Estimation of the Two Parameters for Generalized Rayleigh Distribution Function Using Simulation Technique

A. A.AL- Naqeeb, A. M. Hamed College of Health, of Medical Technology, Baghdad Mathematical Department College of Education, Ibn Al-Haitham, University of Baghdad

Abstract

In this paper, suggested formula as well a conventional method for estimating the twoparameters (shape and scale) of the Generalized Rayleigh Distribution was proposed. For different sample sizes (small, medium, and large) and assumed several contrasts for the two parameters a percentile estimator was been used. Mean Square Error was implemented as an indicator of performance and comparisons of the performance have been carried out through data analysis and computer simulation between the suggested formulas versus the studied formula according to the applied indicator. It was observed from the results that the suggested method which was performed for the first time (as far as we know), had highly advantage than the studied method, since the whole suggested outcomes of statistics in the suggested method are registered.

Introduction

During the last century, vast activities have been observed in generalizing the distributions. These distributions were formulated by statisticians, mathematicians, and engineers to mathematically model or represent certain behavior. Recently Surles and Padgett (2001) introduced two-parameter Burr Type X distribution, which can also be described as Generalized Rayleigh Distribution and It was observed that this particular skewed distribution can be used quite effectively in analyzing lifetime data [1]. Raqab and Kundu; (2003) considered this distribution and discussed its different properties and employed different methods of estimators. It was concluded that the two-parameter Generalized Rayleigh Distribution is a particular member of the generalized Weibull Distribution, originally proposed by Mudholkar and Srivastava [2]. Rayleigh distribution, which is a special case of Weibull distribution, is widely used to model events that occur in different fields such as medicine, social and natural sciences [3]. The Generalized Rayleigh with two parameters (shape and scale) (*GR*) distribution has the distribution function as following:-

Here α and λ , are the shape and scale parameters respectively, and the two-parameter GR distribution will be denoted by GR (α ; λ).

Therefore, GR distribution has the density function

$$f(t) = 2\alpha\lambda^{2}(1 - e^{-(\lambda t)^{2}})^{\alpha - 1}te^{-(\lambda t)^{2}}$$
 ------(2)

And the survival function $S(t; \alpha, \lambda) = 1 - (1 - e^{-(\lambda t)^2})^{\alpha}$

Where the hazard function is

$$h(t;\alpha,\lambda) = \frac{f(t;\alpha,\lambda)}{1 - F(t;\alpha,\lambda)} = \frac{2\alpha\lambda^2 e^{-(\lambda t)^2} (1 - e^{-(\lambda t)^2})^{\alpha - 1}}{1 - (1 - e^{-(\lambda t)^2})^{\alpha}}$$

It was observed that for $\alpha < 1/2$, the probability density function (p.d.f) of a *GR* distribution is a decreasing function and it is a right skewed unimodal function for $\alpha > 1/2$. And the hazard function of a GR distribution can be either bathtub type or an increasing function, depending on the shape parameter α . for $\alpha < 1/2$, the hazard functions of *GR* (α , λ) are bathtub type and for $\alpha > 1/2$, it has an increasing hazard function [2].

The main aim of this paper is to study how the estimator of the different unknown parameters behaves for different sample sizes and for different parameter values. We mainly compare percentiles estimator between the suggested formulas versus the studied methods according to the applied indicator by using extensive simulation techniques.

The rest of the paper is organized as follows, a briefly description of the percentile estimator methods (PCE) and their implementations including the conventional and the suggested methods. Followed by an empirical work, as well as the simulation results and conclusion.

Estimators Based on Percentiles

If the data come from a distribution function which has a closed form, then it is quite natural to estimate the unknown parameters by fitting straight line to the theoretical percentile points obtained from the distribution function and the sample percentile points. This method was originally proposed by Kao (1958, 1959) and it has been used quite successfully for Weibull distribution and for the generalized exponential distribution. In this paper, we apply the same technique for the GR distribution [2].

