On Generalized Gamma Distribution and Its Application to Survival Data

Kiche J[1], Oscar Ngesa[2] & George Orwa[3]

[1] Mathematics Department, Pan African University, Institute of Basic Sciences, Technology, and Innovation, Kenya

[2] Mathematics and Informatics Department, Taita-Taveta University, Kenya

[3] Statistics and Actuarial Sciences Department, Jomo Kenyatta University of Agriculture and Technology, Kenya

Correspondence: Kiche J, Mathematics Department, Pan African University, Institute of Basic Sciences, Technology, and Innovation, Kenya. E-mail: kicheopiyo07@gmail.com

Abstract

The generalized gamma distribution is a continuous probability distribution with three parameters. It is a generalization of the two-parameter gamma distribution. Since many distributions commonly used for parametric models in survival analysis (such as the Exponential distribution , the Weibull distribution and the Gamma distribution) are special cases of the generalized gamma, it is sometimes used to determine which parametric model is appropriate for a given set of data. Generalized gamma distribution is one of the distributions used in frailty modeling. In this study , it is shown that generalized gamma distribution has three sub-families and its application to the analysis of a survival data has also been explored. The parametric modeling approach has been carried out to find the expected results.

Keywords: generalized gamma, sub-families, Kullback-Leibler divergence

1. Introduction

The early generalization of gamma distribution can be traced back to Amoroso(1925)who discussed a generalized gamma distribution and applied it to fit income rates. Johnson et al. gave a four parameter generalized gamma distribution which reduces to the generalized gamma distribution defined by Stacy (1962) when the location parameter is set to zero.The generalized gamma defined by Stacy (1962) is a three-parameter exponentiated gamma distribution. Mudholkar and Srivastava (1993) introduced the exponentiated method to derive a distribution. Agarwal and Al-Saleh (2001) applied generalized gamma to study hazard rates. Balakrishnan and Peng (2006) applied this distribution to develop generalized gamma frailty model.Nadarajah and Gupta (2007) proposed another type of generalized gamma distribution with application to fitting drought data. Cordeiro et al. (2012) derived another generalization of Stacy's generalized gamma distribution using exponentiated method, and applied it to life time and survival analysis.

The generalized gamma distribution presents a flexible family in the varieties of shapes and hazard functions for modeling duration. Distributions that are used in duration analysis in economics include exponential, log-normal, gamma, and Weibull . The generalized gamma family, which encompasses exponential, gamma, and Weibull as subfamilies, and log-normal as a limiting distribution, has been used in economics by Jaggia, Yamaguchi, and Allenby et al . Some authors have argued that the flexibility of generalized gamma makes it suitable for duration analysis, while others have advocated use of simpler models because of estimation difficulties caused by the complexity of generalized gamma parameter structure. Obviously, there would be no need to endure the costs associated with the application of a complex generalized gamma model if the data do not discriminate between the generalized gamma and members of its subfamilies, or if the fit of a simpler model to the data is as good as that for the complex generalized gamma. Hager and Bain inhibited applications of the generalized gamma model.

Prentice resolved the convergence problem using a nonlinear transformation of generalized gamma model. Maximum-likelihood estimation of the parameters and quasi maximum likelihood estimators for its subfamily (two-parameter gamma distribution) can be found. Hwang, T. et al introduced a new moment estimation of parameters of the generalized gamma distribution using it's characterization. In information theory, thus far a maximum entropy (ME) derivation of generalized gamma is found in Kapur, where it is referred to as generalized Weibull distribution, and the entropy of GG has appeared in the context of flexible families of distributions. Some concepts of this family in information theory has introduced by

Dadpay et al .

2. Parametric Distributions and Their Properties

2.1 Log-Normal Distribution

A log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. If the random variable X is log-normally distributed, then $Y = ln(X)$ has a normal distribution. Likewise, if Y has a normal distribution, then the exponential function of Y, $X = exp(Y)$, has a log-normal distribution. A random variable which is log-normally distributed takes only positive real values. A log-normal process is the statistical realization of the multiplicative product of many independent random variables, each of which is positive. This is justified by considering the central limit theorem in the log domain. The log-normal distribution is the maximum entropy probability distribution for a random variate X for which the mean and variance of $ln(X)$ are specified.

A positive random variable X is log-normally distributed if the logarithm of X is normally distributed, $ln(X) \sim N(\mu, \sigma^2)$ The probability density function of the log-normal distribution is given by:

$$f_X(x) = \frac{1}{x} \frac{1}{\sigma \sqrt{2\pi}} exp(-\frac{(lnx-\mu)^2}{2\sigma^2}) \tag{1}$$

The cumulative distribution is given as

$$F_X(x) = \Phi(\frac{(lnx) - \mu}{\sigma}) \tag{2}$$

which may also be expressed as :

$$\frac{1}{2}[1 + erf(\frac{lnx - \mu}{\sigma \sqrt{2}})] \tag{3}$$

where erf is the error function.

Approximating formular for the characteristic function $\varphi(t)$ can be given as

$$\varphi(t) \approx \frac{exp(\frac{-\omega^2(-it\sigma^2 e^\mu) + 2\omega(-it\sigma^2 e^\mu)}{2\sigma^2})}{\sqrt{1 + \omega(-it\sigma^2 e^\mu)}} \tag{4}$$

where ω is the Lambert W function.

For a log-normal random variable, the partial expectation is given by

$$g(k) = e^{\mu + \frac{1}{2}\sigma^2} \Phi(\frac{\mu + \sigma^2 - lnk}{\sigma}) \tag{5}$$

The quantile for the log-normal distribution is given by

$$exp(\mu + \sqrt{2\sigma^2} erf^{-1}(2F - 1)) \tag{6}$$

while the variance is given as

$$[exp(\sigma^2) - 1]exp(2\mu + \sigma^2) \tag{7}$$

The skewness is given by the formular

$$(e^{\sigma^2} + 2) \sqrt{e^{\sigma^2} - 1} \tag{8}$$

The Ex.kurtosis is then denoted by

$$exp(4\sigma^2) + 2exp(3\sigma^2) + 3exp(2\sigma^2) - 6 \tag{9}$$

The Support is over $x \in (0, +\infty)$ while the mean is $exp(\mu + \frac{\sigma^2}{2})$, the mode is $exp(\mu - \sigma^2)$ and the Entropy is given as $log_2(\sigma e^{\mu + \frac{1}{2}} \sqrt{2\pi})$

2.2 Weibull Distribution

The Weibull distribution is a continuous probability distribution. It is named after Swedish mathematician Waloddi Weibull, who described it in detail in 1951, although it was first identified by Frechet (1927) and first applied by Rosin and Rammler (1933) to describe a particle size distribution.

The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering. It is a versatile distribution that can take on the characteristics of other types of distributions, based on the value of the shape parameter. The probability density function of a Weibull random variable is

$$f(x, \xi, \nu) = \frac{\nu}{\xi} (\frac{x}{\xi})^{\nu-1} e^{-(x/\xi)^{\nu}} \quad , x \geq 0 \tag{10}$$

where $\nu > 0$ is the shape parameter and $\xi > 0$ is the scale parameter of the distribution.

