

Original Article

The zero-inflated negative binomial-Erlang distribution: An application to highly pathogenic avian influenza H5N1 in Thailand

Winai Bodhisuwan*, Siriporn Samutwachirawong, and Prasit Payakkapong

*Department of Statistics, Faculty of Science, Kasetsart University,
Chatuchak, Bangkok, 10900 Thailand*

Received: 19 December 2016; Revised: 1 September 2017; Accepted: 7 September 2017

Abstract

In this study, we propose the zero-inflated negative binomial-Erlang distribution for modeling the number of highly pathogenic avian influenza H5N1 outbreaks that were reported at the subdistrict level in Thailand. We provide the probability mass function, mean, variance, skewness, and kurtosis for the zero-inflated negative binomial-Erlang distribution. In addition, this distribution provides a better fit compared to the zero-inflated Poisson and the conventional zero-inflated negative binomial distributions. In Thailand, a highly pathogenic avian influenza virus of the H5N1 subtype was first confirmed in humans in 2004. The epidemic has only caused sporadic outbreaks in Thailand during the second epidemic wave, particularly in the central and northern regions. The zero-inflated Poisson distribution can be used efficiently in epidemiologic count data with an excess number of zeros. However, when the data exhibit overdispersion, an alternative to the zero-inflated Poisson distribution may be considered such as the zero-inflated negative binomial-Erlang distribution.

Keywords: negative binomial-Erlang distribution, zero-inflated distribution, count data analysis, overdispersion, H5N1

1. Introduction

The Poisson distribution plays a central role in count data analysis. For example, when a reliable manufacturing process is in control, the number of defects on an item should be Poisson distributed (Shewhart & Wilks, 2005). However, many count data often exhibit overdispersion where variance is greater than the mean (Rainer, 2008). Greenwood and Yule (1920) proposed the negative binomial (NB) distribution which is a mixture of Poisson distributions around the mean distributed as a gamma distribution. The NB distribution is often employed to analyze overdispersed data (Rainer, 2008).

In addition, the NB distribution can be mixed with other distributions to be an alternative distribution of overdispersed data. It has been shown that the mixed negative binomial distribution provided a better fit to count data compared to the NB distribution, such as the negative binomial-Pareto

(Meng *et al.*, 1999), the negative binomial-inverse Gaussian (Gomez *et al.*, 2008), the negative binomial-Lindley distribution (Zamani & Ismail, 2010), and the negative binomial-beta exponential distribution (Pudprommarat & Bodhisuwan, 2012).

Another mixed NB distribution is the negative binomial-Erlang (NB-EL) distribution, which was presented by Kongrod *et al.* (2014). This distribution was obtained by mixing the distribution of NB (r, p), where $p = \exp(-\lambda)$ and λ is an Erlang (EL) distribution. The NB-EL distribution provides a better fit than the Poisson and the NB distributions and represents the alternative distribution for overdispersed data. Furthermore, the fit improves as the number of zeros increase for some real data sets (Kongrod *et al.*, 2014).

Count data with excess zeros, is frequently described by the zero-inflated (ZI) distribution. The ZI distribution is a mixture model that may also be referred to as a type of finite mixture model. The ZI distribution is used when one has a theory for why there are so many zeros (Hilbe, 2014). The ZI distribution has been extensively used in many fields, e.g., Lambert (1992) proposed the zero-inflated Poisson (ZIP) regression model with an application of defects in

*Corresponding author
Email address: fsciwnb@ku.ac.th

manufacturing. The ZIP distribution is assumed to be a mixture between a Poisson distribution and a distribution to a point mass of zero. Greene (1994) proposed the zero-inflated negative binomial (ZINB) distribution which is assumed to be a mixture between a negative binomial distribution and a distribution to a point mass of zero. The ZINB distribution was applied to fit the number of derogatory reports to a credit reporting agency for a group of credit card applicants.

In this paper, we propose the zero-inflated negative binomial-Erlang (ZINB-EL) distribution, which is assumed to be distributed as a mixture of a point mass at zero and the NB-EL distribution. Some statistical properties such as mean, variance, skewness, and kurtosis will be derived. The parameters of the ZINB-EL distribution are estimated by the maximum likelihood estimation (MLE). The usefulness of the ZINB-EL distribution is that it can be an alternative distribution for describing the distribution of the number of reported outbreaks per subdistrict in Thailand, which is an example of count data.

