STATISTICS IN TRANSITION new series, September 2021
Vol. 22, No. 3, pp. 157–174, DOI 10.21307/stattrans-2021-032
Received – 24.02.2020; accepted – 07.10.2020

A new count data model applied in the analysis of vaccine adverse events and insurance claims

Showkat Ahmad Dar[1], Anwar Hassan[2], Peer Bilal Ahmad[3], Sameer Ahmad Wani[4]

ABSTRACT

The article presents a new probability distribution, created by compounding the Poisson distribution with the weighted exponential distribution. Important mathematical and statistical properties of the distribution have been derived and discussed. The paper describes the proposed model's parameter estimation, performed by means of the maximum likelihood method. Finally, real data sets are analyzed to verify the suitability of the proposed distribution in modeling count data sets representing vaccine adverse events and insurance claims.

Key words: poisson distribution, weighted exponential distribution, compound distribution, count data, maximum likelihood estimation.

1. Introduction

Compounding a discrete distribution with a continuous distribution is a valuable method for creating flexible distributions to assist modelling of count data. Count data distributions play a key role in several applications for applied fields and theoretical research like health, transport, insurance and engineering, etc. Barreto-Souza and Bakouch (2013) obtained a new class of compound distribution with decreasing failure rate by compounding zero-truncated Poisson Lindley distribution and exponential distribution. Hajebi et al. (2013) obtained a new lifetime model by compounding exponential distribution with negative binomial distribution. Mohmoudi and Jafari (2014) introduced a new lifetime compound probability distribution which generalizes the linear failure rate of distribution. Ghitany et al. (2011) obtained weighted Lindley

[1] Department of Statistics, University of Kashmir, Srinagar (J&K), India. Corresponding Author. E-mail: darshowkat2429@gmail.com. ORCID: https://orcid.org/0000-0002-3661-822X.
[2] Department of Statistics, University of Kashmir, Srinagar (J&K), India.
[3] Department of Mathematical Sciences, Islamic University of Science & Technology, Awantipora, Pulwama (J&K), India.
[4] Department of Statistics, University of Kashmir, Srinagar (J&K), India.

distribution and pointed that Lindley distribution is valuable in exhibiting biological data from mortality studies. Asgharzadeh et al. (2014) created a new class of distribution by mixing any continuous distribution and Poisson Lindley distribution through a compounding technique. Chesneau et al. (2020) introduced Cosine geometric distribution for count data modelling. Bourguigon et al. (2014) obtained the Birnbaum-Saunders power series distribution. The new lifetime distribution has a decreasing, increasing or constant hazard rate. Silva and Cordeiro (2015) created a new lifetime distribution by mixing Burr XII and power series distribution through a compounding technique. Pinho et al. (2015) obtained a new distribution by assuming that simple size distribution as Harris distribution. Bardbar and Nematollahi (2016) obtained a modified exponential distribution-geometric distribution with increasing or decreasing failure rate. Flores et al. (2013) obtained the complementary exponential power series distribution by considering the distribution of vectors through maximum components.

In this paper, we propose a new compounding distribution by compounding the Poisson distribution with the weighted exponential distribution, as there is a need to find a more flexible model for analysing statistical data. This model has over-dispersed nature so it will become most appropriate for analysing over-dispersed count data sets. This property makes this model unique as compared to other compounding models already in the statistical literature.

2. Definition of the proposed model (Poisson weighted exponential distribution)

If $Z|\nu \sim \mathrm{P}(\nu)$, where ν is itself a random variable following weighted exponential distribution with parameter (μ,σ), then determining the distribution that results from marginalizing over ν will be known as a compound of the Poisson distribution with that of weighted exponential distribution, which is denoted by $PWED(Z;\mu,\sigma)$. It may be noted that the proposed model will be a discrete one since the parent distribution is discrete.

Theorem 2.1: The probability mass function of a Poisson weighted exponential distribution, i.e. $PWED(Z;\mu,\sigma)$ is given by

$$P(Z=z)=\frac{\mu^2}{\mu+\sigma}\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}}\right];z=0,1,2,3,...,;\mu>0,\sigma\geq 0\ .$$

Proof: Using the definition (2), the pmf of a PWED $(Z;\mu,\sigma)$ can be obtained as

$$g(z\,|\,\nu)=\frac{e^{-\nu}\nu^z}{(z)!}\quad ;\ z=0,1,2,3,...;\nu>0\,.$$

When its parameter ν follows weighted exponential distribution (WED) with pdf

$$h(\nu;\mu,\sigma)=\frac{\mu^2}{\mu+\sigma}(1+\sigma\nu)e^{-\mu\nu}\,;\nu>0,\mu>0,\sigma\geq 0\,.$$

We have

$$P(Z=z)=\int_0^\infty g(z\,|\,\nu).h(\nu;\mu,\sigma)d\nu$$

$$P(Z=z)=\frac{\mu^2}{\mu+\sigma}\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}}\right];z=0,1,2,3,...,;\mu>0,\sigma\geq 0 \tag{2.1}$$

which is the p.m.f. of PWED.

