



Some properties of generalized gamma distribution

Morteza Khodabin^{a,1}, Alireza Ahmadabadi^b

^aDepartment of Mathematics, Islamic Azad University-Karaj Branch, Karaj, Iran

^bDepartment of Mathematics, Payam Nur Branch, Orumiieh, Iran

Abstract

In this paper, the generalized gamma (GG) distribution that is a flexible distribution in statistical literature, and has exponential, gamma, and Weibull as subfamilies, and lognormal as a limiting distribution is introduced. The power and logarithmic moments of this family is defined. A new moment estimation method of parameters of GG family using it's characterization is presented, this method is compared with MLE method in gamma subfamily for small and large sample size. Here we study GG entropy representation and its estimation. In addition Kullback-Leibler discrimination , Akaike and Bayesian information criterion is discussed. In brief, this paper consist of presentation of general review of important properties in GG family.

Keywords: Generalized gamma distribution, Generalized normal distribution, Logarithmic moment, Entropy, Akaike criterion, Bayesian criterion, Kullback-Leibler discrimination.

© 2010 Published by Islamic Azad University-Karaj Branch.

1 .Introduction

The generalized gamma(GG)distribution presents a flexible family in the varieties of shapes and hazard functions for modeling duration. It was introduced by Stacy [23]. Distributions that are used in duration analysis in economics include exponential [6,15]

¹Corresponding Author. E-mail Address:m-khodabin@kiau.ac.ir

, lognormal [7], gamma [19], and Weibull [8]. The GG family, which encompasses exponential, gamma, and Weibull as subfamilies, and lognormal as a limiting distribution, has been used in economics by Jaggia [13], Yamaguchi [26], and Allenby et al [3]. Some authors [13] have argued that the flexibility of GG makes it suitable for duration analysis, while others [3] have advocated use of simpler models because of estimation difficulties caused by the complexity of GG parameter structure. Obviously, there would be no need to endure the costs associated with the application of a complex GG model if the data do not discriminate between the GG and members of its subfamilies, or if the fit of a simpler model to the data is as good as that for the complex GG. Hager and Bain [9] inhibited applications of the GG model. Prentice [21] resolved the convergence problem using a nonlinear transformation of GG model. However, despite its long history and growing use in various applications, the GG family and its properties has been remarkably presented in different papers. Maximum-likelihood estimation of the parameters and quasi maximum likelihood estimators for its subfamily (two-parameter gamma distribution) can be found in [10,11,24,25]. Hwang, T. et al [12] introduced a new moment estimation of parameters of the generalized gamma distribution using its characterization. In information theory, thus far a maximum entropy (ME) derivation of GG is found in Kapur [14], where it is referred to as generalized Weibull distribution, and the entropy of GG has appeared in the context of flexible families of distributions [20]. Some concepts of this family in information theory has introduced by Dadpay et al [5]. The main objective of this paper is to review by more details the important works on GG family. The paper is organized as follows:

Section 2 defines generalized gamma distribution and subfamily of this distribution. Section 2 gives new moment estimation of parameters of GG family, using its characterization. Section 4 discusses entropy of GG distribution and its estimation. Section 5 presents Kullback-Leibler discrimination in GG family. Section 6 illustrates Akaike and Bayesian information criterion in GG family. Section 7 gives some brief concluding.

2 .Generalized gamma distribution

The probability density of the generalized gamma distribution ($GG(\alpha, \tau, \lambda)$) is given by

$$f(y|\alpha, \tau, \lambda) = \frac{\tau}{\lambda\Gamma(\alpha)} \left(\frac{y}{\lambda}\right)^{\alpha\tau-1} e^{-\left(\frac{y}{\lambda}\right)^\tau} \quad y \geq 0, \quad \tau, \alpha, \lambda > 0 \quad (1)$$

where $\Gamma(\cdot)$ is the gamma function, α and τ are shape parameters, and λ is the scale parameter. The GG family is flexible in that it includes several well-known models as subfamilies. The subfamilies of GG thus far considered in the literature are exponential ($\alpha = \tau = 1$), gamma for ($\tau = 1$), and Weibull for ($\alpha = 1$). The lognormal distribution is also obtained as a limiting distribution when $\alpha \rightarrow \infty$. By letting $\tau=2$ we obtain a subfamily of GG which is known as the generalized normal distribution (GN). The GN is itself a flexible family and includes half-normal ($\alpha = 1/2, \lambda^2 = 2\sigma^2$), Rayleigh ($\alpha = 1, \lambda^2 = 2\sigma^2$), Maxwell-Boltzmann ($\alpha = 3/2$), and chi ($\alpha = k/2, k = 1, 2, \dots$). An important property of GG family for information analysis is that the family is closed under power transformation [5]. That is, if $X \sim GG(\alpha, \tau, \lambda)$, then

