The zero-distorted Topp-Leone geometric distribution: some properties and its applications with biological data

Areeya Sudsuk, Winai Bodhisuwan* and Boonorm Chomtee

Department of Statistics, Faculty of Science, Kasetsart University, P.O.Box 1086, Chatuchak, Bangkok, 10900, Thailand
*Corresponding author. E-mail: fsciwnb@ku.ac.th

ABSTRACT

In this paper, the zero-distorted Topp-Leone geometric distribution is introduced. It belongs to the k-distorted generalized discrete family of distributions. This family is useful to fit both zero-inflated and zero-deflated data. In addition, the proposed distribution has many special cases including the Topp-Leone geometric, the discrete zero-truncated Topp-Leone geometric, the zero-deflated Topp-Leone geometric and the zero-inflated Topp-Leone geometric distributions. We also derive the first four moments and index of dispersion for the zero-distorted Topp-Leone geometric distribution. For parameter estimation, the most well-known method called the maximum likelihood estimation is utilized. In application study, we apply the proposed model to fit with three biological datasets. Furthermore, the fitted results of zero-distorted Topp-Leone geometric distribution are compared with the Topp-Leone geometric, the zero-distorted generalized geometric and the negative binomial distributions. In conclusion, the Anderson-Darling test statistic for discrete distributions shows that the zero-distorted Topp-Leone geometric distribution is the most appropriate model for these datasets.

Keywords: zero-distorted distributions, geometric distribution, zero-inflated, maximum likelihood estimation, biological data, T-X family

INTRODUCTION

Count data frequently occur in many research problems. For example, the number of plant in biological study (Bliss and Fisher, 1953), the number of automobile claims in actuarial application (Gómez-Déniz *et al*., 2011), and the number of hospitalizations per family member in clinical trials (Klugman *et al*., 2012) are recorded in the form of count data. Basically, these phenomena can be described by the Poisson distribution. However, the Poisson distribution restricts values of mean and variance to be equal. Frequently, observed data do not meet that restriction as count data exhibits either overdispersion, i.e., the variance is greater than the mean or zero-inflated, i.e., the presence of a high percentage of zero values (Gómez-Déniz *et al*., 2011). In addition, the observed overdispersion may be the result of excessive zeros in the distribution (Perumean-Chaney *et al*., 2013). Some examples of overdispered and zero-inflated data are number of claims of automobile liability policies and hospitalizations per family member per year that the observed variance is greater than the observed mean, where there also exits high percentage of zero claim and zero hospitalization respectively. In literature, traditional discrete distributions have been developed to deal with these issues such as the zero-inflated Poisson distribution (Winkelmann, 2008), the generalized Poisson distribution (Chandra *et al.*, 2013), the Poisson-weighted exponential distribution (Zamani *et al*., 2014), the zero-modified Poisson-Lindley distribution (Xavier *et al*., 2018), and etc.

In contrast to zero-inflated data mentioned above, sometimes count data has a low percentage of zero values. The situation when count data have zero values less than the probability of the underlying model, called zero-deflated. Correspondingly, one needs to meditate a count data model by inflating or deflating. Recently, in 2016, Sastry *et al*. (Sastry *et al*., 2016) considered the model that can analyze both zero-inflated and zero-deflated data. They proposed the framework based on the mixed-Poisson method with cumulative density function $\int_0^\infty \frac{e^{(-\lambda)}\lambda^x}{x!}F(\lambda)d\lambda$ where λ is a continuous distribution with probability density function $F(.)$ to improve performance of the geometric distribution and created zero-distorted generalized geometric (ZDGG) distribution by introducing a new parameter called distortion parameter (β). This parameter represents inflation and deflation of the distribution (Shirke *et al*., 2017). Consequently, the ZDGG distribution has probability mass function (pmf) and associated cumulative density function (cdf) as respectively

$$f_\beta(x) = \begin{cases} 1-q^{\beta+1} & \text{if } x=0, \\ (1-q)q^{(x+\beta)} & \text{if } x=1,2,3,... \end{cases} \tag{1}$$

and

$$F_\beta(x) = \begin{cases} 0 & \text{if } x<0, \\ 1-q^{(x+\beta+1)} & \text{if } x=1,2,3,..., \end{cases} \tag{2}$$

where $0<q<1$ and $-1\le\beta<\infty$. Recently, Shirke *et al*. (Shirke *et al*., 2017) introduced the k-distorted generalized discrete family of distributions by generalizing ZDGG distribution (Sastry *et al*., 2016). This family includes the ZDGG distribution as a special case.

The geometric distribution is a special case of the negative binomial distribution. It is commonly known to be a discrete analog of the exponential distribution (Johnson *et al*., 2005). Many properties of the geometric distribution are relevant to the properties of the exponential distribution (Akinsete *et al*., 2014). In addition to the work of Sastry *et al*. (Sastry *et al*., 2016), there are various attempts to modify the geometric distribution. Some examples are the Marshall and Olkin geometric distribution (Gómez-Déniz, 2010), discrete generalized exponential distribution (Nekoukhou *et al*., 2012), the Kumaraswamy-geometric distribution (Akinsete *et al*., 2014) and the weighted geometric distribution (Bhati and Joshi, 2018). Recently, the Topp-Leone geometric (TLG) distribution was proposed by Sudsuk and Bodhisuwan (Sudsuk and Bodhisuwan, 2016). It is a member of the T-X family of distributions (Aljarrah *et al*., 2014) and has two parameters. This distribution has flexibility over the geometric distribution in terms of density and hazard shape. Sudsuk and Bodhisuwan (Sudsuk and Bodhisuwan, 2016) showed that the TLG distribution could improve on the goodness-of-fit test result of the geometric distribution and provided better fit than the Poisson and geometric distributions in application study.

