Reliability theory

Gunnar Englund
Matematisk statistik
KTH

2010

1 Calculation of moments

Theorem: If $T \geq 0$ has the distribution function $F(t),\ t \geq 0$ then we have if $E(T^r) < \infty$ that

$$E(T^r) = r \int_0^\infty t^{r-1}(1 - F(t))dt$$

□

Proof: T has the density $f(t) = F'(t)$ and

$$E(T^r) = \int_0^\infty t^r f(t) = \text{(partial integration)} = \Big|_0^\infty t^r(F(t) - 1) - \int_0^\infty rt^{r-1}(F(t) - 1)dt.$$

But since $E(T^r) < \infty$ we have

$$\int_A^\infty t^r f(t)dt \to 0 \text{ when } A \to \infty$$

and we get

$$\int_A^\infty t^r f(t)dt \geq A^r \int_A^\infty f(t)dt = A^r P(T > A).$$

Therefore we have $A^r(1 - F(A)) \to 0$ when $A \to \infty$, i.e. the first part is 0. □

Useful formulas are

$$E(T) = \int_0^\infty (1 - F(t))dt \text{ and } E(T^2) = 2\int_0^\infty t(1 - F(t))dt.$$

We let $1 - F(t) = R(t) = P(T > t)$ and $R(t)$ is called the survival function.

2 Failure rate

We have

$$P(x < T \leq x + h | T > x) = \frac{P(x < T \leq x + h)}{P(T > x)} = \frac{F(x+h) - F(x)}{R(x)} \approx \frac{f(x)h}{R(x)}.$$

We therefore define the failure rate as this proportionality factor.

$$\lambda(x) = \frac{f(x)}{R(x)},$$

and the interpretation is that $P(x < T \leq x+h|T > x) \approx \lambda(x)h$ i.e. $\lambda(x)$ is the probability /time unit for failures at time x when the component has survived until the time x.
Since

$$\lambda(x) = \frac{f(x)}{R(x)} = -\frac{d}{dx}\ln(R(x)),$$

we see that

$$R(x) = \exp(-\int_0^x \lambda(u)du) \text{ and } -\ln(R(x)) = \int_0^t \lambda(u)du.$$

For the $Exp(\lambda)$-distribution we have the density $f(x) = \lambda e^{-\lambda x}$ and the survival function $R(x) = e^{-\lambda x}$ implying $\lambda(x) = \lambda$, i.e the failure rate is constant.

If the life length has a Weibull-distribution, i.e. has the survival function $R(x) = \exp(-(\lambda x)^c)$ we get the density $f(x) = c\lambda^c x^{c-1}\exp(-(\lambda x)^c)$ and we have the failure rate $\lambda(x) = c\lambda^c x^{c-1}$. This decreases in x if $0 < c < 1$ and increases in x if $c > 1$. For $c = 1$ it is constant and this corresponds to the exponential distribution.

3 Increasing and decreasing failure rate – IFR and DFR

We say that the distribution has increasing failure rate or is IFR (Increasing Failure Rate) if the failure rate $\lambda(x)$ increases in x. In the same manner we define DFR (Decreasing Failure Rate) if $\lambda(x)$ decreases in x.

A more general definition is to say that the distribution is IFR if $-\ln(R(t))$ is convex in t. Convex functions lie under their cordasm i.e. they lie under straight lines between points on the curve.

Unfortunately a system consisting of independent components which are IFR may not have an increasing failure rate. We define the class
$IFRA$ (A for Average) if

$$\frac{1}{t}\int_0^t \lambda(u)du \text{ increases in } t.$$

We will late see that systems made from independent $IFRA$-components are $IFRA$.

Theorem: $IFR \Rightarrow IFRA$

Proof 1: Let $t_1 \leq t_2$. The distribution is IFR means that $\lambda(t_1) \leq \lambda(t_2)$. We want to prove that

$$\frac{1}{t_1}\int_0^{t_1} \lambda(u)du \leq \frac{1}{t_2}\int_0^{t_2} \lambda(u)du.$$

We get

$$\frac{1}{t_2}\int_0^{t_2} \lambda(u)du = (\text{change of variable } v = \frac{t_1}{t_2}u) = \frac{1}{t_1}\int_0^{t_1} \lambda(\frac{t_2}{t_1}v)dv \geq$$

$$\geq (\text{ty } t_2/t_1 \geq 1 \text{ and } \lambda(u) \text{ increasing}) \geq \frac{1}{t_1}\int_0^{t_1} \lambda(v)dv$$