An estimator is statistic that specifies how to use the sample data to estimate an unknown parameter of the population [4]. In the following sections three estimation procedures are considered, the percentiles estimators, and compare their performances through numerical simulation for different sample sizes and for different parameters values. In percentile methods the Generalized Rayleigh Distribution has the explicit distribution function, therefore in this case the unknown parameters α and λ , can be estimated by equating the sample percentile points with the population percentile points and it is known as the percentile method [5].

Among the most easily obtained estimators of the parameters of the Weibull distribution are the graphical approximation to the best linear unbiased estimators. It can be obtained by fitting a straight line to the theoretical points obtained from the distribution function and the sample percentile points. In case of a *GR* distribution also it is possible to use the same concept to obtain the estimators of α and λ based on the percentiles, because of the structure of its distribution function, when both the parameters are unknown [6].

Since

$$F(x;\alpha,\lambda) = \left(1 - e^{-(\lambda x)^2}\right)^{\alpha}$$

Therefore

$$X^{2} = -\frac{1}{\lambda^{2}} \ln \left(1 - \left[F(x; \alpha, \lambda) \right]^{\frac{1}{\alpha}} \right)$$
(3)

If P_i denotes some estimate of $F(X_{(i)}; \alpha, \lambda)$, then the estimate of α and λ can be obtained by minimizing (2) as following:-

$$\sum_{i=1}^{n} \left[x_{(i)}^{2} + \lambda^{-2} \ln(1 - p_{i}^{(1/\alpha)}) \right]^{2} \qquad (4)$$

$$\sum_{i=1}^{n} 2 \left[x_{(i)}^{2} + \lambda^{-2} \ln(1 - p_{i}^{(1/\alpha)}) \right] \left[-2\lambda^{-3} \ln(1 - p_{i}^{(1/\alpha)}) \right]$$

Note that (4) is a non-linear function and it has to be minimized using some non-linear optimization technique. We call the corresponding estimators as the percentile estimators *PCE*'s. Several estimators of P_i can be used here and in this paper, we mainly consider $P_i = i/n+1$, which is the expected value of F(T(i)).

Where $P_i = \frac{i}{n+1}$ represent the studied formula And $E(F(t_i)) = \frac{i}{n+1}$ the expected value Which $E(t_i) = F^{-1}(\frac{i}{n+1})$

F(t) represents cumulative distribution function (*c.d.f*) for distribution, and $E(t_i)$ named (inverse probability of the cumulative sampling distribution) [7].

Then the suggested formula Pi will be [8] as follows:-

$$P_i = \frac{i - 0.5}{n + 0.5}$$

-Algorithms of the Suggested Methods

The cumulative distribution function of the Generalized Rayleigh distribution can be written in the form:-

 $F(t) = (1 - e^{-(\lambda t)^2})^{\alpha}$ ------(5)

Since the model (5) involves α and λ in a nonlinear way as shown in (4) so it can be transformed and taking its logarithms to the base *e* as follows:-

$$1 - e^{-(\lambda \hat{t})^{2}} = F(\hat{t})^{\frac{1}{\alpha}}$$

$$e^{-(\lambda \hat{t})^{2}} = 1 - F(\hat{t})^{\frac{1}{\alpha}}$$

$$-(\lambda \hat{t})^{2} = \ln(1 - F(\hat{t})^{\frac{1}{\alpha}})$$
Therefore
$$\hat{t} = -\frac{1}{\lambda}\sqrt{\ln(1 - F(\hat{t})^{\frac{1}{\alpha}})} \qquad (6)$$

Using uniform distribution and generating U where

IBN AL- HAITHAM J. FOR PURE & APPL. SCI. VOL.22 (4) 2009

 $U = \begin{cases} 1 & t \in [0,1] \\ 0 & \text{otherwise} \end{cases}$

Since U = 1 - U in-case of generating continues uniform random variable, then

$$F(t)^{1/\alpha} = 1 - F(t)^{1/\alpha}$$
 -----(7)

Taking the logarithm, for the both side of equation [5], then the following equation will produce:

$$\ln(\hat{t}) = \ln(\frac{1}{\sqrt{\alpha}\lambda}) + \frac{1}{2}\ln\left[-\ln F(\hat{t})\right] \quad ----- \quad (8)$$

