HURDLE POISSON – SUSHILA DISTRIBUTION AND ITS APPLICATION

Chookait Pudprommarat

Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand
E-Mail: chookait.pu@ssru.ac.th

ABSTRACT

This study proposes hurdle Poisson - Sushila distribution and its application. The probability mass function of random variable for hurdle Poisson - Sushila which have three parameters. The probability mass function of hurdle Poisson - Sushila has the mathematical properties such as moments about the origin, first - fourth moments about the origin, mean, variance, coefficient of skewness, and coefficient of kurtosis. Next, study of estimating parameters for hurdle Poisson - Sushila distribution with maximum likelihood method. Finally, application study are used a real data set which real data are fitted by the Poisson distribution, Poisson - Sushila distribution and hurdle Poisson - Sushila distribution by using the chi-square goodness of fit test. The maximum likelihood method provides very poor fit for the Poisson distribution and acceptable fits for hurdle Poisson - Sushila distribution.

Keywords: Poisson distribution, Poisson - Sushila distribution, Hurdle Poisson - Sushila distribution

INTRODUTION

Poisson distribution is a basic distribution for count data which mean and variance are equal. In some data set of count data, mean and variance are not equal. If variance is more than mean, it calls over-dispersion. Many statisticians construct a new distribution for count data to solve over-dispersion problem such as Poisson-uniform [1], Poisson - lognormal [2], Poisson- gamma [3], Poisson-Pareto [4], Poisson - Shifted Pareto [4], Poisson-Lomax [5], Poisson-Power Variance [6], Poisson – Sushila distribution [7], etc.

Poisson – Sushila distribution is mixed distribution between Poisson and Sushila distribution [8] was introduced Saratoon [7] which probability mass function is given by

$$f_Y(y) = \frac{\theta^2 \alpha^y}{(\theta+1)(\theta+\alpha)^{y+2}}\left[\theta+\alpha+y+1\right] \quad y = 0,1,2,\ldots$$

and 1st – 4th moment about the origin are given by

$$\mu_1' = \frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2 - \theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right), \qquad \mu_2' = \frac{(2\theta+6)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2 - \theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^2,$$

$$\mu_3' = \frac{(6\theta+24)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2 - \theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^3 \text{and} \qquad \mu_4' = \frac{(24\theta+120)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2 - \theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^4.$$

Hurdle model of count data is the hurdle at zero which is alternative ways to new hurdle model in applied probability such as Poisson hurdle model [9], hurdle negative binomial model [10], etc.

The aim of this study is to present hurdle Poisson - Sushila distribution. Next, we consider some properties of the hurdle Poisson - Sushila distribution which are the moment about the origin, mean, variance, skewness and kurtosis. The parameters estimating of hurdle Poisson - Sushila distribution are maximum likelihood method. Finally, we have shown fitting distribution between Poisson and hurdle Poisson - Sushila distribution based on real data set.

OBJECTIVES

The objectives of this study are

1) To present the probability mass function of hurdle Poisson - Sushila random variable.
2) To study some properties of hurdle Poisson - Sushila distribution.
3) To derive parameters estimation of hurdle Poisson - Sushila distribution.
4) To study application of hurdle Poisson - Sushila distribution.

METHODOLOGY

The methodologies of this study are as follows:

1) Investigate hurdle Poisson - Sushila distribution.

2) Derive some properties of hurdle Poisson - Sushila distribution such as moment about the origin, mean, variance, skewness and kurtosis.

3) Derive parameters estimation of hurdle Poisson - Sushila distribution by using maximum likelihood method.

4) Application study is used of real data set fitting hurdle Poisson - Sushila distribution.

RESULTS

The results of hurdle Poisson - Sushila distribution and its application are as follow:

I. The probability mass function of random variable for hurdle Poisson – Sushila

Theorem 1: Suppose $X \sim HDPS\ (p,\theta,\alpha)$ be a random variable of Poisson - Sushila with parameter $0 < p < 1$, $\theta > 0$ and $\alpha > 0$, then the probability mass function of random variable for hurdle Poisson – Sushila is given by

$$f(x)=\begin{cases} p & ; x=0 \\ (1-p)\cdot\dfrac{\left(\dfrac{\theta^2\alpha^x}{(\theta+1)(\theta+\alpha)^{x+2}}\left[\theta+\alpha+x+1\right]\right)}{1-\left(\dfrac{\theta^2}{(\theta+1)(\theta+\alpha)^2}\left[\theta+\alpha+1\right]\right)} & ; x=1,2,3... \end{cases}$$