□

Proof 2: We assume that $-\ln(R(t))$ is convex. It starts in 0 since $R(0) = 1$ and therefore (draw a figure)

$$-\ln(R(t_1)) \leq \frac{t_1}{t_2}(-\ln(R(t_2)))$$

which easily implies

$$\frac{1}{t_1}\int_0^{t_1} \lambda(u)du \leq \frac{1}{t_2}\int_0^{t_2} \lambda(u)du$$

since $\ln(R(t)) = -\int_0^t \lambda(u)du$. □

4 Memorylessness for the exponential distribution

For the exponential distribution $Exp(\lambda)$ we have the failure rate $\lambda(t) = \lambda$, i.e. is does not depent on t. The exponetial distribution is the only distribution which satisfies this. We see this since the distribution is determined by the failre rate.
We have an interesting memorylessness property for the exponential distribution:

$$P(T > x + y|T > x) = \frac{P(T > x + y; T > x)}{P(T > x)} = \frac{P(T > x + y)}{P(T > x)} =$$

$$\frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda y} = P(T > y).$$

This can be interpreted as follows: A component that has an exponential distribution does not age – if it has survived until time x the probability that is will function at least y time units more, is the same as the probability that a new component will function at least y time units. This follows from the fact that the failure rate is constant. It is therefore idiotic to change a component with exponential life length since the "old" component is just as good as the "new" one.

An alternative way of expressing this equality is to use the definition of conditional probability and write is as $P(T > x + y) = P(T > x)P(T > y)$.

5 The classes NBU and NWU

We define a class of distributions NBU (New Better than Used) for distributions where we gain something by switching a used component.

Definition: The distribution is called NBU if $P(T > x+y) \leq P(T > x)P(T > y)$.

In conditional form we write this as $P(T > x + y|T > x) \leq P(T > y)$ i.e. with the interpretation that we have a greater probability of no failure in y time units if we switch to a new component. It pays off if we have strict inequality.

There is a dual class NWU (New Worse than Used) if the inequality goes the other way.

If the life length of the component is NWU it is stupid to switch since a new component is worse that the "old" one. This should of course not be interpreted as if the component get better as time goes by, but in stead the population initially consists of both good and bad components. As time passes without the component failing we obviously have got a good component. If we replace it with a new one we run the risk of installing a bad component.

Sats: $IFRA \Rightarrow NBU$

If the distribution is $IFRA$ we have (for $x, y > 0$)

$$\frac{1}{x+y}\int_0^{x+y} \lambda(u)du \geq \frac{1}{x}\int_0^x \lambda(u)du$$

We now assume that $y \leq x$ (otherwise we interchange them). We get

$$\int_0^{x+y} \lambda(u)du \geq \frac{x+y}{x}\int_0^x \lambda(u)du =$$
$$\int_0^x \lambda(u)du + \frac{y}{x}\int_0^x \lambda(u)du \geq \text{(ty } IFRA \text{ and } x \geq y)$$
$$\int_0^x \lambda(u)du + \frac{y}{y}\int_0^y \lambda(u)du = \int_0^x \lambda(u)du + \int_0^y \lambda(u)du.$$

This yields $P(T > x + y) \leq P(T > x)P(T > y)$, i.e. it is NBU.

Therefore we have $IFR \Rightarrow IFRA \Rightarrow NBU$ and $DFR \Rightarrow DFRA \Rightarrow NWU$. There are counterexamples if we try to change the implications.

6 The Poisson-process

We want to produce a model for the concept "event which occur completely ramdomly in time". We divide the interval $(0, t]$ in $n = t/h$ intervals of length h each. The idea is to let $h \to 0$.

a) In each such interval we assume that the probability for an event is $\lambda h + o(h)$, i.e. in principle proportional to the interval length. $o(h)$ denotes a remainder term which is smaller in magnitude than h.

b) In each such intervalthe probability for no occurrance is $1 - \lambda h + o(h)$

c) The probability of 2 or more occurrances in an interval is $o(h)$ and can therefore be ignored.

d) The number of occurances in disjoint interval are independent.

Let $X(t)$ =the numer of occurrances in the interval $(0, t]$. We have the standard situation for the binomial distribution and therefore $X(t)$ is $Bin(n, \lambda h + o(h))$ which is approximately $Po(n\lambda h) = Po(\lambda t)$. In the same manner we se that the number of occurrances in the interval $(s, t]$ is $Po(\lambda(t - s))$. We also see that the increments on disjoint time intervals are independent of each other.
If we let T =time until the first occurrance we get

$$P(T > t) = P(X(t) = 0) = e^{-\lambda t}$$

and therefore the time until the first occurrance is $Exp(\lambda)$. The times between occurances are indpoendent Tiderna mellan händelserna är oberoende $Exp(\lambda)$-distributed random variables. If we let

$$S_n = T_1 + T_2 + \cdots + T_n = \text{time until the } n\text{:th occurrance}$$

we get in the same manner

$$P(S_n > t) = P(X(t) \leq n - 1) = \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

If we take the derivative with respect to t the expression cancels out and we get

$$f_{S_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}.$$

which is called the $\Gamma(n, \lambda)$-distribution. It can therefore be interpreted as the distribution of the sum of n independent $Exp(\lambda)$-distributed random variables $T_1, T_2, \ldots, T_n$.

