Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 25, Sayı 1, 49-53, 2021

Süleyman Demirel University
Journal of Natural and Applied Sciences
Volume 25, Issue 1, 49-53, 2021

DOI: 10.19113/sdufenbed.699703

Parameter Estimation for Uniform-Geometric Distribution Based on Censored Sample

Mehtap Koca YILMAZ[1], Yunus AKDOĞAN*[1], Kadir KARAKAYA[1]

[1] Selcuk University, Science Faculty, Department of Statistics, 42100, Konya, Turkey

(Alınış/ Received: 06.03.2020, Kabul/ Accepted: 11.01.2021, Online Yayınlanma/ Published Online: 15.04.2021)

Keywords
Uniform-Geometric distribution,
Maximum likelihood,
Modified maximum likelihood,
Method of moments,
Method of proportions.

Abstract: Recently, many new discrete distributions have been obtained. The uniform-geometric distribution is a newly obtained discrete distribution. In literature, parameter estimation is rare in the case of censored samples for new discrete distributions. In this study, the parameter estimation based on type-I censored sampling for the unknown parameter of the uniform geometric distribution is obtained using the maximum likelihood, methods of proportions, methods of moments, and modified maximum likelihood estimation methods. The performance of estimation methods is compared using the Monte Carlo simulation via biases and mean squared errors. Finally, two real data applications are given.

Düzgün Geometrik Dağılımının Sansürlü Örneklem Durumunda Parametre Tahmini

Anahtar Kelimeler
Düzgün Geometrik dağılım,
En çok olabilirlik,
Modifiye edilmiş en çok olabilirlik,
Momentler yöntemi,
Oranlar yöntemi.

Özet: Son zamanlarda birçok yeni kesikli dağılım elde edilmiştir. Düzgün geometrik dağılım, yeni elde edilen kesikli bir dağılımdır. Yeni kesikli dağılımlar için sansürlü örneklem durumunda parametre tahmininin eksikliği oldukça fazladır. Bu çalışmada düzgün geometrik dağılımın bilinmeyen parametresi için tip-I sağdan sansürlü örnekleme dayalı parametre tahmini elde edilmiştir. Parametre tahmini en çok olabilirlik yöntemi, oranlar yöntemi, momentler yöntemi ve modifiye edilmiş en çok olabilirlik yöntemleri kullanarak elde edilmiştir. Yöntemlerin parametre tahminindeki performanslarını kıyaslamak için parametre tahminlerinden elde edilen yan ve hata kareler ortalaması Monte Carlo simülasyonu ile elde edilmiştir. Son olarak çalışmada gerçek iki veri uygulaması verilmiştir.

1. Introduction

In life testing experiments, the failure-time distribution is used to describe mathematically, the life of a machine, or an electronic device. The exponential, gamma, Weibull, and lognormal distributions are well-known in failure analysis. These models have been proposed by [1, 2]. However, it is challenging and unfeasible to evaluate the life-time of a unit in continuous-time. In practice, one can find situations in which a device's lifecycle is discrete random variables such as the number of shots of the copier until the toner runs out, electrical shocks that cause an electronic device to malfunction, the number of revolutions of a machine gear until it breaks down.

In recent years, popular discrete distributions (such as geometric, negative binomial, etc.) have been worked to model lifetime data. Still, popular discrete distributions have limited practicality as models for reliability. Therefore, there is a need to obtain more sensible discrete lifetime distributions to comply with diverse types of lifetime data. So, many discrete lifetime distributions have been obtained with the discretized of the known continuous distributions in recent studies.

These studies are given as follows, Nakagawa and Osaki (1975) studied discrete Weibull (DW) distribution [3]. Stein and Dattero (1984) discretized other types of Weibull (DW2) distribution [4]. Discrete normal (DN) and Rayleigh (DR) distributions are considered by Dilip Roy [5, 6]. Krishna and Pundir (2009) discretized Burr (DB) and Pareto (DP) distributions [7]. Another continuous distribution, the inverse Weibull distribution, is discretized by Jazi et al. [8] as discrete inverse Weibull (DIW) distribution.

In recent studies, new discrete distributions have been obtained by compounding two discrete distributions. These studies are given by Hu et al. [9] and Déniz [10]. Also, Akdoğan et al. [11] proposed uniform-Geometric (UG) distribution by using this methodology. Finally, using the same method, Binomial-Discrete Lindley (BDL) distribution is suggested by Kuş et. al. [12]. Additionally, these authors have tried to determine

* Corresponding author: yakdogan@selcuk.edu.tr

the distributional characteristics of discrete distributions in recent years. Although there are so many new discrete distributions, the parameter estimation in the case of censored sampling has not been studied except for the discrete Weibull distribution.

