CHAPTER 3

Exponential-Gamma $(\mathbf{3}, \theta)$ Distribution: A Bathtub Shaped Failure Rate Model

3.1 Introduction

Modeling and analyzing lifetime data using mixture distribution is a prominent practice in many applied sciences, such as medicine, engineering, and finance. Mixture distributions are useful when dealing with lifetime data analysis. When a new component switches on for the first time, it may fail at the same instant, or it may fail due to overvoltage, jerking, or any such shocks. Failure due to random shocks can be modeled using an exponential distribution, while failure due to the degradation of components occurs. Failure time may be distributed as a Gamma distribution, Weibull distribution, or any other lifetime distribution if it is fitted to the data. When a group of lifetimes consists of lifetimes due to both types of failures, such as random failures and failures due to degradation, one should use a mixture.

A variety of distributions can be used to model lifetime data, though the failure rate functions of the majority of them do not exhibit bathtub shapes. However, many real-life systems demonstrate BFR functions. To address this discrepancy, distributions like the exponentiated Weibull by Pal et al.(2006), exponentiated gamma by Nadarajah and Gupta (2007), generalized Lindley by Nadarajah et al. (2011), and X-exponential by Chacko (2016) have been proposed to model lifetime data with bathtub-shaped failure rate models.

Models with bathtub-shaped failure rate functions apply to reliability analysis, particularly in reliability-related decision-making, cost analysis, and burn-in analysis. It is necessary to use exponential distributions when dealing with random failures and other lifetime distributions when dealing with failures due to ageing in such situations. The purpose of this chapter is to examine a mixture of an exponential distribution and a gamma distribution that has a BFR function. Real-world problems can be accurately modeled by this distribution.

The introduction of a mixture distribution uses gamma and exponential distributions in many different areas. This modeling strategy is useful when working with populations, systems, or datasets that have intrinsic differences in their properties. The exponential distribution is used to represent constant failure rates, whereas the gamma distribution, with a shape value of 3, describes wear-out failure mechanisms. The occurrence of various behaviors within a population or system can be explained by using these two distributions as a mixture. The fact that we can produce bathtub-shaped failure rate behavior for this combination distribution is a major concern. This is beneficial in reliability analysis, health research, financial modeling, quality control, and other fields.

This chapter is organized as follows. Section 3.2 considers the exponential-gamma $(\mathbf{3},\theta)$ distribution. In section 3.3, various statistical properties of the exponential-gamma $(\mathbf{3},\theta)$ distribution are derived. The estimation procedure is given in section 3.4. Section 3.5 provides a comprehensive simulation study. Additionally, section 3.6 provides data analysis. At the end of the chapter, a summary is given.

3.2 Exponential-Gamma $(\mathbf{3},\theta)$ Distribution

A mixture of exponential (θ) and gamma (3,θ) distributions are considered. It is denoted as EGD(θ). The PDF of the mixture of the exponential (θ) and gamma (3,θ) distribution is as follows:

$$f(x;\theta) = p\, f_1(x;\theta) + (1-p)\, f_2(x;3,\theta),$$

where $p = \frac{\theta}{1+\theta}$, $f_1(x;\theta) = \theta e^{-\theta x}$ and $f_2(x;3,\theta) = \theta^3 \frac{x^2}{2} e^{-\theta x}$.
Then,

$$f(x;\theta) = \frac{\theta^2}{1+\theta}(1 + \frac{\theta}{2}x^2)e^{-\theta x}, x > 0, \theta > 0. \qquad (3.2.1)$$

The CDF corresponding to the EGD(θ) distribution is

$$F(x;\theta) = 1 - \frac{(\theta(x(\theta x + 2) + 2) + 2)e^{-\theta x}}{2(1+\theta)}; x > 0, \theta > 0. \qquad (3.2.2)$$

The Survival function associated with Eq.(3.2.2) is

$$\bar{F}(x;\theta) = 1 - F(x;\theta) = \frac{(\theta(x(\theta x + 2) + 2) + 2)e^{-\theta x}}{2(1+\theta)}; x > 0, \theta > 0. \qquad (3.2.3)$$