Clinical signs and symptoms caused by the avian influenza virus, commonly described as bird flu, resulted in great economic losses (WHO, 2014). The avian influenza pandemic caused thousands of illnesses, hundreds of fatalities, and it was a threat to global health. The highly pathogenic avian influenza is an example of important diseases whose emergence can, in part, be attributed to rapid changes in poultry farming conditions.

H5N1 was first reported in Guangdong Province, China in 1996 (Xu *et al.*, 1999). The spread of H5N1 virus assumed global dimensions. The outbreak began in 2003, predominantly affecting poultry farms in many Southeast Asian countries such as Cambodia, Indonesia, Lao, Thailand, and Vietnam (Gilbert *et al.*, 2007). The H5N1 outbreaks in Thailand were detected in the second epidemic wave from July 2004 to May 2005 (Gilbert *et al.*, 2006). Data on the outbreaks of H5N1 in Thailand have been collected since January 2004 by the Department of Livestock Development, Thailand. The outbreaks of H5N1 virus infection resulted in a high number of affected animals, losses in domestic and international trade of poultry products, and socioeconomic impacts on the livelihoods of farmers and public health (Tiensin *et al.*, 2009).

A common feature of the ecological data sets is their tendency to contain many zero values. Zero-inflation is often the result of a large number of true observed zeros caused by the real ecological effect of interest. However, the term can also be applied to data sets with excess zeros caused by false observations of zeros because of sampling or observer errors in the course of data collection (Martin *et al.*, 2005). A variety of mechanisms may generate excess zeros relative to otherwise standard distributions of count data, e.g., Poisson and NB distributions (Heilbron, 1994).

The number of H5N1 outbreaks reported in each subdistrict during the second epidemic wave from July 3, 2004 to May 5, 2005 in Thailand was likely to be structurally zero-inflated. The subdistricts where no outbreaks were detected could indeed be the subdistricts where no outbreaks occurred (true zero), but it could also be the subdistricts where at least one outbreak occurred, but none were reported (false zero) (Vergne *et al.*, 2014).

In our application study, we compared the performance of the proposed distribution with ZIP and ZINB distributions. We use the log-likelihood, the Akaike Information Criterion (AIC) (Akaike, 1974), the Bayesian Information

Criterion (BIC) (Schwarz, 1978), and P-values of the K-S test for the goodness of fit for model selection.

2. Materials and Methods

In this study, the ZINB-EL distribution is a mixture between a point mass at zero and NB-EL distribution. The parameters of the ZINB-EL distribution were estimated by using numerical optimization with the optim function in stats package of R language (R Core Team, 2015). We used the criteria of the log-likelihood, AIC, and BIC to evaluate goodness of fit. The real data set is applied to compare the efficiencies of fitting distributions based on the goodness of fit test, the Kolmogorov-Smirnov (K-S) from the dgof package (Arnold & Emerson, 2011) in the R language.

3. Zero-inflated Distributions

The ZI distribution is a mixture between a point mass at zero and any other count distribution supported on non-negative integers.

Let X be distributed as the ZI, then the probability mass function (pmf) of X is

$$h(x) = \begin{cases} \omega + (1-\omega)g(x; \Theta), & x=0 \\ (1-\omega)g(x; \Theta), & x=1,2,3,\dots, \end{cases} \quad (1)$$

where ω is a zero-inflation parameter ($0 < \omega < 1$), and $g(x; \Theta)$ is the pmf of X with a vector of parameter, $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ (Johnson *et al.*, 1992).

The ZIP distribution is a class of models for count data with excess zeros, described in Lambert (1992). The ZIP distribution is assumed to be distributed as a mixture of a Poisson distribution and a distribution to a point mass at zero, with mixing probability ω (Hall, 2000).

Let X be a ZIP random variable with parameters ω and λ , denoted as $X \sim ZIP(\omega, \lambda)$, the pmf of X is

$$g(x) = \begin{cases} \omega + (1-\omega)\exp(-\lambda), & x=0 \\ (1-\omega)\frac{\exp(-\lambda)\lambda^x}{x!}, & x=1,2,3,\dots, \end{cases}$$

where $\lambda > 0$ and $0 < \omega < 1$.