Thus the equation (8) is intrinsically linear form, in which

$$y_i = \ln(\hat{t}_i)$$
, $x_i = \frac{1}{2}\ln\left[-\ln F(\hat{t})\right]$ and $\hat{\beta}_0 = \ln(\frac{1}{\sqrt{\alpha\lambda}})$

In equation (8), the slope is a constant and equals to 1, which indicates that $\Delta x = \Delta y$, using simple linear regression equation, then

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$
 Where $\beta_1 = 1$

Employing the initial value of $\alpha \& \lambda$ in the right side of (8) with substitution of the generating uniform values in $F(t) = u_i$, to obtain the left side

 $\hat{t} = \exp(\ln(\hat{t}))$

And if error is added to this model, then

$$t_i = t_i + e_i$$

Since that E(e) = 0, where $e \sim exp(1)$, so the errors are independent and uncorrelated [9].In order to make a comparison between the two methods (suggested and studied methods) for PCE the same procedure in finding equation (8) will be repeated twice time and the

equation of the straight line will produce. Through out solving them the value of $\hat{\alpha} \& \hat{\lambda}$ will be founded.

Hence

$$\ln(\hat{t}_{i}) + e_{i} = \ln(t_{i}) = y_{i} \&x_{i} = \frac{1}{2}\ln[-\ln(u_{i})]$$

And the estimator $\hat{\beta}_0 = \ln\left(\frac{1}{\sqrt{\alpha\lambda}}\right)$ represents the suggested method which is obtained for the first time (as far as we know).

-Percentile Estimator (studied)

According to equation (8) we have

$$t_1' = -\ln\sqrt{\alpha} - \ln\lambda + \frac{1}{2}\ln\left[-\ln\left(\frac{i}{n+1}\right)\right]$$
$$t_2' = -\ln\sqrt{\alpha} - \ln\lambda + \frac{1}{2}\ln\left[-\ln\left(\frac{i}{n+1}\right)\right]$$

$$t_1' + t_2' = -2 \ln \sqrt{\alpha} - 2 \ln \lambda + \ln \left[-\ln \left(\frac{i}{n+1} \right) \right]$$

IBN AL- HAITHAM J. FOR PURE & APPL, SCI. VO

L.22 (4) 2009

And the same method is used to estimate $\hat{\lambda}(\alpha)$ and obtain:

$$\hat{\lambda}(\alpha) = \exp\left\{\frac{1}{2}\ln\left[-\ln\left(\frac{i}{n+1}\right)\right] - \ln\sqrt{\alpha} - \frac{1}{2}(t_1 + t_2)\right\}$$
(10)

-Percentile Estimator (suggested)

 $\hat{\alpha}(\lambda) \& \hat{\lambda}(\alpha)$ in suggested formula of P_i will be as follows:-

$$t_1' = -\ln\sqrt{\alpha} - \ln\lambda + \frac{1}{2}\ln\left[-\ln\left(\frac{i-0.5}{n+0.5}\right)\right]$$
$$t_2' = -\ln\sqrt{\alpha} - \ln\lambda + \frac{1}{2}\ln\left[-\ln\left(\frac{i-0.5}{n+0.5}\right)\right]$$

$$t_{1}'+t_{2}' = -2\ln\sqrt{\alpha} - 2\ln\lambda + \ln\left[-\ln\left(\frac{i-0.5}{n+0.5}\right)\right]$$

$$\ln\sqrt{\alpha} = \ln\left[-\ln\left(\frac{i-0.5}{n+0.5}\right)\right] - 2\ln\lambda - (t_{1}'+t_{2}')$$

$$\hat{\alpha}(\lambda) = \left[\exp\left\{\ln\left[-\ln\left(\frac{i-0.5}{n+0.5}\right)\right] - 2\ln\lambda - (t_{1}+t_{2})\right\}\right]^{\frac{1}{2}}$$
(11)
Where:

Therefore, it can be concluded that the same procedure can be applied for obtaining the estimators ($\alpha \& \lambda$) of GR distribution to other estimation methods such as (Ordinary Least Square (OLS), Maximum Likelihood (MLH)...) as in example in case of OLS the obtaining equations will be flowing:

According to equation (8) we have

$$t_1' = -\ln\sqrt{\alpha} - \ln\lambda + \frac{1}{2}\ln\left[-\ln\left(1 - \frac{i}{n+1}\right)\right]$$

The same method is used to estimate $\hat{\alpha}(\lambda) \& \hat{\lambda}(\alpha)$, hence

$$\hat{\alpha}(\lambda) = \left[\exp\left\{ \ln\left[-\ln\left(1 - \frac{i}{n+1}\right) \right] - 2\ln\lambda - (t_1 + t_2) \right\} \right]^{\frac{1}{2}}$$
$$\hat{\lambda}(\alpha) = \exp\left\{ \frac{1}{2}\ln\left[-\ln\left(1 - \frac{i}{n+1}\right) \right] - \ln\sqrt{\alpha} - \frac{1}{2}(t_1 + t_2) \right\}$$

And $\hat{\alpha}(\lambda) \& \hat{\lambda}(\alpha)$ in suggested formula of P_i will be

$$\hat{\alpha}(\lambda) = \left[\exp\left\{ \ln\left[-\ln\left(1 - \frac{i - 0.5}{n + 0.5}\right) \right] - 2\ln\lambda - (t_1 + t_2) \right\} \right]^{\lambda}$$

$$\hat{\lambda}(\alpha) = \exp\left\{\frac{1}{2}\ln\left[-\ln\left(1-\frac{i-0.5}{n+0.5}\right)\right] - \ln\sqrt{\alpha} - \frac{1}{2}(t_1+t_2)\right\}$$

Empirical work

One of the most important applications of computer science is Computer simulation. It is an attempt to model a real-life on a computer so that it can be studied to see how the system works. By changing variables, predictions may be made about the behavior of the system Computer simulation has become a useful part of modeling many systems in economics, finance, and several applications [10]. Simulation approaches offer great opportunities for working out probabilities, confidence intervals and similar concepts [11]. This analysis may be done, sometimes, through analytical or numerical methods, but the model may be too complex to be dealt with. Essentially, simulation process consists of building a computer model that describes the behavior of a system and experimenting with this computer model to reach conclusions that support decisions [12]. Sometimes, it is not feasible or possible, to build a prototype, yet we may obtain a mathematical model describing, through equations and constraints, the essential behavior of the system. In such extreme cases, we may use simulation to replicate real world studies that cannot be done, simulation exercises may encounter statistical pitfalls that degrade their performance, or fail to take advantage of the opportunities statistics can provide for controlling simulation error and producing statistically reliable results [12]. In order to make comparison of the two methods of PCE for the parameters of the Generalized Rayleigh Distribution which were studied in the previous chapters to reach into the best estimated method of the shape parameter and scale parameter we make a simulation prototype and provide assumption of many cases which it can be existed in real world and use the basic step process in any simulation experiment once we have estimated the corresponding simulation model.

Algorithms steps

First step: - specified the assumed values by choosing different sample size of Generalized Rayleigh distribution, such as small sample size (n=20) and medium sample size (n=50) and large sample size (n=100). And then choosing the values of assumption parameters (α, λ) in each several contrasts and choosing for the initial values of the two parameters (shape and scale) parameters as shown in.

Assumed contrast parameters [Introduced by the researcher].

α	λ
1	1
1	2
2	1
2	2

Second step:- Generation of data which include :-

-Generated the random data which was taken from the uniform distribution in the interval (0, 1) using Excel, and SPSS, software computer package.

- The generation of errors for all data and in methods the random errors have been generated using the standard exponential distribution instead of normal distribution which has been used in conventional methods introduced by Gupta and Kundu.

Third step:- This step contains the following-

-Using the same value of t for three methods and applying the equation $\hat{t} = -\frac{1}{\lambda}\sqrt{\ln(1-U^{\prime/\alpha})}$ as mentioned in [5].

-finding the time event (t) by using the equation $t_i = t_i + e_i$ Where $i=1, \dots, n$

-The values of $\hat{\alpha} \& \hat{\lambda}$ of the generalized Rayleigh distribution can be determined according to the estimator of each method using the equations (9), (10), (11), and (12) respectively.