7 Analysis of life length data

We will in this section analyze data on life lengths $X_1, X_2, \ldots, X_n$ for n identical components where we let the X_i:s be independent and identically distributed. Sometimes we have complete data, sometimes only "censored" data,

i.e. just some of the life lengths. E.g. we can imagine that the experiment is stopped at a fixed time t_0 (Type I-censoring) or that the test is stopped when we observe r failures (Type II-censoring).

We shall produce the so-called TTT-plot and the TTT-transform but also estimate the failure rate (if we assume it is constant) and we will also estimate the survival function.

Let $X_1, \ldots, X_n$ be the lifelengths of n components which we assume are independent and identically distributed. Let $X(1), X(2), \ldots, X(n)$ be these sorted in increasing size. Therefore $X(1)$ is the smallest of $X_1, X_2, \ldots, X_n$ and $X(2)$ the second smallest etc untill $X(n)$ which is the largest. These $X(i)$ are called the ordes statistics.

We assume that F has a density and is strictly increasing and that $E(X_i) = \theta$.

Definition: The Total Time on Test at time x where $X(i) \leq x < X(i+1)$ is

$$T(x) = \sum_{j=1}^{i} X(j) + (n-i)x.$$

□

The first term is the testing time of those components which have failed before time x and the last term $(n-i)x$ is the testing time for the components which have survived past time. If we have numerical data we denote them $x_1, x_2, \ldots, x_n$ which we will see as outcomes of random variables $X_1, X_2, \ldots, X_n$ but we will feel free to switch between these as feel appropriate.

We have

$$T(X(i)) = \sum_{j=1}^{i} X(j) + (n-i)X(i)$$

and in particular $T(X(n)) = \sum_{j=1}^{n} X(j) = \sum_{j=1}^{n} X_n$. Therefore we have $T(X(n))/n = \bar{X}$.
The relative test time when the i:th failure occurs is

$$\frac{T(X(i))}{T(X(n))}$$

and the TTT-plotis a plot of $(i/n, T(X(i))/T(X(n)))$. We will show that

1) If data is from an exponential distribution the TTT-plot is approximately linear.

2) If data is from an IFR-distribution the TTT-plot is approximately concave.

3) If data is from a DFR-distribution the TTT-plot is approximately convex.

The word "approximately" is used since we have a certain randomness in the TTT-plot.

8 More properties of the exponential distribution

If we know that $X(a) = 1$ we know that one event has occurred somewhere during the time interval $(0, a]$. Letting T =time for the first occurrance we get for $0 \leq t \leq a$

$$P(T \leq t | X(a) = 1) = \frac{P(T \leq t; X(a) = 1)}{P(X(a) = 1)} =$$
$$\frac{P(X(t) = 1; X(a) = 1)}{P(X(a) = 1)} = \frac{P(X(t) = 1; X(a) - X(t) = 0)}{P(X(a) = 1)} =$$
$$\text{(independence for dijoint intervals)} = \frac{P(X(t) = 1)P(X(a) - X(t) = 1)}{P(X(a) = 1)} =$$
$$\frac{\lambda t e^{-\lambda t} e^{-\lambda(a-t)}}{\lambda a e^{-\lambda a}} = \frac{t}{a},$$

which means that $T|X(a) = 1$ is uniformly distributed $U(0, a)$. In the same manner we can show that if we have n occurrances on the interval $(0, a]$ these times have the same distribution as the order statistics for n independent uniformly distributed independent random variables. Therefore the times occur completely randomly in the interval.

Theorem: If X_i is $Exp(\lambda_i)$ for $i = 1, 2, \ldots, n$ and independent
then $T = \min(X_1, X_2, \ldots, X_n)$ is exponentially distributed $Exp(\sum_1^n \lambda_i)$. □

Proof:

$$P(T > t) = P(\min(X_1, X_2, \ldots, X_n) > t) = P(X_1 > t; X_2 > t; \ldots; X_n > t) =$$
$$\text{(independence)} = P(X_1 > t)P(X_2 > t) \ldots P(X_n > t) =$$
$$\exp(-\lambda_1 t) \exp(-\lambda_2 t) \ldots \exp(-\lambda_n t) = \exp(-t \sum_{i=1}^{n} \lambda_i),$$

which shows that T is $Exp(\sum_1^n \lambda_i)$. □

The intensities in the exponential distributions add up when we form the minimum of independent exponential distributions.

One can also show that if $L =$ numbe rof the one which is the smallets we have $P(L = j) = \lambda_j / \sum_1^n \lambda_i$ and that L (surprisingly enough) is independent of T.

