Lecture Notes on
Censoring and Estimation in Statistical Sampling

Yakov Ben-Haim
Yitzhak Moda'i Chair in Technology and Economics
Faculty of Mechanical Engineering
Technion — Israel Institute of Technology
Haifa 32000 Israel
yakov@technion.ac.il
http://info-gap.com http://www.technion.ac.il/yakov

Source material: J.F. Lawless, 1982, *Statistical Models and Methods for Lifetime Data,* Section 1.4, pp.31–44; Chapters 3, 4.

A Note to the Student: These lecture notes are not a substitute for the thorough study of books. These notes are no more than an aid in following the lectures.

Contents

1 Preliminary Example: Weibull Reliability Analysis

In this section we consider a preliminary example to motivate our later discussion.

1.1 The Weibull Distribution

The two-parameter Weibull cdf is:

$$F(t) = 1 - \mathrm{e}^{-(\lambda t)^\alpha} \tag{1}$$

where λ and α are positive parameters. Typically, $0.5 < \alpha < 4$. The pdf is found by differentiating $F(t)$:

$$f(t) = \frac{\mathrm{d}F}{\mathrm{d}t} = \alpha\lambda^\alpha t^{\alpha-1}\mathrm{e}^{-(\lambda t)^\alpha} \tag{2}$$

To understand the physical meaning of the parameters, let us examine the failure rate function based on adopting the Weibull density as the pdf for failure:

$$z(t) = \frac{f(t)}{1 - F(t)} = \alpha\lambda^\alpha t^{\alpha-1} \tag{3}$$

We see that λ scales the magnitude of the failure rate function: the failure rate increases as λ increases, so large λ implies large failure rate. α controls the time dependence. $z(t)$ is increasing, constant, or decreasing in time depending on whether $\alpha > 1$, $\alpha = 1$ or $\alpha < 1$.

1.2 Weibull Analysis

§ We have a population of M items of the same sort. $N < M$ items have failed at times $t_1 \le t_2 \le , \ldots , \le t_N$. We wish to determine the best coefficients of a Weibull distribution. Why? These coefficients tell us what stage of the bathtub curve the population is in: burn-in (phase 1), central (phase 2) or burn-out (phase 3).

§ The general approach is to use this data to estimate $F(t)$ empirically, and then fit the parameters. In this section we use only the data on the failed items, and ignore the fact that $M - N$ items are still functional because we do not know when these items will fail. The remainder of this lecture is devoted to learning how to incorporate this additional data.

§ To establish the empirical distribution function, note that the times are ranked in increasing order. Thus the fraction $1/N$ of the sample failed in time t_1, a fraction $2/N$ failed in time t_2 and so on. We could thus estimate $F(t)$ as:

$$\widehat{F}(t_i) = \frac{i}{N}, \quad i = 1, \ldots, N \tag{4}$$

This is not a good approximation to the cdf because the data are not at random times but rather at failure times at which the empirical function makes a step increase.

§ Note that:

$$\widehat{F}(t) = 1 \quad \text{for} \quad t \ge t_n \tag{5}$$

That clearly is not correct.

§ A better approximation, which is strictly *ad hoc,* is:

$$\widehat{F}(t_i) = \frac{i - 0.3}{N + 0.4}, \quad i = 1, \ldots, N \tag{6}$$

§ We now have an estimate of the distribution at N instants: $\widehat{F}(t_i)$, $i = 1, \ldots, N$. We can use either graphical methods or algebraic methods to find the best estimates for α and λ.

§ The graphical method for estimating the Weibull parameters could be based on noting that:

$$\ln\left(-\ln[1 - F(t)]\right) = \alpha \ln t + \alpha \ln \lambda \tag{7}$$

Thus, using log-log paper to plot $-\ln[1 - F(t)]$ versus $\ln t$ would result in a straight line whose slope is α and whose intercept is $\alpha \ln \lambda$.

§ An algebraic method could be based on choosing α and λ to minimize the deviation of the data from the fitted function. This least-squares approach is to choose α and λ to minimize:

$$S^2 = \sum_{i=1}^{N} \left[\widehat{F}(t_i) - F(t_i)\right]^2 \tag{8}$$

where $\widehat{F}(t_i)$ is the estimate of the cpf, eq.(4) or eq.(6), and $F(t_i)$ is the Weibull distribution evaluated at time t_i. We choose α and λ to minimize S^2.