Fourth step:- smoothing the obtained values

-In this step the iteration of data will be repeated 100 times to generate a new different error, so we obtain 100 value of $\hat{\alpha}$, and 100 value of $\hat{\lambda}$ for each contrast. Then the mean of each

case will be calculated to find the estimated α and λ .

<u>Fifth</u> step:- In this step the following comparison indicator will be employed to make a comparison between different methods by M ean Square Error *(MSE)*.

Results and Conclusion

As a consequence for practical work and taking the mean square error as the indicator of preference between the different estimator methods, the following results are obtained:-

1-For the conventional methods and for different sample sizes the following results are obtained :-

(i)S mall sample size (n=20)

For the assumed contrast parameters (2, 1), (1, 2), (2, 2) and (1, 1) the *PCE* estimator method was given the best results.

(ii)Medium sample size (n=50).

For the assumed contrast parameters (2, 2), (2, 1), (1, 2), and (1, 1) the *PCE* estimator method was given the best result.

(iii)Large sample size (n=100).

For the assumed contrast parameters (1, 1), (2, 2), (1, 2), and (2, 1) the *PCE* estimator method was given the best result.

2-For the suggested method the following results are obtained:-

(i)S mall sample size (n=20)

For the assumed contrast parameters (1, 1), (2, 2), (1, 2), and (2, 1), the *PCE* estimator method was given the best result.

(ii)Medium sample size (n=50)

For the assumed contrast parameters (2, 1), (1, 1), (1, 2) and (2, 2) the *PCE* estimator method was given the best results.

(iii)Large sample size.(n=100)

For the assumed contrast parameters (2, 1), (2, 2), (1, 1), and (1, 2) the *PCE* estimator method was given the best results.

3-The comparison between the studied and suggested methods are summarized as follows :-

(i) S mall sample size (n=20)

•For the assumed contrast parameters (1, 1) and (2, 2) the *PCE* (suggested) estimator method was given the best results in studied.

•For the assumed contrast parameters (2, 1), and (1, 2) the *PCE* (studied) estimator method was given the best results in suggested method.

(iii)Medium sample size (n=50)

•For the assumed contrast parameters (2, 2), (1, 1), and (2, 1) the *PCE* (studied) estimator method was given the best results in suggested method.

•For the assumed contrast parameters (1, 2), the PCE (suggested) estimator method was given the best results in studied method.

(iv)Large sample size.(n=100)

For the assumed contrast parameters (2, 2), (1, 1), (1, 2), and (2, 1) the (suggested) *PCE* estimator method was given the best results in studied methods.

4-It can be mentioned that when the sample size increased the mean square error decreased.

5-It can be noticed that when the assumed values are equal, the values of $\hat{\alpha}$ and $\hat{\lambda}$ are equal also, and that consistent with the general mathematical concept.

6-In this work the white nose error is generated in Rayleigh Distribution and it was followed the distribution and gave best results.

The results of simulation (estimation of scale and shape parameters) of the generalized Rayleigh distribution for studied and suggested methods for different sample size (n=20, 50, and 100) are listed in the table (1)

References

- 1. Surles, J. G. and Padgett W. J., (2005), "Some properties of a scaled Burr type X distribution", Journal of Statistical Planning and Inference, <u>128</u>: 271-280.
- 2. D.Kundu, D. and Raqab M., (2005)," Generalized Rayleigh Distribution: Different Methods of Estimations", Computational Statistics & Data Analysis, <u>49</u>: 18.
- 3. Mahdil, S. and Cenac M., (2006), "Estimating and Assessing the Parameters of the Logistic and Rayleigh Distributions from Three Methods of Estimation", Journal of Mathematical Computer Science, <u>13</u>: 25-34.