The cumulative distribution function for the Weibull distribution is

$$F(x, \xi, \nu) = 1 - e^{-(x/\xi)^{\nu}} \tag{11}$$

The failure rate h (or hazard function) is given by

$$h(x, \xi, \nu) = \frac{\nu}{\xi} (\frac{x}{\xi})^{\nu-1} \tag{12}$$

The quantile (inverse cumulative distribution) function for the Weibull distribution is

$$Q(p; \xi, \nu) = \xi(-ln(1-p))^{1/\nu} \tag{13}$$

The moment generating function of the logarithm of a Weibull distributed random variable is given by

$$E[e^{t \log X}] = \xi^{t} \Gamma(\frac{t}{\nu} + 1) \tag{14}$$

The mean and variance of a Weibull random variable can be expressed respectively as

$$E(X) = \xi \Gamma(1 + \frac{1}{\nu}) \tag{15}$$

and

$$Var(X) = \xi^2 [\Gamma(1 + \frac{2}{\nu}) - (\Gamma(1 + \frac{1}{\nu}))^2] \tag{16}$$

The skewness is given by

$$\gamma_1 = \frac{\Gamma(1 + \frac{3}{\nu})\xi^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3} \tag{17}$$

where the mean is denoted by μ and the standard deviation is denoted by σ. The excess kurtosis is given by

$$\gamma_2 = \frac{-6\Gamma_1^4 + 12\Gamma_1^2\Gamma_2 - 3\Gamma_2^2 - 4\Gamma_1\Gamma_3 + \Gamma_4}{[\Gamma_2 - \Gamma_1^2]^2} \tag{18}$$

The information entropy is given by

$$H(\xi, \nu) = \gamma(1 - \frac{1}{\nu}) + ln(\frac{\xi}{\nu}) + 1 \tag{19}$$

where γ is the EulerMascheroni constant.

The Kullback-Leibler divergence for Weibull distribution is given by

$$D_{KL}(Weib_1 \parallel Weib_2) = log\frac{\nu_1}{\xi_1^{\nu_1}} - log\frac{\nu_2}{\xi_2^{\nu_2}} + (\nu_1 - \nu_2)[log\xi_1 - \frac{\gamma}{\nu_1}] + (\frac{\xi_1}{\xi_2})^{\nu_2}\Gamma(\frac{\nu_2}{\nu_1} + 1) - 1 \tag{20}$$

The variances and covariances of $\hat{\nu}$ and $\hat{\xi}$ are estimated from the inverse local Fisher matrix as

$$\begin{pmatrix} \hat{Var}(\hat{\nu}) & \hat{Cov}(\hat{\nu}, \hat{\xi}) \\ \hat{Cov}(\hat{\nu}, \hat{\xi}) & \hat{Var}(\hat{\xi}) \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 \Lambda}{\partial \nu^2} & -\frac{\partial^2 \Lambda}{\partial \nu \partial \xi} \\ -\frac{\partial^2 \Lambda}{\partial \nu \partial \xi} & -\frac{\partial^2 \Lambda}{\partial \xi^2} \end{pmatrix}^{-1}_{\nu=\hat{\nu}, \xi=\hat{\xi}} \tag{21}$$

2.3 Exponential Distribution

The exponential model, with only one unknown parameter, is the simplest of all life distribution models. The exponential distribution is one of the widely used continuous distributions. It is often used to model the time elapsed between events. That is the exponential distribution is often concerned with the amount of time until some specific event occurs. For example, the amount of time (beginning now) until an earthquake occurs has an exponential distribution. Other examples include the length, in minutes, of long distance business telephone calls, and the amount of time, in months, a car battery lasts. An interesting property of the exponential distribution is that it can be viewed as a continuous analogue of the geometric distribution.

The probability density function of exponential distribution is given by

$$f(t, \theta) = \theta e^{-\theta t}, \quad t \geq 0 \tag{22}$$

The cumulative distribution function is thus denoted by

$$F(t) = 1 - e^{-\theta t} \tag{23}$$

The reliability and failure rate of the exponential distribution are given respectively as

$R(t) = e^{-\theta t}$ and $h(t) = \theta$

while the mean and variance are given by *mean* $= 1/\theta$ and *variance* $= 1/\theta^2$.
Note that the failure rate reduces to the constant θ for any time. The exponential distribution is the only distribution to have a constant failure rate. Also, another name for the exponential mean is the Mean Time To Failure or MTTF and we have MTTF $= 1/\theta$.

Kullback-Leibler divergence is given by

$$D_{KL}(exp_0 \parallel exp_1) = log(\theta_0) - log(\theta) + \frac{\theta}{\theta_0} - 1 \tag{24}$$

The most important property is that the exponential distribution is memoryless.Specifically, the memoryless property says that an exponentially distributed random variable T obeys the relation:

$$Pr(T > a + t/T > a) = Pr(T > t), \quad \forall a, t \geq 0 \tag{25}$$

Proof

$$Pr(T > a + t/T > a) = \frac{Pr(T > a + t \cap T > a)}{Pr(T > a)} \tag{26}$$

$$= \frac{Pr(T > a + t)}{Pr(T > a)} \tag{27}$$

$$= \frac{e^{-\theta(a+t)}}{e^{-\theta a}} \tag{28}$$

$$= e^{-\theta t} \tag{29}$$

$$= Pr(T > t) \tag{30}$$

hence the proof.

The memoryless property says that knowledge of what has occurred in the past has no effect on future probabilities.

2.4 Gamma Distribution

The gamma distribution is another widely used distribution. Its importance is largely due to its relation to exponential and normal distributions. The gamma distribution is a two-parameter family of continuous probability distributions. The exponential distribution, Erlang distribution, and chi-squared distribution are special cases of the gamma distribution.

The gamma distribution is used to model errors in multi-level Poisson regression models, because the combination of the Poisson distribution and a gamma distribution is a negative binomial distribution. The gamma distribution is widely used as a conjugate prior in Bayesian statistics. It is the conjugate prior for the precision (i.e. inverse of the variance) of a normal distribution. It is also the conjugate prior for the exponential distribution. The gamma distribution can be

parameterized in terms of a shape parameter $\varphi = \kappa$ and an inverse scale parameter $\beta = 1/\theta$, called a rate parameter. A random variable X that is gamma-distributed with shape φ and rate β is denoted by

$X \sim \Gamma(\varphi, \beta) \equiv \text{Gamma}(\varphi, \beta)$

The corresponding probability density function in the shape-rate parametrization is

$$f(x; \varphi, \beta) = \frac{\beta^{\varphi} x^{\varphi-1} e^{-\beta x}}{\Gamma(\varphi)} \quad \text{for } x > 0 \quad \varphi, \beta > 0,$$

$$f(x; \varphi, \theta) = \frac{x^{\varphi-1} e^{-x/\theta}}{\theta^{\varphi} \Gamma(\varphi)} \quad \text{for } x > 0, \quad \varphi, \theta > 0.$$

(31)

where $\Gamma(\varphi)$ is the gamma function.

Both parameterizations are common because either can be more convenient depending on the situation.

The cumulative distribution function is the regularized gamma function:

$$F(x; \varphi, \beta) = \int_0^x f(u; \varphi, \beta)\, du = \frac{\gamma(\varphi, \beta x)}{\Gamma(\varphi)}$$

(32)

where $\gamma(\varphi, \beta x)$ is the lower incomplete gamma function.

If φ is a positive integer (i.e., the distribution is an Erlang distribution), the cumulative distribution function has the following series expansion:

$$F(x; \varphi, \beta) = 1 - \sum_{i=0}^{\varphi-1} \frac{(\beta x)^i}{i!} e^{-\beta x} = e^{-\beta x} \sum_{i=\varphi}^{\infty} \frac{(\beta x)^i}{i!}.$$

(33)

The KullbackLeibler divergence (KL-divergence), of $\Gamma(\varphi_p, \beta_p)$ ("true" distribution) from $\Gamma(\varphi_q, \beta_q)$ ("approximating" distribution) is given by

$$D_{\text{KL}}(\varphi_p, \beta_p; \varphi_q, \beta_q) = (\varphi_p - \varphi_q)\psi(\varphi_p) - \log\Gamma(\varphi_p) + \log\Gamma(\varphi_q)$$
$$+ \varphi_q(\log\beta_p - \log\beta_q) + \varphi_p \frac{\beta_q - \beta_p}{\beta_p}.$$

(34)

Written using the $k,$ parameterization, the KL-divergence of $\Gamma(k_p, \theta_p)$ from $\Gamma(k_q, \theta_q)$ is given by

$$D_{\text{KL}}(k_p, \theta_p; k_q, \theta_q) = (k_p - k_q)\psi(k_p) - \log\Gamma(k_p) + \log\Gamma(k_q)$$
$$+ k_q(\log\theta_q - \log\theta_p) + k_p \frac{\theta_p - \theta_q}{\theta_q}.$$

(35)

The Laplace transform of the gamma probability density function is

$$F(s) = (1 + \theta s)^{-k} = \frac{\beta^{\varphi}}{(s + \beta)^{\varphi}}.$$

(36)

Parameterization 1: If $X_k \sim \Gamma(\varphi_k, \theta_k)$ are independent, then, $\frac{\varphi_2 \theta_2 X_1}{\varphi_1 \theta_1 X_2} \sim F(2\varphi_1, 2\varphi_2)$, or equivalently, $\frac{X_1}{X_2} \sim \beta'(\varphi_1, \varphi_2, 1, \frac{\theta_1}{\theta_2})$.

