoint intervals $I_1, \ldots, I_L$ of equal size and determine the number of points falling in I_l, i.e.

$$m_l := \sharp\{n;\ x_n \in I_l\},$$

for $l = 1, \ldots, L$. The numbers $m_1, \ldots, m_L$ are realization of random variables $M_1 = M(I_1), \ldots, M_L = M(I_L)$. The nullhypothsis is then that $M_1, \ldots, M_L$ are i.i.d. with Poisson distribution. The parameter λ of this Poisson distribution can be estimated by the maximum likelihood estimator. Having the estimator $\widehat{\lambda}$, the expected number of observations in each interval is $\widehat{\lambda}$. If $\widehat{\lambda} > 5$, then the χ^2 goodness-of-test given by Definition 2.6.1 can be used by replacing Np_l by $\widehat{\lambda}$ and N_l by m_l. Since only one parameter, namely λ must be estimated, the parameter dimension r is here 1.

Chapter 5

Crack growth and degradation

5.1 Fatigue crack growth equations

The most famous fatige crack growth equation is the Paris-Erdogan equation given by (see e.g. Sobczyk and Spencer 1992)

$$\frac{d\,a}{d\,N} = C\,(\Delta\,K)^m, \tag{5.1}$$

where a is the crack length, N is the number of load cycles, C and m are material constants, K is the stress intensity factor and $\Delta\,K = K_{\max} - K_{\min}$ is the stress intensity factor range. Thereby it is used

$$K = \sigma Y \sqrt{\pi a},$$

where Y is a geometry parameter and σ is a uniform tensile stress perpendicular to the crack plane. Using $\Delta\sigma = \sigma_{\max} - \sigma_{\min}$ as the range of the cyclic stress amplitude we obtain

$$\Delta\,K = \Delta\,\sigma Y \sqrt{\pi a}$$

so that (5.1) becomes

$$\frac{d\,a}{d\,N} = C\,(\Delta\,\sigma Y \sqrt{\pi})^m\; a^{m/2}. \tag{5.2}$$

Since the number of load cycles N is a measurement of the time, we will use here t instead of N. Instead of a, we will use l for the length. Moreover, we set $\theta_1 = C\,(\Delta\,\sigma Y \sqrt{\pi})^m$ and $\theta_2 = \frac{m}{2}$. Then (5.2) becomes

$$\frac{d\,l}{d\,t} = \theta_1\, l^{\theta_2}. \tag{5.3}$$

This is an ordinary differential equations.

5.1.1 Lemma
Equation (5.3) has the following solutions depending on θ_2:

$$\begin{array}{rl} (i) & \theta_2 = 1 \Rightarrow l = l(t) = \theta_0 \cdot exp(\theta_1 \cdot t) \\ & \text{with } \ \theta_0 > 0, \\ (ii) & \theta_2 < 1 \Rightarrow l = l(t) = \alpha_1 \cdot (t - \alpha_0)^{\alpha_2} \\ & \text{with } \ \alpha_0 < t, \ \ \alpha_1 = (\theta_1 \cdot (1-\theta_2))^{\frac{1}{1-\theta_2}}, \ \ \alpha_2 = \dfrac{1}{1-\theta_2} > 0, \\ (iii) & \theta_2 > 1 \Rightarrow l = l(t) = \alpha_1 \cdot (\alpha_0 - t)^{-\alpha_2} \\ & \text{with } \ \alpha_0 > t, \ \ \alpha_1 = (\theta_1 \cdot (\theta_2 - 1))^{\frac{-1}{\theta_2-1}}, \ \ \alpha_2 = \dfrac{1}{\theta_2 - 1} > 0. \end{array}$$

Proof

$$\begin{array}{rl} \text{(i)} & \dfrac{\partial l}{\partial t} = \dfrac{\partial}{\partial t}\left[\theta_0 \cdot \exp(\theta_1 \cdot t)\right] = \theta_0 \cdot \exp(\theta_1 \cdot t) \cdot \theta_1 = \theta_1 \cdot l, \\ \text{(ii)} & \dfrac{\partial l}{\partial t} = \dfrac{\partial}{\partial t}\left[\alpha_1 \cdot (t-\alpha_0)^{\alpha_2}\right] = \dfrac{\partial}{\partial t}\left[(\theta_1 \cdot (1-\theta_2))^{\frac{1}{1-\theta_2}} \cdot (t-\alpha_0)^{\frac{1}{1-\theta_2}}\right] \end{array}$$

$$\begin{aligned} &= (\theta_1 \cdot (1-\theta_2))^{\frac{1}{1-\theta_2}} \cdot (1-\theta_2)^{-1} \cdot (t-\alpha_0)^{\frac{1}{1-\theta_2}-1} \\ &= \theta_1^{\frac{1}{1-\theta_2}} \cdot (1-\theta_2)^{\frac{1}{1-\theta_2}} \cdot (1-\theta_2)^{-\frac{1-\theta_2}{1-\theta_2}} \cdot (t-\alpha_0)^{\frac{1}{1-\theta_2}-\frac{1-\theta_2}{1-\theta_2}} \\ &= \theta_1^{\frac{1-\theta_2}{1-\theta_2}} \cdot \theta_1^{\frac{\theta_2}{1-\theta_2}} \cdot (1-\theta_2)^{\frac{\theta_2}{1-\theta_2}} \cdot (t-\alpha_0)^{\frac{\theta_2}{1-\theta_2}} \\ &= \theta_1 \cdot \left((\theta_1 \cdot (1-\theta_2))^{\frac{1}{1-\theta_2}} \cdot (t-\alpha_0)^{\frac{1}{1-\theta_2}}\right)^{\theta_2} \\ &= \theta_1 \cdot (\alpha_1 \cdot (t-\alpha_0)^{\alpha_2})^{\theta_2} = \theta_1 \cdot l^{\theta_2}. \end{aligned}$$

Part (iii) is an exercise. □

However, in Example 1.0.3, the independent variable is the length l and the dependent variable is the time t. Hence we have to calculate the inverse $t(l)$ of the functions $t(l)$ of Lemma 5.1.1.

5.1.2 Lemma
The inverse function $t(l)$ of the functions $l(t)$ of Lemma 5.1.1 are:

$$\begin{array}{rl} (i) & \theta_2 = 1 \Rightarrow t = t(l) = \beta_0 + \beta_1 \cdot ln(l) \\ & \text{with } \ \beta_0 = -\dfrac{1}{\theta_1} \cdot ln(\theta_0) \ , \ \beta_1 = \dfrac{1}{\theta_1}, \\ (ii) & \theta_2 < 1 \Rightarrow t = t(l) = \beta_0 + \beta_1 \cdot l^{\beta_2} \\ & \text{with } \ \beta_0 = \alpha_0 < t \ , \ \beta_1 = \left(\dfrac{1}{\alpha_1}\right)^{\frac{1}{\alpha_2}} > 0 \ , \ \beta_2 = \dfrac{1}{\alpha_2} > 0, \\ (iii) & \theta_2 > 1 \Rightarrow t = t(l) = \beta_0 + \beta_1 \cdot l^{\beta_2} \\ & \text{with } \ \beta_0 = \alpha_0 > t \ , \ \beta_1 = -\left(\dfrac{1}{\alpha_1}\right)^{\frac{1}{-\alpha_2}} < 0 \ , \ \beta_2 = -\dfrac{1}{\alpha_2} < 0. \end{array}$$

Proof

(i)
$$l = l(t) = \theta_0 \cdot \exp(\theta_1 \cdot t) \Leftrightarrow \ln(l) = \ln(\theta_0) + \theta_1 \cdot t$$
$$\Leftrightarrow t = \frac{1}{\theta_1} \cdot \ln(l) - \frac{1}{\theta_1} \cdot \ln(\theta_0) \Leftrightarrow t = t(l) = \beta_0 + \beta_1 \cdot \ln(l),$$

(ii)
$$l = l(t) = \alpha_1 \cdot (t - \alpha_0)^{\alpha_2} \Leftrightarrow \left(\frac{l}{\alpha_1}\right)^{\frac{1}{\alpha_2}} = t - \alpha_0$$
$$\Leftrightarrow t = \alpha_0 + \left(\frac{l}{\alpha_1}\right)^{\frac{1}{\alpha_2}} \Leftrightarrow t = t(l) = \beta_0 + \beta_1 \cdot l^{\beta_2},$$

(iii)
$$l = l(t) = \alpha_1 \cdot (\alpha_0 - t)^{-\alpha_2} \Leftrightarrow \left(\frac{l}{\alpha_1}\right)^{\frac{1}{-\alpha_2}} = \alpha_0 - t$$
$$\Leftrightarrow t = \alpha_0 - \left(\frac{l}{\alpha_1}\right)^{\frac{-1}{\alpha_2}} \Leftrightarrow t = t(l) = \beta_0 + \beta_1 \cdot l^{\beta_2}. \qquad \square$$

5.2 Crack growth prediction via (non)linear models

All models obtained by Lemma 5.1.1 and Lemma 5.1.2 are linear or nonlinear in the unknown parameters. Since crack growth is not deterministic, a simple stochastic version of these function can be obtained by adding a random error. This will lead to the following models.

Linear model:

$$Y_n = \theta_0 + \theta_1 \log(x_n) + E_n$$

Nonlinear models:

$$Y_n = \theta_0 \exp(\theta_1 x_n) + E_n, \tag{5.4}$$
$$Y_n = \theta_1 \cdot (x_n - \theta_0)^{\theta_2} + E_n, \tag{5.5}$$
$$Y_n = \theta_1 \cdot (\theta_0 - x_n)^{-\theta_2} + E_n, \tag{5.6}$$
$$Y_n = \theta_0 + \theta_1 \, x_n^{\theta_2} + E_n, \tag{5.7}$$

where we have in model (5.7)

unbounded growth if $\theta_1 > 0,\ \theta_2 > 0$
bounded growth if $\theta_1 < 0,\ \theta_2 < 0$.

A linearized model of model (5.7) is

$$Y_n = \theta_0 + \theta_1 \, x_n + \theta_2 \, x_n \, \log(x_n) + E_n.$$

If the errors are normally distributed, exact predictions intervals for the linear models, are given by Theorem 3.4.3. Theorem 3.4.9 provide approximate prediction intervals for the nonlinear models. Thereby the errors must be not normally distributed, since the Theorem 3.4.6 holds for more general distributions.

5.2.1 Example (Virkler data)
For the Virkler data reasonable models are the nonlinear model

$$T_n = \theta_0 + \theta_1 \, l_n^{\theta_2} + E_n$$

and its linearization

$$T_n = \theta_0 + \theta_1 \, l_n + \theta_2 \, l_n \, \log(x_n) + E_n,$$

where T_n is the time variable and l_n are the given length values. Figure 5.1 shows the results for prediction for one series.

Figure 5.2 shows the main drawback of using independent additive errors. These additive errors can only model measurements errors so that shortly after the last available observation the prediction intervals are already very large. However one would not expect a big jump after the last observation since the crack length depends very much on the crack length of the last time point. Hence there is a stochastic dependence between observations. Therefore it is better to use an approach based on a stochastic processes like the Wiener process, the Gamma process or processes given by stochastic differential equations.

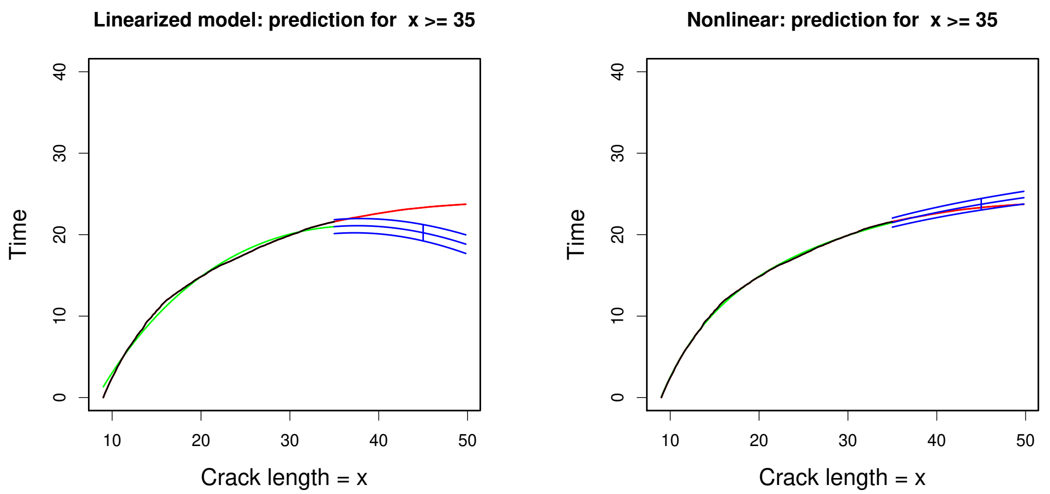

Figure 5.1: Prediction for the Virkler data with a linearized and a nonlinear model

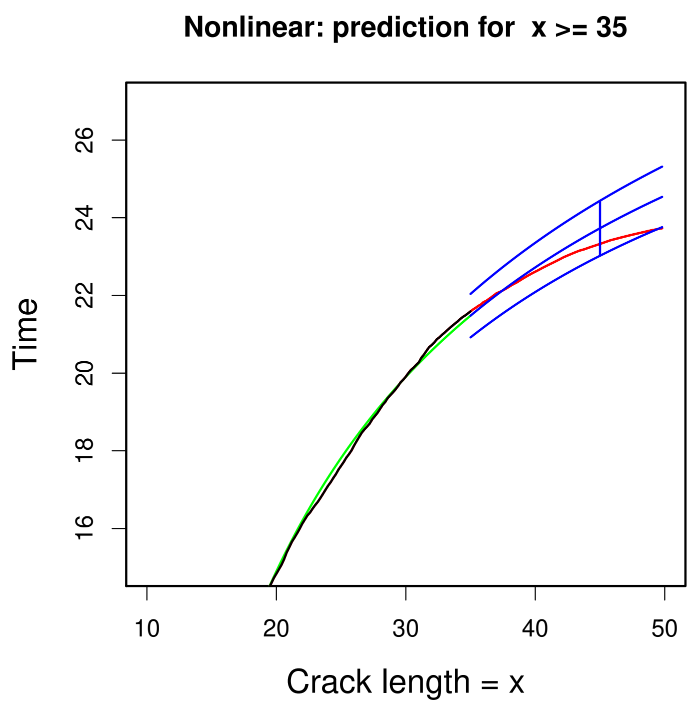

Figure 5.2: Prediction for the Virkler data with a nonlinear model

5.3 Birnbaum-Saunders model

Birnbaum and Saunders (1969) proposed a very simple model for crack growth where the crack growth process is a deterministic functions disturbed only by one random variable.

5.3.1 Definition
The Birnbaum-Saunders degradation process $Y = \{Y(x);\, x \geq 0\}$ is given by

$$Y(x) = \mu\, x + \sigma\, Z\, \sqrt{x},$$

where $\mu > 0$, $\sigma > 0$, and Z is a random variable with standard normal distribution.

5.3.2 Definition (First passage time / time to failure, see e.g. Kahle et al. 2016, p. 12)
If $c > 0$ is a given critical value and $Y = \{Y(x);\, x \geq 0\}$ is a time-contimuous process then

$$X_c := \inf\{x \geq 0;\; Y(x) \geq c\}$$

is called the first passage time of the value c (time to failure at c) of the process Y.

Since Y is random also X_c is a random variable. Therefore, it has cumulative distribution function F_{X_c} and a density f_{X_c} given by $f_{X_c}(x) = F'_{X_c}(x)$. Moreover $Y(x) > y$ with $y \geq c$ implies $X_c < x$ so that we get with the formel of total probability for any $y \geq c$

$$P(Y(x) > y) = P(Y(x) > y,\; X_c < x) = \int_0^x P(Y(x) > y \mid X_c = z)\, f_{X_c}(z)\, dz. \tag{5.8}$$

5.3.3 Theorem
The cumulative distribution function F_{X_c} of the first passage time of c of a Birnbaum-Saunders degradation process $Y = \{Y(x);\, x \geq 0\}$ is given by

$$F_{X_c}(x) = 1 - \Phi\left(\frac{c - \mu\, x}{\sigma\, \sqrt{x}}\right),$$

where Φ is the cumulative distribution function of the standard normal distribution.

Proof. Since

$$\frac{\partial}{\partial x} Y(x) = \mu + \frac{\sigma\, Z}{2\, \sqrt{x}}$$

becomes positive for large x for any Z, the process is possibly at first decreasing but then always strictly increasing if x is large enough. Hence it holds $X_c \leq x \Leftrightarrow Y(x) \geq c$ so that

$$F_{X_c}(x) = P(X_c \leq x) = P(Y(x) \geq c) = P(\mu\, x + \sigma\, Z\, \sqrt{x} \geq c) = P\left(Z \geq \frac{c - \mu\, x}{\sigma\, \sqrt{x}}\right). \;\square$$

Since

$$\frac{\partial}{\partial x} \frac{c - \mu\, x}{\sigma\, \sqrt{x}} = \frac{\partial}{\partial x}\left(\frac{c}{\sigma\, \sqrt{x}} - \frac{\mu\, \sqrt{x}}{\sigma}\right) = -\frac{c}{2\, \sigma\, \sqrt{x}^3} - \frac{\mu}{2\, \sigma\, \sqrt{x}}$$

is always negative for $x > 0$ and

$$\lim_{x \downarrow 0} \frac{c - \mu\, x}{\sigma \sqrt{x}} = \lim_{x \downarrow 0} \left(\frac{c}{\sigma \sqrt{x}} - \frac{\mu \sqrt{x}}{\sigma} \right) = \infty, \quad \lim_{x \to \infty} \frac{c - \mu\, x}{\sigma \sqrt{x}} = \lim_{x \to \infty} \left(\frac{c}{\sigma \sqrt{x}} - \frac{\mu \sqrt{x}}{\sigma} \right) = -\infty,$$

the function

$$1 - \Phi \left(\frac{c - \mu\, x}{\sigma \sqrt{x}} \right)$$

is a strictly increasing function from 0 at $x = 0$ to 1 for $x \to \infty$. Hence it is a cumulative distribution function of a lifetime distribution.

5.3.4 Definition (Birnbaum-Saunders distribution)
T has a Birnbaum-Saunders distribution if its cumulative distribution function F_T is given by

$$F_T(t) = 1 - \Phi \left(\frac{c - \mu\, t}{\sigma \sqrt{t}} \right)$$

for $t \geq 0$.

The Birnbaum-Saunders distribution is a lifetime distribution with three parameters μ, σ, and c. If T has a Birnbaum-Saunders distribution then

$$E(T) = \frac{c}{\mu} + \frac{\sigma^2}{2\mu^2}, \quad \mathrm{var}(T) = E(T^2) - (E(T))^2 = c\frac{\sigma^2}{\mu^3} + \frac{5\,\sigma^4}{4\,\mu^4},$$

see Kahle and Liebscher (2013).

5.4 Wiener processes for modeling degradation processes

Crack grwoth is a special form of degradation. A simple model for degradation is given by the Wiener process. It bases on the Brownian motion which is a time-contimuous process with normal distribution.

5.4.1 Definition (Brownian motion / Wiener process without drift, see e.g. Iacus 2008, p. 18) *A real valued stochastic process $B = \{B(x);\ x \geq 0\}$ is called a Brownian motion (Wiener process without drift) if*

(i) $B(0) = 0$,

(ii) $B(x_2) - B(x_1) \sim \mathcal{N}(0, x_2 - x_1)$ *for all* $0 \leq x_1 \leq x_2$,

(iii) the increments $B(x_{i+1}) - B(x_i)$, $i = 1, \dots, n$, *are independent for all* $n \in I\!N$, $n > 1$, *and* $0 \leq x_1 < x_2 < \dots < x_{n+1}$.

The Brownian motion is a Gaussian process. In particular,

$$(B(x_1), \dots, B(x_n)) \sim \mathcal{N}_n(0_n, \Sigma),$$

where 0_n is the n-dimensional vector consisting of zeros and the matrix Σ is given by $\Sigma = (\min(x_i, x_j))_{i,j=1,\dots,n}$. This implies $B(x) \sim \mathcal{N}(0, x)$ and $B(x_2 - x_1) \sim \mathcal{N}(0, x_2 - x_1)$ so that $B(x_2 - x_1) \sim B(x_2) - B(x_1)$.

A realization $b = \{b(x);\ x \geq 0\}$ of $B = \{B(x);\ x \geq 0\}$ is a continuous function $b : [0, \infty) \to \mathbb{R}$ and is called path.

A Brownian motion on $[0, \xi]$ can be simulated with the following algorithm (see e.g. Iacus 2008, p. 19): Devide the interval $[0, \xi]$ into a grid such that $0 = x_1 < x_2 < \dots < x_{N-1} < x_N = \xi$ with $\Delta x := x_{i+1} - x_i$ for $i = 2, \dots, N$ and set $B(0) = B(x_1) = 0$ and $i = 1$. Iterate then the following steps:

1. $i = i + 1$.
2. Generate a random number z from the standard normal distribution, i.e. from $\mathcal{N}(0, 1)$.
3. Set $B(x_i) = B(x_{i-1}) + z\sqrt{\Delta x}$.
4. If $i < N$, go to step 1.

The simulated process will approach a realization of the Brownian motion the better the smaller Δx is.

5.4.2 Definition (Wiener Process / Brownian motion with drift, see e.g. Kahle et al. 2016, p. 11)
A real valued stochastic process $W = \{W(x);\ x \geq 0\}$ *is called a Wiener process (Brownian motion with drift) if*

$$W(x) = \mu x + \sigma B(x),$$

where $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$, $B = \{B(x);\ x \geq 0\}$ *is a Brownian motion. The parameter* μ *is called drift parameter and the parameter* σ^2 *is called volatility parameter. The Wiener process is called Wiener process with positive (negative) drift if* $\mu > 0$ $(\mu < 0)$.

5.4.3 Lemma
The Wiener process satisfies for any $x_4 > x_4 > x_2 > x_1 \geq 0$

$$\begin{array}{rl} (i) & W(0) = 0, \\ (ii) & W(x_2) - W(x_1) \sim \mathcal{N}(\mu(x_2 - x_1), \sigma^2(x_2 - x_1)) \text{ for all } x_2 > x_1 \geq 0, \\ (iii) & W(x_2) - W(x_1) \sim W(x_2 - x_1) \text{ for all } x_2 > x_1 \geq 0, \\ (iv) & \text{the increments } W(x_{i+1}) - W(x_i),\ i = 1, \ldots, n, \text{ are independent for all } n \in I\!N, n > 1, \\ & 0 \leq x_1 < x_2 < \ldots < x_{n+1}. \end{array}$$

Proof. (i) is clear. For (ii) and (iii) note that

$$\begin{aligned} & W(x_2) - W(x_1) \\ = \; & \mu x_2 - \mu x_1 + \sigma B(x_2) - \sigma B(x_1) \sim \mu(x_2 - x_1) + \sigma B(x_2 - x_1) = W(x_2 - x_1). \end{aligned}$$

(iv) follows from the independence of $B(x_{i+1}) - B(x_i)$, $i = 1, \ldots, n$. $\square$

Here we will assume without loss of generality that the drift parameter satisfies $\mu \geq 0$.

Now we consider the first passage time x_c of a value c of a Wiener process W. The continuity of any path $\{w(x);\ x \geq 0\}$ implies $w(x_c) = c$ so that $X_c = z$ if and only if $W(z) = c$.
More precisely, $X_c = z \Rightarrow W(z) = c$, $W(z) = c \Rightarrow X_c \leq z$.
Hence we obtain with (5.8) for any $y \geq c$

$$\begin{aligned} P(W(x) > y) &= \int_0^x P(W(x) > y \mid W(z) = c)\ f_{X_c}(z)\ dz \\ &\overset{Lemma\ 5.4.3\ (i)}{=} \int_0^x P(W(x) - W(z) > y - c \mid W(z) - W(0) = c)\ f_{X_c}(z)\ dz \\ &\overset{Lemma\ 5.4.3\ (iv)}{=} \int_0^x P(W(x) - W(z) > y - c)\ f_{X_c}(z)\ dz \\ &\overset{Lemma\ 5.4.3\ (iii)}{=} \int_0^x P(W(x - z) > y - c)\ f_{X_c}(z)\ dz \\ &= \int_0^x (1 - F_{W(x-z)}(y - c))\ f_{X_c}(z)\ dz, \end{aligned} \tag{5.9}$$

where $F_{W(x)}$ is the cumulative distribution function of $W(x)$. Differentiation with respect to y leads to the integral equation

$$f_{W(x)}(y) = \int_0^x f_{W(x-z)}(y-c)\, f_{X_c}(z)\, dz. \tag{5.10}$$

5.4.4 Lemma (See e.g. Kahle et al. 2016, p. 14)
If $W = \{W(x);\ x \geq 0\}$ is a Wiener process without drift, i.e. with $\mu = 0$, then the cumulative distribution function F_{X_c} and the density f_{X_c} of the the first passage time X_c of the value c are given for any $x \geq 0$ by

$$F_{X_c}(x) = 2\left(1 - \Phi\left(\frac{c}{\sigma\sqrt{x}}\right)\right)$$

and

$$f_{X_c}(x) = \frac{c}{\sqrt{2\pi\sigma^2 x^3}} \exp\left(-\frac{c^2}{2\sigma^2 x}\right),$$

respectively, where Φ denotes the cumulative distribution function of the standard normal distribution.

Proof. At first note that

$$F'_{X_c}(x) = -2\Phi'\left(\frac{c}{\sigma\sqrt{x}}\right)\frac{-1}{2}\frac{c}{\sigma\sqrt{x^3}} = \frac{c}{\sqrt{2\pi\sigma^2 x^3}} \exp\left(-\frac{c^2}{2\sigma^2 x}\right).$$

To prove the form of $F_{X_c}(x)$, we use the fact that $W(x) = \sigma B(x)$ holds for all $x \geq 0$. This implies with $\frac{B(x)}{\sqrt{x}} \sim \mathcal{N}(0,1)$

$$P(W(x) > y) = P(\sigma B(x) > y) = P\left(\frac{B(x)}{\sqrt{x}} > \frac{y}{\sigma\sqrt{x}}\right) = 1 - \Phi\left(\frac{y}{\sigma\sqrt{x}}\right). \tag{5.11}$$

In particular, we get for $y = c$ and $z < x$

$$P(W(x-z) > y-c) = 1 - \Phi\left(\frac{y-c}{\sigma\sqrt{x-z}}\right) = 1 - \Phi(0) = \frac{1}{2}.$$

Using this in equation (5.9) leads to

$$P(W(x) > c) = \int_0^x P(W(x-z) > y-c)\, f_{X_c}(z)\, dz = \int_0^x \frac{1}{2}\, f_{X_c}(z)\, dz = \frac{1}{2}\, F_{X_c}(x)$$

which provides with (5.11) the assertion. $\square$

Obviously, the function F_{X_c} is strictly increasing with

$$\lim_{x\to\infty} F_{X_c}(x) = \lim_{x\to\infty} 2\left(1-\Phi\left(\frac{c}{\sigma\sqrt{x}}\right)\right) = 2(1-\Phi(0)) = 1$$

and

$$\lim_{x\downarrow 0} F_{X_c}(x) = \lim_{x\downarrow 0} 2\left(1-\Phi\left(\frac{c}{\sigma\sqrt{x}}\right)\right) = 2\left(1-\lim_{z\to\infty}\Phi(z)\right) = 0.$$

Hence F_{X_c} is indeed a cumulative distribution function, the cumulative distribution function of a special *Lévy distribution* and a special *Inverse Gamma distribution*. It is also the limiting case of the *inverse Gaussian distribution* with mean parameter $\mu \to \infty$.

5.4.5 Definition (Lévy distribution)
T has a Lévy distribution, shortly $T \sim \mathcal{L}(\mu,\lambda)$, with mean parameter $\mu \geq 0$ and shape parameter $\lambda > 0$ if the density f_T is given by

$$f_T(t) = \sqrt{\frac{\lambda}{2\pi(t-\mu)^3}}\exp\left(-\frac{\lambda}{2(t-\mu)}\right)$$

for $t > \mu$.

5.4.6 Definition (Inverse Gaussian distribution / Wald distribution, see e.g. Kahle et al. 2016, p. 16)
T has a inverse Gaussian distribution or Wald distribution, shortly $T \sim \mathcal{IN}(\mu,\lambda)$, with mean parameter $\mu > 0$ and shape parameter $\lambda > 0$ if the density f_T is given by

$$f_T(t) = \sqrt{\frac{\lambda}{2\pi t^3}}\exp\left(\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right) = \sqrt{\frac{\lambda}{2\pi t^3}}\exp\left(\frac{-\lambda\left(\frac{t}{\mu}-1\right)^2}{2t}\right)$$

for $t > 0$.

The expectation and variance of a random variable T with a inverse Gaussian distribution can be given explicitly and are simple as those of the Birnbaum-Saunders distribution, see Kahle and Liebscher (2013).

5.4.7 Theorem (See Kahle et al. 2016, p. 16)
If $W = \{W(x);\ x \geq 0\}$ is a Wiener process with drift μ and volatility σ^2, then the density f_{X_c} of the first passage time X_c of the value c is given for any $x \geq 0$ by

$$f_{X_c}(x) = \frac{c}{\sqrt{2\pi\sigma^2 x^3}} \exp\left(-\frac{(c-\mu x)^2}{2\sigma^2 x}\right) = \frac{\frac{c}{\sigma}}{\sqrt{2\pi x^3}} \exp\left(-\frac{\frac{c^2}{\sigma^2}\left(1-\frac{\mu}{c}x\right)^2}{2x}\right),$$

i.e. $X_c \sim \mathcal{IN}\left(\frac{c}{\mu}, \frac{c^2}{\sigma^2}\right)$.