- 4. Mcclave, T., 1990,"Probability and Statistics for Engineers", 3rd Edition, Pws-Kent publishing company, Boston.
- Gupta, R. D. and Kundu D., (2007), "Generalized Exponential Distribution: Existing results and some recent developments", Journal of Statistical Planning and Inference, <u>137(11)</u>: 3537-3547.
- 6. Gupta, R. D. and Kundu D., (2001), "Generalized Exponential Distribution: Different methods of estimation ", Journal of Statistical Computations and Simulations, <u>69(4)</u>: 315-338.
- 7. Al –Naqeeb, A., (1990)," Estimation of the Residual Variance in Linear Models within Cases Normal Dist. And Non-Normal dist.", College of Administration & Economics-University of Al-Mustansiriyah.
- 8. Al –Naqeeb, A., (2007)," Estimation some non-Linear Type by Using Min Ch square method for Binary Data by simulation method", Research Journal of Aleppo University, Basic Science Series.
- Al –Naqeeb, A., 1993,"Measures The Coincidence Power of the Two Random Variables (Dependant And Residual) In The Linear Regression Models By Using The Probability Plot Method" ,5 th international conference of statistical sciences, Baghdad, Iraq.
- 10. "Simulation ", (2009), From Wikipedia, the free encyclopedia, Wikimedia Foundation, Inc., All text is available under the terms of the GNU Free Documentation License
- 11. Michael Wood, (2005),"The Role of Simulation Approaches in Statistics", University of Portsmouth, U.K., Journal of Statistics Education, <u>13(3)</u>.
- 12. Insua, D. R. and etal, (2005), "Simulation in Industrial Statistics ", Statistical and Applied Mathematical Sciences Institute, Technical Report, Research Triangle Park, NC 27709-4006, PO Box 14006, available at www.samsi.info.

IBN AL- HAITHAM J. FOR PURE & APPL. SCI. VOL.22 (4) 2009

		Assumed Parameter		Estimators		Indicator
Sample	Method	α	λ	$\hat{\alpha}$	$\hat{\lambda}$	M SE
20 PCE studied PCE suggested	-	1	1	0.130645	0.130645	3.10325
		1	2	0.205795	0.346934	2.446409
		2	1	0.260983	0.181641	1.614332
		2	2	0.125566	0.177578	2.594462
		1	1	0.13282	0.13282	1.456683
		1	2	0.184391	0.359877	2.504632
	suggested	2	1	0.221384	0.158	2.702558
		2	2	0.127188	0.179871	2.051962
50 Stud	PCE studied	1	1	0.168913	0.168913	2.667632
		1	2	0.141036	0.283179	2.082187
		2	1	0.209621	0.146927	1.617342
		2	2	0.149932	0.212036	1.308038
	PCE suggested	1	1	0.1721	0.1721	2.473099
		1	2	0.144298	0.431658	2.478395
	suggested	2	1	0.210311	0.149445	1.817146
		2	2	0.153072	0.2164759	3.208128
100	PCE studied	1	1	0.194146	0.194146	1.593123
		1	2	0.118901	0.237803	2.388022
		2	1	0.217849	0.152913	2.73559
		2	2	0.1536	0.2172234	2.011512
	PCE suggested	1	1	0.196815	0.196815	1.753823
		1	2	0.119837	0.240805	2.22064
		2	1	0.217955	0.154118	1.214269
		2	2	0.155098	0.2193424	1.385011

Tables (1) Estimation of S cale and Shape Parameters of GED For suggested Table (4-8) The Results of Simulation methods

تقدير معلمتي توزيع رالي العام بأستخدام تقنية المحاكاة

عبد الخالق النقيب ، الأء ماجد حمد الكلية التقنية الطبية ، بغداد قسم الرياضيات ، كلية التربية ، ابن الهيثم ، جامعة بغداد

الخلاصة

في هذا البحث اقترحت صيغة جديدة فضلاً عن الصيغة النقليدية لغرض تقدير معلمتي الشكل و القياس لتوزيع رالي العام ولأحجام وعينات مختلفة (صغيرة ، متوسطة ، كبيرة) مع توليفات عديدة افتراضية للمعلمتين بأستخدام طريقة (PCE) . و استخدم مؤشر معدل مربعات الخطأ مؤشراً لأفضل اداء ومقارنة بين الصيغة النقليدية والمقترحة من خلال تحليل القيم والمحاكاة الحاسوبية. لقد تمت الملاحظة من خلال النتائج التي تم الحصول عليها بأفضلية الصيغة المقترحة (وتم استخدامها لأول مرة على حديدة المتولية.