

ON THE COMPARISON OF THE FISHER INFORMATION OF THE LOG-NORMAL AND GENERALIZED RAYLEIGH DISTRIBUTIONS

FAWZIAH S. ALSHUNNAR¹, MOHAMMAD Z. RAQAB¹ and DEBASIS KUNDU²

Abstract

Surles and Padgett (2001) recently considered two-parameter Burr Type X distribution by introducing a scale parameter and called it the generalized Rayleigh distribution. It is observed that the generalized Rayleigh and log-normal distributions have many common properties and both distributions can be used quite effectively to analyze skewed data set. In this paper we mainly compare the Fisher information matrices of the two distributions for complete and censored observations. Although, both distributions may provide similar data fit and quite similar in nature in many aspects, the corresponding Fisher information matrices can be quite different. We compute the total information measures of the two distributions for different parameter ranges and also compare the loss of information due to censoring. Real data analysis has been performed for illustrative purpose.

KEY WORDS AND PHRASES: Fisher information matrix; Burr Type X distribution; generalized Rayleigh distribution; log-normal distribution; left censoring; right censoring.

ADDRESS OF CORRESPONDENCE: Mohammad Z. Raqab

¹ Department of Mathematics, University of Jordan Amman 11942, JORDON.

² Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India.

1 INTRODUCTION

Recently, Surlles and Padgett [21] considered the two parameter Burr Type X distribution by introducing a scale parameter and correctly named it as the generalized Rayleigh (GR) distribution. The two-parameter GR distribution for $\alpha > 0$ and $\lambda > 0$ has the cumulative distribution function (CDF) and probability density function (PDF)

$$F_{GR}(x; \alpha, \lambda) = \left(1 - e^{-(\lambda x)^2}\right)^\alpha; \quad \text{for } x > 0, \quad (1)$$

and

$$f_{GR}(x; \alpha, \lambda) = 2\alpha\lambda^2 x e^{-(\lambda x)^2} \left(1 - e^{-(\lambda x)^2}\right)^{\alpha-1}; \quad \text{for } x > 0, \quad (2)$$

respectively. Here α and λ are the shape and scale parameters respectively. From now on the generalized Rayleigh distribution with the shape parameter α and scale parameter λ will be denoted by $GR(\alpha, \lambda)$. Several aspects of the GR distribution have been studied by Surlles and Padgett [21], Raqab and Kundu [19] and Kundu and Raqab [15]. For some general references on Burr Type X distribution, the readers are referred to Sartawi and Abu-Salih [20], Jaheen [9, 10], Ahmad *et al.* [2], Raqab [18] and the references cited there. It is observed that the GR distribution is always right skewed and they can be used quite effectively to analyze any skewed data. Shapes of the different PDFs of the GR distribution can be found in Raqab and Kundu [19].

Among several other right skewed distributions, two parameter log-normal distribution also is used quite effectively to analyze lifetime skewed data. In this paper it is assumed that the two-parameter log-normal distribution for $\beta > 0$, $\sigma > 0$ has the PDF

$$f_{LN}(x; \beta, \sigma) = \frac{1}{\sqrt{2\pi}x\sigma} e^{-\frac{(\ln x - \ln \beta)^2}{2\sigma^2}}; \quad x > 0. \quad (3)$$

Here σ and β are the shape and scale parameters respectively. The corresponding CDF can

be expressed as

$$F_{LN}(x; \sigma, \beta) = \Phi\left(\frac{\ln x - \ln \beta}{\sigma}\right) = \frac{1}{2} + \frac{1}{2} \text{Erf}\left(\frac{\ln x - \ln \beta}{\sqrt{2}\sigma}\right), \quad (4)$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution and

$$\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (5)$$

From now on a log-normal distribution with the PDF (3) will be denoted by $\text{LN}(\sigma, \beta)$. A log-normal distribution has always a unimodal PDF. The shapes of the different log-normal PDFs can be obtained in Johnson *et al.* [11].

The Fisher information matrix of a distribution function, plays a significant role in any statistical inference. Therefore, computation of the Fisher information matrix is quite important both from the theoretical and applications point of view. It can be used to compute the asymptotic variance of any function of the estimators and in turn it can be used to construct confidence intervals also. Recently it is observed by Kundu and Raqab [14] that for certain ranges of the shape and scale parameters the PDFs of the log-normal and the generalized Rayleigh distributions can be very close. For example, see Figures 1 and 2 of Kundu and Raqab [14], where it is observed that the the PDFs and CDFs of $\text{GR}(15, 1)$ and $\text{LN}(0.1822, 1.7620)$ are almost indistinguishable.

Computation of the Fisher information matrix for both complete and censored sample, is very important in the statistical literature. In this paper we compute and compare the Fisher information matrices of both distributions for (i) complete sample and (ii) censored sample (fixed time). We compare the total information measures and also the loss of information due to truncation. It is interesting to observe that although the PDFs of the log-normal and GE distributions can be quite close but the total information measures or the loss of information measures due to truncation can be quite different, which is quite counter intuitive.

The rest of the paper is organized as follows. We provide the necessary preliminaries in

Section 2. The results for complete sample and censored sample are provided in Sections 3 and 4. We analyze one data set in Section 5 and finally conclude the paper in Section 6.

2 PRELIMINARIES

Let $X > 0$ be a continuous random variable with the CDF, survival function and PDF as $F(x; \boldsymbol{\theta})$, $\bar{F}(x; \boldsymbol{\theta})$ and $f(x; \boldsymbol{\theta})$ respectively. Here $\boldsymbol{\theta}$ is a vector parameter and for brevity it is assumed that $\boldsymbol{\theta} = (\theta_1, \theta_2)$, although all the results in this section are valid for any finite dimensional vector. The hazard function and reversed hazard function of X will be denoted by

$$h(x; \boldsymbol{\theta}) = \frac{f(x; \boldsymbol{\theta})}{F(x; \boldsymbol{\theta})} = -\frac{d}{dx} \ln \bar{F}(x; \boldsymbol{\theta}) \quad \text{and} \quad r(x; \boldsymbol{\theta}) = \frac{f(x; \boldsymbol{\theta})}{\bar{F}(x; \boldsymbol{\theta})} = \frac{d}{dx} \ln F(x; \boldsymbol{\theta}) \quad (6)$$

respectively. Under the standard regularity conditions (see Lehmann [16]), the Fisher information matrix for the parameter vector $\boldsymbol{\theta}$ is

$$I(\boldsymbol{\theta}) = E \left(\begin{bmatrix} \frac{\partial}{\partial \theta_1} \ln f(X : \boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_2} \ln f(X : \boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \theta_1} \ln f(X : \boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} \ln f(X : \boldsymbol{\theta}) \end{bmatrix} \right). \quad (7)$$

