2.1 INTRODUCTION

In this chapter we consider an outline of some distributions that are useful in the field of reliability analysis. Any distribution of non-negative random variables could be used to describe durations. The distributions to be discussed here are all continuous. Through the literature on reliability analysis, certain parametric models have been used repeatedly such as exponential and Weibull models. These distributions have closed form expression for reliability and hazard functions. Log-normal and gamma distributions are generally less convenient computationally, but are still frequently applied.

Below we discuss some of the standard failure time models for homogeneous populations. The properties and the theoretical bases of these distributions are considered here. The distributions will be studied in the simplest case of independently and identically distributed random variables. Let \( T \) be a non-negative random variable representing the lifetimes of individuals in some population, which we are interested in making inference about. We will apply parametric models, to the breast cancer censored data and complex amplifier real data. We also give all the corresponding results and compare the models: Exponential, Weibull, Gamma, Log-normal, Generalized gamma.

2.2 RELIABILITY DISTRIBUTION

2.2.1 Exponential distribution
The exponential model \((T \sim \text{E}(\lambda))\) is the simplest parametric model and assumes a constant risk over time, which reflects the property of the distribution approximately called ‘lack of memory’. The probability to die within a particular time interval depends only on the length but not on the location of this interval. This means that the distribution of \(T - t\) conditional on \(T \geq t\) is the same as the original distribution. In other words, it holds that

\[
P(t \leq T < t + \Delta t \mid T \geq t) = P(T < \Delta t)
\]

for any positive \(\Delta t\). As a consequence, the exponential distribution (as the only one) is not influenced by the definition of time zero. The parameter \(\lambda\) attains all positive values and the distribution with \(\lambda = 1\) is called the unit or standard exponential. Therefore, the following formulae can be derived by one simple algebraic calculation:

- probability density function \( f(t) = \lambda e^{-\lambda t} \)
- reliability function \( R(t) = e^{-\lambda t} \)
- hazard function \( h(t) = \lambda \)
- cumulative hazard function \( F(t) = \lambda t \)

for all \( t \geq 0, \lambda > 0 \).

The exponential distribution was widely used on the reliability of electronic components and technical systems. The distribution of \( C \) with a positive constant \( c \) is again exponentially distributed with parameter \( \frac{\lambda}{c} \). The minimum of \( n \) independent exponential random variables with parameter \( \lambda \) is still exponential with parameter \( n \).
The model is very sensitive to even a modest variation because it has only one adjustable parameter, the inverse of which is both mean and standard deviation. Recent works have overcome this limitation by using more flexible distribution. Applications of the exponential distribution as a suitable model for reliability distribution include the area of cancer (Galli et al., 1983; Niederjohn et al., 1986), bacterial carriage (Bolin, 1986), Sensory-evoked potentials (Meyer et al., 1991), trends of smoking cessation (Spratt et al., 1992), national infant and child mortality (Riggs, 1990).

2.2.2 Weibull distribution

The Weibull model (introduced by Waloddi Weibull in 1939) is an important generalization of the exponential model with two positive parameters. The second parameter in the model allows great flexibility of the model and different shapes of the hazard function. The convenience of the Weibull model for empirical work stems on the one hand from this flexibility and on the other from the simplicity of hazard and reliability function.

probability density function \( f(t) = \lambda \beta (\lambda \beta t)^{\beta - 1} e^{-(\lambda \beta t)^\beta} \)

reliability function \( R(t) = e^{-(\lambda \beta t)^\beta} \)

hazard function \( h(t) = \lambda \beta (\lambda \beta t)^{\beta - 1} \)

cumulative hazard function \( F(t) = \lambda t^\beta \)

for all \( t > 0, \lambda > 0, \beta > 0 \).
Where Γ denotes the Gamma function with \( \Gamma(k) = \int_0^\infty u^{k-1}e^{-u} \, du \), \( k > 0 \). We abbreviate the distribution as \( W(\lambda, \beta) \). In the case of \( \beta = 1 \), the exponential distribution is obtained. The hazard function is monotone increasing if \( \beta > 1 \), decreasing if \( \beta < 1 \), and constant for \( \beta = 1 \).

![Mortality hazard](image)

Figure 2.1: Weibull hazard functions with different shape parameters.

