ALARI Statistics Course

Part II

Claudio Ortelli

Università della Svizzera italiana

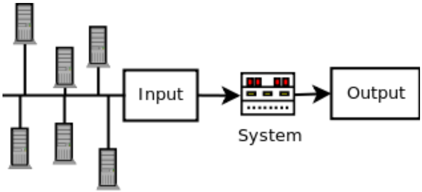
April 2011

Function of random variables

- Let X be a continous random variable with distribution function F_X, Ψ a function and
$$Y = \Psi(X).$$
- Under regularity conditions on Ψ, Y is a random variable!
- Continuity or stepwise continuity of Ψ are sufficient conditions for Y to be a random variable
-

- We are interested in simulating a model of a real system (network, electronic device, ...).
- The output Y of the model depends on a stochastic input, i.e. a random variable X with known distribution function F_X:

$$Y = g(X).$$

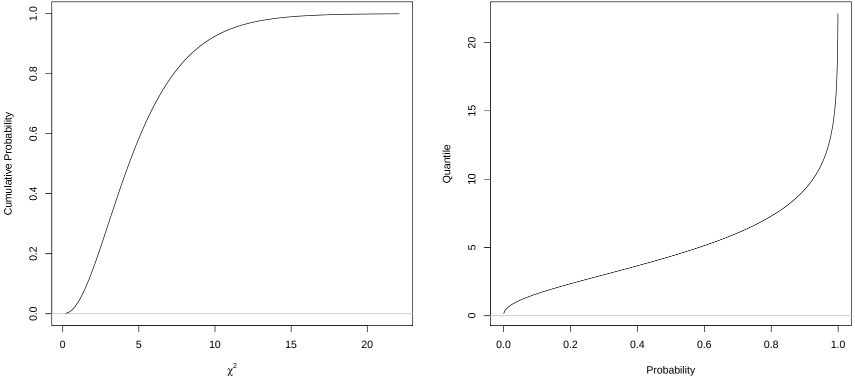


Simulation of random variables

- The model is too complex in order to analytically derive the probabilistic properties of the output, i.e. F_Y.
- Idea: simulate a possible outcome x of the input X and evaluate the corresponding outcome $y = g(x)$ of the output. Repeat the experiment N times and analyse the results.
- The simulated values of the input must be drawn from the distribution of X.
- Question: how is it possible to simulate independent realizations from a given distribution F_X?
- Answer: different methods available. The simplest of them requires simulating from the uniform distribution on the open interval (0,1), i.e. $U(0,1)$. In Matlab use the function "rand".

If the distribution function F_X is continuous and strictly increasing then $F_X^{-1}: (0,1) \to \mathbb{R}$ exists.

Example: Chi-Squared Distribution with 5 degrees of freedom



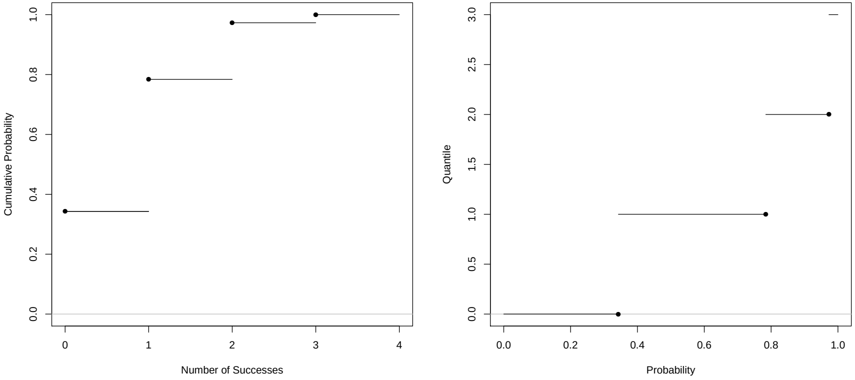
Simulation of random variables

For a general distribution function not necessarily strictly increasing define

$$F_X^{-1}(p) = \inf\{x : p \leq F_X(x)\} \quad 0 < p < 1.$$

It then follows that $F_X^{-1}(p) \leq x \iff p \leq F_X(x)$.

Example: Binomial distribution $Bin(n = 3, p = 0.3)$



Theorem (Inverse Transform Method)

Let U a continuous Unif(0,1) distributed random variable. The random variable $Y = F_X^{-1}(U)$ has distribution function F_X.