The mean and variance of the ZIP distribution are respectively,

$$E(X) = (1-\omega)\lambda \text{ and } Var(X) = (1-\omega)(\omega\lambda^2 + \lambda).$$

Count data are also often overdispersed. Consequently, other possible distributions such as the ZINB distribution may be more appropriate than the ZIP distribution. The ZINB distribution was discussed by Greene (1994). The ZINB distribution is a mixture distribution assigning a mass of ω to extra zeroes and mass $1-\omega$ to an NB distribution (Yang, 2006). Note that the NB distribution can also arise as a mixture of Poisson distributions around the mean distributed as a gamma distribution with scale parameter $\frac{(1-p)}{p}$ and shape parameter r .

Let X be a ZINB random variable with parameters ω , r , and p . It is denoted as $X \sim ZINB(r, p, \omega)$, and its pmf is

$$g(x) = \begin{cases} \omega + (1-\omega)p^r, & x = 0 \\ (1-\omega) \binom{r+x-1}{x} p^r (1-p)^x, & x = 1, 2, 3, \dots, \end{cases}$$

where $0 < \omega < 1$, $r > 0$, and $0 < p < 1$.

The mean and variance of the ZINB distribution are respectively,

$$E(X) = \frac{r(1-\omega)(1-p)}{p} \quad \text{and} \quad \text{Var}(X) = r(1-\omega) \left[\frac{(1+r\omega)(1-p)^2}{p^2} + \frac{(1-p)}{p} \right].$$

4. Results and Discussion

4.1 Zero-inflated negative binomial-Erlang distribution

We propose the ZINB-EL distribution for count data with excess zeros to describe the number of reported outbreaks per subdistrict in Thailand.

Definition 1:

Let X have a NB distribution with parameters $r > 0$ and $p = \exp(-\lambda)$, where λ is distributed as EL distribution with positive parameters c and k , then X is said to follow the NB-EL distribution with parameters r, c , and k (Kongrod *et al.*, 2014). The pmf of X is given by

$$g(x) = \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \left(\frac{c}{c+r+j} \right)^k, \quad x = 0, 1, 2, \dots, \tag{2}$$

where r, c , and $k > 0$.

Let X be a random variable of the NB-EL distribution with pmf $g(x)$ defined as in (2) and $h(x)$ in (1) is a pmf of the ZI distribution. Then $f(x)$ is a pmf of ZINB-EL distribution with the parameters r, c, k , and ω .

$$f(x) = \begin{cases} \omega + (1-\omega) \left(\frac{c}{c+r} \right)^k, & x = 0 \\ (1-\omega) \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \left(\frac{c}{c+r+j} \right)^k, & x = 1, 2, 3, \dots, \end{cases} \tag{3}$$

The ZINB-EL pmf plots with some specified parameter values of r, c, k , and ω are provided in Figure 1.

Theorem 1:

If $X \sim ZINB-EL(r, c, k, \omega)$, then some characteristics of X are as follows

(i) The mean and variance of X are respectively,

$$E(X) = r(1-\omega) \left(\left(\frac{c}{c-1} \right)^k - 1 \right), \tag{4}$$

$$\begin{aligned} \text{Var}(X) = r(1-\omega) & \left((r+1) \left(\frac{c}{c-2} \right)^k - (2r+1) \left(\frac{c}{c-1} \right)^k + r \right) \\ & - \left(r(1-\omega) \left(\left(\frac{c}{c-1} \right)^k - 1 \right) \right)^2. \end{aligned} \tag{5}$$

