"Science Stays True Here"
Journal of Mathematics and Statistical Science, Volume 2016, 490-497 | Science Signpost Publishing

Comparison of Parameter Estimation in the Exponentiated Gumbel Distribution based on Ranked Set Sampling and Simple Random Sampling

Hossein Jabbari Khamnei[1], Sanaz Ravandeh Mayan[2]

1. Department of Statistics, Faculty of Mathematical Science, University of Tabriz, Tabriz, Iran.

2. Department of Statistics, Faculty of Mathematical Science, Alborz Insittution of Higher Education, Qazvin, Iran.

Abstract

Various parametric families of distribution in lifetime data analysis and failure models are used. In this condition, the gamma distribution, Weibull and log-normal named because of the shape and scale parameters flexible high plasticity analysis of different types of data, especially lifetime data are skew. Recently, a new distributed Exponential Gumbel distribution by Mudholkarand Srivastava (1993) and Nadarajah (2006) presented and analyzed the characteristics to estimate the parameters of the distributions. In this paper, we estimate the parameters of Gumbel distribution based on Simple Random Sampling, and Ranked Set Sampling, also we will compare these two methods.

Keywords: Exponentiated Gumbel distribution, Order Statistics, Simple Random Sampling, Ranked Set Sampling, Maximum Likelihood Estimation.

Introduction

Ranked set Sampling ($\boldsymbol{RSS}$) method was used to estimate the population mean by McIntyre (1952) for the first time in 1952. Takahasi and Wakimoto (1968) studied unbiased estimators for the population mean based on $\boldsymbol{RSS}$ in 1968. The theories and applications of $\boldsymbol{RSS}$ has expressed by authors such as Dell and Clutter (1972). Stokes (1980) estimated the population variance using of the data from $\boldsymbol{RSS}$ method. In later years, $\boldsymbol{RSS}$ was used to estimate the involved parameter in the statistical distribution. Chuvi et al (1994) studied the estimator of the location parameter of the Cauchy distribution, and Lam et al (1994) estimated the parameters of two-parameter exponential distribution with the using of RSS.

Lam et al (1996) studied the estimators for locaction and scale parameters of the logistic distribution, Bhoj and Ahsanullah (1996) estimated the parameters of the Geometric distribution based on RSS. Adatia (2000) achieved estimating the parameters of the Half logistic distribution, and Abu-Dayyeh (2004) studied the estimators of parameters of the two-parameter logistic distribution with data of $\boldsymbol{RSS}$. The method of RSS was used for estimating the parameters of Bivariate Exponential distribution by Chacko and Yageen Thomas (2007) and for the parameters of Bivariat Exponential distribution by AL-Saleh and Diab (2009) and also for the parameter of Wibull Modified distribution by AL- Hadhrami (2010).

In this paper, we will introduce the Exponentiated Gumbel Distribution, and study the maximum likelihood estimate $(\boldsymbol{MLE})$ of the parameters of Exponentiated Gumbel Distribution based on a Simple Random Sample $(\boldsymbol{SRS})$. Then describe the $\boldsymbol{RSS}$, and we will study the estimation of the parameters based on the $\boldsymbol{RSS}$, and finally compare these two sampling methods.

Definition

If $\boldsymbol{X}$ is a random variable with the following cumulative distribution function $(\boldsymbol{cdf})$

$$\boldsymbol{F}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\alpha}) = [\boldsymbol{G}(\boldsymbol{x},\boldsymbol{\lambda})]^{\boldsymbol{\alpha}} = [\boldsymbol{exp}(-\,\boldsymbol{exp}(-\boldsymbol{\lambda x}))]^{\boldsymbol{\alpha}} \quad \boldsymbol{\alpha} > 0 \;, -\infty < x < \infty \;, \lambda > 0$$

then $\boldsymbol{X}$ has the Exponentiated Gumbel Distribution with scale parameter $\boldsymbol{\lambda}$, and shape parameter $\boldsymbol{\alpha}$. Its density function $(\boldsymbol{df})$ is as follows

$$\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\alpha}) = \boldsymbol{\lambda\alpha}[\boldsymbol{exp}\{-\boldsymbol{exp}(-\boldsymbol{\lambda x})\}]^{\boldsymbol{\alpha}}\boldsymbol{exp}(-\boldsymbol{\lambda x}).$$