Proof. We have to show that f_{X_c} satisfies the equation (5.10) given by

$$f_{W(x)}(y) = \int_0^x f_{W(x-z)}(y-c)\, f_{X_c}(z)\, dz.$$

Since $W(x) \sim \mathcal{N}(\mu x, \sigma^2 x)$ according to Lemma 5.4.3, equation (5.10) has the form

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi\sigma^2 x}} \exp\left(-\frac{(y-\mu x)^2}{2\sigma^2 x}\right) \\
&= \int_0^x \frac{1}{\sqrt{2\pi\sigma^2(x-z)}} \exp\left(-\frac{(y-c-\mu(x-z))^2}{2\sigma^2(x-z)}\right) \cdot \frac{c}{\sqrt{2\pi\sigma^2 z^3}} \exp\left(-\frac{(c-\mu z)^2}{2\sigma^2 z}\right) dz.
\end{aligned}$$

Dividing this equation by the left-hand side leads to

$$1 = \int_0^x \frac{1}{\sqrt{2\pi\sigma^2}} \frac{c\sqrt{x}}{\sqrt{(x-z)z^3}} \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot A(z)\right) dz, \tag{5.12}$$

where

$$\begin{aligned}
A(z) &:= -\frac{(y-\mu x)^2}{x} + \frac{(y-c-\mu(x-z))^2}{x-z} + \frac{(c-\mu z)^2}{z} \\
&= \frac{-z(x-z)(y-\mu x)^2 + xz(y-\mu x-(c-\mu z))^2 + x(x-z)(c-\mu z)^2}{xz(x-z)} \\
&= \frac{-z(x-z)(y-\mu x)^2 + xz((y-\mu x)^2 - 2(y-\mu x)(c-\mu z) + (c-\mu z)^2) + x(x-z)(c-\mu z)^2}{xz(x-z)} \\
&= \frac{z^2(y-\mu x)^2 - 2xz(y-\mu x)(c-\mu z) + x^2(c-\mu z)^2}{xz(x-z)} \\
&= \frac{(zy-\mu xz)^2 - 2(zy-\mu xz)(xc-\mu xz) + (xc-\mu xz)^2}{xz(x-z)} \\
&= \frac{(zy-\mu xz - xc + \mu xz)^2}{xz(x-z)} = \frac{(zy-xc)^2}{xz(x-z)} = \frac{(zy-zc+zc-xc)^2}{xz(x-z)} \\
&= \frac{(z(y-c)-(x-z)c)^2}{xz(x-z)} = \frac{(\frac{z}{x-z}(y-c)-c)^2}{x\frac{z}{x-z}}.
\end{aligned}$$

Now we use the substitution $u = \frac{z}{x-z}$ leading to $z = \frac{ux}{1+u} =: \varphi(u)$ with $\varphi(0) = 0$, $\lim_{u\to\infty} \varphi(u) = x$,

$$\varphi'(u) = \frac{(1+u)x - ux}{(1+u)^2} = \frac{x}{(1+u)^2},$$

and

$$x - z = x - \frac{ux}{1+u} = \frac{x + xu - ux}{1+u} = \frac{x}{1+u}.$$

The substitution rule

$$\int_{\varphi(a)}^{\varphi(b)} g(z)\, dz = \int_a^b g(\varphi(u))\, \varphi'(u)\, du$$

provides then for the right-hand side of equation (5.12)

$$\begin{aligned}
&\int_0^x \frac{1}{\sqrt{2\pi\sigma^2}} \frac{c\sqrt{x}}{\sqrt{(x-z)z^3}} \cdot \exp\left(-\frac{1}{2\sigma^2}\cdot A(z)\right)\, dz \\
&= \int_0^x \frac{1}{\sqrt{2\pi\sigma^2}} \frac{c\sqrt{x}}{\sqrt{(x-z)^4\left(\frac{z}{x-z}\right)^3}} \cdot \exp\left(-\frac{1}{2\sigma^2}\cdot A(z)\right)\, dz \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \frac{c\sqrt{x}}{\sqrt{\left(\frac{x}{1+u}\right)^4 u^3}} \cdot \exp\left(-\frac{1}{2\sigma^2}\cdot \frac{(u(y-c)-c)^2}{x\,u}\right) \frac{x}{(1+u)^2}\, du \\
&= \int_0^\infty \frac{c}{\sqrt{2\pi\sigma^2 x\, u^3}} \cdot \exp\left(-\frac{\left(\frac{y-c}{c}\right)^2 c^2 \left(u - \frac{c}{y-c}\right)^2}{2\,\sigma^2 x\, u}\right)\, du.
\end{aligned}$$

This integral equals to one since it is the integral of the density of the inverse Gaussian distribution with mean $\frac{c}{y-c}$ and shape $\frac{c^2}{\sigma^2 x}$. Hence the equation (5.10) holds. For the proof that f_{X_c} is the unique solution of (5.10) see Kahle et al. (2016), p. 17/18. □

Up to here, we have assumed that the stochastic process Y starts at $x_0 = 0$ with $Y(0) = 0$. However, the first passage time of the value c can be also of interest when the Wiener process starts at $x_0 > 0$ with $W(x_0) = w_0$. Then we define the first passage time X_c as

$$X_c := \inf\{x \geq x_0;\ Y(x) \geq c\}$$

and have a similar result about its distribution as for the Wiener process.

5.4.8 Definition (Shifted inverse Gaussian distribution)
T has a shifted inverse Gaussian distribution, shortly $T \sim \mathcal{SIN}(\mu, \lambda, t_0)$, with mean parameter $\mu > 0$, shape parameter $\lambda > 0$, and shift $t_0 > 0$ if the density f_T is given by

$$f_T(t) = \sqrt{\frac{\lambda}{2\pi(t-t_0)^3}} \exp\left(\frac{-\lambda(t-t_0-\mu)^2}{2\mu^2(t-t_0)}\right) = \sqrt{\frac{\lambda}{2\pi(t-t_0)^3}} \exp\left(\frac{-\lambda\left(\frac{(t-t_0)}{\mu}-1\right)^2}{2(t-t_0)}\right)$$

for $t > t_0$.

5.4.9 Theorem (See Kahle et al. 2016, p. 16)
If $W = \{W(x);\ x \geq x_0\}$ is a Wiener process with drift μ and volatility σ^2 starting at $W(x_0) = w_0$, then the density f_{X_c} of the first passage time X_c of the value c is given for any $x \geq x_0$ by

$$\begin{aligned} f_{X_c}(x) &= \frac{c-w_0}{\sqrt{2\pi\sigma^2(x-x_0)^3}} \exp\left(-\frac{(c-w_0-\mu(x-x_0))^2}{2\sigma^2(x-x_0)}\right) \\ &= \frac{\frac{c-w_0}{\sigma}}{\sqrt{2\pi(x-x_0)^3}} \exp\left(-\frac{\frac{(c-w_0)^2}{\sigma^2}\left(1-\frac{\mu}{c-w_0}(x-x_0)\right)^2}{2(x-x_0)}\right), \end{aligned}$$

i.e. $X_c \sim \mathcal{SIN}\left(\frac{c-w_0}{\mu}, \frac{(c-w_0)^2}{\sigma^2}, x_0\right)$.

Proof. See Kahle et al. 2016, p. 16/17. □

5.5 Statistical inference for Wiener processes

Although the path (realization) of a Wiener process $w = \{w(x);\ x \geq x_0\}$ is a continuous function $w : [0, \infty) \to \mathbb{R}$ it can be observed only at given points $x_1 < \ldots < x_I$ so that we have only observations $w(x_1), \ldots, w(x_I)$. To be more general, we assume that eventually several processes, J processes, are observed, eventually at different time points $x_{1,j} < \ldots < x_{I_j,j}$ for $j = 1, \ldots, J$. Then the available observations are $w^{(j)}(x_{1,j}), \ldots, w^{(j)}(x_{I_j,j})$ with $j = 1, \ldots, J$. $w^{(j)}(x_{1,j}), \ldots, w^{(j)}(x_{I_j,j})$ are realizations of $W^{(j)}(x_{1,j}), \ldots, W^{(j)}(x_{I_j,j})$ for $j = 1, \ldots, J$. We assume here again $x_{0,j} = 0$ and $W^{(j)}(x_{0,j}) = 0$ for simplicity. Set

$$s_{i,j} = x_{i,j} - x_{i-1,j}$$

for $i = 1, \ldots, I_j$, $j = 1, \ldots, J$. Then the *increments*

$$y_{i,j} = w^{(j)}(x_{i,j}) - w^{(j)}(x_{i-1,j})$$

are realizations of independent variables

$$Y_{i,j} = W^{(j)}(x_{i,j}) - W^{(j)}(x_{i-1,j}) \sim \mathcal{N}(\mu(x_{i,j} - x_{i-1,j}), \sigma^2(x_{i,j} - x_{i-1,j})) = \mathcal{N}(\mu\, s_{i,j}, \sigma^2\, s_{i,j})$$

for $i = 1, \ldots, I_j$, $j = 1, \ldots, J$ according to Lemma 5.4.3. Set

$$W_* := (W^{(1)}(x_{1,1}), \ldots, W^{(1)}(x_{I_1,1}), \ldots, W^{(J)}(x_{1,J}), \ldots, W^{(J)}(x_{I_J,J}))^\top,$$

with realization

$$w_* := (w^{(1)}(x_{1,1}), \ldots, w^{(1)}(x_{I_1,1}), \ldots, w^{(J)}(x_{1,J}), \ldots, w^{(J)}(x_{I_J,J}))^\top.$$

Then according to Lemma 5.4.3, the likelihood function for the data set w_* is given by

$$l(\mu, \sigma^2 | w_*) = \prod_{j=1}^{J} \prod_{i=1}^{I_j} \frac{1}{\sqrt{2\pi\sigma^2 s_{i,j}}} \exp\left(-\frac{(y_{i,j} - \mu s_{i,j})^2}{2\sigma^2 s_{i,j}} \right).$$

5.5.1 Theorem (See Kahle et al. 2016, p. 27)
Set $N = \sum_{j=1}^{J} I_j$. *Then the maximum likelihood estimator* $(\widehat{\mu}, \widehat{\sigma}^2)$ *for* (μ, σ^2) *based on* w_* *is given by*

$$\begin{aligned}
\widehat{\mu} &= \frac{\sum_{j=1}^{J}\sum_{i=1}^{I_j} y_{i,j}}{\sum_{j=1}^{J}\sum_{i=1}^{I_j} s_{i,j}} = \frac{\sum_{j=1}^{J} w_{I_j,j}^{(j)}}{\sum_{j=1}^{J} x_{I_j,j}}, \\
\widehat{\sigma}^2 &= \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{I_j} \frac{(y_{i,j} - \widehat{\mu} s_{i,j})^2}{s_{i,j}}.
\end{aligned}$$

Proof. The loglikelihood function is

$$L(\mu, \sigma^2 | w_*) = \sum_{j=1}^{J}\sum_{i=1}^{I_j} \left(-\frac{1}{2}\ln(2\pi\sigma^2 s_{i,j}) - \frac{(y_{i,j} - \mu s_{i,j})^2}{2\sigma^2 s_{i,j}} \right).$$

Hence we get

$$\begin{aligned}
&\frac{\partial}{\partial \mu} L(\mu, \sigma^2 | w_*) = \sum_{j=1}^{J}\sum_{i=1}^{I_j} \frac{2(y_{i,j} - \mu s_{i,j}) s_{i,j}}{2\sigma^2 s_{i,j}} = \sum_{j=1}^{J}\sum_{i=1}^{I_j} \frac{y_{i,j} - \mu s_{i,j}}{\sigma^2} = 0 \\
&\iff \sum_{j=1}^{J}\sum_{i=1}^{I_j} y_{i,j} = \mu \sum_{j=1}^{J}\sum_{i=1}^{I_j} s_{i,j} \iff \mu = \frac{\sum_{j=1}^{J}\sum_{i=1}^{I_j} y_{i,j}}{\sum_{j=1}^{J}\sum_{i=1}^{I_j} s_{i,j}}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial}{\partial \sigma^2} L(\mu, \sigma^2 | w_*) = \sum_{j=1}^{J}\sum_{i=1}^{I_j} \left(-\frac{1}{2}\frac{1}{\sigma^2} + \frac{(y_{i,j} - \mu s_{i,j})^2}{2\sigma^4 s_{i,j}} \right) \\
&\iff N\sigma^2 = \sum_{j=1}^{J}\sum_{i=1}^{I_j} \frac{(y_{i,j} - \mu s_{i,j})^2}{s_{i,j}} \iff \sigma^2 = \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{I_j} \frac{(y_{i,j} - \widehat{\mu} s_{i,j})^2}{s_{i,j}}.
\end{aligned}$$

That the maximum is attained at $(\widehat{\mu}, \widehat{\sigma}^2)$ follows as usually for the normal distribution. □

5.5.2 Theorem
The information matrix for a single increment $Y_{i,j}$ is

$$I_{(\mu,\sigma^2)}(Y_{i,j}) = \begin{pmatrix} \frac{s_{i,j}}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

and the information matrix for the whole data set W_*, setting $N = \sum_{j=1}^J I_j$, is

$$I_{(\mu,\sigma^2)}(W_*) = \begin{pmatrix} \frac{\sum_{j=1}^J \sum_{i=1}^{I_j} s_{i,j}}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{pmatrix} = \begin{pmatrix} \frac{\sum_{j=1}^J x_{I_j,j}}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{pmatrix}.$$

Proof. The proof of Theorem 5.5.1 provides

$$\frac{\partial}{\partial \mu} \ln f_{\mu,\sigma^2}(y_{i,j}) = \frac{y_{i,j} - \mu s_{i,j}}{\sigma^2}$$

and

$$\frac{\partial}{\partial \sigma^2} \ln f_{\mu,\sigma^2}(y_{i,j}) = -\frac{1}{2}\frac{1}{\sigma^2} + \frac{(y_{i,j} - \mu s_{i,j})^2}{2\sigma^4 s_{i,j}}$$

so that

$$E_{(\mu,\sigma^2)}\left(\left[\frac{\partial}{\partial \mu} \ln f_{\mu,\sigma^2}(Y_{i,j})\right]^2\right) = E_{(\mu,\sigma^2)}\left(\left[\frac{Y_{i,j} - \mu s_{i,j}}{\sigma^2}\right]^2\right) = \frac{\sigma^2 s_{i,j}}{\sigma^4} = \frac{s_{i,j}}{\sigma^2},$$

$$\begin{aligned} & E_{(\mu,\sigma^2)}\left(\left[\frac{\partial}{\partial \mu} \ln f_{\mu,\sigma^2}(Y_{i,j})\right]\left[\frac{\partial}{\partial \sigma^2} \ln f_{\mu,\sigma^2}(Y_{i,j})\right]\right) \\ = \; & E_{(\mu,\sigma^2)}\left(\left[\frac{Y_{i,j} - \mu s_{i,j}}{\sigma^2}\right]\left[-\frac{1}{2}\frac{1}{\sigma^2} + \frac{(Y_{i,j} - \mu s_{i,j})^2}{2\sigma^4 s_{i,j}}\right]\right) \\ = \; & E_{(\mu,\sigma^2)}\left(-\frac{(Y_{i,j} - \mu s_{i,j})}{2\sigma^4} + \frac{(Y_{i,j} - \mu s_{i,j})^3}{2\sigma^6 s_{i,j}}\right) = 0, \end{aligned}$$

$$\begin{aligned} & E_{(\mu,\sigma^2)}\left(\left[\frac{\partial}{\partial \sigma^2} \ln f_{\mu,\sigma^2}(Y_{i,j})\right]^2\right) = E_{(\mu,\sigma^2)}\left(\left[-\frac{1}{2}\frac{1}{\sigma^2} + \frac{(Y_{i,j} - \mu s_{i,j})^2}{2\sigma^4 s_{i,j}}\right]^2\right) \\ = \; & E_{(\mu,\sigma^2)}\left(\frac{1}{4\sigma^4} - \frac{(Y_{i,j} - \mu s_{i,j})^2}{2\sigma^6 s_{i,j}} + \frac{(Y_{i,j} - \mu s_{i,j})^4}{4\sigma^8 s_{i,j}^2}\right) \\ = \; & \frac{1}{4\sigma^4} - \frac{\sigma^2 s_{i,j}}{2\sigma^6 s_{i,j}} + \frac{3\sigma^4 s_{i,j}^2}{4\sigma^8 s_{i,j}^2} = \frac{1}{2\sigma^4}. \end{aligned}$$

The form of $I_{(\mu,\sigma^2)}(W_*)$ follows from the fact that

$$I_{(\mu,\sigma^2)}(W_*) = \sum_{j=1}^{J}\sum_{i=1}^{I_j} I_{(\mu,\sigma^2)}(Y_{i,j})$$

because of the independence of the $Y_{i,j}$. □

Kahle et al. (2016) also consider the situation that $x_{0,j} > 0$ and $w^{(j)}(x_{0,j})$ are unknown nuisance parameters for $j = 1, \ldots, J$. In such case, the maximum likelihood estimators and the information matrix are more complicated.

5.5.3 Theorem (Compare Kahle et al. 2016, p. 27)
If $\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{J} x_{I_j,j} = M > 0$ *and* $(\widehat{\mu}, \widehat{\sigma}^2)$ *is the maximum likelihood estimator for* (μ, σ^2) *then* $\mathbb{C}_N$ *given by*

$$\mathbb{C}_N(w_*) = \left\{ (\mu,\sigma^2);\ \frac{(\widehat{\mu}-\mu)^2}{\sigma^2}\sum_{j=1}^{J} x_{I_j,j} + \frac{(\widehat{\sigma}^2-\sigma^2)^2}{2\sigma^4} N \le \chi^2_{2;1-\alpha} \right\}$$

or

$$\mathbb{C}_N(w_*) = \left\{ (\mu,\sigma^2);\ \frac{(\widehat{\mu}-\mu)^2}{\widehat{\sigma}^2}\sum_{j=1}^{J} x_{I_j,j} + \frac{(\widehat{\sigma}^2-\sigma^2)^2}{2\widehat{\sigma}^4} N \le \chi^2_{2;1-\alpha} \right\}$$

are asymptotic $(1-\alpha)$*-confidence sets for* (μ, σ^2).

5.5.4 Remark
Note that $N = \sum_{j=1}^{J} I_j \to \infty$ means $J \to \infty$ or $I_j \to \infty$, compare Kahle et al. (2016), p. 24/25. However the second case implies $\frac{1}{J}\sum_{j=1}^{J} x_{I_j,j} \to \infty$ since $\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{J} x_{I_j,j} = M$. Hence the second case is usually not satisfied for degradation processes,

Proof. Because of $\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{J} x_{I_j,j} = M > 0$ we have

$$\lim_{N\to\infty} \frac{1}{N} I_{(\mu,\sigma^2)}(W_*) = \begin{pmatrix} \frac{M}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} =: I_{(\mu,\sigma^2)} \tag{5.13}$$

so that

$$\sqrt{N}\left(\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma}^2 \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0_2, I^{-1}_{(\mu,\sigma^2)})$$

implying

$$\sqrt{N} I^{1/2}_{(\mu,\sigma^2)} \left(\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma}^2 \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0_2, I_{2\times 2})$$

or with (5.13) and the Lemma of Slutzky

$$
\begin{aligned}
&\left(I_{(\mu,\sigma^2)}(W_*)\right)^{1/2}\left(\begin{pmatrix}\widehat{\mu}\\ \widehat{\sigma}^2\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma^2\end{pmatrix}\right)\\
&= \left(\frac{1}{N}I_{(\mu,\sigma^2)}(W_*)\right)^{1/2}\sqrt{N}\left(\begin{pmatrix}\widehat{\mu}\\ \widehat{\sigma}^2\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma^2\end{pmatrix}\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0_2, I_{2\times 2}).
\end{aligned}
$$

Hence we have

$$
\left(\begin{pmatrix}\widehat{\mu}\\ \widehat{\sigma}^2\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma^2\end{pmatrix}\right)^{\top} I_{(\mu,\sigma^2)}(W_*)\left(\begin{pmatrix}\widehat{\mu}\\ \widehat{\sigma}^2\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma^2\end{pmatrix}\right)\xrightarrow{\mathcal{D}}\chi_2^2
$$

or using again the Lemma of Slutzky and the convergence of maximum likelihood estimators

$$
\left(\begin{pmatrix}\widehat{\mu}\\ \widehat{\sigma}^2\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma^2\end{pmatrix}\right)^{\top} I_{(\widehat{\mu},\widehat{\sigma}^2)}(W_*)\left(\begin{pmatrix}\widehat{\mu}\\ \widehat{\sigma}^2\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma^2\end{pmatrix}\right)\xrightarrow{\mathcal{D}}\chi_2^2
$$

which provides the assertion. □

Now we want to determine a prediction interval for the first passage time $X_c^{(1)}$. In particular, we want to determine this if the first process satisfies $v := \max\{w^{(1)}(x_{i,1});\ i = 1,\dots,I_1\} < c$ so that we want to determine a prediction interval for the first passage time $X_c^{(1)}$ of c after $x_0 := x_{I_1,1}$.

5.5.5 Theorem
If $w_0 := w^{(1)}(x_{I_1,1})$, $q_\alpha(\mu,\lambda,t_0)$ is the α-quantile of the shifted inverse Gaussian distribution with mean μ, shape λ, and shift t_0, $0 \le \eta_1 < \eta_2 \le 1$ with $\eta_2 - \eta_1 = 1 - \alpha_1$, $\mathbb{C}_N$ is an asymptotic $(1-\alpha_2)$-confidence set for (μ,σ^2) then $\mathbb{P}_N$ given by

$$
\mathbb{P}_N(w_*) = \begin{cases} \bigcup_{(\mu,\sigma)\in\mathbb{C}(w_*)}\left[q_{\eta_1}\left(\frac{c-w_0}{\mu},\frac{(c-w_0)^2}{\sigma^2},x_0\right), q_{\eta_2}\left(\frac{c-w_0}{\mu},\frac{(c-w_0)^2}{\sigma^2},x_0\right)\right], & \text{if } v < c\\ [x_{i_0-1,1}, x_{i_0,1}] \quad \text{with } \ i_0 = \arg\min\{i;\ w^{(1)}(x_{i,1}) \ge c\}, & \text{if } v \ge c, \end{cases}
$$

is an asymptotic $(1-\alpha_1-\alpha_2)$-prediction interval for the first passage time $X_c^{(1)}$ of c of the process $W^{(1)}$ which is observed until $x_0 = x_{I_1,1}$ with $w^{(1)}(x_0) = w_0$.

5.5.6 Remark
Note that here the level of the prediction interval is $(1-\alpha_1-\alpha_2)$ instead of $(1-\alpha_1)(1-\alpha_2)$ as we had before. This is due to the fact that $X_c^{(1)}$ depends on $W^{(1)}(x_{1,1}),\dots,W^{(1)}(x_{I_1,1})$ and thus on W_*. Note also that nothing is to predict if $v \ge c$.

Proof. Set $V := \max\{W^{(1)}(x_{i,1});\ i = 1,\dots,I_1\}$, $W_0 = W^{(1)}(x_0)$, and

$$
q_{\eta_i}(\mu,\sigma^2,w_0) := q_{\eta_i}\left(\frac{c-w_0}{\mu},\frac{(c-w_0)^2}{\sigma^2},x_0\right)
$$

for $i = 1, 2$. In particular $V < c$ implies $W_0 < c$. Then we have

$$
\begin{aligned}
& P_{(\mu_*,\sigma_*^2)}\left(X_c^{(1)} \notin \mathbb{P}_N(W_*)\right) \\
&= P_{(\mu_*,\sigma_*^2)}\left(X_c^{(1)} \notin \mathbb{P}_N(W_*), V \geq c\right) + P_{(\mu_*,\sigma_*^2)}\left(X_c^{(1)} \notin \mathbb{P}_N(W_*), V < c\right) \\
&= P_{(\mu_*,\sigma_*^2)}\left(X_c^{(1)} \notin \bigcup_{(\mu,\sigma)\in\mathbb{C}_N(W_*)} \left[q_{\eta_1}(\mu,\sigma^2,W_0), q_{\eta_2}(\mu,\sigma^2,W_0)\right], V < c\right) + 0 \\
&= P_{(\mu_*,\sigma_*^2)}\Bigg(X_c^{(1)} \notin \bigcup_{(\mu,\sigma)\in\mathbb{C}_N(W_*)} \left[q_{\eta_1}(\mu,\sigma^2,W_0), q_{\eta_2}(\mu,\sigma^2,W_0)\right], \\
&\qquad\qquad (\mu_*,\sigma_*^2) \in \mathbb{C}_N(W_*), V < c\Bigg) \\
&\quad + P_{(\mu_*,\sigma_*^2)}\Bigg(X_c^{(1)} \notin \bigcup_{(\mu,\sigma)\in\mathbb{C}_N(W_*)} \left[q_{\eta_1}(\mu,\sigma^2,W_0), q_{\eta_2}(\mu,\sigma^2,W_0)\right], \\
&\qquad\qquad (\mu_*,\sigma_*^2) \notin \mathbb{C}_N(W_*), V < c\Bigg) \\
&\leq P_{(\mu_*,\sigma_*^2)}\left(X_c^{(1)} \notin \left[q_{\eta_1}(\mu_*,\sigma_*^2,W_0), q_{\eta_2}(\mu_*,\sigma_*^2,W_0)\right], V < c\right) \\
&\qquad + P_{(\mu_*,\sigma_*^2)}\left((\mu_*,\sigma_*^2) \notin \mathbb{C}_N(W_*)\right).
\end{aligned}
$$

Since V depends only on $W_{x\leq x_0}^{(1)}$, we obtain with the tower rule

$$
\begin{aligned}
& \lim_{N\to\infty} P_{(\mu_*,\sigma_*^2)}\left(X_c^{(1)} \notin \mathbb{P}_N(W_*)\right) \\
&\leq \lim_{N\to\infty} P_{(\mu_*,\sigma_*^2)}\left(X_c^{(1)} \notin \left[q_{\eta_1}(\mu_*,\sigma_*^2,W_0), q_{\eta_2}(\mu_*,\sigma_*^2,W_0)\right], V < c\right) + \alpha_2 \\
&= \lim_{N\to\infty} E_{(\mu_*,\sigma_*^2)}\left(\mathbb{I}\{X_c^{(1)} \notin \left[q_{\eta_1}(\mu_*,\sigma_*^2,W_0), q_{\eta_2}(\mu_*,\sigma_*^2,W_0)\right]\} \cdot \mathbb{I}\{V < c\}\right) + \alpha_2 \\
&= \lim_{N\to\infty} E_{(\mu_*,\sigma_*^2)}\Big(E_{(\mu_*,\sigma_*^2)}\Big(\mathbb{I}\{X_c^{(1)} \notin \left[q_{\eta_1}(\mu_*,\sigma_*^2,W_0), q_{\eta_2}(\mu_*,\sigma_*^2,W_0)\right]\} \\
&\qquad\qquad \cdot \mathbb{I}\{V < c\} \big| W^{(1)}(x)_{x\leq x_0}\Big)\Big) + \alpha_2 \\
&= \lim_{N\to\infty} E_{(\mu_*,\sigma_*^2)}\left(E_{(\mu_*,\sigma_*^2)}\left(\mathbb{I}\{X_c^{(1)} \notin \left[q_{\eta_1}(\mu_*,\sigma_*^2,W_0), q_{\eta_2}(\mu_*,\sigma_*^2,W_0)\right]\} \big| W^{(1)}(x)_{x\leq x_0}\right)\right. \\
&\qquad\qquad \left.\cdot \mathbb{I}\{V < c\}\right) + \alpha_2 \\
&\overset{(*)}{\leq} E(\mu_*,\sigma_*^2)\left(\alpha_1 \cdot \mathbb{I}\{V < c\}\right) + \alpha_2 \leq \alpha_1 + \alpha_2.
\end{aligned}
$$

Thereby $(*)$ follows from Theorem 5.4.9 and the fact that the stochastic part of $W^{(1)}$ given $W^{(1)}(x)_{x\leq x_0} = w^{(1)}(x)_{x\leq x_0}$ is a Wiener process $\widetilde{W}$ starting at x_0 with $\widetilde{W}(x_0) = w^{(1)}(x_0) = w_0$. $\square$

5.6 Gamma processes for modeling degradation processes

Although the Wiener process tends to increase for $\mu > 0$, it is not strictly increasing. A strictly increasing stochastic process is given by the gamma process. It bases on the gamma distribution so that we give at first some properties of the Gamma distribution. These can be shown easily by the Laplace transform (also called moment-generating function) of a real-valued random variable which provides a unique characterization of the distribution.

5.6.1 Definition
If $T : \Omega \to \mathbb{R}$ is a random variable then $\mathcal{L} : \mathbb{R} \to \mathbb{R}$ is a Laplace transform of T if

$$\mathcal{L}_T(s) = E(e^{-sT})$$

for all $s \in \mathbb{R}$.

5.6.2 Lemma
If $T \sim \mathcal{G}(\lambda, \beta)$ then the Laplace transform of T for $s \geq 0$ is given by

$$\mathcal{L}_T(s) = \left(\frac{\lambda}{\lambda + s}\right)^{\beta}.$$

Proof.

$$\mathcal{L}_T(s) = \int_0^\infty e^{-st}\,\frac{\lambda^\beta\, t^{\beta-1}}{\Gamma(\beta)}\,e^{-\lambda t}\,dt = \frac{\lambda^\beta}{(\lambda+s)^\beta}\int_0^\infty \frac{(\lambda+s)^\beta\, t^{\beta-1}}{\Gamma(\beta)}\,e^{-(\lambda+s)t}\,dt = \frac{\lambda^\beta}{(\lambda+s)^\beta}. \square$$

5.6.3 Corollary
If $T_1, \dots, T_N$ are independent with $T_n \sim \mathcal{G}(\lambda, \beta_n)$ for $n = 1, \dots, N$ then $\sum_{n=1}^N T_n \sim \mathcal{G}(\lambda, \sum_{n=1}^N \beta_n)$.

Proof. The independence of $T_1, \dots, T_N$ implies

$$\begin{aligned}
&\mathcal{L}_{\sum_{n=1}^N T_n}(s) = E\left(e^{-s\sum_{n=1}^N T_n}\right) \\
&= \quad E\left(\prod_{n=1}^N e^{-s\,T_n}\right) = \prod_{n=1}^N E\left(e^{-s\,T_n}\right) = \prod_{n=1}^N \left(\frac{\lambda}{\lambda+s}\right)^{\beta_n} = \left(\frac{\lambda}{\lambda+s}\right)^{\sum_{n=1}^N \beta_n}
\end{aligned}$$

which is the Laplace transform of a random variable with $\mathcal{G}(\lambda, \sum_{n=1}^N \beta_n)$ distribution. $\square$

5.6.4 Corollary
If $T \sim \mathcal{G}(\lambda, \beta)$ and $c > 0$ then $cT \sim \mathcal{G}\left(\frac{\lambda}{c}, \beta\right)$.

Proof.

$$\mathcal{L}_{cT}(s) = E\left(e^{-scT}\right) = \left(\frac{\lambda}{\lambda + sc}\right)^{\beta} = \left(\frac{\frac{\lambda}{c}}{\frac{\lambda}{c} + s}\right)^{\beta}$$

which is the Laplace transform of a random variable with $\mathcal{G}\left(\frac{\lambda}{c}, \beta\right)$ distribution. □

5.6.5 Definition (Gamma process, see Kahle et al. 2016, p. 68)
Let $B : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing and right-continous function with $B(0) = 0$ and $\lambda > 0$. A real valued stochastic process $Y = \{Y(x);\ x \geq 0\}$ is called a Gamma process with shape function B and rate λ, shortly $Y \sim \mathcal{GP}(\lambda, B)$, if

(i) $Y(0) = 0$,

(ii) $Y(x_2) - Y(x_1) \sim \mathcal{G}(\lambda, B(x_2) - B(x_1))$ *for all* $0 \leq x_1 \leq x_2$,

(iii) the increments $Y(x_{i+1}) - Y(x_i)$, $i = 1, \ldots, n$, *are independent for all* $n \in I\!N$, $n > 1$, *and* $0 \leq x_1 < x_2 < \ldots < x_{n+1}$.

5.6.6 Lemma
$Y \sim \mathcal{GP}(\lambda, B)$ has the following properties:

(i) $cY \sim \mathcal{GP}\left(\frac{\lambda}{c}, B\right)$ *for all* $c > 0$.

(ii) $Y(x_3) - Y(x_1) = Y(x_3) - Y(x_2) + Y(x_2) - Y(x_1) \sim \mathcal{G}(\lambda, B(x_3) - B(x_2) + B(x_2) - B(x_1))$ *for all* $x_3 > x_2 > x_1 \geq 0$.

(iii) $Y(x_2) \geq Y(x_1)$ *for all* $x_2 > x_1 \geq 0$.

(iv) $Y(x) \sim \mathcal{G}(\lambda, B(x))$ *for all* $x \geq 0$.

(v) If B satisfies $B(x) = bx$ for all $x \geq 0$ then $Y(x_2) - Y(x_1) \sim Y(x_2 - x_1)$ for all $x_2 > x_1 \geq 0$, which means that it has stationary increments.