Proof.

By definition

$$F_Y(c) := P(Y \leq c) = P(F_X^{-1}(U) \leq c).$$

But the last equality is equivalent to (see previous slide)

$$P(U \leq F_X(c)) = F_X(c).$$

□

From the previous theorem we derive the following *two steps* simulation algorithm

- Simulate a realization u from a Unif(0,1) random variable U.
- Compute $x = F_X^{-1}(u)$.

Example: Simulation of $Y \sim Exp(\lambda)$

- The distribution function is $F_Y(x) = 1 - \exp(-\lambda x)$
- Compute $F_Y^{-1}(p) = -\frac{1}{\lambda} \ln(1-p)$
- Sample a random draw u from $U \sim$ Unif(0,1)
- Set $y = -\frac{1}{\lambda} \ln(1-u)$

Remark: if $U \sim$ Unif(0,1) then $1 - U$ is also Unif(0,1). Therefore we can also write $Y = -\frac{1}{\lambda}\ln(U)$.

For a descrete random variable Y with probability mass function $P(Y = x_i) = p_i$, $i = 1, ..., m$ consider the following algorithm:

1. Generate a Unif(0,1) random variable U
2. Compute Y as follows

$$Y = x_j \qquad \text{if } \sum_{i=1}^{j-1} p_i < U \leq \sum_{i=1}^{j} p_i.$$

i.e. $Y = x_j$ if $F_Y(x_{j-1}) < U \leq F_Y(x_j)$.

Example: let Y have the following mass function

p_1	p_2	p_3	p_4
0.1	0.2	0.4	0.3

The simulation algorithm is the following:

1. If $u \leq p_1 = 0.1$ then $y = x_1$. Stop.
2. if $u \leq p_1 + p_2 = 0.3$ then $y = x_2$. Stop.
3. if $u \leq p_1 + p_2 + p_3 = 0.7$ then $y = x_3$. Stop.
4. $y = x_4$. Stop.

This algorithm is correct but inefficient. In fact, probabilities of one, two, ... comparisons are equal to the probabilities of x_1, x_2, ..., respectively. The expected number of comparisons is

$$p_1 + 2p_2 + 3p_3 + 4p_4 = 2.9.$$

We can improve efficiency by sorting the values of x_i by decreasing order of probabilities p_i's: x_3, x_4, x_2 and x_1.

1. If $u \leq p_3 = 0.4$ then $y = x_3$. Stop.
2. if $u \leq p_3 + p_4 = 0.7$ then $y = x_4$. Stop.
3. if $u \leq p_3 + p_4 + p_2 = 0.9$ then $y = x_2$. Stop.
4. $y = x_1$. Stop.

The expected number of comparisons is now equal to

$$p_3 + 2p_4 + 3p_2 + 4p_1 = 2.$$

The Laplace distribution is a continuous distribution with density function

$$f(x;\mu,b) = \begin{cases} \frac{1}{2b}\exp(-\frac{\mu-x}{b}) & \text{if } x < \mu \\ \frac{1}{2b}\exp(-\frac{x-\mu}{b}) & \text{if } x \geq \mu \end{cases}$$

where μ is the mean and b a scale parameter.

1. Plot the density, distribution and quantile functions of the Laplace distribution with parameter $\mu = 1$ and $b = 0.5$.
2. Using the previous values of μ and b simulate $N = 1000$ independent realizations of a Laplace distributed random variable Y.
3. Plot the histogram of the simulated random variables and compare it with the density function of Y.
4. Plot the empiric distribution function of the simulated sample and compare it with the distribution function of Y.

The theorem on *Distributions of functions of continuous random variables* allows us to generate random variables by means of ad hoc transformations of Unif(0,1) random variables. The following theorem generalizes the previous theorem to the multivariate case.