The corresponding c.d.f of PWED is obtained as:

$$F_Z(z)=\sum_{n=0}^{z}\left[\frac{\mu^2}{\mu+\sigma}\left[\frac{(1+\mu)+\sigma(n+1)}{(1+\mu)^{n+2}}\right]\right]$$

$$1-\frac{\mu+\mu^2+\sigma+2\mu\sigma+\mu\sigma x}{(1+\mu)^{z+2}(\mu+\sigma)}\;;z>0,\mu>0,\sigma\geq 0 \tag{2.2}$$

3. Special Cases

Case 1: If we put $\sigma=0$ the PWED reduces to the Poisson exponential distribution with pmf as

$$P_1(Z=z)=\left[\frac{\mu}{(1+\mu)^{z+2}}\right].$$

Case 2: If we put $\sigma=1$ the PWED reduces to the Poisson size biased exponential distribution with pmf as

$$P_2(Z=z)=\frac{\mu^2}{\mu+1}\left[\frac{(1+\mu)+(z+1)}{(1+\mu)^{z+2}}\right].$$

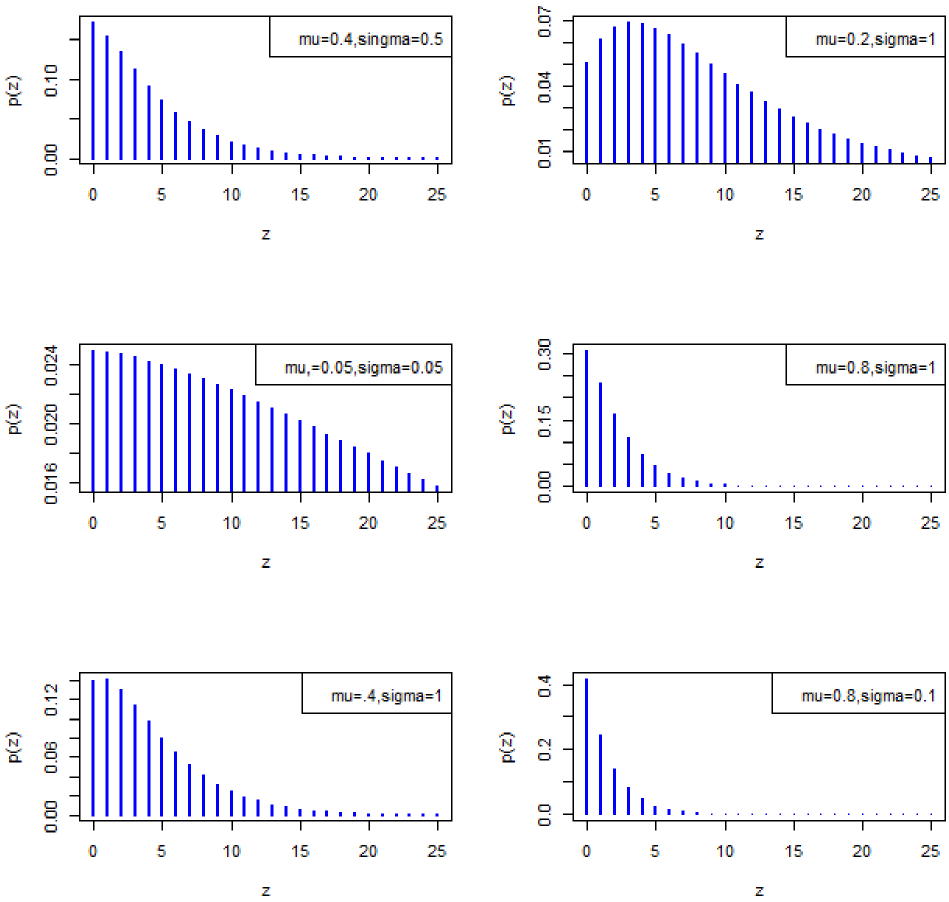

Figure 1. The above figures show the pmf plot for different values of μ and σ

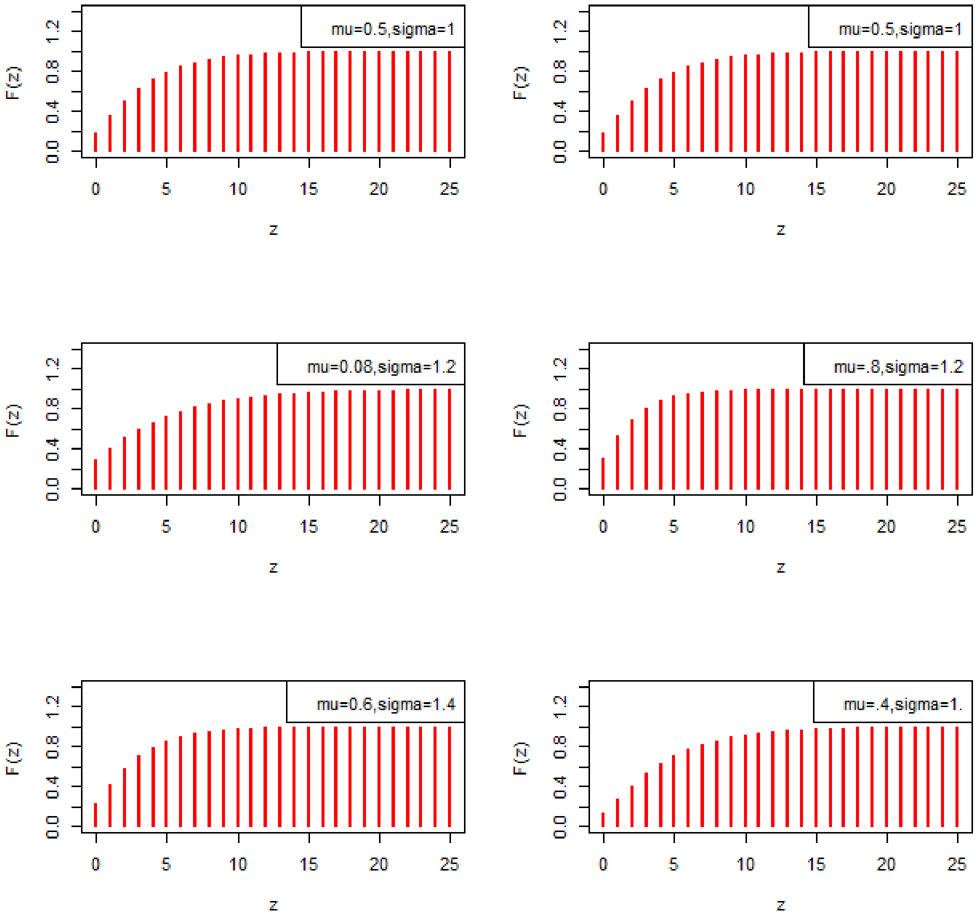

Figure 2. The above figures show the cdf plot for different values of μ and σ

4. Collective risk model

Theorem 4.1: Let Z follow PWED (μ,σ), be a primary distribution with exponential distribution (ζ) as a secondary distribution, then the aggregate loss U= $\sum_{i=0}^{M} Z_i$ has p.d.f given as