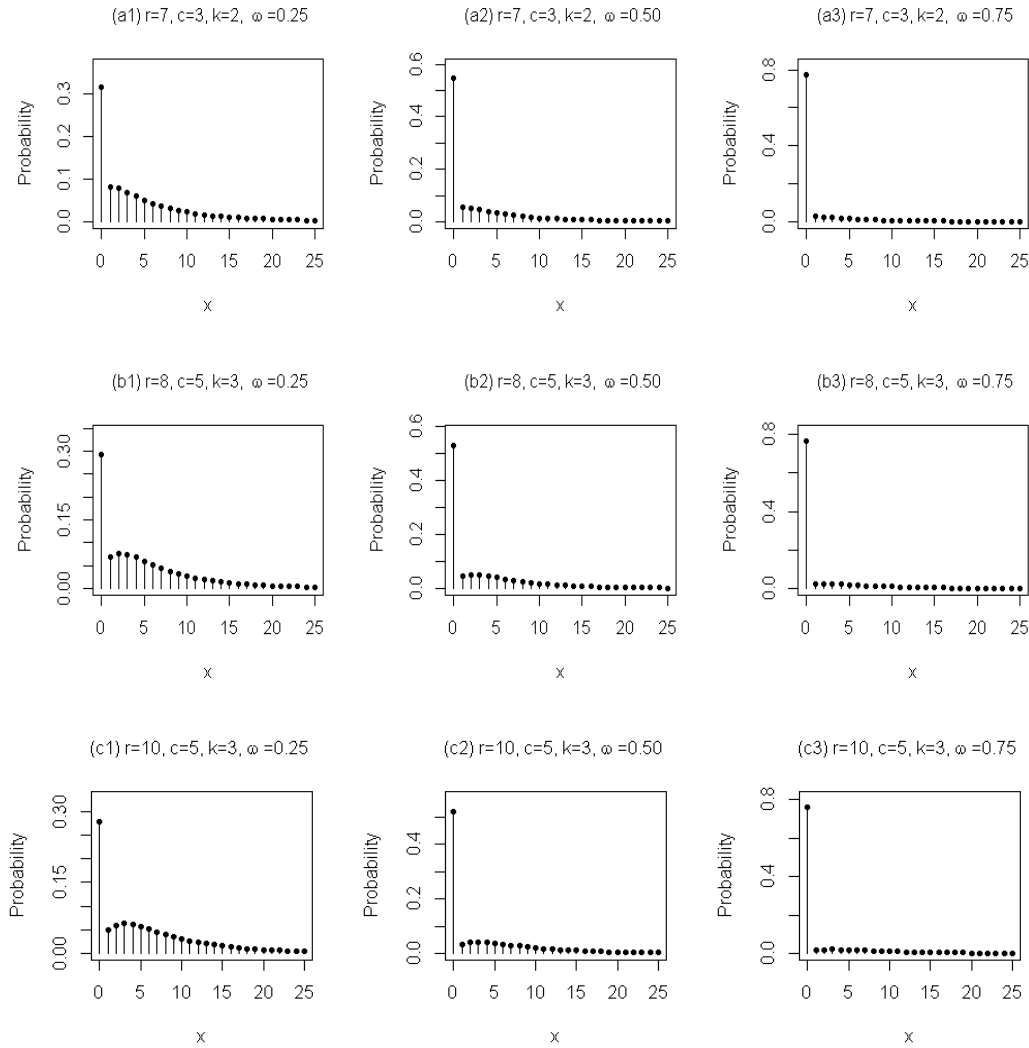


Figure 1. The ZINB-EL pmf plots on some specified values of parameters r, c, k and ω .

(ii) The skewness and kurtosis of X are respectively,

$$\begin{aligned}
 \text{Skewness}(X) &= \left\{ r(1-\omega) \left[(r+1)(r+2) \left(\frac{c}{c-3} \right)^k - 3(r+1)^2 \left(\frac{c}{c-2} \right)^k \right. \right. \\
 &\quad \left. \left. + (3r^2 + 3r + 1) \left(\frac{c}{c-1} \right)^k - r^2 \right] - 3r^2(1-\omega)^2 \left[\left(\frac{c}{c-1} \right)^k - 1 \right] \right. \\
 &\quad \left. \times \left[(r+1) \left(\frac{c}{c-2} \right)^k - (2r+1) \left(\frac{c}{c-1} \right)^k + r \right] \right. \\
 &\quad \left. + 2r^3(1-\omega)^3 \left[\left(\frac{c}{c-1} \right)^k - 1 \right]^3 \right\} / [\text{Var}(X)]^{3/2}, \tag{6} \\
 \text{Kurtosis}(X) &= \left\{ r(1-\omega) \left[(r+1)(r+2)(r+3) \left(\frac{c}{c-4} \right)^k \right. \right.
 \end{aligned}$$