Interestingly it may be observed that $I(\boldsymbol{\theta})$ can be expressed as

$$I(\boldsymbol{\theta}) = E \left(\begin{bmatrix} \frac{\partial}{\partial \theta_1} \ln h(X : \boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_2} \ln h(X : \boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \theta_1} \ln h(X : \boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} \ln h(X : \boldsymbol{\theta}) \end{bmatrix} \right) \quad (8)$$

or

$$I(\boldsymbol{\theta}) = E \left(\begin{bmatrix} \frac{\partial}{\partial \theta_1} \ln r(X : \boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_2} \ln r(X : \boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \theta_1} \ln r(X : \boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} \ln r(X : \boldsymbol{\theta}) \end{bmatrix} \right). \quad (9)$$

respectively. The derivations of (8) and (9) are quite simple and they can be obtained using the definition of the hazard function and reversed hazard function in terms of the density function, see for example Efron and Johnstone [5] and Gupta *et al.* [8]. It may be mentioned that sometimes it may be convenient to use (8) or (9) than (7).

Now consider the case when the observation X is both left and right censored at fixed time points T_1 and T_2 respectively, *i.e.* one observes Y , where

$$Y = \begin{cases} X & \text{if } T_1 < X < T_2 \\ T_1 & \text{if } X < T_1 \\ T_2 & \text{if } X > T_2. \end{cases}$$

In this case the Fisher information of one observation Y for the parameter vector $\boldsymbol{\theta}$ is

$$\begin{aligned} I_C(\boldsymbol{\theta}; T_1, T_2) &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \frac{1}{\bar{F}(T_2; \boldsymbol{\theta})} \begin{bmatrix} \frac{\partial}{\partial \theta_1} \bar{F}(T_2; \boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_2} \bar{F}(T_2; \boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \theta_1} \bar{F}(T_2; \boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} \bar{F}(T_2; \boldsymbol{\theta}) \end{bmatrix} \\ &\quad + \frac{1}{F(T_1; \boldsymbol{\theta})} \begin{bmatrix} \frac{\partial}{\partial \theta_1} F(T_1; \boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_2} F(T_1; \boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \theta_1} F(T_1; \boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} F(T_1; \boldsymbol{\theta}) \end{bmatrix} \\ &= I_M(\boldsymbol{\theta}; T_1, T_2) + I_R(\boldsymbol{\theta}; T_2) + I_L(\boldsymbol{\theta}; T_1) \quad (\text{say}) \end{aligned} \quad (10)$$

where

$$a_{ij} = \int_{T_1}^{T_2} \left(\frac{\partial}{\partial \theta_i} \ln f(x; \boldsymbol{\theta}) \right) \left(\frac{\partial}{\partial \theta_j} \ln f(x; \boldsymbol{\theta}) \right) f(x; \boldsymbol{\theta}) dx,$$

for $i, j = 1, 2$. Note that (10) can be obtained easily using the definition of the Fisher information of a random sample.

Therefore, the Fisher information for complete sample or for fixed right censored (at time T_2) sample or for fixed left censored (at time T_1) sample can be obtained as

$$I_M(\boldsymbol{\theta}; 0, \infty), \quad I_M(\boldsymbol{\theta}; 0, T_2) + I_R(\boldsymbol{\theta}; T_2) \quad \text{and} \quad I_M(\boldsymbol{\theta}; T_1, \infty) + I_L(\boldsymbol{\theta}; T_1)$$

respectively, by observing the fact $I_R(\boldsymbol{\theta}; \infty) = \mathbf{0}$ and $I_L(\boldsymbol{\theta}; 0) = \mathbf{0}$.

3 FISHER INFORMATION MATRICES: COMPLETE SAMPLE

Let us denote the Fisher information matrix of the GR distribution as

$$I_G(\alpha, \lambda) = \begin{bmatrix} f_{11G} & f_{12G} \\ f_{21G} & f_{22G} \end{bmatrix}. \quad (11)$$

Then

$$\begin{aligned} f_{11G} &= \frac{1}{\alpha^2}, \\ f_{12G} &= f_{21G} = \frac{2}{\alpha\lambda} \int_0^1 \left(1 + \frac{\ln y^{\frac{1}{\alpha}}}{\alpha}\right) \eta(y) dy, \\ f_{22G} &= \frac{4}{\lambda^2} \int_0^1 \eta^2(y) dy, \end{aligned}$$

where

$$\eta(y) = 1 + \ln(1 - y^{\frac{1}{\alpha}}) - \frac{(\alpha - 1)(1 - y^{\frac{1}{\alpha}}) \ln(1 - y^{\frac{1}{\alpha}})}{y^{\frac{1}{\alpha}}}. \quad (12)$$

It may be mentioned that f_{12G} and f_{22G} can be expressed in terms of Beta and digamma functions and they are presented in the Appendix.

If we denote the Fisher information matrix of the log-normal distribution as

$$I_L(\sigma, \beta) = \begin{bmatrix} f_{11L} & f_{12L} \\ f_{21L} & f_{22L} \end{bmatrix}, \quad (13)$$

then

$$f_{11L} = \frac{2}{\sigma^2}, \quad f_{12L} = f_{21L} = 0, \quad f_{22L} = \frac{1}{\beta^2 \sigma^2}.$$

Some of the interesting features are observed by comparing the Fisher information matrices of the two distributions. In both cases if the shape (scale) parameter is known, the Fisher information of the scale (shape) parameter is inversely proportional to itself. Moreover, the maximum likelihood estimators (MLEs) of the shape and scale parameters are asymptotically independent for the log-normal distribution, but for GR distribution they are dependent.