Theorem: If $X_1, X_2, \ldots, X_n$ are independent $Exp(\lambda)$ then $T(X(i))/T(X(n))$ for $i = 1, 2, \ldots, n-1$ are distributed as the order statistics $U(1), U(2), \ldots, U(n-1)$ of $n - 1$ independent $U(0, 1)$-distributed $U_1, U_2, \ldots, U_{n-1}$. □

Sketch of proof: Note that $X(1)$ is the minimum of n independent $Exp(\lambda)$-distributed variables and therefore is $Exp(n\lambda)$. Furthermore we have $T(X(1)) = nX(1)$ so $T(X(1)) - 0$ is $Exp(\lambda)$. In the same manner we see that $X(2) - X(1)$ is

the minimum of $n-1$ independent $Exp(\lambda)$ and therefore $T(X(2))-T(X(1)) = (n-1)(X(2)-X(1))$ is $Exp(\lambda)$ etc. We see thatatt

$$T(X(1)), T(X(2)) - T(X(1)) \ldots, T(X(n)) - T(X(n-1)) \text{ etc}$$

are independent $Exp(\lambda)$ and therefore $T((X(j))$ is distributed as the sum of j independent $Exp(\lambda)$ and therefore consist of the times for the successive events in a Poisson process. If we fix $T(X(n)) = a$ we can use the result above to see that we should place $n-1$ events in the interval $(0, a]$. □

From this we see:

$$E(\frac{T(X(i))}{T(X(n))}) = \frac{i}{n}$$

and we can show that the variance is small. This means that

$$T(X(i))/T(X(n)) \approx i/n.$$

For the exponential distribution the TTT-plot should give approximately a straight line. This can be used as a graphical test of whether the data come from an exponential distribution.

9 Empirical distribution

With data $x_1, x_2, \ldots, x_n$ which are outcomes of independent identically distributed variables $X_1, X_2, \ldots, X_n$ we place the probability mass $1/n$ in each of the points $x_1, \ldots, x_n$. If 2 x_i:s are identical we get the mass $2/n$ in that point, etc. This distribution is called the empirical distribution and we denote its distribution function by $F_n(x)$. If n is large it resembles the true distribution F of the X_i:s, i.e. we have $F_n(x) \approx F(x)$. Formally we have

$$F_n(x) = \begin{cases} 0 \text{ if } x < x(1) \\ i/n \text{ if } x(i) \leq x < x(i+1) \\ 1 \text{ if } x > x(n) \end{cases}$$

We now have the following theorem which connects the TTT-plot with the empirical distribution.

Theorem: We have

$$T(x(i)) = n \int_0^{x(i)} (1 - F_n(u)) du.$$

Proof: From a graph we see that the area under $F_n(x)$ up to time $x(i)$ is

$$\frac{1}{n} \sum_{k=1}^{i} (x(i) - x(k)) = \frac{i}{n} x(i) - \frac{1}{n} \sum_{k=1}^{i} x(k),$$

where we calculate the area as the sum of the horizontal strips. The areas under $1 - F_n(x)$ up to $x(i)$ is

$$x(i) \cdot 1 - \text{Area under } F_n \text{ up to } x(i)$$

and therefore

$$n \int_0^{x(i)} (1 - F_n(u))du = nx(i) - ix(i) + \sum_{j=1}^{i} x(j) = T(x(i)).$$

If we now let $n \to \infty$ so we have F_n close to F and take $i/n = v$ with $0 \leq v \leq 1$ we see that $x(i) \to F^{-1}(v)$, i.e. the v-fractile in the distribution. We get

$$\frac{T(x(i))}{n} \to \int_0^{F^{-1}(v)} (1 - F(u))du.$$

In particular with $v = 1$ i.e. $i = n$ we get $F^{-1}(v) = \infty$ and the left hand side is $\bar{x}$ while the right hand side is $E(X)$ corresponding to the law of large numbers.

Definition: The TTT-transform of a distribution is

$$H_F^{-1}(v) = \int_0^{F^{-1}(v)} (1 - F(u))du.$$

In particular $H_F^{-1}(1)$ is the expected value of the distribution F. □

The scaled TTT-transform is $H_F^{-1}(v)/H_F^{-1}(1)$.

Example: If $F(x) = 1 - e^{-\lambda x}$, i.e. the $Exp(\lambda)$-distribution we get the inverse function F^{-1} by solving for x in $y = F(x)$. We get $x = -\frac{1}{\lambda}\ln(1 - y)$ so $F^{-1}(v) = -\frac{1}{\lambda}\ln(1 - v)$. We have

$$H_F^{-1}(v) = \int_0^{-\frac{1}{\lambda}\ln(1-v)} e^{-\lambda u} du = v/\lambda.$$

Since $H_F^{-1}(1) = 1/\lambda$ we get $H_F^{-1}(v)/H_F^{-1}(1) = v$ for $0 \leq v \leq 1$. □