Parameterization 2: If $X_k \sim \Gamma(\varphi_k, \beta_k)$ are independent, then $\frac{\varphi_2 \beta_1 X_1}{\varphi_1 \beta_2 X_2} \sim F(2\varphi_1, 2\varphi_2)$, or equivalently, $\frac{X_1}{X_2} \sim \beta'(\varphi_1, \varphi_2, 1, \frac{\beta_2}{\beta_1})$.

If $X \sim \Gamma(\varphi, \theta)$ and $Y \sim \Gamma(\beta, \theta)$ are independently distributed, then $X/(X + Y)$ has a beta distribution with parameters φ and β, and $X/(X + Y)$ is independent of $X + Y$, which is $\Gamma(\varphi + \beta, \theta)$-distributed.

3. Methods

3.1 Generalized Gamma Distribution

The generalized gamma has three parameters given in this case as : $\kappa > 0$, $\phi > 0$, and $\beta > 0$.
For non-negative x, the probability density function of the generalized gamma is:

$$f(x; \kappa, \phi, \beta) = \frac{(\beta/\kappa^{\phi}) x^{\phi-1} e^{-(x/\kappa)^{\beta}}}{\Gamma(\phi/\beta)}$$

(37)

where $\Gamma(.)$ denotes the gamma function.

3.2 The Cumulative Distribution Function

The cumulative distribution function (CDF) of a random variable X denoted by $F(x)$, is defined as $F(x) = Pr(X \leq x)$. Using identity for the probability of disjoint events, if X is a discrete random variable, then

$$F(x) = \sum_{k=1}^{n} Pr(X = x_k) \tag{38}$$

where x_n gives the largest possible value of X that is less than or equal to x.

The cumulative distribution function for a random variable at x gives the probability that the random variable X is less than or equal to that number x. Note that in the formula for CDFs of discrete random variables, we always have $n \leq N$, where N is the number of possible outcomes of X.

This function, CDF(x), simply tells us the odds of measuring any value up to and including x.

The cumulative distribution function ("c.d.f.") of a continuous random variable X is defined as:

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad for \quad -\infty < x < \infty. \tag{39}$$

3.3 Properties of the CDF

- $0 \leq F(x) \leq 1, \forall x$

- $lim_{x \to -\infty} F(x) = 0$

- $lim_{x \to \infty} F(x) = 1$

- $F(x)$ is a non-decreasing function of x.

  A function $f(x)$ is said to be nondecreasing if $f(x_1) \leq f(x_2)$ whenever $x_1 < x_2$.

The cumulative distribution function of generalized gamma distribution is given by:

$$F(x; \kappa, \phi, \beta) = \frac{\gamma(\phi/\beta, (x/\kappa)^\beta)}{\Gamma(\phi/\beta)} \tag{40}$$

where $\gamma(.)$ denotes the lower incomplete gamma function.

The GG is a three-parameter $(\beta, \sigma > \theta, k)$ family whose survival function is given as

$$S_{GG}(t) = 1 - \Gamma[k^{-2}exp(k[log(t) - \beta]/\sigma); k^{-2}]. \tag{41}$$

$$S_{GG}(t) = 1 - \Gamma[k^{-2}(e^{-\beta}t)^{k/\sigma}; k^{-2}]. \tag{42}$$

3.4 Kullback-Leibler Divergence

The Kullback-Leibler divergence (also called relative entropy) is a measure of how one probability distribution diverges from a second, expected probability distribution. Applications include characterizing the relative (Shannon) entropy in information systems, randomness in continuous time-series, and information gain when comparing statistical models of inference. In contrast to variation of information, it is a distribution-wise asymmetric measure and thus does not qualify as a statistical metric of spread.

In the simple case, a Kullback-Leibler divergence of 0 indicates that we can expect similar, if not the same, behavior of two different distributions, while a Kullback-Leibler divergence of 1 indicates that the two distributions behave in such a different manner that the expectation given the first distribution approaches zero.

If f_1 and f_2 are the probability density functions of two generalized gamma distributions, then their Kullback-Leibler divergence is given by;

$$D_{KL}(f_1 \| f_2) = \int_0^{\infty} f_1(x; \kappa_1, \phi_1, \beta_1) ln \frac{f_1(x; \kappa_1, \phi_1, \beta_1)}{f_2(x; \kappa_2, \phi_2, \beta_2)} dx \tag{43}$$

$$= ln \frac{\beta_1 \kappa_2^{\phi_2} \Gamma(\phi_2/\beta_2)}{\beta_2 \kappa_1^{\phi_1} \Gamma(\phi_1/\beta_1)} + [\frac{\psi(\phi_1/\beta_1)}{\beta_1} + ln\kappa_1](\phi_1 - \phi_2) + \frac{\Gamma((\phi_1 + \phi_2)/\beta_1)}{\Gamma(\phi_1/\beta_1)}(\frac{\kappa_1}{\kappa_2})^{\beta_2} - \frac{\phi_1}{\beta_1} \tag{44}$$

3.5 Moments of Generalized Gamma Distribution

If X has a generalized gamma distribution as above, then

$$E(x)^r = \kappa^r \frac{\Gamma((\phi + r)/\beta)}{\Gamma(\phi/\beta)} \tag{45}$$

The mean of a generalized gamma distribution is given by

$$\kappa \frac{\Gamma((\phi + 1)/\beta)}{\Gamma(\phi/\beta)} \tag{46}$$

The mode of a generalized gamma distribution is given by

$$\kappa(\frac{\phi - 1}{\beta})^{1/\beta} for \phi > 1, otherwise, 0 \tag{47}$$

The variance of generalized gamma distribution is given by

$$\kappa^2(\frac{\Gamma((\phi + 2)/\beta)}{\Gamma(\phi/\beta)} - (\frac{\Gamma((\phi + 1)/\beta)}{\Gamma(\phi/\beta)})^2) \tag{48}$$

While the entropy is given by

$$ln\frac{\kappa\Gamma(\phi/\beta)}{\beta} + \frac{\phi}{\beta} + (\frac{1}{\beta} - \frac{\phi}{\beta})\psi(\frac{\phi}{\beta}) \tag{49}$$

Skewness is given by

$$\frac{\Gamma(\phi + \frac{3}{\beta})\Gamma(\phi)^2 - 3\Gamma(\phi + \frac{2}{k})\Gamma(\phi + \frac{1}{\beta})\Gamma(\phi) + 2\Gamma(\phi + \frac{1}{\beta})^3}{(\Gamma(\phi + \frac{2}{\beta})\Gamma(\phi) - \Gamma(\phi + \frac{1}{\beta})^2)^{3/2}} \tag{50}$$

3.6 Properties of Generalized Gamma Distribution

The generalized gamma family is flexible in that it includes several well-known models as subfamilies. The subfamilies of generalized gamma thus far considered in the literature are exponential ($\phi = \beta = 1$), gamma for $\beta = 1$, and Weibull for $\phi = 1$. The log-normal distribution is also obtained as a limiting distribution when $\phi \longrightarrow \infty$. An important property of generalized gamma family for information analysis is that the family is closed under power transformation.