$$\begin{aligned}
 & -2(r+1)(2r^2 + 7r + 6)\left(\frac{c}{c-3}\right)^k + (r+1)(6r^2 + 12r + 7)\left(\frac{c}{c-2}\right)^k \\
 & - (2r+1)(2r^2 + 2r + 1)\left(\frac{c}{c-1}\right)^k + r^3 \left] - 4r^2(1-\omega)^2 \left[\left(\frac{c}{c-1}\right)^k - 1 \right] \right. \\
 & \times \left[(r+1)(r+2)\left(\frac{c}{c-3}\right) - 3(r+1)^2\left(\frac{c}{c-2}\right) \right. \\
 & \left. \left. + (3r^2 + 3r + 1)\left(\frac{c}{c-1}\right) - r^2 \right] + 6 \left[r^4(1-\omega)^4 \left[\left(\frac{c}{c-1}\right)^k - 1 \right]^2 \right. \right. \\
 & \left. \left. \times \left[(r+1)\left(\frac{c}{c-2}\right) - (2r+1)\left(\frac{c}{c-1}\right) + r \right]^2 \right] \right. \\
 & \left. - 3 \left[r(1-\omega) \left[\left(\frac{c}{c-1}\right)^k - 1 \right] \right]^4 \right\} / [\text{Var}(X)]^2. \tag{7}
 \end{aligned}$$

Proof. Let X have a NB distribution with parameters $r > 0$ and $p = \exp(-\lambda)$, where λ is distributed as EL distribution. As in the ZINB moments, can be treating λ as an EL random variable. Then taking $X | \lambda$ as ZINB we have

$$\begin{aligned}
 E[X | \lambda] &= E \left[r(1-\omega) \frac{1-p}{p} \right], \\
 &= E \left[r(1-\omega)(e^\lambda - 1) \right],
 \end{aligned}$$

so to find $E[X]$ we just need to find

$$\begin{aligned}
 E[(e^\lambda - 1)] &= E[e^\lambda] - 1, \\
 &= M_\lambda(1) - 1, \\
 &= \frac{1}{\left(1 - \frac{1}{c}\right)^k} - 1, \\
 &= \left(\frac{c}{c-1}\right)^k - 1,
 \end{aligned}$$

thus,

$$E(X) = r(1-\omega) \left[\left(\frac{c}{c-1}\right)^k - 1 \right],$$

where $M_\lambda(X)$ is the moment generating function of the Erlang distribution. Higher order moments just require $E[e^{2\lambda}] = M_\lambda(2)$, etc. Therefore, it is possible to show that variance, skewness, and kurtosis of X can be written in the form of Eq. (5), Eq. (6), and Eq. (7), respectively.

4.2 Maximum likelihood estimation

We present the MLE method for estimating parameters of the ZINB-EL distribution. If $X \sim \text{ZINB-EL}(r, c, k, \omega)$, then the likelihood function of parameters r, c, k , and ω , respectively, is

$$L(r, c, k, \omega) = \prod_{i=1}^n \left[I_{(x_i=0)} \left(\omega + (1-\omega) \left(\frac{c}{c+r} \right)^k \right) \right] \\ \times \prod_{i=1}^n \left[I_{(x_i=1,2,3,\dots)} \left((1-\omega) \binom{r+x_i-1}{x_i} \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left(\frac{c}{c+r+j} \right)^k \right) \right],$$

and the associated log-likelihood function is expressed as

$$\ell(r, c, k, \omega) = \log L(r, c, k, \omega) \\ = \sum_{i=1}^n \left[I_{(x_i=0)} \log \left(\omega + (1-\omega) \left(\frac{c}{c+r} \right)^k \right) \right. \\ \left. + I_{(x_i=1,2,3,\dots)} \left[\log(1-\omega) + \log \Gamma(r+x_i-1) - \log \Gamma(r) \right. \right. \\ \left. \left. - \log \Gamma(x_i-1) + \log \left(\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left(\frac{c}{c+r+j} \right)^k \right) \right] \right]$$