Proof. If Y is random variable of Poisson - Sushila with parameter θ and α, the probability mass function of random variable for Poisson – Sushila is given by

$$f_Y(y)=\frac{\theta^2\alpha^y}{(\theta+1)(\theta+\alpha)^{y+2}}\left[\theta+\alpha+y+1\right] \quad y=0,1,2,\ldots$$

Using hurdle model, it is given by

$$f(x)=\begin{cases} p & ; x=0 \\ (1-p)\cdot f_y(x) & ; x=1,2,3... \end{cases}$$

By substituting Poisson – Sushila distribution into hurdle model, then the probability mass function of random variable for hurdle Poisson – Sushila is given by

$$f(x)=\begin{cases} p & ; x=0 \\ (1-p)\cdot\dfrac{\left(\dfrac{\theta^2\alpha^x}{(\theta+1)(\theta+\alpha)^{x+2}}\left[\theta+\alpha+x+1\right]\right)}{1-\left(\dfrac{\theta^2}{(\theta+1)(\theta+\alpha)^2}\left[\theta+\alpha+1\right]\right)} & ; x=1,2,3... \end{cases}$$

Next, we display some of probability mass function of random variable for hurdle Poisson – Sushila in Figure 1 and 2.

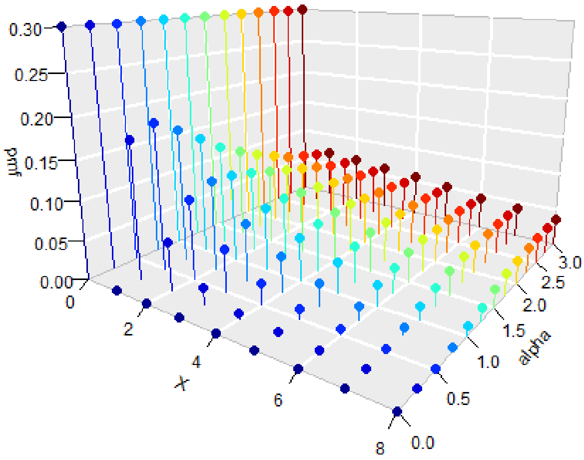

Figure 1 The probability mass function of random variable for hurdle Poisson – Sushila with some parameters: $\theta = 0.7$ and $p = 0.3$.

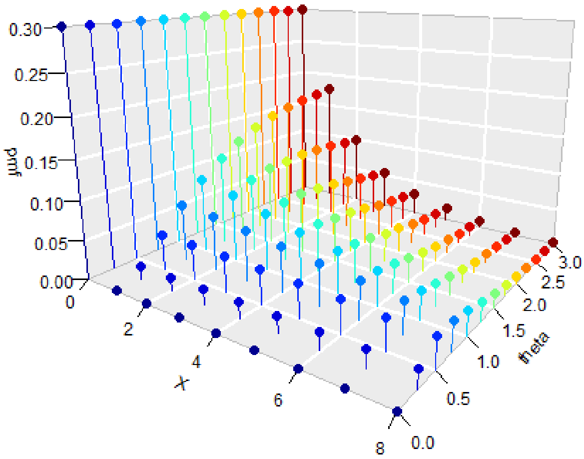

Figure 2 The probability mass function of random variable for hurdle Poisson – Sushila with some parameters: $\alpha = 3$ and $p = 0.3$.

II. The mathematical properties of hurdle Poisson - Sushila distribution

The probability mass function of hurdle Poisson - Sushila has the mathematical properties such as the r th-order moment about the origin, mean, variance, skewness and kurtosis.

Theorem 2: Let $X \sim HDPS\ (p,\theta,\alpha)$ then the r th moment about the origin is given by

$$\mu_r' = E(X^r) = (1-p)\frac{r!(\theta+r+1)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^r, \quad 0<p<1, \alpha>0, \theta>0, r=1,2,\ldots$$