Now we would like to compare the total information measures contained in the corresponding Fisher information matrices. Since $\boldsymbol{\theta}$ is a vector parameter the comparison is not a trivial issue. Several methods can be adopted to compute the total information measure contained in a given Fisher information matrix. One of the measures can be the trace of the Fisher information matrix. It is similar to the E-optimality of the design of experiment problem. Note that the trace of Fisher information matrix is the sum of the Fisher information measures of the shape parameter (assuming scale parameter to be known) and the scale parameter (assuming shape parameter to be known). Another measure (inverse) can be the sum of the asymptotic variances of the MLEs of the shape and scale parameters, *i.e.* the trace of the inverse of the Fisher information matrix.

To compare the total information measures of the two distribution functions, it is quite natural to compare them at their *closest* values. The closeness (distance) between the two distribution functions can be defined in several ways. We have used the Kolmogorov-Smirnov distance as the distance measure between the two distribution functions. Kundu and Raqab [14] recently reported different values of $\tilde{\alpha}$, $\tilde{\lambda}$ for given σ so that $\text{GR}(\tilde{\alpha}, \tilde{\lambda})$ is closest to $\text{LN}(\sigma, 1)$ in terms of the Kolmogorov-Smirnov distance. We compare the trace and variance of the corresponding Fisher information matrices and the results are reported in Table 1. From Table 1 it is quite interesting to observe that even if the two distributions functions are quite close to each other, but the corresponding Fisher informations can be quite different.

One point should be mentioned here that although, the trace and the total variance have been used in the literature to represent the total information content in a random sample X regarding the parameter vector $\boldsymbol{\theta}$ but the trace or the total variance are not scale invariant. For example, if we make a scale change then not only the corresponding values but the trend also might change, (see for example Gupta and Kundu [7]), which may not be desirable. Similarly, if we re-parametrize, then also the total information measure with respect to one

Table 1: The traces and variances of the Fisher information matrices of $GR(\tilde{\alpha}, \tilde{\lambda})$ and $LN(\sigma, 1)$.

σ^2	$\tilde{\alpha}$	$\tilde{\lambda}$	Trace (GR)	Trace (LN)	Total Var (GR)	Total Var (LN)
0.10	3.285	1.255	6.904	30.419	30.043	0.150
0.15	2.065	1.068	6.746	20.031	10.030	0.225
0.20	1.518	0.933	7.037	15.014	4.916	0.300
0.25	1.207	0.828	7.571	12.000	2.938	0.375
0.30	1.003	0.742	8.281	10.026	1.961	0.449
0.35	0.861	0.671	9.126	8.618	1.417	0.522

set of parameters might be different than the other set. Moreover, although α, λ and σ, β are the shape and scale parameters of the GR and log-normal distributions respectively, but the parameters may not characterize the same features of the corresponding distributions. It is more reasonable to identify the same feature of both distributions. Instead of the individual shape and scale parameters, some function of the parameters may be more appropriate in this respect. For example the corresponding percentile points represent a common feature. Therefore, comparing the asymptotic variance of the corresponding percentile estimators is more meaningful, see for example Gupta and Kundu [7]. They are scale invariant and they maintain the trend also. It may be mentioned that similar measure has been considered to find the Fisher information content in a censoring scheme and hence to find the optimal censoring scheme, see for example Zhang and Meeker [23], Ng *et al.* [17], Kundu [12, 13] or Banerjee and Kundu [3].

The p -th ($0 < p < 1$) percentile points of the $GR(\alpha, \lambda)$ and $LN(\sigma, \beta)$ distributions are

$$p_{GE}(\alpha, \lambda) = \frac{1}{\lambda} \left[-\ln(1 - p^{\frac{1}{\alpha}}) \right]^{\frac{1}{2}} \quad \text{and} \quad p_{LN}(\sigma, \beta) = \beta e^{\sigma \Phi^{-1}(p)},$$

respectively. The asymptotic variances of the p -th percentile estimators of the GR and

log-normal distributions are

$$V_{GR}(p) = \begin{bmatrix} \frac{\partial p_{GR}}{\partial \alpha} & \frac{\partial p_{GR}}{\partial \lambda} \end{bmatrix} \begin{bmatrix} f_{11G} & f_{12G} \\ f_{21G} & f_{22G} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial p_{GR}}{\partial \alpha} \\ \frac{\partial p_{GR}}{\partial \lambda} \end{bmatrix}$$

and

$$V_{LN}(p) = \begin{bmatrix} \frac{\partial p_{LN}}{\sigma\beta} & \frac{\partial p_{LN}}{\sigma\beta} \end{bmatrix} \begin{bmatrix} f_{11L} & f_{12L} \\ f_{21L} & f_{22L} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial p_{LN}}{\sigma\beta} \\ \frac{\partial p_{LN}}{\sigma\beta} \end{bmatrix}$$

respectively. Now to compare the information measures of the two distributions the asymptotic variances of the median or 99% percentile estimators may be considered. Using the idea of Gupta and Kundu [7] we propose to compare

$$AV_{GR,W} = \int_0^1 V_{GR}(p)dW(p) \quad \text{and} \quad AV_{LN,W} = \int_0^1 V_{LN}(p)dW(p) \quad (14)$$

here $W(\cdot) \geq 0$ is a weight function such that

$$\int_0^1 dW(p) = 1.$$

Note the above criterion is a more general criterion than the criterion proposed by Gupta and Kundu [7]. The choice of $W(\cdot)$ depends on the problem at hand. For example if we are interested to compare the variances of the median estimators, then $W(\cdot)$ can be chosen as a point mass at 0.5. If we are interested about the average asymptotic variance of all the percentile point estimators, then the choice of $W(p) = 1$, for all $0 < p < 1$. Similarly, if we are interested about the central portion of the distribution or toward the tail, $W(\cdot)$ can be chosen accordingly. We have computed $AV_{GR,W}$ and $AV_{LN,W}$ for different $W(\cdot)$ and the results are reported in Table 2.

