

# Parametric probability distributions in reliability\*

F.P.A. Coolen

Department of Mathematical Sciences, Durham University  
Durham, DH1 3LE, UK

## Abstract

In this paper, we present an overview of basic parametric probability distributions which are frequently used in reliability. We present some main characteristics of these distributions, and briefly discuss underlying assumptions related to their suitability as models for specific reliability scenarios.

**Keywords:** Binomial distribution, Exponential distribution, Gamma distribution, Normal distribution, Poisson distribution, Weibull distribution.

## 1 Introduction

In many applications of reliability theory and quantitative risk assessment, one is interested in random quantities for which one assumes a parametric probability distribution, as is generally often done in statistics and stochastics. In this paper, we briefly introduce some of the most commonly used parametric probability distributions in reliability, restricting ourselves to uni-variate random quantities and the most basic forms of the distributions. Far more detailed presentations of these distributions, with historical notes, discussions of properties and applications, and further generalizations, can be found in the *Wiley Encyclopedia of Statistical Sciences* [1]. Also key text books in reliability and related topics tend to present these (and more) parametric probability distributions in more detail, see e.g. Hougaard [2], Lawless [3], Martz and Waller [4],

---

\*Contribution to *Wiley Encyclopedia of Quantitative Risk Analysis and Assessment*

or Meeker and Escobar [5]. We present these main distributions in a very brief manner, in many other articles in this encyclopedia these distributions are used, providing examples of their applications.

Parametric probability distributions are used both in stochastic analyses of reliability of systems, where they are mostly assumed to be fully known and corresponding properties of the system are analyzed, and in statistical inference, where process data are used to estimate the parameters of the distribution, often followed by a specific inference of interest. In the latter case, one must take care to also diagnose the assumptions underlying the specific parametric distribution assumed, to ensure a reasonable fit with the empirical data [1].

We present some main probability distributions for both continuous and discrete random quantities. For the former, parametric probability distributions can be uniquely specified via the probability density function (pdf)  $f(t)$ , the cumulative distribution function (cdf)  $F(t)$ , the survival function  $S(t)$ , or the hazard rate  $h(t)$ . For example, for a non-negative random quantity, as often used in reliability if one is interested in a random lifetime, these functions are related as follows:  $F(t) = \int_0^t f(u)du$ ,  $S(t) = 1 - F(t)$  and  $h(t) = f(t)/S(t)$ . The hazard rate is often considered to be particularly attractive in reliability, as it represents an instantaneous rate of the occurrence of a failure, conditioned on no failure until now. For example, wear-out of a unit over time can be represented by an increasing hazard rate.

In Section 2 we briefly consider the Normal distribution, and some related distributions, which, although important, play perhaps a less central role in reliability than in other application areas of statistics. In Section 3 we present the Exponential distribution, and the Weibull and Gamma distributions which are popular models in reliability and which can be considered as generalizations of the Exponential distribution. Section 4 presents some important distributions for discrete random quantities, in Section 5 some further important distributions for reliability are briefly mentioned, and the paper is concluded with some further remarks in Section 6.

## 2 Normal and related distributions

The Normal distribution (also known as ‘Gaussian distribution’) is arguably the most important probability distribution in Statistics, as it occurs as the limiting distribution when a sum of random quantities is considered (Central Limit Theorem). In general, it also plays a big role in quantitative risk assessment, e.g. when risks of portfolios

of investments are considered, but its role in reliability is somewhat less important. However, it is frequently used as a suitable probability distribution for the (natural) logarithm of a random lifetime, in which case the lifetime's distribution is called 'Log-normal'. It is also related to the Inverse Gaussian distribution, which is important in some processes in reliability.