Proof. (i) follows from Corollary 5.6.4 and (ii) from Corollary 5.6.3. Property (iii) follows from (ii) which implies $Y(x_2) - Y(x_1) \geq 0$. Because of $Y(x) = Y(x) - Y(0) \sim \mathcal{G}(\lambda, B(x) - B(0)) = \mathcal{G}(\lambda, B(x))$, the property (iv) holds. For (v) note that

$$\begin{aligned} & Y(x_2) - Y(x_1) \sim \mathcal{G}(\lambda, B(x_2) - B(x_1)) \\ = \;\; & \mathcal{G}(\lambda, b(x_2 - x_1)) = \mathcal{G}(\lambda, B(x_2 - x_1) - B(0)) \sim Y(x_2 - x_1) - Y(0). \;\square \end{aligned}$$

If B is a continuous function then, according to Lemma 5.6.6 (ii), a Gamma process $Y \sim \mathcal{GP}(\lambda, B)$ satisfies

$$Y(x) = \sum_{i=1}^{I} \left(Y\left(\frac{i}{I}\, x\right) - Y\left(\frac{i-1}{I}\, x\right)\right).$$

However, a path (realization) y of a Gamma process is a pure jump process with infinite countable jumps in a bounded interval. In particular,

$$Y(x) = \sum_{n=1}^{\infty} V_n \mathbb{1}_{[0,t]}(U_n)$$

where $M = \sum_{n=1}^{\infty} e_{(U_n,V_n)}$ is a Poisson (counting) random measure on $\mathbb{R}_+^2$ with intensity measure

$$\nu(du, dv) = \varphi(du)\frac{e^{-v}}{v}dv$$

where φ is the measure on $\mathbb{R}_+$ with $\varphi([0,x]) = B(x)$ for all $x \geq 0$, see Kahle et al. (2016), p. 71-74.

A path (realization) y of a Gamma process $Y \sim \mathcal{GP}(\lambda, B)$ on $[0,\xi]$ can be simulated with the following algorithm (see Kahle et al. 2016, p. 81): Divide the interval $[0,\xi]$ into a grid such that $0 = x_0 < x_1 < x_2 < \ldots < x_{I-1} < x_I = \xi$ and set $Y(0) = Y(x_0) = 0$ and $i = 1$. Iterate then the following steps:

1. $i = i + 1$.
2. Generate a random number z from $\mathcal{G}(\lambda, B(x_i) - B(x_{i-1}))$.
3. Set $y(x_i) = y(x_{i-1}) + z$.
4. If $i < I$, go to step 1.

The simulated process will approach a realization of the Gamma process the better the smaller $\max\{x_i - x_{i-1};\ i = 1, \ldots, I\}$ is.

The points $y(0), y(x_1), \ldots, y(x_I)$ can be interpolated to get the complete path. However, since the path of a Gamma process is a pure jump process, a better approximations are

$$y^{(I-)}(x) := \sum_{i=1}^{I} y(x_{i-1})\mathbb{1}_{[x_{i-1},x_i)}(x)$$

or

$$y^{(I+)}(x) := \sum_{i=1}^{I} y(x_i)\mathbb{1}_{(x_{i-1},x_i]}(x)$$

so that $y^{(I-)}(x) \leq y(x) \leq y^{(I+)}(x)$. Since these approximations are piecewise constant, it is difficult to determine first passage times. Therefore Kahle et al. (2016) provides improved approximations on the Pages 84-89.

Since the Gamma process is nondecreasing, the distribution of the first passage time X_c of a critical c is easy to obtain.

5.6.7 Theorem
The survival function $\overline{F}_{X_c}$ of the first passage time X_c of a Gamma process $Y \sim \mathcal{GP}(\lambda, B)$ is given by

$$\overline{F}_{X_c}(x) = F_{\mathcal{G}(\lambda,B(x))}(c),$$

where $F_{\mathcal{G}(\lambda,B(x))}$ is the cumulative distribution function of the $\mathcal{G}(\lambda, B(x))$-distribution.

Proof. Since $X_c := \inf\{x;\ Y(x) \geq c\}$, it holds

$$\overline{F}_{X_c}(x) = P(X_c > x) = P(Y(x) < c) = P(Y(x) \leq c) = F_{\mathcal{G}(\lambda,B(x))}(c)$$

because of $Y(x) \sim \mathcal{G}(\lambda, B(x))$ according to Lemma 5.6.6 (iv). □

The distribution of X_c has no nice form. However, an α-quantile can be easily calculated using the fact that the cumulative distribution function $F_{X_c}(x) = 1 - \overline{F}_{X_c}(x) = 1 - F_{\mathcal{G}(\lambda,B(x))}(c)$ is strictly increasing in x. The following algorithm with an given small $\epsilon > 0$ can be used:
Set $x_0 = 0$ and $x_1 > 0$ arbitrary.
While $F_{X_c}(x_1) < \alpha$ set $x_0 = x_1$ and $x_1 = 2x_1$.
While $|F_{X_c}(x_1) - \alpha| > \epsilon$ do:
if $F_{X_c}\left(\frac{x_0+x_1}{2}\right) < \alpha$ set $x_0 = \frac{x_0+x_1}{2}$,
if $F_{X_c}\left(\frac{x_0+x_1}{2}\right) > \alpha$ set $x_1 = \frac{x_0+x_1}{2}$.

For the statistical inference, again the path (realization) of a Gamma process $y = \{y(x);\ x \geq x_0\}$ can be observed only at given points $x_1 < \ldots < x_I$ so that we have only observations $y(x_1), \ldots, y(x_I)$. To be more general, we assume here aigain that eventually several processes, J processes, are observed, eventually at different time points $x_{1,j} < \ldots < x_{I_j,j}$ for $j = 1, \ldots, J$. Then the available observations are $y^{(j)}(x_{1,j}), \ldots, y^{(j)}(x_{I_j,j})$ which are realizations of $Y^{(j)}(x_{1,j})$, $\ldots, Y^{(j)}(x_{I_j,j})$ for $j = 1, \ldots, J$. We assume here again $x_{0,j} = 0$ and $Y^{(j)}(x_{0,j}) = 0$ for simplicity. Then the *increments*

$$z_{i,j} = y^{(j)}(x_{i,j}) - y^{(j)}(x_{i-1,j})$$

are realizations of independent variables

$$Z_{i,j} = Y^{(j)}(x_{i,j}) - Y^{(j)}(x_{i-1,j}) \sim \mathcal{G}(\lambda, B(x_{i,j}) - B(x_{i-1,j}))$$

for $i = 1, \ldots, I_j$, $j = 1, \ldots, J$. Set

$$Y_* := (Y^{(1)}(x_{1,1}), \ldots, Y^{(1)}(x_{I_1,1}), \ldots, Y^{(J)}(x_{1,J}), \ldots, Y^{(J)}(x_{I_J,J}))^\top,$$

with realization

$$y_* := (y^{(1)}(x_{1,1}), \ldots, y^{(1)}(x_{I_1,1}), \ldots, y^{(J)}(x_{1,J}), \ldots, y^{(J)}(x_{I_J,J}))^\top.$$

Then the likelihood function for the data set w_* is given by

$$l(\lambda, B|y_*) = \prod_{j=1}^{J} \prod_{i=1}^{I_j} \frac{\lambda^{B(x_{i,j})-B(x_{i-1,j})} z_{i,j}^{B(x_{i,j})-B(x_{i-1,j})-1} e^{-\lambda z_{i,j}}}{\Gamma(B(x_{i,j}) - B(x_{i-1,j}))}.$$

Maximum likelihood estimators can be only obtained if the function $B : [0, \infty) \to [0, \infty)$ is given by a parametric function $B(\theta)$ with $\theta \in \mathbb{R}^p$. The simplest function is given by $B(t) = bt$ so that $\theta = b \in \mathbb{R}$. With

$$v_{i,j} := x_{i,j} - x_{i-1,j}$$

for $i = 1, \ldots, I_j$, $j = 1, \ldots, J$ we get then

$$l(\lambda, b|y_*) = \prod_{j=1}^{J} \prod_{i=1}^{I_j} \frac{\lambda^{bv_{i,j}} z_{i,j}^{bv_{i,j}-1} e^{-\lambda z_{i,j}}}{\Gamma(bv_{i,j})}.$$

However, also in this case, the maximum likelihood estimator has no simple form and must be calculated numerically. Also the information matrix is complicated. Therefore we present here only the naive (plug-in) prediction interval. For more explicit forms of the maximum likelihood estimator and the information matrix, see Kahle et al. (2016), p.113-117.

5.6.8 Theorem (Naive (plug-in) prediction interval)
If $q_\alpha(\lambda, \theta)$ is the α-quantile of the first passage time $X_c^{(0)}$ of c of the process $Y^{(0)} \sim \mathcal{GP}(\lambda, B(\theta))$, $0 \le \eta_1 < \eta_2 \le 1$ with $\eta_2 - \eta_1 = 1 - \alpha$, and $(\widehat{\lambda}, \widehat{\theta})$ is a consistent estimator for (λ, θ) based on y_* then $\mathbb{P}$ given by

$$\mathbb{P}(y_*) = \left[q_{\eta_1}\left(\widehat{\lambda}, \widehat{\theta}\right), q_{\eta_2}\left(\widehat{\lambda}, \widehat{\theta}\right) \right]$$

is an asymptotic $(1-\alpha)$-prediction interval for the first passage time $X_c^{(0)}$ of c of the process $Y^{(0)}$.

5.6.9 Remark
If a Gamma process Y starts at $x_0 > 0$ with $Y(x_0) = y_0$, i.e. only (i) of Definition 5.6.5 is not necessarily satisfied, then the process Y_0 with $Y_0(x) = Y(x + x_0) - y_0$ is a Gamma process in the sense of Definition 5.6.5 with B_0 given by $B_0(x) = B(x) - B(x_0)$. In particular, for the first passage time, we have $X_c := \inf\{x \ge x_0;\ Y(x) \ge c\}$ and

$$\begin{aligned} \overline{F}_{X_c}(x) &= P(X_c > x) = P(Y(x) < c) = P(Y(x) - y_0 < c - y_0) \\ &= P(Y(x) - y_0 \le c - y_0) = P(Y_0(x - x_0) \le c - y_0) = F_{\mathcal{G}(\lambda, B_0(x-x_0))}(c - y_0). \end{aligned}$$

Hence also prediction intervals for the first passage time of $c > y_0$ of the process $Y^{(1)}$ can be constructed if the process is observed already until $y^{(1)}(x_0) = y_0$.

5.7 Crack growth via stochastic differential equations

The Paris-Erdogan law describes crack growth by the differential equation

$$\frac{dl}{dt} = \theta_1 l^{\theta_2}.$$

5.7.1 Lemma
The differential equation given by the Paris-Erdogan law has a solution for $\theta_2 \neq 1$ given by

$$l = l(t) = \alpha_1 (s(t-\alpha_0))^{s\alpha_2} \tag{5.14}$$

with $s = sign(1-\theta_2),\ s\alpha_0 < s\,t,\ \alpha_1 = (\theta_1[s(1-\theta_2)])^{\frac{1}{1-\theta_2}},\ \alpha_2 = \frac{1}{s(1-\theta_2)} > 0.$

Proof. Since

$$\frac{s\alpha_2 - 1}{s\alpha_2} = \frac{\frac{1}{1-\theta_2} - 1}{\frac{1}{1-\theta_2}} = \theta_2 \quad \text{and} \quad \frac{1}{s\alpha_2} = 1 - \theta_2$$

we get

$$\begin{aligned} l'(t) &= s\alpha_2\alpha_1 (s(t-\alpha_0))^{s\alpha_2 - 1}\, s = \alpha_2 \alpha_1^{\frac{1}{s\alpha_2}} \alpha_1^{\frac{s\alpha_2-1}{s\alpha_2}} \left[(s(t-\alpha_0))^{s\alpha_2}\right]^{\frac{s\alpha_2-1}{s\alpha_2}} = \alpha_2 \alpha_1^{\frac{1}{s\alpha_2}} l(t)^{\theta_2} \\ &= \frac{1}{s(1-\theta_2)} (\theta_1[s(1-\theta_2)])^{\frac{1}{1-\theta_2}(1-\theta_2)} l(t)^{\theta_2} = \theta_1 l(t)^{\theta_2}. \ \square \end{aligned}$$

5.7.2 Lemma
(i) The inverse form of the solution (5.14) is

$$t = t(l) = \beta_0 + \beta_1\, l^{\beta_2}$$

with $\beta_0 = \alpha_0,\ \beta_1 = s\left(\frac{1}{\alpha_1}\right)^{\frac{s}{\alpha_2}},\ \beta_2 = \frac{s}{\alpha_2}.$

(ii) The function t is then a solution of the differential equation

$$\frac{dt}{dl} = \theta_1 (s(t-\theta_0))^{\theta_2}$$

with $s = sign(\beta_1) = sign(\beta_2),\ \theta_0 = \beta_0,\ \theta_1 = s\,\beta_2 (s\,\beta_1)^{\frac{1}{\beta_2}} > 0,\ \ \theta_2 = 1 - \frac{1}{\beta_2}.$

Proof. We prove only (ii) for $s = -1$ since the case $s = 1$ is an exercise. For $s = -1$, we get

$$\begin{aligned} \frac{dt}{dl} &= \beta_1\,\beta_2\, l^{\beta_2 - 1} = -(-\beta_1)\,\beta_2 \left(l^{\beta_2}\right)^{\frac{\beta_2-1}{\beta_2}} = -(-\beta_1)^{\frac{1}{\beta_2}} (-\beta_1)^{1-\frac{1}{\beta_2}} \beta_2 \left(l^{\beta_2}\right)^{1-\frac{1}{\beta_2}} \\ &= -(-\beta_1)^{\frac{1}{\beta_2}} \beta_2 \left(\beta_0 - \beta_0 - \beta_1\, l^{\beta_2}\right)^{1-\frac{1}{\beta_2}} = -(-\beta_1)^{\frac{1}{\beta_2}} \beta_2\, (\beta_0 - t)^{1-\frac{1}{\beta_2}} \\ &= \theta_1\, (-(t-\theta_0))^{\theta_2}. \ \square \end{aligned}$$

Hence we have in both cases a deterministic differential equation of the form

$$\frac{dy}{dx} = \theta_1(s(y-\theta_0))^{\theta_2}, \quad s \in \{-1, 1\},$$

where $y = l$, $x = t$ in the original form and $y = t$, $x = l$ in the inverse form.

We obtain a **stochastic differential equation (SDE)** if we add to the deterministic differential equation an error term

$$dY(x) = \theta_1(s(Y(x)-\theta_0))^{\theta_2}dx + \sigma_{\theta_3}(x, Y(x))\, dE(x). \tag{5.15}$$

Often $E(x) = B(x)$ is the Brownian motion. However, it could be also the Gamma process or some other stochastic process. Then the process given by $Y = \{Y(x);\ x \geq 0\}$ is also a stochastic process.

The process given by (5.15) with $E(x) = B(x)$ is a special case of a diffusion process.

5.7.3 Definition (Diffusion process, see e.g. Iacus 2008, p. 33)
A real valued stochastic process $Y = \{Y(x);\ x \geq 0\}$ is called a diffusion process if

$$dY(x) = b(x, Y(x))\, dx + \sigma(x, Y(x))\, dB(x), \tag{5.16}$$

with drift $b : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$, diffusion $\sigma : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ and initial value $Y(0)$.

Hence the drift and diffusion of the process given by (5.15) are given by

$$b(x, y) = \theta_1(s(y-\theta_0))^{\theta_2}, \ \ \sigma(x, y) = \sigma_{\theta_3}(x, y).$$

Special cases are the Ornstein-Uhlenbeck (Vasicek) process, the Black-Scholes-Merton process (geometric Brownian motion), and the Cox-Ingersoll-Ross process.

5.7.4 Definition (Ornstein-Uhlenbeck / Vasicek process, Black-Scholes-Merton process / geometric Brownian motion, Cox-Ingersoll-Ross process, see e.g. Iacus 2008, p. 43-48)
A diffusion process $Y = \{Y(x);\ x \geq 0\}$ is a

1. *Ornstein-Uhlenbeck / Vasicek process if*
$$dY(x) = (\beta_1 - \beta_2 Y(x))\,dx + \beta_3\,dB(x), \tag{5.17}$$
i.e. $b(x,y) = \beta_1 - \beta_2\,y$, $\sigma(x,y) = \beta_3$,

2. *Black-Scholes-Merton process / geometric Brownian motion if*
$$dY(x) = -\beta_2 Y(x)\,dx + \beta_3\,Y(x)\,dB(x), \tag{5.18}$$
i.e. $b(x,y) = -\beta_2\,y$, $\sigma(x,y) = \beta_3\,y$,

3. *Cox-Ingersoll-Ross process if*
$$dY(x) = (\beta_1 - \beta_2 Y(x))\,dx + \beta_3\,\sqrt{Y(x)}\,dB(x), \tag{5.19}$$
i.e. $b(x,y) = \beta_1 - \beta_2\,y$, $\sigma(x,y) = \beta_3\,\sqrt{y}$.

The stoachastic differential equation (5.16) is interpreted as

$$Y(x) = Y(0) + \int_0^x b(u, Y(u))\,du + \int_0^x \sigma(u, Y(u))\,dB(u). \tag{5.20}$$

The first ingral is the usual Riemann integral, only applied to a random function so that this integral is a random variable in $\mathbb{R}$. Its realization is the Riemann integral of $\int_0^x b(u, Y(u)(\omega))\,du$. The second integral is a stochastic integral.

5.7.5 Definition (Stochastic integral / Itô integral, see e.g. Iacus 2008, p. 30, 31)
Let $\Pi_n([a,b])_{n\in I\!N}$ be a sequence of partitions of $[a,b] \subset \mathbb{R}$ such that $\Pi_n([a,b]) = (x_0^n, x_1^n, \ldots, x_n^n)$, $a = x_0^n < x_1^n < \ldots, x_{n-1}^n < x_n^n = b$ and $\lim_{n\to\infty} \max\{x_{j+1}^n - x_j^n;\ j = 0, 1, \ldots, n-1\} = 0$. If $Z : \Omega \to \mathbb{R}^{[a,b]}$ is a stochastic process on $[a,b]$, then the stochastic integral

$$I(Z) = \int_a^b Z(u)\,dB(u)$$

with respect to the Brownian motion is defined as the limit in quadratic mean of $I(Z^{(n)})$, where $Z^{(n)}$ is a simplified process of Z defined by

$$Z^{(n)}(x)(\omega) := Z(x_j^n)(\omega), \quad \text{for} \quad x_j^n \leq x < x_{j+1}^n,$$

and

$$I(Z^{(n)}) := \sum_{j=0}^{n-1} Z^{(n)}(x_j^n)\,\{B(x_{j+1}^n) - B(x_j^n)\} = \sum_{j=0}^{n-1} Z(x_j^n)\,\{B(x_{j+1}^n) - B(x_j^n)\}.$$

The convergence $I(Z^{(n)}) \to I(Z)$ is not in the usual sense since B has no finite variation. However, $I(Z^{(n)})$ and $I(Z)$ are random variables so that $\mathsf{E}\left([I(Z^{(n)}) - I(Z)]^2\right) \to 0$ must be satisfied. If this is satisfied, then Z is called *Itô integrable*. Necessary for this is that

$$\int_a^b \mathsf{E}(Z(u)^2)\, du < \infty \tag{5.21}$$

is satisfied and that

$$Z \text{ is a stochastic process adapted to the natural filtration of the Brownian motion} \tag{5.22}$$

which means that $Z(x_j^n)$ and $B(x_{j+1}^n) - B(x_j^n)$ are independet for all j and n. This a reason why the step function $Z^{(n)}(\cdot)(\omega)$ is defined via the values of $Z(\cdot)(\omega)$ at the beginning of the intervals of $[x_j^n, x_{j+1}^n)$ and not in the middle of these intervals.

5.7.6 Lemma (See e.g. Iacus 2008, p. 31-33)
If (5.21) and (5.22) are satisfied so that the stochastic process $Z : \Omega \to \mathbb{R}^{[a,b]}$ is Itô integrable, then

a) $\mathsf{E}\left([I(Z^{(n)})]^2\right) = \sum_{j=0}^{n-1} \mathsf{E}\left([Z^{(n)}(x_j^n)]^2\right)(x_{j+1}^n - x_j^n),$

b) $\mathsf{E}\left(\int_a^b Z(u)\, dB(u)\right) = 0,$

c) $\operatorname{var}\left(\int_a^b Z(u)\, dB(u)\right) = \int_a^b \mathsf{E}\left(Z(u)^2\right) du$ *(Itô isometry).*

Proof.
a) Property (5.22), $\mathsf{E}\left(B(x_{j+1}^n) - B(x_j^n)\right) = 0$, and $\mathsf{E}\left(\left[B(x_{j+1}^n) - B(x_j^n)\right]^2\right) = x_{j+1}^n - x_j^n$ yield

$$\begin{aligned}
\mathsf{E}\left([I(Z^{(n)})]^2\right) &= \mathsf{E}\left(\left(\sum_{j=0}^{n-1} Z^{(n)}(x_j^n)\ \{B(x_{j+1}^n) - B(x_j^n)\}\right)^2\right)\\
&\overset{(5.22)}{=} 2\sum_{j=0}^{n-1}\sum_{l=j+1}^{n-1} \mathsf{E}\left(Z^{(n)}(x_j^n)\ Z^{(n)}(x_l^n)\ \{B(x_{j+1}^n) - B(x_j^n)\}\right)\cdot \mathsf{E}\left(B(x_{l+1}^n) - B(x_l^n)\right)\\
&\quad + \sum_{j=0}^{n-1} \mathsf{E}\left(Z^{(n)}(x_j^n)^2\right)\cdot \mathsf{E}\left(\left[B(x_{j+1}^n) - B(x_j^n)\right]^2\right)\\
&= \sum_{j=0}^{n-1} \mathsf{E}\left([Z^{(n)}(x_j^n)]^2\right)(x_{j+1}^n - x_j^n).
\end{aligned}$$

b) The assertion follows from

$$\mathsf{E}\left(I(Z)\right) = \lim_{n\to\infty} \mathsf{E}\left(I(Z^{(n)})\right) \overset{(5.22)}{=} \lim_{n\to\infty} \sum_{j=0}^{n-1} \mathsf{E}(Z(x_j^n))\ \mathsf{E}\left(B(x_{j+1}^n) - B(x_j^n)\right) = 0.$$

c) The assertion follows from a) and b) using Riemann integration in the last step

$$\begin{aligned}\operatorname{var}(I(Z)) &\overset{(b))}{=} \mathsf{E}\left(I(Z)^2\right) = \lim_{n\to\infty} \mathsf{E}\left(I(Z^{(n)})^2\right)\\ &\overset{(a))}{=} \lim_{n\to\infty} \mathsf{E}\left([Z^{(n)}(x_j^n)]^2\right)(x_{j+1}^n - x_j^n) = \int_a^b \mathsf{E}\left(Z(u)^2\right)\, du. \square\end{aligned}$$

5.7.7 Lemma (See e.g. Iacus 2008, p. 33)
If $Y : \Omega \to \mathbb{R}^{[a,b]}$ and $Z : \Omega \to \mathbb{R}^{[a,b]}$ are Itô integrable stochastic processes and $k, l \in \mathbb{R}$, then

$$\int_a^b (k\, Y(x) + l\, Z(x))\, dB(x) = k \int_a^b Y(x)\, dB(x) + l \int_a^b Z(x)\, dB(x).$$

5.7.8 Lemma (See e.g. Iacus 2008, p. 33)
If $a, \xi \in \mathbb{R}$ then

a) $$\int_0^\xi a\, dB(x) = a \int_0^\xi dB(x) = a\, B(\xi),$$

b) $$\int_0^\xi B(x)\; dB(x) = \frac{1}{2}\, B(\xi)^2 - \frac{1}{2}\, \xi.$$

Using the Itô formula, the process given by (5.16) or (5.20), respectively, can be given more explicitely for special diffusion processes.

5.7.9 Lemma (Itô formula, see e.g. Iacus 2008, p. 38)
Let be $Y : \Omega \to \mathbb{R}^{[a,b]}$ a stochastic process and $f : \mathbb{R}^+ \times \mathbb{R} \ni (x, y) \to f(x, y) \in \mathbb{R}$ a twice differentiable function on both x and y with

$$f_x(x, y)) := \frac{\partial f(x, y)}{\partial x}, \quad f_y(x, y)) := \frac{\partial f(x, y)}{\partial y}, \quad f_{yy}(x, y)) := \frac{\partial^2 f(x, y)}{\partial y \partial y}.$$

Then

$$f(x, Y(x)) = f(0, Y(0)) + \int_0^x f_x(u, Y(u))\, du + \int_0^x f_y(u, Y(u))\, dY(u) + \frac{1}{2} \int_0^x f_{yy}(u, Y(u))\, (dY(u))^2$$

or in differential form

$$df(x, Y(x)) = f_x(x, Y(x))\, du + f_y(x, Y(x))\, dY(x) + \frac{1}{2} f_{yy}(x, Y(x))\, (dY(x))^2.$$

Integrals with respect to $dY(u)$ are defined as for the Brownian motion. However integrals with respect to $(dY(u))^2$ are more complicated. But $(dB(u))^2$ can be treated like du. Moreover terms of the form $(du \cdot dB(u))$ and $(du)^2$ are of order $o(du)$ so that they can be neglected.

5.7.10 Lemma (Black-Scholes-Merton process / geometric Brownian motion, see e.g. Iacus 2008, p. 39,40)
The SDE given by (5.18) has the solution

$$Y(x) = Y(0)\,\exp\left\{\left(-\beta_2 - \frac{\beta_3^2}{2}\right)x + \beta_3\,B(x)\right\}$$

for $x > 0$.

Proof. Set

$$f(x,y) = Y(0)\,\exp\left\{\left(-\beta_2 - \frac{\beta_3^2}{2}\right)x + \beta_3\,y\right\}.$$

Thus, $f(x,B(x)) = Y(x)$ and

$$f_x(x,y)) = \left(-\beta_2 - \frac{\beta_3^2}{2}\right)\,f(x,y),\;\; f_y(x,y)) = \beta_3\,f(x,y),\;\; f_{yy}(x,y)) = \beta_3^2\,f(x,y).$$

The Itô formula yields

$$\begin{aligned} dY(x) = df(x,B(x)) &= \left(f_x(x,B(x)) + \frac{1}{2}f_{yy}(x,B(x))\right)dx + f_y(x,B(x))\,dB(x)\\ = &\left(\left(-\beta_2 - \frac{\beta_3^2}{2}\right)Y(x) + \frac{1}{2}\beta_3^2\,Y(x)\right)dx + \beta_3\,Y(x)\,dB(x) = -\beta_2\,Y(x)\,dx + \beta_3\,Y(x)\,dB(x), \end{aligned}$$

which is (5.18). □

This result is a special case of a more general result which is obtained by using $f(x,y) = \log(y)$ in the Itô formula, namely that

$$Y(x) = Y(0)\,\exp\left\{\int_0^x \left(b_1(u) - \frac{1}{2}\sigma_1(u)^2\right)du + \int_0^x \sigma_1(u)\,dB(u)\right\}$$

is the solution of the SDE

$$dY(x) = b_1(x)\,Y(x)\,dx + \sigma_1(x)\,Y(x)\,dB(x),$$

see Iacus (2008), p. 39.

5.7.11 Lemma (Ornstein-Uhlenbeck / Vasicek process, Cox-Ingersoll-Ross process, see e.g. Iacus 2008, p. 44 and 47)
The SDE given by

$$dY(x) = (\beta_1 - \beta_2 Y(x))\, dx + \beta_3\, Y(x)^\gamma\, dB(x),$$

with $\gamma \in \mathbb{R}$ *has the solution*

$$Y(x) = \frac{\beta_1}{\beta_2} + \left(Y(0) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2\, x} + \beta_3 \int_0^x e^{-\beta_2 (x-u)}\, Y(u)^\gamma\, dB(u) \tag{5.23}$$

for $x > 0$. *In particular the SDE given by (5.17) has the solution*

$$Y(x) = \frac{\beta_1}{\beta_2} + \left(Y(0) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2\, x} + \beta_3 \int_0^x e^{-\beta_2 (x-u)}\, dB(u)$$

and the the SDE given by (5.19) has the solution

$$Y(x) = \frac{\beta_1}{\beta_2} + \left(Y(0) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2\, x} + \beta_3 \int_0^x e^{-\beta_2 (x-u)}\, \sqrt{Y(u)}\, dB(u)$$

for $x > 0$.

Proof. Set

$$f(x, y) = y\, e^{\beta_2\, x}.$$

Then

$$f_x(x, y)) = \beta_2\, f(x, y), \quad f_y(x, y)) = e^{\beta_2\, x}, \quad f_{yy}(x, y) = 0,$$

so that the Itô formula provides

$$\begin{aligned}
Y(x)\, e^{\beta_2\, x} = f(x, Y(x)) &= f(0, Y(0)) + \int_0^x \beta_2\, Y(u)\, e^{\beta_2\, u}\, du + \int_0^x e^{\beta_2\, u}\, d(Y(u)) \\
&= Y(0) + \int_0^x \beta_2\, Y(u)\, e^{\beta_2\, u}\, du + \int_0^x e^{\beta_2\, u}\, (\beta_1 - \beta_2 Y(u))\, du + \int_0^x e^{\beta_2\, u}\, \beta_3\, Y(u)^\gamma\, dB(u) \\
&= Y(0) + \frac{\beta_1}{\beta_2}\left(e^{\beta_2\, x} - 1\right) + \beta_3 \int_0^x e^{\beta_2\, u}\, Y(u)^\gamma\, dB(u) \\
&= \frac{\beta_1}{\beta_2} e^{\beta_2\, x} + Y(0) - \frac{\beta_1}{\beta_2} + \beta_3 \int_0^x e^{\beta_2\, u}\, Y(u)^\gamma\, dB(u).
\end{aligned}$$

Multiplying by $e^{-\beta_2\, x}$ yields the assertion. □

The simulation of the Black-Scholes-Merton process (geometric Brownian motion) is easy according to Lemma 5.7.10: as soon as the Brownian motion has been simulated, it can be directly used to get the Black-Scholes-Merton process (geometric Brownian motion) (see Iacus 2008, p. 25). For the simulation of the Ornstein-Uhlenbeck / Vasicek process and the Cox-Ingersoll-Ross process, also a Brownian motion must be simulated, but then it must be used in the Itô integrals appearing in Lemma 5.7.11. This is done by using the approximation of the Itô integral given by Definition 5.7.5, see Iacus (2008), p.45.

Hence to simulate the solution $Y(x)$ given by (5.23) use a fine partition $0 = x_0^n < x_1^n < \ldots, x_{n-1}^n < x_n^n = x$ and calculate

$$
\begin{aligned}
Y(x_{j+1}) &= \frac{\beta_1}{\beta_2} + \left(Y(x_j) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2 x_{j+1}} + \beta_3 \int_{x_j}^{x_{j+1}} e^{-\beta_2(x_{j+1}-u)} Y(u)^\gamma \, dB(u) \\
&\approx \frac{\beta_1}{\beta_2} + \left(Y(0) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2 x_{j+1}} + \beta_3 e^{-\beta_2 x_{j+1}} \sum_{i=1}^{j} e^{\beta_2 x_j} Y(x_i)^\gamma Z_i \\
&\overset{(*)}{=} \frac{\beta_1}{\beta_2} + \left(Y(x_j) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2 (x_{j+1}-x_j)} + \beta_3 e^{-\beta_2(x_{j+1}-x_j)} Y(x_j)^\gamma Z_j.
\end{aligned}
$$

Thereby (*) follows by the recursion

$$
\begin{aligned}
Y(x_{j+1}) &= \frac{\beta_1}{\beta_2} + \left(Y(x_j) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2 (x_{j+1}-x_j)} + \beta_3 e^{-\beta_2(x_{j+1}-x_j)} Y(x_j)^\gamma Z_j \\
&= \frac{\beta_1}{\beta_2} + \left[\frac{\beta_1}{\beta_2} + \left(Y(x_{j-1}) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2 (x_j - x_{j-1})} + \beta_3 e^{-\beta_2(x_j - x_{j-1})} Y(x_{j-1})^\gamma Z_{j-1} \right. \\
&\quad \left. - \frac{\beta_1}{\beta_2}\right] e^{-\beta_2 (x_{j+1}-x_j)} + \beta_3 e^{-\beta_2(x_{j+1}-x_j)} Y(x_j)^\gamma Z_j \\
&= \frac{\beta_1}{\beta_2} + \left(Y(x_{j-1}) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2 (x_{j+1}-x_{j-1})} \\
&\quad + \beta_3 e^{-\beta_2(x_{j+1}-x_{j-1})} Y(x_{j-1})^\gamma Z_{j-1} + \beta_3 e^{-\beta_2(x_{j+1}-x_j)} Y(x_j)^\gamma Z_j.
\end{aligned}
$$

However, the solution of a diffusion process given by a stochastic differential equation is not always known. Then a solution can be simulated by the Euler-Maruyama approximation of the SDE. It uses the approximations of the integrals in the interpretation of the SDE in (5.20) via a stochastic integral.