$$f_u(z) = \frac{\mu^2 e^{-\zeta\mu z} \zeta(\mu^2 + 2\mu(\sigma+1) + \sigma(\zeta z + 2) + 1)}{(1+\mu)^4(\mu+\sigma)},$$

whereas

$$f_u(0)=\frac{\mu^2((\mu+1)+\sigma)}{(\mu+\sigma)(1+\mu)^2}.$$

Proof: Let claim severity follow an exponential distribution $\zeta > 0$, we know that gamma (n, v) distribution is n^{th} fold convolution of exponential distribution, which is given as

$$f^{*n}(z)=\frac{\zeta^n}{(n-1)!}z^{n-1}e^{-\zeta z}, \qquad n=1,2,\dots.$$

Therefore, the random variable U has p.d.f given as:

$$f_u(z)=\frac{\mu^2}{(1+\mu)(\mu+\sigma)}e^{-\zeta z}\sum_{n=1}^{\infty}\left\{\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^z}\right\}\frac{\zeta^n}{(n-1)!}z^{n-1}e^{-\zeta z}$$

$$f_u(z)=\frac{\mu^2 e^{-\zeta\mu z}\zeta(\mu^2+2\mu(\sigma+1)+\sigma(\zeta z+2)+1)}{(1+\mu)^4(\mu+\sigma)}.$$

The probability of no claim is given by

$$f_u(0)=\frac{\mu^2((\mu+1)+\sigma)}{(\mu+\sigma)(1+\mu)^2}.$$

Theorem 4.2: For collective risk model with PWED (μ,σ) as a primary distribution and Erlang $(2,\zeta)$ as secondary distribution, then the probability density function of aggregate loss random variable $U=\sum_{i=0}^{M} Z_i$ is given as

$$fu(z)=\frac{\mu^2 e^{-\zeta z}}{(\mu+\sigma)(1+\mu)^{\frac{7}{2}}2z}\begin{pmatrix} z^2\zeta^2\sqrt{1+\mu}\sigma\cosh\frac{z\zeta}{\sqrt{1+\mu}}+z\zeta\sinh\frac{z\zeta}{\sqrt{1+\mu}} \\ \left(2\mu^2+3\mu\sigma+2+4\mu\right) \end{pmatrix},$$

whereas

$$f_u(0)=\frac{\mu^2((\mu+1)+\sigma)}{(\mu+\sigma)(1+\mu)^2}.$$

Proof: Let claim severity follow Erlang $(2,\zeta)$ and we know that gamma $(2n,\zeta)$ distribution is n^{th} fold convolution of Erlang $(2,\zeta)$ distribution with pdf given as

$$f^{*n}(z)=\frac{\zeta^{2n}}{(2n-1)!}z^{2n-1}e^{-\zeta z}, \qquad n=1,2,\dots.$$

So, the aggregate loss random variable U has pdf given as

$$fu(Z)=\sum_{n=1}^{\infty}\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}(\mu+\sigma)}\right]\frac{\zeta^{2n}}{(2n-1)!}z^{2n-1}e^{-\zeta z}$$

$$fu(z)=\frac{\zeta^{3}}{(\zeta\eta+2\beta)(1+\zeta)^{3}}e^{-vz}\sum_{n=1}^{\infty}\frac{((1+\mu)+\sigma(z+1))}{(1+\zeta)^{n}}\frac{v^{2n}}{(2n-1)!}z^{2n-1}$$

$$fu(z)=\frac{\mu^{2}e^{-\zeta z}}{(\mu+\sigma)(1+\mu)^{\frac{7}{2}}2z}\begin{pmatrix} z^{2}\zeta^{2}\sqrt{1+\mu}\sigma\cosh\frac{z\zeta}{\sqrt{1+\mu}}+z\zeta\sinh\frac{z\zeta}{\sqrt{1+\mu}} \\ \left(2\mu^{2}+3\mu\sigma+2+4\mu\right) \end{pmatrix}.$$

The probability of no claim is given by

$$f_{u}(0)=\frac{\mu^{2}((\mu+1)+\sigma)}{(\mu+\sigma)(1+\mu)^{2}}.$$

5. Reliability Analysis

5.1. Reliability Function R(z): The reliability function is defined as the probability that a system survives beyond a certain time. The reliability function or the survival function of PWED is given as

$$R(z,\mu,\sigma)=\left(\frac{\mu+\mu^{2}+\sigma+2\mu\sigma+\mu\sigma z}{(\mu+\sigma)^{z+2}(\mu+\sigma)}\right).$$

5.2. Hazard Function: The hazard function, also known as the hazard rate, is given as

$$\text{H.R}=h(z,\mu,\sigma)=\frac{f(z;\mu,\sigma)}{R(z;\mu,\sigma)}=\frac{\mu^{2}[(1+\mu)+\sigma(z+1)]}{\mu+\mu^{2}+\sigma+2\mu\sigma+\mu\sigma z}.$$

5.3. Reverse Hazard Rate and Mills Ratio: The reverse hazard rate and the Mills ratio of PWED are respectively given as

$$\text{R.H.R}=h_{r}(z,\mu,\sigma)=\frac{\mu^{2}[(1+\mu)+\sigma(z+1)]}{(\mu+\sigma)^{z+2}(\mu+\sigma)-(\mu+\mu^{2}+\sigma+2\mu\sigma+\mu\sigma z)},$$

$$\text{Mills ratio}=\frac{(\mu+\sigma)^{z+2}(\mu+\sigma)-(\mu+\mu^{2}+\sigma+2\mu\sigma+\mu\sigma z)}{\mu^{2}[(1+\mu)+\sigma(z+1)]}.$$

6. Statistical properties

In this section, structural properties of the PWE model have been evaluated.