$$Y = X^s \sim GG\left(\alpha, \frac{\tau}{s}, \lambda^s\right), \quad s > 0.$$

In particular,

$$Y = X^\tau \sim \text{Gamma}(\alpha, \lambda^\tau).$$

It also has the property that $Z = \eta X$ has $GG(\eta\alpha, \tau, \lambda)$ distribution. In the below we introduce subfamily of this distribution briefly.

2.1 Exponential distribution

The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process. Exponential variables can also be used to model situations where certain events occur with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between road kills on a given road. In queuing theory, the service times of agents in a system

(e.g. how long it takes for a bank teller etc. to serve a customer) are often modeled as exponentially distributed variables. Reliability theory and reliability engineering also make extensive use of the exponential distribution. Because of the memoryless property of this distribution, it is well-suited to model the constant hazard rate portion of the bathtub curve used in reliability theory.

Failure rate is the frequency with which an engineered system or component fails, expressed for example in failures per hour. It is important in reliability engineering. By calculating the failure rate for smaller and smaller intervals of time Δt , the interval becomes infinitely small. This results in the hazard function, which is the instantaneous failure rate at any point in time:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{R(t) - R(t + \Delta t)}{\Delta t \cdot R(t)}.$$

Continuous failure rate depends on a failure distribution, $F(t)$, which is a cumulative distribution function that describes the probability of failure prior to time t ,

$$P(\mathbf{T} \leq t) = F(t) = 1 - R(t), \quad t \geq 0.$$

The hazard function can be defined now as

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

Many probability distributions can be used to model the failure distribution. A common model is the exponential failure distribution,

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

For an exponential failure distribution the hazard rate is a constant with respect to time (that is, the distribution is "memoryless"). For other distributions, such as a Weibull distribution or a log-normal distribution, the hazard function may not be constant with respect to time.

2.2 Gamma distribution

In probability theory and statistics, the gamma distribution is a two-parameter family of continuous probability distributions. It has a scale parameter λ and a shape parameter α . If α is an integer then the distribution represents the sum of α independent exponentially distributed random variables, each of which has a mean of λ (which is equivalent to a rate parameter of λ^{-1}). The gamma distribution is frequently a probability model for waiting times; for instance, in life testing, the waiting time until death is a random variable that is frequently modeled with a gamma distribution.

2.3 Weibull distribution

The Weibull distribution is a continuous probability distribution. It is named after Waloddi Weibull who described it in detail in 1951, although it was first identified by Frchet in 1927 and first applied by Rosin and Rammler in 1933 to describe the size distribution of particles. The Weibull distribution is often used in the field of life data analysis due to its ability to fit the exponential distribution and the normal distribution and interpolate a range of shapes in between them.

2.4 Generalized normal distribution

The generalized normal distribution or generalized Gaussian distribution is either of parametric continuous probability distributions on the real line. The GN family includes the below well-known models as subfamilies.

2.4.1 Half normal distribution

The half-normal distribution is the probability distribution of the absolute value of a random variable that is normally distributed with expected value 0 and variance σ^2 , i.e. if X is normally distributed with mean 0, then $Y = |X|$ is half-normally distributed.

2.4.2 Rayleigh distribution

In statistic literature, the Rayleigh distribution is a continuous probability distribution. As an example of how it arises, the wind speed will have a Rayleigh distribution if the components of the two-dimensional wind velocity vector are uncorrelated and normally distributed with equal variance. The distribution is named after Lord Rayleigh.

2.4.3 Maxwell–Boltzmann distribution

The Maxwell–Boltzmann distribution applies to ideal gases close to thermodynamic equilibrium, negligible quantum effects, and non-relativistic speeds. It forms the basis of the kinetic theory of gases, which explains many fundamental gas properties, including pressure and diffusion. The Maxwell–Boltzmann distribution is usually thought of as the distribution for molecular speeds, but it can also refer to the distribution for velocities, momenta, and magnitude of the momenta of the molecules, each of which will have a different probability distribution function, all of which are related. The original derivation by Maxwell assumed all three directions would behave in the same fashion, but a later derivation by Boltzmann dropped this assumption using kinetic theory. The Maxwell–Boltzmann distribution can now most readily be derived from the Boltzmann distribution for energies.

