International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 2 (2018) pp. 1261-1264

Parameter Estimation Method for the Two Parameter Gamma Distribution Based on Transformation

Manad Khamkong[1, 2]

[1]*Department of Statistics, Chiang Mai University, Chiang Mai, Thailand.*
[2]*Environmental Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand.*

Abstract

Two parameter gamma distribution is widely used to model positively-skewed distributions. In this paper, an alternative closed-form estimator based on transformation to a normal distribution is proposed. The results of a simulation study to compare the new estimator with previously reported ones show that it performed well in terms of root mean square errors, and parameter evaluation using two real-life datasets supported these findings. It is evident that the proposed TM estimator is a useful tool for determining gamma distribution parameters.

Keywords: closed-form estimator; gamma; positively-skewed; transformation**.**

INTRODUCTION

In environmental science data such as for precipitation and air quality monitoring by measuring airborne particulate matter (PM10 and PM2.5), variables most frequently display asymmetric distribution patterns with various levels of skewness and kurtosis, and are often positively skewed [1]. In these situations, a gamma distribution is often used for statistical inference in practice. Suppose $X_1, X_2, \ldots, X_n$ are from a random sample of size n from a two parameter gamma distribution denoted as GM(α, β) with a probability density function (pdf) defined as

$$f(x \mid \alpha, \beta) = \begin{cases} x^{\alpha-1}\left(\beta^{\alpha}\Gamma(\alpha)\right)^{-1} e^{-\frac{x}{\beta}}, & \text{for } x, \alpha, \beta > 0 \\ 0 & , \quad \text{otherwise} \end{cases} \quad (1)$$

where α is the shape parameter, β is the scale parameter, and $\Gamma(.)$ is the gamma function. Here, the expectation and variance of a gamma variable are $\alpha\beta$ and $\alpha\beta^2$, respectively. The problem of estimating the unknown parameters α and β is very important and several methods have been proposed to obtain them for different uses in the literature. Nevertheless, there has not previously been a comparison of closed-form estimators of the parameters. A number of studies have compared different estimation methods, such as Amrir [2], Son and Oh [3], and Zerda [4], and the maximum likelihood estimation (MLE) method has been shown to be quite efficient but computationally demanding, and although the Bayesian method outperformed the other methods, it was not considered.

In this paper, an alternative closed-form estimator based on transforming a GM(α, β) distribution to a normal distribution is presented. Subsequently, a comprehensive comparison of the closed-form estimator with other previously reported ones was carried out via a simulation study with different parameters and sample sizes. Moreover, the proposed method was applied to estimate the parameters using two pertinent real-life datasets.

METODS FOR ESTIMATING GAMMA PARAMETERS

In this section, alternatives to the new closed-form estimator based on transformation to a normal distribution are considered to obtain the estimators α and β of a GM (α, β) distribution.

The method of moments (MM) estimator

The MM estimator is derived by equating the population moment to the sample moments for the unknown parameters of a GM(α, β) distribution to obtain

$$\hat{\alpha}_{MM} = \frac{\bar{X}^2}{S'^2} \text{ and } \hat{\beta}_{MM} = \frac{S'^2}{\bar{X}}. \quad (1)$$

$$\text{where } S'^2 = \frac{\sum_{1=i}^{n}(X_i - \bar{X})^2}{n}, \quad \bar{X} = \frac{\sum_{1=i}^{n} X_i}{n}.$$

Hwang and Huang [5] proposed a bias correction of the MM estimator (denoted MMB) for small samples based on the characterization of a GM(α, β) distribution as

$$\hat{\alpha}_{MMB} = \frac{1}{V_X^2} - \frac{1}{n}, \text{ and } \hat{\beta}_{MMB} = \frac{\bar{X}}{\left(\frac{1}{V_X^2} - \frac{1}{n}\right)}. \quad (2)$$

$$\text{where } V_X^2 = \frac{S_X^2}{\bar{X}^2}, \quad S_X^2 = \frac{\sum_{1=i}^{n}(X_i - \bar{X})^2}{n-1}.$$