5.7.12 Definition (Euler-Maruyama approximation, Iacus 2008, p.62)
Let Y be a diffusion process given by

$$dY(x) = b(x, Y(x))\, dx + \sigma(x, Y(x))\, dB(x)$$

and $0 = x_0 < x_1 < \ldots < x_{N-1} < x_N = \xi$ a partition of $[0, \xi]$. Then the Euler-Maruyama approximation of Y on $[0, \xi]$ is given by

$$Y(x_{i+1}) \approx Y(x_i) + b(x_i, Y(x_i))\,(x_{i+1} - x_i) + \sigma(x_i, Y(x_i))\,(B(x_{i+1}) - B(x_i))$$

for $i = 0, 1, \ldots, N-1$.

The Euler-Maruyama approximation is the better the smaller the step lengths $x_{i+1} - x_i$ are.

5.8 Crack growth prediction via stochastic differential equations

Assume that a diffucion process Y is observed at points $0 = x_0 < x_1 < \ldots < x_{N-1} < x_N = \xi$ so that $y_n := y(x_n)$ for $n = 0, \ldots, N$ are the observations. The observations $y_n := y(x_n)$ are realizations of $Y_n := Y(x_n)$ for $n = 0, \ldots, N$. The aim is to predict $Y_F := Y(x_F)$ with $x_F > x_N$.

If a diffusion process has a explicit solution with known conditional distributions, then an unknown parameter θ can be estimated by maximum likelihood estimation. This is the case for the Black-Scholes-Merton process / geometric Brownian motion, the Ornstein-Uhlenbeck / Vasicek process and the Cox-Ingersoll-Ross process.

5.8.1 Lemma (Ornstein-Uhlenbeck / Vasicek process, see e.g. Iacus 2008, p. 45)
If $Y(x)$ is the solution of the SDE given by (5.17) then its conditional distribution given $Y(0) = y(0)$ is a normal distribution with

$$E(Y(x)|Y(0) = y(0)) = \frac{\beta_1}{\beta_2} + \left(y(0) - \frac{\beta_1}{\beta_2}\right) e^{-\beta_2 x}, \quad \text{var}(Y(x)|Y(0) = y(0)) = \frac{\beta_3^2 \left(1 - e^{-2\beta_2 x}\right)}{2\beta_2}.$$

Proof. This follows from Lemma 5.7.6 b) and c) using the explicit solution of the process given in Lemma 5.7.11. In particular, we have with 5.7.6 b)

$$\mathsf{E}\left(\int_0^x e^{-\beta_2(x-u)}\, dB(u)\right) = 0$$

and with 5.7.6 c)

$$\begin{aligned}\text{var}\left(\int_0^x e^{-\beta_2(x-u)}\, dB(u)\right) &= \int_0^x \left(e^{-\beta_2(x-u)}\right)^2 du \\ &= e^{-2\beta_2 x} \int_0^x e^{2\beta_2 u} du = e^{-2\beta_2 x}\, \frac{1}{2\beta_2}\, e^{2\beta_2 u}\Big|_0^x = \frac{e^{-2\beta_2 x}}{2\beta_2} \left(e^{2\beta_2 x} - 1\right) = \frac{1 - e^{-2\beta_2 x}}{2\beta_2}. \; \square\end{aligned}$$

5.8.2 Lemma (Black-Scholes-Merton process / geometric Brownian motion, see e.g. Iacus 2008, p. 39,40)
If $Y(x)$ is the solution of the SDE given by (5.18) then its conditional distribution given $Y(0) = y(0)$ is a lognormal distribution with

$$E(\ln(Y(x))|Y(0) = y(0)) = \ln(y(0)) + \left(\beta_1 - \frac{1}{2}\beta_2^2\right) x, \quad \text{var}(\ln(Y(x))|Y(0) = y(0)) = \beta_2^2 x.$$

Proof. The assertion follows at once from the logarithm of the explicit solution given by Lemma 5.7.10

$$\ln(Y(x)) = \ln(Y(0)) + \left(\beta_1 - \frac{\beta_2^2}{2}\right) x + \beta_2\, B(x)$$

and $B(x) \sim \mathcal{N}(0, x)$. $\square$

The conditional distribution for the Cox-Ingersoll-Ross process is more complicated and can be found for example in Iacus (2008), p.47/48.

5.8.3 Definition (Maximum likelihood estimation, see e.g. Iacus 2008, p. 111)
If $f_{\theta,Y_n|Y_{n-1}=y_{n-1}}$ is the density of the conditional distribution of Y_n given Y_{n-1} for $n = 1, \ldots, N$, then the maximum likelihood estimator for θ based on the oberservations $y_0, y_1, \ldots, y_N$ is given by

$$\widehat{\theta} \in \arg\max \prod_{n=1}^{N} f_{\theta,Y_n|Y_{n-1}=y_{n-1}}(y_n).$$

5.8.4 Definition (Point prediction for SDEs)
If $f_{\theta,Y_F|Y_N=y_N}$ is the density of the conditional distribution of Y_F given Y_N and $\widehat{\theta}$ is an estimator for θ, then the point predictor for the expectation $E_\theta(Y_F|Y_N)$ of the process at future time $x_F > x_N$ is given by

$$E_{\widehat{\theta}}(Y_F|Y_N) = \int y\, f_{\widehat{\theta},Y_F|Y_N=y_N}(y)dy.$$

It is more difficult to find prediction intervals. Moreover, the conditional distribution of Y_n given Y_{n-1} is only known in rar cases. Therefore we use now an appromiation strategy for SDEs of the form

$$dY(x) = b(\theta, Y(x))\, dx + \sigma\, \widetilde{b}(\theta, Y(x))\, dE(x) \tag{5.24}$$

with $\theta = (\theta_0, \theta_1, \theta_2)^\top$ and a process with independent increments, i.e. $E(x_2) - E(x_1)$ and $E(x_4) - E(x_3)$ are independent for all (x_1, x_2), (x_3, t_4) with $0 \le x_1 \le x_2 \le x_3 \le x_4$. This could be the Brownian motion but other processes are also possible.

In particular, $b(\theta, y) = \widetilde{b}(\theta, y)$ is possible. This process makes sense if the volatility increases proportional to the drift term. It has the advantage that it has only a four-dimensional unknown parameter vector ζ, namely $\zeta = (\theta_0, \theta_1, \theta_2, \sigma)$.

5.8.5 Theorem (Prediction intervals based on confidence sets for dependent observations)
If the distributions of $Y_1, \ldots, Y_N, Y_F$ are continuous and only dependening on an unknown parameter ζ, $F_{Y_F,\zeta}$ is the cumulative distribution function of Y_F, $0 \le \eta_1 < \eta_2 \le 1$ with $\eta_2 - \eta_1 = 1 - \alpha_1$, and $\mathbb{C}$ given by $\mathbb{C}(y_1, \ldots, y_N)$ is a $(1-\alpha_2)$-confidence set function for ζ based on $y_1, \ldots, y_N$, then $\mathbb{P}$ given by

$$\mathbb{P}(y_1, \ldots, y_N) = \bigcup_{\zeta \in \mathbb{C}(y_1,\ldots,y_N)} \left[F^{-1}_{Y_F,\zeta}(\eta_1), F^{-1}_{Y_F,\zeta}(\eta_2)\right]$$

is a $(1 - \alpha_1 - \alpha_2)$-prediction interval function for Y_F.

Proof. At first note that for any underlying parameter ζ_*

$$P_{\zeta_*}\left(Y_F \in \left[F^{-1}_{Y_F,\zeta_*}(\eta_1), F^{-1}_{Y_F,\zeta_*}(\eta_2)\right]\right) = F_{Y_F,\zeta_*}\left(F^{-1}_{Y_F,\zeta_*}(\eta_2)\right) - F_{Y_F,\zeta_*}\left(F^{-1}_{Y_F,\zeta_*}(\eta_1)\right) = \eta_2 - \eta_1 = 1 - \alpha_1$$

is satisfied. Then we obtain for any ζ_*:

$$\begin{aligned} P_{\zeta_*}(Y_F \notin \mathbb{P}(Y_1,\ldots,Y_N)) = P_{\zeta_*}\left(Y_F \notin \bigcup_{\zeta\in\mathbb{C}(Y_1,\ldots,Y_N)} \left[F^{-1}_{Y_F,\zeta}(\eta_1), F^{-1}_{Y_F,\zeta}(\eta_2)\right]\right) \\ = \; P_{\zeta_*}\left(Y_F \notin \bigcup_{\zeta\in\mathbb{C}(Y_1,\ldots,Y_N)} \left[F^{-1}_{Y_F,\zeta}(\eta_1), F^{-1}_{Y_F,\zeta}(\eta_2)\right],\ \zeta_* \in \mathbb{C}(Y_1,\ldots,Y_N)\right) \\ + P_{\zeta_*}\left(Y_F \notin \bigcup_{\zeta\in\mathbb{C}(Y_1,\ldots,Y_N)} \left[F^{-1}_{Y_F,\zeta}(\eta_1), F^{-1}_{Y_F,\zeta}(\eta_2)\right],\ \zeta_* \notin \mathbb{C}(Y_1,\ldots,Y_N)\right) \\ \le \; P_{\zeta_*}\left(Y_F \notin \left[F^{-1}_{Y_F,\zeta_*}(\eta_1), F^{-1}_{Y_F,\zeta_*}(\eta_2)\right]\right) + P_{\zeta_*}\left(\zeta_* \notin \mathbb{C}(Y_1,\ldots,Y_N)\right) \le \alpha_1+\alpha_2. \;\Box \end{aligned}$$

5.8.6 Remark
If $Y_1,\ldots,Y_N,Y_F$ are independent distributed, then $\mathbb{P}$ given by Theorem 5.8.5 is a $(1-\alpha_1)(1-\alpha_2)$-prediction interval function for Y_F according to Theorem 2.7.6. However, the independence not given here.

Strategy for calculating an approximative $(1-2\alpha)$-prediction interval for Y_F based on $y_1,\ldots,y_N$:

1) Find a $(1-\alpha)$-confidence set $\mathbb{C}_{SDE}(y_1,\ldots,y_N)$ for $\zeta=(\theta_0,\theta_1,\theta_2,\sigma)^\top$ by grid search.

2) Fix a number M of simulations.

3) For each $\zeta\in\mathbb{C}_{SDE}(y_1,\ldots,y_N)$:
   3.1) simulate paths $y(\zeta,x_N)=y_N,\ y^m(\zeta,x_{N+1}),\ \ldots,\ y^m(\zeta,x_F)$ of the SDE
   for $m=1,\ldots,M$,

   3.2) calculate the median $q(0.5,\zeta)$, the $\alpha/2$-quantile $q(\alpha/2,\zeta)$,
   and $(1-\alpha/2)$-quantile $q(1-\alpha/2,\zeta)$ of $\{y^1(\zeta,x_F),\ldots,y^M(\zeta,x_F)\}$.

3) Prediction of Y_F: median of $\{q(0.5,\zeta);\ \zeta\in\mathbb{C}_{SDE}(y_1,\ldots,y_N)\}$.

4) Prediction interval: $\mathbb{P}_{SDE}(y_1,\ldots,y_N)=\bigcup_{\zeta\in\mathbb{C}_{SDE}(y_1,\ldots,y_N)}[q(\alpha/2,\zeta),q(1-\alpha/2,\zeta)]$.

Construction of the confidence set
For the construction of the confidence set, we use the duality between statistical tests and confidence sets, i.e. the $(1-\alpha)$-confidence set $\mathbb{C}_{SDE}$ for $\zeta=(\theta_0,\theta_1,\theta_2,\sigma)^\top$ is the set all ζ_* so that a α-test for $H_0:\zeta=\zeta_*$ is not rejected.

Define the following residuals

$$Res_n(\zeta_*) := \frac{Y_n - Y_{n-1} - b(\theta_*,Y_{n-1})\,(x_n-x_{n-1})}{\sqrt{(x_n-x_{n-1})}\,\sigma_*\,\widetilde{b}(\theta_*,Y_{n-1})}$$

and the following sums

$$\begin{aligned}
S_1(\zeta_*) &= \frac{1}{\sqrt{N-1}}\sum_{n=2}^{N} Res_n(\zeta_*),\\
S_2(\zeta_*) &= \frac{1}{\sqrt{N-1}\sqrt{3-1}}\sum_{n=2}^{N}(Res_n(\zeta_*)^2-1),\\
S_3(\zeta_*) &= \frac{1}{\sqrt{N-1}\sqrt{15}}\sum_{n=2}^{N} Res_n(\zeta_*)^3,\\
S_4(\zeta_*) &= \frac{1}{\sqrt{N-1}\sqrt{105-3^2}}\sum_{n=2}^{N}(Res_n(\zeta_*)^4-3).
\end{aligned}$$

5.8.7 Definition (Residual-Moment-Test (ResMom-Test))
Reject $H_0 : \zeta = \zeta_*$ if

there exists $j = 1, 2, 3, 4$ with $|S_j(\zeta_*)| > q_{\mathcal{N}(0,1),1-\alpha/8}$.

5.8.8 Theorem
Let be $E_n = E(x_n)$. If $E_n - E_{n-1} \sim E_0$ for all $n = 1, \ldots, N$, $\mathsf{E}(E_0) = 0$, $\mathsf{E}(E_0^2) = 1$, $\mathsf{E}(E_0^3) = 0$, $\mathsf{E}(E_0^4) = 3$, $\mathsf{E}(E_0^6) = 15$, $\mathsf{E}(E_0^8) = 105$ then the residual-moment-test is an asymptotic α-level test.

Proof. The Euler-Maruyama approximation (see Definition 5.7.12) provides

$$Y_n - Y_{n-1} \approx b(\theta, Y_{n-1})\,(x_n - x_{n-1}) + \sigma\,\widetilde{b}(\theta, Y_{n-1})\,(E_n - E_{n-1})$$

so that it holds approximately

$$\begin{aligned}
Res_n(\zeta_*) &= \frac{Y_n - Y_{n-1} - b(\theta_*, Y_{n-1})\,(x_n - x_{n-1})}{\sqrt{(x_n - x_{n-1})}\,\sigma_*\,\widetilde{b}(\theta_*, Y_{n-1})}\\
&\approx \frac{1}{\sqrt{(x_n - x_{n-1})}}(E_n - E_{n-1}) \sim E_0
\end{aligned}$$

and $Res_2(\zeta_*), \ldots, Res_N(\zeta_*)$ are independent. The central limit theorem applied to $Res_2(\zeta_*)^j$, $\ldots, Res_N(\zeta_*)^j$ provides

$$S_j(\zeta_*) \longrightarrow \mathcal{N}(0, 1)$$

for $N \to \infty$ and $j = 1, 2, 3, 4$ and thus the assertion using Bonferroni adjustment. □

The assumptions of Theorem 5.8.8 are in particular satisfied if the error process E is the Brownian motion B.

To reduce the moment conditions for E_0 used in Theorem 5.8.8 and the number of test statistics, we can regard also test statistics based on data depth. Kustosz et al. (2015) have shown that data depth reduces in many cases to alternating signs of residuals. Since the signs of residuals are only important we can regard

$$R_n(\theta) = Y_n - Y_{n-1} - b(\theta, Y_{n-1})\,(x_n - x_{n-1}).$$

Then the depth of θ in the data $Y_1, \ldots, Y_N$ with respect to the SDE (5.24) can be measured by

$$\begin{aligned} & d_3(\theta, Y_0, Y_1, \ldots, Y_N) \\ = & \frac{1}{\binom{N}{3}} \sum_{1 \le n_1 < n_2 < n_3 \le N} \big(\mathbb{I}\{R_{n_1}(\theta) > 0, R_{n_2}(\theta) < 0, R_{n_3}(\theta) > 0\} \\ & \qquad + \mathbb{I}\{R_{n_1}(\theta) < 0, R_{n_2}(\theta) > 0, R_{n_3}(\theta) < 0\} \big) \end{aligned}$$

or

$$\begin{aligned} & d_4(\theta, Y_0, Y_1, \ldots, Y_N) \\ = & \frac{1}{N-4} \sum_{n=2}^{N-3} \big(\mathbb{I}\{R_n(\theta) > 0, R_{n+1}(\theta) < 0, R_{n+2}(\theta) > 0, R_{n+3}(\theta) < 0\} \\ & \qquad + \mathbb{I}\{R_n(\theta) < 0, R_{n+1}(\theta) > 0, R_{n+2}(\theta) < 0, R_{n+3}(\theta) > 0\} \big). \end{aligned}$$

Kustosz, Müller and Wendler (2016) showed that a depth measure with $K+1$ residuals should be used for a K-dimensional parameter θ so that only d_4 will be appropriate here since θ is three-dimensional. However it turned out that the depth measure d_3 works also very good for parameters of dimension higher than 2. The reason is that d_3 is based on much more subsets than d_4, namely on $\binom{N}{3}$ subsets instead of $N-4$ subsets. However, the computation of d_3 takes much more time. The `R` Package `GSignTest` of Melanie Horn provides in the function `calcDepth` a fast algorithm to compute K-depth. However, this package needs `Rtools` so that the newest `Rtools` should be implemented.

```
> library(devtools)
> devtools::install$\underline{\mbox{ }}$github("melaniehorn/GSignTest")
> library(GSignTest)}
```

For example, the depth d_3 of a residual vector $(1, -1, 1, -1, 1)^\top$ is then calculated by

```
> calcDepth(resy=c(1,-1,1,-1,1),K=3)
[1] 0.5
```

Define the following statistics

$$\begin{aligned} S_2^*(\zeta_*) &= S_2(\zeta_*) = \frac{1}{\sqrt{N-1}\sqrt{3-1}} \sum_{n=2}^{N} (Res_n(\zeta_*)^2 - 1). \\ S_3^*(\zeta_*) &= N\left(d_3(\theta_*, Y_1, \ldots, Y_N) - \frac{1}{4}\right). \\ S_4^*(\zeta_*) &= \frac{\sqrt{N-4}}{\sqrt{15/64}} \left(d_4(\theta_*, Y_1, \ldots, Y_N) - \frac{1}{8}\right). \end{aligned} \tag{5.25}$$

While the data depths and thus $S_3^*(\zeta_*)$ and $S_4^*(\zeta_*)$ are outlier robust, this is not the case for $S_2^*(\zeta_*)$ given by (5.25). Hence it would make sense to replace $S_2^*(\zeta_*)$ by an outlier robust version.

5.8.9 Definition (Depth-Test 1)
Reject $H_0 : \zeta = \zeta_*$ *if*

$$|S_2^*(\zeta_*)| > q_{\mathcal{N}(0,1),1-\alpha/4} \quad or \quad S_4^*(\zeta_*) < q_{\mathcal{N}(0,1),\alpha/2}.$$

5.8.10 Theorem
If $E_n - E_{n-1} \sim E_0$ *for all* $n = 1, \ldots, N$, $med(E_0) = 0$, $\mathsf{E}(E_0^2) = 1$, $\mathsf{E}(E_0^4) = 3$ *then the Depth-Test 1 is an asymptotic* α*-level test.*

Note the conditions $\mathsf{E}(E_0^2) = 1$, $\mathsf{E}(E_0^4) = 3$ are only necessary because of the use of the nonrobust statistic $S_2^*(\zeta_*)$ given by (5.25).

Proof of Theorem 5.8.10. In Kustosz, Müller and Wendler (2016) it is shown $S_4^*(\zeta_*) \longrightarrow \mathcal{N}(0,1)$ for $N \to \infty$. Moreover the test statistic $S_4^*(\zeta_*)$ indicates against the null hypothesis if it is too small, i.e. the depth of the parameter θ in the data set is too small. The convergence $S_2^*(\zeta_*) \longrightarrow \mathcal{N}(0,1)$ follows as in the proof of Theorem 5.8.8. $\square$

The asymptotic distribution of $S_3^*(\zeta_*)$ follows from the result of Kustosz, Leucht and Müller (2016) which was shown for an $AR(1)$ process. Since the proof bases only on the independence and identical distribution of the residuals, the result holds for any i.i.d. residuals. In particular, the asymptotic distribution is not a normal distribution. The α-quantiles of this distribution can be obtained via the `rexpar` package. For example, the α-quantiles for $\alpha = 0.05,\ 0.01,\ 0.001$ can be obtained as follows:

```
> SimQuants[round(SimQuants[, 1], digits = 3) == round((0.05), digits = 3), 2]
    qvals
-1.254541
> SimQuants[round(SimQuants[, 1], digits = 3) == round((0.01), digits = 3), 2]
    qvals
-2.240396
> SimQuants[round(SimQuants[, 1], digits = 3) == round((0.001), digits = 3), 2]
   qvals
-3.71403
```

5.8.11 Definition (Depth-Test 2)
Reject $H_0 : \zeta = \zeta_*$ *if*

$$|S_2^*(\zeta_*)| > q_{\mathcal{N}(0,1),1-\alpha/4} \quad or \quad S_3^*(\zeta_*) < q^*_{\alpha/2},$$

where q^*_α *is the* α*-quantile of the asymptotic distribution of* $S_3^*(\zeta_*)$.

5.8.12 Theorem
If $E_n - E_{n-1} \sim E_0$ for all $n = 1, \ldots, N$, $med(E_0) = 0$, $\mathsf{E}(E_0^2) = 1$, $\mathsf{E}(E_0^4) = 3$ then the Depth-Test 2 is an asymptotic α-level test.

Having an α-level test for $H_0 : \zeta = \zeta_*$, then, by the well known relationship between tests and confidence sets, a $(1-\alpha)$-confidence set for ζ can be constructed by

$$\mathbb{C}(y_1, \ldots, y_n) := \{\zeta_*;\ H_0 : \zeta = \zeta_* \text{ is not rejected}\}.$$

Prediction intervals for Y_F at $x_F > x_N$ can be constructed by simulating the stochastic processes $y_1^m(\zeta), \ldots, y_L^m(\zeta)$ with Euler-Maruyama approximation at L points $x_N < x_1^* < x_2^* < \ldots < x_L^* = x_F$ using $\zeta \in \mathbb{C}(y_1, \ldots, y_n)$ and setting $y_0^m(\zeta) = y_N$ for $m = 1, \ldots, M$. Let $\widehat{q}_\alpha^M(\zeta)$ be the α-quantile of the simulated observations $y_L^1(\zeta), \ldots, y_L^M(\zeta)$. If M is large enough and $\mathbb{C}(y_1, \ldots, y_n)$ is a $(1-\alpha_2)$-confidence set then

$$\mathbb{P}(y_1, \ldots, y_n) := \bigcup_{\zeta \in \mathbb{C}(y_1, \ldots, y_N)} \left[\widehat{q}_{\eta_1}^M(\zeta), \widehat{q}_{\eta_2}^M(\zeta)\right]$$

with $0 \leq \eta_1 < \eta_2 \leq 1$ and $\eta_2 - \eta_1 = 1 - \alpha_1$ is an approximate $(1-\alpha_1-\alpha_2)$-prediction interval for Y_F at x_F.

Some prediction intervals for $\alpha_1 = \alpha_2 = \alpha/2$ and $\eta_1 = \alpha/4$, $\eta_2 = 1 - \alpha/4$ based on the residual-moment-test and the Depth-Test 1 using $b(\theta, y) = \widetilde{b}(\theta, y) = \theta_1(\theta_0 - y)^{\theta_2}$ are shown in Figure 5.3. Figure 5.4 provides a comparison of the treated prediction intervals using 34 series of the data of Virkler et al. (1979). However, these prediction intervals were calculated by an old method which provides smaller prediction intervals. Hence it would be good to repeat the calculation as described above!

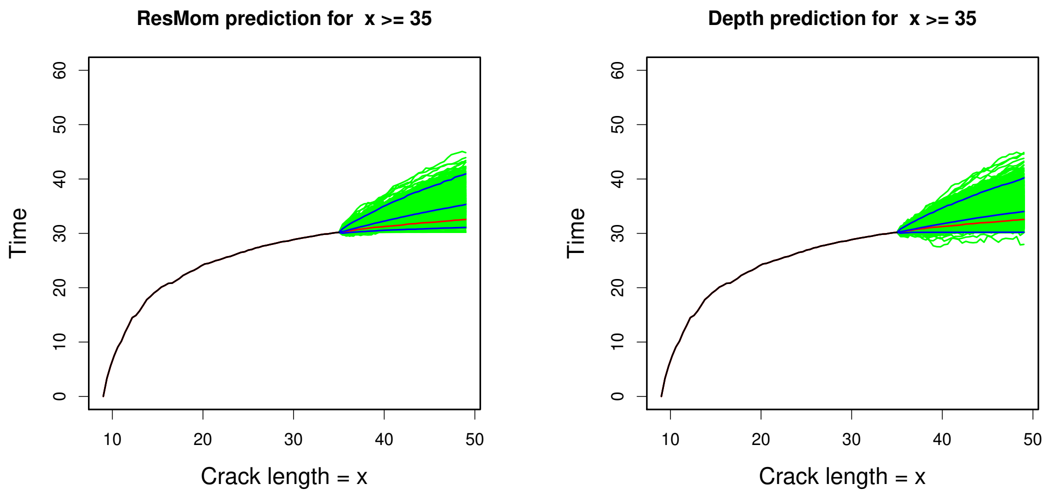

Figure 5.3: Prediction intervals for the Virkler data with the residual-momemt-test and the depth-test

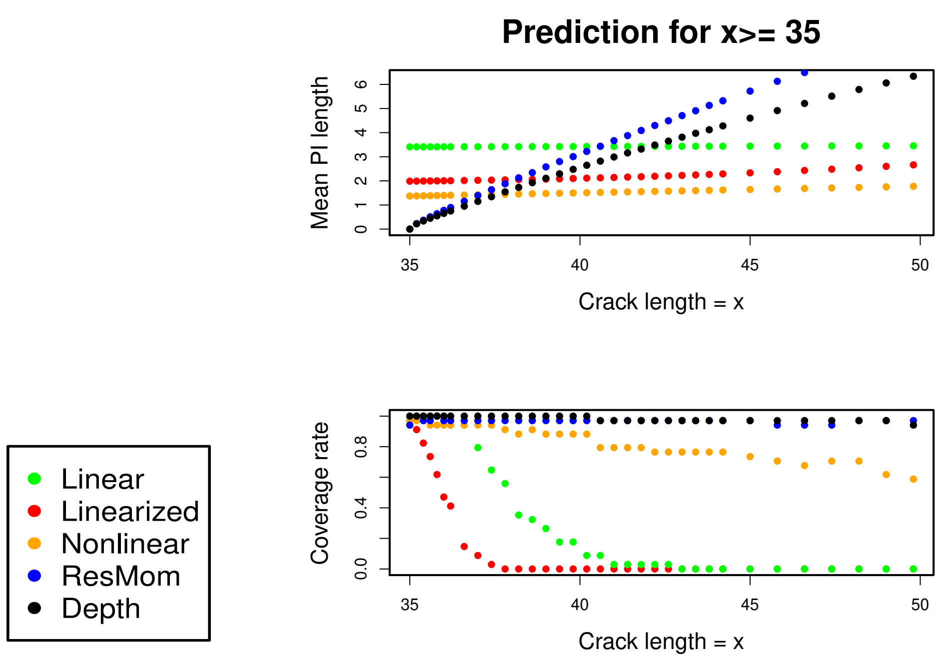

Figure 5.4: Comparison of prediction intervals using 34 series of the Virkler data

Chapter 6

Reliability of systems

6.1 Types of systems

We assume that a system has I components. Let T_* the lifetime of the system, the time up to failure of the system. The lifetimes of the components are denoted by T_i, $i = 1, \dots, I$. Of special interest is often the availability of the system at some time t_*.

6.1.1 Definition (Availability of a system and of components)
The availability of a system with lifetime T_* at time t_* is given by the random variable

$$Z_* := \mathbb{1}_{(t_*,\infty)}(T_*)$$

with realization z_* and the corresponding event is denoted by

$$D_* := \{T_* > t_*\} = \{Z_* = 1\}.$$

The availability of a component $i \in \{1, \dots, I\}$ with lifetime T_i at time t_* is given by the random variable

$$Z_i := \mathbb{1}_{(t_*,\infty)}(T_i)$$

with realization z_i and the corresponding event is denoted by

$$D_i := \{T_i > t_*\} = \{Z_i = 1\}.$$

A system with components in series connection is available at time t_* if all components are available at time t_*.

6.1.2 Definition (Series system)
A system is called a system with components in series connection at time t_* if

$$D_* = \bigcap_{i=1}^{I} D_i \quad \text{or} \quad Z_* = \prod_{i=1}^{I} Z_i,$$

respectively. A system with components in series connection at any time $t_* \geq 0$ is called a series system.

A system with components in parallel connection is available at time t_* if at least one of the components is available at time t_*.

6.1.3 Definition (Parallel system)
A system is called a system with components in parallel connection at time t_* if

$$D_* = \bigcup_{i=1}^{I} D_i \quad \text{or} \quad Z_* = 1 - \prod_{i=1}^{I}(1 - Z_i),$$

respectively. A system with components in parallel connection at any time $t_* \geq 0$ is called a parallel system.

A system may be available at time t_* if at least K of the components is available at time t_*. These system are usually called k-out-of-n systems, see e.g. Kahle and Liebscher (2013), since often the number of components is denoted by n. However, here we will call them K-out-of-I systems.

6.1.4 Definition (K-out-of-I system)
A system is called a K-out-of-I system at time t_* if

$$D_* = \bigcup_{1 \leq i_1 < i_2 < \ldots < i_K \leq I} \bigcap_{k=1}^{K} D_{i_k} \quad \text{or} \quad Z_* = 1 - \prod_{1 \leq i_1 < i_2 < \ldots < i_K \leq I} (1 - \prod_{k=1}^{K} Z_{i_k}),$$

respectively. If this holds for any time $t_* \geq 0$ then the system is called K-out-of-I system.

6.1.5 Example
A more complex system with 6 components is for example given by

$$D_* = (D_1 \cap (D_2 \cup D_6) \cap D_3) \cup (D_1 \cap D_4) \cup (D_3 \cap D_5).$$

Then Z_* has the form

$$Z_* = 1 - [1 - Z_1 \cdot (1 - (1 - Z_2) \cdot (1 - Z_6)) \cdot Z_3]\,[1 - Z_1 \cdot Z_4]\,[1 - Z_3 \cdot Z_5].$$

More complex system are also possible. In particular, complex systems are given by a function $\varphi : \{0,1\}^I \longrightarrow \{0,1\}$ so that $z_* = \varphi(z_1, \ldots, z_I)$ for all $(z_1, \ldots, z_I)^\top \in \{0,1\}^I$. Define $\varphi_i : \{0,1\}^{I+1} \longrightarrow \{0,1\}$ by

$$\varphi_i(y, z_1, \ldots, z_I) := \varphi(z_1, \ldots, z_{i-1}, y, z_{i+1}, \ldots, z_I)$$

for $i = 1, \ldots, I$.

6.1.6 Definition (Irrelevant component)
A component $i \in \{1, \ldots, I\}$ is called irrelevant in the system given by $\varphi : \{0,1\}^I \longrightarrow \{0,1\}$ if

$$\varphi_i(0, z_1, \ldots, z_I) = \varphi_i(1, z_1, \ldots, z_I)$$

for all $(z_1, \ldots, z_I)^\top \in \{0,1\}^I$.

6.1.7 Definition (Coherent system)
A system given by $\varphi : \{0,1\}^I \longrightarrow \{0,1\}$ is coherent if
(i) Each component $i \in \{1, \ldots, I\}$ is not irrelevant.
(ii) $\varphi_i(0, z_1, \ldots, z_I) \leq \varphi_i(1, z_1, \ldots, z_I)$ for all $i \in \{1, \ldots, I\}$ and all $(z_1, \ldots, z_I)^\top \in \{0,1\}^I$.