That is, if

$X \sim GG(\alpha, \tau, \lambda)$ then

$$Y = X^s \sim GG(\alpha, \frac{\tau}{s}, \lambda^s), s > 0 \tag{51}$$

It also has the property that

$Z = \eta X$ has $GG(\eta\alpha, \tau, \lambda)$ distribution.

We use the transformed probability density function of the generalized gamma distribution to prove that generalized gamma is flexible with the subfamilies exponential, Weibull and Gamma distributions.

That is we let the pdf of the generalized gamma distribution ($GG(\alpha, \tau, \lambda)$) to be given by

$$f(\alpha, \tau, \lambda) = \frac{\tau}{\lambda\Gamma(\alpha)}(\frac{x}{\lambda})^{\alpha\tau-1}e^{-(x/\lambda)^\tau} \tag{52}$$

then if $\alpha = \tau = 1$ the subfamily is the exponential distribution

Proof I

From the probability density function of generalized gamma distribution given by

$$f(\alpha, \tau, \lambda) = \frac{\tau}{\lambda\Gamma(\alpha)}\left(\frac{x}{\lambda}\right)^{\alpha\tau-1}e^{-(x/\lambda)^\tau} \tag{53}$$

if we replace $\alpha = \tau = 1$, then we have that

$$= \frac{1}{\lambda\Gamma(1)}\left(\frac{x}{\lambda}\right)^{1-1}e^{-(x/\lambda)^1} \tag{54}$$

$$= \frac{1}{\lambda}e^{-x/\lambda} \tag{55}$$

which is probability density form of an exponential distribution with parameter λ hence the proof.

Proof II

Similarly, if we now let $\tau = 1$ then, from the probability density function of generalized gamma distribution given by:

$$f(\alpha, \tau, \lambda) = \frac{\tau}{\lambda\Gamma(\alpha)}\left(\frac{x}{\lambda}\right)^{\alpha\tau-1}e^{-(x/\lambda)^\tau} \tag{56}$$

we have that

$$= \frac{1}{\lambda\tau(\alpha)}\left(\frac{x}{\lambda}\right)^{\alpha-1}e^{-(x/\lambda)} \tag{57}$$

simplifying gives;

$$= \frac{1}{\lambda\Gamma(\alpha)}x^{\alpha-1}\left(\frac{1}{\lambda}\right)^\alpha\left(\frac{1}{\lambda}\right)^{-1}e^{-(x/\lambda)} \tag{58}$$

$$= \frac{\left(\frac{1}{\lambda}\right)^\alpha x^{\alpha-1}e^{-(x/\lambda)}}{\Gamma(\alpha)} \tag{59}$$

which is a probability density function of a gamma distribution with the parameters α and λ

Proof III

In the last proof ,if we now let $\alpha = 1$

then from the probability density function of generalized gamma distribution given by

$$f(\alpha, \tau, \lambda) = \frac{\tau}{\lambda\Gamma(\alpha)}\left(\frac{x}{\lambda}\right)^{\alpha\tau-1}e^{-(x/\lambda)^\tau} \tag{60}$$

we show that the generalized gamma distribution reduces to Weibull distribution as a sub-family.

That is we have

$$= \frac{\tau}{\lambda\Gamma(1)}\left(\frac{x}{\lambda}\right)^{\tau-1}e^{-(x/\lambda)^\tau} \tag{61}$$

This reduces to

$$= \frac{\tau}{\lambda}x^{\tau-1}\left(\frac{1}{\lambda}\right)^\tau\left(\frac{1}{\lambda}\right)^{-1}e^{-(x/\lambda)^\tau} \tag{62}$$

$$= \tau\left(\frac{1}{\lambda}\right)^\tau x^{\tau-1}e^{-(x/\lambda)^\tau} \tag{63}$$

$$= \frac{\tau}{\lambda}\left(\frac{x}{\lambda}\right)^{\tau-1}e^{-(x/\lambda)^\tau} \tag{64}$$

which is a probability density function of a Weibull distribution with the parameters λ and τ hence the proof.

3.7 Parameter Estimation Using Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a method of estimating the parameters of a distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. The point in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate. The logic of maximum likelihood is both intuitive and flexible, and as such the method has become a dominant means of statistical inference. If the likelihood function is differentiable, the derivative test for determining maxima can be applied. In some cases, the first-order conditions of the likelihood function can be solved explicitly; for instance, the ordinary least squares estimator maximizes the likelihood of the linear regression model. Under most circumstances, however, numerical methods will be necessary to find the maximum of the likelihood function.

From a statistical standpoint, the observations $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ are a random sample from an unknown population. The goal is to make inferences about the population that is most likely to have generated the sample, specifically the probability distribution corresponding to the population. Associated with each probability distribution is a unique vector $\theta = [\theta_1, \theta_2, \ldots, \theta_k]^\mathsf{T}$ of parameters that index the probability distribution within a parametric family $\{f(\cdot\,;\theta) \mid \theta \in \Theta\}$. As θ changes in value, different probability distributions are generated. The idea of maximum likelihood is to re-express the joint probability of the sample data $f(y_1, y_2, \ldots, y_n; \theta)$ as a likelihood function $\mathcal{L}(\theta\,;\mathbf{y})$ that treats θ as a variable. For independent and identically distributed random variables, the likelihood function is defined as

$$L(\theta\,;\mathbf{y}) = f(y_1;\theta)f(y_2;\theta)\cdots f(y_n;\theta) = \prod_{i=1}^{n} f(y_i;\theta) \tag{65}$$

and evaluated at the observed data sample. The goal is then to find the values of the model parameter that maximize the likelihood function over the parameter space Θ. Intuitively, this selects the parameter values that make the observed data most probable. The problem is thus to find the supremum value of the likelihood function by choice of the parameter

$\mathcal{L}(\widehat{\theta}\,;\mathbf{y}) = \sup_{\theta \in \Theta} L(\theta\,;\mathbf{y})$ where the estimator $\widehat{\theta} = \widehat{\theta}(\mathbf{y})$ is function of the sample. A sufficient but not necessary condition for its existence is for the likelihood function to be continuous over a parameter space Θ that is compact. For an open Θ the likelihood function may increase without ever reaching a supremum value.

In practice, it is often convenient to work with the natural logarithm of the likelihood function, called the log-likelihood:

$\ell(\theta\,;\mathbf{y}) = \ln \mathcal{L}(\theta\,;\mathbf{y})$. Since the logarithm is a monotonic function, the maximum of $\ell(\theta\,;\mathbf{y})$ occurs at the same value of θ as does the maximum of $\mathcal{L}$. If $\ell(\theta\,;\mathbf{y})$ is differentiable in θ, the necessary conditions for the occurrence of a maximum (or a minimum) are

$$\frac{\partial \ell}{\partial \theta_1} = 0, \quad \frac{\partial \ell}{\partial \theta_2} = 0, \quad \ldots, \quad \frac{\partial \ell}{\partial \theta_k} = 0, \tag{66}$$

known as the likelihood equations. For some models, these equations can be explicitly solved for $\widehat{\theta}$, but in general no closed-form solution to the maximization problem is known or available, and an MLE can only be found via numerical optimization. Another problem is that in finite samples, there may exist multiple roots for the likelihood equations. Whether the identified root $\widehat{\theta}$ of the likelihood equations is indeed a (local) maximum depends on whether the matrix of second-order partial and cross-partial derivatives,