6.1. Moments

6.1.1. Factorial Moments

Using (2.1), the rth factorial moment about origin of the PWED (2.1) can be obtained. $\mu_{(r)}' = E[E(Z^{(r)} \mid \nu)],\ where\ Z^{(r)} = Z(Z-1)(Z-2)...(Z-r+1)$

$$\mu_{(r)}' = \frac{\mu^2}{\mu+\sigma}\int_0^\infty \left[\nu^r\left(\sum_{z=r}^{\infty}\frac{e^{-\nu}\nu^{z-r}}{(z-r)!}\right)\right](1+\sigma\nu)e^{-\mu\nu}d\nu$$

Taking $u = z - r$, *we get*

$$\mu_{(r)}' = \frac{\mu^2}{\mu+\sigma}\int_0^\infty \left[\nu^r\left(\sum_{u=0}^{\infty}\frac{e^{-\nu}\lambda^{u}}{u!}\right)\right](1+\sigma\nu)e^{-\mu\nu}d\nu$$

$$\mu_{(r)}' = \frac{\mu^2 r!}{\mu+\sigma}\left[\frac{\mu+\sigma(r+1)}{\mu^{r+2}}\right] \tag{6.1}$$

Taking r=1,2,3,4 in (5.1), the first four factorial moments about origin of the PWED can be obtained as

$$\mu_{(1)}' = \frac{1}{\mu+\sigma}\left[\frac{\mu+2\sigma}{\mu}\right], \qquad \mu_{(3)}' = \frac{6}{\mu+\sigma}\left[\frac{\mu+4\sigma}{\mu^3}\right],$$

$$\mu_{(2)}' = \frac{2}{\mu+\sigma}\left[\frac{\mu+3\sigma}{\mu^2}\right], \qquad \mu_{(4)}' = \frac{24}{\mu+\sigma}\left[\frac{\mu+5\sigma}{\mu^4}\right].$$

6.1.2. Moments about origin (Raw moments)

Using the relationship between the factorial moments about origin and the moments about origin, the first four moments about origin of the PWED (2.1) can be obtained as:

$$\mu_1' = \frac{1}{\mu+\sigma}\left[\frac{\mu+2\sigma}{\mu}\right]$$

$$\mu_2' = \frac{\mu(\mu+2\sigma)+2(\mu+3\sigma)}{\mu^2(\mu+\sigma)}$$

$$\mu_3' = \frac{6(\mu+4\sigma)+6\mu(\mu+3\sigma)+\mu^2(\mu+2\sigma)}{\mu^3(\mu+\sigma)}$$

$$\mu_4' = \frac{24(\mu+5\sigma)+36\mu(\mu+4\sigma)+14\mu^2(\mu+3\sigma)+\mu^3(\mu+2\sigma)}{\mu^4(\mu+\sigma)}.$$

6.1.3. Moments about the Mean (Central moments)

Using the relationship $\mu_r = E(Y-\mu_1')^r = \sum_{H=0}^{r}\binom{r}{h}\mu_h'(-\mu_1')^{r-h}$ between the moments about the mean and the moments about origin, the moments about the mean of the

PWED (2.1) can be obtained as $\mu_2 = \frac{\mu(\mu+2\sigma)+2(\mu+3\sigma)}{\mu^2(\mu+\sigma)} - \left[\frac{1}{\mu+\sigma}\left[\frac{\mu+2\sigma}{\mu}\right]\right]^2$

$$\mu_3 = \frac{6(\mu+4\sigma)+6\mu(\mu+3\sigma)+\mu^2(\mu+2\sigma)}{\mu^3(\mu+\sigma)} - 3\frac{(\mu(\mu+2\sigma)+2(\mu+3\sigma))}{\mu^2(\mu+\sigma)}\frac{1}{\mu+\sigma}\left[\frac{\mu+2\sigma}{\mu}\right]$$

$$\mu_4 = \frac{24(\mu+5\sigma)+36\mu(\mu+4\sigma)+14\mu^2(\mu+3\sigma)+\mu^3(\mu+2\sigma)}{\mu^4(\mu+\sigma)} - 4\frac{6(\mu+4\sigma)+6\mu(\mu+3\sigma)+\mu^2(\mu+2\sigma)}{\mu^3(\mu+\sigma)}\frac{1}{\mu+\sigma}\left[\frac{\mu+2\sigma}{\mu}\right]$$
$$+6\frac{\mu(\mu+2\sigma)+2(\mu+3\sigma)}{\mu^2(\mu+\sigma)}\left[\frac{1}{\mu+\sigma}\left[\frac{\mu+2\sigma}{\mu}\right]\right]^2 - 3\left[\frac{1}{\mu+\sigma}\left[\frac{\mu+2\sigma}{\mu}\right]\right]^4$$

6.2. Coefficient of variation (c.v) , skewness $(\sqrt{\beta_1})$, , kurtosis (β_2) and Index of Dispersion (γ)

$$C.V = \frac{\sqrt{(\mu(\mu+2\sigma)+2(\mu+3\sigma))(\mu+\sigma)-(\mu+2\sigma)^2}}{(\mu+2\sigma)}$$

$$\sqrt{\beta_1} = \frac{(\mu+\sigma)(6(\mu+4\sigma)+6\mu(\mu+3\sigma)+\mu^2(\mu+2\sigma))-3(\mu(\mu+2\sigma)+2(\mu+3\sigma)(\mu+2\sigma))(\mu+2\sigma)}{((\mu+\sigma)(\mu(\mu+2\sigma)+2(\mu+3\sigma))-(\mu+2\sigma)^2)^{3/2}}$$

$$\beta_2 = \frac{\begin{array}{l}((24(\mu+5\sigma)+36\mu(\mu+4\sigma)+14\mu^2(\mu+3\sigma)+\mu^3(\mu+2\sigma))-4(6(\mu+4\sigma)+6\mu(\mu+3\sigma)+\mu^2(\mu+2\sigma))(\mu+2\sigma)\\ -6(\mu(\mu+2\sigma)+2(\mu+3\sigma))(\mu+2\sigma)^2-3(\mu+\sigma)(\mu+2\sigma))(\mu+\sigma)^2\end{array}}{(\mu(\mu+2\sigma)+2(\mu+3\sigma))^2(\mu+\sigma)(\mu+2\sigma)^4}$$

$$\gamma = \frac{(\mu(\mu+2\sigma)+2(\mu+3\sigma))(\mu+2\sigma)-(\mu+2\sigma)}{\mu(\mu+\sigma)(\mu+2\sigma)}.$$