The first derivative of the PDF is

$$f'(x) = \frac{\theta^3 e^{-\theta x}}{1+\theta}\left(x - 1 - \frac{\theta x^2}{2}\right).$$

The second derivative of the PDF is

$$f''(x) = \frac{\theta^3 e^{-\theta x}}{1+\theta}\left(1 - 2\theta x + \theta + \frac{\theta^2 x^2}{2}\right).$$

The mode of $f(x)$ is the point $x = x_0$ satisfying $f'(x_0) = 0$. Here $f'(x_0) = 0$ is at the point $x_0 = \frac{1 \pm \sqrt{1 - \frac{\theta}{2}}}{\theta}$, $f''(x) < 0$ for $0 < x < 1$ and $f''(x) > 0$ for $1 \le x \le 2$.

The shape of the PDF is given in figure 3.1 and 3.2.
From the above figures, it is apparent that the PDF can be decreasing or unimodal.
The HRF of EGD(θ) is given below.

$$h(x) = \frac{f(x,\theta)}{\bar{F}(x,\theta)} = \frac{2(1+\theta)\theta^2(1 + \frac{\theta x^2}{2})}{(\theta(x(\theta x + 2) + 2) + 2)}; x > 0, \theta > 0. \qquad (3.2.4)$$

The first derivative of HRF is

$$h'(x) = 2(1+\theta)\theta^2 \frac{\theta x(\theta(x(\theta x + 2) + 2) + 2) - 2\theta(\theta x + 1)(1 + \frac{\theta x^2}{2})}{(\theta(x(\theta x + 2) + 2) + 2)^2}.$$