### **Normal distribution**

The Normal distribution has two parameters,  $-\infty < \mu < \infty$  and  $\sigma^2 > 0$ , which are equal to its mean and variance, so its standard deviation is  $\sigma$ , and its pdf is

$$f(x|\mu, \sigma^2) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right) \quad \text{for } -\infty < x < \infty$$

The cdf of the Normal distribution is not available in closed form, so computations often used tables of the cdf of the Standard Normal distribution, which has  $\mu = 0$  and  $\sigma^2 = 1$ , using the fact that  $(X - \mu)/\sigma$  is Standard Normally distributed if  $X$  is Normally distributed with parameters  $\mu$  and  $\sigma^2$ . Such tables are included in statistics text books, but all main statistical and mathematical software nowadays have good routines for the computation of the Normal distribution cdf. Its direct use in reliability is often restricted to modelling residual or error terms in regression models, but it plays an important role as a model for log-transformed lifetimes.

### **Lognormal distribution**

A random quantity  $T > 0$  is Lognormally distributed if  $X = \ln T$  has a Normal distribution, so that  $T = \exp(X)$  has pdf

$$f(t|\mu, \sigma^2) = \frac{1}{(2\pi)^{1/2}\sigma t} \exp\left(\frac{-(\ln t - \mu)^2}{2\sigma^2}\right) \quad \text{for } t > 0$$

The mean and variance of  $T$  are  $\exp(\mu + \sigma^2/2)$  and  $[\exp(\sigma^2) - 1][\exp(2\mu + \sigma^2)]$ , respectively. This distribution is quite popular as a model for lifetimes, even though its hazard rate has the somewhat unattractive property that it increases, from  $h(0) = 0$ , to a maximum, and thereafter decreases to 0 for  $t \rightarrow \infty$ . However, if attention is particularly to early failure times, yet with a wear-out effect, then this model might be appropriate. One possible argument justifying the use of this model is related to the Central Limit Theorem argument for the Normal distribution. Informally stated, the latter implies that, if a random quantity can be considered to be the sum of many independent random quantities, then its distribution will be approximately Normal. Hence, the same property holds for  $\ln T$  if  $T$  has a Lognormal distribution, with the

log-transform here meaning that  $T$ , for this argument, can be interpreted as the product of many independent random quantities, which might be attractive in certain types of failure processes. An obvious argument for the popularity of the Lognormal distribution was always the wide availability of statistical tables for the Normal distribution, an argument that is less relevant nowadays.

### Inverse Gaussian distribution

Contrary to what the name might perhaps suggests, this is not a distribution for a simple transformation of a Normally distributed random quantity. However, it becomes ever more important in reliability theory due to the fact that it is appropriate for stopping times in Brownian motion ('Gaussian') processes, and these (and related) processes are playing an increasingly important role in reliability modelling. Suppose that a Brownian motion, starting at 0 at time  $t = 0$ , has drift  $\psi \geq 0$  and variance  $\sigma^2$ , then the time to reach the value  $a > 0$  for the first time has an Inverse Gaussian distribution with parameters  $\beta = \frac{\psi}{2\sigma^2}$  and  $\alpha = \frac{a}{\sigma\sqrt{2}}$ . The pdf of this distribution is

$$f(t|\alpha, \beta) = \frac{\alpha \exp(2\alpha\beta^{1/2}) \exp(-\beta t - \alpha^2/t)}{\pi^{1/2} t^{3/2}} \quad \text{for } t > 0$$

Padgett and Tomlinson [6] present an excellent example of the use of such processes to describe degradation, providing a model of continuous cumulative damage. They present a general accelerated test model in which failure times and degradation measures are combined for inference about system lifetime, with the drift of the process depending on the acceleration variable. Their paper includes an illustrative example using degradation data observed in carbon-film resistors.

## **3 Exponential and related distributions**

The probability distributions in this section, for a continuous positive random quantity  $T$ , are very popular models for lifetimes, used in a wide variety of applications in reliability and beyond. We briefly introduce these distributions in basic forms, they can be generalized in a variety of ways. The most obvious generalization is inclusion of a location parameter  $\tau$ , such that  $T > \tau$ , so effectively this shifts the start of the distribution from 0 to  $\tau$ . Mathematically this is straightforward, but one must be careful in case of statistical inference based on the likelihood function, as for some models the maximum likelihood estimator of  $\tau$  is equal to the smallest observation in the available data, which is unreasonable from several perspectives.