In this work, we will apply the zero-distorted generalized discrete family of distribution (Sastry *et al*., 2016) with an extension of the geometric distribution called the TLG distribution. The rest of this paper are organized as follows. Firstly, we describe the k-distorted generalized discrete family of distributions together with its special case called the zero-distorted generalized discrete family of distributions. Secondly, the zero-distorted Topp-Leone geometric (ZDTLG) distribution is created based on analogue of the zero-distorted generalized discrete family of distributions. Therefore, its some properties are derived such as the first four moments and index of dispersion. Next section, the maximum likelihood estimation are considered for the proposed model. Finally, we compare the proposed distribution with the TLG, the ZDGG, and the negative binomial (NB) distributions by fitting with the biological datasets.

PRELIMINARIES

The essential components to create the ZDTLG distribution are discussed. There are the k-distorted generalized discrete family of distributions, the zero-distorted generalized discrete family of distributions, and the TLG distribution. Indeed, the k-distorted generalized discrete family of distributions includes the zero-distorted generalized discrete family of distributions as a special case (Shirke *et al*., 2017). Moreover, the TLG distribution is an extension of the geometric distribution that will be capable of model both zero-inflated and zero-deflated data (Sastry *et al*., 2016) when it cooperated with the zero-distorted generalized discrete family of distributions (Shirke *et al*., 2017).

The k-distorted generalized discrete family of distributions

Based on the definition of Shirke *et al*. (Shirke *et al*., 2017), let $P_\theta(x)$ be the pmf of a discrete random variable X with parameter θ. Then, the k-distorted generalized discrete family of distributions has pmf and cdf as respectively

$$f_\beta(x;\theta)=\begin{cases}1-(1-f_k)^{\beta+1} & \text{if } x=k,\\ (1-f_k)^\beta f_\theta(x) & \text{if } x\neq k,\end{cases} \tag{3}$$

and

$$F_\beta(x;\theta)=\begin{cases}0 & \text{if } x<k,\\ 1-(1-f_k)^{\beta+1} & \text{if } x=k,\\ 1-(1-f_k)^{\beta+1}+(1-f_k)^\beta\sum_{r=1}^{x} f_\theta(x) & \text{if } x=k+1,k+2,...,\end{cases} \tag{4}$$

where $f_k=f_\theta(X=k)$, $-1\le\beta<\infty$.

Consequently, if $k=0$, we have the zero-distorted distribution with the pmf.

$$f_\beta(x;\theta)=\begin{cases}1-(1-f_0)^{\beta+1} & \text{if } x=0,\\ (1-f_0)^\beta f_\theta(x) & \text{if } x\neq 0,\end{cases} \tag{5}$$

where $f_0=f_\theta(X=0)$, $-1\le\beta<\infty$. Thus, the cdf is

$$F_\beta(x;\theta)=\begin{cases}0 & \text{if } x<0,\\ 1-(1-f_0)^{\beta+1} & \text{if } x=0,\\ 1-(1-f_0)^{\beta+1}+(1-f_0)^\beta\sum_{r=1}^{x} f_\theta(x) & \text{if } x=1,2,...\end{cases} \tag{6}$$

According to Shirke *et al*. (Shirke *et al*., 2017), the zero-distorted generalized geometric distribution (Sastry *et al*., 2016) can be obtained by changing only $f_\theta(\mathrm{x})$ in Equations (5) and (6) to be pmf of the geometric distribution.

The Topp-Leone geometric distribution

Sudsuk and Bodhisuwan (Sudsuk and Bodhisuwan, 2016) allowed the cdf of the geometric distribution being a random variable from the Topp-Leone distribution (Topp and Leone, 1955). Then, the TLG distribution with pmf and cdf are

$$f(x)=(1-q^{2(x+1)})^\alpha-(1-q^{2x})^\alpha \tag{7}$$

and

$$F(x)=(1-q^{x+1})^\alpha(1+q^{x+1})^\alpha$$

where $\alpha > 0$, $0 < q < 1$ and $x = 0,1,2,...$

Some mathematical properties of the zero-distorted generalized discrete family of distributions can be derived based on the same measurement of the discrete distribution with pmf $f_\theta(x)$. So, we provide the r^{th} moment of the TLG distribution because it will be useful to obtain the r^{th} moment of the proposed distribution. Let X be distributed as the TLG random variable. The r^{th} moment can be expressed as

$$E(X^r) = \sum_{i=0}^{\infty} (-1)^i \binom{\alpha}{i} (q^{2i} - 1) L_r(q^{2i}),$$

where $L_r(x) = \sum_{j=1}^{\infty} \frac{x^j}{j_r}$ is the polylogarithm function. Thus, the first four moments are respectively