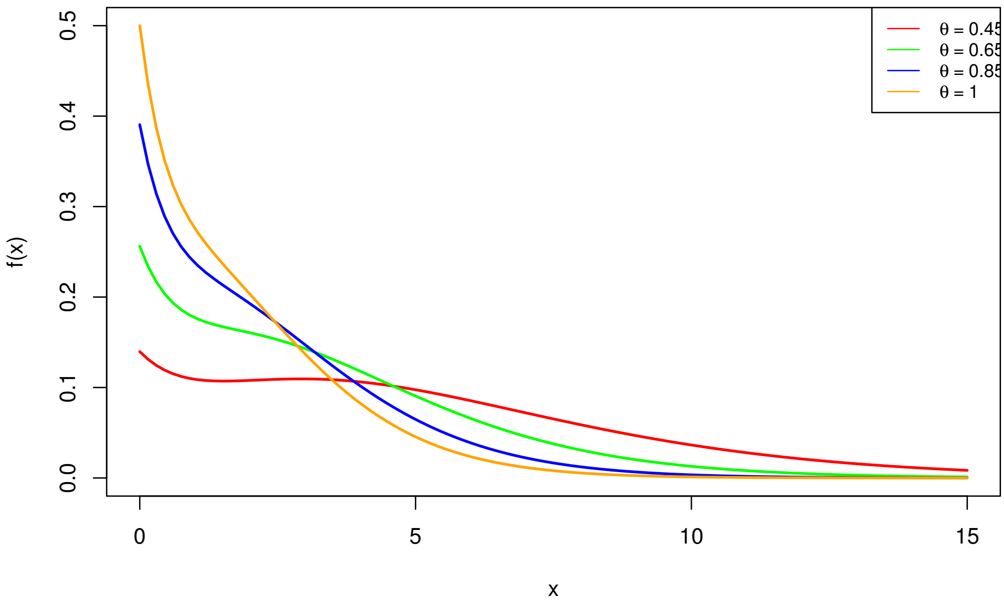

Figure 3.1: PDF plot for $\theta \leq 1$

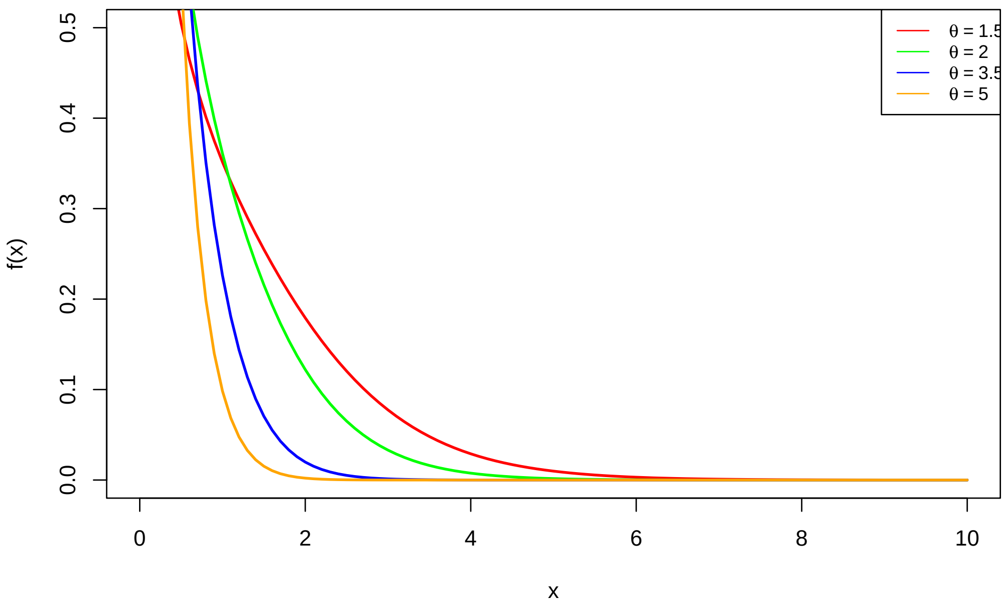

Figure 3.2: PDF plot for $\theta > 1$

The second derivative of HRF is given by

$$h''(x) = \frac{4\theta^3(\theta x + 1)(-\theta^2 x^2 + 6\theta - 2\theta x + 2)(1 + \theta)}{(\theta(x(\theta x + 2) + 2) + 2)^3}.$$