2.4.4 Chi-square distribution

In probability theory, the chi-square distribution ($\lambda = 1$ in the Chi distribution) with k degrees of freedom is the distribution of a sum of squares of k independent standard normal random variables. It is one of the most widely used probability distributions in inferential statistics, e.g. in hypothesis testing, goodness of fit tests, independence of two criteria of classification of qualitative data, Friedman's analysis of variance by ranks, estimating variances, estimating the slope of a regression line via its role in Student's t -distribution, analysis of variance problems via its role in the F -distribution and so on.

The sum of squares of statistically independent unit-variance Gaussian variables which do not have mean zero yields a generalization of the chi-square distribution called the noncentral chi-square distribution.

2.5 Two important moments

The power and logarithmic moments of GG distribution are played important role in statistical inference and information theory.

Theorem 2.1 *Let $X \sim GG(\alpha, \tau, \lambda)$, then*

$$E(X^s) = \frac{\lambda^s \Gamma(\frac{s}{\tau} + \alpha)}{\Gamma(\alpha)}; \tag{2}$$

and

$$E(\log(X^s)) = s \log \lambda + \frac{s}{\tau} \psi(\alpha); \tag{3}$$

where $\psi(\alpha) = \frac{d \log \Gamma(\alpha)}{d\alpha}$ is the digamma function.

proof: According to expected value definition and by using $y = (\frac{x}{\lambda})^\tau$ we have

$$E(X^s) = \int_0^\infty x^s \frac{\tau}{\lambda \Gamma(\alpha)} (\frac{x}{\lambda})^{\alpha\tau-1} e^{-(\frac{x}{\lambda})^\tau} dy = \frac{\lambda^s}{\Gamma(\alpha)} \int_0^\infty y^{(\frac{s}{\tau} + \alpha) - 1} e^{-y} dy,$$

then from definition of gamma function (2) is obtained. Similarly we can prove (3).

Corollary 2.2 *If X has GG distribution, then*

$$i) E(X) = \frac{\lambda \Gamma(\alpha + \frac{1}{\tau})}{\Gamma(\alpha)}$$

$$ii) Var(X) = \frac{\lambda^2 \Gamma(\alpha + \frac{2}{\tau})}{\Gamma(\alpha)} - (\frac{\lambda \Gamma(\alpha + \frac{1}{\tau})}{\Gamma(\alpha)})^2$$

Then, we can have the below table.

Distribution name	α	τ	λ	Mean	Variance
Exponential	1	1	λ	λ	λ^2
Gamma	α	1	λ	$\alpha\lambda$	$\alpha\lambda^2$
Weibull	1	τ	λ	$\lambda\Gamma(1 + \frac{1}{\tau})$	$\lambda^2\Gamma(1 + \frac{2}{\tau}) - (Mean)^2$
Generalized normal	α	2	λ	$\frac{\lambda\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)}$	$\lambda^2\alpha - (Mean)^2$
Half normal	0.5	2	$\sqrt{2\sigma^2}$	$\sigma\sqrt{\frac{2}{\pi}}$	$\sigma^2(1 - \frac{2}{\pi})$
Rayleigh	1	2	$\sqrt{2\sigma^2}$	$\sigma\sqrt{\frac{\pi}{2}}$	$\sigma^2(2 - \frac{\pi}{2})$
Maxwell Boltzmann	$\frac{3}{2}$	2	λ	$\frac{2\lambda}{\sqrt{\pi}}$	$\lambda^2(1 - \frac{4}{\pi})$
Chi	$\frac{k}{2}$	2	λ	$\frac{\lambda\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}$	$\frac{\lambda^2\Gamma(\frac{k+2}{2})}{\Gamma(\frac{k}{2})} - (Mean)^2$

Table 1. Mean and variance for subfamilies of GG distribution

3 Moment method estimation of parameters

In this section, we recall the new method based on moments, using it's characterization for estimation of three parameter of GG family. Also the results of this method is compared with MLE method via simulation. The results shows that this new method is easy and more efficient than MLE method in small sample.

The proves of the following two theorems can be found in [12].