$$\mu_1' = \sum_{i=1}^{\infty} (-1)^i \binom{\alpha}{i} \left[\frac{-q^{2i}}{(1-q^{2i})} \right],$$

$$\mu_2' = \sum_{i=1}^{\infty} (-1)^i \binom{\alpha}{i} \left[\frac{-q^{2i}(1+q^{2i})}{(1-q^{2i})^2} \right],$$

$$\mu_3' = \sum_{i=1}^{\infty} (-1)^i \binom{\alpha}{i} \left[\frac{-q^{2i}(1+4q^{2i}+q^{4i})}{(1-q^{2i})^3} \right],$$

$$\mu_4' = \sum_{i=1}^{\infty} (-1)^i \binom{\alpha}{i} \left[\frac{-q^{2i}(1+q^{2i})(1+10q^{2i}+q^{4i})}{(1-q^{2i})^4} \right].$$

THE ZERO-DISTORTED TOPP-LEONE GEOMETRIC DISTRIBUTION

The ZDTLG distribution can be obtained by replacing Equation (7) to Equation (5) and (6). Then, the pmf and cdf of the ZDTLG distribution can be written respectively as

$$f_\beta(x;\theta) = \begin{cases} 1-(1-(1-q^2)^\alpha)^{\beta+1} & \text{if } x = 0, \\ (1-(1-q^2)^\alpha)^\beta \left((1-q^{2(x+1)})^\alpha - (1-q^{2x})^\alpha\right) & \text{if } x \neq 0, \end{cases} \tag{8}$$

and

$$F_\beta(x;\theta)=\begin{cases}0 & \text{if } x<0,\\ 1-(1-(1-q^2)^\alpha)^{\beta+1} & \text{if } x=0,\\ 1-(1-(1-q^2)^\alpha)^{\beta+1}+ & \\ (1-(1-q^2)^\alpha)^\beta\sum_{r=1}^{x}\left(1-q^{2(x+1)})^\alpha-(1-q^{2x})^\alpha\right) & \text{if } x=1,2,...,\end{cases} \tag{9}$$

where $\alpha>0$, $0<q<1$, $-1\le\beta<\infty$ and $x=0,1,2,...$

Some pmf plots of the ZDTLG distribution are shown in Figure 1 with various values of its parameters. These are categorized into three shapes which are decreasing (1), unimodal (2-5), mixed-shape (6-8). It is obvious that the proposed distribution includes both greater and lesser zero tendency in the single model.

According to Shirke *et al*. (Shirke *et al*., 2017), the ZDTLG distribution has several special cases including the TLG, the discrete zero-truncated Topp-Leone geometric, the zero-deflated Topp-Leone geometric and the zero-inflated Topp-Leone geometric distributions as follows;

1) when $\beta=0$, the model reduces to the TLG distribution,
2) when $\beta=-1$, it corresponds to the discrete zero truncated Topp-Leone geometric distribution,
3) when $-1\le\beta<b$, the distribution is the zero-deflated Topp-Leone geometric distribution, and when $\beta>b$, it is the zero-inflated Topp-Leone geometric distribution where b is a constant that shows in Equation (14) and (15).

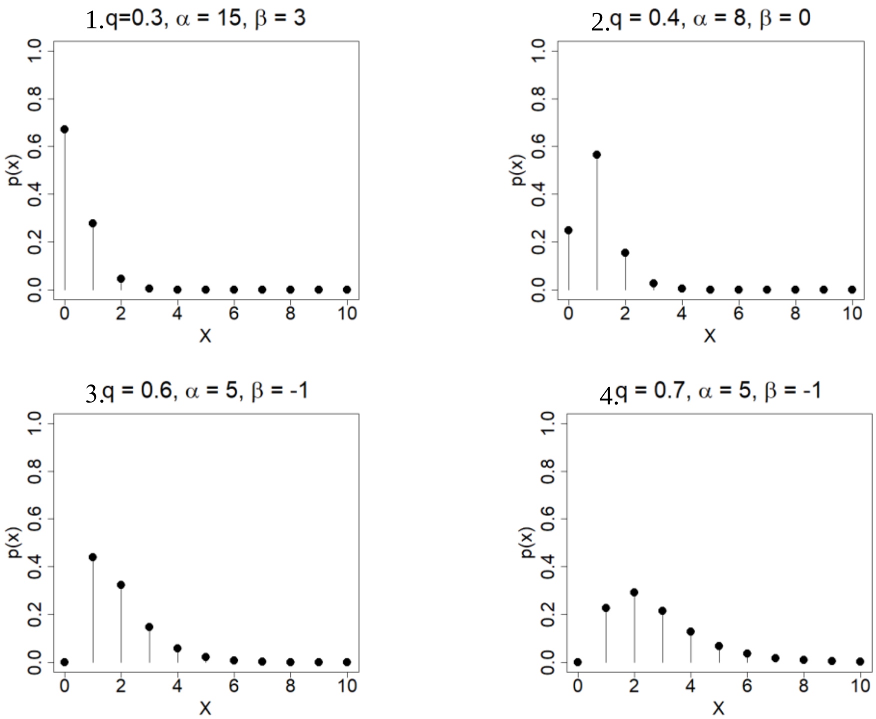
Furthermore, we use stochastic order to show effect of adjustment based on adding distortion parameter. The relation between the ZDTLG distribution and the TLG distribution are illustrated as follows. Let $F_X(t)$ and $F_Y(t)$ be random variable of the TLG distribution and the ZDTLG distribution respectively