Type-I censored sample is widespreadly used in real life and is widely used in reliability analysis [13]. Sometimes the experiment may be limited in terms of time and cost when a lifetime is discrete. For example, if a three-year warranty is required for the life of a part, the test may not continue after three years, which is known as type-I right censored data. Kulasekera [13] studied the parameter estimation in the case of a censored sample for DW distribution.

In this study, the UG distribution introduced by Akdoğan et al. [11] is given in Section 2. In Section 3, method of moments (MM), maximum likelihood (ML), method of Proportion (MP), and modified maximum likelihood (MML) estimates of UG distribution parameter under the type I censored sample are obtained. The performance of the estimators in Section 4, according to the bias and mean square errors (MSE) criteria, are examined. In Section 5, two numerical examples are given to show the applicability.

2. Uniform-Geometric Distribution

Akdoğan et al. [11] defined the probability mass function (pmf) and cumulative distribution function (cdf) of the UG distribution as

$$f(t) = p(1-p)^{t-1} LerchPhi[(1-p),1,t], t = 1,2,\ldots \quad (1)$$

and

$$F(t) = 1 - p(1-p)^{t}\left[\frac{1}{p} - t\ LerchPhi(1-p,1,t+1)\right] \quad (2)$$

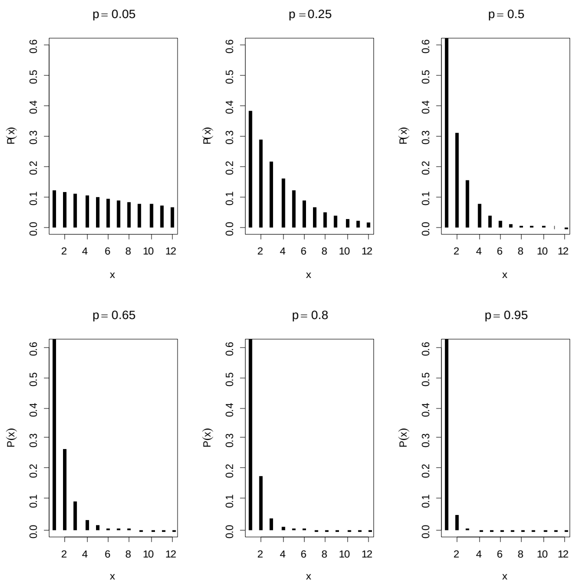

Figure 1. Plots for pmf of $UGD(p)$ distribution.

respectively, where $0 < p < 1$. It will be denoted as $UGD(p)$ and $LerchPhi(z,a,v)$ a Lerch zeta function is given by (see [14])

$$LerchPhi(z,a,v) = \sum_{n=0}^{\infty} \frac{z^n}{(v+n)^a}. \quad (3)$$

If T has the pmf given in Eq. (1), then it is denoted by $T \sim UGD(p)$. Figure 1, shows the graphs of the $UGD(p)$ for some different p.

3. Parameter Estimation Techniques

3.1. Maximum likelihood estimation

Let $T_1^0, T_2^0, \ldots, T_n^0$ be i.i.d. observations from $UGD(p)$ and let each investigation be Type-I censored by K_1,K ,K_n, respectively. So, we observe

$$T_i = min(T_i^0, K_i), i = 1,2,\ldots,n \quad (4)$$

along with the indicator variables

$$\delta_i = \begin{cases} 1, & if\ T_i = T_i^0 \\ 0, & o.w. \end{cases} \quad (5)$$

Afterward, the log-likelihood function for $UGD(p)$ distribution based on the type-I censored sample is given by

$$\begin{aligned} \ell(p) &= log\left\{\prod_{i=1}^{n} f(t_i)^{\delta_i}\left(1-F(t_i)\right)^{1-\delta_i}\right\} \\ &= \sum_{i=1}^{n} \delta_i\, log(f(t_i)) + \sum_{i=1}^{n}(1-\delta_i)\, log(1-F(t_i)) \\ &= \sum_{i=1}^{n} \delta_i\, log(p(1-p)^{t_i-1}\varphi_i) \\ &+ \sum_{i=1}^{n}(1-\delta_i)\, log\left(p(1-p)^{t_i}\vartheta_i\left[\frac{1}{p}-\vartheta_i\right]\right) \end{aligned} \quad (6)$$