We need optimal estimators for the parameters of this distribution. In the beginning, we estimate the parameters of the distribution with the maximum likelihood method based on a $\boldsymbol{SRS}$ of size$\boldsymbol{n}$.

Maximum Likelihood Estimators

In this section we consider the maximum likelihood estimators $(\boldsymbol{MLE's})$ of $\boldsymbol{GE}(\boldsymbol{\alpha},\boldsymbol{\lambda})$. First we consider the case when both $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$, are unknown. Let $\boldsymbol{X_1},\boldsymbol{X_2},\dots,\boldsymbol{X_n}$ be a random sample of size $\boldsymbol{n}$ from $\boldsymbol{GE}(\boldsymbol{\alpha},\boldsymbol{\lambda})$. Then the log-likelihood function $\boldsymbol{L}(\boldsymbol{\alpha},\boldsymbol{\lambda})$ can be written as

$$\boldsymbol{l}(\boldsymbol{\alpha},\boldsymbol{\lambda}|\boldsymbol{x}) = (\boldsymbol{\alpha\lambda})^{\boldsymbol{n}}\,\boldsymbol{exp}\left(-\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\boldsymbol{\lambda}\,\boldsymbol{x_i}\right)\boldsymbol{exp} - \left(\boldsymbol{exp}(-\boldsymbol{\lambda}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\boldsymbol{x_i})\right)^{\boldsymbol{\alpha}}$$

$$L(\alpha,\lambda) = lnl(\alpha,\lambda|x) = nln(\alpha) + nln(\lambda) - \lambda\sum_{i=1}^{n} xi - \alpha\sum_{i=1}^{n} e^{-\lambda xi} \tag{1}$$

The normal equations become

$$\frac{\partial lnl}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} e^{-\lambda xi} = 0 \tag{2}$$

$$\frac{\partial lnl}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i + \alpha\sum_{i=1}^{n} x_i e^{-\lambda xi} = 0 \tag{3}$$

From (2), we obtain the MLE of α as a function of λ, say $\hat{\alpha}(\lambda)$, as

$$\hat{\alpha}(\lambda) = n\left(\sum_{i=1}^{n} e^{-\lambda xi}\right)^{-1} \tag{4}$$

Substituting $\hat{\alpha}(\lambda)$ in (1), we obtain the profile log-likelihood of λ as

$$g(\lambda) = L(\hat{\alpha}(\lambda),\lambda) = c - nln\left(\sum_{i=1}^{n} e^{-\lambda xi}\right) + nln\lambda - \lambda\sum_{i=1}^{n} x_i = 0 \tag{5}$$

Therefore, the MLE of λ, say $\hat{\lambda}_{MLE}$, can be obtained by maximizing (5) with respect to λ. It can be shown that the maximum of (5) can be obtained as a fixed point solution $v(\lambda) = \lambda$,

$$v(\lambda) = \left[\frac{\sum_{i=1}^{n} x_i\,(1 - n\lambda)}{n}\right]^{-1} \tag{6}$$

Now consider the MLE of α, when the scale parameter λ, is known. Without loss of generality, we can assume that $\lambda = 1$. If λ is known, the MLE of α say $\hat{\alpha}_{MLE}$ is

$$\hat{\alpha}_{ML} = n\left(\sum_{i=1}^{n} e^{-\lambda xi}\right)^{-1} \tag{7}$$