Unfortunately, the moment estimators are non-unique and depend on the selected sample moment. Wiens, Cheng, and Beaulieu [6] considered three moments of a sample (denoted MMW) and suggested estimation of the two parameters given by

$$\hat{\alpha}_{MMW} = (n-1)\left[\frac{\sum_{i=1}^{n} X_i}{n\sum_{i=1}^{n} X_i \ln X_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \ln X_i}\right] \text{ and}$$

$$\hat{\beta}_{MMW} = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} X_i \ln X_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \ln X_i\right]. \quad (3)$$

The maximum likelihood estimation (MLE) estimator

The MLE estimator is derived from the log-likelihood function based on the observed sample $\mathbf{x} = (x_1, x_2, ..., x_n)^t$ from a GM(α, β) distribution such that

$$\ell(\alpha,\beta) = -n\alpha \ln\beta - n\ln\Gamma(\alpha) + (\alpha-1)\sum_{i=1}^{n} \ln(x_i) - \frac{\sum_{i=1}^{n} x_i}{\beta}. \quad (4)$$

The MLE estimates for the gamma parameters are obtained by taking the partial derivatives of the log-likelihood function with respect to α and β, respectively. Subsequently, the MLE estimators of α and β are obtained by equating the resulting expressions to zero as follows:

$$\frac{\partial}{\partial\alpha}\ell(\alpha,\beta) = -n\ln\beta - \frac{n}{\Gamma(\alpha)}\left(\frac{\partial}{\partial\alpha}\Gamma(\alpha)\right) + \sum_{i=1}^{n}\ln(x_i) = 0. \quad (5)$$

$$\frac{\partial}{\partial\beta}\ell(\alpha,\beta) = -\frac{n\alpha}{\beta} + \frac{\sum_{i=1}^{n} x_i}{\beta^2} = 0. \quad (6)$$

In particular, equation (5) is a nonlinear equation without a closed-form solution, and so iterative methods are used to find an approximation, which has been proposed by many authors to obtain MLE estimators. Son and Oh [3] showed that the Greenwood and Durand approximate MLE estimators for GM(α, β) are highly efficient compared to other approximation solutions. Provided that $B = \ln\bar{X} - \overline{\ln X}$ can be expressed, the Greenwood and Durand approximate MLE estimators are obtained by

$$\hat{\alpha}_{MLE} = \begin{cases} \dfrac{0.5000876 + 0.1648852B - 0.0544274B^2}{B}, & 0 < B \le 0.5772 \\ \dfrac{8.898919 + 9.05990B - 0.9775373B^2}{B(17.79728 + 11.968477B + B^2)}, & 0.5772 < B < 17, \end{cases}$$

and $$\hat{\beta}_{MLE} = \frac{\bar{X}}{\hat{\alpha}_{MLE}}. \quad (7)$$

Furthermore, Ye and Chen [7] proposed closed-form estimators (denoted as MLEC) for the gamma parameters based on two out of the three likelihood equations for the generalized gamma distribution:

$$\hat{\alpha}_{MLEC} = \frac{n\sum_{i=1}^{n} X_i}{n\sum_{i=1}^{n} X_i \ln X_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \ln X_i} \text{ and}$$

$$\hat{\beta}_{MLEC} = \frac{1}{n^2}\left[n\sum_{i=1}^{n} X_i \ln X_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \ln X_i\right]. \quad (8)$$

The proposed (TM) method

In the paper by Chaito, Khamkong, and Bookkamana [8], the authors applied an appropriate data transformation technique to convert a gamma distribution to a normal distribution to evaluate rainfall and drought potential based on a standardized precipitation index. They found that the cube-root (Wilson-Hilferty approximation) appropriate for an approximation transformation to a normal distribution. Certainly, Olive [9] suggested the Wilson-Hilferty approximation to transform a gamma distribution to a normal distribution with the expectation and variance being $(\alpha\beta)^{1/3}\left(1-\frac{1}{9\alpha}\right)$ and $(\alpha\beta)^{2/3}\frac{1}{9\alpha}\alpha\beta^2$, respectively. Thus,

$$Y = X^{1/3} \approx N\left((\alpha\beta)^{1/3}\left(1-\frac{1}{9\alpha}\right), (\alpha\beta)^{2/3}\frac{1}{9\alpha}\right).$$

