Poisson-Beta Exponential Distribution: Properties and Applications

Eisa Mahmoudi[†,*], Hossein Zamani[‡] and RahmatSadat Meshkat[†]

[†] Yazd University
[‡] Hormozgan University

Abstract. A new generalized version of the mixed Poisson distribution, called the Poisson-beta exponential (PBE) distribution, is obtained by mixing the Poisson and the beta exponential (BE) distributions. Estimation of the parameters, using the method of moments and maximum likelihood estimators, is discussed. We show the consistency of the new model parameters using simulation study. Examples are given for fitting the PBE distribution to data, and the fit model is compared with that obtained using other distributions.

Keywords. Beta exponential distribution; mixed distributions; Poisson mixtures; truncated distributions; weighted distributions.

MSC 2010: 60E05; 62H10; 62H12.

1 Introduction

Count data are often encountered in real world applications. In recent years, many discrete distributions have been developed for modelling count data. Poisson, generalized Poisson (GP) and negative binomial (NB) are the commonly distributions. Poisson distribution has the same mean and variance,

---

[*] Corresponding author

thus it is equi-dispersion. However, count data often exhibit under-dispersion our over-dispersion according to the case where the sample variance is substantially in excess of the sample mean or vice versa. One way to describe the dispersion characteristic is to introduce an additional parameter into Poisson distribution or handling over or under-dispersion phenomena.

During the last decade, a vast activity has been observed in generalizing of the classical discrete distributions. The main idea was to apply the extended versions for modelling different kinds of dependent count or frequency data structure in various fields. In literature, several extensions of the Mixed Poisson distribution have been proposed such as

- Poisson-Gamma (Negative Binomial) presented by Greenwood and Yule (1920).

- Poisson Beta with specific parameter values presented by Simon (1955).

- Poisson Beta Type 2 presented by Gurland (1958).

- Poisson-Exponential Beta presented by Pielou (1962).

- Poisson Truncated Poisson presented by Patil (1964).

- Poisson Beta Type 1 presented by Holla and Bhattacharya (1965).

- Poisson Truncated Gamma presented by Bhattacharya (1966).

- Poisson Linear Exponential presented by Sankaran (1969).

- Poisson Lindley presented by Sankaran (1970).

- Poisson Power Function presented by Rai (1971).

- Poisson Lognormal presented by Bulmer (1974).

- Poisson Generalized Inverse Gaussian presented by Sichel (1974).

- Poisson Inverse Gaussian presented by Sichel (1975).

- Poisson Gamma Product Ratio (Generalized Waring) presented by Irwin (1975).

- Poisson Generalized Pareto presented by Kempton (1975).

- Poisson-Poisson Distribution (Neyman) presented by Douglas (1980).

- Poisson Pearson's Family of Distribution presented by Albrecht (1982).

- Poisson Generalized Gamma presented by Albrecht (1984).

- Poisson Truncated Beta Type 2 presented by Willmot (1986).

- Poisson Log-Student presented by Gaver and O' Muircheartaigh (1987).

- Poisson Shifted Gamma presented by Ruohonen (1988).

- Poisson Linear Exponential presented by Kling and Goovaerts (1993).

- Poisson Inverse Gamma presented by Willmot (1993).

- Poisson Truncated Gamma presented by Willmot (1993).

- Poisson Pareto presented by Willmot (1993).

- Poisson Shifted Pareto presented by Willmot (1993).

- Poisson Modified Bessel presented by Ong and Muthaloo (1995).

- Poisson-Power Variance presented by Hougard (1997).

- Poisson-Lomax presented by Al-Awadhi and Ghitany (2001).

- Poisson Exponential (Geometric) presented by Johnson et al. (2005).

- Poisson-Other Discrete Distribution (Neyman) presented by Johnson et al. (2005).

- Poisson-generalized gamma, -generalized inverse Gaussian, -general exponential, -lognormal and -inverse gamma presented by Gupta and Ong (2005).

- Poisson Lindley Distribution (Size Biased) presented by Ghitany and Al-Mutairi (2008).

- Zero Truncated Poisson-Lindley presented by Ghitany et al. (2008).

- Weighted generalized Poisson presented by Chakraborty (2010).

- Poisson Generalized-Lindley presented by Mahmoudi and Zakerzadeh (2010).

- Poisson-Marshall-Olkin-Generalized Exponential presented by Gómez-Déniz and Calderín-Ojeda (2015).

- Poisson Lindley-Beta prime presented by Gómez-Déniz et al. (2016).

This paper offers a two-parameter PBE distribution, as a model for count data. In our opinion, this new distribution offers several advantages. (i) The present distribution provides enough flexibility for analysing different types of count data in comparing with Poisson, NB, PL and GPL distributions. (ii) It has a natural interpretation in terms of zero-inflation. (iii) The proposed new discrete distribution is candidates for modelling of correlated count data which exhibit over-dispersion.

This study examines various properties of this model. The paper is organized as follows:
Section 2, gives a brief review on BE distribution and studies some properties of this distribution. Section 3 provides the definition of the PBE distribution and represents its basic properties including: the behaviour of the density function, the expressions for the moments, the distribution of the sum of random variables and truncated and weighted PBE distributions. Section 4, discusses the different methods for parameters estimation. In Section 5, we present two algorithms for simulating random data from the PBE distribution. In this section, a simulation study is carried out to investigate the average bias and average mean square error (MSE) of the simulated estimators. An application of the PBE distribution, including two examples of fitting PBE distribution to real data, and comparing with other discrete distributions, is given in Section 6. Finally, Section 7 concludes the paper.

2   Beta Exponential Distribution: Brief Review

The pdf of the $Beta(\beta, \alpha)$ distribution is given by

$$g(z) = \frac{1}{B(\beta, \alpha)} z^{\beta-1}(1 - z)^{\alpha-1}, \qquad 0 < z < 1,$$

where $B(\beta, \alpha)$ denotes the beta function. The cdf and pdf of BE distribution which is introduced and analysed by Nadarajah and Kotz (2006) is given by

$$F(x) = \frac{1}{B(\beta, \alpha)} \int_0^{1-e^{-\lambda x}} t^{\beta-1}(1 - t)^{\alpha-1} dt,$$

and

$$f(x) = \frac{\lambda}{B(\beta, \alpha)} e^{-\alpha\lambda x} (1 - e^{-\lambda x})^{\beta-1}, \qquad x > 0, \ \alpha, \beta, \lambda > 0.$$

Putting $\beta = 2$ in the above formula, we have the pdf of $BE(2, \alpha, \lambda)$ distribution as

$$f(x) = \alpha(\alpha + 1)\lambda e^{-\alpha\lambda x} (1 - e^{-\lambda x}), \qquad x > 0, \ \alpha, \lambda > 0.$$

The survival and hazard rate functions of BE distribution are given respectively by

$$S(x) = 1 - I_{e^{-\lambda x}}(\alpha, 2),$$

and

$$h(x) = \frac{\alpha(\alpha + 1)\lambda e^{-\alpha\lambda x}(1 - e^{-\lambda x})}{B_{1-e^{-\lambda x}}(2, \alpha)},$$

where $I_x(2, \alpha) = \frac{B_x(2,\alpha)}{B(2,\alpha)}$ with $B_x(2, \alpha) = \int_0^x w(1 - w)^{\alpha-1} dw$ denotes the incomplete beta function ratio. The mgf of $BE(\alpha, \lambda)$ is given by

$$M(t) = \frac{\alpha(\alpha + 1)\lambda^2}{(\alpha\lambda - t)((\alpha + 1)\lambda - t)} = \frac{1}{\left(1 - \frac{t}{\alpha\lambda}\right)} \frac{1}{\left(1 - \frac{t}{(\alpha+1)\lambda}\right)}.$$

It is easy to verify that $X = U + V$ where $U \sim \exp(\alpha\lambda)$ and $V \sim \exp((\alpha+1)\lambda)$. By differentiating from $\log(M(t))$ and setting $t = 0$, the first three central moments of $X \sim BE(\alpha, \lambda)$ are given by

$$E[X] = \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha + 1}\right),$$
$$E[(X - \mu)^2] = \frac{1}{\lambda^2}\left(\frac{1}{\alpha^2} + \frac{1}{(\alpha + 1)^2}\right),$$
$$E[(X - \mu)^3] = \frac{2}{\lambda^3}\left(\frac{1}{\alpha^3} + \frac{1}{(\alpha + 1)^3}\right).$$

The kth central moment of X can be calculated as

$$E[(X - \mu)^k] = \frac{(k - 1)!}{\lambda^k}\left(\frac{1}{\alpha^k} + \frac{1}{(\alpha + 1)^k}\right).$$

3   Definition and Some Properties of the PBE Distribution

Suppose that the random variable $Y|\theta$ follows the Poisson distribution with parameter θ and $\theta \sim BE(\alpha, \lambda)$, then the mixed Poisson distribution can be given by

$$
\begin{aligned}
P(y; \alpha, \theta) &= \int_0^\infty \frac{e^{-\theta}\theta^y}{y!}\alpha(\alpha+1)\lambda e^{-\alpha\lambda\theta}(1-e^{-\lambda\theta})d\theta \\
&= \alpha(\alpha+1)\lambda\big[(1+\alpha\lambda)^{-(y+1)} - (1+(\alpha+1)\lambda)^{-(y+1)}\big], \quad y = 0, 1, \ldots,
\end{aligned}
\tag{1}
$$

where $\lambda > 0$ and $\alpha > 0$. We call this distribution as the PBE distribution and denote it by $PBE(\alpha, \lambda)$. Note that the pdf of the PBE can be expressed in terms of sum of two geometric random variables.