4 FISHER INFORMATION MATRIX: CENSORED SAMPLE

In this section first we provide the Fisher information matrix for the GR distribution and then for the log-normal distribution. Let us denote $I_M(\boldsymbol{\theta}; T_1, T_2)$, $I_R(\boldsymbol{\theta}; T_2)$ and $I_L(\boldsymbol{\theta}; T_1)$ for

Table 2: The asymptotic variances of the 5-th percentile estimators, median estimators, 95-th percentile estimators, the average asymptotic variances over all percentile estimators are reported for LN($\sigma, 1$) and GR($\tilde{\alpha}, \lambda$)

$\sigma^2 \rightarrow$	0.10	0.15	0.20	0.25	0.30	0.35
$V_{GR}(0.05)$	0.026	0.054	0.082	0.106	0.122	0.131
$V_{LN}(0.05)$	0.083	0.099	0.108	0.113	0.116	0.117
$V_{GR}(0.5)$	0.017	0.043	0.078	0.123	0.179	0.246
$V_{LN}(0.5)$	0.099	0.150	0.199	0.250	0.299	0.348
$V_{GR}(0.95)$	0.064	0.160	0.302	0.496	0.751	1.070
$V_{LN}(0.95)$	0.667	1.264	2.055	3.061	4.278	5.734
$\int_0^1 V_{GR}(p) dp$	0.028	0.067	0.121	0.190	0.276	0.379
$\int_0^1 V_{LN}(p) dp$	0.207	0.362	0.562	0.818	1.130	1.513

the GR distribution by

$$I_{MG}(\boldsymbol{\theta}; T_1, T_2) = \begin{bmatrix} a_{11G} & a_{12G} \\ a_{21G} & a_{22G} \end{bmatrix}, \quad I_{RG}(\boldsymbol{\theta}; T_2) = \begin{bmatrix} b_{11G} & b_{12G} \\ b_{21G} & b_{22G} \end{bmatrix}, \quad I_{LG}(\boldsymbol{\theta}; T_1) = \begin{bmatrix} c_{11G} & c_{12G} \\ c_{21G} & c_{22G} \end{bmatrix}$$

respectively. Note that for $p_1 = (1 - e^{-(\lambda T_1)^2})^\alpha$ and $p_2 = (1 - e^{-(\lambda T_2)^2})^\alpha$, we have

$$\begin{aligned} a_{11G} &= \frac{1}{\alpha^2} \int_{p_1}^{p_2} (1 + \ln y)^2 dy = \frac{1}{\alpha^2} [p_2(1 + (\ln p_2)^2) - p_1(1 + (\ln p_1)^2)] \\ a_{22G} &= \frac{4}{\lambda^2} \int_{p_1}^{p_2} \eta^2(y) dy \\ a_{12G} &= a_{21G} = \frac{2}{\alpha \lambda} \int_{p_1}^{p_2} \left(1 + \frac{\ln y^{1/\alpha}}{\alpha}\right) \eta(y) dy, \\ b_{11G} &= \frac{p_2^2 (\ln p_2)^2}{\alpha^2 (1 - p_2)} \\ b_{22G} &= \frac{4\alpha^2}{\lambda^2 (1 - p_2)} p_2^{\frac{2\alpha-2}{\alpha}} (1 - p_2^{1/\alpha})^2 (-\ln(1 - p_2^{1/\alpha}))^2 \\ b_{12G} &= b_{21G} = \frac{2}{\lambda(1 - p_2)} p_2^{\frac{2\alpha-1}{\alpha}} (\ln p_2) (1 - p_2^{1/\alpha}) (-\ln(1 - p_2^{1/\alpha})) \\ c_{11G} &= \frac{1}{\alpha^2} p_1 (\ln p_1)^2 \end{aligned}$$

$$\begin{aligned}
c_{22G} &= \frac{4\alpha^2}{\lambda^2} p_1^{\frac{2\alpha-2}{\alpha}} \left(1 - p_1^{1/\alpha}\right)^2 \left(-\ln(1 - p_1^{1/\alpha})\right)^2 \\
c_{12G} &= c_{21G} = \frac{2}{\lambda} p_1^{\frac{\alpha-1}{\alpha}} (\ln p_1) \left(1 - p_1^{1/\alpha}\right) \left(-\ln(1 - p_1^{1/\alpha})\right).
\end{aligned}$$

Similarly, let us denote the $I_M(\boldsymbol{\theta}; T_1, T_2)$, $I_R(\boldsymbol{\theta}; T_2)$ and $I_L(\boldsymbol{\theta}; T_1)$ matrices for the log-normal distribution by

$$I_{ML}(\boldsymbol{\theta}; T_1, T_2) = \begin{bmatrix} a_{11L} & a_{12L} \\ a_{21L} & a_{22L} \end{bmatrix}, \quad I_{RL}(\boldsymbol{\theta}; T_2) = \begin{bmatrix} b_{11L} & b_{12L} \\ b_{21L} & b_{22L} \end{bmatrix}, \quad I_{LL}(\boldsymbol{\theta}; T_1) = \begin{bmatrix} c_{11L} & c_{12L} \\ c_{21L} & c_{22L} \end{bmatrix}$$

respectively.

$$\text{If } F_{LN}(T_1; \sigma, \beta) = \Phi\left(\frac{\ln T_1 - \ln \beta}{\sigma}\right) = p_1 \text{ and } F_{LN}(T_2; \sigma, \beta) = \Phi\left(\frac{\ln T_2 - \ln \beta}{\sigma}\right) = p_2,$$