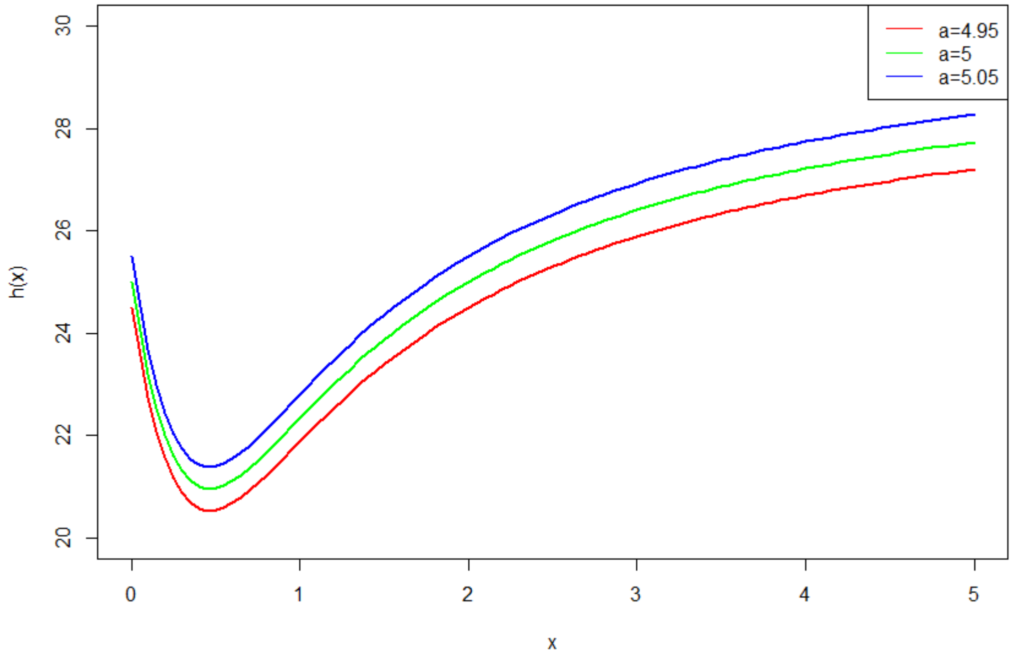

Figure 3.3: HRF plot of $EGD(\theta)$ for θ =4.95, 5, 5.15

The extremum of h(x) is the point $x = x_0$ satisfying $h'(x_0) = 0$, and these points correspond to a maximum or a minimum or a point of inflection according to $h''(x) < 0, h''(x) > 0$ and $h''(x) = 0$ respectively. Here $h'(x) = 0$ at the point $x_0 = \frac{-1+\sqrt{1+2\theta}}{\theta}$ and $h''(x) > 0$ for $\theta > 0$. So h(x) must attain a unique minimum at $x = x_0$.

Initially, the plot of $h(x)$ decreases monotonically and then increases, giving a bathtub shape. Fig.3.3 provides the HRFs of $EGD(\theta)$ for different parameter values.

3.3 Statistical Properties of EGD$(3, \theta)$

Here, the statistical measures for the $EGD(\theta)$ distribution, such as moments, skewness, kurtosis, MGF, CF, quantile function, median, Rènyi entropy, Lorenz curve, and Gini index are discussed.

3.3.1 Moments

In the statistical literature, the concept of moments is of paramount importance. We can measure the central tendency of a population by using moments. Moments also help in measuring the scatteredness, asymmetry, and peakedness of a curve for a particular distribution.

The rth raw moment (about the origin) of $EGD(\theta)$ is

$$\mu'_r = p\frac{r!}{\theta^r} + (1-p)\frac{\Gamma(r+3)}{2\theta^r} = \frac{2\theta r! + \Gamma(r+3)}{2(1+\theta)\theta^r}.$$

Therefore, the mean and variance of $EGD(\theta)$ are respectively given by

$$\mu = \frac{\theta+3}{\theta(1+\theta)},$$

and

$$\sigma^2 = \frac{\theta^2+8\theta+3}{\theta^2(1+\theta)^2}.$$

The skewness and kurtosis can be obtained using these raw moments as

$$Skewness = \frac{2\theta^3+30\theta^2-63\theta+16}{\theta^2+8\theta+3},$$

and

$$Kurtosis = \frac{9\theta^4+192\theta^3+306\theta^2+216\theta+45}{(\theta^2+8\theta+3)^2}.$$

3.3.2 Moment Generating Function and Characteristic Function

Let X has $EGD(\theta)$ distribution, then the MGF of X, $M_X(t) = E(e^{tX})$, is

$$M_X(t) = \frac{\theta^2}{1+\theta}\left(-\frac{(t-\theta)^2+\theta}{(t-\theta)^3}\right),$$

for $t > 0$. Similarly, the CF of X becomes $\phi(t) = M_X(it)$,

$$\phi(t) = \frac{\theta^2}{1+\theta}\left(-\frac{(it-\theta)^2+\theta}{(it-\theta)^3}\right),$$

where $i = \sqrt{-1}$.

3.3.3 Quantile Function and Median

Here, the quantile and median formulas of $EGD(\theta)$ distribution are determined. The quantile x_p of the $EGD(\theta)$ is given from

$$F(x_p) = p, 0 < p < 1$$

The 100 p^{th} percentile can be obtained as,

$$(\theta(x(\theta x + 2) + 2) + 2)e^{-\theta x} = 2(1 - p)(1 + \theta). \tag{3.3.1}$$

Setting $p = 0.5$ in Eq. (3.3.1), the median of $EGD(\theta)$ is obtained as follows.

$$(\theta(x(\theta x + 2) + 2) + 2)e^{-\theta x} = (1 + \theta).$$

The $x_{0.5}$ is the solution of the above monotone increasing function. Using different statistical software, the quantiles or percentiles can be obtained.