Table 1. Index of Dispersion, Mean and Variance of PWED (μ,σ) for different values of parameters

	μ	**0.2**	**0.5**	**0.8**	**1**	**1.2**	**1.5**	**1.8**
$\sigma = 1.2$	IOD	5.3	2.8	3.19	4.13	5.56	8.64	12.95
	VAR	50.6	10.09	6.83	6.88	7.53	9	10.98
	MEAN	9.55	3.6	2.14	1.66	1.35	1.05	0.84
$\sigma = 1.5$	IOD	5.35	2.7	2.89	3.65	4.85	7.52	11.34
	VAR	50.3	9.5	5.97	5.84	6.3	7.52	9.16
	MEAN	9.4	3.5	2.06	1.6	1.3	1	0.8
$\sigma = 1.8$	IOD	5.4	2.66	2.71	3.36	4.42	6.85	10.37
	VAR	50	9.1	5.43	5.19	5.53	6.6	8.06
	MEAN	9.3	3.41	2	1.54	1.25	0.96	0.77
$\sigma = 2$	IOD	5.3	2.76	3.07	3.93	5.27	8.19	12.31
	VAR	50.5	9.86	6.49	6.47	7.03	8.44	10.26
	MEAN	9.5	3.56	2.11	1.64	1.33	1.03	0.85
$\sigma = 2.2$	IOD	5.3	2.84	3.31	4.32	5.84	9.09	13.6
	VAR	50.7	10.3	7.17	7.3	8.02	9.66	11.7
	MEAN	9.6	3.62	2.16	1.68	1.37	1.06	0.86

6.3. Moment generating function and probability generating function of Poisson weighted Exponential Distribution

Theorem 6.3.1: If Z has the PWED (μ,σ), then the probability generating function $P_z(t)$ has the following form:

$$P_X(t) = \frac{\mu^2}{(\mu+\sigma)(1+\mu)}\left[\frac{(\mu+1-t)((1+\mu)+\sigma)+\sigma t}{(\mu+1-t)^2}\right]$$

Proof: We begin with the well-known definition of the probability generating function given by

$$P_Z(t) = \sum_{x=0}^{\infty} t^z \left[\frac{\mu^2}{\mu+\sigma}\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}}\right]\right]$$

$$P_Z(t) = \frac{\mu^2}{(\mu+\sigma)(1+\mu)}\left[\frac{(\mu+1-t)((1+\mu)+\sigma)+\sigma t}{(\mu+1-t)^2}\right]$$

Theorem 6.3.2: If X has the PWED (μ,σ), then the moment generating function $M_Z(t)$ has the following form:

$$M_Z(t) = \frac{\mu^2}{(\mu+\sigma)(1+\mu)}\left[\frac{(\mu+1-e^t)((1+\mu)+\sigma)+\sigma e^t}{(\mu+1-e^t)^2}\right]$$

Proof: We begin with the well-known definition of the moment generating function given by $M_Z(t) = \sum_{z=0}^{\infty} e^{tz}\left[\frac{\mu^2}{\mu+\sigma}\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}}\right]\right]$

$$M_Z(t) = \frac{\mu^2}{(\mu+\sigma)(1+\mu)}\left[\frac{(\mu+1-e^t)((1+\mu)+\sigma)+\sigma e^t}{(\mu+1-e^t)^2}\right]$$

Similarly Laplace and Fourier transforms as calculated as

$$L_Z(t) = \frac{\mu^2}{(\mu+\sigma)(1+\mu)}\left[\frac{(\mu+1+e^t)((1+\mu)+\sigma)+\sigma e^t}{(\mu+1+e^t)^2}\right]$$

$$F_Z(t) = \frac{\mu^2}{(\mu+\sigma)(1+\mu)}\left[\frac{(\mu+1+t)((1+\mu)+\sigma)+\sigma t}{(\mu+1+t)^2}\right]$$

6.4 Recurrence Relation between Probabilities

The PWED can be written as

$$P(Z=z) = \frac{\mu^2}{\mu+\sigma}\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}}\right]$$

$$P(Z=z+1) = \frac{\mu^2}{\mu+\sigma}\left[\frac{(1+\mu)+\sigma(z+2)}{(1+\mu)^{z+3}}\right].$$

Dividing P(Z=z+1) by P(Z=z), we find the recurrence relation between probabilities

$$P(Z=z+1)=\frac{(z+2)(\eta(1+\zeta)+\beta(z+3))}{(1+\zeta)(z+1)(\eta(1+\zeta)+\beta(z+2))}P(Z=z) \quad .$$

6.5. Quantile function

Theorem 6.5: The quantile function of the PWED (μ,σ) is

$$Q_Z(u)=-\frac{(\mu+\mu^2+\sigma+2\mu\sigma)}{\mu\sigma}-\frac{1}{\log(1+\mu)}W_{-1}\left[\frac{(\mu+\sigma)(u-1)\log(1+\mu)}{\mu\sigma(1+\mu)^{\frac{\mu+\mu^2+\sigma}{\mu\sigma}}}\right].$$