then

$$\begin{aligned}
a_{11L} &= \frac{1}{\sigma^2} \int_{\Phi^{-1}(p_1)}^{\Phi^{-1}(p_2)} (1 - y^2)^2 \phi(y) dy, \\
a_{22L} &= \frac{1}{\beta^2 \sigma^2} \int_{\Phi^{-1}(p_1)}^{\Phi^{-1}(p_2)} y^2 \phi(y) dy, \\
a_{12L} &= a_{21L} = \frac{1}{\beta \sigma^2} \int_{\Phi^{-1}(p_1)}^{\Phi^{-1}(p_2)} (y^2 - 1) y \phi(y) dy, \\
b_{11L} &= \frac{1}{(1 - p_2) \sigma^2} \left[\phi\left(\Phi^{-1}(p_2)\right) \right]^2 \left(\Phi^{-1}(p_2)\right)^2 \\
b_{22L} &= \frac{1}{\beta^2 (1 - p_2) \sigma^2} \left[\phi\left(\Phi^{-1}(p_2)\right) \right]^2 \\
b_{12L} &= b_{21L} = \frac{1}{\beta (1 - p_2) \sigma^2} \left[\phi\left(\Phi^{-1}(p_2)\right) \right]^2 \left(\Phi^{-1}(p_2)\right) \\
c_{11L} &= \frac{1}{p_1 \sigma^2} \left[\phi\left(\Phi^{-1}(p_1)\right) \right]^2 \left(\Phi^{-1}(p_1)\right)^2 \\
c_{22L} &= \frac{1}{\beta^2 p_1 \sigma^2} \left[\phi\left(\Phi^{-1}(p_1)\right) \right]^2 \\
c_{12L} &= c_{21L} = \frac{1}{\beta p_1 \sigma^2} \left[\phi\left(\Phi^{-1}(p_1)\right) \right]^2 \left(\Phi^{-1}(p_1)\right).
\end{aligned}$$

It may be observed that as $p_1 \downarrow 0$ and $p_2 \uparrow 1$, then

$$a_{11G} \rightarrow f_{11G}, \quad a_{12G} \rightarrow f_{12G}, \quad a_{22G} \rightarrow f_{22G}, \quad a_{11L} \rightarrow f_{11L}, \quad a_{12L} \rightarrow f_{12L}, \quad a_{22L} \rightarrow f_{22L}.$$

Some of the interesting points can be observed by comparing the two Fisher information matrices. First of all in both cases, it is observed that for fixed scale parameter, the Fisher

information of the shape parameter is inversely proportional to its square and both of them are independent of the corresponding scale parameters. In both the cases the Fisher information of the scale parameters are also inversely proportional to its square, but they depend on the corresponding shape parameters also.

Now we would like to see which portion of the distribution contain the maximum information of the corresponding shape and scale parameters. For the log-normal distribution from the shapes of $(1 - y^2)^2\phi(y)$ and $y^2\phi(y)$, it can be easily seen that the maximum information of the parameters are towards the ends. Therefore, if $p_1 = 1 - p_2$, then both the left and right censored data containing the same Fisher information about the shape and scale parameters for the log-normal distribution.

For GR distribution, from the behavior of $(1 + \ln y)^2$, it is clear that the Fisher information regarding the shape parameter is more on the left tail of the data than on the right tail. For the scale parameter, from the behavior of $\eta^2(y)$, it can be seen that the Fisher information of the scale parameter depends on the values of the shape parameter α . It is also observed that more Fisher information is towards the right tail than left tail for all values of the shape parameter.

For illustrative purposes, we have presented the total Fisher information of the log-normal and GR distributions for three different censoring schemes, namely Scheme 1: $p_1 = 0.0$, $p_2 = 0.7$, Scheme 2: $p_1 = 0.15$, $p_2 = 0.85$, Scheme 3: $p_1 = 0.30$, $p_2 = 1.0$. Note that Scheme 1, Scheme 2 and Scheme 3 represent left censored, interval censored and right censored respectively. The results are reported in Table 3.

Now we would like to discuss the loss of information due to truncation in one parameter, when the other parameter is known. In all these discussions, it is assumed that p_1 and p_2 are fixed. First let us consider the GR distribution. If the scale parameter is known, the loss

Table 3: The traces and variances of the Fisher information matrices of $\text{GR}(\tilde{\alpha}, \tilde{\lambda})$ and $\text{LN}(\sigma,1)$ for three different censoring schemes.

σ^2	$\tilde{\alpha}$	$\tilde{\lambda}$	Scheme	Trace (GR)	Trace (LN)	Total Var (GR)	Total Var (LN)
0.10	3.285	1.255	1	5.838	17.616	40.595	0.343
			2	6.382	6.231	50.829	0.702
			3	6.663	17.616	63.871	0.343
0.15	2.065	1.068	1	5.433	11.745	13.424	0.516
			2	6.093	4.154	16.296	1.052
			3	6.537	11.745	20.118	0.516
0.20	1.518	0.933	1	5.457	8.803	6.627	0.688
			2	6.215	3.114	7.815	1.404
			3	6.794	8.803	9.471	0.688
0.25	1.207	0.828	1	5.699	7.036	4.004	0.861
			2	6.553	2.489	4.592	1.757
			3	7.261	7.036	5.461	0.861
0.30	1.003	0.742	1	6.084	5.879	2.712	1.029
			2	7.036	2.079	3.026	2.103
			3	7.877	5.879	3.531	1.030
0.35	0.861	0.671	1	6.576	5.053	1.993	1.198
			2	7.628	1.787	2.166	2.446
			3	8.607	5.053	2.480	1.198

of information of the shape parameter is

$$Loss_G(\alpha) = 1 - \frac{a_{11G} + b_{11G} + c_{11G}}{f_{11G}} = 1 - \left[p_2 - p_1 + \frac{p_2(\ln p_2)^2}{1 - p_2} \right]. \quad (15)$$

Therefore, for fixed p_1 and p_2 , the loss of information is independent of α . It is interesting to observe that for Scheme 1, Scheme 2 and Scheme 3, the loss of information are 3.0%, 15% and 70% respectively. Therefore, it is clear that the initial portion of the data contains the maximum information of the shape parameter of the GR distribution and it is clear from the behavior of $(1 + \ln y)^2$ also.

For known shape parameter, the loss of information of the scale parameter is

$$\begin{aligned} Loss_G(\lambda) &= 1 - \frac{a_{22G} + b_{22G} + c_{22G}}{f_{22G}} \\ &= 1 - \left[\int_{p_1}^{p_2} \eta^2(y) dy + \frac{\alpha^2}{(1 - p_2)} p_2^{\frac{2\alpha-2}{\alpha}} (1 - p_2^{1/\alpha})^2 (-\ln(1 - p_2^{1/\alpha}))^2 \right. \\ &\quad \left. + \alpha^2 p_1^{\frac{2\alpha-2}{\alpha}} (1 - p_1^{1/\alpha})^2 (-\ln(1 - p_1^{1/\alpha}))^2 \right] / \int_0^1 \eta^2(y) dy. \end{aligned} \quad (16)$$

Here the function $\eta(\cdot)$ is defined in (12). In this case even for fixed p_1 and p_2 (16) depends on the shape parameter. It is clear that the loss of information of the scale parameter is quite different than that of the shape parameter.