Theorem 3.1 *Let $n \geq 3$ and X_1, X_2, \dots, X_n be nonnegative, independent and identical distributed with $f(x)$, then*

$$\bar{X}_n \text{ and } V_n = \frac{S_n}{\bar{X}_n} \text{ are independent iff } f(x) \sim GG \tag{4}$$

Theorem 3.2 *Let $n \geq 3$ and X_1, X_2, \dots, X_n be nonnegative, independent and identical distributed according (1), then*

$$\begin{aligned}
 i) \quad E(S_n^2) &= \frac{\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau}) - \Gamma^2(\alpha + \frac{1}{\tau})}{\alpha^2\Gamma^2(\alpha)} \\
 ii) \quad E(\frac{S_n^2}{\bar{X}_n^2}) &= \frac{n[\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau}) - \Gamma^2(\alpha + \frac{1}{\tau})]}{\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau}) + (n - 1)\Gamma^2(\alpha + \frac{1}{\tau})}
 \end{aligned} \tag{5}$$

Furthermore, in GG distribution we have

$$\frac{\sigma^2}{\mu^2} = \frac{\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau})}{\Gamma^2(\alpha + \frac{1}{\tau})} - 1,$$

and it can be shown that[12]

$$E\left(\frac{S_n^2}{\bar{X}_n^2}\right) \rightarrow \frac{\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau})}{\Gamma^2(\alpha + \frac{1}{\tau})} - 1.$$

Therefore, $\frac{S_n^2}{\bar{X}_n^2} \rightarrow \frac{\sigma^2}{\mu^2}$ asymptotically. From (2) and corollary 2.2 we know that

$$i) \quad E(\bar{X}_n) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = E(X) = \frac{\lambda\Gamma(\alpha + \frac{1}{\tau})}{\Gamma(\alpha)}$$

$$ii) \quad E(\bar{X}_n^\tau) = E\left(\frac{\sum_{i=1}^n X_i^\tau}{n}\right) = E(X^\tau) = \alpha\lambda^\tau$$

$$iii) \quad E\left(\frac{S_n^2}{n\bar{X}_n^2}\right) = \frac{\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau}) - \Gamma^2(\alpha + \frac{1}{\tau})}{\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau}) + (n-1)\Gamma^2(\alpha + \frac{1}{\tau})}$$

Then, we can solve numerically via moment method the below equations for estimating of GG parameters

$$\left\{ \begin{array}{l} \frac{\sum_{i=1}^n X_i}{n} = \frac{\lambda\Gamma(\alpha + \frac{1}{\tau})}{\Gamma(\alpha)}; \\ \frac{\sum_{i=1}^n X_i^\tau}{n} = \alpha\lambda^\tau; \\ \frac{S_n^2}{n\bar{X}_n^2} = \frac{\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau}) - \Gamma^2(\alpha + \frac{1}{\tau})}{\Gamma(\alpha)\Gamma(\alpha + \frac{2}{\tau}) + (n-1)\Gamma^2(\alpha + \frac{1}{\tau})}. \end{array} \right. \quad (6)$$

3.1 Application in gamma subfamily

Here, we generate

- 200 sample of size 10 from $gamma(\alpha = 2, \lambda = 0.5)$ distribution (G1)
- 200 sample of size 10 from $gamma(\alpha = 4, \lambda = 2)$ distribution (G2)
- 200 sample of size 10 from $gamma(\alpha = 7, \lambda = 3)$ distribution (G3)
- 200 sample of size 50 from $gamma(\alpha = 3, \lambda = 2)$ distribution (G4)

The below tables, summarizes the results. In these tables, "Std" is used for standard deviation,"CI" shows 95 percent confidence interval ,"L" is used for length of confidence interval and "MSE" shows mean square error.

Method	$\hat{\alpha}$	Std	CI(95p)	L	MSE
MLE (G1)	2.7056	1.1342	(2.5484 , 2.8647)	0.3163	1.7791
MME (G1)	2.5353	1.0499	(2.3838,2.6816)	0.2928	1.3833
MLE (G2)	5.5703	2.7507	(5.1868,5.9539)	0.7671	9.9946
MME (G2)	5.1056	2.5495	(4.7501,5.4611)	0.711	7.6898
MLE (G3)	9.4345	3.9137	(8.8888,9.9803)	1.0915	21.1672
MME (G3)	8.5682	3.4193	(8.0914,9.045)	0.9536	14.0923
MLE (G4)	3.1351	0.5568	(3.0575,3.2128)	0.1553	0.3267
MME (G4)	3.1476	0.6847	(3.0521,3.2431)	0.191	0.4882

Table 2. Comparison results for α estimation

The above table shows that in G1,G2 and G3 states where sample sizes are small (n=10), the MME is better than MLE method. In the G4 state that sample size is nearly large (n=50), the MLE method is better than MME.