$$\begin{aligned}F_Y(t)&=1-(1-(1-q^2)^\alpha)^{\beta+1}+(1-(1-q^2)^\alpha)^\beta\sum_{r=1}^{[t]}\left(1-q^{2(x+1)})^\alpha-(1-q^{2r})^\alpha\right)\\&=1-(1-(1-q^2)^\alpha)^{\beta+1}+(1-(1-q^2)^\alpha)^\beta\left(F_X(t)-(1-q^2)^\alpha\right)\\&=1-(1-(1-q^2)^\alpha)^\beta\left(1-F_X(t)\right).\end{aligned}$$

Consequently, it leads to

$$\frac{F_Y(t)}{F_X(t)}=\frac{1-(1-(1-q^2)^\alpha)^\beta\left(1-F_X(t)\right)}{F_X(t)}.$$

Therefore, if $\beta > 0$, $F_Y(t)$ of the proposed distribution is larger than $F_X(t)$ of the TLG distribution and Y is stochastically smaller than X. On the other hand, when $\beta < 0$, $F_Y(t)$ of the proposed distribution is smaller than $F_X(t)$ of TLG distribution and Y is stochastically larger than X.



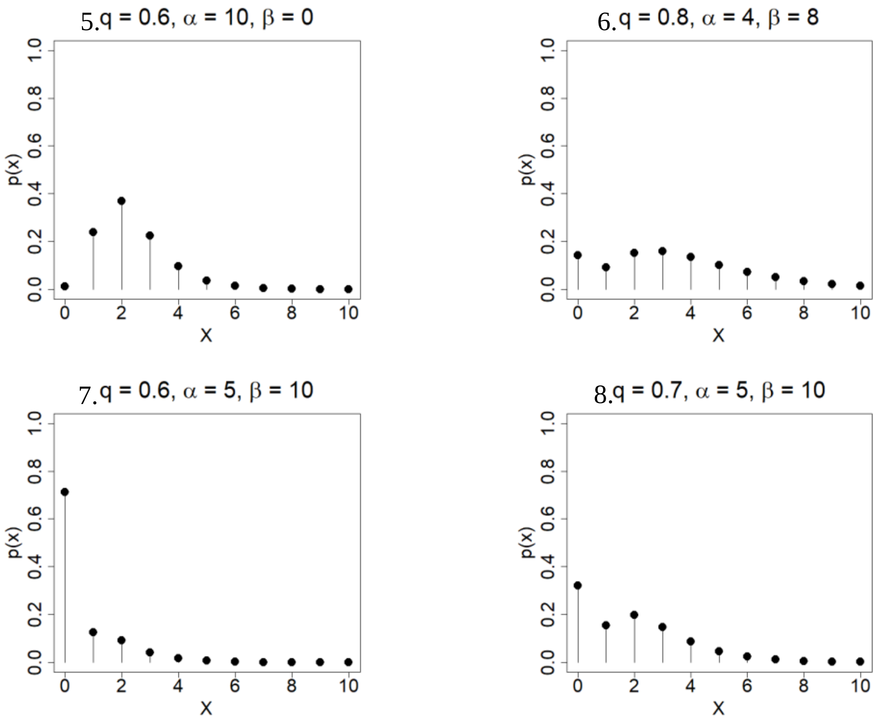


Figure 1 Plots of the ZDTLG pmf for some values of q, α and β

Moments

Shirke *et al*. (Shirke *et al*., 2017) have shown a relation between the moment generating function of the zero-distorted generalized discrete family of distributions and the associated discrete distribution. Then, they utilized it to obtain the r^{th} moment of the zero-distorted generalized discrete family of distributions. In this paper, we firstly simplify the r^{th} moment of the zero-distorted generalized discrete family of distributions in order to obtain the r^{th} moment of the proposed distribution. The r^{th} moment of the zero-distorted generalized discrete family of distributions is

$$\begin{aligned} E^*(X^r) &= \sum_{x=0}^{\infty} x^r f_\beta(x), \\ &= \sum_{x=1}^{\infty} x^r f_\beta(x), \\ &= (1-f_0)^\beta \sum_{x=1}^{\infty} x^r f_\theta(x) \end{aligned} \tag{10}$$

For a random variables X with pmf $f_\theta(x)$, its r^{th} moment $(E(X^r))$ can be rewritten as

$$\begin{aligned} E(X^r) &= \sum_{x=0}^{\infty} x^r f_\theta(x), \\ &= \sum_{x=1}^{\infty} x^r f_\theta(x), \end{aligned} \tag{11}$$

By substituting Equation (11) into Equation (10), the r^{th} moment of the zero distorted generalized discrete family of distributions can be defined as

$$E^*(X^r) = (1-f_0)^\beta E(X^r) \tag{12}$$

When random variable X has the ZDTLG distribution, its r^{th} moment can be obtained from the r^{th} moment of the TLG distribution and Equation (12).