Number of cycles to failure (t)	Failure number (i)
430	1
900	2
1090	3
1220	4
1500	5
1910	6
1915	7
2250	8
2600	9
2610	10
3000	11
3390	12
3430	13
3700	14
4050	15

Table 1: Failure data for example in section 1.3.

1.3 Example: Weibull Distribution and Failure Rate Function

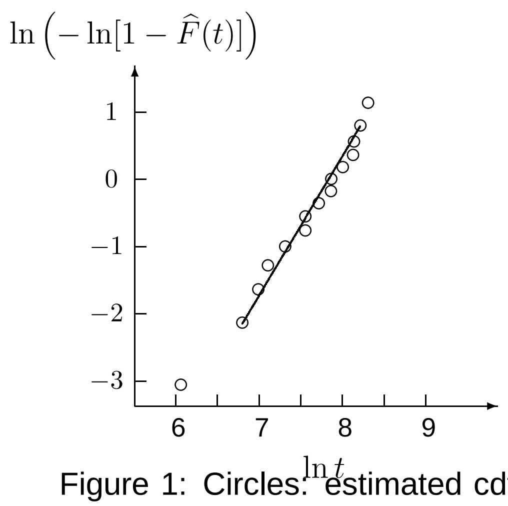

Figure 1: Circles: estimated cdf versus failure time, eqs.(6) and (7). Line: approximate linear fit with slope = 2.1

§ In table 1 are listed the lifetimes of 15 mechanical switches which failed, from among a population of 20 switches. The term ‘lifetime’ refers to the number of cycles performed before failure occurs. All other switches in the population survived more than 4050 cycles. Is the probability of failure of these switches constant, increasing or decreasing in time?

§ From the 15 data points of table 1 we calculate the empirical cdf $\widehat{F}(t_i)$ according to eq.(6). Then we plot $\ln[-\ln(1-F(t))]$ versus $\ln t$ as shown in fig. 1. An approximate “eyeball” linear fit shows the slope to be $\alpha = 2.1$. So the failure rate function is increasing in time, since $\alpha - 1 > 0$.

2 Introduction to Censoring

(p. 31)

¶ Briefly, "censoring" of data arises when exact lifetimes are known only for a portion of the individuals under study.

¶ Formally: an observation is censored at L if the exact value of the observation is not known except that it is $\geq L$.

This is called "right censoring", which is the relevant type of censoring for lifetime data. "Left censoring" rarely occurs in lifetime data.

¶ Different types of censoring occur, and in each case, for any given pdf (probability density function), we must determine:

- The sampling distribution ("funkziat dgima").
- The likelihood function ("funkziat svirut" or "nirut").

Then we determine the properties of statistical estimators. Usually we have a large sample so that we can exploit asymptotic properties.

¶ We will consider:

- Type II censoring (section 3).
- Type I censoring (section 7).
- Random censoring (section 8).

3 Type II Censoring

(pp.32–34)

¶ A type II censored sample is one for which:

1. Only the r smallest observations in a sample of size n are observed, $1 \le r \le n$.
2. r is determined **before** the data are collected.

¶ Let the n lifetimes of the size-n sample be $T_1, \ldots, T_n$.
Their order statistics are:

$$T_{(1)} \le T_{(2)} \le \cdots \le T_{(n)} \tag{9}$$

In type II censoring we know only the values:

$$T_{(1)}, \ldots, T_{(r)} \tag{10}$$

¶ Let $f(t)$ be the pdf of the lifetime:

$$f(t)\,\mathrm{d}t = \text{probability of end-of-life } T \in [t,\ t + \mathrm{d}t] \tag{11}$$

The “survivor function” or “probabilistic reliability” is:

$$S(t) = \text{Prob}(T \ge t) \tag{12}$$

$$= \int_t^\infty f(s)\,\mathrm{d}s \tag{13}$$

¶ If $T_1, \ldots, T_n$ are iid (independent and identically distributed) with lifetime pdf $f(t)$ and survivor function $S(t)$, then the joint pdf of $T_{(1)}, \ldots, T_{(r)}$ is:

$$f_n(t_{(1)}, \ldots, t_{(r)}) = \frac{n!}{(n-r)!} f(t_{(1)}) \cdots f(t_{(r)}) [S(t_{(r)})]^{n-r} \tag{14}$$

Explanation:

1. $\dfrac{n!}{(n-r)!}$ = number of ways of choosing $n-r$ out of n items, without regard to the order in which the items are chosen. The order-free choice holds only for the non-failed items.