The likelihood equation is given by

$$\frac{\partial\, log(\ell(p))}{\partial p} = 0 \quad (7)$$

$$\begin{aligned} &\sum_{i=1}^{n} \delta_i\left[\frac{t_i(1-p)+2p-1}{p(1-p)} - \frac{1}{p(1-p)\varphi_i}\right] \\ &\sum_{i=1}^{n}(1-\delta_i)\left[\frac{p+p^2[t_i-1-(t_i+1)\vartheta_i]}{p^3(1-p)\left[\frac{1}{p}-\vartheta_i\right]}\right] \\ &+\sum_{i=1}^{n}(1-\delta_i)\left[\frac{t_i D(p)}{p}\right] = 0, \end{aligned} \quad (8)$$

where $\varphi_i = LerchPhi[(1-p),1,t_i]$ and $\vartheta_i = LerchPhi[(1-p),1,t_i]$. The solution of the likelihood equation Eq. (8) will provide the ML estimators of p. It will be denoted by $\hat{p}_*$. Iterative methods can be used for this purpose.

3.2. Method of proportions

The MP is proposed by Khan et al. [15] to estimate the parameters of DW distribution. Now we proposed the same method for $UGD(p)$ distribution. Let $T_1, T_2, \dots, T_n$ be a random sample from $UGD(p)$ distribution and define indicator function,

$$\upsilon(T_i) = \begin{cases} 1, T_i = 0 \\ 0, T_i > 0 \end{cases} \tag{9}$$

then $Y = \frac{1}{n}\sum_{i=1}^{n} \upsilon(T_i)$ shows the proportion of 1's in the censored sample. The proportion Y is a consistent, unbiased estimator of the probability $f(1) = pLerchPhi[(1-p), 1{,}1]$. Therefore, an estimate p can be offered as

$$Y = p[LerchPhi(1-p), 1{,}1] = 0 \tag{10}$$

$$\frac{p\,log(p)}{p-1} - Y = 0 \tag{11}$$

the solution of Eq. (11) will provide the MP estimator of p. It will be denoted by $\hat{p}$.

3.3. Modified maximum likelihood estimation

Let $T_1, T_2, \dots, T_n$ be a censored sample from $UGD(p)$ distribution and the log-likelihood function given by Eq. (6). The ML estimation of p is obtained by solving Eq. (8) for p. An iterative method to get the ML estimates. To avoid this duration, we used Kulasekera's [13] approximation in the following fashion. Assuming $K_i > 1, i = 1,2, \dots, n$. Let $f(p) = LHS$ of Eq. (7). Then to achieve the ML estimator of p, the following equation should be solved:

$$f(p) = \frac{\partial\, log(\ell(p))}{\partial p} = 0. \tag{12}$$

Eq. (12) can not be solved numerically. Now, instead of solving Eq. (12), one can solve the system

$$f(\hat{p}) + (p - \hat{p}) \left.\frac{\partial f}{\partial p}\right|_{p=\tilde{p}} = 0 \tag{13}$$

where $\hat{p}$ is the proportion estimate. The function f_1 all the partial derivative is evaluated at $\hat{p}$. This procedure gives an estimator, which is called MML estimator of p. It will be denoted by $\tilde{p}$.

3.4. Methods of moments

To estimate the parameter of $UGD(p)$ distribution by this method, we need the first sample moments, which is given below:

$$p\sum_{x=1}^{\infty} t((1-p)^{t-1}LerchPhi[(1-p), 1, t]) = \frac{1}{n}\sum_{i=1}^{n} T_i \tag{14}$$

ome numerical methods such as the Newton-Raphson method can be used to solve Eq. (14). The solution of Eq. (14) is the MM estimator $\tilde{p}_*$ of parameter p.

4. Simulation Study

In the simulation study, we have generated random samples and calculated the ML, MM, MML, and MP estimates of p and the performance of the estimators according to the bias and MSEs criteria are compared. In Tables 1-4, for different values of (n, p, K), we calculated these criteria of parameter estimates over 5000 iterations. The MATLAB software is used in the simulation study.