Method	$\hat{\lambda}$	Std	CI(95p)	L	MSE
MLE (G1)	0.4182	0.1638	(0.3954 , 0.4411)	0.0457	0.0334
MME (G1)	0.449	0.1807	(0.4238,0.4743)	0.0505	0.0351
MLE (G2)	1.756	0.8041	(1.6439,1.8681)	0.2242	0.7028
MME (G2)	1.9514	0.9765	(1.8153,2.876)	0.2723	0.9511
MLE (G3)	2.584	1.0786	(2.4336,2.7344)	0.3008	1.3307
MME (G3)	2.828	1.1788	(2.6637,2.9924)	0.3287	1.4121
MLE (G4)	1.9976	0.4617	(1.9332,2.0619)	0.1287	0.2121
MME (G4)	2.0128	0.5242	(1.9397,2.0858)	0.1461	0.2735

Table 3. Comparison results for λ estimation

The above table shows that in all states the MLE method is better than MME. In other words, very likely we can say two methods are capable.

4 Entropy and its estimation

The concept of Shannon's entropy [22] is the central role of information theory, sometimes referred as measure of uncertainty. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty. Henceforth we assume that log is to the base 2 and entropy is expressed in bits. For deriving entropy of the generalized gamma distribution, we need the following two definitions, that more details of them can be found in [4].

Definition 4.1 *The entropy of a discrete alphabet random variable f defined on the probability space (Ω, β, P) is defined by*

$$H_P(f) = - \sum_{a \in A} p(f = a) \log(p(f = a)). \quad (7)$$

It is obvious that $H_P(f) \geq 0$.

Definition 4.2 *The obvious generalization of the definition of entropy for a probability density function f defined on the real line is*

$$H(f) = - \int_{-\infty}^{+\infty} f(x) \log f(x) dx = E(-\log f(x)). \quad (8)$$

provided this integral exists.

Theorem 4.3 *Let $X \sim GG(\alpha, \tau, \lambda)$, then*

$$H(GG) = \log \lambda + \log \Gamma(\alpha) + \alpha - \log \tau + \left(\frac{1}{\tau} - \alpha\right) \Psi(\alpha) \quad (9)$$

proof: By definition (8) we can write

$$H(GG) = -E(\log f(y|\alpha, \tau, \lambda)) = -\log \tau + \alpha \tau \log \Gamma(\alpha) - \alpha \tau E(\log Y) + E(\log Y) + \frac{1}{\lambda \tau} E(Y^\tau). \quad (10)$$

Furthermore, from (2) and (3) we have

$$i) E(Y^\tau) = \frac{\lambda^\tau \Gamma(1 + \alpha)}{\Gamma(\alpha)}$$

$$ii) E(\log(Y)) = \log \lambda + \frac{1}{\tau} \psi(\alpha)$$

then, by substitute these relations in (9) the theorem is proved.

Corollary 4.4 For all values of τ , $H(GG)$ is increasing in α .

Corollary 4.5 For values of $\alpha < 1.5$, $H(GG)$ is increasing in τ .

We can summarize the entropy of subfamilies of GG distribution as below table.

Distribution name	α	τ	λ	Entropy
Exponential	1	1	λ	$\log \lambda + 1$
Gamma	α	1	λ	$\log \lambda + \log \Gamma(\alpha) + \alpha + (1 - \alpha)\Psi(\alpha)$
Weibull	1	τ	λ	$\log \lambda + 1 - \log \tau + (\frac{1}{\tau} - 1)\Psi(1)$
Generalized normal	α	2	λ	$\log \lambda + \log \Gamma(\alpha) + \alpha - 1 + (\frac{1}{2} - \alpha)\Psi(\alpha)$
Half normal	0.5	2	$\sqrt{2\sigma^2}$	$\log \sigma + \log \sqrt{\pi}$
Rayleigh	1	2	$\sqrt{2\sigma^2}$	$\frac{1}{2} + \log \sigma - \frac{1}{2}\Psi(1)$
Maxwell Boltzmann	$\frac{3}{2}$	2	λ	$\log \lambda + \log \frac{\sqrt{\pi}}{2} - \frac{1}{2} - \Psi(\frac{3}{2})$
Chi	$\frac{k}{2}$	2	λ	$\log \lambda + \log \Gamma(\frac{k}{2}) + \frac{k-2}{2} + (\frac{1-k}{2})\Psi(\frac{k}{2})$