Proof. If $X \sim HDPS\ (p,\theta,\alpha)$ then the r th moment about the origin is given by

$$\mu_r' = \sum_{i=1}^{\infty} x^r f(x) \quad = \sum_{i=1}^{\infty} x^r (1-p)\cdot\frac{f(x)}{1-f(0)} \quad = (1-p)\sum_{i=1}^{\infty} x^r \frac{f(x)}{1-f(0)}$$

$$\mu_r' = E(X^r) = (1-p)\frac{\dfrac{r!(\theta+r+1)}{(\theta+1)}\left(\dfrac{\alpha}{\theta}\right)^r}{1-\left(\dfrac{\theta^2}{(\theta+1)(\theta+\alpha)^2}\left[\theta+\alpha+1\right]\right)} = (1-p)\frac{r!(\theta+r+1)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^r.$$

Moment about the origin: r th moment about the origin in **Theorem 2**, then 1st – 4th moment about the origin are given by

$$\mu_1' = (1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right), \quad \mu_2' = (1-p)\frac{(2\theta+6)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^2,$$

$$\mu_3' = (1-p)\frac{(6\theta+24)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^3 \text{ and } \mu_4' = (1-p)\frac{(24\theta+120)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^4.$$

Mean: $\mu_1' = (1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right).$

Variance: $Var(X) = E(X^2)-\left[E(X)\right]^2 = (1-p)\frac{(2\theta+6)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right) - \left[(1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right]^2.$

Skewness: $Skewness = \frac{E(X^3)-3E(X)(Var(X))-E(X^3)}{\sqrt{Var(X)^3}}$

$$= \left((1-p)\frac{(6\theta+24)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^3\right) - 3\left((1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right)$$
$$\times\left((1-p)\frac{(2\theta+6)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right) - \left((1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right) -$$
$$\times\left((1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right)^3$$
$$/\sqrt{\left[\left(1-p)\frac{(2\theta+6)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right) - \left((1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right)\right]^3}.$$

Kurtosis: $Kurtosis = \frac{E(X^4)-4E(X^3)E(X)+6E(X^2)\left[E(X)\right]^2-3E(X)^4}{(Var(X))^2}$

$$= \left((1-p)\frac{(24\theta+120)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^4\right) - 4\left((1-p)\frac{(6\theta+24)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^3\right)$$
$$\times\left((1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right) + 6\left((1-p)\frac{(2\theta+6)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)^2\right.$$
$$\left.\times\left((1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right)^2\right) - 3\left((1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right)^4$$
$$/\left[(1-p)\frac{(2\theta+6)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right) - \left[(1-p)\frac{(\theta+2)(\theta+\alpha)^2}{(\theta+1)(\theta+\alpha)^2-\theta^2(\theta+\alpha+1)}\left(\frac{\alpha}{\theta}\right)\right]\right]^2.$$

III. Estimating parameters of hurdle Poisson - Sushila distribution with maximum likelihood method

This section, we consider estimating parameters of hurdle Poisson - Sushila distribution with maximum likelihood method.

Let $X_1, X_2, ..., X_n$ be observed value from hurdle Poisson - Sushila distribution, the likelihood function of hurdle Poisson - Sushila distribution is given by

$$L(\theta,\alpha,p\,|\,x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} p^{1-a_i}\left((1-p)\frac{\theta^2\alpha^x(\theta+\alpha+x+1)}{(\theta+1)(\theta+\alpha)^{x+2}}\right)^{a_i}, \text{ where } a_i = \begin{cases}1 \ , & x_i = 0 \\ 0 \ , & x_i \neq 0\end{cases}$$

and log- likelihood function:

$$\log L = n_0 \log p + \left(\sum_{i=1}^{x_i} a_i \log(1-p) + \sum_{i=1}^{x_i} a_i 2\log\theta - \sum_{i=1}^{x_i} a_i \log(\theta+1) + \sum_{i=1}^{x_i} a_i (x+2)\log(\theta+\alpha)\right.$$
$$\left. + \sum_{i=1}^{x_i} a_i x \log\alpha + \sum_{i=1}^{x_i} a_i \log(\theta+\alpha+x+1)\right)$$

The estimating parameters can be obtained by the first partial derivatives of log- likelihood function with respect to θ, α and p. Then, it gives rise to following equations:

$$\frac{d}{d\theta}\log L = \sum_{i=1}^{x_i} a_i \frac{2n}{\theta} - \sum_{i=1}^{x_i} a_i \frac{1}{(\theta+1)} + \sum_{i=1}^{x_i} a_i \frac{(x_i+2)}{(\theta+\alpha)} + \sum_{i=1}^{x_i} a_i \frac{1}{(\theta+\alpha+x_i+1)},$$

$$\frac{d}{d\alpha}\log L = \sum_{i=1}^{x_i} a_i \frac{(x_i+2)}{(\theta+\alpha)} + \sum_{i=1}^{x_i} a_i \frac{x_i}{\alpha} + \sum_{i=1}^{x_i} a_i \frac{1}{(\theta+\alpha+x_i+1)},$$

$$\frac{d}{dp}\log L = \left(\frac{n_0}{p}\right) - \left(\sum_{i=1}^{x_i} a_i \frac{1}{1-p}\right) \text{ ,where } n_0 = \sum_{i=1}^{n} a_i .$$