Proof: The cdf of the distribution is

$$F_Z(z)=1-\frac{\mu+\mu^2+\sigma+2\mu\sigma+\mu\sigma z}{(1+\mu)^{z+2}(\mu+\sigma)}.$$

The u^{th} quantile function is obtained by solving $F_Z(z)=u$

$$\frac{-(\mu+\mu^2+\sigma+2\mu\sigma+\mu\sigma z)}{(\mu\sigma)}=Z+\frac{(u+1)(1+\mu)^{z+2}(\mu+\sigma)}{\mu\sigma}$$

$$Q_Z(u)=-\frac{(\mu+\mu^2+\sigma+2\mu\sigma)}{\mu\sigma}-\frac{1}{\log(1+\mu)}W_{-1}\left[\frac{(\mu+\sigma)(u-1)\log(1+\mu)}{\mu\sigma(1+\mu)^{\frac{\mu+\mu^2+\sigma}{\mu\sigma}}}\right].$$

7. Order statistics

Let $Z_{(1)},Z_{(2)},Z_{(3)},...,Z_{(n)}$ be the ordered statistics of the random sample $Z_1,Z_2,Z_3,...Z_n$ drawn from the discrete distribution with cdf $F_Z(z)$ and pmf $P_Z(z)$, then the pmf of the rth order statistics $Z_{(r)}$ is given by:
$f_{z(r)}(z,\mu,\sigma)=\dfrac{n!}{(r-1)!(n-r)!}P(z)[F(z)]^{r-1}[1-F(z)]^{n-r}$. r=1, 2, 3,…,n

Using the equations (2.1) and (2.2), the probability density function of the rth order statistics of the Poisson weighted exponential distribution is given by
$$f_{(r)}(z,\mu,\sigma)=\frac{n!}{(r-1)!(n-r)!}\frac{\mu^2}{\mu+\sigma}\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}}\right]\left[1-\frac{\mu+\mu^2+\sigma+2\mu\sigma+\mu\sigma x}{(\mu+\sigma)^{x+2}(\mu+\sigma)}\right]^{r-1}\left[\frac{\mu+\mu^2+\sigma+2\mu\sigma+\mu\sigma z}{(\mu+\sigma)^{z+2}(\mu+\sigma)}\right]^{n-r}$$

Then, the pmf of the first order $Z_{(1)}$ Poisson weighted exponential distribution is given by

$$f_1(Z;\mu,\sigma)=n\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}}\right]\left[\frac{\mu+\mu^2+\sigma+2\mu\sigma+\mu\sigma z}{(\mu+\sigma)^{z+2}(\mu+\sigma)}\right]^{n-1}.$$

And the pmf of the nth order $Z_{(n)}$ Poisson weighted exponential model is given as $f_{(n)}(z,\mu,\sigma)=n\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}}\right]\left[1-\frac{\mu+\mu^2+\sigma+2\mu\sigma+\mu\sigma z}{(\mu+\sigma)^{z+2}(\mu+\sigma)}\right]^{n-1}$.

8. Estimation of Parameters

In this section, we estimate the parameters of the Poisson weighted exponential distribution using methods of moments and the method of maximum likelihood estimation.

8.1. Method of Moments

In order to obtain sample moments, we replace population moments with sample moments:

$$a=\mu_1'=\frac{\mu+2\sigma}{\mu(\mu+\sigma)}$$

$$b=\mu_2'=\frac{\mu(\mu+2\sigma)+2(\mu+3\sigma)}{\mu^2(\mu+\sigma)}.$$

Solving the above equations of sample moments, we get

$$\hat{\sigma}=\frac{\hat{\mu}-a\hat{\mu}^2}{a\hat{\mu}-2}$$

$$\hat{\mu}=\frac{2a+\sqrt{2a-2b+4a^2}}{b-a}.$$

Theorem 8.1: The MOM estimator $\hat{\mu}$ *of* μ is positively biased.

$Proof: Let\ \hat{\mu}=h(\overline{Z})\ where\ \ h(t)=\frac{1-\sigma t+\sqrt{\sigma^2t^2+6\sigma t+1}}{2t}$ $,t>0$

$$Since\ \ h''(t)=\frac{1}{t^3}+\frac{9\sigma t+15\sigma^2t^2+3\sigma^3t^3+}{t^3(1+\sigma^2t^2+6\sigma)^{\frac{3}{2}}}>0$$

Then, h(t) is strictly convex. Hence, by Jensen's inequality, we have

$E\{h(\bar{z})\}>h\{E(\bar{z})\}$

$$finally, since\ h\{E(\bar{z})\}=h(\mu)=h\left(\frac{2\sigma+\mu}{\mu(\mu+\sigma)}\right)=\mu,\ we\ obtain\ E(\hat{\mu})>\mu.$$

Theorem 8.2: The MOM estimator $\hat{\mu}$ *of* μ is consistent and asymptotically normal.

$$\sqrt{n}\,(\hat{\mu}-\mu)\rightarrow_d N(0,\upsilon^2(\mu))$$

$$\upsilon^2(\mu)=\frac{\mu^2(\sigma+\mu)^2(2\sigma^2+2\mu\sigma^2+3\sigma\mu^2+4\sigma\mu+\mu^3+\mu^2)}{(2\sigma^2+\mu^2+4\sigma\mu)^2}.$$

Where

prof: Consistency Since $\mu < \infty$, *then* $\bar{Z} \xrightarrow{p} \mu$. *Also,* sin *ceh(t) is continuous funcationat t=*μ

$$then\ h(\bar{z}) \xrightarrow{p} h(\mu), i.e, \hat{\mu} \xrightarrow{p} \mu$$

Asymptotic normality: Since $\sigma^2 < \infty$, then by the central limit theorem, we have

$$\sqrt{n}\left(\bar{Z} - \mu\right) \xrightarrow{d} N(0, \sigma^2)$$

Also, since $h(\mu)$ *is differentiable and* $h'(\mu) \neq 0$, by the delta-method, we have

$$\sqrt{n}(h(\bar{z}) - h(\mu)) \rightarrow_d N(0, \upsilon^2(\eta))$$

$$\upsilon^2(\mu) = \frac{\mu^2(\sigma+\mu)^2(2\sigma^2 + 2\mu\sigma^2 + 3\sigma\mu^2 + 4\sigma\mu + \mu^3 + \mu^2)}{(2\sigma^2 + \mu^2 + 4\sigma\mu)^2}.$$

Where

The theorem follows.