Theorem: I F is continuous and strictly increasing we have

$$\frac{d}{dv} H_F^{-1}(v)\Big|_{v=F(x)} = \frac{1}{\lambda(x)}$$

where $\lambda(x)$ is the failure rate, i.e. $\lambda(x) = F'(x)/(1 - F(x))$. □

Proof: If we let $G(y) = \int_0^y (1 - F(u))du$ we have $G'(y) = 1 - F(y)$ so by the chain rule

$$\frac{d}{dv} G(F^{-1}(v)) = G'(F^{-1}(v)) \frac{d}{dv} F^{-1}(v),$$

but $u = F^{-1}(v)$, i.e. $v = F(u)$ so $dv = F'(u)du$ and therefore

$$\frac{d}{dv}F^{-1}(v) = \frac{du}{dv} = 1/F'(u) = 1/F'(F^{-1}(v)).$$

In all this yields (with $f = F'$)

$$\frac{d}{dv}H_F^{-1}(v) = (1 - F(F^{-1}(v)))\frac{1}{f(F^{-1}(v))} = \frac{1-v}{f(F^{-1}(v))}.$$

Letting $v = F(x)$ we get

$$\frac{d}{dv}H_F^{-1}(v)\Big|_{v=F(x)} = \frac{1-F(x)}{f(x)} = \frac{1}{\lambda(x)}.$$

□

Therefore we get the following theorem

Theorem:
F is IFR if and only if $H_F^{-1}(v)$ is concave
F is DFR if and only if $H_F^{-1}(v)$ is convex. □

Theorem: F is $IFR \Leftrightarrow \lambda(x)$ increases in $x \Leftrightarrow 1/\lambda(x)$ decreases in $x \Leftrightarrow \frac{d}{dv}H_F^{-1}(v)|_{v=F(x)}$ decreases in $x \Leftrightarrow \frac{d}{dv}H_F^{-1}(v)$ decreases in $v \Leftrightarrow H_F^{-1}(v)$ is concave. □

If we now estimate H_F^{-1} from data using the empiriska distribution function and let

$$H_n^{-1}(v) = \int_0^{F_n^{-1}(v)} (1 - F_n(u))du,$$

we get since $F_n^{-1}(i/n) = x(i)$ and since, as previously shown, $T(x(i)) = n\int_0^{x(i)}(1 - F_n(x))dx$ so we get

$$\frac{H_n^{-1}(i/n)}{H_n^{-1}(1)} = \frac{T(x(i))}{T(x(n))}$$

and we see that the TTT-plot is concave if the distribution is IFR and convex if the distributiojn is DFR. We can use the empirical TTT-plot to decide if the distribution is IFR, DFR or neither. If it has an S-shape it is neither IFR not DFR. The bathtub curve corresponds to the TTT-plot first being convex, then linear and then concave. Note though that the randomness in data can destroy this nice pattern.

10 Estimation of failure rate

We now assume that the life lengths are exponentially distributed $Exp(\lambda)$ and we want to estimate the failure rate λ and calculate confidence intervals for this parameter.

We have the density function $f(x)\lambda e^{-\lambda x}$, $x > 0$. We get from observed data $x_1, x_2, \ldots, x_n$ the likelihood-function

$$L(\lambda) = f(x_1)f(x_2)\ldots f(x_n) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i)$$

and by taking logarithms

$$\ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

showing

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i,$$

so maximum is produced by $\lambda = n/\sum_{i=1}^{n} x_j$. Therefore the ML-estimate of λ

$$\lambda^* = \frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{n}{T(X(n))}.$$

We know that $T(X(n)) = \sum_1^n X_j$ is $\Gamma(n, \lambda)$, i.e. has the density

$$f_{T(X(n))}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} \exp(-\lambda t).$$

Itg is suitable to introduce the Γ-function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

where we can easily show that for positive integers n we have $\Gamma(n) = (n-1)!$. We therefore get with $Y = \sum_1^n X_i$

$$E(\lambda^*) = E(\frac{n}{\sum_{i=1}^{n} X_i}) = nE(\frac{1}{Y}) = n \int_0^\infty \frac{\lambda^n y^{n-1} e^{-\lambda y}}{y\Gamma(n)} dy =$$
$$\frac{n\lambda}{\Gamma(n)} \int_0^\infty (\lambda y)^{n-2} e^{-\lambda y} \lambda dy = (\text{change of variable } \lambda y = t) =$$
$$\frac{n\lambda}{\Gamma(n)} \int_0^\infty t^{n-2} e^{-t} dt = \frac{(n-2)! n\lambda}{(n-1)!} = \frac{n}{n-1}\lambda.$$

Hence λ^* is not unbiased but asymptotically unbiased. The estimate

$$\widehat{\lambda} = \frac{n-1}{T(X(n))}$$

is therefore unbiased. Using the same technique we can show that

$$V(\widehat{\lambda}) = \frac{\lambda^2}{n-2}.$$

11 Confidence interval using complete data

Theorem:

a) X is $Exp(\lambda) \Leftrightarrow \lambda X$ is $Exp(1)$.