3.3.4 Rènyi Entropy

An important entropy measure is Rènyi entropy (Rènyi (1980)). If X has the $EGD(\theta)$ then Rènyi entropy is defined by

$$\Im_R(\nu) = \frac{1}{1-\nu} \log \Big\{ \int f^{\nu}(x)dx \Big\},$$

where $\nu > 0$ and $\nu \neq 1$. Then we can calculate, for $EGD(\theta)$,

$$\int f^{\nu}(x)dx = \int_0^{\infty} \left\{ \frac{\theta^2}{1+\theta} e^{-\theta x\left(1+\frac{\theta x^2}{2}\right)} \right\}^{\nu} dx$$

$$= \left(\frac{\theta^2}{1+\theta}\right)^{\nu} \int_0^{\infty} \left(1 + \frac{\theta x^2}{2}\right)^{\nu} e^{-\nu\theta x}$$

$$= \left(\frac{\theta^2}{1+\theta}\right)^{\nu} \sum_{k=0}^{\infty} \binom{\nu}{k} (-1)^k \int_0^{\infty} x^{2k} e^{-\nu\theta x} dx$$

$$= \left(\frac{\theta^2}{1+\theta}\right)^{\nu} \sum_{k=0}^{\infty} \binom{\nu}{k} (-1)^k \frac{\Gamma(2k+1)}{(\nu\theta)^{2k+1}}.$$

Therefore, Rènyi entropy is given by

$$\Im_R(\nu) = \frac{1}{1-\nu}\log\left\{\left(\frac{\theta^2}{1+\theta}\right)^{\nu}\sum_{k=0}^{\infty}\binom{\nu}{k}(-1)^k\frac{\Gamma(2k+1)}{(\nu\theta)^{2k+1}}\right\}$$

$$= \frac{\nu}{1-\nu}\log\left(\frac{\theta^2}{1+\theta}\right) + \frac{1}{1-\nu}\log\left\{\sum_{k=0}^{\infty}\binom{\nu}{k}(-1)^k\frac{\Gamma(2k+1)}{(\nu\theta)^{2k+1}}\right\}.$$

3.3.5 Lorenz Curve and Gini Index

The Lorenz curve and the Gini index have applications not only in economics but also in reliability.
The Lorenz curve is defined by

$$L(p) = \frac{1}{p}\int_0^q xf(x)dx$$

or equivalently,

$$L(p) = \frac{1}{p}\int_0^q xF^{-1}(x)dx,$$

where $p = E(X)$ and $q = F^{-1}(p)$.
The Gini index is given by

$$G = 1 - 2\int_0^1 L(p)dp.$$

If X has $EGD(\theta)$ then

$$L(p) = \frac{1}{p}\left[\frac{\theta+3}{\theta(\theta+1)} - \frac{(\theta(q(\theta(q(\theta q+3)+2)+6)+2)+6)e^{-\theta q}}{2\theta(1+\theta)}\right].$$

Gini Index is

$$G = 1 - \frac{2}{p\theta(1+\theta)}\left[\theta+3-\frac{(\theta(q(\theta(q(\theta q+3)+2)+6)+2)+6)e^{-\theta q}}{2}\right], \theta > 0.$$

3.3.6 Distribution of Maximum and Minimum

Let $X_1, X_2, \ldots, X_n$ be a simple random sample from $EGD(\theta)$. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics obtained from this sample. The

PDF of $X_{(r)}$ is given by,

$$f_{r:n}(x) = \frac{1}{B(r, n-r+1)}[F(x;\theta)]^{r-1}[1-F(x;\theta)]^{n-r}f(x;\theta),$$

where $F(x;\theta), f(x;\theta)$ are the CDF and PDF given by Eq. (3.2.2) and Eq. (3.2.1), respectively. That is,

$$f_{r:n}(x) = \frac{1}{B(r, n-r+1)}\left[1-\frac{(\theta(x(\theta x+2)+2)+2)e^{-\theta x}}{2(1+\theta)}\right]^{r-1}\left[\frac{(\theta(x(\theta x+2)+2)+2)e^{-\theta x}}{2(1+\theta)}\right]^{n-r}$$

$$\frac{\theta^2}{1+\theta}\left(1+\frac{\theta}{2}x^2\right)e^{-\theta x}. \quad (3.3.2)$$