$$\mathbf{H}(\widehat{\theta}) = \begin{bmatrix} \left.\frac{\partial^2 \ell}{\partial \theta_1^2}\right|_{\theta=\widehat{\theta}} & \left.\frac{\partial^2 \ell}{\partial \theta_1\, \partial \theta_2}\right|_{\theta=\widehat{\theta}} & \cdots & \left.\frac{\partial^2 \ell}{\partial \theta_1\, \partial \theta_k}\right|_{\theta=\widehat{\theta}} \\ \left.\frac{\partial^2 \ell}{\partial \theta_2\, \partial \theta_1}\right|_{\theta=\widehat{\theta}} & \left.\frac{\partial^2 \ell}{\partial \theta_2^2}\right|_{\theta=\widehat{\theta}} & \cdots & \left.\frac{\partial^2 \ell}{\partial \theta_2\, \partial \theta_k}\right|_{\theta=\widehat{\theta}} \\ \vdots & \vdots & \ddots & \vdots \\ \left.\frac{\partial^2 \ell}{\partial \theta_k\, \partial \theta_1}\right|_{\theta=\widehat{\theta}} & \left.\frac{\partial^2 \ell}{\partial \theta_k\, \partial \theta_2}\right|_{\theta=\widehat{\theta}} & \cdots & \left.\frac{\partial^2 \ell}{\partial \theta_k^2}\right|_{\theta=\widehat{\theta}} \end{bmatrix}, \tag{67}$$

known as the Hessian matrix is negative semi-definite at $\widehat{\theta}$, which indicates local concavity. Conveniently, most common probability distributions—in particular the exponential family—are logarithmically concave.

Note that the probability density function using the shape-scale parameterization for gamma distribution is given by

$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \quad \text{for } x > 0 \text{ and } k, \theta > 0. \tag{68}$$

Here $\Gamma(k)$ is the gamma function evaluated at k. The likelihood function for N iid observations $(x_1, \ldots, x_N)$ is

$L(k, \theta) = \prod_{i=1}^{N} f(x_i; k, \theta)$ from which we calculate the log-likelihood function

$$\ell(k, \theta) = (k - 1) \sum_{i=1}^{N} \ln(x_i) - \sum_{i=1}^{N} \frac{x_i}{\theta} - Nk \ln(\theta) - N \ln(\Gamma(k)) \tag{69}$$

Finding the maximum with respect to θ by taking the derivative and setting it equal to zero yields the maximum likelihood estimator of the θ parameter:

$\hat{\theta} = \frac{1}{kN} \sum_{i=1}^{N} x_i.$

Substituting this into the log-likelihood function gives

$$\ell = (k - 1) \sum_{i=1}^{N} \ln(x_i) - Nk - Nk \ln\left(\frac{\sum x_i}{kN}\right) - N \ln(\Gamma(k)) \tag{70}$$

Finding the maximum with respect to k by taking the derivative and setting it equal to zero yields

$$\ln(k) - \psi(k) = \ln\left(\frac{1}{N} \sum_{i=1}^{N} x_i\right) - \frac{1}{N} \sum_{i=1}^{N} \ln(x_i) \tag{71}$$

An initial value of k can be found either using the method of moments, or using the approximation

$$\ln(k) - \psi(k) \approx \frac{1}{2k}\left(1 + \frac{1}{6k + 1}\right) \tag{72}$$

If we let

$$s = \ln\left(\frac{1}{N} \sum_{i=1}^{N} x_i\right) - \frac{1}{N} \sum_{i=1}^{N} \ln(x_i) \tag{73}$$

then k is approximately

$$k \approx \frac{3 - s + \sqrt{(s - 3)^2 + 24s}}{12s} \tag{74}$$

Suppose random variable X is exponentially distributed with rate parameter θ, and $x_1, \ldots, x_n$ are n independent samples from X, with sample mean $\bar{x}$.

The likelihood function for θ, given an independent and identically distributed sample $x = (x1, ..., xn)$ drawn from the variable, is:

$$L(\theta) = \prod_{i=1}^{n} \theta \exp(-\theta x_i) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) = \theta^n \exp(-\theta n \bar{x}) \tag{75}$$

, where:

$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the sample mean.

The derivative of the likelihood function's logarithm is:

$$\frac{d}{d\theta} \ln(L(\theta)) = \frac{d}{d\theta}(n \ln(\theta) - \theta n \bar{x}) = \frac{n}{\theta} - n\bar{x} \begin{cases} > 0, & 0 < \theta < \frac{1}{\bar{x}}, \\ = 0, & \theta = \frac{1}{\bar{x}}, \\ < 0, & \theta > \frac{1}{\bar{x}}. \end{cases} \tag{76}$$

Consequently, the maximum likelihood estimate for the rate parameter is:

$$\widehat{\theta} = \frac{1}{\bar{x}} = \frac{n}{\sum_i x_i} \tag{77}$$

Although this is not an unbiased estimator of θ, $\bar{x}$ is an unbiased MLE estimator of $1/\theta = \beta$, where β is the scale parameter defined in the 'Alternative parameterization' section above and the distribution mean.

The bias of $\hat{\theta}_{\text{mle}}$ is equal to

$$b \equiv E\left[(\hat{\theta}_{\text{mle}} - \theta)\right] = \frac{\theta}{n} \tag{78}$$

which yields the bias-corrected maximum likelihood estimator

$$\hat{\theta}^*_{\text{mle}} = \hat{\theta}_{\text{mle}} - \hat{b} \tag{79}$$

The Fisher information, denoted $\mathcal{I}(\theta)$, for an estimator of the rate parameter θ is given as:

$$\mathcal{I}(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 \middle| \theta\right] = \int \left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta)\,dx \tag{80}$$

Substituting in the distribution and solving gives:

$$\mathcal{I}(\theta) = \int_0^\infty \left(\frac{\partial}{\partial\theta}\log\theta e^{-\theta x}\right)^2 \theta e^{-\theta x}\,dx = \int_0^\infty \left(\frac{1}{\theta} - x\right)^2 \theta e^{-\theta x}\,dx = \theta^{-2}. \tag{81}$$

This determines the amount of information each independent sample of an exponential distribution carries about the unknown rate parameter θ.

3.8 Joint Moments of i.i.d. Exponential Order Statistics

Let $X_1, \ldots, X_n$ be n independent and identically distributed exponential random variables with rate parameter θ. Let $X_{(1)}, \ldots, X_{(n)}$ denote the corresponding order statistics. For $i < j$, the joint moment $E[X_{(i)}X_{(j)}]$ of the order statistics $X_{(i)}$ and $X_{(j)}$ is given by

$$E[X_{(i)}X_{(j)}] = \frac{1}{(n-j+1)\theta}E[X_{(i)}] + E[X_{(i)}^2] = \frac{1}{(n-j+1)\theta}\frac{1}{(n-i+1)\theta} + \frac{2}{((n-i+1)\theta)^2}. \tag{82}$$

This can be seen by invoking the law of total expectation and the memoryless property:

$$\begin{aligned}
\mathrm{E}[X_{(i)}X_{(j)}] &= \int_0^\infty \mathrm{E}[X_{(i)}X_{(j)} \mid X_{(i)} = x]f_{X_{(i)}}(x)\,dx \\
&= \int_{x=0}^\infty x\,\mathrm{E}[X_{(j)} \mid X_{(j)} \geq x]f_{X_{(i)}}(x)\,dx \qquad (\text{since } X_{(i)} = x \implies X_{(j)} \geq x) \\
&= \int_{x=0}^\infty x\left[\frac{1}{(n-j+1)\theta} + x\right]f_{X_{(i)}}(x)\,dx \qquad (\text{by the memoryless property}) \\
&= \frac{1}{(n-j+1)\theta}\,\mathrm{E}[X_{(i)}] + \mathrm{E}[X_{(i)}^2].
\end{aligned} \tag{83}$$

The first equation follows from the law of total expectation. The second equation exploits the fact that once we condition on $X_{(i)} = x$, it must follow that $X_{(j)} \geq x$. The third equation relies on the memoryless property to replace $\mathrm{E}[X_{(j)} \mid X_{(j)} \geq x]$ with $\mathrm{E}[X_{(j)}] + x$.