Table 4. Entropy of subfamilies of GG distribution

4.1 Entropy estimation

Consider another form of (1) as

$$f(y|\alpha, \tau, \lambda) = \frac{\tau}{\lambda^{\alpha\tau}\Gamma(\alpha)} e^{(\alpha\tau-1)\log y - (\frac{y}{\lambda})^\tau} \quad y \geq 0, \quad \tau, \alpha, \lambda > 0. \quad (11)$$

Then, the likelihood function is given by

$$L(y_1, \dots, y_n|\alpha, \tau, \lambda) = \left(\frac{\tau}{\lambda^{\alpha\tau}\Gamma(\alpha)}\right)^n e^{(\alpha\tau-1)\sum_{i=1}^n \log y_i - \sum_{i=1}^n (\frac{y_i}{\lambda})^\tau} \quad y \geq 0, \quad \tau, \alpha, \lambda > 0. \quad (12)$$

Consequently,

$$l(\alpha, \tau, \lambda) = \log L(y_1, \dots, y_n | \alpha, \tau, \lambda) = n(\log \tau - \alpha\tau \log \lambda - \log \Gamma(\alpha) + (\alpha\tau - 1)\overline{\log y} - \frac{\overline{y^\tau}}{\lambda^\tau}), \quad (13)$$

where, $\overline{\log y} = \frac{\sum_{i=1}^n \log y_i}{n}$ and $\overline{y^\tau} = \frac{\sum_{i=1}^n y_i^\tau}{n}$.

By taking derivative to parameters we have

$$\begin{cases} \frac{\partial l(\alpha, \tau, \lambda)}{\partial \alpha} = -n\tau \log \lambda - n\Psi(\alpha) + n\tau\overline{\log y} = 0 \\ \frac{\partial l(\alpha, \tau, \lambda)}{\partial \lambda} = -\frac{n\alpha\tau}{\lambda} + \frac{n\tau\overline{y^\tau}}{\lambda^{\tau+1}} = 0 \end{cases} \quad (14)$$

By solving this equation and from (2) and (3) we have

$$\overline{\log y} = \log \lambda + \frac{1}{\tau}\psi(\alpha) = E(\log Y) \quad \text{and} \quad \overline{y^\tau} = \lambda^\tau \alpha = E(Y^\tau). \quad (15)$$

Then, by replacement (15) in (10) we get

$$\hat{H}(GG) = -(\log \hat{\tau} - \hat{\alpha}\hat{\tau} \log \hat{\lambda} - \log \Gamma(\hat{\alpha}) + (\hat{\alpha}\hat{\tau} - 1)\overline{\log y} - \frac{\overline{y^{\hat{\tau}}}}{\hat{\lambda}^{\hat{\tau}}}), \quad (16)$$

From (13),(16) we can write

$$\hat{H}(GG) = -\frac{l(\hat{\alpha}, \hat{\tau}, \hat{\lambda})}{n} \quad (17)$$

5 Kullback-Leibler discrimination

In information theory, the Kullback - Leibler (KL)divergence (also information divergence, discrimination information, or relative entropy) is a non-symmetric measure of the difference between two probability distributions P and Q. KL divergence is a special case of a broader class of divergences called f-divergences. Originally introduced by Solomon Kullback and Richard Leibler [16] as the directed divergence between two distributions, it is not the same as a divergence in calculus. Although it is often intuited as a distance metric, the KL divergence is not a true metric - for example, the KL from P to Q is not necessarily the same as the KL from Q to P. For more details see [17,18].

For probability distributions P and Q of a discrete random variable the KL divergence of Q from P is defined to be

$$K(P : Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

For distributions P and Q of a continuous random variable the summations give way to integrals, so that

$$K(P : Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx,$$

where p and q denote the densities of P and Q.

Let $GG_0 = GG(\alpha_0, \tau_0, \lambda_0)$ be a given GG distribution. Authors in [5] showed that the discrimination information function between GG_0 and GG is given by

$$K(GG : GG_0) = \log \frac{\phi_\tau}{\phi_\lambda^{\alpha\phi_\tau}} - \log \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)} - \alpha + \mu(\alpha, \phi_\tau, \phi_\lambda) + (\alpha\phi_\tau - \alpha_0)v(\alpha, \phi_\tau, \phi_\lambda), \quad (18)$$

