TPK4120 - Lecture summary

Jørn Vatn
eMail: jorn.vatn@ntnu.no

Updated 2022-09-07

Chapter 5 - Probability Distributions in Reliability Analysis

What to know from this chapter

- CDF, PDF, $z(t)$, $R(t)$, $R(x|t)$, MRL(t)
- Important distributions, Binomial, Exponential, Weibull, Normal, log-normal and Poisson
- Relation between Exponential lifetimes and a Poisson process
- Relation between Gamma distribution and the k'th event in the Poisson process

.... $\square$

This chapter deals with time-to-failure distributions for non-repairable items. The results are also valid for repairable items if we only consider what is happening up to the first failure. The reliability metrics to be covered are:

- The survivor function $R(t)$
- The failure rate function $z(t)$
- The mean time-to-failure (MTTF)
- The conditional survivor function
- The mean residual lifetime (MRL)

The textbook covers a wide range of distributions, whereas we will only consider the following distributions:

- The exponential distribution
- The gamma distribution

- The Weibull distribution
- The normal and lognormal distributions
- The Poisson distribution
- The binomial distribution

State variable

The state variable of an item was introduced in chapter 2 and chapter 4 where we only considered that an item might be in various states. Now we will also treat the stochastic behaviour of the item, hence the state variable will be a stochastic variable:

$$X_i(t) = \begin{cases} 1 \text{ if the item is functioning at time } t \\ 0 \text{ if the item is in a failed state at time } t \end{cases} \tag{1}$$

Time-to-failure

The *time-to-failure*, or *lifetime* of an item is the time elapsing from when the item is put into operation until it fails for the first time. If we denote the time-to-failure with T then

$$T = \min\{t : X(t) = 0\}$$

Note that "time" sometimes is measured indirectly, e.g., by the number of kilometres driven by a car, the number of times a switch is operated, and the number of rotations of a bearing.

Since $X(t)$ is a stochastic variable, the time-to-failure, T is also a stochastic variable. To grasp the reliability metrics we will introduce, we could relate these metrics to what we would observe if did experiments and collected the true lifetimes of the items. In the textbook several examples of such "empirical metrics" are given.

PDF and CDF

Assume that the time-to-failure T is a continuous distributed stochastic variable with probability density function (PDF) $f(t)$ and cumulative (probability) distribution function (CDF) $F(t)$. Then we have:

$$F(t) = \Pr(T \le t) = \int_0^t f(u)du$$

To interpret the PDF we have for small Δt:

$$\Pr(t < T \le t + \Delta t) \approx f(t)\Delta t$$

i.e., the probability that a new item will fail in the interval t to $t + \Delta t$ equals the PDF at time t multiplied with the length of the interval.

Survivor Function

The survivor function of an item is defined by:

$$R(t) = 1 - F(t) = \Pr(T > t) = \int_t^\infty f(u)du$$

i.e., the probability that a new item will survive the time interval $(0, t]$.

Failure Rate Function

The failure rate function is essentially the conditional probability that an item will fail in a small time interval given that it has not failed up till now. The probability that an item will fail in $(t, t+\Delta t]$ when we know that the item is functioning at time t is:

$$p(t, \Delta t) = \Pr(t < T \leq t + \Delta t | T > t) = \frac{\Pr(t < T \leq t + \Delta t)}{\Pr(T > t)} = \frac{F(t+\Delta t) - F(t)}{R(t)}$$

If we investigate the ratio $p(t, \Delta t)/\Delta t$ we get the failure rate function $z(t)$ of the item:

$$\begin{aligned} z(t) = \lim_{\Delta t \to 0} \frac{p(t, \Delta t)}{\Delta t} &= \lim_{\Delta t \to 0} \frac{\Pr(t < T \leq t + \Delta t | T > t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} \frac{1}{R(t)} = \frac{f(t)}{R(t)} \end{aligned}$$

And for small Δt:

$$\Pr(t < T \leq t + \Delta t | T > t) \approx z(t)\Delta t$$

Typical shape of $z(t)$ is shown in Figure 1:

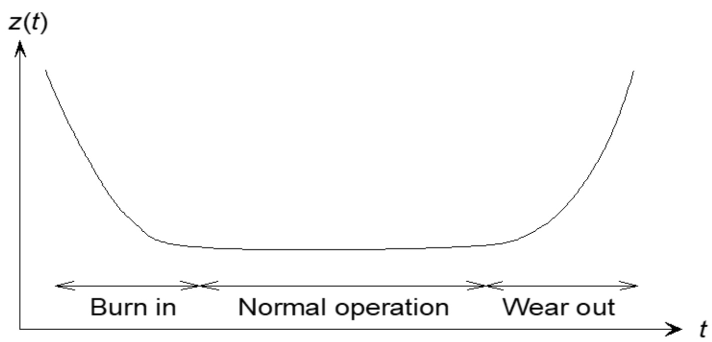

Figure 1: Typical shape of $z(t)$

hence the failure rate function is often denoted the bathtub curve. It should be noted that not all items will follow the bathtub curve!

Note, $f(t)$ is the probability of failing at time t, whereas $R(t)$ is the probability of surviving time t. In $z(t) = \frac{f(t)}{R(t)}$ we divided by $R(t)$, so even if the probability of failing at large times t is low, the fraction becomes very high since we hardly survive t.

Conditional survivor function

To express the conditional survivor function we introduce $R(x|t) = \Pr(T > x+t|T > t)$, i.e., the conditional probability of surviving another x time units given that the component has survived t time units. It follows that

$$R(x|t) = \frac{\Pr(T > x+t \cap T > t)}{\Pr(T > t)} = \frac{\Pr(T > x+t)}{\Pr(T > t)} = \frac{R(t+x)}{R(t)}$$

MTTF

The mean time to failure is the expected value of the lifetime of an item. The formula $\mathrm{E}(T) = \int_0^\infty R(t)dt$ is often easier to apply than $\mathrm{E}(T) = \int_0^\infty t \cdot f(t)dt$ when we are seeking the mean time to failure. To prove this result we use the partial integration, i.e.,

$$\int_a^b u(x)v'(x)dx = u(x)v(x)\Big|_a^b - \int_a^b u'(x)v(x)dx$$

Using t rather than x, and letting $u(t) = t$ and $v(t) = -R(t)$, we have

$$\int_0^\infty t \cdot f(t)d = \int_0^\infty u(t)v'(t)dt = u(t)v(t)\Big|_0^\infty - \int_0^\infty u'(t)v(t)dt =$$
$$-t \cdot R(t)\Big|_0^\infty - \int_0^\infty -1 \cdot R(t)dt = \int_0^\infty R(t)dt$$

where we use that $f(t) = F'(t) = -R'(t)$

Mean residual life

To obtain the mean residual life we use the conditional survivor function, i.e.,

$$\mathrm{MRL}(t) = \int_0^\infty R(x|t)dx = \int_0^\infty \frac{R(t+x)}{R(t)}dx = \frac{1}{R(t)}\int_t^\infty R(x)dx$$

Relationships between the functions $F(t)$, $f(t)$, $R(t)$, and $z(t)$

For easy reference the relationships between the functions $F(t)$, $f(t)$, $R(t)$, and $z(t)$ are given in Table 1.

Some distribution classes

Binomial Distribution

Bernoulli trial: (i) n independent trials, (ii) observing A or A^C, and (iii) $\Pr(A) = p$ in all trials. Let X be the number of A's in the experiment. X is then binomially distributed:

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n$$

Table 1: Relationship between the functions $F(t), f(t), R(t)$, and $z(t)$

Expressed by	$F(t)$	$f(t)$	$R(t)$	$z(t)$
$F(t) =$	–	$\int_0^t f(u)\,du$	$1 - R(t)$	$1 - \exp\left(-\int_0^t z(u)\,du\right)$
$f(t) =$	$\frac{d}{dt}F(t)$	–	$-\frac{d}{dt}R(t)$	$z(t) \cdot \exp\left(-\int_0^t z(u)\,du\right)$
$R(t) =$	$1 - F(t)$	$\int_t^\infty f(u)\,du$	–	$\exp\left(-\int_0^t z(u)\,du\right)$
$z(t) =$	$\frac{dF(t)/dt}{1 - F(t)}$	$\frac{f(t)}{\int_t^\infty f(u)\,du}$	$-\frac{d}{dt}\ln R(t)$	–

$$\mathrm{E}(X) = np, \qquad \mathrm{var}(X) = np(1-p)$$

To calculate $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ by hand we use $\binom{n}{x} = \frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-x)}{1\cdot 2\cdot\ldots\cdot x}$.