If \( T \sim W(\lambda, \beta) \), then it holds that \( cT \sim W(\lambda, \beta) \), when \( c \) is a positive constant. Furthermore, the minimum of \( n \) independently and identically distributed variables from this distribution is \( W(n, \beta) \) (minimum – stable distribution). The Weibull distribution can also be generated as the limiting distribution of the minimum of a sample from a continuous distribution with support on \([0, u]\) for some \( u \) \((0 < u < \infty)\). This extreme value character makes the Weibull distribution appropriate for the distribution of individual time to death, because there are different causes of death which compete with each other and the first one to strike will kill the individual.
The Weibull hazard has been theoretically derived for cancer incidence by Pike (1966), but it is unknown whether it has relevance for other diseases. The Weibull distribution is inappropriate when the hazard rate is indicated to be unimodal or bathtub-shaped. A generalization of the Weibull distribution to include such kind of shapes was proposed by Mudholkar et al. (1996). Application of lifetime or durability of manufactured items is common, and it is used as a model with diverse types of items, such as ball bearings, automobile components, and electrical insulation. It is also used in biological and medical applications, for example, in studies on the time to the occurrence of tumors in human population or in laboratory animals. The Weibull has been used, among other applications, to investigate carcinogenesis experiments by Williams (1978), to characterize radiation response by Scott & Hahn (1980), to investigate the time between release and return to prison for criminal recidivists by Schmidt & Witte (1988), to model human mortality distribution by Juckett & Rosenberg (1993), to analyze trends of smoking cessation by Elketroussi & Fan (1977), and to model the treatment-free incubation period of AIDS by Hendriks et al. (1993).

2.2.3 Gamma distribution

The gamma distribution includes the exponential distribution as a special case. The gamma distribution is of limited use in reliability analysis because the gamma models do not have closed form expression for reliability and hazard functions. Both include the incomplete gamma integral

\[ I(k, x) = \frac{1}{\Gamma(k)} \int_0^x u^{k-1} e^{-u} \, du \]
Consequently, traditional maximum likelihood estimation is difficult and requires the calculation of such incomplete gamma integrals, which imposes additional numerical problems in parameter estimation. A random variable $T$ is gamma distributed with parameter $k$ and $\Lambda(T \sim \Gamma(k, \lambda))$, if the following holds:

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

probability density function

$$h(t) = 1 - l(k, \lambda)$$

reliability function

$$h(t) = \frac{\lambda^n e^{k^k - k\lambda}}{(1 - l(k, \lambda)) \Gamma(k)}$$

hazard function

$$F(t) = -l \left(1 - l(k, \lambda)\right)$$

cumulative hazard function

for all $k > 0$, $\lambda > 0$, $t > 0$.

If $k = 1$, the gamma distribution is reduced to the exponential distribution. With integer $k$, the gamma distribution is sometimes called a special Erlangian distribution. It can be derived as the distribution of waiting time to the $k$-th emission from a Poisson source with intensity parameter $\lambda$. Consequently, the sum of $k$ independent exponential variates with parameter $\lambda$ has a gamma distribution with parameters $k$ and $\lambda$ and can be used to model life times of technical systems with repeated repairing after failure. An extension of this idea was used by Aalen (1992) dealing with the compound poisson distribution. In spite of, the fact that the gamma distribution is of limited value as a lifetime distribution and the gamma distribution is a widely used frailty (mixture) distribution because of some neat mathematical features. It is a flexible distribution that takes on a variety of different shapes as $k$ varies. Further, frailty cannot be negative and the
gamma distribution is, along with the log-normal and Weibull distribution, one of the most commonly used distributions to model positive random variables. The assumption that frailty is gamma distributed yields some useful mathematical results, indicating that the frailty among survivors of any age \( t \) is again gamma distributed and frailty among those who die at any time \( t \) (Vaupel et al., 1979). The gamma distribution has been applied to hepatograms in normal adults and in patients with cirrhosis and obstructive jaundice (Galli et al., 1983), speech amplitude (Niederjohn et al., 1986), platelet survival (Bolin, 1986), genetic influences on the age-at-menarche (Meyer et al., 1991) and treatment-free incubation time of AIDS Hendriks et al. (1993).

2.2.4 Log-normal distribution

In the log-normal model \( \{ T \sim \ln \ (\mu, \sigma^2) \} \), the natural logarithm \( \ln(T) \) of the lifetime \( T \) is assumed to be normally distributed \( \{ \ln(T) \sim N(\mu, \sigma^2) \} \). A log normal distribution results when the variable is the product of a large number of independent, identically distributed variables in the same way that a normal distribution results when the variable is the sum of a large number of independent, identically distributed variables. The reliability and hazard function include the incomplete normal integral

\[
\Phi(x) = \int_{-\infty}^{x} \phi(u) \, du ,
\]

where \( \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \) denotes the probability density function of a standard normal distribution. Consequently,

probability density function \( f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\ln(t) - \mu)^2}{2\sigma^2}} \)

reliability function \( H(t) = 1 - \Phi \left( \frac{\ln(t) - \mu}{\sigma} \right) \)
hazard function \( h(t) = \frac{e^{-\frac{(\ln(t)-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi} \sigma \left(1 - \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)\right)} \)

cumulative hazard function \( F(t) = -\ln \left(1 - \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)\right) \)

for all \( t > 0 \).

The log-normal distribution may be convenient to use with non-censored data, but when this distribution is applied to censored data, the computations quickly become formidable. Unfortunately, the hazard function has a strange form: it has value zero at \( t = 0 \), increases to a maximum and then decreases, approaching zero as \( t \) heads to infinity. Because of the decreasing form of the hazard function for older ages, the distributions seem implausible as a lifetime model in most situations. Nevertheless, it makes sense if interest is focused on time periods of younger ages. Despite its unattractive features, the log-normal distribution has been widely used as failure distribution in diverse situations, such as the analysis of electrical insulation or time to occurrence of lung cancer among smokers.