b) Y ia $\Gamma(n,1) \Leftrightarrow 2Y$ is $\chi^2(2n)$. □

Proof: a) $P(\lambda X > x) = P(X > x/\lambda) = \exp(-\lambda x/\lambda) = \exp(-x)$.

b) The $\chi^2(f)$-distribution has the density

$$\frac{x^{f/2-1}e^{-x/2}}{\Gamma(f/2)2^{f/2}}, \text{ for } x > 0,$$

which proves it since $P(2Y \leq x) = P(Y \leq x/2)$ and taking derivatives

$$f_{2Y}(x) = f_Y(x/2)/2 = \frac{(x/2)^{n-1}}{2\Gamma(n)}\exp(-x/2).$$

□

This means that we do not need to have a table of percentiles for the Γ-distributions but canuse the tables for the χ^2-distributions. In Matlab we have the Γ-distribution percentiles using the function gaminv, but since this is in the Stats-module we may not have access to it.

If $X_1, X_2, \ldots, X_n$ are independent $Exp(\lambda)$ we have

$$2\lambda\sum_{i=1}^{n} X_i \text{ är } \chi^2(2n).$$

Therefore

$$1-\alpha = P\left(\chi^2_{1-\alpha/2}(2n) \leq 2\lambda\sum_{i=1}^{n} X_i \leq \chi^2_{\alpha/2}(2n)\right),$$

which produces a confidence interval for λ with the confidence lever $1-\alpha$

$$\left(\frac{\chi^2_{1-\alpha/2}(2n)}{2\sum_{i=1}^{n} x_i}, \frac{\chi^2_{\alpha/2}(2n)}{2\sum_{i=1}^{n} x_i}\right) = \left(\frac{\chi^2_{1-\alpha/2}(2n)}{2T(x(n))}, \frac{\chi^2_{\alpha/2}(2n)}{2T(x(n))}\right).$$

Note that if we want a confidence interval for $m = 1/\lambda$ we only have to take 1/the limits above. In general we see that if we have a confidence interval for a parameter θ and want a confidence interval for $\psi = g(\theta)$ where g is a monotone (increasing or decreasing function) we just takes g of the limits in the interval for θ. Possible the limits can be switched as for $1/\lambda$ above.

12 Censored data

We describe 4 different types of censoring, i.e. situations where we do not have complete data.

Type I-censoring: The test is ended at a fixed time t_0.

Type II-censoring: The test is ended when we have r failures in the units, i.e. at time $X(r)$.

Type III-censoring: The test is ended at $\min(t_0, X(r))$, i.e. at time t_0 or (if it occurs earlier) when r have failed.

Type IV-censoring: Independent stochastic censoring of the units, i.e. each observation is censored randomly. For a censored observation we only know that the life length is larger than the censoring time.

We will in this section treat Type I and Type II.We will also for these study two cases:

a) With replacement where failed components are replaced by new components.

b) Without replacement where failed components are not replaced.

We are interested in the total time on test TTT, since it can be used for estimation.

Type IV-censoring is treated in the next section which treats estimation of survival function and failure rate.

12.1 Type II censoring without replacement

The test is stopped when r units have failed. We therefore observe $x(1), x(2), \ldots, x(r)$ where r is fixed. The rest of the units are censored – we only know that they fail after $X(r)$ when the test was stopped. We therefore have r failures in n tested units and the failed units have not been replaced. The total time on test is

$$T = \sum_{j=1}^{r} x(j) + (n-r)x(r).$$

We observe the r smallest of $x_1, \ldots, x_n$ and for the remaining $n-r$ we only observe that they are greater that the r:th largest, i.e. $x(r)$. The likelihoodfunction consists first of the product of the densities in $x(1), x(2), \ldots, x(r)$ and second the probability that the remaining $n-r$ are greater than $x(r)$. This last probability is $P(X_i > x(r)) = \exp(-\lambda x(r))$ and we get the likelihoodfunction

$$L(\lambda) = \lambda^r \exp(-\lambda \sum_{j=1}^{r} x(j)) \exp(-\lambda(n-r)x(r)) = \lambda^r \exp(-\lambda T(x(r)))$$

where (as before) $T(x(r))$ is the total time on test up to the r:th failure. The ML-estimate is

$$\lambda^* = \frac{r}{T(x(r))},$$

where $T(X(r)) = \sum_{j=1}^{r} X(j) + (n-r)X(r)$. The time from 0 to $X(1)$ is the minimum of n independent $Exp(\lambda)$ and is therefore $Exp(n\lambda)$. In the same

manner $X(2) - X(1)$ is the minimum of $n-1$ independent $Exp(\lambda)$ and is therefore $Exp((n-1)\lambda)$. In general $D_j = X(j) - X(j-1)$ is the minimum of $n-j+1$ independent $Exp(\lambda)$ and is therefore $Exp((n-j+1)\lambda)$ for $j = 1, 2, \ldots, r$. But

$$T(X(r)) = nD_1 + (n-1)D_2 + \cdots + (n-r+1)D_r$$

and this means that $\lambda T(X(r))$ is $\Gamma(r,1)$. We have (as for complete data) that $2\lambda T(X(r))$ is $\chi^2(2r)$. We also see that λ^* is not unbiased but that $E(\lambda^*) = r\lambda/(r-1)$. We get the same type of confidence interval for λ as for complete data with minor modifications

$$\left(\frac{\chi^2_{1-\alpha/2}(2r)}{2T(x(r))}, \frac{\chi^2_{\alpha/2}(2r)}{2T(x(r))}\right).$$