By differentiating the log-likelihood function of the ZINB-EL, the partial derivatives of the log-likelihood function with respect to $r, c, k,$ and $\omega,$ respectively, are

$$\frac{\partial}{\partial r} \ell(r, c, k, \omega) \\ = \sum_{i=1}^n \left[I_{(x_i=0)} \left(\frac{(\omega-1)k \left(\frac{c}{c+r} \right)^k}{\omega + (1-\omega) \left(\frac{c}{c+r} \right)^k} \right) \right. \\ \left. + I_{(x_i=1,2,3,\dots)} \left(\psi(r+x_i) - n\psi(r) + \frac{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \frac{\partial}{\partial r} \left(\left(\frac{c}{c+r+j} \right)^k \right)}{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left(\frac{c}{c+r+j} \right)^k} \right) \right], \tag{8}$$

$$\frac{\partial}{\partial c} \ell(r, c, k, \omega) \\ = \sum_{i=1}^n \left[I_{(x_i=0)} \left(\frac{(1-\omega)kr \left(\frac{c}{c+r} \right)^{k-1}}{\omega + (1-\omega) \left(\frac{c}{c+r} \right)^k} \right) \right. \\ \left. + I_{(x_i=1,2,3,\dots)} \left(\psi(r+x_i) - n\psi(r) + \frac{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \frac{\partial}{\partial c} \left(\left(\frac{c}{c+r+j} \right)^k \right)}{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left(\frac{c}{c+r+j} \right)^k} \right) \right], \tag{9}$$

$$\begin{aligned} & \frac{\partial}{\partial k} \ell(r, c, k, \omega) \\ &= \sum_{i=1}^n \mathbb{I}_{(x_i=0)} \left[\frac{\left((1-\omega) \left(\frac{c}{c+r} \right)^k \log \left(\frac{c}{c+r} \right) \right)}{\omega + (1-\omega) \left(\frac{c}{c+r} \right)^k} \right] \\ &+ \mathbb{I}_{(x_i=1,2,3,\dots)} \left[\frac{\left(\psi(r+x_i) - n\psi(r) + \frac{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \frac{\partial}{\partial k} \left(\left(\frac{c}{c+r+j} \right)^k \right)}{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left(\frac{c}{c+r+j} \right)^k} \right)}{\sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left(\frac{c}{c+r+j} \right)^k} \right], \end{aligned} \tag{10}$$

$$\frac{\partial}{\partial \omega} \ell(r, c, k, \omega) = \sum_{i=1}^n \mathbb{I}_{(x_i=0)} \left[\left(1 - \left(\frac{c}{c+r} \right)^k \right) - \mathbb{I}_{(x_i=1,2,3,\dots)} \left(\frac{1}{1-\omega} \right) \right], \tag{11}$$

where $\psi(g) = \frac{\Gamma'(g)}{\Gamma(g)}$ is the digamma function

In this case the derivative equations cannot be solved analytically. The MLE solution of \hat{r} , \hat{c} , \hat{k} and $\hat{\omega}$ can be obtained by solving the resulting equations simultaneously using `optim` function in R language (R Core Team, 2015).

4.3 Application

We illustrated ZIP, ZINB and ZINB-EL distributions by application to a real data set. The data set is the number of reported outbreaks per subdistrict in Thailand from July 3, 2004 to May 5, 2005 (Vergne, 2014). This data set contains 89.42% zeros. The mean and variance are 0.2274 and 0.7783, respectively. The index of dispersion is 3.423 which indicates a high percentage of zeros and the variance is greater than the mean.

We used the `optim` function to estimate the parameters to describe the assumed distributions (i.e. ZIP, ZINB, ZINB-EL). Using those parameters we can conduct the K-S test for the discrete goodness of fit test from the `dgoF` package (Arnold & Emerson, 2011) in the R language to estimate whether this real data set is from the same distribution as the assumed distribution and compare the efficiencies of fitting the distributions.

The log-likelihood values, AIC, BIC, and the P-values of K-S test are summarized in Table 1. For illustration, Figure 2 reveals that the proposed distribution is more appropriate to fit the data than the ZIP and ZINB distributions. The expected frequencies of the ZINB-EL distribution are close to the observed frequencies, the values of K-S test of ZINB-EL distribution is smaller than the values of the K-S test of ZIP and ZINB distributions.