Furthermore, the log-normal distribution has often been used as a frailty (mixing) distribution. Especially in the context of unobserved normal distributed covariates in the Cox model, the log-normal frailty distribution provides an appealing interpretation of the model. The log-normal distributions are in practice very close to the inverse Gaussian distributions.

The lognormal distribution has been applied frequently in reliability and biomedical research, partially due to its close relationship with the normal distribution. For
example, Horner (1987) found that the distribution of age at onset of Alzheimer’s disease can be approximated by the lognormal distribution. Larsen et al. (1991) used a lognormal model to relate human lung function (FEV1) decrease to \( C_2 \) exposure, and Ahmed et al. (1993) found that the lognormal distribution was appropriate in a study of human health risk due to consumption of chemically contaminated fishery products. They reported that the lognormal distribution provided good descriptions of the pattern of variation of contaminant concentrations among different species and geographic areas. The results offer an opportunity to reduce exposure through restricting harvested fishery products from certain areas and by excluding certain species.

### 2.2.5 Log-logistic distribution

An alternative model to the Weibull distribution is the log-logistic distribution. The log-logistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped that is it initially increases and then decreases. The distribution imposes the following functional forms on the density, reliability, hazard and cumulative hazard function:

**Probability density function**

\[
    f(t) = \frac{\left(\frac{\beta}{\alpha}\right) \left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1 + \left(\frac{t}{\alpha}\right)^{\beta}\right]^2}
\]

**Reliability function**

\[
    R(t) = \left[1 + \left(\frac{t}{\alpha}\right)^{\beta}\right]^{-1}
\]

**Hazard function**

\[
    h(t) = \frac{\left(\frac{\beta}{\alpha}\right) \left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1 + \left(\frac{t}{\alpha}\right)^{\beta}\right]}
\]
The general shape of the hazard function of a log-logistic distribution is very similar to that of the log-normal distribution. The log-logistic distribution can be obtained as a mixture of Gompertz distributions with a gamma distributed mixture variable with mean and variance equal to one. The special features of the hazard function of the log-logistic distribution provide a good alternative to the Weibull, lognormal and gamma distribution. It has been used to model the time to the return to prison in criminology (Schmidt and Witte, 1988) and the time between smoking cessation and restarting (Elketroussi and Fan, 1977).

2.2.6 Gompertz distribution

In 1825 the British actuary Benjamin Gompertz made a simple but important observation that a law of geometrical progression pervades large portions of different tables of mortality for humans. The simple formula he derived describing the exponential rise in death rates between sexual maturity and old age is commonly referred to as the Gompertz equation—a formula that remains a valuable tool in demography and in other scientific disciplines. Gompertz’s observation of a mathematical regularity in human life tables led him to believe in the presence of a law of mortality that explained why common age patterns of death exist. It has been widely used, especially in actuarial and biological applications and in demography. A random variable follows a Gompertz distribution with parameters \( a > 0 \) and \( b > 0 \) (\( T \sim G \sim (a, b) \)), if the following relation holds:

\[
\text{probability density function } f(t) = ae^b e^{\frac{-a}{t} (e^t - 1)}
\]
reliability function \( K(t) = \frac{u}{b}(e^b - 1) \)

hazard function \( h(t) = u e^b \)

cumulative hazard function \( F(t) = \frac{u}{b}(e^b - 1) \)

The hazard function is increasing from \( a \) at time zero to \( \infty \) at time \( \infty \). The model can be generalized to the Gompertz-Makeham distribution by adding a constant to the hazard:

\[
h(t) = u e^b + \gamma.
\]

Figure 2.2 Gompertz hazard functions with different parameters.

The distribution was compared to the Weibull distribution as a descriptor for human mortality distributions by Juckett & Rosenberg (1993) and to the exponential distribution and other models in breast cancer growth by Spratt et. al. (1992). It was also used by Riggs in a series of papers on adult mortality due to a variety of diseases such as Parkinson’s disease (1990), lung cancer (1991), prostate cancer (1991), stroke and emphysema (1992).
2.2.7 Generalized Gamma distribution

The generalized gamma function is a three parameter distribution. One version of
the generalized gamma distribution uses the parameters $k, \beta, \theta$. The probability density
function for this form of the generalized gamma distribution is given by

$$f(t) = \frac{\beta}{\Gamma(k) \theta} \left( \frac{t}{\theta} \right)^{k-1} e^{-\left( \frac{t}{\theta} \right)^{\beta}}$$

where $\theta > 0$ is a scale parameter, $\beta > 0$ and $k > 0$ are shape parameters.

Reparametrisation with parameters $k, \beta, \theta$,

where $\mu = \ln(\theta) + \frac{1}{\beta} \ln\left( \frac{1}{\lambda^2} \right)$, $\sigma = \frac{1}{\beta \sqrt{k}}$, $\lambda = \frac{1}{\sqrt{k}}$

where $-\infty < \mu < \infty$, $\sigma > 0$, $\lambda > 0$.