12.2 Type II-censoring with replacement

We have total test time $T(x(r)) = nx(r)$ since all n sockets are filled all the time until we stop the test when r failures have been observed. The times between failurs ($D_j = X(j) - X(j-1)$ for $j = 1, 2, \ldots, r$) are the minimum of n independent $Exp(\lambda)$ and therefore $Exp(n\lambda)$ and the total time on test can be written

$$T(X(j)) = n(D_1 + D_2 + \cdots + D_r).$$

Again we see that $T(X(r))$ is $\Gamma(r,\lambda)$ and therefore $2\lambda T(X(r))$ is $\chi^2(2r)$. We get the same confidence interval as in the previous case.

12.3 Type I censoring with replacement

As soon as a unit has failed it is replaced with a new one. Therefore we have n units under testing the whole time. The total time on test is therefore $T = nt_0$. The number of observed failures is s which is random.

We get the ML-estimate $\lambda^* = s/(nt_0)$ where s is the number of observed failures. Note that s is random and an outcome of S which is $Po(\lambda nt_0)$. Therefore λ^* is unbiased since $E(S) = \lambda nt_0$. Furthermoe we have $V(S) = \lambda nt_0$ which is estimated by $\lambda^* nt_0 = s$ and we get the standard error $d(s) = \sqrt{s}$.

We could do an approximate confidence interval by using a normal approximation of the Poisson distribution and get the approximate interval $s \pm \lambda_{\alpha/2}\sqrt{s}$ for λnt_0 i.e. the interval

$$\frac{s}{nt_0} \pm \frac{\lambda_{\alpha/2}\sqrt{s}}{nt_0}$$

for λ at least if s is large, e.g. $s \geq 15$.

Often we will ignore the fact that the censoring is of type I and use the confidence interval for Type II censoring, i.e

$$\left(\frac{\chi^2_{1-\alpha/2}(2s)}{2T(x(s))}, \frac{\chi^2_{\alpha/2}(2s)}{2T(x(s))}\right),$$

which has the approximate confidence level $1-\alpha$.

12.4 Type I-censoring without replacement

The test i stopped at time t_0. Of the n units being tested we have s (random number) failure. We observe $X(1), X(2), \ldots, X(s)$.

We have the total time on test

$$T(X(S)) = \sum_{j=1}^{S} X(j) + (n - S)t_0$$

where S is the random number of failures we observe before time t_0. The estimate $\lambda^* = s/T(x(s))$ is approximately unbiased. To calculate an exact confidence interval is tricky but we will do the same as for Type I-censoring with replacement, i.e. ignore the fact that the number of failures is random and again take the interval

$$\left(\frac{\chi^2_{1-\alpha/2}(2s)}{2T(x(s))}, \frac{\chi^2_{\alpha/2}(2s)}{2T(x(s))} \right).$$

13 Estimates of the survival function

13.1 Complete data

With complete data $x_1, x_2, \ldots, x_n$ we estimate the underlying distribution with the empirical distribution $F_n(x)$, i.e. the distribution which put the mass $1/n$ in each observation. The survival function is estimated by $\widehat{R}(x) = 1 - F_n(x)$.

We will now treat data with censoring of Type IV and will give two different estimates, the Kaplan-Meier-estimate (often called the "product limit estimator") and also the Nelson-estimator.

13.2 The Kaplan-Meier-estimator

We assume that the censoring occurs at random times, i.e. that each individual observation x_i either is the actual time for the failure och is the time up to which the component worked. We therefore have ordered observations $x(1) \leq x(2) \leq \cdots \leq x(n)$ where some of them are censored. Let ν run through those values j where $x(j)$ is a time to failure and $x(j) < x$. The Kaplan-Meier-estimator of the survival function is

$$\widehat{R}(x) = \prod_{\nu} \frac{n - \nu}{n - \nu + 1}.$$

In order to explain this, we do the following. We imagine time between 0 and x to be divided into a large number of short intervals $0 = u_0, u_1, u_2, \ldots, u_m = x$. We assume that this partition is so dense that only one of the individual occurrances failure and censored observation occurs in each interval. We get

$$\begin{aligned} R(x) = P(X > x) = P(X > u_m) = \\ P(X > u_m | X > u_{m-1}) P(X > u_{m-1} | X > u_{m-2}) \ldots \\ \ldots P(X > u_2 | X > u_1) P(X > u_1 | X > u_0). \end{aligned}$$

We will estimate these factors in the following manner

a) In an interval where nothing happens we estimate it by 1.

b) If a censored observation occurs but no failure we also estimate the factor by 1.

c) If we observe a failure and we had $n-r+1$ units working at the beginning of the interval and $n-r$ at the end we estimate the factor by $(n-r)/(n-r+1)$.

