International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 2, February - 2022, Pages:88-91

Maximum Product Spacing Method for Estimation Parameter of Poisson-Exponential Distribution

Andi Isna Yunita[1], Nurtiti Sunusi[2], Erna Tri Herdiani[3]

[1,2,3]Departement of Statistics, Hasanuddin University, Makassar, 90245, Indonesia
andiisnayunita17@gmail.com

***Abstract**: The Poisson-Exponential (PE) distribution is a new distribution for lifetime data with two parameters and an increased hazard function. The PE distribution is a mixed distribution between the Exponential distribution and the zero truncated Poisson distribution. This distribution has an important position on the latent complementary risk problem scenario. PE distribution parameters are estimated using the Maximum Product Spacing (MPS) method with Newton Raphson iterations. The methodology is illustrated in data simulation.*

Keywords—Exponential distribution; zero-truncated Poisson distribution; maximum product spacing

1. INTRODUCTION

Survival analysis is an analysis of the survival of a unit or component under certain conditions [1]. Survival analysis can be used to model survival data. Survival data is divided into censored data and uncensored data. Censored data is data that cannot be observed completely. In survival analysis, there are two kinds of models, namely parametric models and non-parametric models. A parametric model is a survival model with data following a certain distribution. Commonly used survival distributions include Exponential, Weibull, Log-Normal, and Gamma distributions [2]. A new distribution introduced by Kus [3] is the Poisson-Exponential distribution as a model for lifetime data with an increasing hazard function.

The Poisson-Exponential (PE) distribution is a mixed distribution between the Exponential distribution and a zero-truncated Poisson distribution and. Let $Y = \min\{T_1, T_2, .., T_M\}$ where M is a random variable with zero-truncated Poisson distribution and T_i are independent of M and assumed to be independent and identically distributed according to an Exponential distribution, then the random variable Y has PE distribution [3]. Meanwhile, Cancho et al. [4] follows an opposite way and defined the component as $Y = \max\{T_1, T_2, .., T_M\}$ under the same assumptions so that Y has a Complementary PE (CEP) distribution. However, the CEP distribution was then used more in subsequent research as the PE distribution. For example, research by Louzada et al. [5] and Kumar et al. [6] applied these to a data set taken from Collett [7] is data on the survival times of 26 women diagnosed with ovarian cancer submitted to chemotherapy. In addition, there is also research that uses the PE distribution on censored data, such as research by Belaghi et al. [8] on type-II censored data and Monfared et al. [9] on type-I censored data.

In their research, Belaghi et al. [8] and Monfared et al. [9] estimated the parameters of PE distribution using the Maximum Likelihood Estimation (MLE) method. The MLE is the method most often used if the data follows a certain distribution. The MLE method is carried out by maximizing the likelihood function. Another research by Cancho et al. [4] used the Expectation-Maximization (EM) algorithm to complete the estimation of PE parameters using MLE.

However, the MLE does not always give satisfactory estimates of parameters in more general conditions. The method used to solve this problem is the Maximum Product Spacing (MPS). MPS is an alternative method of MLE to estimate parameters of continuous univariate distribution [10]. This method was introduced by Cheng and Amin [10] who proved that the MPS method is efficient and consistent compared to MLE under more general conditions. MPS was also derived independently by Ranneby [11] based on a simple distance from Kullback-Leibler information. Using MPS was also found in research by Almetwally and Almongy [12] which estimated the parameters of Generalized Power Weibull (GPW) distribution under type-II progressive censored data. Research by Alshenawy et al. [13] estimates the reliability parameters of the Exponentiated-Gumbel distribution under type-II progressive censored data. Yalçınkaya et al. [14] estimated the Skew-Normal distribution parameters under doubly type-II censored data.

This article is organized as follows. Section 2 presents the properties of the PE distribution. Section 3 explains the estimation parameter of PE distribution using the MPS method with Newton Raphson's numerical method to solve the implicit function. Section 4 provides conclusions related to the results of this research. Furthermore, the development of Poisson-exponential distribution theory is expected to be a reference for use of data analysis methods, especially on censored data.