where $\phi_\tau = \frac{\tau}{\tau_0}$, $\phi_\lambda = (\frac{\lambda}{\lambda_0})^{\tau_0}$, $\mu(\alpha, \phi_\tau, \phi_\lambda)$ is the first moment and $v(\alpha, \phi_\tau, \phi_\lambda)$ is the geometric mean of a GG distribution with parameters $(\alpha, \phi_\tau, \phi_\lambda)$. The discrimination information $K(GG : GG_0)$ is a complicated function of the parameters, (18) is a general representation that encompasses discrimination information functions between the GG and its subfamilies, between distributions within each subfamily, and between distributions from different subfamilies. The discrimination information between $GG(\alpha, \tau, \lambda)$ and $Gamma(\alpha_0, \lambda_0)$ is given by (18) with $\phi_\tau = \tau$. The discrimination information between $GG(\alpha, \tau, \lambda)$ and $Weibull(\tau_0, \lambda_0)$ is given by (18) with $\alpha_0 = 1$. The discrimination information between $GG(\alpha, \tau, \lambda)$ and $Exponential(\lambda_0)$ is given by (18) with $\phi_\tau = \tau$ and $\alpha_0 = 1$. The discrimination information between $GG(\alpha, \tau, \lambda)$ and $GN(\alpha_0, \lambda_0)$ is given by (18) with $\phi_\tau = \frac{\tau}{2}$ and $\alpha_0 = 2\alpha$.

6 Akaike and Bayesian information criterion

In order to introducing of an approach for model selection, we remember Akaike and Bayesian information criterion based on entropy estimation. Akaike's information crite-

tion, developed by Hirotugu Akaike [1,2] under the name of "an information criterion" (AIC) in 1971 and proposed in Akaike [2], is a measure of the goodness of fit of an estimated statistical model. It is grounded in the concept of entropy, in effect offering a relative measure of the information lost when a given model is used to describe reality and can be said to describe the tradeoff between bias and variance in model construction, or loosely speaking that of precision and complexity of the model. The AIC is not a test of the model in the sense of hypothesis testing, rather it is a test between models - a tool for model selection. Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best. From the AIC value one may infer that e.g. the top three models are in a tie and the rest are far worse, but it would be arbitrary to assign a value above which a given model is "rejected". In the general case, the AIC is

$$AIC = 2K - 2\log(L(\hat{\theta})),$$

where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model.

The Bayesian information criterion (BIC) or Schwarz Criterion is a criterion for model selection among a class of parametric models with different numbers of parameters. Choosing a model to optimize BIC is a form of regularization. It is very closely related to AIC. In BIC, the penalty for additional parameters is stronger than that of the AIC. The formula for the BIC is

$$BIC = K \log n - 2\log(L(\hat{\theta})).$$

The AIC and BIC methodology attempts to find the model that best explains the data with a minimum of their values. from (17) we have

$$l(\hat{\alpha}, \hat{\tau}, \hat{\lambda}) = -n\hat{H}(GG).$$

Then for GG family we have

$$AIC = 2n\hat{H}(GG) + 2K \tag{19}$$

and

$$BIC = 2n\hat{H}(GG) + K \log n \tag{20}$$

6.1 Application in model selection

For description of this manner, we generate two samples of sizes 100 and 10 from Weibull ($\tau = 10, \lambda = 5$) distribution. Thus Weibull distribution is real distribution. Suppose that some subfamilies of GG and normal distribution is considered as an approximate distributions. We will to find best distribution based on these criterions. The below tables, summarizes the results.

Distribution	MLE(θ)	\hat{H}	AIC	BIC
Exponential	9.2216	3.2215	646.3	648.9052
Gamma	(17.6593,0.5222)	2.1857	441.14	446.3503
Weibull	(10.0228,5.1649)	0.337	71.4	76.6103
Normal	(9.2216,2.0621)	2.1426	432.52	437.7303
Rayleigh	6.6801	2.5526	512.52	515.1252
Half normal	9.4471	2.9715	596.3	598.9052

Table 5. The results for n=100

Distribution	MLE(θ)	\hat{H}	AIC	BIC
Exponential	8.8104	3.1759	65.518	65.8206
Gamma	(17.5272,0.5027)	2.1438	46.876	47.4812
Weibull	(9.6449,4.4446)	0.2253	8.506	9.1112
Normal	(8.8104,2.2236)	2.2181	48.362	48.9672
Rayleigh	6.406	2.5107	52.214	52.5166
Half normal	9.0594	2.9296	60.592	60.8946

Table 6. The results for n=10

From above tables, we conclude that the Weibull distribution have smallest AIC and BIC among others, which is exactly what we have been expected.