3.9 Sum of Two Independent Exponential Random Variables

The sum of two independent variables corresponds to the convolution of probability distributions. If X_1 and X_2 are independent exponential random variables of independent observations, with respective rate parameters θ_1 and θ_2, then the probability density of $Z = X_1 + X_2$ is given by,

$$\begin{aligned}
f_Z(z) &= \int_0^z f_{X_1}(x_1)f_{X_2}(z - x_1)\,dx_1 \\
&= \theta_1\theta_2\exp[-\theta_2 z]\int_0^z \exp[(\theta_2 - \theta_1)x_1]\,dx_1 \\
&= \frac{\theta_1\theta_2}{\theta_2 - \theta_1}\left(\exp[-\theta_1 z] - \exp[-\theta_2 z]\right)
\end{aligned} \tag{84}$$

The mean and variance of $Z = X_1 + X_2$ are found to be

$$\begin{aligned}
\mathrm{E}[Z] &= \int z\,df_Z(z) \\
&= \theta_1^{-1} + \theta_2^{-1} \\
\mathrm{var}(Z) &= \mathrm{E}[Z^2] - \mathrm{E}[Z]^2 \\
&= \int z^2\,df_Z(z) - (\theta_1^{-1} + \theta_2^{-1})^2 \\
&= \theta_1^{-2} + \theta_2^{-2}
\end{aligned} \tag{85}$$

The cumulative distribution function of the sum of two independent exponential random variables then follows as,

$$F_Z(z) = \Pr(Z \leq z) = \int_0^z f_Z(t)\, dt$$

$$= 1 + \frac{\theta_1}{\theta_2 - \theta_1} \exp[-\theta_2 z] - \frac{\theta_2}{\theta_2 - \theta_1} \exp[-\theta_1 z] \tag{86}$$

For the Weibull distribution, the maximum likelihood estimator for the ξ parameter given ν is

$$\widehat{\xi^\nu} = \frac{1}{n} \sum_{i=1}^{n} x_i^\nu \tag{87}$$

The maximum likelihood estimator for ν is the solution for ν of the following equation

$$0 = \frac{\sum_{i=1}^{n} x_i^\nu \ln x_i}{\sum_{i=1}^{n} x_i^\nu} - \frac{1}{\nu} - \frac{1}{n} \sum_{i=1}^{n} \ln x_i \tag{88}$$

This equation defining $\widehat{\nu}$ only implicitly, one must generally solve for ν by numerical means.

When $x_1 > x_2 > \cdots > x_N$ are the N largest observed samples from a dataset of more than N samples, then the maximum likelihood estimator for the ξ parameter given ν is

$$\widehat{\xi^\nu} = \frac{1}{N} \sum_{i=1}^{N} (x_i^\nu - x_N^\nu) \tag{89}$$

Also given that condition, the maximum likelihood estimator for ν is

$$0 = \frac{\sum_{i=1}^{N} (x_i^\nu \ln x_i - x_N^\nu \ln x_N)}{\sum_{i=1}^{N} (x_i^\nu - x_N^\nu)} - \frac{1}{N} \sum_{i=1}^{N} \ln x_i \tag{90}$$

Again, this being an implicit function, one must generally solve for ν by numerical means.

4. Statistical Results and Data Analysis

An application of the distributions discussed above were carried out on the Ovarian Cancer Survival Data in which a randomized trial was performed comparing two treatments for ovarian cancer .The specified distributions were then compared parametrically thus the best choice was based on the minimum value of the AIC results.

There are different methods for selecting the most appropriate model in statistical analysis. The most commonly used methods include information and likelihood based criteria. To compare the different sub-families of a distribution used in the study , the information based criteria is applied . The most commonly used model selection criteria are the Akaike information criterion (AIC) and Bayesian information criterion (BIC). AIC is given by the expression,

$$AIC = -2log(L) + 2k \tag{91}$$

where L is the maximized likelihood value and k is the number of parameters in the model.

BIC is given by the expression

$$BIC = -2log(L) + kln(N) \tag{92}$$

where N is the total sample size.

The model with the smallest AIC value is considered a better fit.

Table 1. Table showing results when Generalised Gamma is used as the distribution

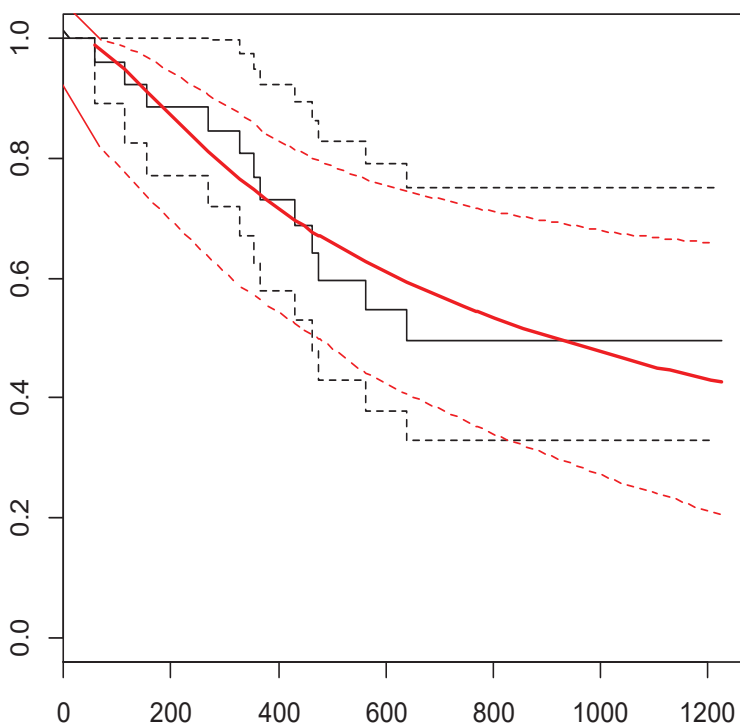| Parameters	Estimates	L95	U95	SE	Log-likelihood	AIC
κ	6.426	4.984	7.868	0.736		
ϕ	1.426	0.888	2.292	0.345		
β	-0.766	-3.340	1.807	1.313		
Log-likelihood					-96.94907	
AIC						199.8981

Figure 1. For gengamma distribution

Table 2. Table showing results when Exponential is used as the distribution

Parameters	Estimates	L95	U95	SE	Log-likelihood	AIC
rate	0.000770	0.000437	0.001356	0.000222		
Log-likelihood					-98.0322	
AIC						198.0644

Table 3. Table showing results when Weibull is used as the distribution

Parameters	Estimates	L95	U95	SE	Log-likelihood	AIC
shape	1.108	0.674	1.822	0.281		
scale	1225.419	690.421	2174.979	358.714		
Log-likelihood					-97.9539	
AIC						199.9078