   For instance, $n = 3$ and $n - r = 2$: $\dfrac{3!}{2!} = 3$. Let the items be A, B and C. We can choose the following three couples: $\{A, B\}$, $\{A, C\}$, $\{B, C\}$.

2. $S(t_{(r)})$ = probability that a specific item will live at least $t_{(r)}$.

3. Thus $[S(t_{(r)})]^{n-r}$ = probability that $n-r$ specific items will have lifetimes $\geq t_{(r)}$.

4. Thus $\dfrac{n!}{(n-r)!}[S(t_{(r)})]^{n-r}$ = probability that $n-r$ items, from a population of size n, will have lifetimes $\geq t_{(r)}$.

5. $f(t_{(1)}) \cdots f(t_{(r)})$ = the joint probability density for the r specific independent items whose lifetimes are known.

¶ The likelihood function for any parameter model is based on $f_n(t_{(1)}, \ldots, t_{(r)})$ from eq.(14).

4 Estimating the MTBF in the Exponential Distribution

In this section we discuss estimation of the MTBF for the exponential distribution based on type II censored data.

4.1 Poisson Process and the Exponential Distribution

¶ Why is the exponential distribution common in lifetime modelling? Because many situations correspond to the assumptions which underlie its derivation.

¶ The assumptions which underlie the Poisson process are:

- Events occur independently as points on a continuum.
- The average rate of events is constant.

With these assumptions one can show that the probability of exactly n events in duration (or space) t is:

$$P_n(t) = \frac{\mathrm{e}^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0,\, 1,\, \ldots \tag{15}$$

¶ The probability of no event occurring in a duration $[0,\, t]$ is:

$$P_0(t) = \mathrm{e}^{-\lambda t} \tag{16}$$

¶ The probability of one event occurring in an infinitesimal interval $[t,\, t + \mathrm{d}t]$ is:

$$\text{Prob}[t,\, t + \mathrm{d}t] = \lambda \mathrm{d}t \tag{17}$$

¶ We can combine eqs.(16) and (17) to find the probability that no event occurs during $[0,\, t]$ and then one event occurs during $[t,\, t + \mathrm{d}t]$:

$$f(t)\mathrm{d}t = P_0(t)\lambda \mathrm{d}t = \mathrm{e}^{-\lambda t}\lambda \mathrm{d}t, \quad t \ge 0 \tag{18}$$

t is a random variable which expresses the waiting time between events. If these events are failures, then t is the time between failures (or the time since recovery from the last failure, if repair takes a positive amount of time).

¶ The **mean time between failures** is the expectation of t:

$$\mathrm{E}(t) = \int_0^\infty t f(t)\, \mathrm{d}t = \frac{1}{\lambda} \tag{19}$$

Thus estimating the parameter of an exponential distribution is precisely the problem of estimating the MTBF.

4.2 Maximum Likelihood Estimate for the Exponential Distribution

Let $t_1, \ldots, t_r$ be a type II censored sample from an exponential distribution:

$$
\begin{aligned}
f(t) &= \lambda \mathrm{e}^{-\lambda t}, \quad t \geq 0 && (20)\\
S(t) &= \mathrm{e}^{-\lambda t} && (21)
\end{aligned}
$$

Thus the joint probability density, eq.(14), is:

$$
f_n(t_{(1)}, \ldots, t_{(r)}) = \lambda^r \frac{n!}{(n-r)!} \exp\left[-\lambda\left((n-r)t_{(r)} + \sum_{i=1}^{r} t_{(i)}\right)\right] \tag{22}
$$

This is the likelihood function for λ: $L(\lambda | t_{(1)}, \ldots, t_{(r)})$.

The maximum likelihood estimate for the parameter λ, given the observations $t_{(1)}, \ldots, t_{(r)})$, is found by:

$$
\widehat{\lambda} = \arg\max_{\lambda} f_n(t_{(1)}, \ldots, t_{(r)}) \tag{23}
$$

Define the “total time on test”:

$$
T = (n-r)t_{(r)} + \sum_{i=1}^{r} t_{(i)} \tag{24}
$$

Denote $b = \dfrac{n!}{(n-r)!}$. Thus:

$$
f_n(\lambda) = b\lambda^r \mathrm{e}^{-\lambda T} \tag{25}
$$

Thus we find the λ which maximizes f_n as:

$$
\begin{aligned}
\frac{\partial f_n}{\partial \lambda} &= b\mathrm{e}^{-\lambda T}\left[r\lambda^{r-1} - \lambda^r T\right] && (26)\\
&= b\lambda^{r-1}\mathrm{e}^{-\lambda T}(r - \lambda T) && (27)
\end{aligned}
$$

Thus the MLE for λ is:

$$
\widehat{\lambda} = \frac{r}{T} \tag{28}
$$

4.3 Maximum Likelihood Estimate for the Exponential Distribution: Continued

Let us repeat the previous example, but ignore the censored data. What impact would this have on the estimate?