Then the PDF of the smallest and largest order statistics, $X_{(1)}$ and $X_{(n)}$, respectively, are

$$f_1(x) = \frac{1}{B(1,n)}\left[\frac{(\theta(x(\theta x+2)+2)+2)}{2(1+\theta)}\right]^{n-1}\frac{\theta^2}{1+\theta}\left(1+\frac{\theta}{2}x^2\right)e^{-\theta x}$$

and

$$f_n(x) = \frac{1}{B(n,1)}\left[1-\frac{(\theta(x(\theta x+2)+2)+2)}{2(1+\theta)}\right]^{n-1}\frac{\theta^2}{1+\theta}\left(1+\frac{\theta}{2}x^2\right)e^{-\theta x}.$$

The CDF of $X_{(r)}$ is

$$F_{r:n}(x) = \sum_{j=r}^{n}\binom{n}{j}\left[1-\frac{(\theta(x(\theta x+2)+2)+2)}{2(1+\theta)}\right]^{j}\left[\frac{(\theta(x(\theta x+2)+2)+2)}{2(1+\theta)}\right]^{n-j}. \quad (3.3.3)$$

Then the CDF of the smallest and largest order statistics $X_{(1)}$ and $X_{(n)}$, respectively, are

$$F_1(x) = 1-\left[\frac{(\theta(x(\theta x+2)+2)+2)}{2(1+\theta)}\right]^{n}, \theta > 0$$

and

$$F_n(x) = \left[1 - \frac{(\theta(x(\theta x + 2) + 2) + 2)}{2(1+\theta)}\right]^n, \theta > 0.$$

These distributions can be used in reliability operations.

3.4 Parametric Estimation

In this section, point estimation of the unknown parameter of the $EGD(\theta)$ is described by using the method of maximum likelihood for complete sample data, as given below.

3.4.1 Maximum Likelihood Estimation

The likelihood function of the $EGD(\theta)$ distribution is

$$L = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \frac{\theta^2(1 + \frac{\theta}{2}x_i^2)e^{-\theta x}}{1+\theta}$$

The log-likelihood function is,

$$\log L(x_i; \theta) = 2n \log \theta - n \log(1+\theta) + \sum_{i=1}^{n} \left[\log\left(1 + \frac{\theta x_i^2}{2}\right) - \theta x_i\right].$$

The first partial derivatives of the log-likelihood function with respect to θ is

$$\frac{\partial L}{\partial \theta} = \frac{2n}{\theta} - \frac{n}{1+\theta} + \sum_{i=1}^{n} \left(\frac{x_i^2}{2(1 + \frac{\theta x_i^2}{2})} - x_i\right) \tag{3.4.1}$$

Setting the left side of the above equation to zero, the likelihood equation as a system of nonlinear equations in θ is obtained. Solving this system in θ gives the MLE of θ. It is easy to obtain numerically by using a statistical software package like the *nlm* package in R programming with arbitrary initial values.

The Fisher information about $\theta, I(\theta)$, is

$$I(\theta) = E\left\{ - \frac{\partial^2}{\partial \theta^2} \log f(X; \theta)\right\} = E\left(\frac{2}{\theta^2} - \frac{1}{(1+\theta)^2} + \frac{x^4}{4}\frac{1}{(1 + \frac{\theta x^2}{2})^2}\right)$$

$$= \frac{2}{\theta^2} - \frac{1}{(1+\theta)^2} + E\left\{\frac{x^4}{4}\frac{1}{(1+\frac{\theta x^2}{2})^2}\right\}.$$