Now let us discuss the loss of information of the shape or scale parameter, for log-normal distribution. The loss of information of the shape parameter is

$$\begin{aligned} Loss_L(\sigma) &= 1 - \frac{a_{11L} + b_{11L} + c_{11L}}{f_{11L}} \\ &= 1 - \frac{1}{2} \left[\int_{\Phi^{-1}(p_1)}^{\Phi^{-1}(p_2)} (1 - y^2)^2 \phi(y) dy + \frac{1}{(1 - p_2)} \left[\phi(\Phi^{-1}(p_2)) \right]^2 (\Phi^{-1}(p_2))^2 \right. \\ &\quad \left. + \frac{1}{p_1} \left[\phi(\Phi^{-1}(p_1)) \right]^2 (\Phi^{-1}(p_1))^2 \right], \end{aligned} \quad (17)$$

and for the scale parameter it is

$$Loss_L(\beta) = 1 - \frac{a_{22L} + b_{22L} + c_{22L}}{f_{22L}}$$

$$= 1 - \left[\int_{\Phi^{-1}(p_1)}^{\Phi^{-1}(p_2)} y^2 \phi(y) dy + \frac{1}{(1-p_2)} \left[\phi(\Phi^{-1}(p_2)) \right]^2 + \frac{1}{p_1} \left[\phi(\Phi^{-1}(p_1)) \right]^2 \right] \quad (18)$$

Both (17) and (18) depend only on p_1 and p_2 . It can be seen that for the shape parameter, the loss of information due to Scheme 1, Scheme 2 and Scheme 3 are 31%, 57% and 31% respectively, whereas for the scale parameter, the corresponding loss of information are 37%, 59% and 37% respectively. Therefore, although for Scheme 2, in both the cases loss of information are quite similar, but for Scheme 1 and Scheme 3, they can be quite different.

5 DATA ANALYSIS

In this section we analyze a real data set for illustrative purpose. The data set set was originally obtained from Bjerkedal [4]. The data set represents the survival times of guinea pigs injected with different doses of tubercle bacilli. It is known that guinea pigs have high susceptibility to human tuberculosis and that is why they were used in this study. Here, we are primarily concerned with the animals in the same cage that were under the same regimen. The regimen number is the common logarithm of the number of bacillary units in 0.5 ml. of challenge solution; i.e., regimen 6.6 corresponds to 4.4×10^6 bacillary units per 0.5 ml. ($\log 4.4 \times 10^6 = 6.6$). Corresponding to regimen 6.6, there were 72 observations listed below:

12, 15, 22, 24, 24, 32, 32, 33, 34, 38, 38, 43, 44, 48, 52, 53, 54, 54, 55, 56, 57, 58, 58, 59, 60, 60, 60, 61, 62, 63, 65, 65, 67, 68, 70, 70, 72, 73, 75, 76, 76, 81, 83, 84, 85, 87, 91, 95, 96, 98, 99, 109, 110, 121, 127, 129, 131, 143, 146, 146, 175, 175, 211, 233, 258, 258, 263, 297, 341, 341, 376.

The mean, standard deviation and the coefficient of skewness are calculated as 99.82, 80.55 and 1.80, respectively. The skewness measure indicates that the data are positively

skewed. From the observed data we try to obtain an estimate of the shape of the hazard function. A device called scaled TTT transform and its empirical version are relevant in this context. For a family with the survival function $S(y) = 1 - F(y)$, the scaled TTT transform, with $H_F^{-1}(u) = \int_0^{F^{-1}(u)} S(y)dy$ defined for $0 < u < 1$ is $\phi_F(u) = H_F^{-1}(u)/H_F^{-1}(1)$. The empirical version of the scaled TTT transform is given by

$$\phi_n(j/n) = H_n^{-1}(j/n)/H_n^{-1}(1) = \left(\sum_{i=1}^j x_{(i)} + (n-j)x_{(j)} \right) / \left(\sum_{i=1}^n x_{(i)} \right),$$

here $j = 1, \dots, n$ and $x_{(i)}$ for $i = 1, \dots, n$ represent the i th order statistics of a sample of size n . Aarset [1] showed that the scaled TTT transform is convex (concave) if the hazard rate is decreasing (increasing), and for bathtub (unimodal) hazard rates, the scaled TTT transform is first convex (concave) and then concave (convex). We have plotted the empirical version of the scaled TTT transform of the data set in Figure 1. From Figure 1 it is clear that

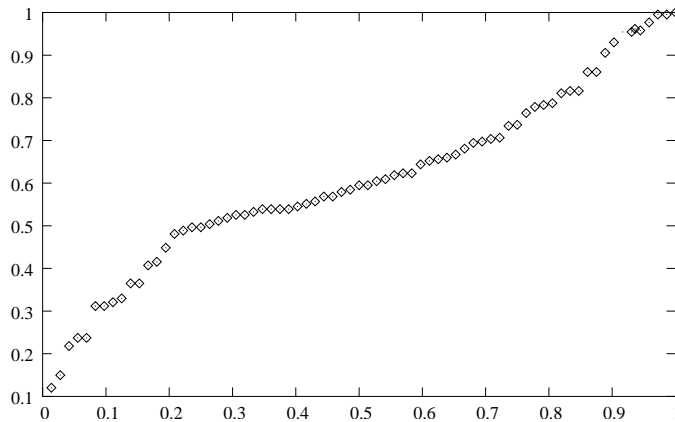


Figure 1: The plot of the scaled TTT transform of the data set.

the scaled TTT transform is first concave and then convex, therefore it indicates that the hazard function is unimodal. In this case, both the log-normal and GR can be used to analyze the data.