Therefore, the r^{th} moment of the ZDTLG distribution is

$$E^*(X^r) = (1-f_0)^\beta \left(\sum_{i=0}^{\infty} (-1)^i \binom{\alpha}{i} (q^{2i}-1) L_r(q^{2i}) \right)$$

and the first four moments are, respectively

$$\mu_1' = (1-f_0)^\beta \sum_{i=1}^{\infty} (-1)^i \binom{\alpha}{i} \left[\frac{-q^{2i}}{(1-q^{2i})} \right],$$

$$\mu_2' = (1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left[\frac{-q^{2i}(1+q^{2i})}{(1-q^{2i})^2}\right],$$
$$\mu_3' = (1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left[\frac{-q^{2i}(1+4q^{2i}+q^{4i})}{(1-q^{2i})^3}\right],$$
$$\mu_4' = (1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left[\frac{-q^{2i}(1+q^{2i})(1+10q^{2i}+q^{4i})}{(1-q^{2i})^4}\right], \tag{13}$$

Index of dispersion

The index of dispersion (ID) is the ratio of the variance to the mean. It is used to determine whether the distribution is overdispersed ($ID>1$) or underdispersed ($ID<1$). Based on μ_1' and μ_2' in Equation (13), we can derive mean and variance of the ZDTLG distribution as

$$E(X) = (1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left(\left[\frac{-q^{2i}}{(1-q^{2i})}\right]\right),$$

and

$$Var(X) = (1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left[\frac{-q^{2i}(1+q^{2i})}{(1-q^{2i})^2}\right] - \left((1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left[\frac{-q^{2i}}{(1-q^{2i})}\right]\right)^2.$$

The ZDTLG distribution is overdispersed when $Var(X) > E(X)$ which

$$(1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left[\frac{-q^{2i}(1+q^{2i})}{(1-q^{2i})^2}\right] - \left((1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left[\frac{-q^{2i}}{(1-q^{2i})}\right]\right)^2 > (1-f_0)^{\beta}\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left[\frac{-q^{2i}}{(1-q^{2i})}\right]$$

$$\beta > \frac{\log\left(\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left(\left[\frac{-2q^{4i}}{(1-q^{2i})^2}\right]\right) \Big/ \left(\sum_{i=1}^{\infty}(-1)^i\binom{\alpha}{i}\left(\left[\frac{-q^{2i}}{(1-q^{2i})}\right]\right)\right)^2\right)}{\log(1-f_0)} \tag{14}$$

On the other hand, the proposed distribution is underdispersed which

$$\beta < \frac{\log\left(\sum_{i=1}^{\infty} (-1)^i \binom{\alpha}{i} \left(\left[\frac{-2q^{4i}}{(1-q^{2i})^2} \right] \right) \Big/ \left(\sum_{i=1}^{\infty} (-1)^i \binom{\alpha}{i} \left(\left[\frac{-q^{2i}}{(1-q^{2i})} \right] \right) \right)^2 \right)}{\log(1-f_0)} \tag{15}$$

PARAMETER ESTIMATION

Parameter estimation plays a central role in statistical inference. It is useful to both describe behavior of population and test hypothesis for making the decision. The maximum likelihood estimation (MLE) is one of the widely-used parameter estimation methods. In this paper, MLE is considered to estimate the unknown parameters of the ZDTLG distribution.

Let $X_1,...,X_n$ be observations of independent random variable from ZDTLG distribution. The maximum likelihood estimator of the vector parameter $\theta = (q,\alpha,\beta)$ can be obtained as follows. The log-likelihood function for the vector parameter θ is

$$\ell(\theta) = m\log\left(1-\left(1-\left(1-q^2\right)^{\alpha}\right)^{\beta+1}\right) + (n-m)\beta\log\left(1-\left(1-q^2\right)^{\alpha}\right) - \sum_{x_i>0}\log\left(\left(1-q^{2(x_i+1)}\right)-\left(1-q^{2x_i}\right)^{\alpha}\right)$$

where *m* is the number of zero in the sample. The corresponding vector scores are obtained as

$$\frac{\partial \ell(\theta)}{\partial q} = -\frac{2mq\alpha(\beta+1)(1-q^2)^{\alpha-1}\left(1-\left(1-q^2\right)^{\alpha}\right)^{\beta}}{1-\left(1-\left(1-q^2\right)^{\alpha}\right)^{\beta+1}} + \frac{2(n-m)q\alpha\beta(1-q^2)^{\alpha-1}}{1-\left(1-q^2\right)^{\alpha}} + 2\alpha\sum_{x_i>0}\frac{x_i q^{2x_i-1}\left(1-q^{2x_i}\right)^{\alpha-1} - (x_i+1)q^{2x_i+1}\left(1-q^{2(x_i+1)}\right)^{\alpha-1}}{\left(1-q^{2(x_i+1)}\right)^{\alpha} - \left(1-q^{2x_i}\right)^{\alpha}} \tag{16}$$

$$\frac{\partial \ell(\theta)}{\partial \alpha} = -\frac{m(\beta+1)(1-q^2)^{\alpha}\left(1-\left(1-q^2\right)^{\alpha}\right)^{\beta}\log\left(1-q^2\right)}{1-\left(1-\left(1-q^2\right)^{\alpha}\right)^{\beta+1}} + \frac{(n-m)\beta(1-q^2)^{\alpha}\log\left(1-q^2\right)}{1-\left(1-q^2\right)^{\alpha}} + \sum_{x_i>0}\frac{\left(1-q^{2(x_i+1)}\right)^{\alpha}\log\left(1-q^{2(x_i+1)}\right)-\left(1-q^{2x_i}\right)^{\alpha}\log\left(1-q^{2x_i}\right)}{\left(1-q^{2(x_i+1)}\right)^{\alpha}-\left(1-q^{2x_i}\right)^{\alpha}} \tag{17}$$