### **Exponential distribution**

The Exponential distribution has a constant hazard rate, say  $h(t|\lambda) = \lambda > 0$ , so its pdf is

$$f(t|\lambda) = \lambda \exp(-\lambda t) \quad \text{for } t > 0$$

It is the unique model, for continuous  $T$ , with the so-called ‘memory-less property’, that is the probability for the event to take place in a future interval does not depend on the current age of the item or individual considered. This distribution models the random times between events in homogeneous Poisson processes.

### **Weibull distribution**

The Weibull distribution is a very widely used probability distribution in reliability, and it has two parameters: scale parameter  $\alpha > 0$  and shape parameter  $\beta > 0$ . It is easiest introduced via its hazard rate, which is

$$h(t|\alpha, \beta) = \alpha\beta(\alpha t)^{\beta-1} \quad \text{for } t > 0$$

The corresponding survival function is

$$S(t|\alpha, \beta) = \exp(-(\alpha t)^\beta) \quad \text{for } t > 0$$

For  $\beta = 1$  this distribution is an Exponential distribution, whereas  $\beta > 1$  leads to an increasing hazard rate, and hence can be considered to model wear-out, as often deemed appropriate for mechanical units, and  $\beta < 1$  leads to decreasing hazard rate, hence modelling wear-in of a product as often advocated for electronic units. One Weibull distribution cannot model both wear-in and wear-out at different stages, so e.g. it cannot resemble a ‘bath-tub shaped hazard rate’, but different Weibull distributions are often advocated for the different stages of a units’ life. Although its form is mathematically straightforward, statistical inference based on the Weibull distribution is less trivial due to the fact that there are no reduced-dimensional sufficient statistics for the shape parameter  $\beta$ . This distribution is routinely available in standard mathematical and statistical software. The Weibull distribution with  $\beta = 2$ , hence with hazard rate a linear function of  $t$ , is also known as the Rayleigh distribution. Weibull distributions are also frequently used in regression-type inferences in reliability, enabling information from covariates to be taken into account. There is also a useful relationship between the Weibull distribution and an Extreme Value distribution (also called the Gumbel distribution), which occurs as a limiting distribution for extreme values of a sample

under some assumptions on the tail of the sample distribution: if  $T$  has a Weibull distribution, then  $\ln T$  has such an Extreme Value distribution.

### **Gamma distribution**

The Gamma distribution is also a useful lifetime model in many reliability applications, and has two parameters: scale parameter  $\lambda > 0$  and shape parameter  $\kappa > 0$ . Its pdf is

$$f(t|\kappa, \lambda) = \frac{\lambda(\lambda t)^{\kappa-1} e^{-\lambda t}}{\Gamma(\kappa)} \quad \text{for } t > 0$$

where  $\Gamma(\kappa)$  is the Gamma function, which of course equals the integral of the numerator of this pdf over  $t > 0$ . Computation of the cdf, survival function and hazard rate is somewhat awkward as it involves the incomplete Gamma function, yet also this distribution is available in the leading mathematical and statistical software. The Gamma distribution with  $\kappa = 1$  is an Exponential distribution, and Gamma distributions with integer  $\kappa$  are also known as Erlang distributions. The pdf of the Gamma distribution can have many different shapes, but it is not as commonly used in reliability as the Weibull distribution, possibly due to the lack of a clearly interpretable hazard rate. An attractive interpretation, however, occurs for integer  $\kappa$ , in which case a Gamma distributed lifetime can be interpreted as the sum of  $\kappa$  independent Exponentially distributed random quantities, which might, for example, be suitable if one wishes to model the lifetime of a system with spare units. This latter property also makes the Gamma distribution popular in models for queueing systems. With increased popularity of Bayesian methods in statistics, the Gamma distribution is used frequently as it is the conjugate prior distribution for both Exponential and Poisson sampling distributions.

## **4 Distributions for discrete random quantities**

In reliability, as in many other application areas of stochastic modelling and statistical inference, one is also regularly interested in discrete random quantities, e.g. when counting the number of components that have performed a task successfully. The Binomial distribution is probably the most frequently used distribution, whereas also the Negative Binomial distribution and the Poisson distribution deserve to be mentioned explicitly. Of course, there are again many variations and generalizations to these and other distributions for discrete random quantities, for these we refer again to the literature as mentioned in the introductory section.