In MS Excel we may use $\Pr(X = x)$ `=binomdist(x,n,p,FALSE)`, $\Pr(X \le x)$ `=binomdist(x,n,p,TRUE)` and $\binom{n}{x}$ `=combin(n,x)`.

Geometric Distribution

We are considering the same situation as for the binomial distribution, but now n is not fixed, and we repeat the experiment until the first time we get an A. Let X be the number of trials until an A is received. X is then geometrically distributed:

$$\Pr(X = x) = (1-p)^{x-1}p \quad \text{for } x = 1, 2, \ldots$$

$$\mathrm{E}(X) = \frac{1}{p}, \qquad \mathrm{var}(X) = \frac{1-p}{p^2}$$

Exponential Distribution

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \\ R(t) &= e^{-\lambda t} \end{aligned}$$

$$\mathrm{E}(T) = \frac{1}{\lambda}, \qquad \mathrm{var}(T) = \frac{1}{\lambda^2}$$

$$z(t) = \lambda \text{ (= constant, independent of age)}$$

Homogeneous Poisson Process

We are considering a time interval from 0 to t. For this interval we have (i) Events, say A's, may occur at any point of time on the interval, and the probability of an event A occurring in a small interval of length Δt is independent on where this small interval is, and is given by $\lambda\Delta t$ plus a "small term", i.e., $o(\Delta t)$, (ii) the probability of two or more A's in such a small interval is "very small" i.e., $o(\Delta t)$, and (iii) the event that A occurs in one interval is independent of whether event A occurs in another non-overlapping interval.

Let $N(t)$ be the number of A's occurring in an interval of length t. $N(t)$ is Poisson distributed, i.e.,

$$\Pr(N(t)=n) = \frac{(\lambda t)^n}{n!}e^{-\lambda t} \quad \text{for } n = 0, 1, 2, \ldots$$

$$\mathrm{E}(N(t)) = \lambda t, \qquad \mathrm{var}(N(t)) = \lambda t$$

It may be proved that the times between the events (A) in an homogeneous Poisson process are *exponentially* distributed with parameter λ. This is easily verified for the first occurrence of A, say T_1. We have: $F_{T_1}(t) = 1 - \Pr(T_1 > t) = 1 - \Pr(\text{No A's in } [0, t >)$. Since $N(t)$ is Poisson distributed, $\Pr(N(t) = 0) = \frac{(\lambda t)^0}{0!}e^{-\lambda t} = e^{-\lambda t}$, and hence : $F_{T_1}(t) = 1 - e^{-\lambda t}$ which is the distribution function in the exponential distribution.

Let S_k denote the k'th occurrence of an A event. We have that $F_{S_k}(t) = 1 - \Pr(S_k > t) = 1 - \Pr(\text{less than } k \text{ failures up to time } t) = 1 - \sum_{j=0}^{k-1}\frac{(\lambda t)^j}{j!}e^{-\lambda t}$. The distribution function is recognized as the distribution function in the gamma distribution.

Gamma Distribution

$$f(t) = \frac{\lambda}{\Gamma(k)}(\lambda t)^{k-1}e^{-\lambda t}$$

where $\Gamma()$ is the gamma function, which is tabulated in Table 2 at the end of this document.

$$R(t) = \sum_{x=0}^{k-1}\frac{(\lambda t)^x}{x!}e^{-\lambda t} \quad \text{when } k \text{ is an integer}$$

When k is not an integer $R(t)$ can be expressed by the *incomplete* gamma function, but numerical methods are required for computation. The failure rate function can be expressed by $z(t) = \frac{f(t)}{R(t)}$, but no simple expression can be found.

$$\mathrm{E}(T) = \frac{k}{\lambda}, \qquad \mathrm{var}(T) = \frac{k}{\lambda^2}$$

With Excel we may use the `=gammadist(t,k,1/lambda,TRUE)` function. Note that the parametrization is deviating from the textbook.