The probability density function of the reparametrised distribution is given by

$$f(t) = \begin{cases} 
\left| \frac{\lambda}{\sigma} \right| \Gamma\left( \frac{1}{\lambda^2} \right) e^{\frac{\mu(t) - \mu + \ln(\frac{1}{\lambda^2})}{\sigma^2}} \frac{e^{\frac{\lambda(t) - \mu}{\lambda^2} - e^{\frac{\lambda(t) - \mu}{\lambda^2}}}}{\lambda^2} & \text{if } \lambda \neq 0 \\
\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln(t) - \mu}{\sigma} \right)^2} & \text{if } \lambda = 0
\end{cases}$$

The Reliability function for the generalized gamma distribution is given by

$$H(t) = \begin{cases} 
1 - \Gamma\left( \frac{\lambda(t) - \mu}{\lambda^2}; \frac{1}{\lambda^2} \right) & \text{if } \lambda > 0 \\
1 - \Phi\left( \frac{\ln(t) - \mu}{\sigma} \right) & \text{if } \lambda = 0 \\
\Gamma\left( \frac{\lambda(t) - \mu}{\lambda^2}; \frac{1}{\lambda^2} \right) & \text{if } \lambda < 0
\end{cases}$$
where $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} \, dx$ and $\Gamma_{\gamma}(k; x)$ is the incomplete gamma function of $k$ and $x$.

The generalized gamma distribution includes other distributions as special cases based on the values of the parameters.

![Generalized Gamma Family Distributions](image)

Fig. 2.3 Generalized Gamma family distributions

The Weibull, the exponential, the lognormal and the gamma distributions are shown to be the particular cases of the generalized gamma distributions when

- (i) $\lambda = 1$, $\alpha = 1$  
- (ii) $\lambda = 1$, $\alpha = 0$  
- (iii) $\alpha = 1$, $\nu = \frac{1}{1-\lambda}$

2.3 ANALYTICAL METHOD

2.3.1 Maximum Likelihood Estimator

Maximum likelihood is a simple method of constructing an estimator for an unknown parameter $\theta$. It was introduced by R. A. Fisher in 1912. If the sample is large, the method will yield an tremendous estimator of $\theta$. For these reasons, the method of maximum likelihood is probably the most widely used method of estimation in statistics. Suppose that the random variables $X_1, X_2, \ldots, X_n$ form a random sample from a distribution $f(x|\theta)$; if $X$ is a continuous random variable, $f(x|\theta)$ is probability density
function, if $X$ is discrete random variable $f(x|\theta)$ is probability mass function. For every observed random sample $x_1, x_2, \ldots, x_n$, we define

$$f(x_1, x_2, \ldots, x_n|\theta) = f(x_1|\theta) \cdots f(x_n|\theta)$$ (2.1)

If $f(x|\theta)$ is probability density function, $f(x_1, x_2, \ldots, x_n|\theta)$ is the joint probability density function; if $f(x|\theta)$ is probability mass function, $f(x_1, x_2, \ldots, x_n|\theta)$ is the joint probability mass function. Now we name $f(x_1, x_2, \ldots, x_n|\theta)$ as the likelihood function. As we can see, the likelihood function depends on the unknown parameter $\theta$, and it is always denoted as $L(\theta)$.

Suppose, for the moment, that the experiential random sample $x_1, x_2, \ldots, x_n$ came from a discrete distribution. If an estimate of $\theta$ must be selected, we would certainly not consider any value of $\theta$ for which it would have been impossible to obtain the data $x_1, x_2, \ldots, x_n$ that was actually experiential. Furthermore, suppose that the probability $f(x_1, x_2, \ldots, x_n|\theta)$ of obtaining the actual experiential data $x_1, x_2, \ldots, x_n$ is very high when $\theta$ has a particular value, say $\theta = \theta_0$, and is very small for every other value of $\theta$. Then we would naturally estimate the value of $\theta$ to be $\theta_0$. When the sample comes from a continuous distribution, it would again be natural to try to find a value of $\theta$ for which the probability density $f(x_1, x_2, \ldots, x_n|\theta)$ is large, and to use this value as an estimate of $\theta$. For any given experiential data $x_1, x_2, \ldots, x_n$, we are led by this reasoning to consider a value of $\theta$ for which the likelihood function $L(\theta)$ is a maximum and to use this value as an estimate of $\theta$.

We choose the parameter that makes the likelihood of having the obtained data at hand maximum. With discrete distributions, the likelihood is the same as the probability.
We choose the parameter for the density that maximizes the probability of the data coming from it. Theoretically, if we had no actual data, maximizing the likelihood function will give us a function of $n$ random variables $X_1, X_2, \ldots, X_n$, which we shall name maximum likelihood estimate, $\hat{\theta}$. When there are actual data, the estimate takes a particular numerically value, which will be the maximum likelihood estimator.