Theorem

Assume that $X = (X_1, X_2)$ *is a random vector with joint density function* $f_X(x_1, x_2)$ *and* $g : \mathbb{R}^2 \to \mathbb{R}^2$ *a one-to-one and continuously differentiable function. Define* $Y = (Y_1, Y_2) = g(X)$*. The density function of* Y *is then equal to*

$$f_Y(y) = f_X(g^{-1}(y))\,|J(g^{-1}(y))|$$

where

$$J(g^{-1}(y)) = \det(M_J) = \det\left[\frac{\partial x_i(y)}{\partial y_j}\right]_{i=1,2;\ j=1,2}$$

Example: Let $X \sim N(\mu, \Sigma)$ be a 2×1 random vector and consider the affine transformation

$$Y = b + A\,X$$

where b is a deterministic 2×1 vector and A a 2×2 invertible matrix. We then have

$$X = A^{-1}(Y - b);\;\; M_J = A^{-1}$$

and $f_Y(y)$ is equal to

$$\frac{1}{\sqrt{2\pi det(\Sigma)}} \exp(-\frac{1}{2}(A^{-1}(Y-b)-\mu)'\Sigma^{-1}(A^{-1}(Y-b)-\mu))|det(A^{-1})|$$

$$\frac{1}{\sqrt{2\pi det(A\Sigma A')}} \exp(-\frac{1}{2}(Y-b-A\mu)'(A\Sigma A')^{-1}(Y-b-A\mu))$$

Example (continued):
Looking at the density function of Y we note that $Y \sim N(\tilde{\mu}, \tilde{\Sigma})$ whith $\tilde{\mu} = b + A\mu$ and $\tilde{\Sigma} = A\Sigma A'$.

Exercise: U_1 and U_2 are two independent Unif(0,1) distributed random variables. Define $Y = (Y_1, Y_2)$ with

$$Y_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \text{ and } Y_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2)$$

Show that $Y \sim N(0, I)$, i.e. Y_1 and Y_2 are independent standard normal distributed random variables.

Exercise: The joint density of the random variables X_1 and X_2 is

$$f(x_1, x_2) = 2\exp(x_1)\exp(x_2), \text{ for } 0 < x_1 < x_2 < \infty$$

and $f(x_1, x_2) = 0$ otherwise.
We define the transformation

$$Y_1 = 2X_1, Y_2 = X_2 - X_1.$$

Find the joint density of Y_1 and Y_2. Are Y_1 and $Y2$ independent?

- Transform methods are transformations of the probability mass function (discrete case) or the density function (continuous case).
- They are particular useful to compute moments of a distribution and in problems involving sums of independent random variables.

Definition

The moment generating function (MGF) $M_X(\theta)$, abbreviated $M(\theta)$, of the random variable X is defined by

$$M(\theta) = E\left[\exp(X\theta)\right]$$

provided the expectation exists ($M(\theta)$ may not exist for all $\theta \in \mathbb{R}$).

Definition

The characteristic function of a random variable X is given by

$$N_X(\tau) = N(\tau) = E\left[\exp(iX\tau)\right] = M_X(i\tau) \text{ where } i = \sqrt{-1}.$$

Note that $N_X(\tau)$ is always defined for any X and all τ.

Definition

Let X be a nonnegative continuous random variable. The Laplace - Stieltjes transform of X is

$$L_X(s) = L(s) = M_X(-s) = \int_0^\infty \exp(-sx) f(x) dx.$$

Definition

Let X be a discrete nonnegative integer-valued random variable. The z transform (or probability generating function) of X is defined as

$$G_X(z) = G(z) = E\left[z^X\right] = M_X(ln(z)) = \sum_{i=0}^{\infty} p_X(i) z^i.$$

Theorem

Affine transformation. Let $Y = aX + b$. Then

$$M_Y(\theta) = \exp(b\theta) M_X(a\theta)$$

Theorem (The Convolution Theorem)

Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables. Define $Y = \sum_{i=1}^{n} X_i$. If $M_{X_i}(\theta)$ exists for all i, then $M_Y(\theta)$ exists, and

$$M_Y(\theta) = \prod_{i=1}^{n} M_{X_i}(\theta).$$

Theorem (Uniqueness Theorem)

If $M_X(\theta) = M_Y(\theta)$ for all θ, then $F_X = F_Y$, i.e. X and Y have the same distribution.