In this case, instead of eq.(22), the joint distribution of the non-censored data is:

$$f_n(t_{(1)}, \ldots, t_{(r)}) = \lambda^r \exp\left[-\lambda \sum_{i=1}^{r} t_{(i)}\right] \tag{29}$$

Define:

$$T_o = \sum_{i=1}^{r} t_{(i)} \tag{30}$$

Thus:

$$f_n(t_{(1)}, \ldots, t_{(r)}) = \lambda^r \mathrm{e}^{-\lambda T_o} \tag{31}$$

Hence, instead of eqs.(26) and (27):

$$\frac{\partial f_n}{\partial \lambda} = \mathrm{e}^{-\lambda T_o}\left[r\lambda^{r-1} - \lambda^r T_o\right] \tag{32}$$

$$= \lambda^{r-1}\mathrm{e}^{-\lambda T_o}(r - \lambda T_o) \tag{33}$$

Thus the MLE for λ is, instead of eq.(28):

$$\widehat{\lambda}_o = \frac{r}{T_o} \tag{34}$$

We see that ignoring the censored data increases the estimate of λ and decreases the estimated MTBF:

$$T_o \leq T = T_o + (n-r)t_{(r)} \quad \Longrightarrow \quad \widehat{\lambda}_o \geq \widehat{\lambda} \quad \Longrightarrow \quad \frac{1}{\widehat{\lambda}_o} \leq \frac{1}{\widehat{\lambda}_o} \tag{35}$$

with strict inequality unless there are no censored data.

5 Confidence Intervals for the Exponential MTBF

¶ Recall our definition in eq.(24) on page 9 of the "total time on test":

$$T = (n-r)t_{(r)} + \sum_{i=1}^{r} t_{(i)} \tag{36}$$

T is a random variable — a statistic — and it has a chi squared distribution. Specifically:

$$2\lambda T \sim \chi^2_{(2r)} \tag{37}$$

That is, $2\lambda T$ is distributed as a chi squared random variable with $2r$ degrees of freedom.

λ in this relation is a constant parameter; the randomness comes from T. However, we will use this relation to derive something like a confidence interval for our estimate of λ.

¶ Let $\chi^2_{(n),p}$ denote the pth quantile of $\chi^2_{(n)}$:

$$\text{Prob}(\chi^2_{(n)} \le \chi^2_{(n),p}) = p \tag{38}$$

Thus $2\lambda T$ has probability $1-\alpha$ to fall between $\chi^2_{(2r),\frac{\alpha}{2}}$ and $\chi^2_{(2r),1-\frac{\alpha}{2}}$. That is:

$$\text{Prob}\left(\chi^2_{(2r),\frac{\alpha}{2}} \;\le\; 2\lambda T \;\le\; \chi^2_{(2r),1-\frac{\alpha}{2}}\right) = 1-\alpha \tag{39}$$

Thus a two-sided $1-\alpha$ confidence interval for the random variable $2\lambda T$ is:

$$\chi^2_{(2r),\frac{\alpha}{2}} \;\le\; 2\lambda T \;\le\; \chi^2_{(2r),1-\frac{\alpha}{2}} \tag{40}$$

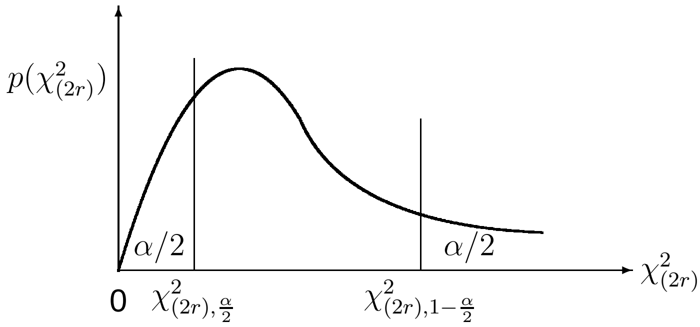

Figure 2: Two-sided confidence interval for $2\lambda T$, eq.(40).