MLE requires us to maximum the likelihood function $L(\theta)$ with respect to the unknown parameter $\theta$. From equation 2.1, $L(\theta)$ is defined as a product of $n$ terms, which is not easy to maximized. Maximizing $L(\theta)$ is equivalent to maximizing $\prod_{i=1}^{n} f(X_i|\theta) = \sum_{i=1}^{n} f(X_i|\theta)$ because log is a monotonic increasing function. We define $L(\theta)$ as log likelihood function, we denote it as $l(\theta)$,

$$
l(\theta) = \log \prod_{i=1}^{n} f(X_i|\theta) = \sum_{i=1}^{n} \log f(X_i|\theta)$$  \hspace{1cm} (2.2)

Maximizing $l(\theta)$ with respect to $\theta$ will give us the MLE estimation.

If is frequently easier to work with the natural log of the likelihood function. For this is simply called the log likelihood. Since $\log$ is an increasing function, the maxima of the likelihood and log likelihood coinciding. The method of Maximum likelihood (Harter and Moore, 1965; Cohen, 1965) is a commonly used procedure because it has very desirable properties.

### 2.3.2 Likelihood-ratio test:

A very accepted form of hypothesis test is the likelihood ratio test, which is a generalization of the best test for simple null and alternative hypotheses that was developed by Neyman and Pearson. The likelihood ratio test is based on the likelihood
function \( f_n(X - 1, \ldots, X_n|\theta) \), and the insight that the likelihood function tends to be highest near the true value of \( \theta \).

A likelihood ratio test is a statistical test used to compare the fit of two models, one of which (the null models) is a special case of other (the alternative model). The test is on the likelihood ratio, which express how many times more likely the data are under one model than the other. This likelihood ratio, or equivalently its logarithm, can then be used to compute a p-value, or compared to a critical value to decide whether to reject the null model in favor of the alternative model. Both models are fitted to the data and their log-likelihoods are recorded. The test statistic (usually denoted D) is twice the difference in these log-likelihoods:

\[
U = -2[\ln(l_i ho f n l m ) - \ln(l_i ho f a m )]
\]

The use of likelihood ratio in statistical inference is common (Edwards, 1972; Royall, 1997) and the role of likelihood in model comparison is well established (Akaike, 1973; Schwartz, 1978). Furthermore, the likelihood ratio plays a pivotal role in most approaches to hypothesis testing. Neyman and Pearson (1928) pointed out that likelihood ratio test, it has become one of the most accepted methods for testing restrictions on the parameters of a statistical model. It is well known that minus twice the likelihood ratio test has a limiting central chi-square distribution under the null hypothesis (Wilks, 1938) and a limiting noncentral chi-square distribution under sequences of local alternatives (Wald, 1943). However, as Foutz and srivastava (1977), Kent (1982), and White (1982) advocated that, then the largest model is misspecified, the likelihood ratio test is no longer necessarily chi-square distributed under the null hypothesis, where the null hypothesis must be redefined in terms of the Pseudo-true values satisfying the specified restrictors.
The likelihood ratio test has also been advocated as a basis for testing non-nested models (Cox (1961, 1962)).

2.4 APPLICATION TO CENSORED DATA

*Parametric Regression Models in the Analysis of Breast Cancer Survival data*

2.4.1 Introduction

The incidence of breast cancer has increased globally over the last several decades (Hortobagyi, 2005; Anderson, 2008; Porter, 2008), the greatest increase has been in Asian countries (Green et al., 2008). In Asia, breast cancer incidence peaks among women in their forties (Agarwal et al., 2007). It is expected that in the coming decades, these countries would account for majority of new breast cancer patients diagnosed globally. Over 1,00,000 new breast cancer patients are estimated to be diagnosed annually in India (Nandakumar, 1995; Agarwal et al., 2007). A study conducted by the Harvard School of Public Health revealed that 1.35 million cases of breast cancer would be diagnosed worldwide in 2009 accounting for 10.5% of new cancers, second only to lung cancer. Breast cancer cases are expected to increase by 26% by 2020 and most of these will be seen in developing countries (Anonymous, 2009).

Data from the International Agency for Research on Cancer (IARC) registry suggest that 45% of newly diagnosed cases of breast cancer and 55% of breast cancer related mortality currently occur in low- and middle income countries. IARC trends also show a 20-30% increase in the incidence of breast cancer in developing countries during the past decade (Curado et al., 2009). As per the ICMR-PBCR data, breast cancer is the commonest cancer among women in urban registries of Delhi, Mumbai, Ahmedabad,
Kolkata, and Trivandrum where it constitutes >30% of all cancers in females (National Cancer Registry Programme, 2001). In the rural PBCR of Barshi, breast cancer is the second commonest cancer in women after cancer of the uterine cervix (National Cancer Registry Programme, 2001).