Weibull Distribution

$$
\begin{aligned}
f(t) &= \alpha\lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\alpha} \\
R(t) &= e^{-(\lambda t)^\alpha}
\end{aligned}
$$

α is a *shape* parameter and λ is an *intensity* parameter.

$$
\mathrm{E}(T) = \frac{1}{\lambda}\Gamma\left(\frac{1}{\alpha}+1\right), \quad \mathrm{var}(T) = \frac{1}{\lambda^2}\left(\Gamma\left(\frac{2}{\alpha}+1\right) - \Gamma^2\left(\frac{1}{\alpha}+1\right)\right)
$$

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

where we observe that the failure rate function is increasing for $\alpha > 1$, and decreasing for $\alpha < 1$.

Example

For Norwegian women MTTF = 84 years and for Norwegian men MTTF = 81. Assume we will model the life times with a Weibull distribution with $\alpha = 5$. For men we have:

$$
\lambda = \Gamma(1+1/\alpha)/\mathrm{MTTF} = \Gamma(1+1/5)/81 \approx 0.0113
$$

For a new born boy the probability of surviving 100 years is:

$$
R(100) = e^{-(100\lambda)^\alpha} \approx 15\%
$$

For a 59 years old man the probability of becoming 100 years or more ($x = 100-59 = 41$) we have:

$$
R(x = 100-59|59) = R(100)/R(59) \approx 17\%
$$

The mean residual life,MRL$(t) = \dfrac{1}{R(t)}\displaystyle\int_t^\infty R(x)\,dx$, for this man is found by numerical methods to be approximately 27 years and he will die at the age of 86 (expected value !) which is four years more than for the new born boy.

Parametrization of the Weibull distribution

For may distributions we can use several reasonable ways to parametrize the distribution. Each way to do this could have pros and cons. For the Weibull distribution it is common to use an alternative parametrization:

$$
R(t) = e^{-\left(\frac{1}{\theta}\right)^\alpha}
$$

where θ is a *scale* parameter measured in time units, and α is still the *shape* parameter. The last version of the textbook is using the parametrization with the scale parameter, whereas previous versions used the parametrization with the intensity parameter ($\lambda = 1/\theta$).

Normal Distribution

$$f(t) = \frac{1}{\sqrt{2\pi}\cdot\tau}\, e^{-(t-\mu)^2/2\tau^2}$$

$$\mathrm{E}(T) = \mu, \qquad \mathrm{var}(T) = \tau^2$$

The distribution function for T could not be written on closed from. Numerical methods are required to find $F_T(t)$. It is convenient to introduce a standardised normal distribution for this purpose. We say that U is standard normally distributed if it's probability density function is given by:

$$f_U(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

We then have

$$F_U(u) = \Phi(u) = \int_{-\infty}^{u} \phi(t)\mathrm{d}t = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,\mathrm{d}t$$

and we observe that the distribution function of U does not contain any parameters. We therefore only need one look-up table or function representing $\Phi(u)$. A look-up table is given in the table of formulas. To calculate probabilities in the non-standardised normal distribution we use the following result:

If T is normally distributed with parameters μ and σ, then

$$U = \frac{T-\mu}{\sigma} \tag{2}$$

is standard normally distributed.

If we have access to for example MS Excel we may use $\Phi(u)$=`normdist(u,0,1,TRUE)`, or even $\Pr(T \le t)$ =`normdist(t,mu,sigma,TRUE)`.