¶ We have stressed that T is the random variable; 2λ is just a constant. So we really cannot simply manipulate eq.(40) and say that a $1-\alpha$ confidence interval for λ is:

$$\frac{\chi^2_{(2r),\frac{\alpha}{2}}}{2T} \;\le\; \lambda \;\le\; \frac{\chi^2_{(2r),1-\frac{\alpha}{2}}}{2T} \qquad (41)$$

We do not have a probability distribution for λ which in fact is not a random variable.

But what **can** we say about the interval in eq.(41)?

¶ We can think of eq.(41) as a **likelihood interval** for λ in the same way that we think of r/T as a maximum likelihood estimate of λ, eq.(28) on p. 9.

That is, given an observed value of T, the "likely" or "reasonable" value for λ is the MLE of eq.(28) and the $1-\alpha$ likelihood interval of λ is eq.(41).

Example 1 Consider a type-II censored sample of size $r = 8$ from a population of size $N = 12$, with censored data $t_{(1)}, \ldots, t_{(8)} =$ 31, 58, 157, 185, 300, 470, 497, 673 hours. The total time on test, eq.(24), is $T = 5063$ hours. Thus the MLE estimate of λ is $\widehat{\lambda} = r/T = 1.58 \times 10^{-3}$.

We construct a 0.95 likelihood interval as follows. The quantiles of $\chi^2_{(16)}$ are:

$$\chi^2_{(16),0.025} = 6.91 \qquad (42)$$

$$\chi^2_{(16),0.975} = 28.8 \qquad (43)$$

Thus, since $2\lambda T \sim \chi^2_{(16)}$:

$$0.95 = \text{Prob}\left(6.91 \le \chi^2_{(16)} \le 28.8\right) = \text{Prob}\left(6.91 \le 2\lambda T \le 28.8\right) \qquad (44)$$

Hence the 0.95 likelihood interval for λ is:

$$\left(\frac{6.91}{2T} \le \lambda \le \frac{28.8}{2T}\right) = \left(1.36 \times 10^{-3} \le \lambda \le 2.84 \times 10^{-3}\right) \qquad (45)$$

Note that $\widehat{\lambda}$ is in the 0.95 likelihood interval, but it is **not** in the middle. ∎

¶ Sometimes we wish to estimate a likelihood interval for a quantity related to λ. Two examples:

- The survival function:

$$S(t) = 1 - F(t) = 1 - \mathrm{e}^{-\lambda t} \tag{46}$$

- The pth quantile of the lifetime distribution, t_p:

$$\text{Prob}(t \leq t_p) = p \tag{47}$$

So, for the exponential distribution:

$$F(t_p) = p \quad \Longrightarrow \quad 1 - \mathrm{e}^{-\lambda t_p} = p \quad \Longrightarrow \quad t_p = -\frac{\ln(1-p)}{\lambda} \tag{48}$$

In each case, $S(t)$ and t_p, the new quantity is a 1-to-1 function of λ. So, let an α likelihood interval for λ be:

$$A \leq \lambda \leq B \tag{49}$$

Then an α likelihood interval for $S(t)$ is:

$$1 - \mathrm{e}^{-Bt} \leq S(t) \leq 1 - \mathrm{e}^{-At} \tag{50}$$

Similarly an α likelihood interval for t_p is:

$$-\frac{\ln(1-p)}{B} \leq t_p \leq -\frac{\ln(1-p)}{A} \tag{51}$$

Example 2 Consider an hypothesis test based on the data from example 1. Examine the following null and alternative hypotheses:

$$H_0: \quad \lambda = 0.001 \text{ hr}^{-1} \tag{52}$$

$$H_1: \quad \lambda > 0.001 \text{ hr}^{-1} \tag{53}$$

From the data we have $T = 5063$ and we know that $2\lambda T \sim \chi^2_{(16)}$.

We will evaluate the level of significance, which is the probability, conditioned on H_0, of a more extreme result than that which was observed. A small value of T is evidence against H_0.

So the level of significance is:

$$\alpha = \text{Prob}(T \leq 5063 | H_0) \tag{54}$$

$$= \text{Prob}(2\lambda T \leq 2 \times 0.001 \times 5063 | H_0) \tag{55}$$

$$= \text{Prob}(\chi^2_{(16)} \leq 10.126 | H_0) = 0.15 \tag{56}$$

This is not small, so we cannot reject H_0. ■

6 Estimating the Weibull Distribution

6.1 Weibull and Gumbel Extreme Value Distributions

¶ The pdf and cdf of the Weibull distribution are:

$$\begin{aligned} f(t) &= \lambda\beta(\lambda t)^{\beta-1}\mathrm{e}^{-(\lambda t)^\beta}, \quad t \geq 0 && (57)\\ F(t) &= \int_0^t f(\tau)\,\mathrm{d}\tau = 1 - \mathrm{e}^{-(\lambda t)^\beta} && (58)\end{aligned}$$