5. Conclusions

In this paper, we considered the distribution for count data with excess zeros, which is called the ZINB-EL distribution. The ZINB-EL distribution is a mixture of a point mass at zero and NB-EL distribution. We have derived some properties of the ZINB-EL distribution, including mean, variance, skewness, and kurtosis. Moreover, we derived the parameter estimator of ZINB-EL distribution by using the MLE method. We have compared efficiencies of fitting distributions based on the goodness of fit test and some information criteria. The usefulness of the ZINB-EL distribution is illustrated by the number of H5N1 outbreaks reported in each subdistrict during the second epidemic wave from July 3, 2004 to May 5, 2005 in Thailand. Finally, the results of this study show that the ZINB-EL distribution provides a better fit compared to the ZIP and ZINB distributions. Obviously, the ZINB-EL distribution is an alternative distribution to the other.

Acknowledgements

We are thankful to the Department of Statistics, Faculty of Science, Kasetsart University. The first author would like to thank the Research Professional Development Project under the Science Achievement Scholarship of Thailand (SAST) for the funding. The authors also wish to thank the reviewer for their detailed comments and suggestions. These suggestions helped us very much to improve the manuscript, specifically in a simple way to calculate the ZINB moments and the computation in terms of the fitted frequencies to obtain realistic answers.

Table 1. Observed and expected frequencies for the number of reported outbreaks per subdistrict in Thailand from July 3rd 2004 to May 5th 2005.

Number of outbreaks	Number of subdistricts	Expected value by fitting distribution		
		ZIP	ZINB	ZINB-EL
0	6587	6586.87	6587.13	6587.03
1	410	279.70	500.41	414.21
2	161	250.19	178.53	165.91
3	87	149.20	64.01	80.72
4	46	66.73	23.00	44.47
5	26	23.88	8.28	26.72
6	21	7.12	2.98	17.13
7	8	1.82	1.08	11.54
8	4	0.41	0.39	8.08
9	6	0.08	0.14	5.85
10+	10	0.01	0.05	4.35
Estimated parameters		$\hat{\lambda} = 1.789$ $\hat{\omega} = 0.873$	$\hat{r} = 0.971$ $\hat{p} = 0.551$ $\hat{\omega} = 0.701$	$\hat{r} = 1.052$ $\hat{c} = 2.695$ $\hat{k} = 1.311$ $\hat{\omega} = 0.692$
log-likelihood		-3778.073	-3633.407	-3007.488
AIC		7560.146	7272.814	6022.976
BIC		7563.880	7278.416	6030.445
K-S test		0.018	0.015	0.003
p-value		0.020	0.084	0.998

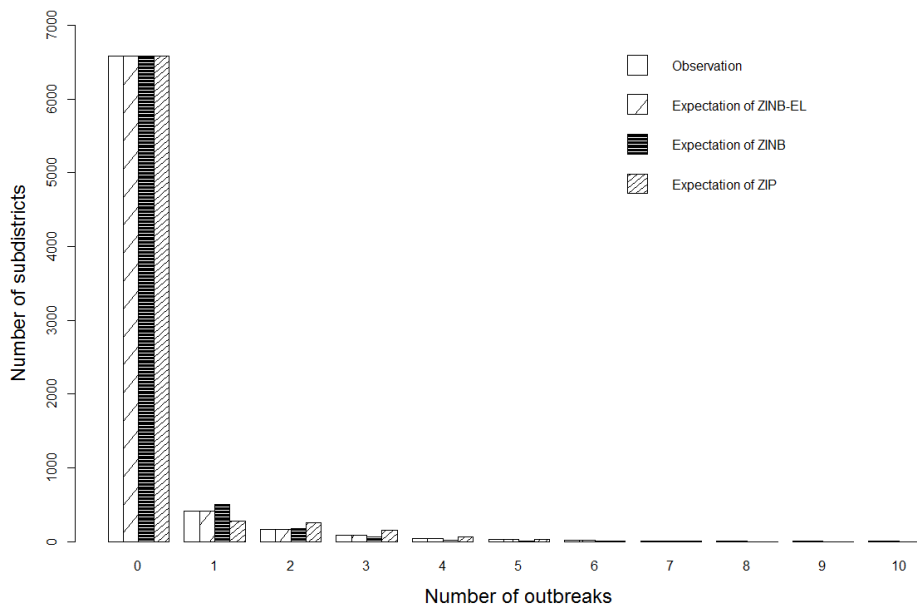


Figure 2. Comparison of observed and expected frequencies of the number of reported outbreaks per subdistrict in Thailand from July 3rd 2004 to May 5th 2005.