5. Discussion and Conclusion

Because of its constant failure rate property, the exponential distribution is an excellent model for the long flat "intrinsic failure" portion of the Bathtub Curve. Since most components and systems spend most of their lifetimes in this portion of the Bathtub Curve, this justifies frequent use of the exponential distribution (when early failures or wear out is not a concern). Just as it is often useful to approximate a curve by piecewise straight line segments, we can approximate any failure rate curve by week-by-week or month-by-month constant rates that are the average of the actual changing rate during the respective time durations. That way we can approximate any model by piecewise exponential distribution segments patched together. Some natural phenomena have a constant failure rate (or occurrence rate) property; for example, the arrival rate of cosmic ray alpha particles or Geiger counter tics. The exponential model works well for inter arrival times (while the Poisson distribution describes the total number of events in a given period). When these events trigger failures, the exponential life distribution model will naturally apply. A probability plot of normalized exponential data was generated, so that a perfect exponential fit is a diagonal line with slope 1. The probability plot for 200 normalized
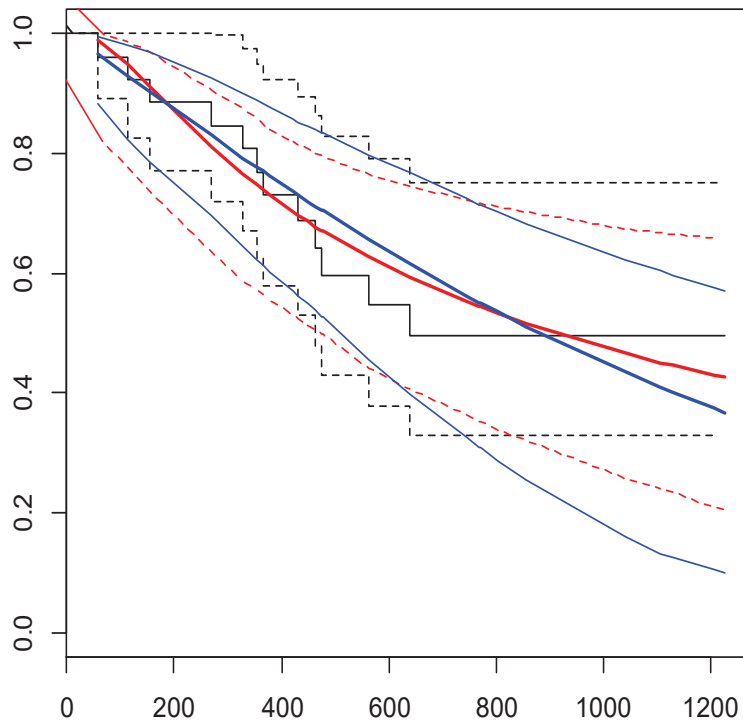
Figure 2. Weibul distribution superimposed to generalised gamma distribution
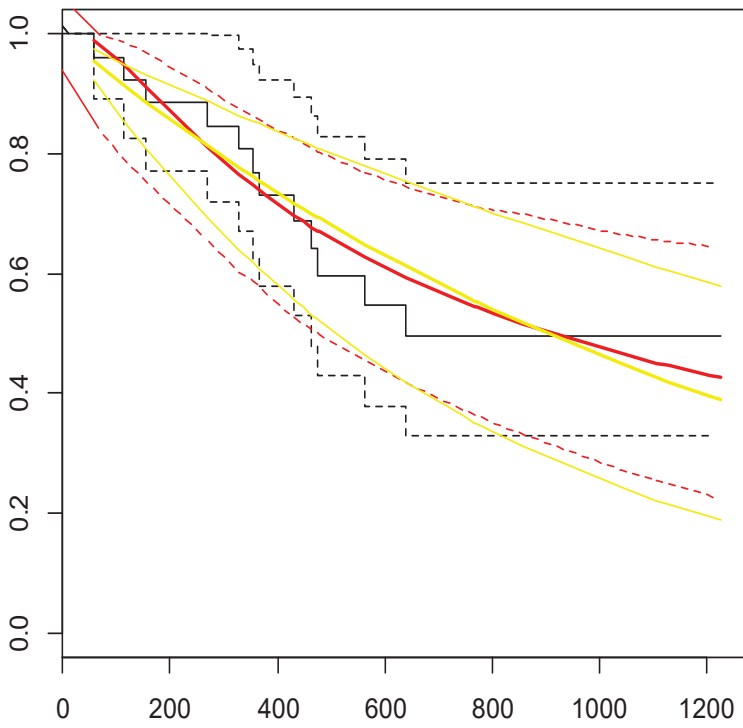
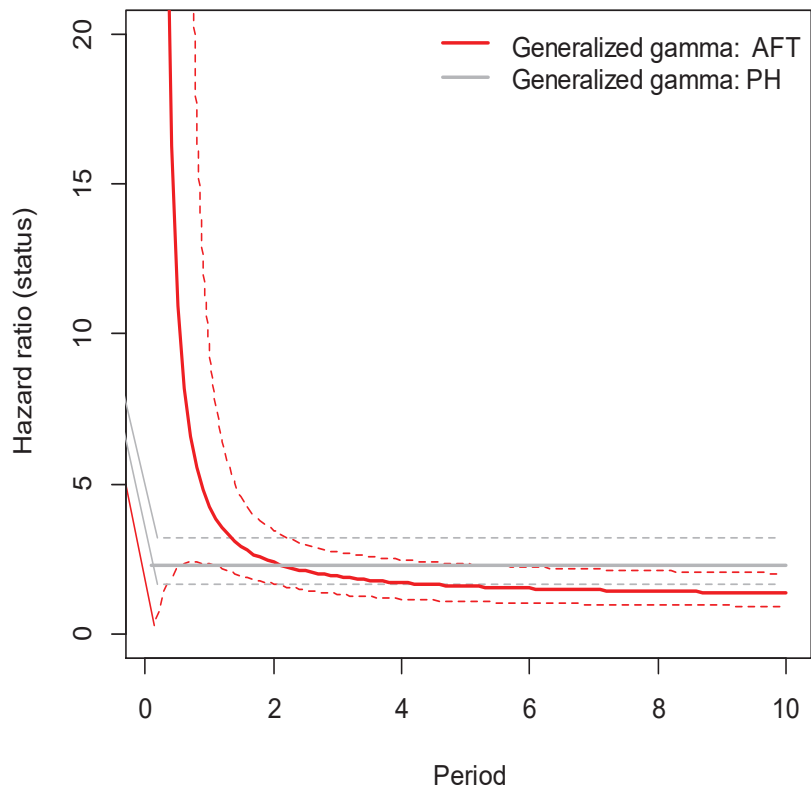Figure 3. Exponential distrbution superimposed to generalised gamma distribution

Figure 4. Generalized-gamma-graph with AFT and PH
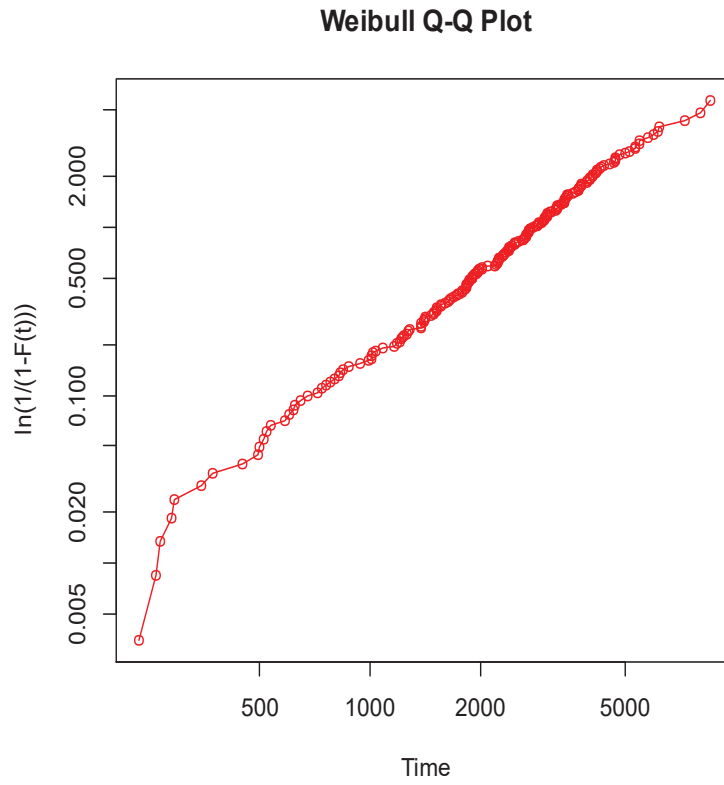
Weibull Q-Q Plot
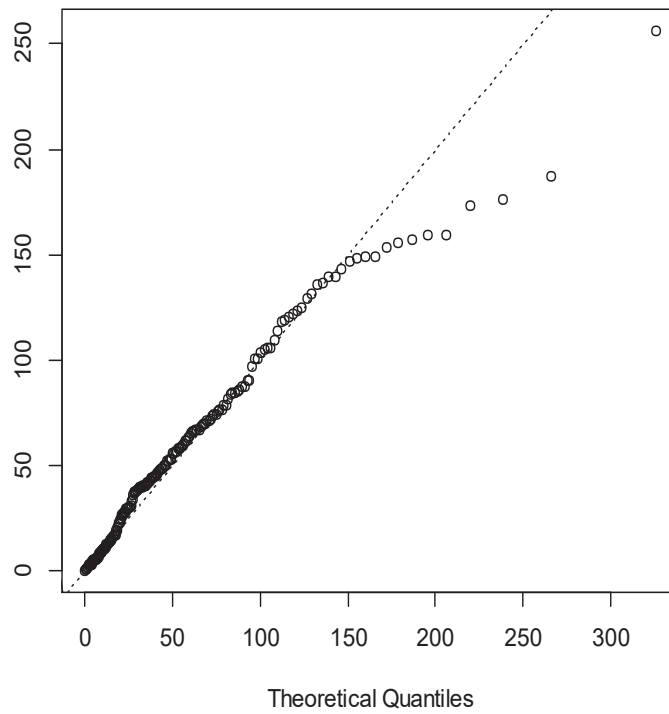