A control study conducted at a government medical college at Nagpur, lack of or less duration of breast feeding associated with the risk of breast cancer. Risk of breast cancer was more for women who had menopause after 50 years compared to women who had menopause before 45 years of age. Although breast cancer can be detected at earlier stages by simple breast examination, maximum (> 90%) cases are diagnosed in advance stages i.e., stage II, III and IV (Meshram et al., 2009). Radiotherapy is an important role in the treatment of breast cancer at every stage. With appropriate treatment, many women are cured of breast cancer, while many other live longer with the disease and have a better quality of life. International guidelines recommend one megavoltage therapy equipment for every 120,000 population (Ravichandran, 2009).

The objective is to compare the performance of parametric models using German Breast Cancer data.

2.4.2 Description of the data set

The data obtained from the website, ftp://ftp.wiley.com/public/scitech_med/survivaland http://www.umass.edu/statdata/statdata. Cancer clinical trials are a rich source for examples of applications of methods for the analysis of time to event. Willi Sauerbrei and Patrick Royston have graciously provided the data from the German Breast Cancer Study Group, which they used to illustrate methods for building prognostic models (Sauerbrei and Royston, 1999). In the main study,
a total of 720 patients with primary node positive breast cancer were recruited between July 1984, and December 1989, (Schmoor, Olschewski and Schumacher M. 1996 and Schumacher et al. (1994)). This data consist of the German Breast Cancer study of 686 patients with 16 variables. The event of interest is survival time. These are the covariates considered here, age (years), menopausal status (1 = yes and 2 = no), hormone Therapy (1 = yes and 2 = no), tumor size (mm), Number of nodes involved (1-51), tumor grade (1–3), number of progesterone receptors (1–2380) and number of estrogen receptors (1–1144). Event is coded as 1 and censoring is coded as 0.
2.4.3 Description of Variables

The variables and codes for the data provided in the following table:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Name</th>
<th>Description</th>
<th>Codes/ Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>id</td>
<td>Study ID</td>
<td>1-686</td>
</tr>
<tr>
<td>2</td>
<td>diagdate</td>
<td>Date of Diagnosis</td>
<td>dd/mm/yyyy</td>
</tr>
<tr>
<td>3</td>
<td>recdate</td>
<td>Date of Recurrence or of Recurrence</td>
<td>dd/mm/yyyy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Free Survival</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>deathdate</td>
<td>Date of Death</td>
<td>dd/mm/yyyy</td>
</tr>
<tr>
<td>5</td>
<td>age</td>
<td>Age of Diagnosis</td>
<td>Years</td>
</tr>
<tr>
<td>6</td>
<td>menopause</td>
<td>Menopausal Status</td>
<td>1 = Yes, 2 = No</td>
</tr>
<tr>
<td>7</td>
<td>hormone</td>
<td>Hormone Therapy</td>
<td>1 = Yes, 2 = No</td>
</tr>
<tr>
<td>8</td>
<td>Size</td>
<td>Tumor Size</td>
<td>mm</td>
</tr>
<tr>
<td>9</td>
<td>grade</td>
<td>Tumor Grade</td>
<td>1 - 3</td>
</tr>
<tr>
<td>10</td>
<td>nodes</td>
<td>Number of Nodes involved</td>
<td>1 - 51</td>
</tr>
<tr>
<td>11</td>
<td>prog_recp</td>
<td>Number of Progesterone Receptors</td>
<td>1 - 2380</td>
</tr>
<tr>
<td>12</td>
<td>estrg_recp</td>
<td>Number of Estrogen Receptors</td>
<td>1 - 1144</td>
</tr>
<tr>
<td>13</td>
<td>rectime</td>
<td>Time of Recurrence</td>
<td>Days</td>
</tr>
<tr>
<td>14</td>
<td>censrec</td>
<td>Recurrence Censoring</td>
<td>0 = Censored</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 = Recurrence</td>
</tr>
<tr>
<td>15</td>
<td>survtime</td>
<td>Time to Death</td>
<td>Days</td>
</tr>
<tr>
<td>16</td>
<td>censdead</td>
<td>Death Censoring</td>
<td>0 = Censored</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 = Death</td>
</tr>
</tbody>
</table>
### 2.4.4 Model Results