Proposition 1. The PBE distribution has a unique mode at $x^* = \lfloor y \rfloor$, where

$$
y = \frac{\log(1+\lambda(\alpha+1)) + \log\left[\frac{\log(1+\alpha\lambda)}{(1+\alpha\lambda)\log(1+\lambda(\alpha+1))}\right]}{\log(1+\alpha\lambda) - \log(1+\lambda(\alpha+1))},
$$

when $\lfloor . \rfloor$ denotes the smallest integer $\geqslant x^*$.

Proof. Mode is the point which maximize the pmf or pdf, thus we must search a point for which $\frac{d}{dy}f(y;\alpha,\lambda) = 0$. The derivative of $f(y;\alpha,\lambda)$ w.r.t. y is given by

$$
\begin{aligned}
\frac{d}{dy}f(y;\alpha,\lambda) = \alpha\lambda(\alpha+1)[&\log(1+\lambda(\alpha+1))(1+\lambda(\alpha+1))^{-(y+1)} \\
&- \log(1+\alpha\lambda)(1+\alpha\lambda)^{-(y+1)}].
\end{aligned}
$$

Solving $\frac{d}{dy}f(y;\alpha,\lambda) = 0$ gives $\frac{\log(1+\lambda(\alpha+1))}{\log(1+\alpha\lambda)} = \left(1 + \frac{\lambda}{1+\alpha\lambda}\right)^{y+1}$. Thus we have

$$
y = \frac{\log(1+\lambda(\alpha+1)) + \log\left[\frac{\log(1+\alpha\lambda)}{(1+\alpha\lambda)\log(1+\lambda(\alpha+1))}\right]}{\log(1+\alpha\lambda) - \log(1+\lambda(\alpha+1))}.
$$

Since y takes only the integer values, thus the mode is equal to $x^* = \lfloor y \rfloor$ and the proof is completed. $\square$

***Theorem* 1.** *The cumulative distribution function of the PBE distribution*

is given by

$$F(y; \alpha, \lambda) = (\alpha + 1)(1 - (1 + \alpha\lambda)^{-(y+1)}) - \alpha(1 - (1 + (\alpha + 1)\lambda)^{-(y+1)}), \quad (2)$$

for $y = 0, 1, 2, \ldots$.

Proof. If Y has $PBE(\alpha, \lambda)$ distribution, then its cumulative distribution function is obtained as

$$
\begin{aligned}
F(y; \alpha, \lambda) &= P(Y \leqslant y) \\
&= \sum_{t=0}^{y} \alpha(\alpha + 1)\lambda \left[(1 + \alpha\lambda)^{-(t+1)} - (1 + (\alpha + 1)\lambda)^{-(t+1)} \right] \\
&= \alpha(\alpha + 1)\lambda \left[\sum_{t=0}^{y} (1 + \alpha\lambda)^{-(t+1)} - \sum_{t=0}^{y} (1 + (\alpha + 1)\lambda)^{-(t+1)} \right] \\
&= \alpha(\alpha + 1)\lambda \left[\frac{1}{\alpha\lambda} \left(1 - (1 + \alpha\lambda)^{-(y+1)} \right) \right. \\
&\quad \left. - \frac{1}{(\alpha + 1)\lambda} \left(1 - (1 + (\alpha + 1)\lambda)^{-(y+1)} \right) \right] \\
&= (\alpha + 1)(1 - (1 + \alpha\lambda)^{-(y+1)}) - \alpha(1 - (1 + (\alpha + 1)\lambda)^{-(y+1)}).
\end{aligned}
$$

$\square$

The graphs of pmf of the PBE distribution for $\lambda = 2$ and different values α are given in the Figure 1.

3.1 Moments, Skewness and Kurtosis

If Y has $PBE(\alpha, \lambda)$ distribution, then the direct calculation shows that the corresponding moment generating function (mgf) and the probability generating function (pgf) can be defined respectively as

$$M_Y(t) = M_\theta(e^t - 1) = \left(\frac{\alpha\lambda}{\alpha\lambda + 1 - e^t} \right) \left(\frac{(\alpha + 1)\lambda}{(\alpha + 1)\lambda + 1 - e^t} \right),$$

and

$$G_Y(t) = \left(\frac{\alpha\lambda}{\alpha\lambda + 1 - t} \right) \left(\frac{(\alpha + 1)\lambda}{(\alpha + 1)\lambda + 1 - t} \right).$$
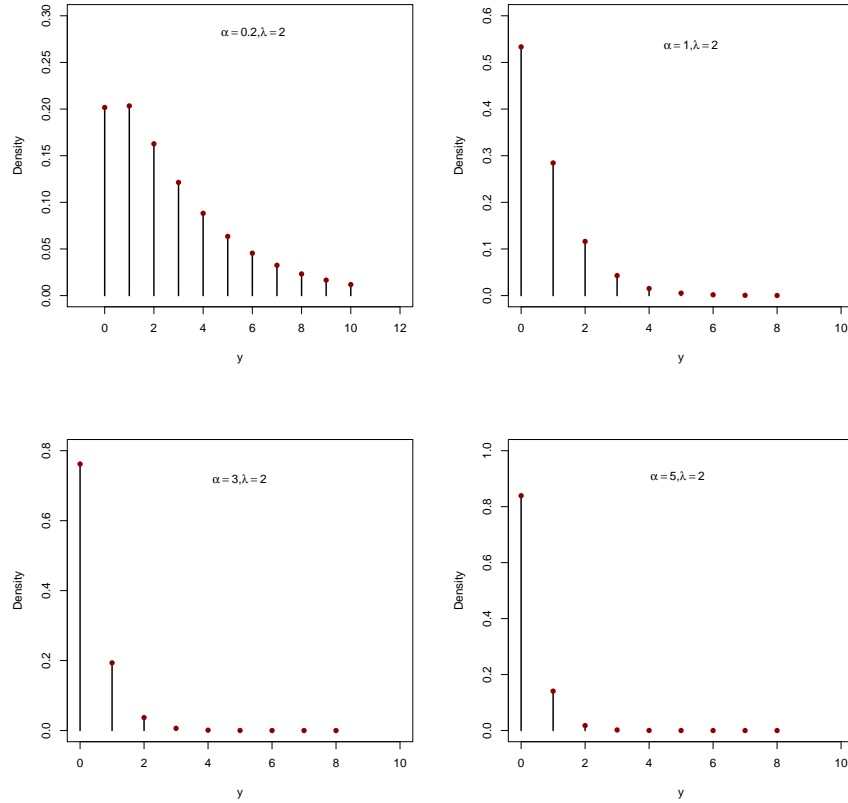
Figure 1. Graphs of pmf of PBE distribution for different values (α, λ).

Using $M_Y(t)$ and $G_Y(t)$ in the above formula we see that the random variable Y can be written as the sum of two random variables W and Z, where $W \sim Ge(\frac{\alpha\lambda}{\alpha\lambda+1})$ and $Z \sim Ge(\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1})$. We have

$$\log M_Y(t) = \log(\alpha(\alpha+1)\lambda^2) - \log(\alpha\lambda + 1 - e^t) - \log((\alpha+1)\lambda + 1 - e^t).$$

The four cumulants of Y can be given by differentiating $\log M_Y(t)$ w.r.t. t and setting $t = 0$. We have

$$\mu = E[Y] = \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right),$$

$$\mu_2 = \sigma^2 = E[Y-\mu]^2 = \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right) + \frac{1}{\lambda^2}\left(\frac{1}{\alpha^2} + \frac{1}{(\alpha+1)^2}\right),$$

$$\mu_3 = E[Y-\mu]^3 = \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right) + \frac{3}{\lambda^2}\left(\frac{1}{\alpha^2} + \frac{1}{(\alpha+1)^2}\right) + \frac{2}{\lambda^3}\left(\frac{1}{\alpha^3} + \frac{1}{(\alpha+1)^3}\right),$$

$$\mu_4 = E[Y-\mu]^4 = \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right) + \frac{7}{\lambda^2}\left(\frac{1}{\alpha^2} + \frac{1}{(\alpha+1)^2}\right) + \frac{12}{\lambda^3}\left(\frac{1}{\alpha^3} + \frac{1}{(\alpha+1)^3}\right)$$
$$+ \frac{6}{\lambda^4}\left(\frac{1}{\alpha^4} + \frac{1}{(\alpha+1)^4}\right).$$