6.1.8 Example
Consider the system given by $D_* = D_1 \cup (D_2 \cap D_4)$ with $I = 4$ components. Then we have $Z_* = 1 - [1 - Z_1] \cdot [1 - Z_2 \cdot Z_4]$ and

$$\begin{aligned}
&\varphi(z_1, z_2, z_3, z_4) = 1 - [1 - z_1] \cdot [1 - z_2 \cdot z_4],\\
&\varphi_1(0, z_1, z_2, z_3, z_4) = 1 - [1 - 0] \cdot [1 - z_2 \cdot z_4] = z_2 \cdot z_4,\\
&\varphi_1(1, z_1, z_2, z_3, z_4) = 1 - [1 - 1] \cdot [1 - z_2 \cdot z_4] = 1 \geq \varphi_1(0, z_1, z_2, z_3, z_4),\\
&\varphi_2(0, z_1, z_2, z_3, z_4) = 1 - [1 - z_1] = z_1,\\
&\varphi_2(1, z_1, z_2, z_3, z_4) = 1 - [1 - z_1] \cdot [1 - z_4] \geq \varphi_2(0, z_1, z_2, z_3, z_4),\\
&\varphi_3(0, z_1, z_2, z_3, z_4) = \varphi_3(1, z_1, z_2, z_3, z_4),\\
&\varphi_4(0, z_1, z_2, z_3, z_4) = 1 - [1 - z_1] = z_1,\\
&\varphi_4(1, z_1, z_2, z_3, z_4) = 1 - [1 - z_1] \cdot [1 - z_2] \geq \varphi_4(0, z_1, z_2, z_3, z_4).
\end{aligned}$$

Hence the system given by φ is not coherent. However it would be coherent if the component given by Z_3 is excluded.

6.1.9 Example
A system given by

$$\varphi(z_1, z_2, z_3) = 1 - z_1(1 - z_2 \cdot z_3)$$

is not coherent since $\varphi_1(0, z_1, z_2, z_3) = 1 > 0 = z_2 \cdot z_3 = \varphi_1(1, z_1, z_2, z_3)$ if $z_2 = 0$ or $z_3 = 0$.

6.2 Systems with independent components

In this section we assume that the lifetimes $T_1 \ldots, T_I$ of the components of a system are independent. Let the cumulative distribution function and the survival function of the lifetime of component i be $F_{\theta_i,i}$ and $\overline{F}_{\theta_i,i}$, respectively, for $i = 1, \ldots, I$.

6.2.1 Theorem (Lifetime of series systems)
If the lifetimes $T_1 \ldots, T_I$ are independent, then the survival function $\overline{F}_{\theta_1,\ldots,\theta_I}$ and the cumulative distribution function $F_{\theta_1,\ldots,\theta_I}$ of a series system at t_* are given by

$$\overline{F}_{\theta_1,\ldots,\theta_I}(t_*) = \prod_{i=1}^{I} \overline{F}_{\theta_i,i}(t_*) \quad \textit{and} \quad F_{\theta_1,\ldots,\theta_I}(t_*) = 1 - \prod_{i=1}^{I} (1 - F_{\theta_i,i}(t_*))\,.$$

Proof. Because of the independence of $T_1 \ldots, T_I$ we get

$$\begin{aligned}
&\overline{F}_{\theta_1,\ldots,\theta_I}(t_*) = P_{\theta_1,\ldots,\theta_I}(T_* > t_*) = P_{\theta_1,\ldots,\theta_I}(D_*) \\
&= \quad P_{\theta_1,\ldots,\theta_I}(\bigcap_{i=1}^{I} D_i) = \prod_{i=1}^{I} P_{\theta_1,\ldots,\theta_I}(D_i) = \prod_{i=1}^{I} P_{\theta_1,\ldots,\theta_I}(T_i > t_*) = \prod_{i=1}^{I} \overline{F}_{\theta_i,i}(t_*).
\end{aligned}$$

6.2.2 Corollary
If $T_1 \sim \mathcal{E}(\lambda_1), \ldots, T_I \sim \mathcal{E}(\lambda_I)$ are independent, then the lifetime of a series system satisfies $T_* \sim \mathcal{E}(\sum_{i=1}^{I} \lambda_i)$.

6.2.3 Corollary
If the lifetimes $T_1 \ldots, T_I$ are i.i.d. with $F_\theta = F_{\theta_i,i}$ and $\overline{F}_\theta = \overline{F}_{\theta_i,i}$ for $i = 1, \ldots, I$, then the survival function $\overline{F}_{\theta,*} := \overline{F}_{\theta_1,\ldots,\theta_I}$ and the cumulative distribution function $F_{\theta,*} := F_{\theta_1,\ldots,\theta_I}$ of a series system at t_* are given by

$$\overline{F}_{\theta,*}(t_*) = \overline{F}_{\theta}(t_*)^I \quad \textit{and} \quad F_{\theta,*}(t_*) = 1 - (1 - F_\theta(t_*))^I\,.$$

6.2.4 Theorem (Lifetime of parallel systems)
If the lifetimes $T_1 \ldots, T_I$ are independent, then the survival function $\overline{F}_{\theta_1,\ldots,\theta_I}$ and the cumulative distribution function $F_{\theta_1,\ldots,\theta_I}$ of a parallel system at t_* are given by

$$\overline{F}_{\theta_1,\ldots,\theta_I}(t_*) = 1 - \prod_{i=1}^{I} \left(1 - \overline{F}_{\theta_i,i}(t_*)\right) \quad \text{and} \quad F_{\theta_1,\ldots,\theta_I}(t_*) = \prod_{i=1}^{I} F_{\theta_i,i}(t_*).$$

Proof. Because of the independence of $T_1 \ldots, T_I$ we get

$$\begin{aligned} F_{\theta_1,\ldots,\theta_I}(t_*) &= P_{\theta_1,\ldots,\theta_I}(T_* \leq t_*) = P_{\theta_1,\ldots,\theta_I}(\overline{D_*}) \\ &= P_{\theta_1,\ldots,\theta_I}(\overline{\bigcup_{i=1}^{I} D_i}) == P_{\theta_1,\ldots,\theta_I}(\bigcap_{i=1}^{I} \overline{D_i}) = \prod_{i=1}^{I} P_{\theta_1,\ldots,\theta_I}(\overline{D_i}) = \prod_{i=1}^{I} P_{\theta_1,\ldots,\theta_I}(T_i \leq t_*) = \prod_{i=1}^{I} F_{\theta_i,i}(t_*). \end{aligned}$$

6.2.5 Corollary
If the lifetimes $T_1 \ldots, T_I$ are i.i.d. with $F_\theta = F_{\theta_i,i}$ and $\overline{F}_\theta = \overline{F}_{\theta_i,i}$ for $i = 1, \ldots, I$, then the survival function $\overline{F}_{\theta,*} := \overline{F}_{\theta_1,\ldots,\theta_I}$ and the cumulative distribution function $F_{\theta,*} := F_{\theta_1,\ldots,\theta_I}$ of a parallel system at t_* are given by

$$\overline{F}_{\theta,*}(t_*) = 1 - \left(1 - \overline{F}_\theta(t_*)\right)^I \quad \text{and} \quad F_{\theta,*}(t_*) = F_\theta(t_*)^I.$$

6.2.6 Theorem (Lifetime of K-out-of-I systems)
If the lifetimes $T_1 \ldots, T_I$ are i.i.d. with $F_\theta = F_{\theta_i,i}$ and $\overline{F}_\theta = \overline{F}_{\theta_i,i}$ for $i = 1, \ldots, I$, then the survival function $\overline{F}_{\theta,*} := \overline{F}_{\theta_1,\ldots,\theta_I}$ and the cumulative distribution function $F_{\theta,*} := F_{\theta_1,\ldots,\theta_I}$ of a K-out-of-I system at t_* are given by

$$\overline{F}_{\theta,*}(t_*) = \sum_{i=K}^{I} \binom{I}{i} \overline{F}_\theta(t_*)^i \left(1 - \overline{F}_\theta(t_*)\right)^{I-i} \quad \text{and} \quad F_{\theta,*}(t_*) = 1 - \sum_{i=K}^{I} \binom{I}{i} \left(1 - F_\theta(t_*)\right)^i F_\theta(t_*)^{I-i}.$$

Proof. It holds

$$D_* = \{T_* > t_*\} = \left\{ \sum_{i=1}^{I} Z_i \geq K \right\} = \{Y \geq K\}$$

with $Y := \sum_{i=1}^{I} Z_i$. Since $Z_i = 1\!\!1_{(t_*,\infty)}(T_i) \sim \mathsf{Bin}(1, p)$ with $p := P_\theta(Z_i = 1) = P_\theta(T_i > t_*) = \overline{F}_\theta(t_*)$ we have $Y \sim \mathsf{Bin}(I, p)$ so that

$$\begin{aligned} \overline{F}_{\theta,*}(t_*) &= P_\theta(T_* > t_*) = P_\theta(Y \geq K) \\ &= \sum_{i=K}^{I} \binom{I}{i} p^i (1-p)^{I-i} = \sum_{i=K}^{I} \binom{I}{i} \overline{F}_\theta(t_*)^i \left(1 - \overline{F}_\theta(t_*)\right)^{I-i}. \square \end{aligned}$$

6.2.7 Lemma
If $\lim_{y\to\infty} y\,(1-F(y)) = 0$, *then the expectation of a random variable* Y *with positive support and cumulative distribution function* F *is given by*

$$E(Y) = \int_0^\infty (1-F(y))\,dy.$$

Proof. With partial integration we get with $f = F'$

$$\int_0^\infty (1-F(y))\,dy = y\,(1-F(y))\big|_0^\infty - \int_0^\infty y\,(-f(y))\,dy = \int_0^\infty y\,f(y)\,dy = E(Y).$$

6.2.8 Theorem (See Kahle and Liebscher 2013, p. 191)
If $T_1\ldots,T_I \sim \mathcal{E}(\lambda)$ *are i.i.d., then the expected life time of a* K*-out-of-*I *system* T_* *is given by*

$$E(T_*) = \frac{1}{\lambda}\sum_{i=K}^{I}\frac{1}{i}.$$

Proof. The cumulative distribution function of T_* is $F_{\theta,*}$ given by Theorem 6.2.6 with $\theta=\lambda$ and this satisfied for the exponential distribution of the T_i's

$$\begin{aligned}
1-F_{\theta,*}(t) &= \sum_{i=K}^{I}\binom{I}{i}(1-F_\theta(t))^i\,F_\theta(t_*)^{I-i} = \sum_{i=K}^{I}\binom{I}{i}\left(e^{-\lambda t}\right)^i\left(1-e^{-\lambda t}\right)^{I-i}\\
&= \sum_{i=K}^{I}\binom{I}{i}e^{-i\lambda t}\sum_{l=0}^{I-i}\binom{I-i}{l}1^l\left(-e^{-\lambda t}\right)^{I-i-l}\\
&= \sum_{i=K}^{I}\binom{I}{i}e^{-i\lambda t}\sum_{l=0}^{I-i}\binom{I-i}{l}(-1)^{I-i-l}e^{-(I-i-l)\lambda t}\\
&= \sum_{i=K}^{I}\binom{I}{i}\sum_{l=0}^{I-i}\binom{I-i}{l}(-1)^{I-i-l}e^{-(I-l)\lambda t}.
\end{aligned}$$

Hence we have $\lim_{t\to\infty} t(1-F_{\theta,*}(t)) = 0$ so that we get with Lemma 6.2.7

$$\begin{aligned}
E(T_*) &= \int_0^\infty (1-F_{\theta,*}(t))\,dt = \sum_{i=K}^{I}\binom{I}{i}\sum_{l=0}^{I-i}\binom{I-i}{l}(-1)^{I-i-l}\int_0^\infty e^{-(I-l)\lambda t}dt\\
&= \sum_{i=K}^{I}\binom{I}{i}\sum_{l=0}^{I-i}\binom{I-i}{l}(-1)^{I-i-l}\frac{1}{(I-l)\lambda} = \frac{1}{\lambda}\sum_{i=K}^{I}\binom{I}{i}\sum_{l=0}^{I-i}\binom{I-i}{l}(-1)^{I-i-l}\frac{1}{(I-l)}.
\end{aligned}$$

Hence we have only to show that

$$\binom{I}{i}\sum_{l=0}^{I-i}\binom{I-i}{l}(-1)^{I-i-l}\frac{1}{(I-l)} = \frac{1}{i} \tag{6.1}$$

holds for all $I \in I\!N$ and all $i \in \{1, \dots, I\}$. We do an induction over $m = I - i$. The base case with $m = I - i = 0 \Leftrightarrow i = I$ is satisfies because of

$$\binom{I}{i} \sum_{l=0}^{I-i} \binom{I-i}{l} (-1)^{I-i-l} \frac{1}{(I-l)} = \binom{I}{I} \binom{0}{0} (-1)^0 \frac{1}{(I-0)} = \frac{1}{I}.$$

For the inductive step, we assume that (6.1) holds for all (i, I) with $I - i \leq m - 1$ and prove it for (i, I) with $I - i = m$. In particular (6.1) holds then for $(i, I-1)$ and $(i+1, I)$ since $(I-1) - i = I - (i+1) = I - i - 1 = m - 1$. With the binomial formula

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1},$$

we get

$$\begin{aligned}
&\binom{I}{i} \sum_{l=0}^{I-i} \binom{I-i}{l} (-1)^{I-i-l} \frac{1}{(I-l)} \\
&= \binom{I}{i} \left(\binom{I-i}{0} (-1)^{I-i-0} \frac{1}{(I-0)} + \sum_{l=1}^{I-i-1} \left(\binom{I-i-1}{l-1} + \binom{I-i-1}{l} \right) (-1)^{I-i-l} \frac{1}{(I-l)} \right. \\
&\quad + \left. \binom{I-i}{I-i} (-1)^{I-i-(I-i)} \frac{1}{(I-(I-i))} \right) \\
&= \binom{I}{i} \left((-1)^{I-i} \frac{1}{I} + \sum_{l=1}^{I-i-1} \binom{I-i-1}{l-1} (-1)^{I-i-l} \frac{1}{(I-l)} \right. \\
&\quad + \left. \sum_{l=1}^{I-i-1} \binom{I-i-1}{l} (-1)^{I-i-l} \frac{1}{(I-l)} + \frac{1}{i} \right) \\
&= \binom{I}{i} \left((-1)^{I-i} \frac{1}{I} + \sum_{l=1}^{(I-1)-i} \binom{(I-1)-i}{l-1} (-1)^{(I-1)-i-(l-1)} \frac{1}{(I-1-(l-1))} \right. \\
&\quad + \left. \sum_{l=1}^{I-(i+1)} \binom{(I-(i+1)}{l} (-1)^{I-(i+1)-l+1} \frac{1}{(I-l)} + \frac{1}{i} \right)
\end{aligned}$$

$$
\begin{aligned}
&= \binom{I}{i}\Bigg((-1)^{I-i}\frac{1}{I} + \sum_{l=0}^{(I-1)-i-1}\binom{(I-1)-i}{l}(-1)^{(I-1)-i-l}\frac{1}{(I-1-l)} \\
&\quad - \sum_{l=1}^{I-(i+1)}\binom{(I-(i+1)}{l}(-1)^{I-(i+1)-l}\frac{1}{(I-l)} + \frac{1}{i}\Bigg) \\
&= \binom{I}{i}\Bigg((-1)^{I-i}\frac{1}{I} + \sum_{l=0}^{(I-1)-i}\binom{(I-1)-i}{l}(-1)^{(I-1)-i-l}\frac{1}{(I-1-l)} \\
&\quad - \binom{(I-1)-i}{(I-1)-i}(-1)^{(I-1)-i-((I-1)-i)}\frac{1}{(I-1-((I-1)-i))} \\
&\quad - \sum_{l=0}^{I-(i+1)}\binom{(I-(i+1)}{l}(-1)^{I-(i+1)-l}\frac{1}{(I-l)} + \binom{(I-(i+1)}{0}(-1)^{I-(i+1)-0}\frac{1}{(I-0)} + \frac{1}{i}\Bigg) \\
&= \binom{I}{i}\Bigg(\binom{I-1}{i}^{-1}\binom{I-1}{i}\sum_{l=0}^{(I-1)-i}\binom{(I-1)-i}{l}(-1)^{(I-1)-i-l}\frac{1}{(I-1-l)} \\
&\quad - \binom{I}{i+1}^{-1}\binom{I}{i+1}\sum_{l=0}^{I-(i+1)}\binom{(I-(i+1)}{l}(-1)^{I-(i+1)-l}\frac{1}{(I-l)}\Bigg)
\end{aligned}
$$

Because (6.1) holds for $(i, I-1)$ and $(i+1, I)$, we obtain

$$
\begin{aligned}
&\binom{I}{i}\sum_{l=0}^{I-i}\binom{I-i}{l}(-1)^{I-i-l}\frac{1}{(I-l)} \\
&= \binom{I}{i}\left(\binom{I-1}{i}^{-1}\frac{1}{i} - \binom{I}{i+1}^{-1}\frac{1}{i+1}\right) \\
&= \frac{I!}{(I-i)!\,i!}\left(\frac{(I-1-i)!\,i!}{(I-1)!}\frac{1}{i} - \frac{(I-(i+1))!\,(i+1)!}{I!}\frac{1}{i+1}\right) \\
&= \frac{I}{(I-i)\,i} - \frac{1}{I-i} = \frac{1}{i}. \ \square
\end{aligned}
$$

If we observe the life times $T_{*,1}, \ldots, T_{*,N}$ of several i.i.d. systems with i.i.d. components, then the unknown parameter θ can be estimated by the maximum likelihood method and confidence sets can be obtained by the likelihood ratio tests. The densities which are necessary for the maximum likelihood estimator and the confidence sets based on the likelihood ratio test can be obtained by differentiation of the cumulative distribution function.

6.3 Load sharing systems

Often the failures of the components in a system do not happen independently of the failure of the other components. This happens in particular in systems sharing a common external load. As soon as one component fails then the load has to be redistributed over the remaining components so that the remaining components has to carry more load. In particular the lifetimes of the components are not stochastically independent. Such systems are called load sharing systems.

6.3.1 Example
Example 1.0.1 shows the growth curve of the crack width of an initial crack in a prestressed concrete beam. The jumps in the growth curve are caused by the breaking of the tension wires. Since there are 35 tension wire, up to 35 jumps could be observed. However, usually a much smaller number of breaks are observed since then the failure of the beam happens.

The jumps can be treated as outliers and the remaining process can be analyzed by models derived from the Paris-Erdogan equation, see Capter 5. However, these jumps are innovation outliers and the dynamic of this process is mainly caused by these jumps. Moreover, the time points of these jumps can be detected quite exactly by acoustical measurements. Hence in this chapter, we will consider only these time points of jumps. In particular, we have a load sharing system where the component of the system are the 35 tension wires. As soon as one tension wire breaks then the remaining tension wires has to carry a higher load.

The time points $0 = T_0 < T_1 < T_2 < \ldots < T_I$ of failures of the components of a load sharing system can be modelled by a one-dimensional simple point process as introduced in Section 4.2. The one-dimensional versions of the Poisson point process and the Cox point process defined in Definition 4.2.5 are the homogeneous and the inhomogeneous Poisson process.

6.3.2 Definition (Poisson process, see e.g. Jacobsen 2006, p. 19, or Krengel 1991, p. 222-225)
If $d = 1$, then the Cox point process is called inhomogeneous Poisson process and the Poisson point process is called homogeneous Poisson process or shortly Poisson process.

The failures of the components are then the *events* of the point process. The time between events is sometimes called *interarrival time*. However, we will call here the time between two events the *waiting time*.

6.3.3 Theorem (See e.g. Krengel 1991, p. 225)
Let the homogeneous Poisson process with the parameter λ be given by the event times $0 = T_0 < T_1 < T_2 < \ldots$. Then the waiting times between the event times given by $W_i = T_{i+1} - T_i$ for $i \in \mathbb{N}_0$ are independent and identically distributed and the distribution is the exponential distribution with parameter λ.

Obviously, the waiting times of a load sharing system are not identically distributed since the failure rate increases with time. This could be modeled by an inhomogeneous Poisson process with increasing intensity function $\lambda : [0, \infty] \to \mathbb{R}$ like $\lambda(t) = \theta_1 t^{\theta_2}$. However, the waiting time between failure depends on the number of failures which has happened before. Hence there is no continuous increase of the intensity function but a sudden change of the intensity as soon as an event, the failure, happens. The failure will increase the load on the other remaining components.

6.3.4 Definition (State dependent point process, see e.g. Jacobsen 2006, Example 3.1.4)
a) A simple point process given by $0 = T_0 < T_1 < T_2 < \ldots$ is called state dependent, if the waiting times $W_i = T_{i+1} - T_i$ for $i \in \mathbb{N}_0$ are independent but not identically distributed.
b) A state dependent point process is a (shifted) Birth process, if the waiting times W_i are independent and have an exponential distribution with parameter λ_i for $i \in \mathbb{N}_0$.

The intensity parameter λ_i will not only depend on the number of failures but also on the initial (external) load s exposed to the system. Hence we will make the following assumption.

$$W_i \sim \mathcal{E}(\lambda_\theta(i,s)), \;\; i = 0, \ldots, I_{obs} < I,$$

and $W_0, \ldots, W_{I_{obs}}$ are stochastically independent. Thereby I denotes the number of components of the system and I_{obs} the observed number of failures of the system.

A simple assumption for $\lambda_\theta(i,s))$ is

$$\lambda_\theta(i,s) := h_\theta\left(s \cdot \frac{I}{I-i}\right)$$

for some function h_θ depending on $\theta \in \Theta$. The term $\frac{I}{I-i}$ reflects the increased internal load when a failure has happened. In the beginning where no failure has occurred ($i = 0$), then the load is only given by the initial load s of the system. If, for example, the half of the components are failed ($i = \frac{I}{2}$) then the internal load is doubled. The system fails if all of the I components has been failed so that I is maximum number of events. However, the system may also fail when a critical number $I_c < I$ of components has failed.

6.3.5 Example (Basquin link)
If we set

$$h_\theta(x) = \exp(-g_\theta(x)) \quad \text{with} \quad g_\theta(x) = \theta_0 - \theta_1 \ln(x)$$

then it is the Basquin link and we get

$$\begin{aligned}
\ln(\mathsf{E}(W_i)) &= \ln\left(\frac{1}{\lambda_\theta(i,s)}\right) = \ln\left(\frac{1}{h_\theta\left(s \cdot \frac{I}{I-i}\right)}\right) = \ln\left(\exp\left(g_\theta\left(s \cdot \frac{I}{I-i}\right)\right)\right) \\
&= g_\theta\left(s \cdot \frac{I}{I-i}\right) = \theta_0 - \theta_1 \ln\left(s \cdot \frac{I}{I-i}\right).
\end{aligned}$$

6.3.6 Example
A further example for the function h_θ is

$$h_\theta(x) := \exp\left(-\theta_0 + \theta_1 \cdot x - \theta_2 \cdot x^{-\theta_3}\right) = \exp(-g_\theta(x))$$

with $\theta = (\theta_0, \theta_1, \theta_2, \theta_3)^\top \in \Theta = [0,\infty)^4$ and $g_\theta(x) = \theta_0 - \theta_1 \cdot x + \theta_2 \cdot x^{-\theta_3}$ so that

$$\ln(\mathsf{E}(W_i)) = \ln\left(\frac{1}{\lambda_\theta(i,s)}\right) = \theta_0 - \theta_1 \cdot s \cdot \frac{I}{I-i} + \theta_2 \cdot \left(s \cdot \frac{I}{I-i}\right)^{-\theta_3}$$

is strictly decreasing in s and i.

Further we assume that we have J repetitions of the process coming from J different systems which were exposed to possibly different external load levels $s_1, \ldots, s_J$. I.e. we observe realisations $w_{i,j}$, $i = 0, \ldots, I_j$, $j = 1, \ldots, J$, of

$$W_{i,j} \sim \mathcal{E}(\lambda_\theta(i, s_j)),\ i = 0, \ldots, I_j < I,\ j = 1, \ldots, J.$$

Additionally, we have realisations $w_{i,0}$, $i = 0, \ldots, I_0$, of

$$W_{i,0} \sim \mathcal{E}(\lambda_\theta(i, s_0)),\ i = 0, \ldots, I_0,$$

from a new system, where I_0 is much smaller than I. We set $I_0 = -1$ if no failure of the new system was observed. Otherwise, we have $I_0 = 0, 1, \ldots$. We assume that $I_c + 1$ is a critical number of failed components of the new system. The aim is now to predict

$$T_{I_c+1,0} \ := \ w_{0,0} + \ldots + w_{I_0,0} + W_{I_0+1,0} + \ldots W_{I_c,0},\ \ I > I_c > I_0, \tag{6.2}$$

i.e. the time up to the $(I_c + 1)$'th failure of a component of the new system. If we can predict the waiting time $W := W_{I_0+1,0} + \ldots W_{I_c,0}$, then we can of course predict also (6.2). Thereby I_c could be $I - 1$ but also a smaller value.

The prediction interval for W can only be obtained approximately via an asymptotic law. Therefore we need additional assumptions for the design of the stress levels.

Let $d_J := (s_1, \ldots, s_J)^\top \in [0, s_{max}]^J$ be the concrete design and $\delta_J := \sum_{j=1}^J e_{s_j}$ the corresponding design measure on $[0, s_{max}]$, where e_s denote the Dirac (one-point) measure on s. Then we assume

$$\delta_J \longrightarrow \delta \ \text{ weakly for } \ J \longrightarrow \infty. \tag{6.3}$$

Additionally, set $\tilde{d}_J := \left((0, s_0)^\top, \ldots, (I_0, s_0)^\top, (0, s_1)^\top, \ldots, (I_1, s_1)^\top, \ldots, (0, s_J)^\top, \ldots, (I_J, s_J)^\top\right)^\top$ with corresponding design measure $\tilde{\delta}_J$ on $\{0, \ldots, I_{max}\} \times [0, s_{max}]$. Since $\{0, \ldots, I\}$ is finite, the assumption (6.3) implies at once

$$\tilde{\delta}_J \longrightarrow \tilde{\delta} \ \text{ weakly for } \ J \longrightarrow \infty, \tag{6.4}$$

where $\tilde{\delta}$ is a design measure on $\{0, \ldots, I\} \times [0, s_{max}]$.

6.4 Estimation for load sharing systems

Since $W_{0,0}, \ldots, W_{I_0,0}, W_{0,1}, \ldots, W_{I_1,1}, \ldots, W_{0,J}, \ldots, W_{I_J,J}$ are independent, we can easily estimate θ by the maximum likelihood principle, i.e.

$$\widehat{\theta} \in \arg\max_{\theta \in \Theta} \prod_{j=0}^{J} \prod_{i=0}^{I_j} f_{\lambda_\theta(i,s_j)}(w_{i,j})$$

where $f_\lambda(w) = \lambda e^{-\lambda w}$ is the density of the exponential distribution. The estimated expected additional time $\mathsf{E}_\theta(W)$ until the (I_c+1)'th failure is then

$$\mathsf{E}_{\widehat{\theta}}(W) = \mathsf{E}_{\widehat{\theta}}(W_{I_0+1,0} + \ldots + W_{I_c,0}) = \frac{1}{\lambda_{\widehat{\theta}}(I_0+1, s_0)} + \ldots + \frac{1}{\lambda_{\widehat{\theta}}(I_c, s_0)} = g(\widehat{\theta})$$

with

$$g(\theta) \quad := \quad \mathsf{E}_\theta(W) = \frac{1}{\lambda_\theta(I_0+1, s_0)} + \ldots + \frac{1}{\lambda_\theta(I_c, s_0)}. \tag{6.5}$$

Since $w_{0,0}, \ldots, w_{I_0,0}$ are already observed, we get at once that

$$w_{0,0} + \ldots + w_{I_0,0} + \frac{1}{\lambda_{\widehat{\theta}}(I_0+1, s_0)} + \ldots + \frac{1}{\lambda_{\widehat{\theta}}(I_c, s_0)}$$

is the estimated expected time up to the (I_c+1)'th failure.

6.5 Confidence interval for the expected waiting time in load sharing systems

Of special interest is here the confidence interval for the expected waiting time $g(\theta)$ defined by (6.5). However, in Section 6.6 confidence intervals of other aspects of θ are necessary. Therefore let be $g : \Theta \to \Re$ any function of θ so that $\dot{g}(\theta) := \frac{\partial}{\partial\theta} g(\theta)$ exists.

Property (6.4) implies the following central limit theorem for the maximum likelihood estimator (see e.g. Schervish (1995), p. 421-428)

$$\sqrt{N(J)}\,(\widehat{\theta} - \theta) \longrightarrow \mathcal{N}\left(0, I_\theta(\tilde{\delta})^{-1}\right),$$

where $N(J) = I_0 + I_1 + \ldots + I_J + J + 1$ are the observed events/failures in the $J+1$ experiments and

$$I_\theta(\tilde{\delta}) := \int \mathsf{E}_\theta\left(\frac{\partial}{\partial\theta} \ln f_{\lambda_\theta(i,s)}(\tilde{W}_{j,s}) \left(\frac{\partial}{\partial\theta} \ln f_{\lambda_\theta(i,s)}(\tilde{W}_{j,s})\right)^\top\right) \tilde{\delta}(d(i,s))$$

with $\tilde{W}_{i,s} \sim \text{Exp}(\lambda_\theta(i,s))$ is the information matrix. Since we have exponential distribution, the information matrix equals (see e.g. Müller 2013)

$$I_\theta(\tilde{\delta}) = \int \frac{1}{\lambda_\theta(i,s)^2}\,\dot{\lambda}_\theta(i,s)\,\dot{\lambda}_\theta(i,s)^\top\,\tilde{\delta}(d(i,s))$$

with $\dot{\lambda}_\theta(i,s) := \frac{\partial}{\partial\theta}\lambda_\theta(i,s)$. The δ-method provides then

$$\sqrt{N(J)}\,(g(\widehat{\theta}) - g(\theta)) \longrightarrow \mathcal{N}\left(0, \dot{g}(\theta)^\top I_\theta(\tilde{\delta})^{-1}\,\dot{g}(\theta)\right)$$

or, respectively,

$$\frac{\sqrt{N(J)}\,(g(\widehat{\theta}) - g(\theta))}{\sqrt{\dot{g}(\theta)^\top I_\theta(\tilde{\delta})^{-1}\,\dot{g}(\theta)}} \longrightarrow \mathcal{N}(0,1).$$

$I_\theta(\tilde{\delta})$ is estimated by $\frac{1}{N(J)}I(\widehat{\theta})$, where

$$I(\theta) := \sum_{j=0}^{J}\sum_{i=0}^{I_j}\frac{1}{\lambda_\theta(i,s_j)^2}\,\dot{\lambda}_\theta(i,s_j)\,\dot{\lambda}_\theta(i,s_j)^\top.$$

Then we have

$$\dot{g}(\widehat{\theta})^\top\left(\frac{1}{N(J)}I(\widehat{\theta})\right)^{-1}\,\dot{g}(\widehat{\theta}) \longrightarrow \dot{g}(\theta)^\top I_\theta(\tilde{\delta})^{-1}\,\dot{g}(\theta) \text{ in probability}$$

for $J \to \infty$ and Lemma of Slutzky provides

$$\frac{g(\widehat{\theta}) - g(\theta)}{\sqrt{\dot{g}(\widehat{\theta})^\top I(\widehat{\theta})^{-1}\,\dot{g}(\widehat{\theta})}} = \frac{\sqrt{N(J)}\,(g(\widehat{\theta}) - g(\theta))}{\sqrt{\dot{g}(\widehat{\theta})^\top\left(\frac{1}{N(J)}I(\widehat{\theta})\right)^{-1}\,\dot{g}(\widehat{\theta})}} \longrightarrow \mathcal{N}(0,1).$$

Hence an approximate $(1-\alpha)$-confidence interval for $g(\theta)$ is

$$\left[g(\widehat{\theta}) - q_{1-\alpha/2}\sqrt{\dot{g}(\widehat{\theta})^\top I(\widehat{\theta})^{-1}\,\dot{g}(\widehat{\theta})},\; g(\widehat{\theta}) + q_{1-\alpha/2}\sqrt{\dot{g}(\widehat{\theta})^\top I(\widehat{\theta})^{-1}\,\dot{g}(\widehat{\theta})}\right],$$

where q_α is the α-quantil of the standard normal distribution.