### **Binomial distribution**

Suppose that  $X$  denotes the number of successes in  $n$  independent trials, where each trial is a success with probability  $\theta$  and a failure with probability  $1 - \theta$ . The probability distribution of  $X$ , for a given value of  $n$ , is given by

$$P(X = x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad \text{for } x \in \{0, 1, \dots, n\}$$

The expected value and variance of  $X$  are  $n\theta$  and  $n\theta(1 - \theta)$ , respectively. Computation of these probabilities is straightforward if  $n$  is not too large, for large values of  $n$  two approximations can be used: for values of  $x$  close to 0 or  $n$ , the Poisson distribution with expected value  $n\theta$  can be used as a reasonable approximation, whereas for other values of  $x$  an approximation based on the Normal distribution with the same expected value and variance is suitable. A standard ‘text-book’ example for a situation where the Binomial distribution is appropriate is  $n$  tosses of a possibly biased coin, where the coin lands heads up with probability  $\theta$  on each toss, and where the number of tosses with heads up is counted. In situations where the possible outcomes are in more than two unordered categories, the Multinomial distribution provides a suitable generalization of the Binomial distribution [1].

### **Negative Binomial distribution**

The Negative Binomial distribution is a variation to the Binomial distribution, with the total number of trials not a predetermined constant, but instead one counts the number of trials until a specified number of successes have occurred. Suppose again that trials are independent, each a success with probability  $\theta$  and a failure with probability  $1 - \theta$ , and let  $N$  denote the number of trials required to obtain the  $r$ -th success. Then the probability distribution of  $N$ , for fixed  $r$ , is given by

$$P(N = n|\theta) = \binom{n-1}{r-1} \theta^r (1 - \theta)^{n-r} \quad \text{for } n = r, r + 1, \dots$$

Sometimes the Negative Binomial distribution is defined slightly differently, namely as counting the number of trials prior to the  $r$ -th success. This distribution is important in its own right in probabilistic risk assessment, in reliability but also for example in quality control, but it is also increasingly used in Bayesian statistical inferences [4], as it occurs as the posterior predictive distribution for Poisson sampling with a conjugate Gamma prior distribution. The special case of  $r = 1$  is known as the Geometric distribution, which simply counts the number of trials until the first success has occurred.

### Poisson distribution

A random quantity  $X$  has a Poisson distribution with expected value  $\mu$  if its probability distribution is given by

$$P(X = x|\mu) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{for } x = 0, 1, 2, \dots$$

This distribution is particularly important in stochastic processes, where it counts the number of events in a given period of time. For example, for a Non-Homogeneous Poisson Process with failure rate function  $\lambda(t)$ , the number of events in time interval  $[t_1, t_2)$  has a Poisson distribution with expected value  $\mu = \int_{t_1}^{t_2} \lambda(u)du$ .

## 5 Other distributions

Many other parametric probability distributions play an important role in some specific reliability applications, we briefly mention some examples but refer to the literature, e.g. the sources mentioned in Section 1, for more details. Historically, the Log-logistic distribution was frequently used in reliability, with a lifetime  $T$  having a Log-logistic distribution if  $Y = \ln T$  has a Logistic distribution with pdf

$$f(y|\mu, \sigma) = \frac{\exp[(y - \mu)/\sigma]}{\sigma(1 + \exp[(y - \mu)/\sigma])^2} \quad \text{for } -\infty < y < \infty$$

with parameters  $-\infty < \mu < \infty$  and  $\sigma > 0$ . This distribution is very similar to the Normal distribution, but its survival function is available in closed form and hence it makes it easier to deal with right-censored observations, which often occur in reliability applications.

The Uniform distribution, which simply has a constant pdf over a finite interval, is also useful in some reliability applications. For example, it models the times of events in a Homogeneous Poisson Process if the total number of events over a time interval of pre-determined length is given. It is also frequently used in Bayesian statistics, as a prior distribution which, as is sometimes argued, can reflect quite limited prior knowledge being available. The Beta distribution is also popular in Bayesian statistics, as it is a conjugate prior for Binomial sampling models.