2. DISTRIBUSI POISSON-EXPONENTIAL

Following Cancho et al. [4], the PE model can be derived as follows. Let M be a random variable denoting the number of complementary risks (CR) related to the occurrence of an event of interest. Further, assume that M has a zero truncated Poisson distribution with probability mass function given by

$$p(m;\theta) = \frac{\theta^m e^{-\theta}}{m!\,(1-e^{-\theta})}, \quad m = 1,2,\ldots;\ \theta > 0 \quad (1)$$

Let $T_j,\ j = 1,2,\ldots,m$ denote the time-to-event due to the jth CR, hereafter lifetime. Given $M = m$, the random variables are assumed to be independent and identically distributed according to an Exponential distribution with probability density function given by

$$f(t;\lambda) = \lambda e^{-\lambda t}, \qquad t > 0;\ \lambda > 0 \qquad (2)$$

In the latent CR scenario, the number of causes M and the lifetime T_j associated with a particular cause are not observed, but only the maximum lifetime Y among all causes is usually observed. So, the component lifetime is defined as

$$Y = \max(T_1, T_2, \ldots, T_M) \qquad (3)$$

Therefore, if the random variable Y is defined as in Equation (3), then considering Equation (1) and (2), Y is distributed according to a Poisson-exponential (PE) distribution, denoted with $PE(\theta,\lambda)$ and has probability density function given by

$$f(y;\theta,\lambda) = \frac{\theta\lambda e^{-\lambda y-\theta e^{-\lambda y}}}{1-e^{-\theta}}, \qquad y > 0 \qquad (4)$$

Where $\theta > 0$ and $\lambda > 0$ is shape and scale parameters. When θ approached zero, the PE distribution converges to an exponential distribution with parameter λ. The parameter PE distribution has a direct interpretation in terms of complementary risks. The parameter θ represents the mean of the number of complementary risks, while parameter λ denoted the lifetime failure rate [15].

Mean and variance of the PE distribution are given, respectively, by

$$E[Y] = \frac{\theta}{\lambda(1-e^{-\theta})} F_{2,2}([1,1],[2,2],-\theta)$$

$$Var(Y) = \frac{\theta}{\lambda^2(1-e^{-\theta})}\left\{F_{3,3}([1,1,1],[2,2,2],-\theta) - \frac{\theta}{1-e^{-\theta}} {F_{2,2}}^2([1,1],[2,2],-\theta)\right\}$$

where $F_{p,q}(\boldsymbol{a},\boldsymbol{b},\theta)$ is a generalized hypergeometric function given by

$$F_{p,q}(\boldsymbol{a},\boldsymbol{b},\theta) = \sum_{j=0}^{\infty} \frac{\theta^j \prod_{i=1}^{p} \Gamma(a_i+j)\Gamma(a_i)^{-1}}{\Gamma(j+1)\prod_{i=1}^{p}\Gamma(b_i+j)\Gamma(b_i)^{-1}}$$

where $\boldsymbol{a} = [a_1, a_2, \ldots, a_p]$ and $\boldsymbol{b} = [b_1, b_2, \ldots, b_q]$. It should be stressed that vector $\boldsymbol{a}$ and $\boldsymbol{b}$ have variable dimensions in the above formulas. Meanwhile, the cumulative density function of Y is given by

$$F(y;\theta,\lambda) = \frac{1-e^{\theta-\theta e^{-\lambda y}}}{1-e^{\theta}} \qquad (5)$$

The survival (or reliability) and hazard function of the $PE(\theta,\lambda)$ distribution given by

$$S(y) = \frac{1-e^{-\theta e^{-\lambda y}}}{1-e^{-\theta}} \quad \text{dan} \quad h(y) = \frac{\theta\lambda e^{-\lambda y-\theta e^{-\lambda y}}}{1-e^{-\theta e^{-\lambda y}}}$$

3. RESULT

3.1 Estimation of Poisson-Exponential Distribution Parameters Using Maximum Product Spacing

The parameter of PE distribution to be estimated is θ and λ using the maximum product spacing (MPS) method. The MPS method estimates parameters by maximizing the geometric mean of the spacings. To estimate the PE distribution parameters, the MPS method requires the cumulative distribution function of the PE distribution as Equation (5).

The first step of the MPS method in estimating parameters is to calculate the uniform distance of a random sample from the PE distribution given by

$$D_i(\theta,\lambda) = F(y_i;\theta,\lambda) - F(y_{i-1};\theta,\lambda), \ \ i = 1,2,\ldots,n \quad (6)$$

where

$$D_i(\theta,\lambda) = \begin{cases} D_1 = F(y_1;\theta,\lambda) \\ D_i = F(y_i;\theta,\lambda) - F(y_{i-1};\theta,\lambda), & i = 1,2,\ldots,n \\ D_{n+1} = 1 - F(y_n;\theta,\lambda) \end{cases}$$

Clearly $\sum_{i=1}^{n+1} D_i(\theta,\lambda) = 1$.