Now note that, if $X_i{'}\mathbf{s}$ are independent and identically distributed $GE(\alpha, 1)$, then $-2\alpha\sum_{i=1}^{n} e^{xi}$ follows gamma random variable with shape parameter n and scale parameter $\frac{1}{2}$. Therefore, for $n > 2$

$$E(\hat{\alpha}_{ML}) = \frac{n}{n-1}\alpha, \quad var(\hat{\alpha}_{ML}) = \frac{n^2}{(n-1)^2(n-2)}\alpha^2$$

Using (7), an unbiased estimate of α_u can be easily obtained as

$$\hat{\alpha}_u = \frac{n-1}{n}\hat{\alpha}_{ML} = \frac{n-1}{n}\frac{n}{\sum_{i=1}^{n} e^{x_i}}.$$

Ranked Set Sampling

In environmental and laboratory studies, small number of units, are selected, and are measured, and are assessed chemical variables that have a negative impact on the environment.

Since the measurement of the chemical units are expensive, using SRS is associated with a heavy expenditure. So, it is useful providing a reasonable sampling strategy with fewer units of measurement. In the condition that measurement of the units is difficult or costly, but the units can be ranked simply with the lowest cost, RSS offers efficient methods for estimate population parameters than SRS.

To selecting a sample of size k in RSS method, we select k numbers of the samples of size k. Each of the samples of size k is ranked the variable of interest, and is selected the smallest unit level from first sample of size k, second ranked unit from the second sample of size k and unit has a larger rank from the $k-th$ sample of size k. If a larger sample size is required, In this case, these methods can be repeated s times until be measured a sample of size $n = sk$. This n units are data from RSS method.

$X_{i(ic)}$, $i = 1, 2, \dots, k$ $c = 1, 2, \dots, s$ are sample sets of ranked from Exponentiated Gumbel Distribution of size $n = sk$, that k is the size of the collection, and s is the number of iterations (without loss of generality, in order to compare with SRS, we get zero for the location parameter). For simplicity $Y_{ic} = X_{i(ic)}$ is considered, and c is fixed. Y_{ic}'s are independent random variables with density function of $i-th$ order statistic

$$g(y_{ic}) = \frac{k!}{(i-1)!\,(k-i)!}[F(y_{ic})]^{i-1}[1 - F(y_{ic})]^{k-i} f(y_{ic})$$

Likelihood function of sample $Y_{1c}, Y_{2c}, \dots, Y_{kc}$, and the log-likelihood function are derived as follows

$$L\left(\underline{y}\middle| a, \lambda\right) = \prod_{c=1}^{s}\prod_{i=1}^{k} \frac{k!\, a\lambda e^{-\lambda y_{ic}}\left[exp\left(-e^{-\lambda y_{ic}}\right)\right]^{ai}\left[1 - \left(exp\left(-e^{-\lambda y_{ic}}\right)\right)^{a}\right]^{k-i}}{(i-1)!\,(k-i)!}$$

$$\boldsymbol{LogL = C + ks(ln\alpha + ln\lambda) + \sum_{c=1}^{s}\sum_{i=1}^{k}(\alpha i)e^{-\lambda y_{ic}}}$$
$$\boldsymbol{+\sum_{c=1}^{s}\sum_{i=1}^{k}(k-i)ln\left[1-\left(exp(-e^{-\lambda y_{ic}})\right)^{\alpha}\right]-\lambda\sum_{c=1}^{s}\sum_{s=1}^{k}y_{ic}}$$

where $\boldsymbol{c}$ is fixed. We derivative of the log-likelihood function ratio to $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$, and we put them equal to zero

$$\boldsymbol{\frac{\partial LogL}{\partial \lambda} = \frac{ks}{\lambda} - \sum_{c=1}^{s}\sum_{s=1}^{k}y_{ic} + \sum_{c=1}^{s}\sum_{s=1}^{k}(\alpha i)y_{ic}\,e^{-\lambda y_{ic}} - \alpha\sum_{c=1}^{s}\sum_{s=1}^{k}\frac{(k-i)(e^{-\lambda y_{ic}})(1-e^{-\lambda y_{ic}})^{\alpha}y_{ic}}{1-(exp(-e^{-\lambda y_{ic}}))^{\alpha}}} \quad (8)$$