¶ The failure rate function is:

$$\begin{aligned} z(t)\mathrm{d}t &= \text{Prob of failure in } [t,\, t+\mathrm{d}t] \text{ given survival to } t && (59)\\ &= \frac{f(t)}{1-F(t)} && (60)\\ &= \lambda\beta(\lambda t)^{\beta-1} && (61)\end{aligned}$$

$z(t)$ is constant, increasing and decreasing in time if $\beta = 1$, $\beta > 1$, $\beta < 1$. See Transparency.

¶ The Weibull distribution is a "limit distribution". If $t_1,\, t_2,\, \ldots, t_N$ are independent non-negative random variables, then define:

$$T = \min_i\{t_1,\, t_2,\, \ldots, t_N\} \qquad (62)$$

T has an asymptotic (for large N) Weibull distribution.

The Weibull distribution is a stochastic "weakest link" model.

¶ The Gumbel extreme value pdf and cdf are:

$$\begin{aligned} f(x) &= \frac{1}{b}\exp\left[\frac{x-u}{b} - \exp\left(\frac{x-u}{b}\right)\right], \quad -\infty < x < \infty && (63)\\ F(x) &= 1 - \exp\left[-\exp\left(\frac{x-u}{b}\right)\right] && (64)\end{aligned}$$

where the parameters are $b > 0$ and $-\infty < u < \infty$.

The "standard" Gumbel distribution has $b = 1$, $u = 0$. (Transparency).

¶ The relation between Weibull and Gumbel is:

$$\text{If } T \sim W_{\lambda,\beta} \quad \text{then} \quad \ln T \sim G_{b=\frac{1}{\beta},\ u=-\ln\lambda} \tag{65}$$

¶ Moments of the Weibull distribution:

$$\mathrm{E}(t) = \frac{1}{\lambda}\Gamma\left(1+\frac{1}{\beta}\right) \tag{66}$$

$$\text{var}(t) = \frac{1}{\lambda^2}\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right] \tag{67}$$

¶ Moments of the Gumbel distribution:

$$\mathrm{E}(x) = u - \gamma b, \quad \gamma = \text{Euler's constant} \approx 0.5772 \tag{68}$$

$$\text{var}(x) = \frac{b^2\pi^2}{6} \tag{69}$$

The pth quantile is:

$$x_p = u + b\ln[-\ln(1-p)] \tag{70}$$

That is:

$$\text{Prob}(x \leq x_p) = p \tag{71}$$

6.2 Point Estimates

¶ The Weibull distribution is an important lifetime model because of its relation to extreme value distributions.

There is extensive statistical study of the Weibull distribution because in general there is no 2-dimensional sufficient statistic for estimating its parameters.

¶ Given type II censored data from a sample of size N from a Weibull distribution where the r smallest observations are:

$$t_1 \leq \cdots \leq t_r \tag{72}$$

Equivalently define:

$$x_i = \ln t_i, \quad i = 1, \ldots, r \tag{73}$$

which are the r smallest observations from a sample of size N from a Gumbel distribution.

¶ The joint pdf of $x_1, \ldots, x_r$ is:

$$f(x_1, \ldots, x_r) = \tag{74}$$
$$\frac{N!}{(N-r)!}\left[\prod_{i=1}^{r} \underbrace{\frac{1}{b}\mathrm{e}^{(x_i-u)/b}}_{A} \underbrace{\exp\left(-\mathrm{e}^{(x_i-u)/b}\right)}_{B}\right] \underbrace{\left[\exp\left(-\mathrm{e}^{(x_r-u)/b}\right)\right]^{N-r}}_{C}$$

We consider this function, without the binomial coefficient, the “likelihood function” (LHF) for the parameters u and b, which we denote $L(u, b)$:

$$L(u, b) = \tag{75}$$
$$\underbrace{\frac{1}{b^r}\exp\left(\sum_{i=1}^{r}\frac{x_i-u}{b}\right)}_{A} \exp\left[\underbrace{-(N-r)\mathrm{e}^{\frac{x_r-u}{b}}}_{C} - \underbrace{\sum_{i=1}^{r}\mathrm{e}^{\frac{x_i-u}{b}}}_{B}\right]$$