The maximum product of spacing estimates parameter distribution PE, $\hat{\theta}_{MPS}$ dan $\hat{\lambda}_{MPS}$ of the parameter θ dan λ are obtained by maximizing $G(\theta,\lambda)$ function. $G(\theta,\lambda)$ function is the geometric mean of the spacing given by

$$G(\theta,\lambda) = \left\{\prod_{i=1}^{n+1} D_i(\theta,\lambda)\right\}^{\frac{1}{n+1}} \qquad (7)$$

or, equivalently, by maximizing the function

$$g(\theta,\lambda) = \ln G(\theta,\lambda) = \ln\left\{\prod_{i=1}^{n+1} D_i(\theta,\lambda)\right\}^{\frac{1}{n+1}}$$

$$g(\theta,\lambda) = \frac{1}{n+1}\sum_{i=1}^{n+1} \ln D_i(\theta,\lambda) \qquad (8)$$

The estimates $\hat{\theta}$ and $\hat{\lambda}$ can be obtained by maximizing the function $g(\theta,\lambda)$ i.e. the function $g(\theta,\lambda)$ in Equation (8) is derived concerning θ and λ then solving the nonlinear equation

$$\frac{\partial g(\theta,\lambda)}{\partial\theta}=\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(\theta,\lambda)}\left[\frac{\partial F(y_i;\theta,\lambda)}{\partial\theta}-\frac{\partial F(y_{i-1};\theta,\lambda)}{\partial\theta}\right]=0$$

$$\frac{\partial g(\theta,\lambda)}{\partial\lambda}=\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(\theta,\lambda)}\left[\frac{\partial F(y_i;\theta,\lambda)}{\partial\lambda}-\frac{\partial F(y_{i-1};\theta,\lambda)}{\partial\lambda}\right]=0$$

... (9)

where

$$\frac{\partial F(y_i;\theta,\lambda)}{\partial\theta}=\frac{e^{\theta-\theta e^{-\lambda y_i}-\lambda y_i}}{(e^\theta-1)^2}\left[e^{\theta e^{-\lambda y_i}+\lambda y_i}-e^{\lambda y_i}-e^\theta+1\right]$$

$$\frac{\partial F(y_i;\theta,\lambda)}{\partial\lambda}=\frac{\theta y_i e^{\theta-\theta e^{-\lambda y_i}-\lambda y_i}}{e^\theta-1}$$

Because Equation (9) is a nonlinear or implicit equation and difficult to solve explicitly, the Newton-Raphson method is used as an iteration method to solve the equation. Based on the Newton-Raphson method, the second derivative of Equation (8) is obtained for each parameter. The Newton-Raphson iteration method in estimating parameters, $\hat{\boldsymbol{\phi}}=[\hat{\theta}\quad\hat{\lambda}]^T$ can be given by

$$\hat{\boldsymbol{\phi}}^{(\boldsymbol{m}+\mathbf{1})}=\hat{\boldsymbol{\phi}}^{(\boldsymbol{m})}-\left[\boldsymbol{H}\left(\hat{\boldsymbol{\phi}}^{(\boldsymbol{m})}\right)\right]^{-\mathbf{1}}\cdot\boldsymbol{I}\left(\hat{\boldsymbol{\phi}}^{(\boldsymbol{m})}\right)$$

where $\boldsymbol{I}\left(\hat{\boldsymbol{\phi}}^{(\boldsymbol{m})}\right)$ is the first derivative matrix for each parameter as in Equation (9) and $\boldsymbol{H}\left(\hat{\boldsymbol{\phi}}^{(\boldsymbol{m})}\right)$ is the second derivative for each parameter given by

$$\boldsymbol{H}\left(\hat{\boldsymbol{\phi}}^{(\boldsymbol{m})}\right)=\begin{bmatrix}\frac{\partial^2 g(\theta,\lambda)}{\partial\theta^2} & \frac{\partial^2 g(\theta,\lambda)}{\partial\theta\,\partial\lambda}\\ \frac{\partial^2 g(\theta,\lambda)}{\partial\theta\,\partial\lambda} & \frac{\partial^2 g(\theta,\lambda)}{\partial\lambda^2}\end{bmatrix}$$

where

$$\frac{\partial^2 g(\theta,\lambda)}{\partial\theta^2}=\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{[D_i(\theta,\lambda)]^2}\left\{[D_i(\theta,\lambda)]\left[\frac{\partial^2 D_i(\theta,\lambda)}{\partial\theta^2}\right]-\left[\frac{\partial D_i(\theta,\lambda)}{\partial\theta}\right]^2\right\}$$

$$\frac{\partial^2 g(\theta,\lambda)}{\partial\lambda^2}=\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{[D_i(\theta,\lambda)]^2}\left\{[D_i(\theta,\lambda)]\left[\frac{\partial^2 D_i(\theta,\lambda)}{\partial\lambda^2}\right]-\left[\frac{\partial D_i(\theta,\lambda)}{\partial\lambda}\right]^2\right\}$$

$$\frac{\partial^2 g(\theta,\lambda)}{\partial\theta\,\partial\lambda}=\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{[D_i(\theta,\lambda)]^2}\left\{D_i(\theta,\lambda)\left[\frac{\partial^2 D_i(\theta,\lambda)}{\partial\theta\,\partial\lambda}\right]-\left[\frac{\partial D_i(\theta,\lambda)}{\partial\theta}\right]\left[\frac{\partial D_i(\theta,\lambda)}{\partial\lambda}\right]\right\}$$