$$\boldsymbol{\frac{\partial LogL}{\partial \alpha} = \frac{ks}{\alpha} + \sum_{c=1}^{s}\sum_{s=1}^{k}(i)\,e^{-\lambda y_{ic}} - \sum_{c=1}^{s}\sum_{s=1}^{k}\frac{(k-i)\left(exp\left(-e^{-\lambda y_{ic}}\right)\right)^{\alpha}ln\left(exp\left(-e^{-\lambda y_{ic}}\right)\right)}{1-\left(exp(-e^{-\lambda y_{ic}})\right)^{\alpha}}} \quad (9)$$

Estimators $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ are obtained of nonlinear equations (8) and (9). These equations are solved by using numerical methods.

Simulation

In this section, simulation results based on maximum likelihood estimation of parameters of two simple random sampling and sampling size sets and repetitions per set to rank different (for $\boldsymbol{\lambda = 1}$, $\boldsymbol{\alpha = 1,\ 1.5}$ based on 1000 replications of software Maple) have been done.

Maximum likelihood estimation $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ in simple random sampling of equations (2) and (3) and set the sampling rate from equation (8) and (9) is obtained. In order to compare more accurately estimate the parameters using both Newton-Raphson done.

Bias, mean squared error and sampling efficiency ratings set to a random sample per $\boldsymbol{n = 9, 12, 15, 20, 24, 30, 32, 40}$. The size of sample $\boldsymbol{(3, 4)}$ and repetition rate $\boldsymbol{(3, 5, 8, 10)}$ that $\boldsymbol{(n = ks)}$. $\boldsymbol{\lambda = 1}$, $\boldsymbol{\alpha = 1,\ 1.5}$. Using the calculations presented in Table 1 and 2, given the following tables can be seen. The values of the mean squared error and bias estimators in simple random sampling, the sampling rate is greater than the set. Set the sampling rate for a fixed set size by increasing the repetition rate of the mean square error and bias can be reduced.

Table1. Bias and (MSE) of the exponentiated Gumbel distribution parameters for $\lambda = 1$, $\alpha = 1$ with RSS and SRS

(m,k)	RSS		n	SRS		$Efficiency = \frac{MSE(SRS)}{MSE(RSS)}$	
	α	λ		α	λ	α	λ
(3,3)	1.3940(1.9433)	0.8984(0.8091)	9	2.4746(6.9351)	1.9586(1.5869)	3.5687	1.9613
(4,3)	0.0999(0.0470)	0.2415(0.0647)	12	1.5049(2.3249)	1.7865(1.8744)	49.4695	28.9706
(3,5)	1.3303(0.7030)	0.8995(0.8071)	15	1.3961(2.1103)	1.7965(0.9562)	3.0018	1.1847
(4,5)	0.0873(0.0450)	0.1781(0.0373)	20	1.3773(2.0505)	1.3179(1.7388)	45.5613	46.6162
(3,8)	0.3397(0.3018)	0.3399(0.1163)	24	1.2074(1.4733)	0.5804(0.9454)	4.8817	8.1289
(3,10)	0.3225(0.0510)	0.3216(0.0103)	30	1.0759(1.3379)	0.5577(0.9453)	26.2333	91.776
(4,8)	0.0775(0.0440)	0.1851(0.0348)	32	1.3305(1.8936)	0.7910(1.0077)	4.3036	28.9458
(4,10)	0.0740(0.0410)	0.2453(0.0336)	40	1.2819(1.6790)	0.2911(0.9438)	28.0892	14.8396

Table2. Bias and (MSE) of the exponentiated Gumbel distribution parameters for $\lambda = 1$, $\alpha = 1.5$ with RSS and SRS