6.6 Prediction interval for the waiting time in loadsharing systems

Here we are going to construct an interval $\mathbb{P}(W_*)$ based on

$$W_* := (W_{0,0}, \ldots, W_{I_0,0}, W_{0,1}, \ldots, W_{I_1,1}, \ldots, W_{0,J}, \ldots, W_{I_J,J})^\top$$

so that

$$\lim_{J\to\infty} P_\theta(W_{I_0+1,0} + \ldots W_{I_c,0} \in \mathbb{P}(W_*)) \geq 1-\alpha$$

for all $\theta \in \Theta$. For that we need the distribution of $W := W_{I_0+1,0} + \ldots W_{I_c,0}$.

Since $\lambda_\theta(I_0+1, s_0), \ldots, \lambda_\theta(I_c, s_0)$ are pairwise different, the density of W is given by (see e.g. `https://en.wikipedia.org/wiki/Hypoexponential_distribution`),

$$\begin{aligned} f_{W,\theta}(w) &:= \sum_{i=I_0+1}^{I_c} \lambda_\theta(i,s_0)\, e^{-w\,\lambda_\theta(i,s_0)} \prod_{k=I_0+1,k\neq i}^{I_c} \frac{\lambda_\theta(k,s_0)}{\lambda_\theta(k,s_0)-\lambda_\theta(i,s_0)} \\ &= \sum_{i=I_0+1}^{I_c} \lambda_\theta(i,s_0)\, e^{-w\,\lambda_\theta(i,s_0)}\, a_i(\theta) \end{aligned} \tag{6.6}$$

with $a_i(\theta) := \prod_{k=I_0+1,k\neq i}^{I_c} \frac{\lambda_\theta(k,s_0)}{\lambda_\theta(k,s_0)-\lambda_\theta(i,s_0)}$. Hence the cumulative distribution function is

$$F_{W,\theta}(w) := \sum_{i=I_0+1}^{I_c} a_i(\theta)\ \left(1 - e^{-w\,\lambda_\theta(i,s_0)}\right). \tag{6.7}$$

An α-quantile $b_\alpha(\theta)$ of this distribution can be given only implicitly, namely as

$$H_\alpha(\theta, b_\alpha(\theta)) = 0$$

where

$$H_\alpha(\theta, b) := \sum_{i=I_0+1}^{I_c} a_i(\theta)\ \left(1 - e^{-b\,\lambda_\theta(i,s_0)}\right) - \alpha.$$

An α-quantile can be easily calculated using the fact that the cumulative distribution function is strictly increasing. The following algorithm with an given small $\epsilon > 0$ can be used:
Set $w_0 = 0$ and $w_1 > 0$ arbitrary.
While $F_{W,\theta}(w_1) < \alpha$ set $w_0 = w_1$ and $w_1 = 2w_1$.
While $|F_{W,\theta}(w_1) - \alpha| > \epsilon$ do:
if $F_{W,\theta}\left(\frac{w_0+w_1}{2}\right) < \alpha$ set $w_0 = \frac{w_0+w_1}{2}$,
if $F_{W,\theta}\left(\frac{w_0+w_1}{2}\right) > \alpha$ set $w_1 = \frac{w_0+w_1}{2}$.

As for independent life times, the prediction intervals can be constructed with two general methods.

6.6.1 Theorem (Prediction intervals based on confidence sets for θ)
If $0 \leq \eta_1 < \eta_2 \leq 1$ with $\eta_2 - \eta_1 = 1 - \alpha_1$, $b_\alpha(\theta)$ is the α-quantile of the hypoexponential distribution given in (6.6) or (6.7), and $\mathbb{C}(W_*)$ is a $(1-\alpha_2)$-confidence set for θ then

$$\mathbb{P}(W_*) =:= \bigcup_{\theta \in \mathbb{C}(W_*)} [b_{\eta_1}(\theta), b_{\eta_2}(\theta)]$$

is a $(1-\alpha_1)(1-\alpha_2)$ prediction interval for $W = W_{I_0+1,0} + \ldots W_{I_c,0}$.

The confidence sets for θ can be constructed as in Theorem 3.2.10 or via depth tests as mentioned in Section 5.8 since the waiting times are independent.

As in Theorem 3.2.11, prediction intervals can be contructed with the δ-method for the quantiles as well. For that, note that the implicit function theorem provides for the derivative $\dot{b}_\alpha(\theta) := \frac{\partial}{\partial\theta} b_\alpha(\theta)$ of the quantile $b_\alpha(\theta)$

$$\dot{b}_\alpha(\theta) = -\left(\frac{\partial}{\partial \tilde{b}} H_\alpha(\tilde{\theta}, \tilde{b}) \Big|_{(\tilde{\theta},\tilde{b})=(\theta,b_\alpha(\theta))} \right)^{-1} \frac{\partial}{\partial \tilde{\theta}} H_\alpha(\tilde{\theta}, \tilde{b}) \Big|_{(\tilde{\theta},\tilde{b})=(\theta,b_\alpha(\theta))},$$

which can be calculated explicitly. Setting $g(\theta) = b_{\eta_1}(\theta)$ and $g(\theta) = b_{\eta_2}(\theta)$, respectively, in Section 6.5 provide that

$$\left[b_{\eta_1}(\widehat{\theta}) - q_{1-\alpha_2/2} \sqrt{\dot{b}_{\eta_1}(\widehat{\theta})^\top I(\widehat{\theta})^{-1} \dot{b}_{\eta_1}(\widehat{\theta})}, \infty \right) \tag{6.8}$$

and

$$\left(-\infty, b_{\eta_2}(\widehat{\theta}) + q_{1-\alpha_2/2} \sqrt{\dot{b}_{\eta_2}(\widehat{\theta})^\top I(\widehat{\theta})^{-1} \dot{b}_{\eta_2}(\widehat{\theta})} \right] \tag{6.9}$$

are one-sided asymptotic $1 - \frac{\alpha_2}{2}$-confidence intervals for $b_{\eta_1}(\theta)$ and $b_{\eta_2}(\theta)$, respectively.

6.6.2 Theorem
If $0 \leq \eta_1 < \eta_2 \leq 1$ with $\eta_2 - \eta_1 = 1 - \alpha_1$, then the interval $\mathbb{P}(W_*)$ given by

$$\mathbb{P}(W_*) := \left[b_{\eta_1}(\widehat{\theta}) - \widehat{v}_1, b_{\eta_2}(\widehat{\theta}) + \widehat{v}_2 \right],$$

where

$$\widehat{v}_1 := q_{1-\alpha_2/2} \sqrt{\dot{b}_{\eta_1}(\widehat{\theta})^\top I(\widehat{\theta})^{-1} \dot{b}_{\eta_1}(\widehat{\theta})}$$
$$\widehat{v}_2 := q_{1-\alpha_2/2} \sqrt{\dot{b}_{\eta_2}(\widehat{\theta})^\top I(\widehat{\theta})^{-1} \dot{b}_{\eta_2}(\widehat{\theta})},$$

is an asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval for $W = W_{I_0+1,0} + \ldots W_{I_c,0}$.

Proof. The one-sided confidence intervals given by (6.8) and (6.9) provide

$$P_\theta\left(b_{\eta_1}(\theta) < b_{\eta_1}(\widehat{\theta}) - \widehat{v}_1\right) \leq \frac{\alpha}{2}$$

and

$$P_\theta\left(b_{\eta_2}(\theta) > b_{\eta_2}(\widehat{\theta}) + \widehat{v}_2\right) \leq \frac{\alpha}{2}.$$

Since W and $\widehat{\theta}$ are independent, we obtain

$$\begin{aligned}
&\lim_{J\to\infty} P_\theta(W \in \mathbb{P}(W_*)) \\
&= \lim_{J\to\infty} P_\theta\left(W \geq b_{\eta_1}(\widehat{\theta}) - \widehat{v}_1 \text{ and } W \leq b_{\eta_2}(\widehat{\theta}) + \widehat{v}_2\right) \\
&\geq \lim_{J\to\infty} P_\theta\Big(W \geq b_{\eta_1}(\theta) \text{ and } b_{\eta_1}(\theta) \geq b_{\eta_1}(\widehat{\theta}) - \widehat{v}_1 \text{ and} \\
&\qquad\qquad W \leq b_{\eta_2}(\theta) \text{ and } b_{\eta_2}(\theta) \leq b_{\eta_2}(\widehat{\theta}) + \widehat{v}_2\Big) \\
&= \lim_{J\to\infty} P_\theta\left(b_{\eta_1}(\theta) \leq W \leq b_{\eta_2}(\theta)\right) \\
&\qquad\qquad \cdot \left(1 - P_\theta\left(b_{\eta_1}(\theta) < b_{\eta_1}(\widehat{\theta}) - \widehat{v}_1 \text{ or } b_{\eta_2}(\theta) > b_{\eta_2}(\widehat{\theta}) + \widehat{v}_2\right)\right) \\
&\geq (\eta_2 - \eta_1) \cdot \left(1 - \frac{\alpha}{2} - \frac{\alpha}{2}\right) = (1-\alpha_1)(1-\alpha_2). \square
\end{aligned}$$

Note that $b_\alpha(\theta)$ can be explicitly given if $I_c = I_0 + 1$. Then we have

$$b_\alpha(\theta) = -\frac{\ln(1-\alpha)}{\lambda_\theta(I_0+1, s_0)} = \frac{\ln\left(\frac{1}{1-\alpha}\right)}{\lambda_\theta(I_0+1, s_0)} = \ln\left(\frac{1}{1-\alpha}\right) g(\theta),$$

where $g(\theta)$ is the same as in (6.5) for the case $I_c = I_0 + 1$. Setting

$$\widehat{v} := q_{1-\alpha_2/2}\sqrt{\dot{g}(\widehat{\theta})^\top I(\widehat{\theta})^{-1}\,\dot{g}(\widehat{\theta})},$$

the asymptotic $(1-\alpha_1)(1-\alpha_2)$-prediction interval is then

$$\mathbb{P}(W_*) = \left[\ln\left(\frac{1}{1-\eta_1}\right)\left(g(\widehat{\theta}) - \widehat{v}\right), \ln\left(\frac{1}{1-\eta_2}\right)\left(g(\widehat{\theta}) + \widehat{v}\right)\right].\square$$

Calculation
To calculate the prediction interval, the follwoing steps are needed:
1) Calculation of the maximum likelihood estimate $\widehat{\theta}$.
2) Calculation of $I(\widehat{\theta})^{-1}$.
3) Calculation of $b_{\eta_1}(\widehat{\theta})$ and $b_{\eta_2}(\widehat{\theta})$ as solutions of $H_{\eta_1}(\widehat{\theta}, b_{\eta_1}(\widehat{\theta})) = 0$ and $H_{\eta_2}(\widehat{\theta}, b_{\eta_2}(\widehat{\theta})) = 0$, respectively.
4) Calculation of $\dot{b}_{\eta_1}(\widehat{\theta})$ and $\dot{b}_{\eta_2}(\widehat{\theta})$.

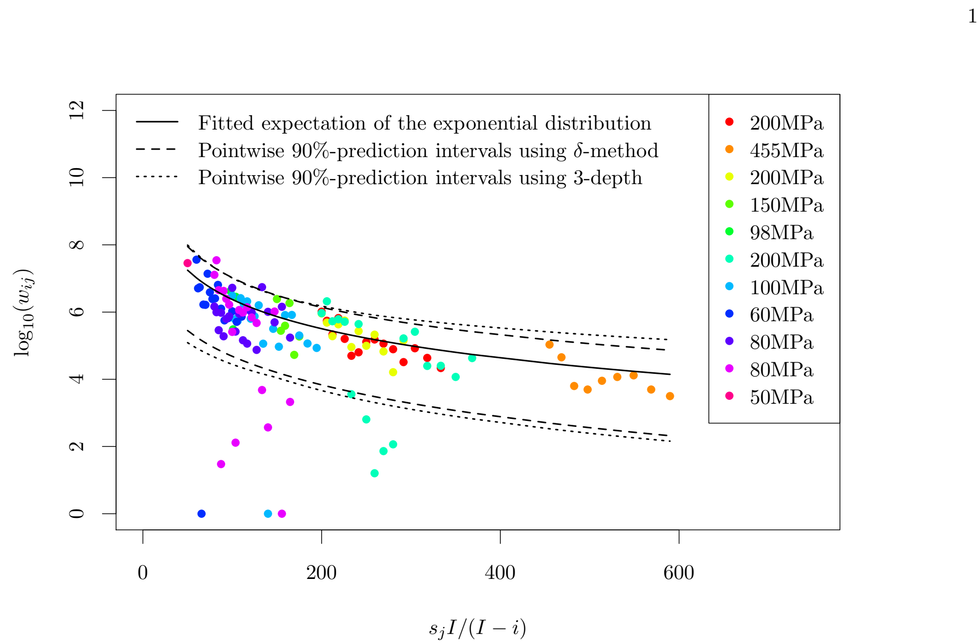

Figure 6.1: Logarithmic waiting times between the breaks of all experiments with pointwise 90%-prediction interval for the next wire break using data depth and the δ-method.

6.6.3 Example

Figure 6.1 shows the waiting times between two breaks of the $I = 35$ tension wires in 11 experiments as that described in Example 1.0.1. It shows the logarithmic expected waiting time for the next break given by

$$\ln(\mathsf{E}(W_{i,j})) = \ln\left(\frac{1}{\lambda_\theta(i, s_j)}\right) = \theta_1 - \theta_2 \ln\left(s_j \cdot \frac{35}{35 - i}\right)$$

where

$$h_\theta(x) := \exp\left(-\theta_1 + \theta_2 \ln(x)\right)$$

was used. It shows also the predictions intervals for $W_{i,j}$ in the logarithmic scale.

6.6.4 Example

Figure 6.2 shows the predictions intervals obtained by the different methods for the 20th break in the experiment SB04 using the $0, 1, \ldots, 19$ breaks observed in the same experiment and the breaks of the other 9 experiments. The horizontal line is the realized number of load cycles before the 20th break. Here the link function is given by

$$\log(\mathsf{E}(W_{i,j})) = \log\left(\frac{1}{\lambda_\theta(i,s_j)}\right) = \theta_1 - \theta_2 \cdot s_j \cdot \frac{35}{35-i} + \theta_3 \cdot \left(s_j \cdot \frac{35}{35-i}\right)^{-\theta_4}$$

with

$$h_\theta(x) := \exp\left(-\theta_1 + \theta_2 \cdot x - \theta_3 \cdot x^{-1}\right).$$

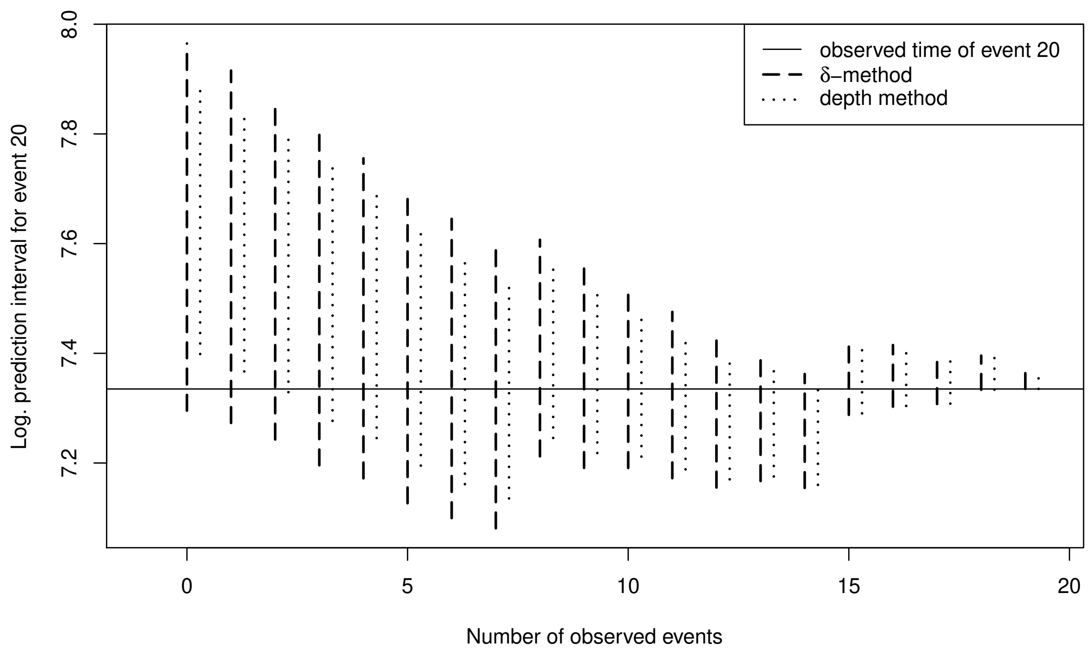

Figure 6.2: 90% Prediction intervals for the 20th break in Experiment SB04 with initial stress of 80 MPa using the results of the other 9 experiments and the breaks $0, 1, 2, \ldots, 19$ of the experiment SB04.

The lefthand side of Figure 6.3 shows the prediction interval for the first breaks in Experiment SB06 with initial stress of 50 MPa using the results of the other 9 experiments with initial stress of 60 to 455 MPa. Unfortunately the Experiment SB06 was stopped after approximately 6 months (marked by the arrow) so that only the first break was observed. This first break lay in the prediction interval. After this, the experiment was continued with a stress of 120 MPa. The righthand side of Figure 6.3 shows the prediction intervals calculated for 120 MPa. As can be seen from this figure, the observed failure times are not lying in the prediction intervals. The failures (breaks) happen earlier than predicted. This means that the first stress of 50 MPa applied to the beam in experiment SB05 in the first half year causes already a damage to the remaining 34 tension failures.

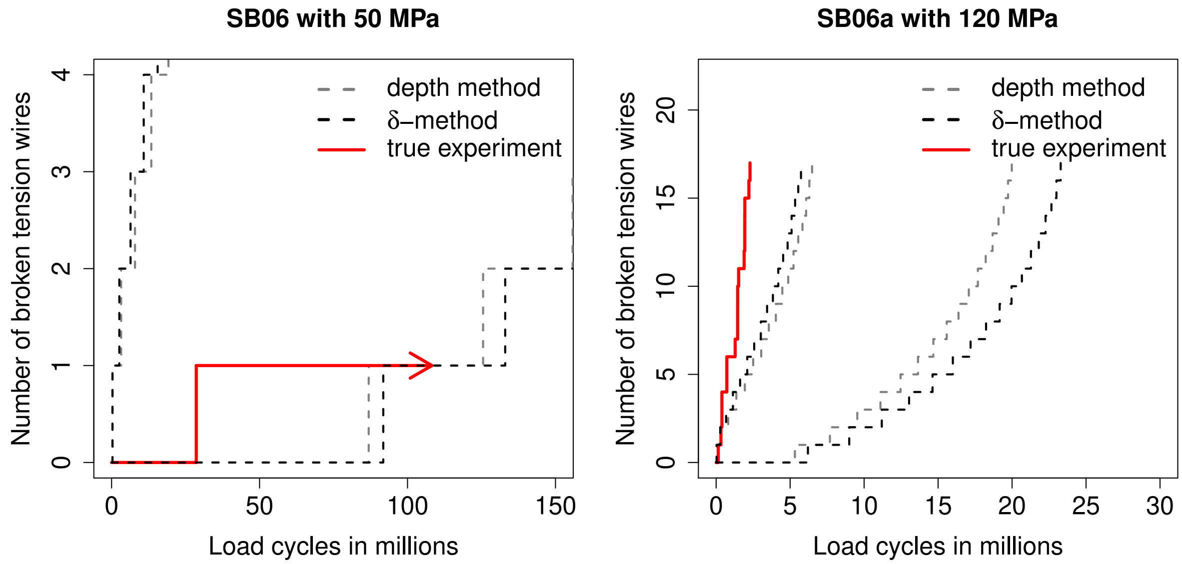

Figure 6.3: Left: Prediction intervals for the first breaks in Experiment SB06 with initial stress of 50 MPa using the results of the other 9 experiments with initial stress of 60 to 455 MPa. Right: Prediction intervals for the first breaks in Experiment SB06a, where Experiment SB06 was continued with stress 120 MPa.

6.7 Load sharing systems with damage accumulation

Again assume that there are J stochastically independent systems where the jth system has I_j components. Moreover, the systems are observed up to different time points τ_j. Then the failure times of the components of the jth system $0 < t_{1,j} < \ldots < t_{I^j,j}$ are realizations of $T_{1,j} < \ldots < T_{I^j,j}$ and thus are realizations of the point process N_j over $[0, \tau_j]$ with $N_j(\tau_j) = I^j$. The point processes N_j are stochastically independent for $j = 1, \ldots, J$. Additionally, the systems are exposed to different initial stress s_j for $j = 1, \ldots, J$.

If the systems are load sharing systems, the following left-continuous conditional intensity function for the j'th system with I_j components and initial stress s_j makes sense:

$$\lambda_j(t) = h_\theta\left(\frac{s_j}{I_j - N_j(t-)}\right), \tag{6.10}$$

where h_θ is an increasing function depending on a parameter vector θ. This means that for t with $N_j(t-) = 0$, i.e. no failure is observed until t, we get $\lambda_j(t) = h_\theta\left(\frac{s_j}{I_j}\right)$, i.e. the initial stress is distributed equally over the I_j components, a reason why this model is called "equal load sharing model". Moreover, for t with $N_j(t-) = \frac{I_j}{2}$, we get $\lambda_j(t) = h_\theta\left(\frac{2\,s_j}{I_j}\right)$. Hence, the stress is doubled for each of the $\frac{I_j}{2}$ components which has not failed.

In particular, model (6.10) means that the conditional intensities between events are constant. Hence the interarrival times (waitung times) $W_{ij} = T_{i,j} - T_{i-1,j}$ have an exponential distribution with parameter $h_\theta\left(\frac{s_j}{I_j-(i-1)}\right)$.

A resonable function h_θ is given by the model of Basquin (1910) who provided a link between the stress σ and the lifetime y by

$$\ln(y) = \theta_1 - \theta_2 \ln(\sigma)$$

with $\theta \in \mathbb{R}$ and $\theta_2 \in [0, \infty$. Since $\left(h_\theta\left(\frac{s_j}{I_j-(i-1)}\right)\right)^{-1}$ is the expected waiting time $E(W_{ij})$, the function h_θ given by

$$h_\theta(x) = \exp(-\theta_1 + \theta_2 \ln(x)) = \exp(-\theta_1) x^{\theta_2}$$

with $\theta = (\theta_1, \theta_2)^\top$ corresponds to the Basquin link and is a link between the waiting time for the next failure and the stress on the non-failed components.

Set

$$a_j(t) := \frac{s_j}{I_j - N_j(t-)} \quad \text{and} \quad a_{ij} := \frac{s_j}{I_j - N_j(t-)} = \frac{s_j}{I_j - (i-1)} \tag{6.11}$$

for $N_j(t-) = i - 1$. In particular, $a_j(t)$ and a_{ij} might be replaced by other types of stress $\tilde{a}_j(N_j(t-))$ and $\tilde{a}_{ij}$ which differ from the equal load sharing model given by (6.11). However, the Basquin link will be kept here. Hence the Basquin load sharing model without damage accumulation is given by

$$\tilde{\lambda}_j^W(t) := \exp(-\theta_1) a_j(t)^{\theta_2}. \tag{6.12}$$

To get a scale invariant estimator in this model, it is approriate to divide the intensity by e.g.

$$\tau := \frac{1}{J} \sum_{j=1}^{J} \tau_j$$

so that one should use

$$\lambda_j^W(t) := \frac{1}{\tau} \exp(-\theta_1) a_j(t)^{\theta_2}. \tag{6.13}$$

One could use also $\frac{1}{\tau_j} \exp(-\theta_1) a_j(t)^{\theta_2}$. However then experiments with short test duration will get too large weights. To extend this model to a load sharing model with damage accumulation, at first not that

$$A_j(t) := \frac{1}{\tau} \int_0^t a_j(x) dx = \frac{1}{\tau} \left(a_j(t) \left(t - \sum_{k=1}^{N_j(t-)} W_{kj} \right) + \sum_{k=1}^{N_j(t-)} a_{kj} W_{kj} \right)$$

accumulates the stress $a_j(x)$ until time t in the sense of load sharing. To avoid the dependence of the intensity function on the mean test duration τ, one can use also a fixed value $\tau_0 > 0$ leading to

$$\tilde{A}_j(t) := \frac{1}{\tau_0} \int_0^t a_j(x) dx = \frac{1}{\tau_0} \left(a_j(t) \left(t - \sum_{k=1}^{N_j(t-)} W_{kj} \right) + \sum_{k=1}^{N_j(t-)} a_{kj} W_{kj} \right).$$

In particular $\tilde{A}_j(t)$ and $A_j(t)$ take into account how long the stress was distributed over the remaining components. Thereby, the factors $\frac{1}{\tau_0}$ and $\frac{1}{\tau}$ are not necessary. However, they prevent too different summands in the load sharing model with additive damage accumulation given below. Moreover, a scale invariant estimator is only possible with τ.

One could use $\tilde{A}_j(t)$ and $A_j(t)$ inside the Basquin link. However, then the pure load sharing model given by (6.12) is not a sepcial case of it. Two models are considered as real extensions of the load sharing model (6.12):
load sharing with multiplicative damage accumulation given by

$$\tilde{\lambda}_j^M(t) := \frac{1}{\tau_0} \exp(-\theta_1) a_j(t)^{\theta_2} \tilde{A}(t)^{\theta_3}, \tag{6.14}$$

and load sharing with additive damage accumulation given by

$$\tilde{\lambda}_j^A(t) := \frac{1}{\tau_0} \exp(-\theta_1) \left(a_j(t) + \theta_3 \tilde{A}(t) \right)^{\theta_2}. \tag{6.15}$$

In both models, the pure load sharing model (6.12) is obtained by setting $\theta_3 = 0$. However, to get scale invariant estimators, it is more appropriate to use

$$\lambda_j^M(t) := \frac{1}{\tau} \exp(-\theta_1) a_j(t)^{\theta_2} A(t)^{\theta_3}, \tag{6.16}$$

and

$$\lambda_j^A(t) :== \frac{1}{\tau} \exp(-\theta_1) \left(a_j(t) + \theta_3 A(t) \right)^{\theta_2} \tag{6.17}$$

instead of the intensity functions given by (6.14) and (6.15), respectively.

6.8 Likelihood function for the load sharing model with damage accumulation

At first we derive the general likelihood function of a point process where $t_1 < t_2 < \ldots$ with $t_i \in< (0, \infty)$ are the realizations of $T_1 < T_2 < \ldots$ of the point process and N given by

$$N(t) := N(t,\omega) := \sum_{i=1}^{\infty} \mathbb{1}_{(0,t]}(t_i)$$

is the realization of the corresponding count process. We follow here the approach for point processes as given by Daley and Vere-Jones (2003). Thereby a count process N is also called a point process, and it is called regular if it has absolute continuous densities on all bounded subsets of $(0, \infty)$. Hence, let N be a regular point process on $[0, \tau]$ for some finite $\tau > 0$, and let $t_1 < t_2 < \ldots < t_{N(\tau)}$ denote a realization of N over $[0, \tau]$. Let $f_i(t|t_1, \ldots, t_{i-1})$ be the conditional density function for an event after the event t_{i-1} and

$$S_i(t|t_1, \ldots, t_{i-1}) := 1 - \int_{t_{i-1}}^{t} f_i(u|t_1, \ldots, t_{i-1})du$$

the associated survival function. The corresponding hazard functions are given by

$$h_i(t|t_1, \ldots, t_{i-1}) := \frac{f_i(t|t_1, \ldots, t_{i-1})}{S_i(t|t_1, \ldots, t_{i-1})}$$

so that

$$f_i(t|t_1, \ldots, t_{i-1}) = h_i(t|t_1, \ldots, t_{i-1}) \exp\left(-\int_{t_{i-1}}^{t} h_i(u|t_1, \ldots, t_{i-1})du\right).$$

The conditional intensity function is then defined by

$$\lambda^*(t) := \begin{cases} h_1(t), & 0 < t \leq t_1, \\ h_i(t|t_1, \ldots, t_{i-1}), & t_{i-1} < t \leq t_i, i \geq 2. \end{cases}$$

Since densities are not unique on subsets with Lebesgue measure equal to zero, they are also not necessarily left-continuous. Therefore, let be $\lambda(t)$ the left-continuous modification of $\lambda^*(t)$, i.e. $\lambda(t) = \lambda^*(t-)$. Then $\lambda(t)_{t>0}$ is a $\mathcal{H}_{t-}$ predictable process where $\mathcal{H}_{t-}$ is the σ-algebra of events at times up to but not including t. Especially, it holds $\lambda(t)dt \approx E[N(dt)|\mathcal{H}_{t-}]$.

Then the likelihood L of $t_1, \ldots, t_{N(\tau)}$ is given by

$$L = \left[\prod_{i=1}^{N(\tau)} \lambda(t_i)\right] \exp\left(-\int_0^{\tau} \lambda(u)du\right), \tag{6.18}$$

see Daley and Vere-Jones (2003) Prop. 7.2.III, p. 232.

For the load sharing system with damage accumulation, we present everything for the factor τ of mean test duration. However, the same holds for a fixed value τ_0 if τ is replaced by τ_0.

Set $W_{(I^j+1)j} := \tau_j - t_{I^j,j}$, $j = 1, \ldots, J$, although it is no waiting time (interarrival time). However, then

$$C_j(0) := 0, \quad C_j(i) := \sum_{k=1}^{i} a_{kj} W_{kj}, \quad i = 1, \ldots, I^j + 1,$$

can be defined as cumulutative stress for $j = 1, \ldots, J$.

6.8.1 Theorem

Let L_M and L_A be the loglikelihood function for the load sharing model with multiplicative and additive damage accumulation given by (6.16) and (6.17), respectively. Then it holds

$$\begin{aligned} &\ln(L_M((\theta_1, \theta_2, \theta_3)^\top)) \qquad (6.19)\\ &= \sum_{j=1}^{J} \Bigg\{ \sum_{i=1}^{I^j} \left[-\theta_1 + \theta_2 \ln(a_{ij}) + \theta_3 \ln\left(\frac{1}{\tau} C_j(i)\right) - \ln(\tau) \right] \\ &\quad - \frac{\exp(-\theta_1)}{\theta_3 + 1} \left[\sum_{i=1}^{I^j+1} a_{ij}^{\theta_2 - 1} \left(\left(\frac{1}{\tau} C_j(i)\right)^{\theta_3+1} - \left(\frac{1}{\tau} C_j(i-1)\right)^{\theta_3+1} \right) \right] \Bigg\} \end{aligned}$$

and

$$\begin{aligned} &\ln(L_A((\theta_1, \theta_2, \theta_3)^\top)) \qquad (6.20)\\ &= \sum_{j=1}^{J} \Bigg\{ \sum_{i=1}^{I^j} \left[-\theta_1 + \theta_2 \ln\left(a_{ij} + \theta_3 \frac{1}{\tau} C_j(i)\right) - \ln(\tau) \right] \\ &\quad - \frac{\exp(-\theta_1)}{\theta_3(\theta_2 + 1)} \left[\sum_{i=1}^{I^j+1} \frac{1}{a_{ij}} \left(\left(a_{ij} + \theta_3 \frac{1}{\tau} C_j(i)\right)^{\theta_2+1} - \left(a_{ij} + \theta_3 \frac{1}{\tau} C_j(i-1)\right)^{\theta_2+1} \right) \right] \Bigg\}. \end{aligned}$$

Proof. According to (6.18), the likelihood function of realizations $t_1, \ldots, t_{N(\tau)}$ of a point process N on $[0, \tau]$ with left-continuous conditional intensity function $\lambda : [0, T] \to \mathbb{R}$ is given by

$$L = \left[\prod_{i=1}^{N(\tau)} \lambda(t_i) \right] \exp\left(- \int_0^\tau \lambda(t) dt \right). \qquad (6.21)$$

At first, we calculate the term $\int_0^\tau \lambda(t)dt$ for the intensity functions of the two load sharing models with damage accumulation. For this, note that it holds for $c > 0$ and $v \geq 0$

$$\int_a^b (ct+d)^v dt = \frac{1}{c(v+1)} (ct+d)^{v+1} \Big|_a^b .$$

Set $t_{0,j} = 0$ for $j = 1, \ldots, J$ and recall $N_j(\tau_j) = I^j$.