Theorem (Moment generating property of the *MGF*)

Let X be a random variable such that all moments exist. Then

$$E\left[X^k\right] = \frac{\partial^k M_X}{\partial \theta^k}\Big|_{\theta=0} \quad k = 1, 2, \ldots$$

Proof.

$$\exp(X\theta) = 1 + X\theta + \frac{X^2\theta^2}{2!} + \cdots + \frac{X^k\theta^k}{k!} + \ldots$$

Taking expectation on both sides

$$M_X(\theta) = E\left[\exp(X\theta)\right] = 1 + E\left[X\right]\theta + \frac{E\left[X^2\right]\theta^2}{2!} + \cdots + \frac{E\left[X^k\right]\theta^k}{k!} + \ldots$$

$\square$

The corresponding properties for the characteristic function N_X, the Laplace - Stieltjes transform L_X and the z transform G_X are

$$E\left[X^k\right] = (-i)^k \frac{\partial^k N_X}{\partial \tau^k}\Big|_{\tau=0} \quad k = 0, 1, \ldots$$

$$E\left[X^k\right] = (-1)^k \frac{\partial^k L_X}{\partial s^k}\Big|_{s=0} \quad k = 0, 1, \ldots$$

$$E\left[\frac{X!}{(X-k)!}\right] = \lim_{\tilde{z}\uparrow 1} \frac{\partial^k G_X}{\partial z^k}\Big|_{z=\tilde{z}} \quad k = 0, 1, \ldots$$

respectively, where $\left[\frac{X!}{(X-k)!}\right] = X(X-1)\ldots(X-k+1)$.

Finally, let X be a discrete nonnegative integer-valued random variable with z transform G_X. The probability mass function of X can be recovered by taking derivatives of G_X:

$$p_k = P(X = k) = \frac{1}{k!}\frac{\partial^k G_X}{\partial z^k}\,|_{z=0}$$

Examples: Let X be exponentially distributed with parameter λ. Then

$$f_X(x) = \lambda \exp(-\lambda x),\ \ x > 0.$$

$$\begin{aligned} L_X(s) &= \int_0^\infty exp(-sx)exp(-\lambda x)dx \\ &= \frac{\lambda}{s+\lambda}\int_0^\infty (\lambda + s)exp(-(\lambda + s)x)dx \\ &= \frac{\lambda}{s+\lambda}. \end{aligned}$$

Example (continued):

$$E[X] = (-1)\frac{\partial L_X}{\partial s}\big|_{s=0} = (-1)\frac{-\lambda}{(\lambda+s)^2}\big|_{s=0} = \frac{1}{\lambda}.$$

$$E\left[X^2\right] = \frac{\partial^2 L_X}{\partial s^2}\big|_{s=0} = \frac{2\lambda}{(\lambda+s)^3}\big|_{s=0} = \frac{2}{\lambda^2}.$$

Example: Let X be a n trials Binomial distributed random variable with probability of success p. The z transform of X is by definition

$$\begin{aligned} G_X(z) &= E(z^X) = \sum_{k=0}^{n} z^k \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} \\ &= (pz+1-p)^n \end{aligned}$$

Exercise:
Let X be a Bernoulli distributed random variable with probability of success p.

1. Compute the MGF M_X.
2. Compute skewness and kurtosis of X.

Exercise:
Let $X_1, X_2, \ldots, X_n$ a sequence of independent Bernoulli distributed random variables.

1. Compute the moment generating function of $Y = \sum_{i=1}^{n} X_i$.
2. Show that M_Y is the MGF of a Bernoulli (n, p) distributed random variable.

Exercise:
Let X be a standard normally distributed random variable.

1. Compute the MGF of X.
2. Compute the Kurtosis of X.
3. Define $Y = \sigma X + \mu$. Derive the MGF of Y and compute its expected value and variance.

Exercise:
Let X be a geometric distributed random variable with probability mass function $p_X(i) = p(1-p)^i$, $i = 1, 2, ...$

1. Compute the z transform of X.
2. Compute the skewness of X.

Exercise:

Let X be a continuous Unif(a, b) distributed random variable with $0 \leq a < b$.

1. Compute the MGF and the Laplace - Stieltjes transform of X.
2. Compute skewness and kurtosis of X.

Let us consider an electronic system with n independent components. Define the event

$$A_i := \text{“The } i - th \text{ component is functioning properly ”.}$$

Definition

The reliability of component i is defined as

$$R_i = P(A_i),$$

i.e. it is the probability that the component is functioning properly.

Definition

A series system is a system such that the entire system fails if any one of its components fails.

Definition

A parallel system is a system such that the entire system fails only if all its components fail.

Theorem (Product law of reliabilities for series systems)

The reliability of a series system decreases "quickly "with an increase in complexity (number of components).

Proof.

Let us consider the event $A =$"The system functions properly". The reliability of a series system of n components is then

$$R = P(A) = P(A_1 \cap A_2 \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i)$$

$\square$

Example

Let a series system have $n = 5$ components and $P(A_i) = 0.970$ for all components.