(m,k)	RSS		n	SRS		$Efficiency = \frac{MSE(SRS)}{MSE(RSS)}$	
	α	λ		α	λ	α	λ
(3,3)	0.5771(0.3335)	0.5679(0.4207	9	1.6164(2.6438)	1.6484(2.2666)	7.9274	5.3876
(4,3)	0.5942(0.3534)	0.6816(0.4697	12	0.9699(0.9675)	0.9338(1.2461)	2.7376	2.6529
(3,5)	0.5578(0.3113	0.4878(0.1704)	15	0.8974(0.8970)	1.5045(2.1076)	2.8814	12.3658
(4,5)	0.5752(0.3323)	0.5099(0.2602)	20	0.8851(0.8512)	0.9032(1.1420)	2.5615	4.3889

(3,8)	0.5543(0.3079)	0.4373(0.1659)	24	0.7716(0.6022)	1.4651(2.0935)	1.9558	12.6190
(3,10)	0.5428(0.2983)	0.4292(0.1565)	30	0.1506(0.5957)	0.8511(1.5217)	1.9969	9.7233
(4,8)	0.5641(0.3183)	0.4853(0.2429	32	0.8536(0.7835)	0.8876(1.1884)	2.4615	2.4615
(4,10)	0.5619(0.3159)	0.3943(0.2404)	40	0.8212(0.6903)	0.7957(1.1460)	2.1851	3.1798

Results

The bias and the MSE of the estimators of the parameter of the exponentiated Gumbel distribution based on RSS are smaller than SRS by comparing simulation results. Thus the RSS gives more efficient estimators than SRS.

References

Abu-Dayyeh, W. A., Al-Subh, S. A., & Muttlak, H. A. (2004). Logistic parameters estimation using simple random sampling and ranked set sampling data. *Applied mathematics and computation,* 150 (2), 543-554.

Adatia, A. (2000). Estimation of parameters of the half-logistic distribution using generalized ranked set sampling. *Computational statistics & data analysis*, 33(1), 1-13.

Al-Hadhrami, S. A. (2010). Parametric estimation on modified Weibull distribution based on ranked set sampling. *European Journal of Scientific Research*, 44(1), 73-78.

Al-Saleh, M. F., & Diab, Y. A. (2009). Estimation of the parameters of Downton's bivariate exponential distribution using ranked set sampling scheme. *Journal of Statistical Planning and Inference,* 139 (2), 277-286.

Bhoj, D. S., & Ahsanullah, M. (1996). Estimation of parameters of the generalized geometric distribution using ranked set sampling. *Biometrics,* 685-694.

Chacko, M., & Yageen Thomas, P. (2007). Estimation of a parameter of bivariate Pareto distribution by ranked set sampling. *Journal of Applied Statistics,* 34 (6), 703-714.

Chuiv, N. N., Sinha, B. K., & Wu, Z. (1994). Estimation of location parameter of a Cauchy distribution using ranked set sample. In *Technical Report*. University of Maryland Baltimore County Blacksburg, Virginia, 24061, USA.

Dell, T. R., & Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics,* 545-555.

Lam, K., Sinha, B. K., & Wu, Z. (1994). Estimation of parameters in a two-parameter exponential distribution using ranked set sample. *Annals of the Institute of Statistical Mathematics*, 46 (4), 723-736.

Lam, K., Sinha, B. K., & Wu, Z. (1996). Estimation of location and scale parameters of a logistic distribution using a ranked set sample. In *Statistical Theory and Applications* (pp. 187-197). Springer New York.

McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Crop and Pasture Science,* 3 (4), 385-390.

Mudholkar, G. S., & Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability,* 42(2), 299-302.

Nadarajah, S. (2006). The exponentiated Gumbel distribution with climate application. *Environmetrics,* 17 (1), 13-23.

Stokes, S. L. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics,* 35-42.

Takahasi, K., & Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics,* 20 (1), 1-31.