7 Conclusion

This paper took the first major step toward reviewing of some important properties in GG model thus far. In brief we discussed :

(a) The generalized gamma distribution, subfamilies and limiting distribution of it. (b) A new moment estimation method of parameters of GG family. (c) Entropy representation and its estimation. (d) Kullback-Leibler discrimination. (e) Akaike and Bayesian information criterion. In addition, we gave two explanatory examples as applications.

References

- [1] Akaike H. (1973) "Information Theory as an extension of the maximum likelihood principle," second international symposium on information theory, Akademiai kiado, Budapest.
- [2] Akaike H. (1974) "A new look at the statistical model identification," IEEE Transactions on Automatic Control, 19(6), 716-723.
- [3] Allenby G.M., Leone R.P., Jen L. (1999) "A Dynamic model of purchase timing with application to direct marketing," Journal of the American Statistical Association, 94, 365-74.
- [4] Cover T.M., Thomas J.A., Elements of Information Theory, New York, Wiley, 1991.
- [5] Dadpay A., Soofi E.S., Soyer R. (2007) "Information Measures for Generalized Gamma Family," Journal of Econometrics, 138, 568-585.
- [6] Diebold F.X., Rudebusch G.D. (1990) "A Nonparametric investigation of duration dependence in the american business cycle," Journal of Political Economy, 98, 596-616.

- [7] Eckstein Z., Wolpin K.I. (1995) "Duration to first job and the return to schooling: estimates from a search-matching model," *Review of Economic Studies*, 62, 263-86.
- [8] Favero C.A., Pesaran M.H., Sharma S. (1994) "A duration model of irreversible oil investment: theory and empirical evidence," *Journal of Applied Econometrics*, 9, 95-112.
- [9] Hager H.W. Bain L.J. (1970) "Theory and methods inferential procedures for the generalized gamma distribution," *Journal of the American Statistical Association*, 65, 1601-1609.
- [10] Harter H.L. (1967) "Maximum-likelihood estimation of the parameters of a four-parameter generalized gamma population from complete and censored samples," *Technometrics*, 9, 159-165.
- [11] Hirose H., Maximum likelihood parameters estimation by model augmentation with application to the extended four-parameters Generalized Gamma distribution, Department of control Engineering and Science Kyushu Institute of technology, Fukuoka, 820-8502, Japan, 1999.
- [12] Hwang T., Huang P. (2006) "On new moment estimation of parameters of the Generalized Gamma distribution using it's characterization," *Taiwanese journal of Mathematics*, vol. 10, No. 4, 1083-1093.
- [13] Jaggia S. (1991) "Specification tests based on the heterogeneous generalized gamma model of duration: with an application to Kennan's strike data," *Journal of Applied Econometrics*, 6, 169-180.
- [14] Kapur, J. N., 1989, *Maximum entropy models in science and engineering* (Wiley, New York).
- [15] Kiefer N.M. (1984) "A simple test for heterogeneity in exponential models of duration," *Journal of Labor Economics* 2, 539-549.

- [16] Kullback S., Leibler R.A. (1951) "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22(1), 7986.
- [17] Kullback S., *Information theory and statistics*, John Wiley and Sons, 1959.
- [18] Kullback S., (1987) "Letter to the Editor: The Kullback-Leibler distance," *The American Statistician*, 41(4), 340341.
- [19] Lancaster T. (1979) "Econometric methods for the duration of unemployment," *Econometrica*, 47, 939-956.
- [20] Nadarajah S., Zografos K. (2003) "Formulas for Renyi information and related measures for univariate distributions," *Information Science*, 155, 119-138.
- [21] Prentice R.L. (1974) "A Log gamma model and its maximum likelihood estimation," *Biometrika*, 61, 539-544.
- [22] Shannon C.E. (1948) "A Mathematical theory of communication," *Bell System Technical Journal*, 27, 623-659.
- [23] Stacy E.W. (1962) "A generalization of the gamma distribution," *The Annals of Mathematical Statistics*, 33, 1187-1192.
- [24] Stacy E.W., Mihram G.A. (1965) "Parameter estimation for a generalized gamma distribution," *Technometrics*, 7, 349-358.
- [25] Stacy E.W. (1973) "Quasimaximum likelihood estimators for two-parameter gamma distribution," *IBM Journal Research and Development*, 17, 115-124.
- [26] Yamaguchi K. (1992) "Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employment" in Japan," *Journal of the American Statistical Association*, 87, 284-92.