The system reliability is then equal to

$$R = P(A_1)^5 = 0.97^5 = 0.859.$$

If we increase $n = 10$ the system reliability decreases to 0.738!

What if $n = 1'000'000$?

In order to mitigate the problem one possible solution is to implement parallel redundancy.

Example

Consider a parallel system of n independent components. The system runs correctly if at least one of its components runs properly, i.e.

$$A = (A_1 \cup A_2 \cdots \cup A_n).$$

But this means that

$$P(\overline{A}) = P(\overline{(A_1 \cup A_2 \cdots \cup A_n)}) = P(\overline{A}_1 \cap \overline{A}_2 \cdots \cap \overline{A}_n)) = \prod_{i=1}^{n} P(\overline{A}_i)$$

Example (continued)

Applying the identity $P(\overline{B}) = 1 - P(B)$ to both sides of the equality and solving w.r.t. $P(A)$, we obtain the final formula

$$P(A) = 1 - \prod_{i=1}^{n}(1 - P(A_i))$$

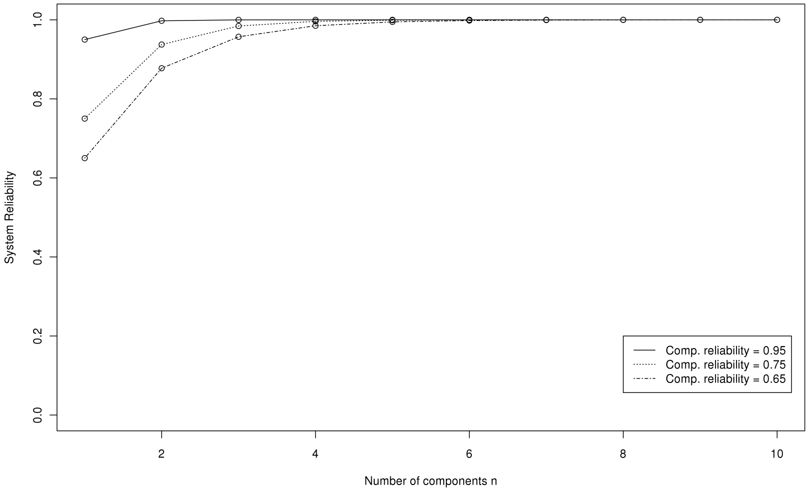
or equivalently, in terms of R and R_i

$$R = 1 - \prod_{i=1}^{n}(1 - R_i).$$

The next picture shows the so called Product Law of Unreliabilities.

System's reliability

Product Law of Unriliabilities



From the previous picture it is evident one characteristic of parallel redundancy: the marginal increase in reliability decreases with increasing number of parallel components. This behaviour is called the **Law of Diminishing Returns**.

Definition (series-parallel system)

A system with both series and parallel parts is called a series-parallel system.

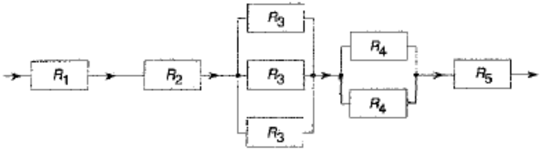


Figure: series-parallel reliability block diagram

Definition

Let X be a state vector of a system with n components so that $X = (x_1, \ldots, x_n)$ where

$$x_i = \begin{cases} x_i = 1 & \text{if component } i \text{ is functioning,} \\ x_i = 0 & \text{if component } i \text{ has failed.} \end{cases}$$

The structure function $\Phi(X)$ is defined by

$$\Phi(X) = \begin{cases} x_i = 1 & \text{if system is functioning,} \\ x_i = 0 & \text{if system has failed.} \end{cases}$$

The reliability of the system is then

$$R = P(\Phi(X) = 1).$$

Example (Communication network)

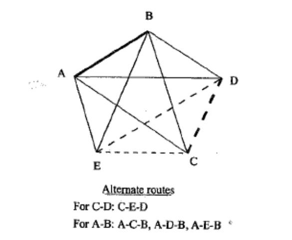


Figure: Communication network with five nodes

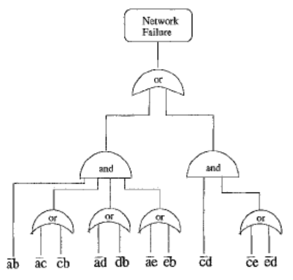


Figure: Fault tree for the communication network

The exponential distribution find its application in reliability theory and queuing theory. The following random variables are often modeled as exponential:

1. Time between two successive job arrivals to a file server (often called **interarrival time**).
2. Service time at a server in a queuing network; the server could be a resource such as a CPU, an I/O device, or a communication channel.
3. Time to failure (lifetime) of a component.
4. Time required to repair a component that has malfunctioned.