¶ Define:

$$\sum_{i=1}^{r}{}^{\star}\, w_i = (N-r)w_r + \sum_{i=1}^{r} w_i \tag{76}$$

Now the likelihood function is:

$$L(u, b) = \frac{1}{b^r}\exp\left[\sum_{i=1}^{r}\frac{x_i-u}{b} - \underbrace{\sum_{i=1}^{r}{}^{\star}\exp\left(\frac{x_i-u}{b}\right)}_{BC}\right] \tag{77}$$

The logarithm is more useful, so the log LHF is:

$$\ln L(u, b) = -r\ln b + \sum_{i=1}^{r}\frac{x_i-u}{b} - \sum_{i=1}^{r}{}^{\star}\exp\left(\frac{x_i-u}{b}\right) \tag{78}$$

¶ The maximum likelihood estimates of u and b are obtained from:

$$0 = \frac{\partial \ln L}{\partial u} \tag{79}$$

$$0 = \frac{\partial \ln L}{\partial b} \tag{80}$$

These relations lead to transcendental equations for which no analytical solutions exist:

$$
\begin{aligned}
\mathrm{e}^{u} &= \frac{1}{r}\left(\sum_{i=1}^{r}{}^{\star}\,\mathrm{e}^{x_i/b}\right)^{b} && (81)\\
0 &= \frac{\sum_{i=1}^{r}{}^{\star} x_i \mathrm{e}^{x_i/b}}{\sum_{i=1}^{r}{}^{\star} \mathrm{e}^{x_i/b}} - b - \frac{1}{r}\sum_{i=1}^{r} x_i && (82)
\end{aligned}
$$

Only numerical solutions are available. The same is true for the Weibull formulation.

7 Type I Censoring

¶ Briefly, type I censoring occurs when the experiments are run only for a fixed duration, L, so the lifetimes are known only for those individuals whose lifetimes are $\leq L$.

¶ More precisely, consider a population of n individuals subjected to periods of *known and predetermined* observation $L_1, \ldots, L_n$, and with lifetimes $T_1, \ldots, T_n$.

The ith individual's lifetime is observed only if $T_i \leq L_i$.

For instance, trials stop on a specified date, but different individuals start at different specified times.

¶ Type I is different from type II in that in type I censoring the number of observed lifetimes is a random variable, unlike in type II censoring, as well as the lifetimes themselves.

¶ **Notation** for type I censoring.
$n =$ number of individuals.

$L_i =$ censoring time for the ith individual.

$T_i =$ lifetime for the ith individual.
We don't necessarily observe T_i. What we observe is t_i:

$$t_i = \min(T_i,\ L_i) \tag{83}$$

$$\delta_i = \begin{cases} 1 & T_i \leq L_i \\ 0 & T_i > L_i \end{cases} \tag{84}$$

¶ **Assumption:** The T_is are iid with pdf $f(t)$ and survivor function $S(t)$.

¶ The joint pdf for t_i and δ_i is:

$$\text{Prob}(t_i,\ \delta_i) = f(t_i)^{\delta_i} S(L_i)^{1-\delta_i} \tag{85}$$

Explanation:

1.

$$\begin{aligned} \text{Prob}(t_i = L_i) &= \text{Prob}(\delta_i = 0) && (86)\\ &= \text{Prob}(T_i > L_i) && (87)\\ &= S(L_i) && (88) \end{aligned}$$

2. For $t_i < L_i$:

$$\begin{aligned} \text{Prob}(t_i|\delta_i = 1) &= \text{Prob}(t_i|T_i < L_i) && (89)\\ &= \frac{f(t_i)}{1 - S(L_i)} && (90) \end{aligned}$$

(Recall the definition of conditional probability: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.)

3. Thus:

$$\text{Prob}(t_i = L_i,\ \delta_i = 0) = \text{Prob}(\delta_i = 0) = S(L_i) \tag{91}$$

and

$$\begin{aligned} \text{Prob}(t_i,\ \delta_i = 1) &= \underbrace{\text{Prob}(t_i|\delta_i = 1)}_{\frac{f(t_i)}{1-S(L_i)}} \underbrace{\text{Prob}(\delta_i = 1)}_{1-S(L_i)} && (92)\\ &= f(t_i) && (93) \end{aligned}$$

Combining eqs.(91) and (93) gives eq.(85). ■

¶ Now given n independent pairs $(t_i,\ \delta_i)$, $i = 1,\ \ldots, n$, the joint pdf is:

$$f_n(t_1, \delta_1,\ \ldots, t_n, \delta_n) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(L_i)^{1-\delta_i} \tag{94}$$

This is the likelihood function, $L(\lambda)$.