Table 1. Biases and MSEs estimate for $p = 0.3$

		MLE		MMLE	
n	K	Bias	Mse	Bias	Mse
50	3	0.0675	0.0052	0.0811	0.0076
100	3	0.0669	0.0051	0.0808	0.0073
200	3	0.0273	0.0025	0.0174	0.0026
300	3	0.0228	0.0021	0.0031	0.0022
500	3	0.0112	0.0015	0.0013	0.0016
50	5	-0.0262	0.0035	-0.0285	0.0042
100	5	-0.0256	0.0021	-0.0284	0.0025
200	5	-0.0250	0.0014	-0.0282	0.0016
300	5	-0.0236	0.0012	-0.0281	0.0014
500	5	-0.0182	0.0010	-0.0251	0.0011
50	∞	0.0057	0.0024	0.0006	0.0025
100	∞	0.0030	0.0011	0.0005	0.0012
200	∞	0.0017	0.0006	0.0003	0.0006
300	∞	0.0012	0.0004	0.0003	0.0004
500	∞	0.0006	0.0002	0.0002	0.0002

Table 2. Biases and MSEs estimate for $p = 0.3$

		MME		MPE	
n	K	Bias	Mse	Bias	Mse
50	3	0.0052	0.0026	0.0041	0.0047
100	3	0.0010	0.0012	0.0011	0.0023
200	3	0.0006	0.0006	0.0008	0.0012
300	3	0.0005	0.0004	0.0005	0.0008
500	3	0.0004	0.0002	0.0004	0.0004
50	5	-0.0052	0.0026	0.0041	0.0047
100	5	-0.0015	0.0012	0.0011	0.0023
200	5	-0.0010	0.0006	0.0008	0.0012
300	5	-0.0006	0.0004	0.0005	0.0008
500	5	-0.0005	0.0002	0.0001	0.0005
50	∞	0.0035	0.0025	0.0025	0.0047
100	∞	0.0010	0.0012	0.0011	0.0023
200	∞	0.0007	0.0006	0.0008	0.0012
300	∞	0.0005	0.0004	0.0005	0.0008
500	∞	0.0004	0.0002	0.0001	0.0003

Table 3. Biases and MSEs estimate for $p = 0.8$

		MLE		MMLE	
n	K	Bias	Mse	Bias	Mse
50	3	0.0259	0.0106	0.0127	0.0049
100	3	0.0106	0.0030	0.0103	0.0026
200	3	0.0026	0.0012	0.0020	0.0012
300	3	0.0012	0.0009	0.0019	0.0009
500	3	0.0006	0.0005	0.0014	0.0005
50	5	0.0159	0.0080	0.0056	0.0060
100	5	0.0068	0.0025	0.0031	0.0025
200	5	0.0012	0.0012	0.0019	0.0012
300	5	0.0009	0.0008	0.0013	0.0008
500	5	0.0004	0.0005	0.0007	0.0005
50	∞	0.0094	0.0089	0.0085	0.0046
100	∞	0.0038	0.0024	0.0051	0.0024
200	∞	0.0016	0.0012	0.0023	0.0012
300	∞	0.0009	0.0008	0.0014	0.0008
500	∞	0.0003	0.0005	0.0005	0.0005

Table 4. Biases and MSEs estimate for $p = 0.8$

		MME		MPE	
n	K	Bias	Mse	Bias	Mse
50	3	0.0053	0.0047	0.0018	0.0055
100	3	0.0037	0.0024	0.0018	0.0028
200	3	0.0013	0.0012	0.0005	0.0014
300	3	0.0008	0.0008	0.0004	0.0009
500	3	0.0006	0.0005	0.0002	0.0006
50	5	0.0039	0.0048	0.0006	0.0056
100	5	0.0019	0.0024	0.0003	0.0029
200	5	0.0013	0.0012	0.0005	0.0014
300	5	0.0009	0.0008	0.0005	0.0009
500	5	0.0005	0.0005	0.0004	0.0006
50	∞	0.0067	0.0047	0.0033	0.0056
100	∞	0.0038	0.0024	0.0020	0.0028
200	∞	0.0016	0.0012	0.0009	0.0014
300	∞	0.0009	0.0008	0.0005	0.0010
500	∞	0.0003	0.0005	0.0001	0.0005

From Tables 1-4, one can see that all estimates are biased, but these estimates asymptotically unbiased. The MM and MP estimates are almost like in terms of MSE and both perform better than ML and MML estimates. Furthermore, when the sample size increases, the values of bias and MSE decreases.

5. Real Data Application

The first data set: This data set analyzed by [16]. The data has been integrated for 12 or more cycles, a total of 586 women. The first real data are given in Table 5.