These three equations are solved iteratively till sufficiently close values of $\hat{\theta},\hat{\alpha}$ and $\hat{p}$ by using function nlm in stats p

IV. Application Study

This application study, we use real data set in [11] which is number of visits and hospital stays for a sample of United States residents aged 66 and over. The Poisson and hurdle Poisson - Sushila distribution have been fitted to a real data set. It is found that to the hurdle Poisson - Sushila distribution is a better fit than Poisson which have shown in Table 1.

Table 1 Observed and expected frequency of number of visits and hospital stays for a sample of United States residents aged 66 and over

Number of visits and hospital stays for a sample of United States residents aged 66 and over	Observed Frequency	Expected Frequency	
		Poisson	Hurdle Poisson - Sushila
0	434	397.55	433.61
1	143	189.91	140.29
2	42	45.36	45.39
3	14	8.18 (3–6)	14.68
4	5		7.03 (4–6)
5	1		
6	2		
Total	641	641	641
Estimated parameters		$\hat{\theta}$ = 0.4777	$\hat{\theta}$ = 146.3787 $\hat{\alpha}$ = 69.5375 $\hat{p}$ = 0.6771
Chi-square		38.5590	0.3097
Degree of freedom		2	1
P-value		<0.0001	0.5779

CONCLUSION

In this study, we have introduced hurdle Poisson - Sushila distribution as an extension to Poisson - Sushila distribution. In particular, some mathematical properties are introduced such as the r th-order moment about the origin, mean, variance, skewness and kurtosis. Parameter estimation is also implemented by using maximum likelihood method. In application study, the hurdle Poisson - Sushila distribution is a better fit than Poisson.

ACKNOWLEDGEMENT

This research was successful by financial support by Suan Sunandha Rajabhat University and many helps from my friends and students from Faculty of Science and Technology, Suan Sunandha Rajabhat University. The author really appreciates and many thanks for all.

REFERENCES

[1] Bhattacharya, S.K. (1966). Confluent Hypergeometric Distributions of Discrete and Continuous type with Application to Accident Proneness, *Bulletin of the Calcutta Statistical Association*. 15, p.p. 20 – 31.

[2] Bulmer, M.G. (1974). On Fitting the Poisson Lognormal Distribution to Species Abundance Data. *Biometrics*. 30, p.p. 101–110.

[3] Greenwood, M. & Yule, G. (1920). An Inquiry 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 A*. 83, p.p. 255–279.

[4] Willmot, G.E. (1993). On Recursive Evaluation of Mixed Poisson Probabilities and Related Quantities. *Scandinavian Actuarial*. 18, pp. 114 – 133.

[5] Al – Awadhi and Ghitany. (2001). Statistical Properties of Poisson – Lomax Distribution and its Application to Repeated Accidents Data. *Applied Statistical Science*. 10, pp. 365 – 372.

[6] Hougaard, P., Lee, M.L.T. and Whitmore, G.A. (1997). Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes. *Biometrics*. 53, pp. 1225–1238.

[7] Saratoon, A. (2017). *Poisson –Sushila distribution and its application*, *Suan Sunandha Rajabhat University*. Thailand.

[8] Rama Shanker et al. (2013). Sushila Distribution and lts Application to Waiting Times data. *Opinion – International Journal of Business Management*. 3 (2), pp. 1-11.

[9] Brian Neelon, Pulak Ghosh and Patrick F. Loebs. (2013). A Spatial Poisson Hurdle Model for Exploring Geographic Variation in Emergency Department Visits. *HHS Public Access*. 172 (2), pp. 389-413.

[10] Seyed Ehsan Saffari et.al. (2012). Hurdle negative binomial regression model with right censored count data. *SORT*. 3 (2), pp.181-194.

[11] Flynn, M. (2009). More flexible GLMs zero-inflated models and hybrid models. *Casualty Actuarial Society*. pp. 148-224.