Since the index of dispersion is given by

$$r = \frac{\sigma^2}{\mu} = 1 + a\mu,$$

where $a = \frac{2\alpha^2+2\alpha+1}{4\alpha^2+4\alpha+1} > 0$, it follows that the $PBE(\alpha, \lambda)$ is suitable for fitting over-dispersed count data with variance $\sigma^2 = \mu + a\mu^2$, $0.5 < a = \frac{1}{2} + \frac{1}{8\alpha^2+8\alpha+2} < 1$. Other important indices of the shape of the distribution are the skewness, $\sqrt{\beta_1} = \frac{\mu_3}{(\mu_2)^{3/2}}$, kurtosis, $\beta_2 = \frac{\mu_4}{(\mu_2)^2}$, and the coefficient of variation, $C.V. = \frac{\sigma}{\mu}$, respectively. For the PBE distribution these indices are given by

$$\sqrt{\beta_1} = \frac{(1 + \alpha(2 + \lambda + \alpha\lambda))(2 + \alpha(1 + \alpha)(2 + \lambda + 2\alpha\lambda))}{(1 + \alpha(1 + \alpha)(2 + \lambda + 2\alpha\lambda))^{3/2}},$$

$$\beta_2 = \frac{6 + \alpha(1 + \alpha)(4 + \alpha(1 + \alpha)(2 + \lambda + 2\alpha\lambda))(6 + \lambda(3 + \alpha(6 + \lambda + \alpha\lambda)))}{(1 + \alpha(1 + \alpha)(2 + \lambda + 2\alpha\lambda))^2},$$

and

$$C.V. = \frac{\sqrt{(1 + \alpha(1 + \alpha)(2 + \lambda + 2\alpha\lambda))}}{1 + 2\alpha}.$$

The graphs of the skewness, kurtosis, dispersion and coefficient of variation of PBE distribution for different values α and λ are shown in Figure 2. The interpretation of these graphs are as below:

(i)   If $\alpha \to 0$, the skewness, kurtosis and coefficient of variation are close to 2, 6 and 1, respectively for each value of λ.

(ii) If $\lambda \to 0$ and $\alpha \to \infty$ the skewness, kurtosis and coefficient of variation
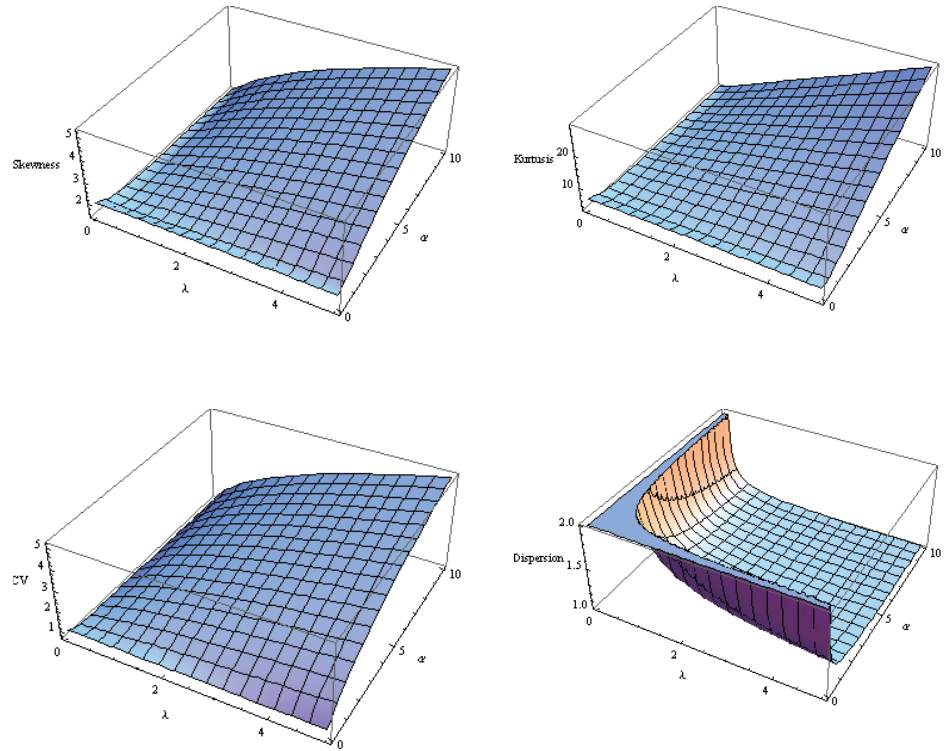
Figure 2. Graphs of skewness, kurtosis, coefficient of variation and dispersion of PBE distribution as a function of (α, λ).

are close to $\sqrt{2}$, 3 and $\sqrt{2}/2$, respectively.

(iii) As λ and α increase, the skewness, kurtosis and coefficient of variation are also increase.

Here, we give the distribution of sum of random variables having the PBE distribution. Truncated and weighted versions of the PBE distribution is presented in the next section.

3.2   Distribution of Sum of the PBE Random Variables

Let $f_{NB}(x; k, p)$ denotes the pmf of the negative binomial distribution defined by, $f(x; k, p) = \binom{x+k-1}{x} p^k (1-p)^x$, $x = 0, 1, \ldots$, $k > 0$ and $0 < p < 1$. It

is well known that the distribution of sum of independent negative binomial random variables with the same probability of success is again a negative binomial distribution. The following result shows that the distribution of sum of independent and identical random variables from PBE distribution, can be written as the sum of two independent negative binomial distributions.

Proposition 2. Let $X_1, \ldots, X_r$ denote r independent and identical random variables from $PBE(\alpha, \lambda)$. Then the mgf of $Z = \sum_{i=1}^{r} X_i$, is given by

$$M_Z(t) = \big(M_{X_1}(t)\big)^r = \Big(\frac{\alpha\lambda}{\alpha\lambda + 1 - e^t}\Big)^r \Big(\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda + 1 - e^t}\Big)^r = M_{Z_1}(t)M_{Z_2}(t),$$

where $Z_1 \sim NB(r, \frac{\alpha\lambda}{\alpha\lambda+1})$ and $Z_2 \sim NB(r, \frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1})$.

3.3 Truncated and Weighted Versions of the PBE Distribution

In the most common form of truncation, the zeroes are not recorded. In this case the zero-truncated distributions can be used as a distribution for the sizes of groups. This situation occurs in applications such as the number of offspring per family, the number of claims per claimant, and the number of occupants per car.

Proposition 3. The density function of the zero-truncated Poisson-beta exponential (ZTPBE) distribution is given by

$$f_{ZT}(y; \alpha, \lambda) = \frac{\alpha(\alpha+1)\lambda(1+(\alpha+1)\lambda)}{1+\lambda} \big[(1+\alpha\lambda)^{-(y+1)} - (1+(\alpha+1)\lambda)^{-(y+1)}\big],$$

for $y = 1, 2, \ldots$.

Proposition 4. The displaced Poisson-beta exponential (DPBE) distribution (left-truncated PBE distribution) is obtained by truncation of the first $r-1$ probabilities of a $PBE(\alpha, \lambda)$ distribution. It has the probability density function of the form

$$f_{LT}(y; \alpha, \lambda) = \frac{(1+\alpha\lambda)^{-(y+1)} - (1+(\alpha+1)\lambda)^{-(y+1)}}{\frac{(1+\alpha\lambda)^{-r}}{\alpha\lambda} - \frac{(1+(\alpha+1)\lambda)^{-r}}{(\alpha+1)\lambda}}, \quad y = r, r+1, \ldots.$$

Proposition 5. The right-truncated Poisson-beta exponential (RTPBE) distribution has support $0, 1, 2, \ldots, m$ and it's probability density function is given by

$$f_{RT}(y; \alpha, \lambda) = \frac{(1 + \alpha\lambda)^{-(y+1)} - (1 + (\alpha + 1)\lambda)^{-(y+1)}}{\frac{1-(1+\alpha\lambda)^{-m}}{\alpha\lambda} - \frac{1-(1+(\alpha+1)\lambda)^{-m}}{(\alpha+1)\lambda}}, \quad y = 0, 1, \cdots, m.$$

Very often, the distribution of recorded observations produced by a certain model may differ from that under the hypothesis model, mainly because of the recording mechanism that may employ unequal probabilities of recording the observations. The observed distribution will thus have a density of the form $f_w(x) = \frac{w(x)f(x)}{E[w(X)]}$, where $f(x)$ is the original density anticipated under the hypothesized model and $w(x)$ is a function proportional to the probability with which an observation n is recorded. These model were introduced by Rao (1965) and are known as weighted models. When the weighted function $w(x)$ is equal to x, these models are known as size-biased models. The observed distribution is then termed size biased distribution.

Proposition 6. Let Y is a random variable with $PBE(\alpha, \lambda)$ distribution. Then the size-biased PBE distribution (SBPBE) of Y is given by

$$f_{SB}(y; \alpha, \lambda) = \frac{(\alpha(\alpha + 1)\lambda)^2}{2\alpha + 1} \left[y(1 + \alpha\lambda)^{-(y+1)} - y(1 + (\alpha + 1)\lambda)^{-(y+1)} \right],$$

for $y = 1, 2, \cdots$.