We standardize the data by dividing each element by the standard deviation of the data. It does not affect the shape parameters in both cases. We obtained the MLEs of the unknown

parameters as follows: $\hat{\alpha} = 0.9362$, $\hat{\lambda} = 0.2147$, $\hat{\sigma} = 0.6290$, $\hat{\beta} = 3.3210$. The difference of the log-likelihood function $l(GR) - L(LN) = 1.2403$. The Kolmogorov-Smirnov distance between the empirical distribution function and the fitted GR (log-normal) is 0.097 (0.110) and the corresponding p value is 0.51 (0.34). Therefore, it is clear that both distributions provide a very good fit to the given data set. The expected Fisher information matrices for the GR and log-normal distributions are

$$\mathbf{I}_{GR}(\hat{\alpha}, \hat{\lambda}) = \begin{bmatrix} 1.1399 & -6.1307 \\ -6.1307 & 81.8826 \end{bmatrix}, \quad \mathbf{I}_{LN}(\hat{\sigma}, \hat{\beta}) = \begin{bmatrix} 5.0551 & 0.0 \\ 0.0 & 0.2292 \end{bmatrix}.$$

Based on the above information matrices, it is clear that the Fisher information of the shape parameter of GR is less than that of the log-normal distribution whereas for the scale parameter it is the opposite. But if we compare the total information measure (trace or determinant), then GR parameters have more information than the log-normal parameters. Now let us look at the variance (inverse of the information) of the p -th percentile estimators for both distributions for different values of p . It is clear from Figure 2 that the informa-

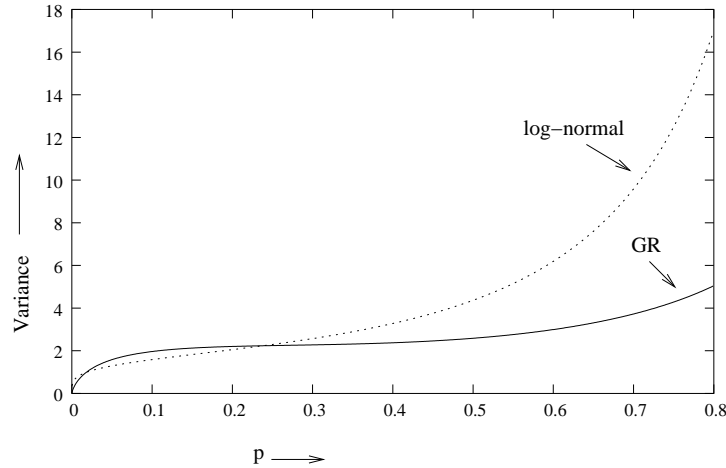


Figure 2: The variances of the p -th percentile estimators for log-normal and GR distribution for different values of p

tion content of the unknown parameters for GR distribution is more than the log-normal distribution.

Now let us look at the case when the data are censored. For illustrative purpose, we are assuming that $p_1 = 0.1$ and $p_2 = 0.9$. Based on the censored sample the expected Fisher information matrices for the GR and log-normal distributions are

$$\mathbf{I}_{GR}(\hat{\alpha}, \hat{\lambda}) = \begin{bmatrix} 1.0267 & -6.0836 \\ -6.0836 & 73.2409 \end{bmatrix}, \quad \mathbf{I}_{LN}(\hat{\sigma}, \hat{\beta}) = \begin{bmatrix} 2.7786 & 0.0039 \\ 0.0039 & 0.1763 \end{bmatrix}.$$

respectively. For comparison purposes we have plotted the asymptotic variances of the p -th percentile estimators for both the complete and censored samples in Figure 3. From Figure

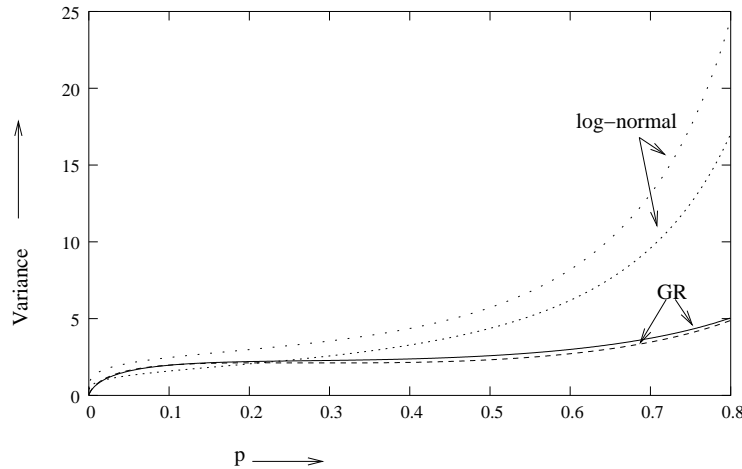


Figure 3: The variances of the p -th percentile estimators for log-normal, GR distribution for different values of p and for both complete and censored samples.

3 it is clear that the loss of information due to truncation for log-normal is much more than GR.

The main aim of this example is to point out that unless we have a very good knowledge about the underlying distribution, drawing inferences about the percentile points can be quite misleading. It is observed that even if two distributions fit the data very well (none of them can be rejected using any standard statistical tests) and the other characteristics (hazard function) also match reasonably well but the asymptotic variances of the percentile estimators based on two different distributions can be quite different. Therefore, in a situation like this some prior information may be used to choose the appropriate model. If no prior information

is available, we suggest to use the non-parametric method to draw inference of the tail behavior of the underlying distribution, rather than any parametric method. More work is needed in this direction.

6 CONCLUSIONS

In this paper we compute and compare the information measures of two closely related distributions, namely log-normal and generalized Rayleigh distributions. We compare their Fisher information measures for complete and truncated samples. It is observed that although the two distribution functions match very well for certain ranges of the parameter values and they can be almost indistinguishable, the total information measures or the loss of information of the two distributions can be very different, which is quite counter intuitive.

ACKNOWLEDGEMENTS:

The authors would like to thank two reviewers for their valuable comments and also the editor Prof. Robert G. Aykroyd for his encouragements.