Then the asymptotic $100(1-\alpha)\%$ confidence interval for θ is given by

$$\hat{\theta} \pm Z_{\alpha/2}\frac{I^{-1/2}(\hat{\theta})}{\sqrt{n}}.$$

3.5 Simulation Study

A simulation study is conducted to illustrate the performance of the accuracy of the estimation method. The following scheme is used:

1. Specify the value of the parameter θ.
2. Specify the sample size n.
3. Generate a random sample with size n from $EGD(\theta)$.
4. Using the estimation method used in this chapter, calculate the point estimate of the parameter θ.
5. Repeat steps 3-4, N=1000 times.
6. Calculate the bias and the MSE.

3.6 Applications

Data analysis is provided to see how the new model works. The data set is taken from Klein and Berger (1997). It shows survival data on the death times of 26 psychiatric inpatients admitted to the University of Iowa hospital during the years 1935-1948.

Different distributions were used, such as ED, EED, and $EGD(\theta)$, to analyze the data. The estimate(s) of the unknown parameter(s), corresponding KS test statistic, and Log L values for three different models are given in table 3.3. The AIC (see Akaike(1974)), BIC, and CAIC are presented in the following table 3.4.

Table 3.3 shows the parameter MLEs, KS test statistic value with p-value, and log-likelihood values of the fitted distributions, and table 3.4 shows the values of AIC, BIC, and CAIC. The values in tables 3.3 and 3.4 indicate that the $EGD(\theta)$

Table 3.1: Simulation study for $\theta = 1, 1.5, 1.85$.

θ	n	Bias	MSE
1	50	-0.0009	3.6485×10^{-05}
	100	0.0004	1.5573×10^{-05}
	500	1.311×10^{-05}	8.599×10^{-08}
	1000	3.6889×10^{-05}	1.491×10^{-09}
1.5	50	-0.0007	2.637×10^{-05}
	100	-0.0006	3.393×10^{-05}
	500	-3.906×10^{-06}	7.628×10^{-09}
	1000	-3.823×10^{-05}	1.462×10^{-06}
1.85	50	0.0017	0.0002
	100	0.0009	8.593×10^{-05}
	500	0.0002	1.410×10^{-05}
	1000	3.296×10^{-05}	1.086×10^{-06}

Table 3.2: The survival data on the death times of Psychiatric inpatients

1	1	2	22	30	28	32	11	14	36	31	33	33
37	35	25	31	22	26	24	35	34	30	35	40	39

distribution is a strong competitor to other distributions used here for fitting the dataset.

P-P plot for ED, EED and $EGD(\theta)$ are given in Figure 3.4 which shows that $EGD(\theta)$ model is more plausible than ED and EED models.

Table 3.3: The estimates, K-S test statistic and log-likelihood for the dataset

Model	**Estimates**	**KS**	**Log L**	**p value**
ED	$\hat{\theta}$ =0.0378	0.3728	-111.1302	0.0015
EED	$\hat{a}$ =1.79724674,$\hat{b}$ =0.0525	0.3146	-108.9871	0.0116
EGD	$\hat{\theta}$ =0.1050	0.2613	-104.5856	0.0574

Table 3.4: AIC, BIC, and CAIC of the models based on the dataset

Model	**AIC**	**BIC**	**CAIC**
ED	224.2604	225.5185	226.5185
EED	221.9741	224.4903	226.4903
EGD	211.1713	212.4294	213.4294

3.7 Summary

A bathtub-shaped failure rate model, Exponential-Gamma$(3, \theta)$ distribution, is discussed, and its properties are studied. Moments, skewness, kurtosis, MGF, CF, Rènyi entropy, Lorenz curve, Gini index, and the distribution of maximum and minimum order statistics are obtained. A simulation study is conducted to illustrate the accuracy of the estimation method that has been obtained using maximum likelihood estimators. The application of $EGD(\theta)$ to real data shows that the new distribution is effective in providing a better fit than the exponential and exponentiated exponential distributions.