Table 2.2 Parametric Regression model Fitted to Breast Cancer Data.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>COVARIATES</th>
<th>EXPONENTIAL</th>
<th>WEIBULL</th>
<th>GOMPertz</th>
<th>LOGNORMAL</th>
<th>LOGLOGISTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Coef.</td>
<td>S.E.</td>
<td>Coef.</td>
<td>S.E.</td>
<td>Coef.</td>
</tr>
<tr>
<td>1</td>
<td>AGE</td>
<td>0.00347</td>
<td>0.01207</td>
<td>0.00670</td>
<td>0.01208</td>
<td>0.00512</td>
</tr>
<tr>
<td>2</td>
<td>MENOPAUSE</td>
<td>0.12147</td>
<td>0.25176</td>
<td>0.10616</td>
<td>0.25280</td>
<td>0.12117</td>
</tr>
<tr>
<td>3</td>
<td>HORMONE</td>
<td>-0.21458</td>
<td>0.16864</td>
<td>-0.28528</td>
<td>0.16828</td>
<td>-0.28545</td>
</tr>
<tr>
<td>4</td>
<td>SIZE</td>
<td>0.01200 *</td>
<td>0.00479</td>
<td>0.01357 *</td>
<td>0.00483</td>
<td>0.01340 *</td>
</tr>
<tr>
<td>5</td>
<td>NODES</td>
<td>0.04702 *</td>
<td>0.00952</td>
<td>0.05501 *</td>
<td>0.00961</td>
<td>0.05345 *</td>
</tr>
<tr>
<td>6</td>
<td>GRADE</td>
<td>0.38754</td>
<td>0.14478</td>
<td>0.43986 *</td>
<td>0.14606</td>
<td>0.44132 *</td>
</tr>
<tr>
<td>7</td>
<td>PROG_RECP</td>
<td>-0.00503 *</td>
<td>0.00116</td>
<td>-0.00541 *</td>
<td>0.00119</td>
<td>-0.00533 *</td>
</tr>
<tr>
<td>8</td>
<td>ESTRG_RECP</td>
<td>-0.00016 *</td>
<td>0.00054</td>
<td>-0.00025</td>
<td>0.00055</td>
<td>-0.00019</td>
</tr>
</tbody>
</table>

DEVIANCE: 866.62246 | 812.7804 | 834.75706 | **798.36414** | 804.0081

* P<0.05

Table 2.3 Difference in deviance of various models compared to Lognormal

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Distributions</th>
<th>Difference in Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exponential</td>
<td>68.25832</td>
</tr>
<tr>
<td>2</td>
<td>Weibull</td>
<td>14.41626</td>
</tr>
<tr>
<td>3</td>
<td>Gompertz</td>
<td>36.39292</td>
</tr>
<tr>
<td>4</td>
<td>Loglogistic</td>
<td><strong>5.64396</strong></td>
</tr>
</tbody>
</table>

49
2.4.5 Results

The parametric models were fitted using STATA 12 and the results are presented in Table 2.2. From the table we see that the four covariates namely size, nodes, grade and progesterone receptors are significantly associated with the survival time under all the model assumptions. Among the models, the lognormal has the lowest level of deviance compared to all other models. The difference in deviance of the all other models compared to lognormal are given in Table 2.3 It is further noted that all other models have significantly higher deviance compared to lognormal.

2.4.6 Summary

A study conducted by Pourhoseingholi et al. (2007), pointed out, in multivariate models, Cox and Exponential are the same with respect to AIC and standardized variability. But in univariate, all parametric ones are better than Cox except for tumor size and the lognormal is the first choice among parametric models. Jiezhi Qi (2009), viewed that, after comparison of all the models and assessment of goodness of fit, found that the loglogistic AFT model fits better, for randomized placebo-controlled trial to prevent Tuberculosis in Uganda adults infected with HIV. Hayat et al. (2010), showed that age was not found as a risk factor, by using AIC, Gompertz model was more suitable, for Breast Cancer Registry Data from Ege University Cancer Research Center. Ponnuraja and Venkatesan (2010), applied likelihood based criteria for model selection indicated that the Gamma model was the best fitting parametric model for tuberculosis clinical trial data. Baghestani et al. (2010), the results indicated that the early detection of a cancer at a young patient age and in primary stages is important to increase survival from gastric cancer. According to statistical criteria, a parametric model can also be useful statistical
model to find prognostic factors in the presence of interval censoring. Deviance supported the loglogistic regression as the best option. Nakhaee and Law (2011), showed that, survival following a diagnosis of HIV infection was modeled by applying parametric survival models on people who were only diagnosed with HIV or HIV and AIDS registered in the national surveillance system from 1997 to 2003. likelihood based criteria for model selection indicated that the Weibull model was the best fitting parametric model for predicting survival following both HIV and AIDS diagnoses.

In spite of different models suggested and recommended by many researchers, who followed different methodologies in their experiments, the result of my present study has a strong inclination for the Lognormal method as the most suitable one; better than Exponential, Weibull, Gompertz, and Loglogistic model, mainly on the scale of “Deviance”.

2.5 APPLICATION TO REAL DATA

Comparison of Reliability Models for Lifetime Data

2.5.1 Introduction

Reliability is defined as the probability that a system, vehicle, machine, device and so on will perform its intended function under operating conditions, for a specified period of time. Improving reliability is an important part of the larger overall picture of improving product quality. There are so many definitions of quality, but in general that, an unreliable product is not a high quality product. Condra (1993) emphasizes that ‘reliability is quality over time’.

Various parametric families of models are used in the analysis of lifetime data and the modeling of aging or failure processes. Among univariate models, a few distributions occupy a central position because of their demonstrated usefulness in a wide range of situations. Foremost in the category are the exponential, Weibull, lognormal, loglogistics and gamma distributions. The relevant literature includes a paper by Dumonceaux, Antle and Hass (1973) who examined maximum likelihood ratio (MLR) test for discriminating between two models with unknown location and scale parameters. Dumonceaux and Antle (1973) gave an MLR procedure for discriminating between Weibull and lognormal distributions. Bain and Engelhordt (1980) considered a likelihood ratio selection statistic for selecting between gamma and Weibull distributions. Al-Garni et al. (2006) use Maximum likelihood method to produce parameter estimates.