Lognormal Distribution

$$f(t) = \frac{1}{\sqrt{2\pi}\cdot\tau t}\, e^{-(\ln t-\nu)^2/2\tau^2}$$

$$\mathrm{E}(T) = e^{\nu+\tau^2/2}, \quad t_{\mathrm{m}} = e^{\nu}, \quad \mathrm{var}(T) = e^{2\nu}\left(e^{2\tau^2} - e^{\tau^2}\right)$$

If T is lognormally distributed with parameters ν and τ, then $Y = \ln T$ is normally distributed[1] with expected value ν and variance τ^2.

[1] $\ln(\cdot)$ is the natural logarithm function

The failure rate function of the lognormal distribution can not be found on closed forms, but is easy to calculate by using the $\Phi(u)$ function. It appears that the failure rate function is first increasing, then decreasing. This is a reason why repair times often are modelled by the lognormal distribution. First the (repair) rate is increasing meaning that it is very likely that the repair is completed in the near future. But if for some reason it takes long time, the repair rate decreases. This is then an indication that problems were encountered, and the likelihood of completing the repair is dropping.

The Gamma Function

The gamma function $\Gamma(\alpha)$ is defined for all real $\alpha > 0$ by the integral

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

By partial integration it is easy to show that

$$\Gamma(\alpha + 1) = \alpha\, \Gamma(\alpha) \quad \text{for all } \alpha > 0 \tag{3}$$

When k is a positive integer

$$\Gamma(k+1) = k \cdot (k-1) \cdots 2 \cdot 1 \cdot \Gamma(1)$$

Since

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

we have

$$\Gamma(k+1) = k!$$

In Table 2 the Gamma function $\Gamma(\alpha)$ is given for values of α between 1.00 and 2.00. $\Gamma(\alpha)$ for other positive values of α may be calculated from formula (3).

If we have access to for example MS Excel we may use `=Gamma()`.

Table 2: Gamma Function $\Gamma(\alpha)$ for α between 1.00 and 2.00.

α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$
1.00	1.00000	1.25	0.90640	1.50	0.88623	1.75	0.91906
1.01	0.99433	1.26	0.90440	1.51	0.88659	1.76	0.92137
1.02	0.98884	1.27	0.90250	1.52	0.88704	1.77	0.92376
1.03	0.98355	1.28	0.90072	1.53	0.88757	1.78	0.92623
1.04	0.97844	1.29	0.89904	1.54	0.88818	1.79	0.92877
1.05	0.97350	1.30	0.89747	1.55	0.88887	1.80	0.93138
1.06	0.96874	1.31	0.89600	1.56	0.88964	1.81	0.93408
1.07	0.96415	1.32	0.89464	1.57	0.89049	1.82	0.93685
1.08	0.95973	1.33	0.89338	1.58	0.89142	1.83	0.93969
1.09	0.95546	1.34	0.89222	1.59	0.89243	1.84	0.94261
1.10	0.95135	1.35	0.89115	1.60	0.89352	1.85	0.94561
1.11	0.94740	1.36	0.89018	1.61	0.89468	1.86	0.94869
1.12	0.94359	1.37	0.88931	1.62	0.89592	1.87	0.95184
1.13	0.93993	1.38	0.88854	1.63	0.89724	1.88	0.95507
1.14	0.93642	1.39	0.88785	1.64	0.89864	1.89	0.95838
1.15	0.93304	1.40	0.88725	1.65	0.90012	1.90	0.96177
1.16	0.92980	1.41	0.88676	1.66	0.90167	1.91	0.96523
1.17	0.92670	1.42	0.88636	1.67	0.90330	1.92	0.96877
1.18	0.92373	1.43	0.88604	1.68	0.90500	1.93	0.97240
1.19	0.92089	1.44	0.88581	1.69	0.90678	1.94	0.97610
1.20	0.91817	1.45	0.88566	1.70	0.90864	1.95	0.97988
1.21	0.91558	1.46	0.88560	1.71	0.91057	1.96	0.98374
1.22	0.91311	1.47	0.88563	1.72	0.91258	1.97	0.98768
1.23	0.91075	1.48	0.88575	1.73	0.91467	1.98	0.99171
1.24	0.90852	1.49	0.88595	1.74	0.91683	1.99	0.99581
						2.00	1.00000