Subsequently, the alternative TM estimator based on moment estimation is proposed as

$$\hat{\alpha}_{TM} = \frac{\left(18+\frac{9}{V_Y^2}\right) + \sqrt{\left(18+\frac{9}{V_Y^2}\right)^2 - 324}}{162} \text{ and}$$

$$\hat{\beta}_{TM} = \frac{\bar{Y}}{\hat{\alpha}_{TM}}. \quad (9)$$

SIMULATION RESULTS

In order to assess the performance of the estimators: MMB, MMW, MLE, MLEC, and the proposed TM estimator, a simulation study was conducted using the R statistical program (R Core Team, [10]) to generate random samples from a GM(α, β) distribution with different parameter values $(\alpha, \beta) = \{(1,1),(3,1),(5,1)\}$ based on 5,000 replications.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 2 (2018) pp. 1261-1264

All of the experiments were performed for different sample sizes n = 10, 30, 50, and 100. The criteria for comparing the performance of the estimators was the sample root mean square error (RMSE) calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{5{,}000}\sum_{i=1}^{5{,}000}\left[\left(\hat{\alpha}_i - \alpha\right)^2 + \left(\hat{\beta}_i - \beta\right)^2\right]}. \quad (10)$$

The smallest RMSE indicates the best performance, and the results are shown in Table 1.

In all of the cases, the proposed TM method outperformed the other methods. The RMSE of the MMW was slightly less than the MLEC. The RMSE of all methods decreased as the sample size increased for all of the parameters.

Table 1: Simulation Results of the Parameter Estimation of GM(α,β =1)

n	Method	α = 1	β	RMSE	α = 3	β	RMSE	α =5	β	RMSE
10	MMB	1.326	1.013	1.189	3.901	1.015	6.894	6.483	1.000	17.900
	MMW	1.230	0.994	0.768	3.799	1.004	6.079	6.366	0.996	16.617
	MLE	1.354	1.503	1.968	4.211	1.444	8.866	7.059	1.442	23.074
	MLEC	1.367	0.894	0.911	4.221	0.903	8.111	7.073	0.896	22.421
	TM	1.270	0.964	**0.729**	3.764	1.010	**5.708**	6.411	0.996	**17.139**
30	MMB	1.120	0.997	0.314	3.273	1.002	1.161	5.450	0.994	2.879
	MMW	1.065	0.995	0.167	3.213	1.000	0.888	5.381	0.995	2.462
	MLE	1.096	1.131	0.221	3.319	1.115	1.024	5.559	1.120	2.814
	MLEC	1.102	0.962	0.173	3.324	0.967	0.998	5.566	0.961	2.792
	TM	1.111	0.950	**0.151**	3.221	0.997	**0.834**	5.397	0.992	**2.437**
50	MMB	1.079	0.992	0.182	3.156	1.004	0.633	5.262	0.998	1.514
	MMW	1.039	0.995	0.094	3.117	1.002	0.469	5.213	0.998	1.247
	MLE	1.056	1.078	0.111	3.178	1.065	0.507	5.315	1.067	1.343
	MLEC	1.061	0.975	0.096	3.180	0.982	0.503	5.320	0.978	1.350
	TM	1.087	0.946	**0.087**	3.129	0.998	**0.450**	5.223	0.996	**1.210**
100	MMB	1.041	0.996	0.090	3.072	1.004	0.285	5.137	0.997	0.679
	MMW	1.021	0.996	0.045	3.050	1.004	0.206	5.113	0.997	0.549
	MLE	1.029	1.039	0.048	3.079	1.028	0.211	5.163	1.034	0.568
	MLEC	1.031	0.986	0.045	3.081	0.994	0.213	5.165	0.987	0.573
	TM	1.070	0.945	**0.044**	3.070	0.998	**0.197**	5.126	0.995	**0.547**

Note. The bold-face values indicate the performance estimation.

APPLICATION TO TWO REAL-LIFE DATASETS

Two real-life datasets were employed to compare the point estimates of the gamma distribution parameters using the TM estimator. The Anderson-Darling (AD) test statistic was used as criteria to fit the model and compare the performance of the estimator with the others. The estimator with the smallest AD value is preferred.