Remark: The choice of the exponential distribution to model the stochastic structure of the upper described variables is an assumption and not a given fact! Experimental verification of the distributional assumption will be therefore necessary before to relying on the results of the analysis.

Le $X \sim Exp(\lambda)$ be the lifetime of a component. Suppose we have observed that it has already been operating for t hours.

- What is the distribution of the remaining (residual) lifetime $Y = X - t$?

Let the conditional probability of $Y \leq y$, given that $X > t$, be denoted by $G_Y(y|t)$. For $y \geq 0$

$$\begin{aligned} G_Y(y|t) &= P(Y \leq y|X > t) = \frac{P(\{Y \leq y\} \text{ and } \{X > t\})}{P(X > t)} \\ &= \frac{P(\{X \leq y + t\} \text{ and } \{X > t\})}{P(X > t)} = \frac{P(t < X \leq y + t)}{P(X > t)} \\ &= \frac{exp(-\lambda t)(1 - exp(-\lambda y))}{exp(-\lambda t)} = 1 - exp(-\lambda y). \end{aligned}$$

Result:
The conditional distribution $G_Y(y|t)$ does not depend on t and is identical to the distribution of X, i.e. $Exp(\lambda)$.

Interpretation:
The distribution of the remaining life does not depend on how long the component has been operating, i.e. the component does not age (it is as good as new). Therefore, the exponential distribution is not suited to model components or devices that gradually deteriorate.

Let the random variable X be the lifetime (or time to failure) of a component.

Definition

The **reliability** $R(t)$ of the component is the probability that the component survives until some time t, i.e.

$$R(t) = P(X > t) = 1 - F_X(t)$$

$F_X(t)$ is often called the **unreliability** of the component.

The conditional probability that the component does not survive for an additional interval of duration x given that it has survived until time t is equal to

$$G_Y(x|t) = \frac{P(t < X \leq t + x)}{P(X > t)} = \frac{F_X(t+x) - F_X(t)}{R(t)}$$

Definition

The instantaneous failure rate $h(t)$ is defined to be

$$h(t) = \lim_{x \to 0} \frac{1}{x} G_Y(x|t) = \lim_{x \to 0} \frac{F_X(t+x) - F_X(t)}{xR(t)},$$

so that

$$h(t) = \frac{f_X(t)}{R(t)}.$$

Alternate terms for $h(t)$ are *hazard rate*, *force of mortality*, *intensity rate*, *conditional failure rate* or **failure rate**.

Interpretation:

- $h(t)\Delta t$ represents the conditional probability that a component having survived to age t will fail in the interval $(t, t + \Delta t]$.

- $f_X(t)\Delta t$ is the *unconditional* probability while $h(t)\Delta t$ is a conditional probability.

Next theorem shows the connection between reliability and failure rate.

Theorem

$$R(t) = \exp\left(-\int_0^t h(x)dx\right)$$

Proof.

$$\int_0^t h(x)dx = \int_0^t \frac{f_X(t)}{R(t)}dx = \int_0^t \frac{-R'(t)}{R(t)}dx = -\ln(R(t))$$

using the fact that $R'(t) = -f_X(t)$ and the boundary contition $R(0) = 1$. □

Definition

The cumulative hazard is defined to be

$$H(t) = \int_0^t h(x)dx$$

Then, reliability can also be written as $R(t) = \exp(-H(t))$.