Table 5. The fecundability data

Cycles	1	2	3	4	5	6	7	8	9	10	11+
N	227	123	72	42	21	31	11	14	6	4	35

Table 6. Results of $UGD(p)$ for first real data

	Complete Data	$K = 3$	$K = 5$
ML	0.1889	0.1178	0.1485
MP	0.1883	0.1883	0.1883
MM	0.1784	0.1784	0.1784
MML	0.1889	0.1178	0.1485

The second data set: The second data set is given from Xie and Goh [17] and presents an industrial process. The data are:

1 1 1 1 1 1 2 2 2 2 3 3 3 4 4 4 5 5 7 9 11 13 14 14 17 18 26 29.

Table 7. Results of $UGD(p)$ for second real data

	Complete Data	$K = 3$	$K = 5$
ML	0.0742	0.0469	0.0683
MP	0.0772	0.0772	0.0772
MM	0.0771	0.0771	0.0771
MML	0.0742	0.0476	0.0763

The ML, MP, MM, and MML estimates of p for the first and second real data are given in Tables 6 and 7. From these Tables, the ML, MP, MM, and MML estimators obtained almost the same for complete data. The MP and the MM estimates are almost identical in terms of estimates and both perform better than ML and MML estimates.

6. Conclusion

The censored sampling is very advantageous in terms of both cost and time. In this study, the ML, MM, MML and MP estimators of the $UGD(p)$ distribution are obtained in the case of a censoring sample. In the simulation and real data, it shows that in the case of a censoring sample and a complete sample, the results are close to each other. In the case of limited time and cost, the units have $UGD(p)$ distribution and can be used in the case of a censored sample, which is very advantageous for both time and cost.

Declaration of Ethical Code

In this study, we undertake that all the rules required to be followed within the scope of the "Higher Education Institutions Scientific Research and Publication Ethics Directive" are complied with, and that none of the actions stated under the heading "Actions Against Scientific Research and Publication Ethics" are not carried out.

References

[1] Lawless, J. F. 1982. Statistical Models and Methods for Lifetime Data. John Wiley & Sons, New York, 580p.

[2] Sinha, S. K. 1986. Reliability and Life testing. Wiley Eastern Ltd, New Delhi, 252p.

[3] Nakagawa, T., Osaki, S. 1975. This discrete Weibull distribution. IEEE Transactions on Reliability, 24, 300-301.

[4] Stein, W. E., Dattero, R. 1984. A new discrete Weibull distribution. IEEE Transactions on Reliability, 33, 196-197.

[5] Roy, D. 2003. The discrete normal distribution. Communications in Statistics Theory and Methods, 32, 1871-1883.

[6] Roy, D. 2004. Discrete Rayleigh distribution. IEEE Transactions on Reliability, 53(2), 255-260.

[7] Krishna, H., Pundir, P. S. 2009. Discrete Burr and discrete Pareto Distributions. Statistical Methodology, 6, 177-188.

[8] Jazi, M. A, Lai, C. D., Alamatsaz, M.H. 2009. A discrete inverse Weibull distribution and estimation of its parameters. Statistical Methodology, 7, 121-132.

[9] Hu, Y., Peng, X., Li, T., Guo, H. 2017. On the Poisson approximation to photon distribution for faint lasers. Phys Lett A., 367, 173–176.

[10] Déniz, E. G. 2007. A new discrete distribution: Properties and applications in medical care. Journal Applied Statistics, 40(12), 2760–2770.

[11] Akdoğan, Y., Kuş, C., Asgharzadeh, A., Kınacı, I., Sharai, F. 2016. Uniform-geometric distribution. Journal of Statistical Computation and Simulation, 86(9), 1754-1770.

[12] Kuş, C., Akdoğan, Y., Asgharzadeh, A., Kınacı, İ., and Karakaya, K. 2018. Binomial-discrete Lindley distribution. Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, 68(1), 401-411.

[13] Kulasekera, K. B. 1994. Approximate MLEs of the Parameters of a Discrete Weibull Distribution with Type I Censored Data. Microelection Reliability, 34, 1185-1188.

[14] Gradshteyn, I. S., Ryzhik, I. M. 2007. Table of Integrals, Series, and Products. 7th ed. San Diego, CA: Academic Press, 1171p.

[15] Khan, M. S. A., Khalique, A., Abouammoh, A.M. 1989. On estimating parameters in a discrete Weibull distribution. IEEE Transactions on Reliability, 38, 348-350.

[16] Weinberg, C. R., Gladen, B. C. 1986. The beta-geometric distribution applied to comparative fecundability studies. Biometrics, 42, 547-560.

[17] Xie, M., Goh, T. N. 1993. Improvement detection by control charts for high yield processes. Int J Qual Reliab Manag, 10, 24-31.