PERFORMANCE OF VARIOUS LOCATION-BASED LINDLEY FAMILY DISTRIBUTIONS BASED ON AKAIKE INFORMATION CRITERION

R. Tharshan[1,2*] and P. Wijekoon[3]

[1]Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka
[2]Department of Mathematics and Statistics, Faculty of Science, University of Jaffna, Jaffna, Sri Lanka
[3]Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka
[*]tharshan10684@gmail.com

In statistical literature, several modifications of the Lindley distribution (LD) have been introduced in terms of having the flexibility of LD's failure rate criteria. These modifications have been applied to several heterogeneous lifetime data sets and used to model the heterogeneous Poisson parameter. Herein, it is needed to examine the performance of these modifications for various datasets that may have various characteristics such as horizontal symmetry, measured by skewness (SK); tail-heaviness measured by excess kurtosis (EK); and dispersion, measured by Fano factor (mean-variance ratio). In this study, we compare the performance of various location-based Lindley family distributions based on the Akaike information criterion. To generate various characteristics of data sets, and accommodate most of the Lindly family distributions, we propose a new location-based generalized Lindley distribution (NLGLD). This is a two-component mixture of exponential (θ, β) and gamma (γ, θ, β) with mixing proportion, $p = \frac{\delta\alpha}{\delta\alpha+\eta}$. Here, α, δ, η are shape parameters, and θ, β are scale and location parameters, respectively. Several possible sub-models of NLGLD are fitted for various characteristic simulated data sets. It was found that sub-models with the mixing proportion, $p = \frac{\alpha^3}{\alpha^3+1}$ perform well for all different settings of γ and β. However, the inclusion of the location parameter β, in these types of distributions resists the flexibility to cover the higher Fano factor (FF) values. Further, among the location-based sub-models of NLGLD, the model at $\gamma = 2$ is recommended for a data set with higher horizontal symmetry (around $SK > 0.85$), tail-heaviness (around $EK > 1.32$), and various ranges of dispersions when compared with the location-based model at $\gamma = 3$. Below this range of SK and EK values, and around $FF \leq 8$, the location-based model at $\gamma = 3$ can perform well. However, the location-based models with $\gamma \geq 4$ are not suitable to fit for various characteristics of the data sets since they have converging issues for the unknown parameter estimation β, and poor performance when compared with simple models without having a location parameter.

Keywords: Finite mixture models, Failure rate, Fano factor, Lindley distribution