Definition

The conditional reliability $R_t(y)$ is the probability that the component survives an additional interval of duration y given that it has survived until time t.

$$R_t(y) = \frac{R(t+y)}{R(t)} \tag{1}$$

Assume a component does not age stochastically, i.e. the survival probability over an additional time interval y is the same regardless of the age t of the component:

$$R_t(y) = R_s(y) \text{ for all } t, s \geq 0.$$

For $s = 0$

$$R_t(y) = R_0(y) = \frac{R(y)}{R(0)} = R(y),$$

so that

$$R(t + y) = R(t)R(y).$$

In particular we obtain

$$\frac{R(t + y) - R(y)}{t} = \frac{(R(t) - 1)R(y)}{t} = \frac{(R(t) - R(0))R(y)}{t}.$$

Taking the limit as $t \to 0$

$$\begin{aligned} R'(y) &= R'(0)R(y) \\ R(y) &= \exp(yR'(0)) = \exp(-\lambda y) \end{aligned}$$

which shows that the lifetime $X \sim Exp(\lambda)$.
If a component has exponential lifetime distribution it follows that

1. A replacement policy of used components based on the lifetime of the components is useless.
2. In estimating mean life and reliability the age of the observed components are of no concern. The number of hours of observed live and the number of observed failures are of interest.

Definition

Increasing (decreasing) failure rate distribution
Let X be the lifetime of a component and $F_X(t)$ the corresponding distribution function. If its failure rate $h(t)$ is an increasing (decreasing) function of t for $t \geq 0$ then F_X is an Increasing (Decreasing) Failure Rate distribution: IFR (DFR) distribution.

The behavior of the failure rate $h(t)$ as a function of age is known as the *mortality curve*, *hazard function*, *life characteristic* or *lambda characteristic*.

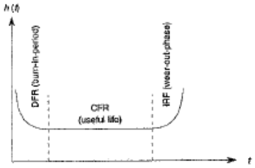


Figure 3.6. Failure rate as a function of time

The hypoexponential distribution is used to model processes that can be divided into sequential phases such that the time the process spends in each phase is independent and exponentially distributed.

- Service times for input-output operations in a computer system often follow this distribution.

A two stage hypoexponential random variable $X \sim Hypo(\lambda_1, \lambda_2)$ has pdf and distribution function equal to

$$\begin{aligned} f(t) &= \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1}(\exp(-\lambda_1 t) - \exp(\lambda_2 t)), \quad t > 0 \\ F(t) &= 1 - \frac{\lambda_2}{\lambda_2 - \lambda_1} \exp(-\lambda_1 t) + \frac{\lambda_1}{\lambda_2 - \lambda_1} \exp(-\lambda_2 t) \end{aligned}$$

When r sequential phases have identical exponential distribution the resulting density is known as $r-$stage Erlang and is given by

$$
\begin{aligned}
f(t) &= \frac{\lambda^r t^{r-1} \exp(-\lambda t)}{(r-1)!} \text{ with } t > 0,\ \lambda > 0,\ r = 1, 2, \ldots \\
F(t) &= 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} \exp(-\lambda t) \text{ with } t \geq 0,\ \lambda > 0,\ r = 1, 2, \ldots
\end{aligned}
$$

Suppose that a process consists of alternate phases, i.e. during any single experiment the process experiences one and only one of the many alternate phases, and thes phases have exponential distributions. The overall distribution is then hyperexponential with density and distribution functions given by

$$
\begin{aligned}
f(t) &= \sum_{i=1}^{k} \alpha_i \lambda_i \exp(-\lambda_i t) \text{ with } t > 0,\ \lambda_i > 0,\ \sum_{i=1}^{k} \alpha_i = 1 \\
F(t) &= \sum_{i} \alpha_i (1 - \exp(\lambda_i t)) \quad t \geq 0
\end{aligned}
$$

The Weibull distribution is the most widely used parametric family of failure distributions. It has been used to describe

- fatigue failure
- electronic component failure
- ballbearing failure

The reason is that by a proper choice of the shape parameter α we can obtain an IFR, DFR or constant failure rate distribution. The corresponding density and distribution functions are given by

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

where $t \geq 0$, $\lambda > 0$ and $\alpha > 0$.

The Pareto (also knonw as double-exponential, hyperbolic or power-law) distribution has been used to model

- the amount of CPU time consumed by an arbitrary process
- the Web file size on the Internet servers
- the thinking time of the web browser
- the number of data bytes in FTP bursts
- the access frequency of Web traffic

The density and distributions functions are given by

$$\begin{aligned} f(x) &= \alpha k^{\alpha} x^{-\alpha-1} \quad x \geq k,\ k > 0,\ \alpha > 0 \\ F(x) &= \begin{cases} 1 - \left(\frac{k}{x}\right)^{\alpha} & x \geq k \\ 0 & x < k \end{cases} \end{aligned}$$