As a result of this, the asymptotic $100(1-\alpha)\%$ confidence interval for μ is given by

$$\hat{\mu} \pm z_{\frac{\alpha}{2}} \frac{\upsilon(\hat{\mu})}{\sqrt{n}},$$

Where $z_{\frac{\alpha}{2}}$ is the (1-$\frac{\alpha}{2}$) percentile of the standard normal distribution.

8.2. Method of Maximum Likelihood Estimation

This is one of the most useful methods for estimating the different parameters of the distribution. Let $Z_1, Z_2, Z_3, \ldots, Z_n$ be the random sample of size n draw from PWED, then the likelihood function of PWED is given as

$$L(z \mid \mu, \sigma) = \frac{\mu^{n2}}{(\mu+\sigma)^n} \prod_{i=1}^{n} \left(\left[\frac{(1+\mu)+\sigma(z+1)}{(1+\mu)^{z+2}} \right] \right)$$

$$\log L = 2n \log \mu - n \log(\mu+\sigma) + \sum_{i=1}^{n} \log((1+\mu)+\sigma(z+1)) - (\sum_{i=1}^{n} z_i + 2n) \log(1+\mu)$$

$$\frac{\delta}{\delta\mu} \log L = \frac{2n}{\mu} - \frac{n}{\mu} + \sum_{i=1}^{n} \frac{1}{((1+\mu)+\sigma(z+1)} - \frac{\sum_{i=1}^{n} z_i + 2n}{(1+\mu)} = 0$$

$$\frac{\delta}{\delta\sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(z+1)}{((1+\mu)+\sigma(z+1)} = 0$$

The above equations can be solved numerically by using R software (3.5.2).

9. Applications of Poisson weighted exponential distribution

In this section, we fit our proposed distribution to a data set representing vaccine adverse event counts and the number of claims in automobile insurance so as to illustrate our claim that our proposed model fits well when compared to other competing models. The data sets are given in Table 2 and 5 respectively. In Table 6 the degree of freedom is zero for some distributions, and hence p-value is not given and thus in such tables comparisons can be done on the basis of the AIC and BIC values.

Table 2. Dataset representing vaccine adverse event counts (see C. E. Rose, S.W. Martain, K. A. Wannemueler, B. D. Plikaytis (2006))

Counts	0	1	2	3	4	5	6	7	8	9	10	11	12
Actual	1437	1010	660	428	236	122	62	34	14	8	4	4	1

We compute the expected frequencies for fitting Poisson Weighted Exponential (PWED), Zero Inflated Poisson (ZIPD), Negative Binomial (NBD), Geometric (GD), Poisson Lindley (PLD), Poisson Akash (PAD), Poisson Distribution (PD) and Discrete Generalized Inverse Weibull Distribution (DGIWD) with the help of R studio statistical software, and Pearson's chi-square test is applied to check the goodness of fit of the models discussed. The calculated figures are given in Table 3 and 6. Based on the chi-square, we observe that the Poisson weighted exponential distribution provides a satisfactorily better fit for the data set representing vaccine adverse event counts in Table 3 and the number of claims in automobile insurance in Table 6 as compared to other distributions. Also the parameters are estimated by using the ML method. We have analysed the data using R software (3.5.2). Parameter estimates along and the model function of the fitted distributions are given in Table 3 and 6.

Table 3. Fitted proposed distribution and other competing models to a dataset representing vaccine adverse event counts

Z	Obs.freq	PD	ZIPD	NBD	GD	PLD	DGIW	PAD	PWED
0	1437	890.75	1436.4	1401.7	1603.5	1500.1	1354	1500.2	1417
1	1010	1342.35	810.6	1065.3	963.9	1003.5	1377.2	977.65	1048
2	660	1011.4	789.6	671.15	579.4	629.2	524.15	632.55	670
3	428	508	529.5	393.75	348.3	378.7	248.9	392	397.4
4	236	191.4	274.6	222.5	209.35	221.6	139	232.1	225.1
5	122	57.7	117.3	122.8	125.85	127	86.45	132.1	123.6
6	62	14.5	42.9	66.7	75.65	71	57.9	72.8	66.3
7	34	3.1	13.8	35.8	45.45	39.9	41	39	35
8	14	0.6	4	19.1	27.35	22	30.2	20.45	18.2

Table 3. Fitted proposed distribution and other competing models to a dataset representing vaccine adverse event counts (cont.)

Z	Obs.freq	PD	ZIPD	NBD	GD	PLD	DGIW	PAD	PWED
9	8	0.1	1	10.1	16.45	12	23	10.5	9.4
10	4	0.1	0.3	5.3	9.9	7	18	5.35	4.8
11	4	0.1	0.1	2.75	5.9	3.55	14	2.65	2.9
12	1	0.1	0.1	1.45	8.95	4	105	2.55	1.7
Total	4020								
d.f		5	6	8	9	9	7	9	8
Chi-square		1901	570	10.4	78.68	20.12	488	14.16	6.21
Parameter estimate (S.E)		$\hat{\lambda}=1.5$ (0.019)	$\hat{\mu}=2.04$ (0.031) $\hat{\sigma}=0.261$ (0.009)	$\hat{r}=1.52$ (0.07) $\hat{p}=0.5$ (0.01)	$\hat{p}=0.398$ (0.004)	$\hat{\theta}=0.99$ (0.016)	$\hat{a}=1.49$ (0.02) $\hat{b}=0.47$ (0.17) $\hat{t}=0.037$ (0.068)	$\hat{\lambda}=1.35$ (0.016)	$\hat{\mu}=1.14$ (0.04) $\hat{\sigma}=2.99$ (0.9)
p-value		0.00	0.00	0.23	0.00	0.017	0.00	0.11	0.62

Furthermore, from Table 4 and 7, it has been observed that the Poisson weighted exponential distribution have the lesser AIC and BIC values as compared to other competing models. Hence, we can conclude that the PWED leads to a best fit as compared to other competing models for analysing the data set given in Table 2 and 5.