We get

$$\widehat{R}(x) = 1 \cdot 1 \cdot 1 \cdot \frac{n-\nu_1}{n-\nu_1+1} \cdot 1 \cdot \cdots \cdot 1 \cdot \frac{n-\nu_2}{n-\nu_2+1} \cdots = \prod_{\nu} \frac{n-\nu}{n-\nu+1},$$

where $x(\nu_1), x(\nu_2), \ldots$ are the times that components fail. If there are no censored observations this is identical to $1 - F_n(x)$. With censored observations the estimated survival curve jumps downward at each observed failure time by a factor which is the proportions of non-failed components of the remaining components still working at that time.

13.3 The Nelson-estimator

We let $\Lambda(x) = \int_0^x \lambda(u)du$ and see that $R(x) = e^{-\Lambda(x)}$ and therefore

$$\Lambda(x) = -\ln(R(x)) = -\ln(1 - F(x)).$$

We will try to estimate $\Lambda(x)$ from data. We star with a simple theorem

Theorem: Let X be continuous with strictly increasing distribution function F. We have

a) $U = F(X)$ is uniformly distributed on $(0, 1)$.

b) $Z = -\ln(1 - F(X))$ is $Exp(1)$-distributed. □

Proof: a) With $0 < x < 1$ we get

$$P(U \leq x) = P(F(X) \leq x) = P(X \leq F^{-1}(x)) = F(F^{-1}(x)) = x,$$

which is the distribution function for the uniform distribution on $(0, 1)$.

b) With $z > 0$ we get

$$P(Z \leq z) = P(-\ln(1 - F(X)) \leq z) = P(1 - F(X) \geq e^{-z}) = \\ P(F(X) \leq 1 - e^{-z}) = P(U \leq 1 - e^{-z}) = 1 - e^{-z},$$

which means that Z is $Exp(1)$. □

Note that the transformation producing Z is the same that appeared above in the expression for $\Lambda(x)$. This means that we in a sense transform data so that they correspond to order statitics for exponentially distributed variables.

We start with complete data, i.e. no censoring. Let $X_1, X_2, \ldots, X_n$ be independent all with the distribution F. The corresponding order statistics we denote as $X(1), X(2), \ldots, X(n)$. We now form $Z_j = -\ln(1 - F(X_j))$ as in the theorem above and note that the Z_j:s are independent $Exp(1)$. In the same way let $Z(j) = -\ln(1 - F(X(j)))$ be the order statistics for the Z:variables. They are therefore order statistics for n independent $Exp(1)$-variables. We get

$$E(Z(j)) = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-j+1}.$$

This holds since $Z(1)$ is the minimum of n independent $Exp(1)$ and therefore is $Exp(n)$, and furthermore $Z(2) - Z(1)$ is the minimum of $n-1$ independent $Exp(1)$ and therefore is $Exp(n-1)$ etc.

Note that $\Lambda(X(j)) = -\ln(1 - F(X(j))) = Z(j)$ so that $E(\Lambda(X(j))) = E(Z(j))$. We estimate $\Lambda(x)$ by

$$\widehat{\Lambda}(x) = \sum_{i=1}^{j} \frac{1}{n-i+1} \text{ when } X(j) \leq x < X(j+1)$$

which gives approximately the "correct" expected value, i.e. $E(\widehat{\Lambda}(X(j))) = E(\Lambda(X(j)))$. We then estimate $R(x)$ by $\widehat{R}(x) = e^{-\widehat{\Lambda}(x)}$.

If we have censored data we modify this so that

$$\widehat{\Lambda}(x) = \sum_{\nu} \frac{1}{n-\nu+1}$$

where ν runs through those j where $x(j)$ is a time of a failure and $x(j) < x$. This means that $\widehat{\Lambda}$ increases at times of observed failures by $1/$(number working at that time). The survival function is estimated by $\widehat{R}(x) = e^{-\widehat{\Lambda}(x)}$. As for the Kaplan-Meier-estimator we see that $\widehat{R}$ goes down by a certain factor at each observed failure. This factor $\exp(-1/(n-\nu+1))$ where $n-\nu+1$ is the number working at that time. A simple Taylor expansion shows that

$$e^{-1/(n-\nu+1)} \approx 1 - \frac{1}{n-\nu+1} = \frac{n-\nu}{n-\nu+1}$$

which means that the Kaplan-Meier-estimate and the Nelson-estimate are quite similar especially if the number of failures and censorings is small compared with n.