and

$$\frac{\partial \ell(\theta)}{\partial \beta} = -\frac{m\left(1-\left(1-q^2\right)^{\alpha}\right)^{\beta+1}\log\left(1-\left(1-q^2\right)^{\alpha}\right)}{1-\left(1-\left(1-q^2\right)^{\alpha}\right)^{\beta+1}} + (n-m)\log\left(1-\left(1-q^2\right)^{\alpha}\right) \tag{18}$$

By setting Equation (16)-(18) to zero, these differential equations cannot be solved algebraically. But, we can use a numerical method in `optim` function in R language (R Core Team, 2017) to maximize the log-likelihood function and then the estimator $\hat{\theta} = (\hat{q}, \hat{\alpha}, \hat{\beta})^T$ will be obtained.

APPLICATION

As mentioned earlier, the ZDTLG distribution has competence in modeling both zero-inflated and zero-deflated data. In this section, we use three biological datasets to show flexibility and capability of the proposed distribution. These datasets from Bliss and Fisher (Bliss and Fisher, 1953) are as follows; the first dataset is the number of Ameria maritina counts per quadrat in a salt marsh, the second dataset is the number of Chenopodium counts per quadrat in a weed on arable land, and the last dataset is the number of Salicornia stricta counts per quadrat in a salt marsh. Moreover, Table 1 displays mean, variance and ID for these three datasets. According to ID, we can conclude that these datasets behave over-dispersed.

Table 1 Mean, variance and ID of three datasets

	Datasets		
	America maritima	Chenopodium album	Salicornia stricta
Mean	1.580	4.032	6.653
Variance	5.418	15.754	31.054
ID	3.430	3.907	3.165

In this article, the proposed distribution called the ZDTLG distribution is compared with the three existing distributions: the TLG, ZDGG, and NB distributions. For parameter estimation, three distributions including the ZDTLG, TLG, ZDGG distributions are obtained by the `optim` function in R language (R Core Team, 2017). The parameters of the NB distribution are estimated by the `fitdist` function of `fitdistrplus` package (Delignette-Muller and Dutang, 2015). In addition, for goodness-of-fit test, the Anderson-Darling (AD) test statistic for discrete distributions is applied to these distributions by using the `dgof` package (Arnold and John, 2011). Tables 2-4 provide observed and expected frequencies of three real datasets and show the Akaike information criterion (AIC), the Bayesian information criterion (BIC), AD test, and *p*-values of AD test, corresponded to the G ZDGG, ZDTLG, and NB distributions.

In addition, plots of observed and expected values for each of the datasets among these distributions are presented in Figures 2-4. For model selection in this study, the proposed model have the smallest AIC and BIC in dataset 1 and dataset 2. In dataset 3, the AIC (556.021) and BIC (561.191) of the NB model are a bit less than the AIC (556.502) and BIC (564.257) of the ZDTLG model. In goodness of fit test, the proposed distribution has the highest *p*-values. In summary, the proposed distribution provides better fit comparing to the TLG and the ZDGG distributions and can be candidate for the NB distribution.

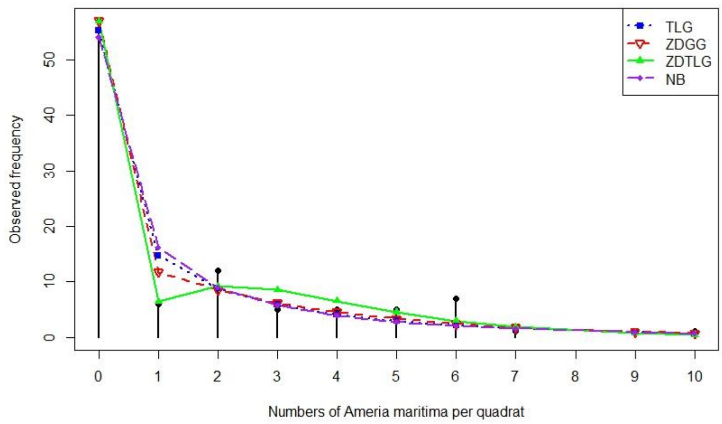


Figure 2 Plots of observed and expected values of Ameria maritima based on the TLG, ZDGG, ZDTLG and NB distributions