Figure 5. Weibull-qq-plot

Figure 6. Exponential-distribution quantile

random exponential observations ($\theta = 0.02$) is as shown in the graph above. The exponential PDF and CDF at 200 hours for the case where $\theta = 0.02$ were computed and found as the PDF value 0.0003663128 and the CDF value 0.9816844.

Because of its flexible shape and ability to model a wide range of failure rates, the Weibull has been used successfully in many applications as a purely empirical model. The Weibull model can be derived theoretically as a form of Extreme Value Distribution, governing the time to occurrence of the "weakest link" of many competing failure processes. This may explain why it has been so successful in applications such as capacitor, ball bearing, relay and material strength failures. Another special case of the Weibull occurs when the shape parameter is 2. The distribution is called the Rayleigh Distribution and it turns out to be the theoretical probability model for the magnitude of radial error when the x and y coordinate errors are independent normals with 0 mean and the same standard deviation. In case the data follow a Weibull distribution, the points should follow a straight line. The PDF and CDF values for failure time T = 2000 were computed, using the Weibull distribution with $\kappa = 1.8$ and $\lambda = 3000$ with 200 generated values. The PDF value is 0.0002678883 and the CDF value is 0.3824452. Special Case: When $\kappa = 1$, the Weibull reduces to the Exponential Model, with $\lambda = 1/\theta =$ the mean time to failure (MTTF).

The Weibull distribution has the ability to assume the characteristics of many different types of distributions. This has made it extremely popular among engineers and quality practitioners, who have made it the most commonly used distribution for modeling reliability data. Most researchers like incorporating the Weibull distribution into their data analysis because it is flexible enough to model a variety of data sets. The Weibull distribution can also model hazard functions that are decreasing, increasing or constant, allowing it to describe any phase of an item's lifetime. The Weibull distribution is used extensively in reliability applications to model failure times.

Acknowledgements

References

Stacy, E. W., & Mihram, G. A. (1965). Parameter estimation for a generalized gamma distribution. *Technometrics, 7,* 349-358. https://doi.org/10.2307/1266594

Shukla, G., & Kumar, V. (2006). Use of generalized gamma type model in life failure data. *Indian Journal of Applied Statistics, 10*, 13-20.

Lawless, J. F. (1980). Inference in the generalized gamma and log gamma distributions. *Technometrics, 22*(3), 409-419. https://doi.org/10.1080/00401706.1980.10486173

Cox, C., Chu, H., Schneider, M. F., & Munoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine, 26,* 4252-4374.

Carroll, R. J., & Stefanski, L. A. (1990). Approximate quasilikelihood estimation in models with surrogate predictor. *J. Amer. Statist. Assoc. 85,* 652-663. https://doi.org/10.2307/2290000

Stacy, E. W. (1962). A generalization of the gamma distribution. *Annals of Mathematical Statistics, 33,* 1187-92. https://doi.org/10.1214/aoms/1177704481

Nadarajah, S., & Gupta, A. K. (2007). A generalized gamma distribution with application to drought data. *Mathem. Comput. Simulation, 74,* 1-7. https://doi.org/10.1016/j.matcom.2006.04.004

Dadpay, A., Soofi, E. S., & Soyer, R. (2007). Information measures for generalized gamma family. *J. Econometrics, 138,* 568-585. https://doi.org/10.1016/j.jeconom.2006.05.010

Ghitany, M. E. (1998). On the recent generalization of gamma distribution. *Communication in Statistics-Theory and Methods, 27*(1), 223-233. https://doi.org/10.1080/03610929808832662

Khodabin, M., & Ahmadabadi A. R. (2010). Some properties of generalized gamma distribution. *Mathematica science, 4*(1).

Leipnik, R. B. (January 1991). On Lognormal Random Variables: I-The Characteristic Function. *Journal of the Australian Mathematical Society Series B, 32*(3), 327-347. https://doi.org/10.1017/S0334270000006901

Jiang, R., & Murthy, D. N. P. (2011). A study of Weibull shape parameter: Properties and significance. *Reliability Engineering and System Safety, 96*(12), 1619-26. https://doi.org/10.1016/j.ress.2011.09.003

Bauckhage, C. (2013). Computing the Kullback-Leibler Divergence between two Weibull Distributions.

Sharif, M. N., & Islam, M. N. (1980). The Weibull distribution as a general model for forecasting technological change.

Technological Forecasting and Social Change, 18(3), 247-56. https://doi.org/10.1016/0040-1625(80)90026-8

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). Continuous Univariate Distributions, volume 1, chapter 21. Wiley, New York.

Weibull, W. (1951). A Statistical Distribution Function of Wide Applicability. Journal of Applied Mechanics.

Park, S. Y., & Bera, A. K. (2009). Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics, 150*(2), 219-230.

Zhi-Sheng, Y., & Nan, C. (2017) Closed-Form Estimators for the Gamma Distribution Derived from Likelihood Equations. *The American Statistician, 71*(2), 177-181.

Ahrens, J. H., & Dieter, U. (1974). Computer methods for sampling from gamma, beta. *Poisson and binomial distributions Computing, 12*(3), 223-246. https://doi.org/10.1007/BF02293108

Choi, S. C., & Wette, R. (1969). Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. *Technometrics, 11*(4), 683-690. https://doi.org/10.1080/00401706.1969.10490731

Marsaglia, G., & Tsang, W. W. (2000). A simple method for generating gamma variables. *ACM Transactions on Mathematical Software, 26*(3), 363-372. https://doi.org/10.18637/jss.v005.i08

Minka, T. P. (2002). Estimating a Gamma distribution.

Rossi, R. J. (2018). Mathematical Statistics : An Introduction to Likelihood Based Inference. New York: John Wiley and Sons. p. 227. https://doi.org/10.1002/9781118771075

Myung, I. J. (2003). Tutorial on Maximum Likelihood Estimation. *Journal of Mathematical Psychology, 47*(1), 90-100. https://doi.org/10.1016/S0022-2496(02)00028-7

Gould, W., Pitblado, J., & Poi, B. (2010). Maximum Likelihood Estimation with Stata (Fourth ed.). College Station: Stata Press. pp. 13-20.

Pratt, J. W. (1976). F. Y. Edgeworth and R. A. Fisher on the efficiency of maximum likelihood estimation. *The Annals of Statistics, 4*(3), 501-514. https://doi.org/10.1214/aos/1176343457

Aldrich, J. (1997). R. A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science, 12*(3), 162-176. https://doi.org/10.1214/ss/1030037906

Copyrights