APPENDIX

$$\begin{aligned}
f_{12G} &= f_{21G} = -4\alpha\lambda^3 \int_0^\infty x^3 e^{-2(\lambda x)^2} \left(1 - e^{-(\lambda x)^2}\right)^{\alpha-2} dx \\
&= \frac{2\alpha}{\lambda} B(2, \alpha - 1) (\psi(2) - \psi(\alpha + 1)) \\
f_{22G} &= \frac{2}{\lambda^2} \left\{ -1 + \psi(1) - \psi(\alpha + 1) - \alpha(\alpha - 1) \int_0^1 \frac{y \ln y}{(1 - y)^{2-\alpha}} \left(1 + 2\alpha \ln y + \frac{2y}{(1 - y)}\right) dy \right\} \\
&\hspace{15em} \text{if } 0 < \alpha \leq 1 \\
&= \frac{2}{\lambda^2} \left\{ -1 + \psi(1) - \psi(\alpha + 1) - 2\alpha \left([\psi(2) - \psi(\alpha + 1)]^2 + \psi'(2) - \psi'(\alpha + 1)\right) \right\}
\end{aligned}$$

$$\begin{aligned}
& \left. (\psi(2) - \psi(\alpha + 1)) + 2\alpha(\alpha - 1) \int_0^1 \frac{y^2(\ln y)^2}{(1 - y)^{3-\alpha}} dy \right\} \quad \text{if } 1 \leq \alpha < 2, \\
= & \frac{2}{\lambda^2} \left\{ -1 + \psi(1) - \psi(\alpha + 1) - 2\alpha \left([(\psi(2) - \psi(\alpha + 1))]^2 + \psi'(2) - \psi'(\alpha + 1) \right) \right. \\
& \left. + (\psi(2) - \psi(\alpha + 1)) - \frac{4}{\alpha - 2} \left([(\psi(3) - \psi(\alpha + 1))]^2 + \psi'(3) - \psi'(\alpha + 1) \right) \right\} \\
& \hspace{15em} \text{if } \alpha > 2.
\end{aligned}$$

Here $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, the Beta function, $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$ and $\psi'(x) = \frac{d}{dx} \psi(x)$ are the digamma and polygamma functions respectively.

References

- [1] Aarset, M. V. (1987), "How to identify a bathtub hazard rate", *IEEE Transactions on Reliability*, vol. 36, 106 - 108.
- [2] Ahmad, K.E., Fakhry, M.E. and Jaheen, Z.F. (1997), "Empirical Bayes estimation of $P(X < Y)$ and characterization of Burr Type X model", *Journal of Statistical Planning and Inference*, vol. 64, 297 - 308.
- [3] Banerjee, A. and Kundu, D. (2008), "Inference based on Type-II hybrid censored data from Weibull distribution", *IEEE Transactions on Reliability*, vol. 57, 369 - 378.
- [4] Bjerkedal, T. (1960), "Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli", *American Journal of Hygiene*, vol. 72, 130-148.
- [5] Efron, B. and Johnstone, I. (1990), "Fisher information in terms of the hazard function", *Annals of Statistics*, vol. 18, 38 - 62.
- [6] Gertsbakh, I. (1995), "On the Fisher information in type I censored and quantile response data", *Statistics and Probability Letters*, vol. 23, 297 - 306.

- [7] Gupta, R.D. and Kundu, D. (2006), "On the comparison of Fisher information of the Weibull and GE distributions", *Journal of Statistical Planning and Inference*, Vol. 136, 3130-3144.
- [8] Gupta, R.D., Gupta, R.C. and Sankaran, P.G. (2004), "Some characterization results based on factorization of the (reversed) hazard rate function", *Communications in Statistics - Theory and Methods*, vol. 33, 3009 - 3031.
- [9] Jaheen, Z.F. (1995), "Bayesian approach to prediction with outliers from the Burr Type X model", *Microelectronic Reliability*, vol. 35, 45 - 47.
- [10] Jaheen, Z.F. (1996), "Bayesian estimation of the reliability and failure rate functions of the Burr Type X model", *Journal of the Applied Statistical Sciences*, vol. 3, 281 - 288.
- [11] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995), *Continuous Univariate Distribution Vol. 1*, 2nd Ed., New York, Wiley.
- [12] Kundu, D (2007), "On hybrid censored Weibull distribution", *Journal of Statistical Planning and Inference*, vol. 137, 2127-2142, 2007.
- [13] Kundu, D (2008), "Bayesian inference and life testing plan for Weibull distribution in presence of progressive censoring", *Technometrics*, vol. 50, 144 - 154.
- [14] Kundu, D. and Raqab, M.Z. (2007), "Discriminating between the log-normal and generalized Rayleigh distribution", *Statistics*, vol. 41, 505 - 515. 49, 187 - 200.
- [15] Kundu, D. and Raqab, M.Z. (2005), "Generalized Rayleigh distribution; different methods of estimation", *Computational Statistics and Data Analysis*, vol. 49, 187 - 200.
- [16] Lehmann, E.L. (1991), *Theory of Point Estimation*, Wiley, New York.
- [17] Ng, H.K.T., Chan, P.S. and Balakrishnan, N. (2004), "Optimal progressive censoring plans for the Weibull distribution", *Technometrics*, vol. 46, 470 - 481.

- [18] Raqab, M.Z. (1998), "Order statistics from the Burr Type X mode", *Computer Mathematics and Applications*, vol. 36, 111 - 120.
- [19] Raqab, M.Z. and Kundu, D. (2006), "Burr type X distribution: Revisited", *Journal of Probability and Statistical Sciences*, vol. 4, 179 - 193.
- [20] Sartawi, H.A. and Abu-Salih, M.S. (1991), "Bayes prediction bounds for the Burr Type X model", *Communications in Statistics - Theory and Methods*, vol. 20, 2307 - 2330.
- [21] Surles, J.G. and Padgett, W.J. (2001), "Inference for reliability and stress-strength for a scaled Burr Type X distribution", *Lifetime Data Analysis*, vol. 7, 187-200.
- [22] Surles, J.G. and Padgett, W.J. (2005), "Some properties of a scaled Burr type X distribution", *Journal of Statistical Planning and Inference*, vol. 128, 271-280.
- [23] Zhang, Y. and Meeker, W. Q. (2005), "Bayesian life test planning for the Weibull distribution with given shape parameter", *Metrika*, vol. 61, 237 - 249.