References

- Akaike, H. (1978). A new look at the statistical model identification. *IEEE-TAC*, 19, 716-723.
- Arnold, B. T., & Emerson, W. J. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, 3, 34-39.
- Gilbert, M., Chaitaweesub, P., Parakamawongsa, T., Premashthira, S., Tiensin, T., Kalpravidh, W., & Slingenbergh, J. (2006). Free-grazing ducks and highly pathogenic avian influenza Thailand. *Emerging Infectious Diseases*, 12, 227-234.
- Gilbert, M., Xiao, X., Chaitaweesub, P., Kalpravidh, W., Premashthira, S., & Boles, S. (2007). Avian influenza domestic ducks and rice agriculture in Thailand. *Agriculture Ecosystems and Environment*, 119, 409-415.
- Gomez, E., Sarabia, M., & Calderin, E. (2008). Univariate and multivariate versions of the negative binomial-Inverse Gaussian distribution with applications. *Insurance: Mathematics and Economics*, 42, 39-49.
- Greene, W. (1994). *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models* (Working Papers No. EC-94-10), New York University, New York, United States of America.
- Greenwood, M., & Yule, G. U. (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*, 83, 255-279.
- Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random Effects: A Case Study. *Biometrics*, 56, 1030-1039.
- Heilbron, D. C. (1994). Zero-Altered and other Regression Models for Count Data with Added Zeros. *Biometrical Journal*, 36, 531-547.
- Hilbe, J. M. (2014). *Modeling Count Data*. New York, NY: Cambridge University Press.
- Johnson, N. L., Kotz, S., & Kemp A. W. (1992). *Univariate discrete distributions*. New York, NY: Wiley.
- Kongrod, S., Bodhisuwan, W., & Payakkapong, P. (2014). The negative binomial-Erlang distribution with applications. *International journal of Pure and Applied Mathematics*, 4, 609-610.
- Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
- Martin, G. T., Wintle, A. B., Rhodes, J. R., Field, A. S., Low-Choy, J. S., Tyre, A. J., & Possingham, P. H. (2005). Zero tolerance ecology: improving ecological-inference by modelling the source of zero observations. *Ecology Letters*, 8, 1235-1246.
- Meng, S., Wei, Y., & Whitmore, G. A. (1999). Accounting for individual overdispersion in a Bonus-Malus system. *Astin Bulletin*, 29, 327-337.
- Pudprommarat, C., & Bodhisuwan, W. (2012). A new mixed negative binomial distribution. *Journal of Applied Sciences*, 12, 1853-1858.
- R Core Team. (2015, May 17). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. Vienna. Retrieved from <http://www.R-project.org>
- Rainer, W. (2008). *Econometric Analysis of Count Data*. New York, NY: Springer.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.
- Shewhart, W. A., & Wilks, S. S. (2005). *Finite Mixture Models*. New York, NY: John Wiley and Sons.
- Tiensin, T., Ahmed, S. S., Rojanasthien, S., Songserm, T., Ratanakorn, P., Chaichoun, K., & Nielen, M. (2009). Ecologic risk factor investigation of clusters of avian influenza A (H5N1) virus infection in Thailand. *The Journal of Infectious Diseases*, 199, 1735-1743.
- Vergne, T. M., Paul, W. B., Chaengprachak, M., Durandb, B., Gilbert, F., Dufourh, S., & Grosboisa, V. (2014). Zero-inflated models for identifying disease risk factors when case detection is imperfect: Application to highly pathogenic avian influenza H5N1 in Thailand. *Preventive Veterinary Medicine*, 114, 28-36.
- World Health Organization. (2016, May 17). Avian influenza fact sheet March 2014. Retrieved from http://www.who.int/mediacentre/factsheets/avian_influenza
- Xu, X., Subbarao, Cox, N. J., & Guo, Y. (1999). Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology*, 261, 15-19.
- Yang, Z. (2006). *Score Tests for Generalization and Zero-inflation in Count Data Modeling*. University of South Carolina, SC.
- Zamani, H., & Ismail, N. (2010). Negative binomial-Lindley distribution and Its application. *Journal of Mathematics and Statistics*, 6, 4-9.