Table 2 Numbers of Ameria maritima counts per quadrat

Numbers of Plant	Observed	Expected			
		TLG	ZDGG	ZDTLG	NB
0	57	55.29	57.00	57.00	54.11
1	6	14.76	11.70	6.40	16.19
2	12	8.81	8.52	9.21	8.98
3	5	5.85	6.20	8.54	5.75
4	5	4.07	4.51	6.52	3.93
5	5	2.92	3.28	4.51	2.78
6	7	2.12	2.39	2.95	2.02
7	1	1.56	1.74	1.87	1.49
8	0	1.16	1.27	1.16	1.11
9	1	0.86	0.92	0.72	0.84
10	1	0.64	0.67	0.44	0.64
Estimates		$\hat{q}=0.869$ $\hat{\alpha}=0.420$	$\hat{q}=0.728$ $\hat{\beta}=1.657$	$\hat{q}=0.778$ $\hat{\alpha}=3.843$ $\hat{\beta}=28.726$	$\hat{r}=0.369$ $\hat{\mu}=1.580$
AIC		330.424	325.645	321.227	333.199
BIC		335.634	330.856	329.043	338.409
AD statistics		0.5978	0.333	0.121	0.748
p-value		0.389	0.604	0.911	0.31

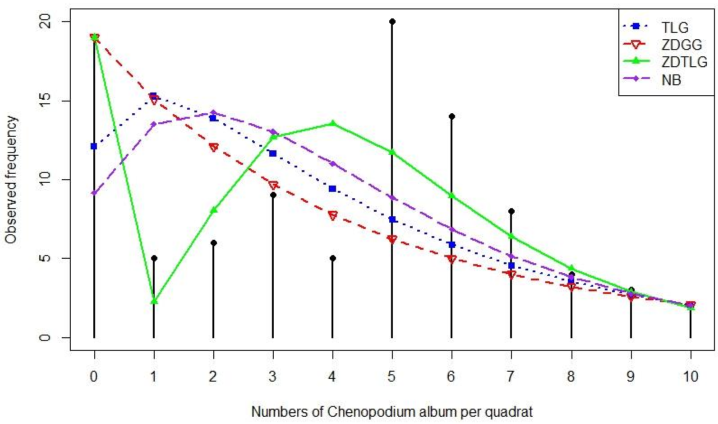


Figure 3 Plots of observed and expected values of Chenopodium album based on the TLG, ZDGG, ZDTLG and NB distributions

Table 3 Numbers of Chenopodium album counts per quadra

Numbers of Plant	Observed	Expected			
		TLG	ZDGG	ZDTLG	NB
0	19	12.08	18.99	18.99	9.16
1	5	15.26	15.08	2.25	13.51
2	6	13.86	12.09	8.04	14.25
3	9	11.63	9.69	12.69	13.03
4	5	9.41	7.77	13.53	11.01
5	20	7.47	6.23	11.69	8.83
6	14	5.85	4.99	8.95	6.84
7	8	4.54	4.00	6.37	5.16
8	4	3.50	3.21	4.35	3.81
9	3	2.69	2.57	2.88	2.77
10	2	2.07	2.06	1.88	1.99
Estimates		$\hat{q}=0.872$ $\hat{\alpha}=1.445$	$\hat{q}=0.802$ $\hat{\beta}=0.009$	$\hat{q}=0.796$ $\hat{\alpha}=6.776$ $\hat{\beta}=201.0385$	$\hat{r}=2.327$ $\hat{\mu}=4.032$
AIC		476.237	480.715	444.44	470.185
BIC		481.345	485.823	452.101	475.293
AD statistics		0.175	0.208	0.077	0.222
p-value		0.741	0.621	0.964	0.709

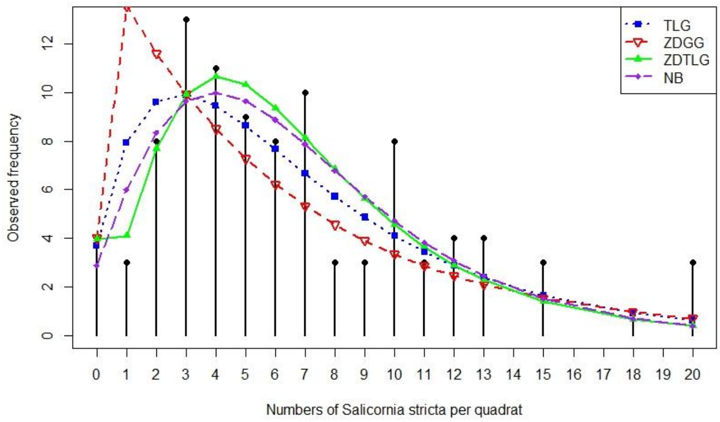


Figure 4 Plots of observed and expected values of Salicornia stricta based on the TLG, ZDGG, ZDTLG and NB distributions

Table 4 Numbers of Salicornia stricta counts per quadrat

Numbers of Plant	Observed	Expected			
		TLG	ZDGG	ZDTLG	NB
0	4	3.68	4.00	3.97	2.87
1	3	7.94	13.55	4.09	6.01
2	8	9.61	11.6	7.69	8.34
3	13	9.91	9.93	9.92	9.63
4	11	9.45	8.49	10.66	9.98
5	9	8.63	7.27	10.32	9.65
6	8	7.66	6.22	9.35	8.88
7	10	6.66	5.32	8.12	7.87
8	3	5.72	4.56	6.83	6.78
9	3	4.86	3.90	5.62	5.71
10	8	4.09	3.34	4.56	4.72
11	3	3.43	2.86	3.65	3.84
12	4	2.86	2.44	2.90	3.08
13	4	2.38	2.09	2.28	2.44
14	0	1.97	1.79	1.79	1.92
15	3	1.63	1.53	1.40	1.50
16	0	1.34	1.31	1.09	1.16
17	0	1.11	1.12	0.85	0.89
18	1	0.91	0.96	0.66	0.68
19	0	0.75	0.82	0.51	0.52
20	3	0.62	0.70	0.40	0.39
Estimates		$\hat{q}=0.905$ $\hat{\alpha}=1.922$	$\hat{q}=0.856$ $\hat{\beta}=-0.732$	$\hat{q}=0.879$ $\hat{\alpha}=3.287$ $\hat{\beta}=4.295585$	$\hat{r}=3.05$ $\hat{\mu}=6.653$
AIC		560.206	575.278	556.502	556.021
BIC		565.376	580.448	564.257	561.191
AD statistics		0.651	3.218	0.211	0.225
p-value		0.519	0.017	0.953	0.946