The objective is to compare the performance of parametric models using Complex amplifier real data set.
2.5.2 Data for study:

We consider a real–life data set from Dimitri Kececioglu (1992, p.76). This data consist of times to failure for 112 amplifiers.

2.5.3 Model Results

Using Weibull++7 software, the distribution of data under study is obtained and is given in Table 2.4. It is observed that the mean and standard deviation of the distribution are determined as 496.04 and 14.117 respectively.

<table>
<thead>
<tr>
<th>Time to failure</th>
<th>No. of Failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>470-490</td>
<td>29</td>
</tr>
<tr>
<td>490-510</td>
<td>74</td>
</tr>
<tr>
<td>510-530</td>
<td>08</td>
</tr>
<tr>
<td>530-550</td>
<td>-</td>
</tr>
<tr>
<td>550-570</td>
<td>-</td>
</tr>
<tr>
<td>570-590</td>
<td>-</td>
</tr>
<tr>
<td>590-610</td>
<td>1</td>
</tr>
</tbody>
</table>

The estimates of the parameters of various models considered here are obtained and tabulated in Table 2.5.
Table 2.5 Estimates of parameter of various models

<table>
<thead>
<tr>
<th>Models</th>
<th>Parameters</th>
<th>Likelihood values</th>
<th>-2LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\lambda = 0.0020$</td>
<td>-807.14</td>
<td>1614.28</td>
</tr>
<tr>
<td>Weibull</td>
<td>$\beta = 19.2117, \eta = 5042247$</td>
<td>-512.13</td>
<td>1024.26</td>
</tr>
<tr>
<td>Gamma</td>
<td>$\mu = -0.9802, k = 13219899$</td>
<td>-451.57</td>
<td>903.14</td>
</tr>
<tr>
<td>Lognormal</td>
<td>$\mu = 6.2063, \sigma = 0.0272$</td>
<td>-449.99</td>
<td>899.98</td>
</tr>
<tr>
<td>Generalized</td>
<td>$\mu = 6.1975, \sigma = 0.0209, \lambda = -0.7680$</td>
<td>-431.73</td>
<td>863.46</td>
</tr>
</tbody>
</table>

Table 2.6 given below presents the deviance of various models when compared to the generalized gamma distribution.

Table 2.6 Deviance of various models when compared to the generalized gamma distribution

<table>
<thead>
<tr>
<th>Models</th>
<th>Likelihood Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>750.82</td>
</tr>
<tr>
<td>Weibull</td>
<td>160.80</td>
</tr>
<tr>
<td>Gamma</td>
<td>39.68</td>
</tr>
<tr>
<td>Lognormal</td>
<td>36.52</td>
</tr>
</tbody>
</table>

2.5.4 Graphs

Fig. 2.4: Probability density function
The curves of the probability density functions of the exponential, Weibull, gamma, lognormal and generalized gamma distributions are given in Fig.2.4.

![Cumulative Distribution Function](image1)

**Fig. 2.5 Cumulative distribution function**

The curves of the Cumulative distribution function of exponential, Weibull, gamma, lognormal and generalized gamma distributions are given in Fig.2.5.

![Hazard Function](image2)

**Fig. 2.6 Hazard function**
The curves of the Hazard function of exponential, Weibull, gamma, lognormal and generalized gamma distributions are given in Fig.2.6.

2.5.5 Summary

From Table 2.5, -2loglikelihood (-2LL) value of the exponential distribution, the parameter value \( \lambda \) of which is 0.0020, is 1614.28, the highest of all other distributions; whereas the Weibull distribution has come second next to the exponential distribution registering 1024.26; Having the highest -2LL values, the exponential and Weibull distributions could be easily assumed to be the most unfitting distributions. The difference between the -2LL values of gamma and lognormal in this case is very small, both recording 903.14 and 899.98 respectively. As there is no significant difference between gamma and lognormal distributions, both can be considered as symmetrical distributions and come as the second best feasible distributions.

The best of all is generalized gamma, -2loglikelihood value of which is 863.46, accommodating the fulfillment of the requirements of a typical distribution to a maximum extent. As having the lowest value, the generalized gamma distribution is the most suitable one. The deviance of various models compare to generalized gamma by the method of Likelihood ratios are given in Table 2.6. The chi-square value for one degree of freedom at 1% level is 3.841. Likelihood ratio values are greater than 3.841. There is significant difference in the models for 3-parameter. For, 2-parameter, the likelihood ratio value of the Weibull distribution is 160.80, the highest of all other distributions; whereas the gamma distribution has come second next to the Weibull distribution registering 39.68. As having the lowest value, is 36.52, the lognormal distribution is the most suitable one for two parameter.