3.4 Entropy

In information theory, entropy is a measure of the uncertainty associated with a random variable.

Theorem 2. *For the random variable Y having PBE distribution, the en-*

tropy is given by

$$H(Y) = -\sum_{y=0}^{\infty} \log P(y; \alpha, \lambda)$$

$$= \log(\alpha(\alpha+1)\lambda) - \log(1+\alpha\lambda)\left(1 + \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right)\right)$$

$$- \sum_{n=1}^{\infty} \frac{t^n}{n}\left(\frac{\alpha\lambda}{\alpha\lambda+1-t^n}\right)\left(\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1-t^n}\right),$$

where $t = \dfrac{1+\alpha\lambda}{1+(\alpha+1)\lambda}$.

Proof. Using this fact that the mean and probability generating function of Y are given, respectively, by

$$E[Y] = \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right),$$

$$\Psi_Y(t) = \left(\frac{\alpha\lambda}{\alpha\lambda+1-t}\right)\left(\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1-t}\right),$$

we can give an expression for $H(y)$. For simplicity, we first obtain an equation for $\log P(y; \alpha, \lambda)$.

$$\log P(y; \alpha, \lambda) = \log\left[\alpha(\alpha+1)\lambda\left((1+\alpha\lambda)^{-(y+1)} - (1+(\alpha+1)\lambda)^{-(y+1)}\right)\right]$$

$$= \log(\alpha(\alpha+1)\lambda)$$

$$+ \log\left((1+\alpha\lambda)^{-(y+1)}\left(1 - \left(\frac{1+\alpha\lambda}{1+(\alpha+1)\lambda}\right)^{-(y+1)}\right)\right)$$

$$= \log(\alpha(\alpha+1)\lambda) - (y+1)\log(1+\alpha\lambda)$$

$$+ \log\left(1 - \left(\frac{1+\alpha\lambda}{1+(\alpha+1)\lambda}\right)^{-(y+1)}\right).$$

Using this fact that the Taylor expansion of $\log(1-x)$ is

$$\log(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}, \qquad \text{for } |x| < 1,$$

we have

$$
\begin{aligned}
\log P(y; \alpha, \lambda) \;=\;& \log(\alpha(\alpha+1)\lambda) - (y+1)\log(1+\alpha\lambda) \\
& - \sum_{n=1}^{\infty} \left(\frac{\left(\dfrac{1+\alpha\lambda}{1+(\alpha+1)\lambda} \right)^n}{n} \right)^{y+1} .
\end{aligned}
$$

Hence, the entropy of the random variable Y is given by

$$
\begin{aligned}
H(Y) =\;& \log(\alpha(\alpha+1)\lambda) - \log(1+\alpha\lambda)E[Y+1] \\
& - \sum_{n=1}^{\infty} \frac{1}{n} \Psi_{Y+1}\left(\left(\frac{1+\alpha\lambda}{1+(\alpha+1)\lambda} \right)^n \right) \\
=\;& \log(\alpha(\alpha+1)\lambda) - \log(1+\alpha\lambda)\left(1 + \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1} \right) \right) \\
& - \sum_{n=1}^{\infty} \frac{t^n}{n}\left(\frac{\alpha\lambda}{\alpha\lambda+1-t^n} \right)\left(\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1-t^n} \right),
\end{aligned}
$$

where $t = \dfrac{1+\alpha\lambda}{1+(\alpha+1)\lambda}$. $\square$

Theorem 3. *For the random variable Y having PBE distribution, we have the following inequality*

$$
\max\left\{ -\frac{1+2\alpha\lambda}{\alpha^2\lambda^2}\log\left(\frac{\alpha\lambda+1}{\alpha\lambda} \right), -\frac{1+2(\alpha+1)\lambda}{(\alpha+1)^2\lambda^2}\log\left(\frac{(\alpha+1)\lambda+1}{(\alpha+1)\lambda} \right) \right\} \leqslant
$$
$$
H(Y) \leqslant -\frac{1+2\alpha\lambda}{\alpha^2\lambda^2}\log\left(\frac{\alpha\lambda+1}{\alpha\lambda} \right) - \frac{1+2(\alpha+1)\lambda}{(\alpha+1)^2\lambda^2}\log\left(\frac{(\alpha+1)\lambda+1}{(\alpha+1)\lambda} \right).
$$

Proof. In Subsection 3.1, we show that $Y = Z+W$ where Z and W are two independent random variable with $Ge\left(\dfrac{\alpha\lambda}{\alpha\lambda+1} \right)$ and $Ge\left(\dfrac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1} \right)$,

respectively. So, the entropy of Z and W are given by

$$H(Z) = -\frac{\frac{\alpha\lambda+1}{\alpha\lambda}\log\left(\frac{\alpha\lambda+1}{\alpha\lambda}\right) + \frac{\alpha\lambda}{\alpha\lambda+1}\log\left(\frac{\alpha\lambda}{\alpha\lambda+1}\right)}{\frac{\alpha\lambda}{\alpha\lambda+1}}$$

$$= -\frac{1+2\alpha\lambda}{\alpha^2\lambda^2}\log\left(\frac{\alpha\lambda+1}{\alpha\lambda}\right),$$

$$H(W) = -\frac{\frac{(\alpha+1)\lambda+1}{(\alpha+1)\lambda}\log\left(\frac{(\alpha+1)\lambda+1}{(\alpha+1)\lambda}\right) + \frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1}\log\left(\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1}\right)}{\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1}}$$

$$= -\frac{1+2(\alpha+1)\lambda}{(\alpha+1)^2\lambda^2}\log\left(\frac{(\alpha+1)\lambda+1}{(\alpha+1)\lambda}\right).$$

Since Y can be written as the sum of two independent geometric random variable, we have

$$\max\{H(Z), H(W)\} \leqslant H(Y) \leqslant H(Z) + H(W).$$

Thus, the proof is complete. See for more details about the entropy of sum of random variable. □

4 Different Methods of Estimation

4.1 Method of Moments

Given a random sample $x_1, \ldots, x_n$ of size n from the PBE distribution, the moment estimates (MOM), $\tilde{\lambda}$ and $\tilde{\alpha}$, of λ and α are given by

$$\tilde{\alpha} = -\frac{1}{2} + \frac{m_1}{2\sqrt{2(m_2 - m_1) - m_1^2}}, \tag{3}$$

and

$$\tilde{\lambda} = \frac{1}{m_1}\left(\frac{1}{\tilde{\alpha}} + \frac{1}{\tilde{\alpha}+1}\right), \tag{4}$$

where $m_1 = \frac{1}{n}\sum_{i=1}^n x_i$ and $m_2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$.

Theorem 4. *For a fixed values of α, the moment estimator $\tilde{\lambda}$ of λ is consistent and asymptotically normal distributed as,*

$$\sqrt{n}(\tilde{\lambda} - \lambda) \sim N(0, \nu^2(\lambda)),$$

where
$$\nu^2(\lambda) = \frac{\alpha(\alpha+1)\lambda^3 + (2\alpha+1)(\alpha^2 + (\alpha+1)^2)\lambda^2}{2\alpha+1}.$$

Proof. We know that $\tilde{\lambda} = \frac{1}{m_1}\left(\frac{1}{\tilde{\alpha}} + \frac{1}{\tilde{\alpha}+1}\right)$. Let $g(t) = \frac{1}{t}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right)$. In this case $g(\bar{X}) = \tilde{\lambda}$ and $g(\mu) = \lambda$ where $E(Y) = \mu = \frac{1}{\lambda}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1}\right)$.

Consistency: Since $\mu < \infty$, $\bar{X} \to \mu$ in probability and $g(t)$ is a continuous function at $t = \mu$, then $g(\bar{X}) \to g(\mu)$ in probability and this means $\tilde{\lambda} \to \lambda$ in probability.

Asymptotic normality: Since $\sigma^2 < \infty$, using central limit theorem we have $\sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2)$. Also since $g'(\mu) \neq 0$, using the delta method we have $\sqrt{n}(g(\bar{X}) - g(\mu)) \sim N(0, g'(\mu)^2\sigma^2)$, which proves the theorem. $\qquad\square$

As a result of the above theorem one can obtain a $100(1-\gamma)$ confidence interval for λ, for a given value α, using $\tilde{\lambda} \pm \frac{\nu(\tilde{\lambda})}{\sqrt{n}}Z_{\gamma/2}$, where $Z_{\gamma/2}$ is the quantile $1 - \gamma/2$ of the standard normal distribution.