¶ **Example.** Suppose, as before that t is exponentially distributed:

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0 \tag{95}$$

$$S(t) = e^{-\lambda t} \tag{96}$$

The likelihood function becomes:

$$L(\lambda) = \prod_{i=1}^{n} \left(\lambda e^{-\lambda t_i}\right)^{\delta_i} e^{-\lambda t_i(1-\delta_i)} \tag{97}$$

$$= \lambda^r \exp\left(-\lambda \sum_{i=1}^{n} t_i\right) \tag{98}$$

where $r = \sum_{i=1}^{n} \delta_i$ is the number of observed "deaths" or failures.

What is the MLE of λ?

Let $T = \sum_{i=1}^{n} t_i$, so $L(\lambda) = \lambda^r e^{-\lambda T}$.

Thus:

$$0 = \frac{dL}{d\lambda} = e^{-\lambda T}\left[r\lambda^{r-1} - \lambda^r T\right] \quad \Longrightarrow \quad \widehat{\lambda} = \frac{r}{T} \tag{99}$$

This is formally the same as eq.(28) on p. 9, though T is defined differently. ■

¶ Compare the likelihood functions for types I and II censoring:

$$L_{II} = \frac{n!}{(n-r)!} f(t_{(1)}) \cdots f(t_{(r)}) [S(t_{(r)})]^{n-r} \tag{100}$$

$$L_I = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(L_i)^{1-\delta_i} \tag{101}$$

L_I from eq.(94) on p. 19. L_{II} from eq.(14) on p. 7.

For L_I:

- Each observed lifetime ($\delta_i = 1$) contributed a factor $f(t_i)$.
- Each censored lifetime ($\delta_i = 0$) contributed a factor $S(L_i)$.

Thus L_I is similar in form to L_{II}, though different in origin and precise structure.

8 Random Censoring

¶ In type I censoring we assume the censoring times $L_1, \ldots, L_n$ are known and predetermined.

In random censoring the individuals start at random times, so both the lifetimes and the censoring times are random.

¶ Define:
$T_i =$ lifetime of ith individual.
$L_i =$ censoring time of ith individual.

Assume:
T_i and L_i are independent random variables.
$T_1, \ldots, T_n$ are iid with pdf $f(t)$ and survivor function $S(t)$.
$L_1, \ldots, L_n$ are iid with pdf $g(t)$ and survivor function $G(t)$.
That is:

$$\text{Prob}(T) = f(t) \tag{102}$$

$$\text{Prob}(T > t) = S(t) \tag{103}$$

$$\text{Prob}(L) = g(t) \tag{104}$$

$$\text{Prob}(L > \ell) = G(\ell) \tag{105}$$

Define as before:

$$t_i = \min(T_i,\ L_i) \tag{106}$$

$$\delta_i = \begin{cases} 1 & T_i \le L_i \\ 0 & T_i > L_i \end{cases} \tag{107}$$

The pdf for ($t_i,\, \delta_t$) is:

$$\begin{aligned} \text{Prob}(t_i = t,\, \delta_i = 0) &= \text{Prob}(L_i = t,\, T_i > L_i) && (108)\\ &= g(t)S(t) && (109)\\ \text{Prob}(t_i = t,\, \delta_i = 1) &= \text{Prob}(L_i = t,\, T_i \le L_i) && (110)\\ &= f(t)G(t) && (111) \end{aligned}$$

Combine eqs.(108)–(111) as:

$$\text{Prob}(t_i = t,\, \delta_i) = [f(t)G(t)]^{\delta_i}[g(t)S(t)]^{1-\delta_i} \tag{112}$$

So, for n individuals with observations $(t_1,\, \delta_1),\, \ldots,(t_n,\, \delta_n)$, the likelihood function is:

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n}[f(t_i)G(t_i)]^{\delta_i}[g(t_i)S(t_i)]^{1-\delta_i} && (113)\\ &= \underbrace{\left(\prod_{i=1}^{n} G(t_i)^{\delta_i} g(t_i)^{1-\delta_i}\right)}_{\text{Depends on censored r.v.s}} \underbrace{\left(\prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}\right)}_{\text{Depends on lifetime r.v.s}} && (114) \end{aligned}$$

It may happen that G and g, which express the censoring random variables, do not depend on parameters of interest. In that case, the likelihood function in eq.(114) is effectively the same as the likelihood function for type I censoring, eq.(94) on p. 19.