CONCLUSIONS

We create the ZDTLG geometric distribution based on the k-distorted generalized discrete family of distributions. The ZDTLG distribution offers flexible benefits in modeling both zero-inflated and zero-deflated data. The ZDTLG geometric distribution includes the Topp-Leone geometric, the discrete zero-truncated Topp-Leone geometric, the zero-deflated Topp-Leone geometric and the zero-inflated Topp-Leone geometric distributions as spatial cases. Its first four moments can be derived based on the TLG distribution. To estimate the ZDTLG parameters, the maximum likelihood estimation is applied. Moreover, real biological datasets are fitted with the proposed distribution, the TLG distribution, the ZDGG distribution, and the NB distribution. As the ZDTLG distribution has the smallest AD statistics and the biggest *p*-values, we can conclude that the ZDTLG distribution is the best model to describe these real datasets.

ACKNOWLEDGEMENTS

There are overwhelming to thank the Department of Statistics, Faculty of Science, Kasetsart University, Graduate School Kasetsart University and the Research Professional Development Project under the Science Achievement Scholarship of Thailand (SAST). As well as, thank you Bliss and Fisher for datasets in application part (Bliss and Fisher, 1953).

REFERENCES

Akinsete, A., Famoye, F., and Lee. C. (2014). The Kumaraswamy-geometric distribution. *Journal of statistical distributions and applications*, *1*(1), 17.

Aljarrah, M.A., Lee, C., and Famoye, F. (2014). On generating T-X family of distributions using quantile functions. *Journal of Statistical Distributions and Applications*, *1*(1), 2.

Arnold, T.A., and John, E.W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, *3*(2), 34-39.

Bhati, D., and Joshi, S. (2018). Weighted geometric distribution with new characterizations of geometric distribution. *Communications in Statistics-Theory and Methods*, *47*(6), 1510-1527.

Bliss, C. I., and Fisher, R. A. (1953). Fitting the Negative Binomial Distribution to Biological Data. *Biometrics*, *9*(2), 176-200. doi:10.2307/3001850.

Chandra, N. K., Roy, D., and Ghosh T. (2013). A generalized Poisson distribution. *Communications in Statistics-Theory and Methods*, *42*(15), 2786-2797.

Delignette-Muller, M., and Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1-34.

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *Test*, *19*(2), 399-415.

Gómez-Déniz, E., Sarabia, J. M., and Calderín-Ojeda, E. (2011). A new discrete distribution with actuarial applications. *Insurance:Mathematics and Economics*, *48*(3), 406-412.

Johnson, N.L., Kemp, A.W., and Kotz, S. (2005). *Univariate discrete distributions*. John Wiley & Sons.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions*. John Wiley & Sons, New York.

Nekoukhou, V., Alamatsaz, M. H., and Bidram, H. (2012). A discrete analog of the generalized exponential distribution. *Communications in Statistics-Theory and Methods*, *41*(11), 2000-2013.

Perumean-Chaney, S. E., Morgan, C., McDowall, D., & Aban, I. (2013). Zero-inflated and overdispersed: what's one to do?. *Journal of Statistical Computation and Simulation*, 83(9), 1671-1683.

R Core Team. (2017). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Sastry, D.V.S., Bhati, D., Rattihalli, R.N., and Gómez-Déniz, E. (2016). On zero-distorted generalized geometric distribution. *Communications in Statistics-Theory and Methods*, *45*(18), 5427-5442.

Shirke, D. T., Supanekar, S. R., and Bhati, D. (2017). On k-distorted generalized discrete family of distributions. *Communications in Statistics-Theory and Methods*, *46*(23), 11591-11603.

Sudsuk, A., and Bodhisuwan, W. (2016). The Topp-Leone geometric distribution. *In Mathematics, Statistics, and Their Applications (ICMSA), 2016 12th International Conference, IEEE*, 108-112.

Topp, C. W., and Leone, F. C. (1955). A family of J-shaped frequency functions. *Journal of the American Statistical Association, 50*(269), 209-219.

Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Science & Business Media.

Xavier, D., Santos-Neto, M., Bourguignon, M., and Tomazella, V. (2018). Zero-modified Poisson-Lindley distribution with applications in zero-inflated and zero-deflated count data. *Biometrics & Biostatistics International Journal*, *7*(2).

Zamani, H., Ismail, N., and Faroughi, P. (2014). Poisson-weighted exponential univariate version and regression model with applications. *Journal of Mathematics and Statistics, 10*(2), 148-154.