4.2   Method of Maximum Likelihood

In this part the maximum likelihood estimators of $PBE(\alpha, \lambda)$ are considered, where both parameters are unknown. Given a random sample $y_1, y_2, ..., y_n$, of size n, from the PBE distribution, the log-likelihood function is of the form

$$l(\alpha, \lambda) = n\log\alpha + n\log(\alpha+1) + n\log\lambda$$
$$+ \sum_{i=1}^{n}\log\left[(1+\alpha\lambda)^{-(y_i+1)} - (1+(\alpha+1)\lambda)^{-(y_i+1)}\right]. \qquad (5)$$

Therefor, the normal equations are

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + \frac{n}{\alpha+1} - \sum_{i=1}^{n}\frac{(1+y_i)\lambda\left((1+\alpha\lambda)^{-2-y_i} + (1+\lambda(\alpha+1))^{-2-y_i}\right)}{(1+\alpha\lambda)^{-1-y_i} - (1+\lambda(\alpha+1))^{-1-y_i}}, \quad (6)$$

and

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n}\frac{(1+y_i)\left(\alpha(1+\alpha\lambda)^{-2-y_i} + (1+\alpha)(1+\lambda(1+\alpha))^{-2-y_i}\right)}{(1+\alpha\lambda)^{-1-y_i} - (1+\lambda(1+\alpha))^{-1-y_i}}. \quad (7)$$

Unfortunately, the two non-linear equations (6) and (7) have no closed form of answers. Also, these equations can be solved using non-linear algorithms such as Newton-Raphson, by taking the moment estimates of α and λ in (3) and (4), as an initial values.

4.3   Asymptotic Variances and Covariances of the MLEs

Applying the usual large sample approximation, MLE of Θ i.e. $\hat{\Theta} = (\hat{\alpha}, \hat{\lambda})$, can be treated as being approximately bivariate normal with mean Θ and variance-covariance matrix, which is the inverse of the expected information matrix $J(\Theta) = E[I(\Theta)]$, i.e., $N_2(\Theta, J(\Theta)^{-1})$, where $I(\Theta)$ is the observed information matrix with elements $I_{ij} = -\frac{\partial^2 l}{\partial \theta_i \theta_j}$ with $i, j = 1, 2$, and the expectation is to be taken with respect to the distribution of Y. When the expectations of I_{ij}, $i, j = 1, 2$, is obtained, we would have the matrix $J(\Theta)$. The inverse of $J(\Theta)$, evaluated at $\hat{\Theta}$ provides the asymptotic variance-covariance matrix of MLEs. Alternative estimates can be obtained from the inverse of the observed information matrix since it is a consistent estimator of $J(\Theta)^{-1}$. The estimated asymptotic multivariate normal $N_2(\Theta, I(\hat{\Theta})^{-1})$ distribution of $\hat{\Theta}$ can be used to construct approximate confidence intervals for the parameters and for the hazard rate and survival functions. An $100(1-\gamma)$ asymptotic confidence intervals for the parameters α and λ are given by $ACI_\alpha = (\hat{\alpha} - Z_{\gamma/2}\sqrt{\hat{I}_{\alpha\alpha}}, \hat{\alpha} + Z_{\gamma/2}\sqrt{\hat{I}_{\alpha\alpha}})$ and $ACI_\lambda = (\hat{\lambda} - Z_{\gamma/2}\sqrt{\hat{I}_{\lambda\lambda}}, \hat{\lambda} + Z_{\gamma/2}\sqrt{\hat{I}_{\lambda\lambda}})$ where $\hat{I}_{\alpha\alpha}$ and $\hat{I}_{\lambda\lambda}$ are the diagonal element of $I(\hat{\Theta})^{-1}$ and $Z_{\gamma/2}$ is the quantile $1 - \gamma/2$ of the standard normal distribution. Differentiating $\frac{\partial l}{\partial \alpha}$ and $\frac{\partial l}{\partial \lambda}$ w.r.t. α and λ, the elements of the symmetric second-order observed information matrix are found. These elements can be seen in Appendix.

5   Random Data Generation

Here we give two methods for random data generation from $PBE(\alpha, \lambda)$.

First method: Note that if $Y \sim PBE(\alpha, \lambda)$, W and Z are two independent random variables having geometric distributions with respective parameters $\frac{\alpha\lambda}{\alpha\lambda+1}$ and $\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1}$, then $Y = W + Z$. For generating a random sample of size n from PBE distribution, one can generate two independent random sample

from $Ge(\frac{\alpha\lambda}{\alpha\lambda+1})$ and $Ge(\frac{(\alpha+1)\lambda}{(\alpha+1)\lambda+1})$ and add them together.

Second method: An algorithm, similar to the one given by Ghitany et al. (2008), can be used to generate random data from PBE distribution. If $Y|\theta \sim Poisson(\theta)$ and $\theta \sim BE(\alpha,\lambda)$, then the density function of the PBE can be written as a mixed form

$$P_{PBE}(y) = \int_0^\infty P_{Y|\theta}(y|\theta) f_{BE}(\theta) d\theta.$$

To generate a random sample y_i, $i = 1, 2, \ldots, n$, from PBE distribution, regarding to the above expression, one can use the following algorithm:

- Generate $\theta_i \sim BE(\alpha,\lambda)$ for $i = 1, 2, \ldots, n$.

- Generate $Y_i \sim Poisson(\theta_i)$ for $i = 1, 2, \ldots, n$.

In the above algorithm one can simulate a random sample from $BE(\alpha,\lambda)$ using the expression $\theta_i = -\frac{1}{\lambda} \log Z_i$ for $i = 1, 2, \ldots, n$, where $Z_i \sim Beta(\alpha,2)$.

5.1   Simulation Study

In this section, a simulation study have been performed to investigate the consistency of the MMEs and MLEs of the parameters of a PBE distribution. A simulation study is carried out $N = 10,000$ times for each $(\alpha,\lambda) = (0.2,0.6),(0.8,0.6),(1,0.8),(0.8,1),(0.8,2), (0.2,0.8),(2,0.8),(2,2)$ and $n = 50(50)300$. The study calculates the following measures:

- Average maximum likelihood and moment estimators of the parameters α, λ; i.e., $\bar{\hat{\alpha}} = \frac{1}{N}\sum_{i=1}^N \hat{\alpha}_i$, $\bar{\hat{\lambda}} = \frac{1}{N}\sum_{i=1}^N \hat{\lambda}_i$, $\bar{\tilde{\alpha}} = \frac{1}{N}\sum_{i=1}^N \tilde{\alpha}_i$, $\bar{\tilde{\lambda}} = \frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i$.

- Average bias of the maximum likelihood and moment estimators of the parameters α, λ; i.e., $\frac{1}{N}\sum_{i=1}^N \hat{\alpha}_i - \alpha$, $\frac{1}{N}\sum_{i=1}^N \hat{\lambda}_i - \lambda$, $\frac{1}{N}\sum_{i=1}^N \tilde{\alpha}_i - \alpha$, $\frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i - \lambda$.

- Average square root of mean square error of the maximum likelihood and moment estimators of the parameters α, λ; i.e., $\sqrt{\frac{1}{N}\sum_{i=1}^N (\hat{\alpha}_i - \alpha)^2}$, $\sqrt{\frac{1}{N}\sum_{i=1}^N (\hat{\lambda}_i - \lambda)^2}$, $\sqrt{\frac{1}{N}\sum_{i=1}^N (\tilde{\alpha}_i - \alpha)^2}$, $\sqrt{\frac{1}{N}\sum_{i=1}^N (\tilde{\lambda}_i - \lambda)^2}$.

The results from simulated data sets are reported in Tables 1 and 2 for MMEs and MLEs, respectively. Tables 1 and 2 indicate that the convergence has been achieved in all cases, even when the sample size is small and emphasizes the robustness of the MLEs and MMEs such that the average square root of mean square errors decrease when sample size increases. Also, the average bias is decreasing when sample size is increased. Figure 3 represents the bias and mse of the MLEs for parameters $(\alpha, \lambda) = (0.2, 0.6)$.
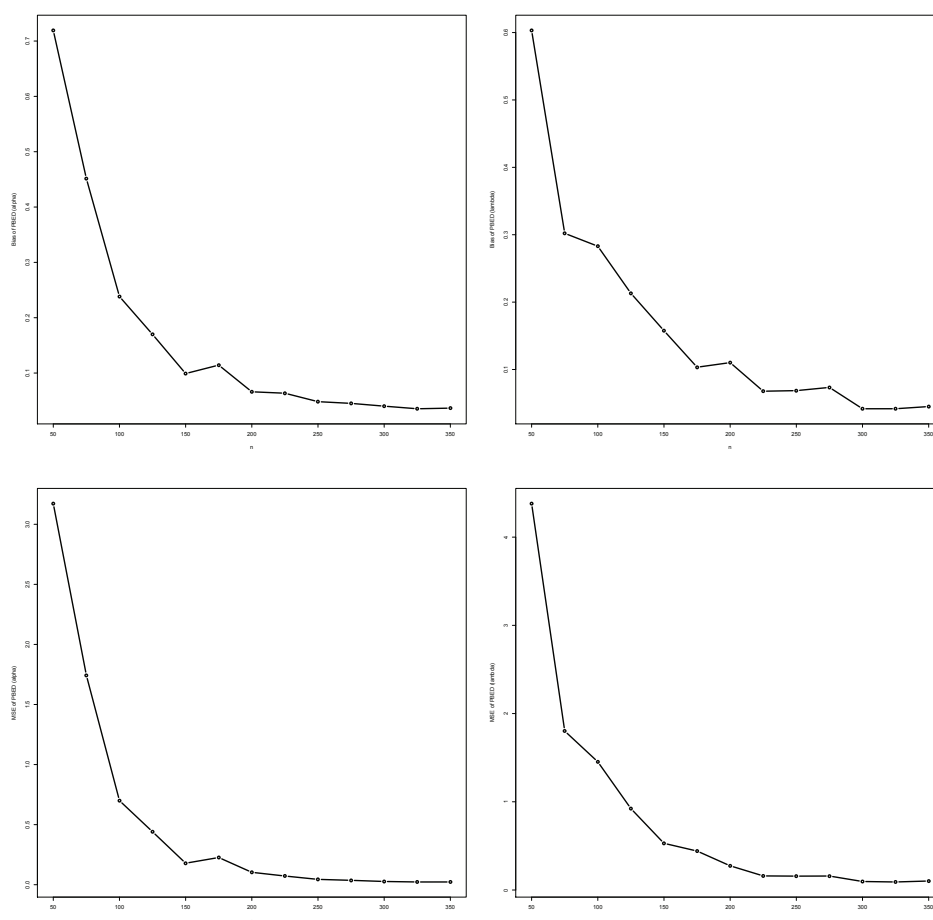