For the load sharing model with multiplicative damage accumulation we get

$$\begin{aligned}
\int_0^{\tau_j} \lambda_j^M(t) dt &= \int_0^{\tau_j} \frac{1}{\tau} \exp(-\theta_1) a_j(t)^{\theta_2} A(t)^{\theta_3}\, dt \\
&= \sum_{i=1}^{I^j} \int_{t_{i-1,j}}^{t_{i,j}} \frac{\exp(-\theta_1)}{\tau} a_{ij}^{\theta_2} A(t)^{\theta_3}\, dt + \int_{t_{I^j,j}}^{\tau_j} \frac{\exp(-\theta_1)}{\tau} a_{(I^j+1)j}^{\theta_2} A(t)^{\theta_3}\, dt \\
&= \frac{\exp(-\theta_1)}{\tau} \left[\sum_{i=1}^{I^j} a_{ij}^{\theta_2} \frac{1}{\tau}^{\theta_3} \int_{t_{i-1,j}}^{t_{i,j}} \left(a_{ij} \left(t - \sum_{k=1}^{i-1} W_{kj} \right) + \sum_{k=1}^{i-1} a_{kj} W_{kj} \right)^{\theta_3} dt \right. \\
&\quad + \left. a_{(I^j+1)j}^{\theta_2} \frac{1}{\tau^{\theta_3}} \int_{t_{I^j,j}}^{\tau_j} \left(a_{(I^j+1)j} \left(t - \sum_{k=1}^{I^j} W_{kj} \right) + \sum_{k=1}^{I^j} a_{kj} W_{kj} \right)^{\theta_3} dt \right] \\
&= \frac{\exp(-\theta_1)}{\tau} \left[\sum_{i=1}^{I^j} a_{ij}^{\theta_2} \frac{1}{\tau^{\theta_3}\, a_{ij}(\theta_3+1)} \left(a_{ij} \left(t - \sum_{k=1}^{i-1} W_{kj} \right) + \sum_{k=1}^{i-1} a_{kj} W_{kj} \right)^{\theta_3+1} \Bigg|_{t_{i-1,j}}^{t_{i,j}} \right. \\
&\quad + \left. a_{(I^j+1)j}^{\theta_2} \frac{1}{\tau^{\theta_3}\, a_{(I^j+1)j}(\theta_3+1)} \left(a_{(I^j+1)j} \left(t - \sum_{k=1}^{I^j} W_{kj} \right) + \sum_{k=1}^{I^j} a_{kj} W_{kj} \right)^{\theta_3+1} \Bigg|_{t_{I^j,j}}^{\tau_j} \right] \\
&= \frac{\exp(-\theta_1)}{\tau^{\theta_3+1}(\theta_3+1)} \left[\sum_{i=1}^{I^j} a_{ij}^{\theta_2-1} \left(\left(\sum_{k=1}^{i} a_{kj} W_{kj} \right)^{\theta_3+1} - \left(\sum_{k=1}^{i-1} a_{kj} W_{kj} \right)^{\theta_3+1} \right) \right. \\
&\quad + \left. a_{(I^j+1)j}^{\theta_2-1} \left(\left(a_{(I^j+1)j} \left(\tau_j - \sum_{k=1}^{I^j} W_{kj} \right) + \sum_{k=1}^{I^j} a_{kj} W_{kj} \right)^{\theta_3+1} - \left(\sum_{k=1}^{I^j} a_{kj} W_{kj} \right)^{\theta_3+1} \right) \right] \\
&= \frac{\exp(-\theta_1)}{\tau^{\theta_3+1}(\theta_3+1)} \left[\sum_{i=1}^{I^j+1} a_{ij}^{\theta_2-1} \left(C_j(i)^{\theta_3+1} - C_j(i-1)^{\theta_3+1} \right) \right]
\end{aligned}$$

with $W_{(I^j+1)j} := \tau_j - t_{I^j,j}$ and $C_j(i) = \sum_{k=1}^{i} a_{kj} W_{kj}$ for $i = 1, \ldots, I^j + 1$.

Fo the load sharing model with additive damage accumulation, it holds

$$
\begin{aligned}
&\int_0^{\tau_j} \lambda_j^A(t)dt = \int_0^{\tau_j} \frac{1}{\tau}\exp(-\theta_1)\left(a_j(t)+\theta_3 A(t)\right)^{\theta_2}\, dt\\
=&\ \sum_{i=1}^{I^j}\int_{t_{i-1,j}}^{t_{i,j}} \frac{\exp(-\theta_1)}{\tau}\left(a_{ij}+\theta_3 A(t)\right)^{\theta_2}\, dt + \int_{t_{I^j,j}}^{\tau_j} \frac{\exp(-\theta_1)}{\tau}\left(a_{(I^j+1)j}+\theta_3 A(t)\right)^{\theta_2}\, dt\\
=&\ \frac{\exp(-\theta_1)}{\tau}\left[\sum_{i=1}^{I^j}\int_{t_{i-1,j}}^{t_{i,j}}\left(a_{ij}+\theta_3\frac{1}{\tau}\left[a_{ij}\left(t-\sum_{k=1}^{i-1}W_{kj}\right)+\sum_{k=1}^{i-1}a_{kj}W_{kj}\right]\right)^{\theta_2} dt\right.\\
&\left.\quad + \int_{t_{I^j,j}}^{\tau_j}\left(a_{(I^j+1)j}+\theta_3\frac{1}{\tau}\left[a_{(I^j+1)j}\left(t-\sum_{k=1}^{I^j}W_{kj}\right)+\sum_{k=1}^{I^j}a_{kj}W_{kj}\right]\right)^{\theta_2} dt\right]\\
=&\ \frac{\exp(-\theta_1)\,\tau}{\tau\,\theta_3(\theta_2+1)}\left[\sum_{i=1}^{I^j}\frac{1}{a_{ij}}\left(a_{ij}+\theta_3\frac{1}{\tau}\left[a_{ij}\left(t-\sum_{k=1}^{i-1}W_{kj}\right)+\sum_{k=1}^{i-1}a_{kj}W_{kj}\right]\right)^{\theta_2+1}\Bigg|_{t_{i-1,j}}^{t_{i,j}}\right.\\
&\left.\quad + \frac{1}{a_{(I^j+1)j}}\left(a_{(I^j+1)j}+\theta_3\frac{1}{\tau}\left[a_{(I^j+1)j}\left(t-\sum_{k=1}^{I^j}W_{kj}\right)+\sum_{k=1}^{I^j}a_{kj}W_{kj}\right]\right)^{\theta_2+1}\Bigg|_{t_{I^j,j}}^{\tau_j}\right]\\
=&\ \frac{\exp(-\theta_1)}{\theta_3(\theta_2+1)}\left[\sum_{i=1}^{I^j+1}\frac{1}{a_{ij}}\left(\left(a_{ij}+\theta_3\frac{1}{\tau}C_j(i)\right)^{\theta_2+1}-\left(a_{ij}+\theta_3\frac{1}{\tau}C_j(i-1)\right)^{\theta_2+1}\right)\right]
\end{aligned}
$$

with $W_{(I^j+1)j} := \tau_j - t_{I^j,j}$ and $C_j(i) = \sum_{k=1}^{i} a_{kj}W_{kj}$ for $i = 1,\dots,I^j+1$.

To calculate the likelihood function, note that $A(t_{i,j}) = \frac{1}{\tau}C_j(i)$ holds for $i = 1,\dots,I^j$. Hence for the load sharing model with multiplicative damage accumulation, we get

$$\lambda_j^M(t_{i,j}) = \frac{1}{\tau}\exp(-\theta_1)a_{ij}^{\theta_2}\left(\frac{1}{\tau}C_j(i)\right)^{\theta_3},$$

and for the load sharing model with additive damage accumulation, we have

$$\lambda_j^A(t_{i,j}) = \frac{1}{\tau}\exp(-\theta_1)\left(a_{ij}+\theta_3\frac{1}{\tau}C_j(i)\right)^{\theta_2}.$$

This completes the proof using the form (6.21) for a likelihood function of a point process and using the fact that the point processes from the J systems are stochastically independent. □

6.8.2 Corollary

Let $L_D = L_M$ or $L_D = L_A$, respectively, be the likelihood function for the load sharing model with multiplicative or additive damage accumulation and $\widehat{\theta}_D \in \mathbb{R}^3$ the corresponding maximum likelihood estimate, let L_W be the likelihood function for the load sharing model without damage accumulation and $\widehat{\theta}_W \in \mathbb{R}^2$ the corresponding maximum likelihood estimate, and let $\chi^2_{1,1-\alpha}$ be the $(1-\alpha)$ quantile of the χ^2 distribution with one degree of freedom. Then the decision rule

$$\textit{reject } H_0 : \theta_3 = 0 \textit{ if } -2\left(\ln(L_W(\widehat{\theta}_W)) - \ln(L_D(\widehat{\theta}_D))\right) > \chi^2_{1,1-\alpha},$$

provides an asymptotically α *level test for* $H_0 : \theta_3 = 0$.

Table 6.1 provides p-values of the likelihood ratio test for the effect θ_3 for damage accumulation. It shows that the effect θ_3 indeed differs significantly from 0. However, it seems that there is a problem with the scale dependent version for the additive damage accumulation.

Table 6.1: P-values of the likelihood ratio tests given by Corollary 6.8.2 based on the scale dependent ML estimators and the scale invariant ML estimators using $\tau_0 = 1\,000$ for the scale dependent version.

Model	scale dependent ML	scale invariant ML
multiplicative damage accumulation	2.46e-06	2.46e-06
additive damage accumulation	1	1.39e-04.

6.9 Systems with repair

This section deals with systems where components which are failed are substituted by a new one. Thereby, it does not matter how many components the system has. In particular, the system can consist of only one component so that the failure of the component is also the failure of the system. The simplest case is a system where the repair happens immediately after failure. Then the time points $0 = T_0 < T_1 < T_2 < \ldots$ of failures of the components of a system with immediate repair can be modelled by a renewal process.

6.9.1 Definition
The point process given by the event times $0 = T_0 < T_1 < T_2 < \ldots$ is called a renewal process if the waiting times between the event times given by $W_i = T_{i+1} - T_i$ for $i \in \mathbb{N}_0$ are independent and identically distributed.

According to Theorem 6.3.3, the Poisson process is a special renewal process, namely a renewal process where W_i has an exponential distribution.

Define again the corresponding counting process $N = (N_t)_{t\geq 0} = (N(t))_{t\geq 0}$ by

$$N(t) := \sum_{i=1}^{\infty} \mathbb{1}_{[0,t]}(T_i).$$

6.9.2 Definition
The function H given by $H(t) = E(N(t))$ (the expected number of failures up to time t) is called the renewal function.

6.9.3 Theorem (See Kahle and Liebscher 2013, p. 57)
The renewal function H is the unique solution of the integral equation

$$H(t) = F(t) + \int_0^t H(t-u)\, f(u)\, du \tag{6.22}$$

if W_i has cumulative distribution function F and density f.

Proof. Set

$$F^{*i}(t) := P(T_i \leq t), \quad f^{*i}(t) := \frac{\partial}{\partial t} F^{*i}(t).$$

Since $T_i = T_{i-1} + W_{i-1}$ and T_{i-1} and W_{i-1} are independent, it holds

$$f^{*i}(s) = \int_{-\infty}^{\infty} f^{*(i-1)}(s-u)\, f(u)\, du = \int_0^s f^{*(i-1)}(s-u)\, f(u)\, du$$

so that with Fubini's theorem

$$
\begin{aligned}
F^{*i}(t) = \int_0^t f^{*i}(s)\,ds = \int_{-\infty}^t \int_{-\infty}^\infty f^{*(i-1)}(s-u)\,f(u)\,du\,ds \\
= \int_{-\infty}^\infty \int_{-\infty}^t f^{*(i-1)}(s-u)\,ds\,f(u)\,du = \int_{-\infty}^\infty \int_{-\infty}^{t-u} f^{*(i-1)}(s)\,ds\,f(u)\,du \\
= \int_{-\infty}^\infty F^{*(i-1)}(t-u)\,f(u)\,du = \int_0^t F^{*(i-1)}(t-u)\,f(u)\,du.
\end{aligned}
$$

This implies with $F^{*1}(t) = P(T_1 \le t) = P(W_0 \le t) = F(t)$ and

$$
H(t) = E(N(t)) = E\left(\sum_{i=1}^\infty \mathbb{1}_{[0,t]}(T_i)\right) = \sum_{i=1}^\infty P(T_i \le t) = \sum_{i=1}^\infty F^{*i}(t)
$$

the integral equation

$$
\begin{aligned}
H(t) = F(t) + \sum_{i=2}^\infty \int_0^t F^{*(i-1)}(t-u)\,f(u)\,du \\
= F(t) + \int_0^t \sum_{i=2}^\infty F^{*(i-1)}(t-u)\,f(u)\,du = F(t) + \int_0^t \sum_{i=1}^\infty F^{*i}(t-u)\,f(u)\,du \\
= F(t) + \int_0^t H(t-u)\,f(u)\,du.
\end{aligned}
$$

Thereby, the exchange of the infinite sum and the integral is possible according to the monotone convergence theorem of Henri Lebesgue and Beppo Levi since all integrands are nonnegative. For the uniqueness of H as solution of the integral equation (6.22) see Kahle and Liebscher (2013). □

6.9.4 Lemma
If $W_i \sim \mathcal{E}(\lambda)$ for $i \in \mathbb{N}_0$ then $H(t) = \lambda\,t$ for $t \ge 0$.

Proof. We start with the righthand side of the integral equation (6.22) and get with partial integration

$$
\begin{aligned}
F(t) + \int_0^t H(t-u)\,f(u)\,du = 1 - e^{-\lambda t} + \int_0^t \lambda\,(t-u)\,\lambda\,e^{-\lambda\,u}\,du \\
= 1 - e^{-\lambda t} + \lambda\,t \int_0^t \lambda\,e^{-\lambda\,u}\,du - \lambda^2 \int_0^t u\,e^{-\lambda\,u}\,du \\
= 1 - e^{-\lambda t} + \lambda\,t\left(1 - e^{-\lambda t}\right) - \lambda^2\left[u\,\frac{-1}{\lambda}\,e^{-\lambda\,u}\Big|_0^t - \int_0^t \frac{-1}{\lambda}\,e^{-\lambda\,u}\,du\right] \\
= 1 - e^{-\lambda t} + \lambda\,t - \lambda\,t\,e^{-\lambda t} + \lambda\,t\,e^{-\lambda\,t} - \left(1 - e^{-\lambda t}\right) \\
= \lambda\,t = H(t).
\end{aligned}
$$

Hence H is a solution of (6.22). Since the solution is unique, H is the renewal function according to Theorem 6.9.3. □

If W_i for $i = 1, \ldots, I$ is distributed according to a distribution with unknown parameter θ then estimators and confidence sets for θ can be obtained as in Chapter 2. If the repair times $W_{i,j}$ for $i = 1, \ldots, I_j$ of several systems $j = 1, \ldots, J$ are observed under possibly different stress levels $s_1, \ldots, s_J$ and the distribution of $W_{i,j}$ is given by a link function $g_\theta(s_j)$ so that θ is the only unknown parameter then estimators and confidence sets for θ can be obtained as in Chapter 3.

For further results about systems with repair see Kahle and Liebscher (2013).

Chapter 7

Bayesian inference

7.1 Foundations

The idea of Bayesian estimation is, unlike the frequentist point of view, that θ is not a fixed, but unknown, value, but a random variable with an unknown distribution to be estimated. If some (expert) knowledge about this distribution is given, this goes in for the so-called **prior distribution** $p(\theta)$. The estimated distribution $p(\theta|Y_1, ..., Y_N)$ is called **posterior distribution**. It can be calculated by the following theorem.

7.1.1 Theorem (Theorem of Bayes)
Let θ be a random variable with prior distribution $p(\theta)$. It is

$$p(\theta|Y_1, ..., Y_N) = \frac{p(Y_1, ..., Y_N|\theta)p(\theta)}{p(Y_1, ..., Y_N)} \tag{7.1}$$

$$\propto p(Y_1, ..., Y_N|\theta)p(\theta),$$

where $\propto$ means proportional up to a constant.

In many cases, the posterior distribution $p(\theta|Y_1, ..., Y_N)$ cannot be calculated explicitly. But in some cases, a calculation is possible and the posterior distribution belongs to a known distribution family.

7.1.2 Definition (Conjugate prior)
If $p(\theta|Y_1, ..., Y_N)$ is analytically available and belongs to the same distribution family as $p(\theta)$, this prior distribution is called conjugate for the likelihood $p(Y_1, ..., Y_N|\theta)$.

7.1.3 Example
Let $Y_1, ..., Y_N$ be independent and identical $\mathcal{E}(\theta)$ distributed. Then, $\theta \sim \mathcal{G}(\alpha, \beta)$ is conjugate to

the likelihood. This can be seen as follows:

$$p(\theta|Y_1,...,Y_N)\cdot p(Y_1,...,Y_N|\theta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\beta\theta)\cdot\prod_{n=1}^N\theta\cdot\exp(-\theta Y_n)$$

$$\propto\theta^{\alpha+N-1}\exp(-(\beta+\sum_{n=1}^N Y_n)\cdot\theta).$$

The posterior distribution is given by $\theta|Y_1,...,Y_N\sim\mathcal{G}(\alpha+N,\beta+\sum_{n=1}^N Y_n)$.

In many cases, the posterior distribution cannot be calculated. In these cases, it is simulated through a sampling algorithm. Very popular is the class of Markov chain Monte Carlo (MCMC) algorithms.

For a low dimensional vector, the following algorithm is suitable.

7.1.4 Algorithm (Metropolis Hastings (MH))
We want to sample from the posterior distribution

$$p(\theta|Y_1,...,Y_N)\propto p(Y_1,...,Y_N|\theta)\cdot p(\theta).$$

At first, choose a proper proposal density $q(\theta^*|\theta)$. A minimal necessary condition is that

$$\bigcup_{\theta\in supp\ p(\cdot|Y_1,...,Y_N)} supp\ q(\cdot|\theta)\supset supp\ p(\cdot|Y_1,...,Y_N)$$

with supp denoting the support of a function. In words, each point of the posterior's support has to be available by the Markov chain based on that proposal density. This would also be true for the density of the dirac measure at θ^*. Therefore, this is only the minimal necessary condition, the proposal density has to be chosen wisely.

At second, choose a starting value θ_0 with $p(\theta_0|Y_1,...,Y_N)>0$. For $k=1,...,K$ draw $\theta^*\sim q(\cdot|\theta_{k-1})$ and set $\theta_k=\theta^*$ with acceptance probability

$$\rho(\theta^*,\theta_{k-1})=\min\left\{\frac{p(Y_1,...,Y_N|\theta^*)\cdot p(\theta^*)}{p(Y_1,...,Y_N|\theta_{k-1})\cdot p(\theta_{k-1})}\cdot\frac{q(\theta_{k-1}|\theta^*)}{q(\theta^*|\theta_{k-1})},\ 1\right\}$$

and $\theta_k^*=\theta_{k-1}$ with probability $1-\rho(\theta^*,\theta_{k-1})$. For further details see Robert and Casella (2004).

7.1.5 Lemma
The stationary distribution of the Markov chain $\{\theta_k,\ k=1,...,K\}$ resulting from the MH algorithm is equal to the posterior $p(\theta|Y_1,...,Y_N)$.

7.1.6 Lemma
If the detailed balance condition

$$p(\theta_k|\theta_{k-1})p(\theta_{k-1}|Y_1,...,Y_N)=p(\theta_{k-1}|\theta_k)p(\theta_k|Y_1,...,Y_N)$$

is satisfied, the chain has stationary distribution $p(\theta|Y_{(n)})$. For the proof, see Robert and Casella (2004), p. 230, Definition 6.45 and Theorem 6.46.

Proof of Lemma 7.1.5
We have a Markov chain with transition density $p(\theta_k|\theta_{k-1}) = \rho(\theta_k, \theta_{k-1}) \cdot q(\theta_k|\theta_{k-1})$.
It is

$$\begin{aligned}
p(\theta_k|\theta_{k-1})p(\theta_{k-1}|Y_1, ..., Y_N) =&\rho(\theta_k, \theta_{k-1})q(\theta_k|\theta_{k-1})p(\theta_{k-1}|Y_1, ..., Y_N)\\
=&\frac{p(\theta_k|Y_1, ..., Y_N)}{p(\theta_{k-1}|Y_1, ..., Y_N)} \cdot \frac{q(\theta_{k-1}|\theta_k)}{q(\theta_k|\theta_{k-1})} \cdot q(\theta_k|\theta_{k-1})p(\theta_{k-1}|Y_1, ..., Y_N)\\
=&p(\theta_k|Y_1, ..., Y_N)q(\theta_{k-1}|\theta_k).
\end{aligned}$$

This means, the detailed balance condition is satisfied and with Lemma 7.1.6 the assumption holds. Compare Theorem 7.2 on p. 272 in Robert and Casella (2004).

7.1.7 Remark

(i) In the special case of a symmetric proposal density, i.e. $q(\theta^*|\theta_{k-1}) = q(\theta_{k-1}|\theta^*)$ the acceptance probability reduces to

$$\rho(\theta^*, \theta_{k-1}) = \min\left\{\frac{p(Y_1, ..., Y_N|\theta^*) \cdot p(\theta^*)}{p(Y_1, ..., Y_N|\theta_{k-1}) \cdot p(\theta_{k-1})},\ 1\right\}.$$

One example is the density of the $\mathcal{N}(\theta_{k-1}, sd^2)$ distribution with mean equal to the last iteration step θ_{k-1} and a fixed standard deviation sd. This is also known as symmetric random walk, see Robert and Casella (2004) p. 206.

(ii) Crucial step of the algorithm is the choice of the proposal density. Theoretically, the assumption given above suffices. But in practice, a good approximation of the posterior density is of interest. If the acceptance probability $\rho(\theta^*, \theta_{k-1})$ is large, many samples are accepted, but the new one only moves little from the old iteration step and the support of the distribution will be filled slowly. On the other hand, if the acceptance probability is small, only few candidates are accepted, which is also not suitable for a continuous distribution. Rosenthal (2011) calculated an optimal acceptance rate of the MH algorithm, which is 0.234. To reach this, an adaptive algorithm might be a solution, also presented in Rosenthal (2011).

(iii) The number of iterations, before the chain has reached the stationary distribution, is called **burn-in phase**, which is dependent on the starting value and the proposal density. To simulate independence of the samples, in many cases, not every chain iteration is taken as random sample of the posterior, because they are very dependent. In particular, if the mixing is bad or the acceptance rate small, this leads to a bad approximation of the posterior. In this case, the chain is thinned, which means, a certain amount of iterations are skipped, called **thinning rate**.

7.1.8 Definition
For parameter vector $\theta = (\theta_1, ..., \theta_d)$, in some cases, a **full conditional posterior** *density $p(\theta_j|Y_1, ..., Y_N, \theta_1, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_d)$ can be calculated.*

7.1.9 Algorithm (Gibbs sampler)
For a high dimensional parameter vector $\theta = (\theta_1, ..., \theta_d)$, where each $\theta_j, j = 1, .., d$, can be also a

vector itself, the idea is to sample iteratively from the full conditional posterior distributions of the components $p(\theta_j|Y_1,...,Y_N,\theta_1,...,\theta_{j-1},\theta_{j+1},...,\theta_d), j=1,...,d$.

Choose starting values $\theta_{2,0},...,\theta_{d,0}$ *and for* $k=1,...,K$ *draw*

$$\begin{aligned}
\theta_{1,k} &\sim p(\theta_1|Y_1,...,Y_N,\theta_{2,k-1},...,\theta_{d,k-1}),\\
\theta_{2,k} &\sim p(\theta_2|Y_1,...,Y_N,\theta_{1,k},\theta_{3,k-1},...,\theta_{d,k-1}),\\
&\vdots\\
\theta_{d-1,k} &\sim p(\theta_{d-1}|Y_1,...,Y_N,\theta_{1,k},...,\theta_{d-2,k},\theta_{d,k-1}),\\
\theta_{d,k} &\sim p(\theta_d|Y_1,...,Y_N,\theta_{1,k},...,\theta_{d-1,k}).
\end{aligned}$$

7.1.10 Lemma
The resulting Markov chain $\{(\theta_{1,k},...,\theta_{d,k}),\ k=1,...,K\}=\{\theta_k,\ k=1,...,K\}$ *of the Gibbs sampler has stationary distribution* $p(\theta|Y_1,...,Y_N)$*, see for the proof Robert and Casella (2004) p. 372.*

7.1.11 Algorithm (Metropolis-within-Gibbs sampler)
In the case of not explicitly available full conditional posteriors, one step of the Gibbs sampler can be conducted by a MH algorithm. The original work of Metropolis et al. (1953) introduced what we now call Metropolis within Gibbs algorithm. We restrict here to the case of only one component $\theta_j, j\in\{1,...,d\}$ *to be sampled by an MH step. Of course, this can be done for several components. In addition, for notation simplicity, we assume the* j*th component to be independent from the others in their prior distribution. This means* $p(\theta)=p(\theta_j)p(\theta_{-j})$*. We choose a proper proposal density* $q(\cdot|\theta_{1,k},...,\theta_{j-1,k},\theta_{j,k-1},...,\theta_{d,k-1})$ *for* θ_j *similar to the MH algorithm and starting values* $\theta_{2,0},...,\theta_{d,0}$*, if* $j=1$ *also* $\theta_{1,0}$*. For* $k=1,...,K$ *draw*

$$\begin{aligned}
\theta_{1,k} &\sim p(\theta_1|Y_1,...,Y_N,\theta_{2,k-1},...,\theta_{d,k-1}),\\
&\vdots\\
\theta_{j-1,k} &\sim p(\theta_{j-1}|Y_1,...,Y_N,\theta_{1,k},...,\theta_{j-2,k},\theta_{j,k-1},...,\theta_{d,k-1}),
\end{aligned}$$

$\theta_j^* \sim q(\theta_j|\theta_{1,k},...,\theta_{j-1,k},\theta_{j,k-1},...,\theta_{d,k-1})$ *and accept* $\theta_{j,k}=\theta_j^*$ *with probability*

$$\rho(\theta_j^*,\theta_{j,k-1}) = \min\left\{1,\ \frac{p(Y_1,...,Y_N|\theta_{1,k},...,\theta_{j-1,k},\theta_j^*,\theta_{j+1,k-1},...,\theta_{d,k-1})\cdot p(\theta_j^*)}{p(Y_1,...,Y_N|\theta_{1,k},...,\theta_{j-1,k},\theta_{j,k-1},...,\theta_{d,k-1})\cdot p(\theta_{j,k-1})}\right.$$
$$\left.\cdot\frac{q(\theta_{j,k-1}|\theta_{1,k},...,\theta_{j-1,k},\theta_j^*,\theta_{j+1,k-1},...,\theta_{d,k-1})}{q(\theta_j^*|\theta_{1,k},...,\theta_{j-1,k},\theta_{j,k-1},...,\theta_{d,k-1})}\right\}$$

and $\theta_{j,k}=\theta_{j,k-1}$ *with probability* $1-\rho(\theta_j^*,\theta_{j,k-1})$,

$$\begin{aligned}
\theta_{j+1,k} &\sim p(\theta_{j+1}|Y_1,...,Y_N,\theta_{1,k},...,\theta_{j,k},\theta_{j+2,k-1},...,\theta_{d,k-1}),\\
&\vdots\\
\theta_{d,k} &\sim p(\theta_d|Y_1,...,Y_N,\theta_{1,k},...,\theta_{d-1,k}).
\end{aligned}$$

See for further details Robert and Casella (2004), Section 10.3.3, p. 392.

7.2 Bayesian Prediction

7.2.1 Definition (Predictive distribution)
The predictive distribution of Y^*, given $Y_1, ..., Y_N$, is given by

$$p(Y^*|Y_1, ..., Y_N).$$

7.2.2 Lemma
Let $Y_1, ..., Y_N$ be independent and identical distributed. The predictive distribution of Y^*, independent and identical distributed as $Y_1, ..., Y_N$, is given by

$$p(Y^*|Y_1, ..., Y_N) = \int p(Y^*|\theta)p(\theta|Y_1, ..., Y_N)\, d\theta.$$

In addition, let $Y_1, ..., Y_N$ stem from a Markov process seen in Section 5, and Y^* be the observation variable in $t^* > t_N$. Then

$$p(Y^*|Y_1, ..., Y_N) = \int p(Y^*|Y_N, \theta)p(\theta|Y_1, ..., Y_N)\, d\theta.$$

Proof In the first case, Y^* and $Y_1, ..., Y_N$ are independent. Therefore, it is

$$\begin{aligned}
&\int p(Y^*|\theta)p(\theta|Y_1, ..., Y_N)\, d\theta = \int p(Y^*|\theta, Y_1, ..., Y_N)p(\theta|Y_1, ..., Y_N)\, d\theta\\
&= \int \frac{p(Y^*, \theta, Y_1, ..., Y_N)}{p(\theta, Y_1, ..., Y_N)} \frac{p(\theta, Y_1, ..., Y_N)}{p(Y_1, ..., Y_N)}\, d\theta = \int \frac{p(Y^*, \theta, Y_1, ..., Y_N)}{p(Y_1, ..., Y_N)}\, d\theta\\
&= \int p(Y^*, \theta|Y_1, ..., Y_N)\, d\theta = p(Y^*|Y_1, ..., Y_N).
\end{aligned}$$

The second case follows analogously.

In many cases, the predictive distribution can not be calculated explicitly. It is common practice to approximate $p(Y^*|Y_1, ..., Y_N)$ by

$$p(Y^*|Y_1, ..., Y_N) = \int p(Y^*|\theta)p(\theta|Y_1, ..., Y_N)\, d\theta \approx \frac{1}{K}\sum_{k=1}^{K} p(Y^*|\theta_k). \tag{7.2}$$

This can be seen by the approximation of the posterior density

$$p(\theta|Y_1, ..., Y_N) \approx \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\{\theta = \theta_k\}.$$

7.2.3 Algorithm (Inversion sampling method)
Beside MCMC sampling methods there is another possibility to draw random samples from a continuous distribution with distribution function $F(y) = \int_{-\infty}^{y} p(z)\, dz$. For $U \sim \mathcal{U}(0, 1)$, the random variable $F^{-1}(U)$ has the distribution of interest. In many cases, this inverse function F^{-1} is not calculable. However, there is one possibility to fix an interval $[y_l, y_u]$ and to choose a vector of points $y_l = y_1 < y_2 < ... < y_C = y_u$. Then, one calculates $F(y_1), ..., F(y_C)$ and for a realization u from the uniform distribution, $\min\{y \in \{y_1, ..., y_C\}|F(y) \geq u\}$ can be seen as a sample from F.

The inversion method can be very suitable for the predictive distribution, if it is not analytically available and has to be approximated by (7.2).