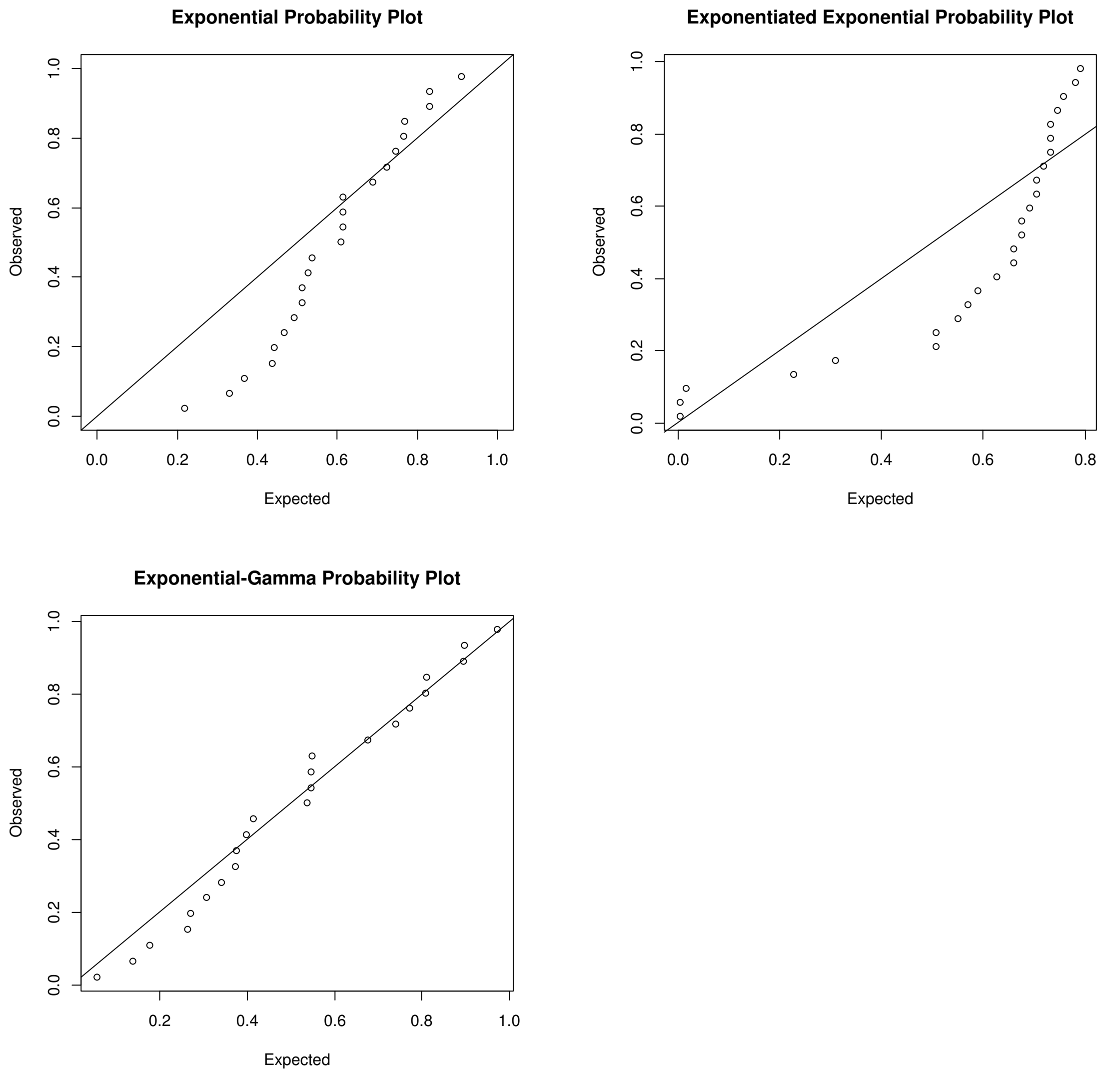

Figure 3.4: P-P Plots