Figure 3. Graphs of bias and mse of the MLEs for parameters $(\alpha, \lambda) = (0.2, 0.6)$; bias of α is top left, bias of λ is top right, mse of α is bottom left and mse of λ is bottom right.

Table 1. The average estimators, average biases and average square root of mean square errors of the MME of the PBE distribution from 10,000 simulations.

	Parameter values		Average estimators		Average bias		Average rmse	
n	α	λ	$\widetilde{\alpha}$	$\widetilde{\lambda}$	$bias(\widetilde{\alpha})$	$bias(\widetilde{\lambda})$	$\sqrt{mse(\widetilde{\alpha})}$	$\sqrt{mse(\widetilde{\lambda})}$
50	0.2	0.6	1.4116	0.1847	1.2116	-0.4153	2.1863	0.4705
	0.8	0.6	1.9986	0.4502	1.1986	-0.1498	2.6619	0.3446
	1.0	0.8	2.0618	0.7977	1.0618	-0.0023	3.1105	0.7169
	0.8	1.0	1.8415	0.9003	1.0415	-0.0997	3.0011	0.6695
	0.8	2.0	1.2772	2.5017	0.4772	0.5017	1.3929	7.1859
	0.2	0.8	1.4129	0.2336	1.2129	-0.5664	1.8698	0.5987
	2.0	0.8	1.5029	2.0946	-0.4971	1.2946	1.8931	5.7338
	2.0	2.0	0.8429	7.3466	-1.1569	5.3466	2.7881	8.6023
100	0.2	0.6	1.2548	0.1655	1.0548	-0.4345	1.6814	0.4398
	0.8	0.6	2.1678	0.3809	1.3678	-0.2191	2.843	0.302
	1.0	0.8	2.1897	0.6463	1.1897	-0.1537	3.3080	0.3961
	0.8	1.0	1.9521	0.7259	1.1500	-0.2741	2.8992	0.5203
	0.8	2.0	1.4232	1.8953	0.6232	-0.1047	1.3128	1.5153
	0.2	0.8	1.3613	0.2159	1.1613	-0.5841	1.9821	0.5924
	2.0	0.8	1.8473	1.3911	-0.1527	0.5911	2.0716	1.0925
	2.0	2.0	0.9231	6.6796	-1.0780	4.6796	2.4009	7.8991
150	0.2	0.6	1.1599	0.1643	0.9599	-0.4357	1.2497	0.4400
	0.8	0.6	2.2920	0.3455	1.4920	-0.2545	2.8497	0.3052
	1.0	0.8	2.2097	0.6022	1.2097	-0.1978	2.5653	0.3852
	0.8	1.0	2.0690	0.6761	1.2690	-0.3239	2.7537	0.4932
	0.8	2.0	1.5777	1.6976	0.7777	-0.3024	1.7192	1.0749
	0.2	0.8	1.1400	0.2226	0.9400	-0.5774	1.2137	0.5838
	2.0	0.8	2.0913	1.1830	0.0913	0.3830	2.0152	0.7909
	2.0	2.0	1.0030	4.7167	-0.9970	2.7167	2.2111	6.2877
200	0.2	0.6	1.0661	0.1651	0.8661	-0.4349	1.0130	0.4378
	0.8	0.6	2.4858	0.3387	1.6858	-0.2613	3.5359	0.3099
	1.0	0.8	2.5861	0.5594	1.5861	-0.2406	3.9634	0.3837
	0.8	1.0	2.6688	0.6010	1.8688	-0.3990	5.1487	0.5088
	0.8	2.0	1.9648	1.5152	1.1648	-0.4848	3.1072	0.9829
	0.2	0.8	1.1324	0.2198	0.9324	-0.5802	1.2068	0.5852
	2.0	0.8	2.1696	1.1275	0.1696	0.3275	2.5004	0.7229
	2.0	2.0	1.3947	4.6709	-0.6053	2.6709	1.4596	4.5187
250	0.2	0.6	1.0475	0.1644	0.8475	-0.4356	1.0159	0.4380
	0.8	0.6	2.9978	0.3111	2.1978	-0.2889	5.7499	0.3270
	1.0	0.8	2.6641	0.5057	1.6641	-0.2943	3.5167	0.3824
	0.8	1.0	2.4619	0.5712	1.6619	-0.4288	3.2209	0.5132
	0.8	2.0	2.2983	1.3750	1.4983	-0.6250	3.2146	1.0445
	0.2	0.8	1.0534	0.2202	0.8534	-0.5798	1.0442	0.5831
	2.0	0.8	2.3363	1.0545	0.3363	0.2545	2.5187	0.6645
	2.0	2.0	1.8244	4.0913	-0.1756	2.0913	3.4323	4.1677
300	0.2	0.6	1.0061	0.1659	0.8061	-0.4341	0.8796	0.4361
	0.8	0.6	2.6473	0.3048	1.8473	-0.2952	3.2048	0.3273
	1.0	0.8	2.6541	0.5000	1.6541	-0.3000	3.9226	0.3798
	0.8	1.0	2.4839	0.5635	1.6839	-0.4365	4.5985	0.5100
	0.8	2.0	2.2079	1.3966	1.4079	-0.6034	3.7668	0.9691
	0.2	0.8	1.0160	0.2204	0.8160	-0.5796	0.8921	0.5828
	2.0	0.8	2.4397	0.9662	0.4397	0.1662	2.1959	0.5356
	2.0	2.0	1.6532	3.8646	-0.3468	1.8646	1.8640	3.1876

Table 2. The average estimators, average biases and average square root of mean square errors of the MLE of the PBE distribution from 10,000 simulations.