Dataset 1: Thai rainy season rainfall data for four months (February to May) from 2001 to 2013 (n = 13) were obtained from the Hydrology and Water Management Center for the Upper Northern Region Chiang Mai Thailand (in millimetres, [11]):

215, 373, 43, 42, 190, 86, 152,
136, 156, 140, 146, 118, 310.

Dataset 2: A random sample of 20 survival times (in weeks) of male mice exposed to 240 rads of gamma radiation (Gross and Clark, [12]):

152, 152, 115, 109, 137, 88, 94, 77, 160, 165,
125, 40, 128, 123, 136, 101, 62, 153, 83, 69.

The numerical results of the parameter estimation for the two real-life datasets are reported in Table 2. This table also shows that the point estimates of the parameters by the TM method outperformed the others, which supports the simulation studies.

Table 2: Comparison Results of GM(α, β) Parameter Estimation Using the Two Real-life Datasets

Method	**Dataset 1** (n = 13)			**Dataset 2** (n = 20)		
	$\hat{\alpha}$	$\hat{\beta}$	AD	$\hat{\alpha}$	$\hat{\beta}$	AD
MMB	2.842	56.979	0.357	9.998	11.347	0.469
MMW	2.831	57.210	0.357	8.725	13.002	0.413
MLE	3.048	53.120	0.364	8.797	12.896	0.415
MLEC	3.067	52.809	0.365	9.185	12.352	0.427
TM	2.841	57.514	**0.355**	8.389	13.565	**0.397**

Note. The bold-face values indicate the performance estimation.

CONCLUSIONS

The proposed TM estimator is an alternative closed-form estimator for parameter estimation of a GM(α, β) distribution based on transformation to a normal distribution, and it was shown to perform better than other previously reported estimators in terms of root mean square errors. Moreover, parameter estimation of a GM(α, β) distribution using two real-life datasets supported these findings. It is evident that the proposed TM estimator is a useful tool for determining gamma distribution parameters.

ACKNOWLEDGMENT

The author acknowledges to the financial and research facilities supported by Center of Excellence Research Fellowship of Environmental Science Research Center.

REFERENCES

[1] Khamkong, M., and Bookkamana, P. 2015, "Development of Statistical Models for Maximum Daily Rainfall in Upper Northern Region of Thailand," Chiang Mai J.Sci., 42(4), 1044 – 1053.

[2] Amrir, M. M. S., 2014, "Comparing different estimators for parameters of two gamma parameters using simulation," Int. J. Comput. Sci. Net. Sc., 14(11), 111-117.

[3] Son, Y. S., and Oh, M., 2006, "Bayesian estimation of the two-parameter gamma distribution," Commun. Stat. Simulat., 35, 285-293.

[4] Zerda, I., 2011, "An experimental comparison of popular estimation methods for the Weibull, gamma and Gompertz distributions," Schedae Informaticae, 20, 67-81.

[5] Hwang, T. Y., and Huang, P. H., 2002, "On new moment estimation of parameters of the gamma distribution using its characterization," Ann. Inst. Stat. Math., 54(4), 840-847.

[6] Wiens, D. P., Cheng, J., and Beaulieu, N. C., 2003, "A class of method of moments estimators for the two-parameter gamma family," Pak. J. Stat., 19(1), 129-141

[7] Ye, Z. S., and Cheng, N., 2016, "Closed-form estimators for the gamma distribution derived from likelihood equations," Am. Stat., 71(2), 177-181.

[8] Chaito, T., Khamkong, M., and Bookkamana, P., 2016, "Application of transformation techniques to evaluate drought," Burapha. Sci. J., 21(2), 86 -98.

[9] Olive, D. J., 2014, Statistical Theory and Inference, New York: Springer.

[10] R Core Team., 2016, R: Language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing.

[11] Hydrology and Water Management Center for Upper Northern Region Chiang Mai Thailand, 2013, Rainfall Data, Offce of Water Management and Hydrology Royal Irrigation Department Thailand, http://hydro-1.net.

[12] Gross, A., and Clark, V., 1975, Survival Distribution: Reliability Applications in the Biomedical Sciences, New York: John Wiley & Sons.