The Gompertz distribution is characterized by an attractive form for the hazard rate, namely  $h(t) = \lambda\phi^t$  for  $t > 0$ , with scale parameter  $\lambda > 0$  and shape parameter  $\phi > 0$ . Clearly,  $\phi = 1$  gives the Exponential distribution,  $\phi > 1$  models an increasing hazard rate (and hence can be used to model wear-out), but for  $\phi < 1$ , so decreasing hazard rate



(‘wear-in’), a problem occurs as the corresponding probability distribution is improper (i.e. its density does not integrate to one for  $t > 0$ ). This latter aspect could be interpreted as if a proportion of the population considered cannot experience the event of interest. We mention this distribution mostly due to its apparently attractive hazard rate, and to emphasize the complications that can occur already with such rather simple mathematical forms. The Gompertz distribution has been used in actuarial mathematics since the early 19th century.

When we presented the Weibull distribution in Section 3, we briefly referred to an Extreme Value distribution. Generally, there are several Extreme Value distributions (3 main functional forms), which occur for the maximum (or minimum) of  $n$  identically distributed real-valued random quantities if  $n \rightarrow \infty$ , and which often provide good approximations for the distribution of this maximum (or minimum) if  $n$  is large. These distributions are useful in a variety of reliability applications, for example related to reliability of systems or to structural reliability and overloads.

In this paper, we have restricted attention to uni-variate random quantities. Of course, in many reliability applications one is interested in multi-variate random quantities, we refer to Hougaard [2] for an excellent presentation of related statistical theory, including useful parametric distributions for multi-variate data.

## 6 Concluding remarks

Basic parametric probability distributions, such as presented in this paper, are also often used as part of more complex statistical models, such as mixture models, Bayesian hierarchical models, models with co-variates or parametric models for processes, we refer to the literature (see Section 1) for further discussion of such models, several examples are provided in other articles in this encyclopedia. Parametric models are certainly not always suitable, in particular if one has many data available the use of non-parametric or semi-parametric methods might be preferable due to their increased flexibility to adapt to specific features of the data. Due to increased computational power, applications of non- and semi-parametric methods have become more widely available, yet parametric distributions are likely to remain important tools in reliability and risk assessment.

It is often an advantage if the parameters have an intuitive interpretation. Several of the parametric probability distributions discussed in this paper, including Normal, Exponential and Gamma (but unfortunately not the Weibull distribution unless its

shape parameter is assumed to be a known constant), belong to the so-called Exponential Family of distributions [7], for which the parameters can be related to one- or two-dimensional sufficient statistics which summarize the data, which has the added benefit of available conjugate priors to simplify computation in Bayesian statistics.

When choosing a parametric distribution as a model in a specific application, one often has to look for a suitable balance between simplicity of the model and corresponding computational aspects, and how flexible and realistic the model is. As mentioned before, for trustworthiness of statistical inferences based on an assumed parametric model, it is important that the model choice is well explained and, where possible, diagnostic methods (e.g. ‘goodness-of-fit tests’ [1]) are used to check if the model fits well with the available data.

## References

- [1] Kotz, S., Read, C.B., Balakrishnan, N. & Vidakovic, B. 2006. *Wiley Encyclopedia of Statistical Sciences*. Wiley, New York.
- [2] Hougaard, P. 2000. *Analysis of Multivariate Survival Data*. Springer, New York.
- [3] Lawless, J.F. 1982. *Statistical Models and Methods for Lifetime Data*. Wiley, New York.
- [4] Martz, H.F. & Waller, R.A. 1982. *Bayesian Reliability Analysis*. Wiley, New York.
- [5] Meeker, W.Q. & Escobar, L.A. 1998. *Statistical Methods for Reliability Data*. Wiley, New York.
- [6] Padgett, W.J. & Tomlinson, M.A. 2004. Inference from accelerated degradation and failure data based on Gaussian process models. *Lifetime Data Analysis*, **10**, 191-206.
- [7] Lee, P.M. 1997. *Bayesian Statistics: An Introduction* (2nd ed.). Arnold, London.