Table 4. Model comparison criterion for fitted models to a data set representing vaccine adverse event counts

Criterion	PD	NBD	ZIPD	GD	PLD	DGIW	PAD	PWED
AIC	14464.2	13485.2	13741.5	13558	13494	14049.8	13487	13480
BIC	14470.5	13486.33	13754.1	13564.3	13500.3	14068.7	13493.3	13481

Table 5. Data set representing the number of claims in automobile insurance (see Klugman et al. (2012))

Claim counts	0	1	2	3	4
Observed frequency	1563	271	32	7	2

Table 6. Fitted proposed distribution and other competing models to a data set representing the number of claims in automobile insurance

Z	Obs.fre	PD	ZIPD	NBD	GD	PLD	DGIW	PAD	PWED
0	1563	1544.1	1562.9	1566.4	1570.1	1569.5	1562.7	1571.9	1564.4
1	271	299.8	265.2	261.5	255.25	256.34	274.6	252.7	268.4
2	32	29.1	42	40.15	41.5	41.34	27.15	41.85	35.7
3	7	1.9	4.45	6.6	6.75	6.6	6.3	7	5.6
4	2	0.1	0.35	1	1.3	1.0444	4.25	1.45	0.75
Total	1875								
d.f		1	-	1	2	2	-	2	1
Chi-square		57.04	-	3.61	3.29	3.87	-	3.72	1.51
Parameter estimate (S.E)		$\hat{\lambda} = 0.194$ (0.01)	$\hat{\alpha} = 0.31$ (.042) $\hat{\beta} = .38$ (.07)	$\hat{p} = 6.13$ (0.4) $\hat{r} = 1.19$ (0.07)	$\hat{p} = 0.83$ (0.007)	$\hat{\theta} = 05.89$ (0.3)	$\hat{a} = 3.16$(. $\hat{b} = .4$(.17 $\hat{t} = .03$(.17	$\hat{\theta} = 05.74$ (0.27)	$\hat{\alpha} = 7.87$ (2.3) $\hat{\beta} = 8.83$ (18.4)
p-value		0.00	-	0.05	0.19	0.14	-	0.15	0.22

Table 7. Model comparison criterion for fitted models to a data set representing the number of claims in automobile insurance counts

Criterion	PD	NBD	ZIPD	GD	PLD	DGIW	PAD	PWED
AIC	2005.5	1991.1	1995.5	1989.8	1989.7	1994.6	1990.25	1987
BIC	2011	1990.3	2006.2	1995.3	1995.25	2011.2	1995.8	1986.3

10. Conclusion

A new over-dispersed probability distribution is introduced using the compounding technique. Statistical properties of the proposed model are studied and applications in handling count data sets representing vaccine adverse counts and insurance claims are analysed.

References:

Berreto-Souza, W., Bakouch, H., (2013). A new lifetime model with decreasing failure rate. *A journal of theoretical and Applied Statistics*, 47(2), pp. 465–476 .

Hajebi,M., Rezaei, S. and Nadarajah, S., (2013). An exponential- negative binomial distribution. *REVSTAT-Statistical journal*, 11(2), pp. 191–210.

Asgharzadeh, A., Bakouch,H.S., Nadarajah, S. and Esmaeili, L., (2014). A new family of compound lifetime distributions. *Kybernetika*, 50, pp. 142–169.

Mohmoudi,E. Jafari, A. A., (2014). The compound class of linear failure rate power series distributions: model, properties and applications. *Communications in statistics-simulation and computation.*

Silva, R. B., Cordeiro, G. M., (2015). The Burr XII power series distribution: A new compounding family. *The Brazilian journal of probability and statistics*, 29, pp. 565–589.

Pinho, L. G. B., Cordeiro, G. M. and Nobre, J. S., (2015). On the Harris-G class of distributions: general results and application. *Brazilian journal of probability and statistics*, 29(4), pp. 813–832.

Bourguignon, M., Silva, R. B. and Cordeiro, G. M., (2014). Anew class of fatigue life distributions. *Journal of statistics Computation and Simulation*,84 , pp. 2619–2635.

Bordbar, F., Nematollahi, A. R., (2016). The modified exponential- geometric distribution. Communications in Statistics –*Theory and Methods*, 45(1), pp. 173–181.

Flores, D, J., Borges, P., Cancho, G. and Louzada, F., (2013). The complementary exponential power series distribution. *Brazilian journal of probability and statistics*, 27, pp. 565–584.

Ghitany M. E., Alqallaf F., Al- Mutairi D. K. and Husain H. A.,(2011). A two parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in Simulations*, 81, pp. 1190–1201.

Christophe Chesneau, Hassan S. Bakouch, Tassaddaq Hussain and Bilal A. Para, (2020). The cosine geometric distribution with count data modeling, Journal of Applied Statistics, DOI: 10.1080/02664763.2019.1711364.

R Core Team, (2019). R version 3.5.3: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Klugman S., Panjer H., Willmot G., (2012). Loss Models. From Data to Decisions. *John Wiley and Sons, New York.*

C. E. Rose, S. W.Martin, K. A. Wannemuehler, and B. D. Plikaytis, (2006). On the use of zero infilated and hurdle models for modeling vaccine adverse events count data, Journal of Biopharmaceutical Statistics, Vol. 16, pp. 463–481.