	Parameter values		Average estimators		Average bias		Average rmse	
n	α	λ	$\widehat{\alpha}$	$\widehat{\lambda}$	$bias(\widehat{\alpha})$	$bias(\widehat{\lambda})$	$\sqrt{mse(\widehat{\alpha})}$	$\sqrt{mse(\widehat{\lambda})}$
50	0.2	0.6	0.9385	1.132	0.7385	0.5307	1.825	1.932
	0.8	0.6	4.648	1.718	3.848	1.118	6.090	4.296
	1.0	0.8	6.387	3.673	5.387	2.873	8.152	9.035
	0.8	1.0	6.201	4.403	5.401	3.403	8.309	10.294
	0.8	2.0	5.972	18.308	5.172	16.308	10.043	31.560
	0.2	0.8	1.143	1.630	0.9429	0.8304	2.250	2.769
	2.0	0.8	9.283	8.443	7.283	7.643	11.103	19.707
	2.0	2.0	3.210	77.678	1.210	75.678	8.111	106.025
100	0.2	0.6	0.4918	0.8272	0.2918	0.2272	0.9783	1.056
	0.8	0.6	3.687	1.037	2.888	0.4374	4.968	2.145
	1.0	0.8	5.335	1.987	4.335	1.187	4.335	1.184
	0.8	1.0	4.862	2.516	4.062	1.516	6.781	5.658
	0.8	2.0	3.806	12.512	3.006	10.512	7.476	21.597
	0.2	0.8	0.5505	1.197	0.3505	0.3965	1.198	1.587
	2.0	0.8	8.004	5.078	6.004	4.278	9.685	12.484
	2.0	2.0	1.348	54.167	-0.6522	52.167	4.968	74.902
150	0.2	0.6	0.3322	0.7374	0.1322	0.1374	0.5380	0.7114
	0.8	0.6	3.212	0.8602	2.412	0.2602	4.386	1.517
	1.0	0.8	4.486	1.439	3.486	0.6386	5.994	2.801
	0.8	1.0	4.245	1.759	3.445	0.7585	6.020	3.499
	0.8	2.0	2.807	9.374	2.007	7.374	6.053	16.496
	0.2	0.8	0.3556	1.030	0.1556	0.2299	0.6296	1.074
	2.0	0.8	7.338	3.621	5.338	2.821	8.938	8.696
	2.0	2.0	1.087	41.719	-0.9126	39.719	3.999	61.064
200	0.2	0.6	0.2757	0.6967	0.0757	0.0968	0.3390	0.5087
	0.8	0.6	2.863	0.7302	2.063	0.1302	3.926	0.938
	1.0	0.8	4.261	1.156	3.261	0.3562	5.652	1.845
	0.8	1.0	3.821	1.530	3.021	0.5303	5.505	2.533
	0.8	2.0	2.275	7.372	1.475	5.372	5.114	12.404
	0.2	0.8	0.3007	0.9546	0.1007	0.1549	0.4429	0.7968
	2.0	0.8	6.638	2.853	4.638	2.053	8.204	6.499
	2.0	2.0	0.8976	33.493	-1.102	31.493	3.358	50.225
250	0.2	0.6	0.2542	0.6751	0.0542	0.0750	0.2563	0.4232
	0.8	0.6	2.616	0.6937	1.816	0.0937	3.595	0.7472
	1.0	0.8	3.997	1.075	2.997	0.2747	5.324	1.559
	0.8	1.0	3.369	1.358	2.569	0.3576	4.953	1.891
	0.8	2.0	1.843	6.085	1.043	4.085	4.250	9.768
	0.2	0.8	0.2622	0.9070	0.0622	0.1070	0.2816	0.6197
	2.0	0.8	6.470	2.316	4.470	1.516	7.920	5.100
	2.0	2.0	0.7630	28.730	-1.237	26.730	2.727	43.627
300	0.2	0.6	0.2409	0.6619	0.0409	0.0619	0.1831	0.3714
	0.8	0.6	2.342	0.6613	1.542	0.0613	3.228	0.5946
	1.0	0.8	3.704	0.9975	2.704	0.1975	4.967	1.205
	0.8	1.0	3.245	1.229	2.445	0.2287	4.739	1.454
	0.8	2.0	1.843	6.085	1.043	4.085	4.250	9.768
	0.2	0.8	0.2529	0.8809	0.0529	0.0809	0.2343	0.5340
	2.0	0.8	6.230	2.057	4.230	1.257	7.632	4.044
	2.0	2.0	0.7893	24.493	-1.211	22.493	2.628	38.057

6 Application of the PBE Distribution

In general, three types of distributions, used for modelling count data, are: under-dispersed, equi-dispersed and over-dispersed. The choice of distributions usually depends on the degrees of dispersion by observing the empirical distribution of the data. However, in the case of the highly-skewed empirical distribution of the data, with a very long-tail or a very heavy-tail, the commonly used distributions, such as Poisson and PL often fail to adequately fit the data.

To show the superiority of the PBE distribution, we compare the results of fitting the PBE distribution to some of other distributions such as Poisson, NB, PL and GPL distributions using two real data sets. The required numerical evaluations are implemented using the SAS and R softwares.

The performances of the model fittings are compared by using chi-squared goodness-of-fit statistic, -2log likelihood statistic and Akaike Information Criteria (AIC). We deliberately choose to combine the adjacent classes when the expected frequency for the chi-square is below 5.

6.1 The Belgian Motor Third Party Liability Insurance Portfolio Observed During the Year 1997

Table 3 presents the Belgian motor third party liability insurance portfolio observed during the year 1997 (Denuit et al., 2007). The data set comprises 14,505 policies with average annual claim frequency 14.6 percent. The majority of policies are in force during the whole year. The data have a sample mean 0.1197, a sample variance of 0.1453, skewness of 3.41 and kurtosis of 16.67. Clearly over-dispersion is present. Our results show that the PBE distribution provides a better fit to this data than the Poisson, NB, PL and GPL models. The pmfs of these distributions ig given in Appendix. Table 3 shows that the Poisson, NB, PL and GPL distributions are not suitable for fiting this set of data. The present two-parameter PBE distribution appears to give a best fit to this data set.

6.2 An Insurance Count Data from Belgium in Year 1993

Table 4 presents an insurance count data from Belgium in year 1993 which considered by Denuit (1997). The data set comprises 63299 observations. The data have a sample mean 0.1057, a sample variance of 0.1249, skewness of 3.52 and kurtosis of 17.59. Clearly over-dispersion is present. Our results

Table 3. Belgian motor third party liability insurance portfolio observed during the year 1997.

Obs.	Observed	Expected frequencies				
		Poisson	NB	PL	GPL	PBE
0	12962	12867.4	13030.4	12874.9	12886.9	12890.3
1	1369	1541.5	1318.7	1448.2	1431.4	1437.5
2	157	92.34	139.25	161.7	164.98	157.7
3	14	3.69	14.91	17.94	19.15	17.4
4	3	1.104	1.6	1.98	2.22	1.9
Total	14505	14505	14505	14505	14505	14505
MLEs		$\hat{\lambda}$=0.1198	$\hat{r}$=0.92	$\hat{\theta}$=8.72	$\hat{\theta}$=8.11	$\hat{\alpha}$=0.015
			$\hat{p}$=0.89		$\hat{\alpha}$=0.913	$\hat{\lambda}$=541.04
Stds		sd($\hat{\lambda}$)=0.0025	sd($\hat{r}$)=0.1187	sd($\hat{\theta}$)=0.1125	sd($\hat{\theta}$)=0.9541	sd($\hat{\alpha}$)=0.0038
			sd($\hat{p}$)=0.0067		sd($\hat{\alpha}$)=0.1067	sd($\hat{\lambda}$)=13.16
-2Log L		11133.9	11050.9	11052.01	11050.95	11046.15
AIC		11135.9	11054.9	11054.01	11054.95	11050.15
χ^2-Pearson		96.28	4.56	5.48	4.44	3.9

Table 4. An insurance count data from Belgium in year 1993.

Obs.	Observed	Expected frequencies				
		Poisson	NB	PL	GPL	PBE
0	57178	56989.6	57183.2	57243.7	57192.9	57171.1
1	5617	5983.9	5584.8	5479.3	5577.3	5610.1
2	446	314.2	486.6	521.5	484.8	475.08
3	50	11	40.69	49.4	40.38	43.22
4	8	0.29	3.33	4.7	3.28	4.71
Total	63299	63299	63299	63299	63299	63299
MLEs		$\hat{\lambda}$=0.105	$\hat{r}$=1.275	$\hat{\theta}$=10.297	$\hat{\alpha}$=1.27	$\hat{\alpha}$=0.223
			$\hat{p}$=0.9234		$\hat{\theta}$=12.72	$\hat{\lambda}$=49.99
Stds		sd($\hat{\lambda}$)=0.0013	sd($\hat{r}$)=0.1208	sd($\hat{\theta}$)=0.1231	sd($\hat{\alpha}$)=0.1239	sd($\hat{\alpha}$)=0.0912
			sd($\hat{p}$)=0.0068		sd($\hat{\theta}$)=1.127	sd($\hat{\lambda}$)=7.081
-2Log L		44301.1	44128.6	44135.7	44128.7	44127.8
AIC		44303.1	44132.6	44137.7	44132.7	44131.8
χ^2-Pearson		411.35	8.03	16.9	8.11	5.15

show that the PBE distribution provides a better fit to this data than the Poisson, NB, PL and GPL models. Table 4 shows that the Poisson, NB, PL and GPL distributions are not suitable for fitting this set of data. The present two-parameter PBE distribution appears to give a best fit to this data set.

7   Conclusion

This paper proposes PBE distribution as an alternatives for modelling event counts. Some properties of the PBE distribution are studied. The flexibility of PBE distribution allows for fitting diverse distributions of count data. We derive expansions for the moments and generating function. The estimation of parameters is approached by the method of moments and method of maximum likelihood and the observed information matrix is derived. Two applications to real data sets serve as illustrations of how to fit the PBE distribution. We hope that this generalization may attract wider applications in the modelling event counts.

Acknowledgment

References

Albrecht, P. (1982). On Some Statistical Methods Connected with the Mixed Poisson Process. *Scandinavian Actuarial Journal*, **9**, 1-14.

Albrecht, P. (1984). Laplace Transforms, Mellin Transforms and Mixed Poisson Process. *Scandinavian Actuarial Journal*, **11**, 58-64.

Al-Awadhi, S.A. and Ghitany, M.E. (2001). Statistical Properties of Poisson Lomax Distribution and Its Application to Repeated Accident Data. *Journal of Applied Statistical Science*, **10**, 365-372.

Bhattacharya, S.K. (1966). Confluent Hypergeometric Distributions of Discrete and Continuous Type with Application to Accident Proneness. *Bulletin of the Calcutta Statistical Association*, **15**, 20-31.

Bulmer, M.G. (1974). On Fitting The Poisson Lognormal Distribution to Species Abundance Data. *Biometrics*, **30**, 101-110.