7.3 Bayesian prediction of crack growth with non-linear regression models

Remember Section 5.2, where non-linear models are introduced. We here consider a general regression function $f(\theta, x)$. This means, we consider the Bayes model

$$\begin{aligned}
Y_n &= f(\theta, x_n) + E_n,\\
E_n &\sim \mathcal{N}(0, \sigma^2),\ n = 1, ..., N,\\
\theta &\sim p(\theta),\\
\frac{1}{\sigma^2} &\sim \mathcal{G}(\alpha, \beta).
\end{aligned}$$

For example, model (5.7) would be nested with $f(\theta, x) = \theta_0 + \theta_1 x^{\theta_2}$.

7.4 Bayesian prediction for the state dependent point process

Remember the state dependent point process from Chapter 6

$$W_{i,j} \sim \text{Exp}(\lambda_\theta(i, s_j)), \;\; i = 0, \ldots, I_j < I_{max}, j = 1, ..., J,$$

stochastically independent and

$$\lambda_\theta(i, s) := h_\theta\left(s \cdot \frac{I_{max}}{I_{max} - i}\right)$$

for some function h_θ. The likelihood is given by

$$p(\{W_{i,j}\}_{i=0,\ldots,I_j<I_{max},j=1,\ldots,J}|\theta) = \prod_{j=1}^{J}\prod_{i=0}^{I_j} \lambda_\theta(i, s_j) \exp\left(\lambda_\theta(i, s_j) W_{i,j}\right).$$

– references:

Robert, Christian, and George Casella. Monte Carlo statistical methods. Springer Science & Business Media, 2004.

Rosenthal, Jeffrey S. "Optimal proposal distributions and adaptive MCMC." Handbook of Markov Chain Monte Carlo (2011): 93-112.

Carlin, Bradley P., and Thomas A. Louis. Bayesian methods for data analysis. CRC Press, 2009.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). "Equation of State Calculations by Fast Computing Machines". The Journal of Chemical Physics 21, pp. 1087–1092.

Chapter 8

Experimental design

In accelerated lifetime experiments as treated in Chapter 3, the interesting life times are for low stress levels. Since experiments at low stress levels last long, most experiments are done under much higher stress level than are of interest. However, observations at too high stress levels are less informative for the main aim. The question is how many observations should be made at low stress levels and how many at high stress levels to obtain the best estimate or best prediction interval of the lifetime at certain low stress level.

We consider here lifetime experiments with exponential distribution as treated in Section 3.3. The problem is to construct optimal designs of the stress levels for the maximum likelihood estimator of θ

$$\widehat{\theta} := \arg\max_{\theta} L_\theta(y_*, d_*, s_*)$$

and for maximum likeihood estimators $\varphi(\widehat{\theta})$ of the nonlinear aspects $\varphi(\theta) = \lambda_\theta^{-1}(k)$ which provide the maximum stress level, where the expected life time is greater than $1/k$. Section 8.1 deals then with the optimal designs for λ_θ given by $\lambda_\theta(s) = \theta s$ and Section 8.2 with the locally D-optimal designs for λ_θ given by $\lambda_\theta(s) = \exp(\theta_0 + \theta_1 s)$.

8.1 Optimal Designs if $\lambda_\theta(s) = \theta s$

Here the information is

$$I_\theta(\delta) = \int \frac{1}{\theta^2}\left(1 - e^{-\theta sc}\right)\delta(ds).$$

Since $1 - e^{-\theta sc}$ is strictly increasing in s, the information is maximized if the design puts all its mass on the largest possible value for the stress, i.e. the optimal design on a design region $\mathcal{S} = [S_l, S_u]$ uses only the upper value S_u. However, as soon as there is no censoring, i.e. $c = \infty$, then it does not matter which stress levels are used.

8.2 D-Optimal Designs if $\lambda_\theta(s) = \exp(\theta_0 + \theta_1 s)$

At first we consider here the existence of a maximum likelihood estimators if $\lambda_\theta(s) = \exp(\theta_0+\theta_1 s)$ is the link funktion.

8.2.1 Theorem
The maximum likelihood estimator exists at (y_*, d_*, s_*) and is unique if and only if there exists at least one uncensored observation and $n \neq m$ with $s_n \neq s_m$.

. **Proof.** A necessary condition for the maximum likelihood estimator $\widehat{\theta} = \widehat{\theta}(y_*, d_*, s_*)$ is

$$\begin{aligned}
0 &= \sum_{n=1}^N \dot{l}(\widehat{\theta}, t_n, s_n)\\
&= \sum_{n=1}^N \frac{\partial}{\partial\theta}\lambda_\theta(s_n)\Big|_{\theta=\widehat{\theta}} \left[\left(\frac{1}{\lambda_{\widehat{\theta}}(s_n)} - t_n\right) 1_{[0,c]}(t_n) - c\,1_{(c,\infty)}(t_n)\right]\\
&= \sum_{t_n \leq c} \left(1 - \exp(\widehat{\theta}_0 + \widehat{\theta}_1\, s_n)\, t_n\right) \begin{pmatrix} 1 \\ s_n \end{pmatrix} - \sum_{t_n > c} \exp(\widehat{\theta}_0 + \widehat{\theta}_1\, s_n)\, c \begin{pmatrix} 1 \\ s_n \end{pmatrix}.
\end{aligned}$$

In particular, we have

$$\begin{pmatrix} \sharp\{n;\ t_n \leq c\} \\ \sum_{t_n \leq c} s_n \end{pmatrix} = \sum_{t_n \leq c} \exp(\widehat{\theta}_0 + \widehat{\theta}_1\, s_n)\, t_n \begin{pmatrix} 1 \\ s_n \end{pmatrix} + \sum_{t_n > c} \exp(\widehat{\theta}_0 + \widehat{\theta}_1\, s_n)\, c \begin{pmatrix} 1 \\ s_n \end{pmatrix}$$

so that at least one observation must be uncensored. The second derivative at $\widehat{\theta}$ is

$$\begin{aligned}
&\sum_{n=1}^N \ddot{l}(\widehat{\theta}, t_n, s_n)\\
&= \sum_{t_n \leq c} - \exp(\widehat{\theta}_0 + \widehat{\theta}_1\, s_n)\, t_n \begin{pmatrix} 1 \\ s_n \end{pmatrix} (1, s_n) - \sum_{t_n > c} \exp(\widehat{\theta}_0 + \widehat{\theta}_1\, s_n)\, c \begin{pmatrix} 1 \\ s_n \end{pmatrix} (1, s_n)\\
&= -\begin{pmatrix} \sum_{n=1}^N a_n & \sum_{n=1}^N a_n\, s_n \\ \sum_{n=1}^N a_n\, s_n & \sum_{n=1}^N a_n\, s_n^2 \end{pmatrix}
\end{aligned}$$

with

$$a_n = \exp(\widehat{\theta}_0 + \widehat{\theta}_1\, s_n)\, t_n\, 1_{[0,c]}(t_n) + \exp(\widehat{\theta}_0 + \widehat{\theta}_1\, s_n)\, c\, 1_{(c,\infty)}(t_n).$$

The Hölder inequality provides with $P(\{s_n\}) = p_n = \frac{a_n}{\sum_{n=1}^N a_n}$

$$\begin{aligned}
\frac{\sum_{n=1}^N a_n\, s_n}{\sum_{n=1}^N a_n} &= \sum_{n=1}^N s_n p_n = \int 1 \cdot s\, dP \leq \sqrt{\int 1^2\, dP \int s^2\, dP}\\
&= \sqrt{(\sum_{n=1}^N p_n)\ (\sum_{n=1}^N s_n^2\, p_n)} = \sqrt{\frac{\sum_{n=1}^N a_n}{\sum_{n=1}^N a_n}\ \frac{\sum_{n=1}^N a_n\, s_n^2}{\sum_{n=1}^N a_n}}
\end{aligned}$$

so that

$$\left(\sum_{n=1}^{N} a_n\, s_n\right)^2 \leq \left(\sum_{n=1}^{N} a_n\right) \left(\sum_{n=1}^{N} a_n\, s_n^2\right).$$

This means

$$\det\left(\sum_{n=1}^{N} \ddot{l}(\widehat{\theta}, t_n, s_n)\right) \leq 0$$

and $\sum_{n=1}^{N} \ddot{l}(\widehat{\theta}, t_n, s_n)$ is negative definite if and only if there exists $n \neq m$ with $s_n \neq s_m$. □

Setting

$$x_\theta(s) := \sqrt{1 - e^{-\exp(\theta_0+\theta_1\, s)\, c}} \begin{pmatrix} 1 \\ s \end{pmatrix} = \sqrt{1 - e^{-k\,\exp(\theta_1\, s)}} \begin{pmatrix} 1 \\ s \end{pmatrix}$$

with $k := c\,\exp(\theta_0)$, the information matrix can be expressed here by

$$I_\theta(\delta) = \int \left(1 - e^{-k\,\exp(\theta_1\, s)}\right) \begin{pmatrix} 1 & s \\ s & s^2 \end{pmatrix}\, \delta(ds) = \int x_\theta(s)\, x_\theta(s)^\top \delta(ds).$$

To derive locally D-optimal two-point designs on $[0, S_u]$, let be $0 \leq s_1 < s_2 \leq S_u$ and set $X_\theta := \begin{pmatrix} x_\theta(s_1)^\top \\ x_\theta(s_2)^\top \end{pmatrix}$. Then $\delta_{s_1,s_2} := \frac{1}{2}e_{s_1} + \frac{1}{2}e_{s_2}$, where e_s is the Dirac measure on s, is D-optimal within all designs with support s_1 and s_2 since with the equivalence theorem of D-optimality (see Kiefer and Wolfowitz 1960) we have

$$x_\theta(s_i)^\top I_\theta(\delta_{s_1,s_2})^{-1} x_\theta(s_i) = u_i^\top X_\theta \left(\frac{1}{2} X_\theta^\top X_\theta\right)^{-1} X_\theta^\top u_i = 2$$

for $i = 1, 2$ (here u_i denotes the i'th unit vector in $\Re^2$). The determinant of the information matrix of a design δ_{s_1,s_2} is given by the following lemma.

8.2.2 Lemma

$$\det(I_\theta(\delta_{s_1,s_2})) = \frac{1}{4}\left(1 - e^{-k\,\exp(\theta_1\, s_1)}\right)\left(1 - e^{-k\,\exp(\theta_1\, s_2)}\right)[s_2 - s_1]^2.$$

Proof. Set $a := \frac{1}{2}\left(1 - e^{-k\,\exp(\theta_1\, s_1)}\right)$, $b := \frac{1}{2}\left(1 - e^{-k\,\exp(\theta_1\, s_2)}\right)$. Then we have

$$\begin{aligned} I_\theta(\delta_{s_1,s_2}) &= \frac{1}{2}\left(1 - e^{-k\,\exp(\theta_1\, s_1)}\right)\begin{pmatrix} 1 & s_1 \\ s_1 & s_1^2 \end{pmatrix} + \frac{1}{2}\left(1 - e^{-k\,\exp(\theta_1\, s_2)}\right)\begin{pmatrix} 1 & s_2 \\ s_2 & s_2^2 \end{pmatrix} \\ &= \begin{pmatrix} a + b & a\, s_1 + b\, s_2 \\ a\, s_1 + b\, s_2 & a\, s_1^2 + b\, s_2^2 \end{pmatrix} \end{aligned}$$

so that

$$\begin{aligned}
&\det(I_\theta(\delta_{s_1,s_2})) = (a+b)(a\,s_1^2 + b\,s_2^2) - (a\,s_1 + b\,s_2)^2 \\
&\quad = \quad a^2\,s_1^2 + ab\,s_2^2 + abs_1^2 + b^2 s_2^2 - a^2\,s_1^2 - 2ab\,s_1\,s_2 - b^2\,s_2^2 \\
&\quad = \quad a\,b\,[s_2^2 - 2s_1 s_2 + s_1^2] = a\,b\,[s_2 - s_1]^2.
\end{aligned}$$

□

8.2.3 Theorem
Let be $k := c\exp(\theta_0) > 0$. *Then* $\delta_{0,S_u} = \frac{1}{2}e_0 + \frac{1}{2}e_{S_u}$ *is the D-optimal design within all two-point designs on* $\mathcal{S} = [0, S_u]$ *if and only if* $\theta_1 \leq \frac{2}{k\,S_u}(e^k - 1)$.

Proof. Since $1 - e^{-k\,\exp(\theta_1\,s)}$ is strictly increasing in s, $\det(I_\theta(\delta_{s_1,s_2}))$ is maximized with respect to $s_2 \in (s_1, S_u]$ for any given $s_1 \in [0, S_u]$ if and only if $s_2 = S_u$. Therefore we have only to determine $s \in [0, S_u]$ so that $\det(I_\theta(\delta_{s,S_u}))$ is maximized. This is equivalent of maximizing

$$g(s) = \left(1 - e^{-k\,\exp(\theta_1 s)}\right)[S_u - s]^2.$$

Since we have

$$g'(s) = e^{-k\,\exp(\theta_1 s)}\,k\,\theta_1\,\exp(\theta_1\,s)[S_u - s]^2 - 2\left(1 - e^{-k\,\exp(\theta_1 s)}\right)[S_u - s],$$

δ_{0,S_u} can be only D-optimal if

$$0 \geq g'(0) = e^{-k}\,k\,\theta_1\,S_u^2 - 2\left(1 - e^{-k}\right)S_u \Longleftrightarrow e^{-k}\,k\,\theta_1\,S_u \leq 2\left(1 - e^{-k}\right).$$

This is equivalent with $\theta_1 \leq \frac{2}{k\,S_u}\,e^k\,\left(1 - e^{-k}\right) = \frac{2}{k\,S_u}(e^k - 1)$. Hence δ_{0,S_u} is not D-optimal if $\theta_1 > \frac{2}{k\,S_u}(e^k - 1)$. To prove that δ_{0,S_u} is indeed the D-optimal two-point design for $\theta_1 \leq \frac{2}{k\,S_u}(e^k - 1)$, it is sufficient to prove that g is strictly decreasing on $[0, S_u]$. This property is given by the following Lemma. □

8.2.4 Lemma
If $\theta_1 \leq \frac{2}{k\,S_u}(e^k - 1)$ *and* $k > 0$, *then* $g : [0, S_u] \to \Re$ *given by* $g(s) = \left(1 - e^{-k\,\exp(\theta_1 s)}\right)[S_u - s]^2$ *is strictly decreasing.*

Proof. To show $g'(s) < 0$, we need the monotonicity of some auxiliary functions.
a) We have for $h_1(k) := 1 - e^k + k\,e^k - k^2\,e^k$

$$h_1'(k) = -e^k + e^k + k\,e^k - 2\,k\,e^k - k^2\,e^k = -k\,e^k - k^2\,e^k < 0$$

so that h_1 is strictly decreasing for $k > 0$. Since obviously $h_1(0) = 0$, it holds $h_1(k) < 0$ for all $k > 0$.
b) Now consider $h_2(k) := \frac{2}{k}(e^k - 1) - 1 - 2e^k$. The rule of L'Hospital provides

$$\lim_{k\downarrow 0} \frac{2(e^k - 1)}{k} = \lim_{k\downarrow 0} \frac{2e^k}{1} = 2$$

so that $\lim_{k\downarrow 0} h_2(k) = -1$. Then $h_2(k) < 0$ for all $k \geq 0$ follows with a) from

$$h_2'(k) = -\frac{2}{k^2}(e^k - 1) + \frac{2}{k}\, e^k - 2\, e^k = 2\, k^2\, h_1(k) < 0.$$

c) $\theta_1 \leq \frac{2}{k\, S_u}(e^k - 1)$ and b) imply for $g_1(s) := \theta_1[S_u - s] - 1 - 2e^{k\, \exp(\theta_1 s)}$

$$\begin{aligned} g_1(0) &= \theta_1\, S_u - 1 - 2\, e^k \\ &\leq \frac{2}{k\, S_u}(e^k - 1)\, S_u - 1 - 2\, e^k = \frac{2}{k}(e^k - 1) - 1 - 2e^k = h_2(k) < 0 \end{aligned}$$

for all $k \geq 0$. Because of

$$g_1'(s) = -\theta_1 - 2\, e^{k\, \exp(\theta_1 s)}\, k\, \theta_1\, \exp(\theta_1 s) < 0,$$

we have $g_1(s) < 0$ for all $k \geq 0$, $s \geq 0$.
d) $\theta_1 \leq \frac{2}{k\, S_u}(e^k - 1)$ implies for $g_2(s) := k\, \theta_1\, \exp(\theta_1\, s)[S_u - s] + 2 - 2\, e^{k\, \exp(\theta_1 s)}$

$$g_2(0) = k\, \theta_1\, S_u + 2 - 2\, e^k \leq k\, \frac{2}{k\, S_u}(e^k - 1)\, S_u + 2 - 2\, e^k = 2\, e^k - 2 + 2 - 2\, e^k = 0.$$

Moreover, with c) we obtain

$$\begin{aligned} g_2'(s) &= k\, \theta_1^2\, \exp(\theta_1\, s)[S_u - s] - k\, \theta_1\, \exp(\theta_1\, s) - 2\, e^{k\, \exp(\theta_1 s)}\, k\, \theta_1\, \exp(\theta_1 s) \\ &= k\, \theta_1\, \exp(\theta_1\, s) \left[\theta_1\, [S_u - s] - 1 - 2\, e^{k\, \exp(\theta_1 s)}\right] = k\, \theta_1\, \exp(\theta_1\, s)\, g_1(s) < 0 \end{aligned}$$

so that g_2 is strictly decreasing from a value $g_2(0) \leq 0$ which implies $g_2(s) < 0$ for all $k > 0$, $s > 0$.
e) Finally, we have

$$\begin{aligned} g'(s) &= e^{-k\, \exp(\theta_1 s)}\, k\, \theta_1\, \exp(\theta_1\, s)[S_u - s]^2 - 2\, \left(1 - e^{-k\, \exp(\theta_1 s)}\right) [S_u - s] \\ &= [S_u - s]\, e^{-k\, \exp(\theta_1 s)}\, [k\, \theta_1\, \exp(\theta_1\, s)[S_u - s] + 2] - 2\, [S_u - s] < 0 \\ &\iff \\ &\quad e^{-k\, \exp(\theta_1 s)}\, [k\, \theta_1\, \exp(\theta_1\, s)[S_u - s] + 2] < 2 \\ &\iff \\ &\quad k\, \theta_1\, \exp(\theta_1\, s)[S_u - s] + 2 < 2\, e^{k\, \exp(\theta_1 s)} \\ &\iff \\ &\quad g_2(s) < 0 \end{aligned}$$

so that d) provides the assertion. □

8.2.5 Lemma

$$x_\theta(s)^\top I_\theta(\delta_{0,S_u})^{-1} x_\theta(s) = \frac{2}{S_u^2}\, \left(1 - e^{-k\, \exp(\theta_1 s)}\right) \left(\frac{(S_u - s)^2}{1 - e^{-k}} + \frac{s^2}{1 - e^{-k\, \exp(\theta_1\, S_u)}}\right).$$

Proof. With $s_1 = 0$ and $s_2 = S_u$ we obtain (see the proof of Lemma 8.2.2)

$$I_\theta(\delta_{0,S_u}) = \begin{pmatrix} a+b & b\,S_u \\ b\,S_u & b\,S_u^2 \end{pmatrix}$$

with $a := \frac{1}{2}\left(1 - e^{-k}\right)$, $b := \frac{1}{2}\left(1 - e^{-k\,\exp(\theta_1\,S_u)}\right)$. Then we have

$$I_\theta(\delta_{0,S_u})^{-1} = \frac{1}{a\,b\,S_u^2} \begin{pmatrix} b\,S_u^2 & -b\,S_u \\ -b\,S_u & a+b \end{pmatrix}$$

so that

$$\begin{aligned}
& x_\theta(s)^\top I_\theta(\delta_{0,S_u})^{-1} x_\theta(s) \\
= \; & \frac{1 - e^{-k\,\exp(\theta_1 s)}}{a\,b\,S_u^2}(1,s) \begin{pmatrix} b\,S_u^2 & -b\,S_u \\ -b\,S_u & a+b \end{pmatrix} \begin{pmatrix} 1 \\ s \end{pmatrix} = \frac{1 - e^{-k\,\exp(\theta_1 s)}}{a\,b\,S_u^2}(1,s) \begin{pmatrix} b\,S_u^2 - b\,S_u\,s \\ -b\,S_u + (a+b)\,s \end{pmatrix} \\
= \; & \frac{1 - e^{-k\,\exp(\theta_1 s)}}{a\,b\,S_u^2}(b\,S_u^2 - b\,S_u\,s - b\,S_u\,s + (a+b)\,s^2) = \frac{1 - e^{-k\,\exp(\theta_1 s)}}{a\,b\,S_u^2}(b\,(S_u - s)^2 + a\,s^2).
\end{aligned}$$

□

To prove that δ_{0,S_u} is D-optimal within all designs on $\mathcal{S} = [0, S_u]$, the property

$$\begin{aligned}
2 \; & \geq \; x_\theta(s)^\top I_\theta(\delta_{0,S_u})^{-1} x_\theta(s) \\
& = \; \frac{2}{S_u^2}\left(1 - e^{-k\,\exp(\theta_1 s)}\right)\left(\frac{(S_u - s)^2}{1 - e^{-k}} + \frac{s^2}{1 - e^{-k\,\exp(\theta_1\,S_u)}}\right)
\end{aligned} \tag{8.1}$$

must be shown for all $s \in [0, S_u]$ according to Kiefer and Wolfowitz (1960) where equality holds only for $s = 0$ and $s = S_u$. The equality is indeed always satisfied for $s = 0$ and $s = S_u$. Set

$$q(s) := \left(1 - e^{-k\,\exp(\theta_1 s)}\right)\left(\frac{(S_u - s)^2}{1 - e^{-k}} + \frac{s^2}{1 - e^{-k\,\exp(\theta_1\,S_u)}}\right).$$

A necessary condition for the D-optimality of δ_{0,S_u} is then $q'(0) \leq 0$.

8.2.6 Lemma

$q'(0) \leq 0$ if and only if $\theta_1 \leq \frac{2}{k\,S_u}(e^k - 1)$.

Proof.

$$\begin{aligned}
q'(s) = e^{-k\,\exp(\theta_1 s)}\,k\,\theta_1\,\exp(\theta_1 s) & \left(\frac{(S_u - s)^2}{1 - e^{-k}} + \frac{s^2}{1 - e^{-k\,\exp(\theta_1\,S_u)}}\right) \\
+ \left(1 - e^{-k\,\exp(\theta_1 s)}\right) & \left(\frac{-2(S_u - s)}{1 - e^{-k}} + \frac{2s}{1 - e^{-k\,\exp(\theta_1\,S_u)}}\right)
\end{aligned}$$

so that

$$
\begin{aligned}
0 \geq q'(0) &= e^{-k}\, k\, \theta_1 \frac{S_u^2}{1-e^{-k}} - \left(1-e^{-k}\right) \frac{2S_u}{1-e^{-k}} \\
\iff \quad k\, \theta_1\, S_u \leq e^k\, 2\,(1-e^{-k}) &\iff \theta_1 \leq \frac{2}{k\, S_u}(e^k-1).
\end{aligned}
$$

□

Hence the condition $\theta_1 \leq \frac{2}{k\,S_u}(e^k-1)$ implies not only that δ_{0,S_u} is the locally D-optimal design within all two-point designs on $[0, S_u]$ but also the necessary condition for D-optimality of δ_{0,S_u} within all designs on $[0,1]$. Several plots of $q(s)$ for different values of θ_1 and k with $\theta_1 \leq \frac{2}{k\,S_u}(e^k-1)$ showed that q is first decreasing and then increasing on $[0, S_u]$ so that (8.1) should be satisfied.

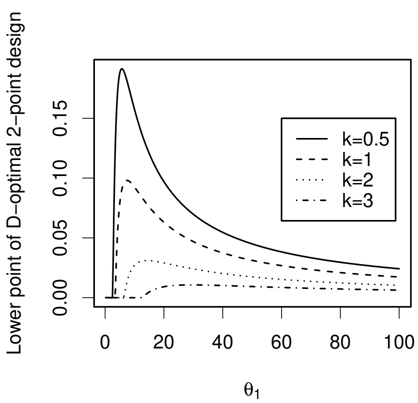

Figure 8.1: Lower points $s(\theta_1, k)$ of the D-optimal two-point designs on [0,1].

As soon as $\theta_1 > \frac{2}{k\,S_u}(e^k-1)$ holds then the locally D-optimal two-point design is of the form $\delta_{s(\theta_1,k),S_u}$ with $0 < s(\theta_1, k) < S_u$. The lower points $s(\theta_1, k)$ depending on θ_1 are shown in Fig. 8.1 for $k = 0.5, 1, 2, 3$ and $S_u = 1$. The condition $\theta_1 > \frac{2}{k\,S_u}(e^k-1)$ is in particular satisfied if k is small. The quantity $k := c\,\exp(\theta_0)$ is small if the censoring variable c or the regression parameter θ_0 is small. A small θ_0 means a high expected lifetime at $s = 0$ which provides a high probability of censoring. Then it is reasonable to make the observations at higher stress levels $s(\theta_1, k) > 0$ so that the probability of censoring is smaller. But since $\frac{2}{k\,S_u}(e^k-1) \geq \frac{2}{S_u}$ for all $k \geq 0$, the censoring variable as well as θ_0 have no influence on the D-optimal design as soon as $\theta_1 \leq \frac{2}{S_u}$. The condition $\theta_1 > \frac{2}{k\,S_u}(e^k-1)$ is also satisfied if θ_1 is large. In this case, the expected lifetime decreases so rapidly with growing stress that observations at $s(\theta_1, k) > 0$ provide more information than at 0 where observations are censored with higher probability.

8.2.7 Lemma
It holds $\frac{e^k-1}{k} \geq 1$ *for all* $k \geq 0$.

Proof.

$$\frac{e^k - 1}{k} \geq 1 \Longleftrightarrow e^k - 1 \geq k \Longleftrightarrow h(k) := e^k - 1 - k \geq 0,$$

But $h(k) := e^k - 1 - k \geq 0$ is satisfied since $h(0) = 0$ and $h'(k) = e^k - 1 \geq 0$ for all $k \geq 0$. □

Chapter 9

References

1. Ansell, J.I. and Phillips, M.J. (1994). *Practical Methods for Reliability Data Analysis.* Oxford University Press, New York.

2. Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (2006). *Case Studies in Spatial Point Process Modeling.* Lecture Notes in Statistics 185, Springer, New York.

3. Balakrishnan, N. (2009). A synthesis of exact inferential results for exponential step-stress models and associated optimal accelerated life-tests. *Metrika* **69**, 351-396.

4. Basquin, O. (1910). The exponential law of endurance tests. *ASTM, Philadelphia, PA.*

5. Birnbaum, Z.W. and Saunders, S.C. (1969). A new family of life distributions. *Journal of Applied Probability* **6**, 319-327.

6. Blischke, W.R. and Murthy, D.N.P. (2000). *Reliability: Modeling, Prediction and Optimization.* Wiley, New York.

7. Bunke, H. and Bunke, O. (1989). *Nonlinear Regression, Functional Relations and Robust Methods. Statistical Methods of Model Building, Volume II.* Wiley, New York.

8. Carlin, B.P. and Louis, T.A. (2009). *Bayesian Methods for Data Analysis.* CRC Press, Boca Raton.

9. Castillo, E. and Fernandez-Canteli, A. (2009). *A Unified Statistical Methodology for Modeling Fatigue Damage.* Springer Science + Business Media, Dordrecht.

10. Daley, D.J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods.* Springer, New York.

11. Deshpande, J.V. and Purohit, S.G. (2015). *Lifetime Data: Statistical Models and Methods.* World Scientific Publishing Co Inc, New Jersey.

12. Gneiting, T. and Raftery, A.E. (2007). Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association* **102**, 359-378.

13. Gunkel, C., Stepper, A., Müller, A.C., and Müller, Ch.H. (2012). Micro crack detection with Dijkstra's shortest path algorithm. *Machine Vision and Applications* **23**, 589-601.

14. Gupta, A.K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika* **39**, 260-273.

15. Haibach, E. (2006). *Betriebsfestigkeit. Verfahren und Daten zur Bauteilberechnung.* Springer, Berlin.

16. Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. and Ostrowski, E. (1994). *A Handbook of Small Data Sets.* Chapman & Hall, London.

17. Hudak, S. J., Saxena, A. Bucci, R. J. and Malcom, R. C. (1978). Development of standard methods of testing and analyzing fatigue crack growth rate data. Technical report. AFML-TR-78-40. Westinghouse R & D Center, Wesinghouse Electric Corp., Pittsburgh, PA.

18. Iacus, S.M. (2008). *Simulation and Inference for Stochastic Differential Equations.* Springer, New York.

19. Jacobsen, M. (2006). *Point Process Theory and Applications. Marked Point and Piecewise Deterministic Processes.* Birkhäuser, Boston.

20. Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.

21. Jiang, R. and Murthy, D.N.P. (1995), Modeling failure-data by mixture of 2 Weibull distributions : A Graphical approach. *IEEE Trans. Reliab.* **44**, 477–488.

22. Kahle, W. and Liebscher, E. (2013). *Zuverlässigkeitsanalye und Qualitätssicherung.* Oldenbourg Verlag, München.

23. Kahle, W., Mercier, S., and Paroissin, C. (2016). *Degradation Processes in Reliability.* Wiley, New York.

24. Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Can. J. Math.* **12**, 363-366.

25. Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer, New York.

26. Krengel, U. (1991). *Einführung in die Wahrscheinlichkeitstheorie und Statistik.* Vieweg, Braunschweig.

27. Kustosz, Ch.P., Leucht, A. and Müller, Ch.H. (2016). Tests based on simplicial depth for AR(1) models with explosion. *Journal of Time Series Analysis* **37**, 763-784.

28. Kustosz, Ch.P., Müller, Ch.H. and Wendler, M. (2016). Simplified simplicial depth for regression and autoregressive growth processes. *Journal of Statistical Planning and Inference* **173**, 125-146.

29. Meeker, W.Q. and Escobar, L.A. (1998). *Statistical Methods for Reliability Data.* Wiley, New York.

30. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller,A.H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**, 1087-1092.

31. Müller, Ch.H. (1995). Maximin efficient designs for estimating nonlinear aspects in linear models. *J. Statist. Plann. Inference* **44**, 117-132.

32. Müller, Ch.H. (1997). *Robust Planning and Analysis of Experiments.* Lecture Notes in Statistics 124, Springer, New York.

33. Müller, Ch.H. (2013). D-optimal designs for lifetime experiments with exponential distribution and censoring. In: *mODa 10 - Advances in Model-Oriented Design and Analysis.* Eds. D. Ucinski, A.C. Atkinson and M. Patan, Physica-Verlag, Heidelberg, 179-186.

34. Müller, Ch.H. and Pázman, A. (1998). Applications of necessary and sufficient conditions for maximin efficient designs. *Metrika* **48**, 1-19.

35. Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods.* Springer Science & Business Media, Dordrecht.

36. Rosenthal, J.S. (2011). Optimal proposal distributions and adaptive MCMC. In: *Handbook of Markov Chain Monte Carlo.* Eds. S. Brooks, A. Gelman, G.L. Jones and X.-L. Meng, Chapman & Hall/CRC, London, 93-112.

37. Schervish, M.J. (1997). *Theory of Statistics.* Springer, New York.

38. Silvey, S.D. (1980). *Optimal Design.* Chapman and Hall, London.

39. Sobczyk, K. and Spencer, B.F. (1992). *Random Fatigue. From Data to Theory.* Academic Press, London.

40. Virkler, D. A., B. M. Hillberry and P. K. Goel (1979). The Statistical Nature of Fatigue Crack Propagation. *Journal of Engineering Materials and Technology* **101**, 148-153.

41. Witting, H. und Müller-Funk, U. (1995). *Mathematische Statistik II.* Teubner, Stuttgart.