The iteration will stop when the parameters have converged, i.e. when $\left\|\hat{\boldsymbol{\phi}}^{(\boldsymbol{m}+\mathbf{1})}-\hat{\boldsymbol{\phi}}^{(\boldsymbol{m})}\right\|\le\varepsilon$, where ε is a very small error value.

4. CONCLUSION

In this paper, we present the Poisson-Exponential (PE) distribution and its properties, such as probability density function, the mean and variance, cumulative density function, until survival and hazard function. The PE distribution has two parameters, namely θ and λ. Both parameters have direct interpretation in terms of CR. The parameter θ represents the mean of the number of complementary risks, while parameter λ denoted the lifetime failure rate. Both parameters are estimated using the maximum product spacing (MPS) method with Newton-Raphson iteration. We hope that the application of this method can be used in real data to get more accurate estimation results.

5. REFERENCES

[1] Kleinbaum, D. G. &. K. M. (2005). Survival Analysis: A Self-Learning Text, 2nd ed. New York: Springer Science Business Media, Inc.

[2] Lee, E. T. & Wang, J. W. (2003). Statistical Method for Survival Data Analysis, 3rd ed. Oklahoma: John Wiley & Sons.

[3] Kus, C., (2007). A new lifetime distribution. *Journal of Computational Statistics & Data Analysis*, pp. 4497-4509.

[4] Cancho, V. G., Louzada-Neto, F., & Barriga, G. D. (2011). The Poisson-Exponential Lifetime Distribution. *Journal of Computational Statistics and Data Analysis*, pp. 677-686.

[5] Louzada, F., Cancho, V. G., & Barriga, G. D. (2012). The Poisson-exponential regression model under different latent activation schemes. *Journal of Computational & Applied Mathematics*, pp. 31(3): 617-632.

[6] Kumar, M., Singh, S. K., & Singh, U. (2014). Bayesian estimation for Poisson-exponential model under Progressive Type-II censoring data with binomial removal and its application to ovarian cancer data. *Journal of Communications in Statistics - Simulation and Computation.*

[7] Collet, D. (2003). Modeling Survival Data in Medical Research, 2nd ed. London: Chapmann & Hall.

[8] Belaghi, R. A. et al. (2018). Estimation and Prediction for the Poisson Exponential Distribution Based on Type-II Censored Data. *American Journal of Mathematical and Management Sciences*, pp. 96-115.

[9] Monfared, M. M., Belaghi, R. A., Behzadi, M. H., & Singh, S. (2019). Estimation and prediction based on type-I hybrid censored data from the Poisson-Exponential distribution. *Journal of Communications in Statistics - Simulation and Computation*.

[10] Cheng, R. C. H. & K. Amin, N. A. (1983). Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin. *Journal of Royal Statistical Society*, pp. 394-403.

[11] Ranneby, B. (1984). The Maximum Spacing Method. An Estimation Method Related to the Maximum Likelihood Method. *Scandinavian Journal of Statistics*, pp. 93-112.

[12] Almetwally, E. M. & Almongy, H. M. (2019). Maximum Product Spacing and Bayesian Method for Parameter Estimation for Generalized Power Weibull Distribution Under Censoring Scheme. *Journal of Data Science*, pp. 407-444.

[13] Alshenawy, R., Sabry, M. A. H., Almetwally, E. M., & Almongy, H. M. (2021). Product Spacing of Stress–Strength under Progressive Hybrid Censored for Exponentiated-Gumbel Distribution. *Journal of Computers, Materials & Continua*, pp. 2973-2995.

[14] Yalçınkaya, A., Yolcu, U., & Şenoglu, B. (2021). Maximum likelihood and maximum product of spacings estimations for the parameters of skew-normal distribution under doubly type II censoring using genetic algorithm. *Journal of Expert Systems With Applications*, pp. 114407.

[15] Tomazella, V. L., Cancho, V. G. & Louzada, F. (2013. Bayesian reference analysis for the Poisson-exponential lifetime distribution. *Chilean Journal of Statistics*, p. 99-113.