Chakraborty, S. (2010). On Some Distributional Properties of the Family of Weighted Generalized Poisson Distribution. *Communications in Statistics-Theory and Methods*, **39**, 2767-2788.

Denuit, M., Marechal, X., Pitrebois, S. and Walhin, J.F. (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Scales*. Wiley, New York.

Denuit, M. (1997). A New Distribution of Poisson-type for the Number of Claims. *ASTIN Bulletin: The Journal of the IAA*, **27**, 229-242.

Douglas, J.B. (1980). Analysis with Standard Contagious Distributions, Statistical Distributions in Scientific Work Series 4, International Cooperative Publishing House.

Gaver, D. and O'Muircheartaigh, I.G., (1987). Robust Empirical Bayes Analysis of Event Rates. *Technometrics*, **29**, 1-15.

Ghitany, M.E. and Al-Mutairi, D.K. (2008). Size-biased Poisson-Lindley Distribution and Its Application. *METRON-International Journal of Statistics*, **66**(3), 299-311.

Ghitany, M.E., Al-Mutari, D.K. and Nadarajah, S. (2008). Zero-truncated Poisson-Lindley Distribution and Its Application. *Mathematics and Computers in Simulation* , **79**, 279-287.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2015). Parameters Estimation for a New Generalized Geometric Distribution. *Communications in Statistics -Simulation and Computation*, **44**(8).

Gómez-Déniz, E., Hernández-Bastida, A. and Fernández-Sánchez, M.P. (2016). A Suitable Discrete Distribution for Modelling Automobile Claim Frequencies. *Bulletin of the Malaysian Mathematical Sciences Society*, **39**, 633-647.

Greenwood, M. and Yule, G. (1920). An Enquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society*, **83**, 255-279.

Gupta, R.C. and Ong, S.H. (2005). Analysis of Long-tailed Count Data by Poisson Mixtures. *Communications in Statistics-Theory and Methods*, **34**, 557-573.

Gurland J. (1958). A Generalized Class of Contagious Distributions. *Biometrics*, **14**, 229-249.

Holla, M.S. and Bhattacharya, S.K. (1965). On a Discrete Compound Distribution. *Annals of the Institute of Statistical Mathematics*, **15**, 377-384.

Hougard, P. (1997). Analysis of Over Dispersed Count Data by Mixture of Poisson Variables and Poisson Processes. *Biometrics*, **53**, 1225-1238.

Irwin, J. (1975). The Generalized Waring Distribution, Parts I, II,III. *Journal of the Royal Statistical Society A,* 18-31(Part I), 204-227(Part II), 374-384(Part III).

Johnson, N.L., Kotz, S. and Kemp A.W. (2005). *Univariate Discrete Distributions*, Second Edition, John Wiley and Sons.

Kempton, R.A. (1975). A Generalized form of Fisher's Logarithmic Series. *Biometrics*, **62**, 29-38.

Kling, B. and Goovaerts, M. (1993). A Note on Compound Generalized Distributions. *Scandinavian Actuarial Journal*, **20**, 60-72.

Mahmoudi, E. and Zakerzadeh, H. (2010). Generalized Poisson-Lindley Distribution. *Communications in Statistics-Theory and Methods*, **39**, 1785-1798.

Nadarajah, S. and Kotz, S. (2006). The Beta Exponential Distribution. *Reliability Engineering and System Safety*, **91**, 689-697.

Ong, S.H. and Muthaloo, S. (1995). A Class of Discrete Distribution Suited to Fitting very Long Tail Data. *Communication in Statistics-Simulation and Computation*, **24**, 929-945.

Patil, G.P. (1964). On Certain Compound Poisson and Compound Binomial Distributions. *Sankha A*, **27**, 929-945.

Pielou, E. (1962). Run of One Species with Respect to Another in Transects Through Plant Population. *Biometrics*, **18**, 579-593.

Rai, G. (1971). A Mathematical Model for Accdent Proneness. *Trabajos Estadistica*, **22**, 207-212.

Rao, C.R. (1965). On Discrete Distributions Arising Out of Methods of Ascertainment. *In Classical and Contagious Discrete Distributions*, Ed. G. P. Patil. Calcutta: Pergamon Press and Statistical Publishing Society, 320-332.

Ruohonen M. (1988). A Model for the Claim Number Process. *ASTIN Bulletin*, **18**, 57-68.

Sankaran, M. (1969). On Certain Properties of a Class of Compound Poisson Distributions. *Sankha B*, **32**, 353-362.

Sankaran, M. (1970). The Discrete Poisson-Lindley Distribution. *Biometrics*, **26**, 145-149.

Sichel, H.S. (1974). On a Distribution Representing Sentence-Length in Written Prose. *Journal of The Royal Statistical Society A*, **137**, 25-34.

Sichel, H.S. (1974). On a Distribution Law for Word Frequencies. *Journal of the American Statistical Association*, **70**, 542-547.

Simon, P. (1955). On a Class of Skew Distributions. *Biometrika*, **42**, 425-440.

Willmot, G.E. (1986). Mixed Poisson Distribution. *ASTIN Bulletin Supplement*, **16**, 59-79.

Willmot, G.E. (1993). On Recursive Evaluation of Mixed Poisson Probabilities and Related Quantities. *Scandinavian Actuarial Journal*, **18**, 114-133.

Appendix

A1. Elements of the symmetric second-order observed information matrix:

$$I_{\alpha\alpha} = \frac{n}{\alpha^2} + \frac{n}{(\alpha+1)^2} - \lambda^2 \sum_{i=1}^{n} \frac{(y_i+1)\left((1+\alpha\lambda)^{-2(y_i+2)}+(1+(\alpha+1)\lambda)^{-2(y_i+2)}\right)}{\left((1+\alpha\lambda)^{-(y_i+1)}-(1+(\alpha+1)\lambda)^{-(y_i+1)}\right)^2}$$

$$+ \lambda^2 \sum_{i=1}^{n} \frac{(y_i+1)(1+\alpha\lambda)^{-(y_i+3)}(1+(\alpha+1)\lambda)^{-(y_i+3)}\left(2+\lambda(2+\lambda(y_i+2)+2\alpha(2+\lambda(\alpha+1)))\right)}{\left((1+\alpha\lambda)^{-(y_i+1)}-(1+(\alpha+1)\lambda)^{-(y_i+1)}\right)^2},$$

$$I_{\alpha\lambda} = \sum_{i=1}^{n} \frac{y_i+1}{(1+\alpha\lambda)^2} - \lambda^2 \sum_{i=1}^{n} \frac{(y_i+1)^2(1+\alpha\lambda)^{2y_i}}{\left(-(1+\alpha\lambda)^{y_i+1}(1+(\alpha+1)\lambda)+(1+(\alpha+1)\lambda)^{y_i+2}\right)}$$

$$+ \lambda \sum_{i=1}^{n} \frac{(y_i+1)(-2+\lambda y_i-2\alpha\lambda)}{(1+\alpha\lambda)^2(1+(\alpha+1)\lambda)^2-(1+\alpha\lambda)^{-(y_i-1)}(1+(\alpha+1)\lambda)^{(y_i+3)}},$$

$$I_{\lambda\lambda} = \frac{n}{\lambda^2} - \sum_{i=1}^{n} \frac{(y_i+1)\left(\alpha^2(1+\alpha\lambda)^{-2(y_i+2)}+(\alpha+1)^2(1+(\alpha+1)\lambda)^{-2(y_i+2)}\right)}{\left((1+\alpha\lambda)^{-(y_i+1)}-(1+(\alpha+1)\lambda)^{-(y_i+1)}\right)^2}$$

$$+ \sum_{i=1}^{n} \frac{(y_i+1)(1+\alpha\lambda)^{-(y_i+3)}(1+(\alpha+1)\lambda)^{-(y_i+3)}\left(y_i+2(1+\alpha(\alpha+1)(1+\alpha\lambda)(1+(\alpha+1)\lambda))\right)}{\left((1+\alpha\lambda)^{-(y_i+1)}-(1+(\alpha+1)\lambda)^{-(y_i+1)}\right)^2}.$$

A2. Pmfs of the NB, PL and GPL are given by

- (NB) $P(X=x) = \binom{r+x-1}{x}p^r(1-p)^x$.

- (PL) $P(X=x) = \frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}}$.

- (GPL) $P(X=x) = \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha+1)}\frac{\theta^{\alpha+1}}{(\theta+1)^{x+\alpha+1}}\left(\alpha+\frac{x+\alpha}{\theta+1}\right)$.

Eisa Mahmoudi
Department of Statistics,
Yazd University,
Yazd, Iran.
email: *emahmoudi@yazd.ac.ir*

Hossein Zamani
Department of Statistics,
Hormozgan University,
Bandarabbas, Iran.
email: *zamani-h@hormozgan.ac.ir*

Rahmatsadat Meshkat
Department of Statistics,
Yazd University,
Yazd, Iran.
email: *r.meshkat@gmail.com*