ATLANTIS PRESS Journal of Statistical Theory and Applications, Vol. 16, No. 3 (September 2017) 322–336

Generalized Sichel Distribution and Associated Inference

Yeh Ching Low

Institute of Mathematical Sciences, University of Malaya,
50603 Kuala Lumpur, Malaysia
Department of Computing and Information Systems, Sunway University,
47500 Selangor, Malaysia
yehchingl@sunway.edu.my

Seng Huat Ong

Institute of Mathematical Sciences, University of Malaya,
50603 Kuala Lumpur, Malaysia
Faculty of Business and Information Science, UCSI University,
56000 Kuala Lumpur, Malaysia
ongsh@um.edu.my

Ramesh C. Gupta*

Department of Mathematics and Statistics, University of Maine,
Orono, ME 04469, USA
Ramesh_Gupta@umit.maine.edu

Received 16 January 2016

Accepted 7 December 2016

Abstract

In this paper, we propose a generalized form of Sichel distribution which is obtained by mixing the Poisson distribution with the extended generalized inverse Gaussian distribution. This distribution models over dispersed, zero-inflated and heavy-tailed count data sets. These characteristics are examined with respect to the dispersion, zero-inflation and the third central moment inflation indices. Examples are provided to compare the extension with several other existing models including the Poisson-inverse Gaussian and the Sichel distributions.

Keywords: Poisson distribution; extended generalized inverse Gaussian distribution; overdispersion; zero-inflation; long-tailed distributions.

2000 Mathematics Subject Classification: 62E15, 62F03, 62F10, 62P30

1. Introduction

It is well known that count data often shows over dispersion relative to Poisson distribution for which variance equals the mean. Over dispersion means that the variance is greater than the mean and a common

measure for it is the index of dispersion. The over dispersion can be due to various situations, for instance, unobserved heterogeneity in the data, or having extra zeros than produced by the model. Mullahy [15] demonstrated that unobserved heterogeneity, commonly assumed to be the source of over dispersion in the count data model, have predictable implications for the probability structures of such models. One way to take care of the heterogeneity is by way of mixture models. In the case of the mixed Poisson distribution, the mean θ of the Poisson distribution is considered as a random variable with an appropriate probability structure.

The simplest choice of the distribution of θ is the gamma density, resulting in a negative binomial distribution (NB) which is introduced by Greenwood and Yule [5]. Some generalizations of the NB have been studied by applying a generalized gamma distribution resulting in a generalized form of NB, see Gupta and Ong [7] and the references therein. Other choices for the distribution of θ include the inverse Gaussian and the generalized inverse Gaussian (GIG), giving rise to the Poisson-inverse Gaussian (PIG) and Poisson-generalized inverse Gaussian or Sichel (PGIG) distribution respectively, see Refs. 1, 10, 19 and 23 for more details. The Sichel distribution is a long-tailed distribution that is found to be suitable for highly skewed data and it has been used, amongst others, to model insurance claim counts, protein abundance, word frequency in a text and consumer purchase behaviour. In addition to the mixing distributions mentioned above, various other distributions such as the lognormal, Lindley and shifted gamma distributions have been used to obtain mixed Poisson distributions; for more examples and illustrations, see Refs. 7, 11, 14 and 16.

In this paper, we consider the distribution of θ as that of extended generalized inverse Gaussian (EGIG). This distribution is briefly mentioned by Jørgensen [12] and further studied by Gupta and Viles [8], [9]. The EGIG model has one additional parameter (δ) than the generalized inverse Gaussian (GIG) model having three parameters, see Ref. 12. Gupta and Viles [9] have provided examples to illustrate that the EGIG model fits the data better than the GIG model. In the same paper, they have also shown the importance of the additional parameter δ. We call the resulting mixed Poisson-EGIG distribution as the generalized Sichel distribution. This includes the PGIG distribution as a special case.

The work in the present paper is motivated by the fact that there are a number of count frequency data sets with very high zero counts or very long right tails which may not be adequately fitted by existing mixed Poisson models. For high zero counts it is customary to use a zero-inflated model if structural zeros are involved. Based on Shaked's [22] Two-Crossings Theorem, a mixed Poisson distribution, relative to the Poisson distribution, has a higher probability for the zero count and a longer right tail. This elevation of probability for zero counts and tail lengthening will vary according to the mixing distributions considered. We shall show that the proposed generalized Sichel distribution fits better than the PGIG and other well-known mixed Poisson distributions when the data has high zero counts and/or a long tail. This is illustrated with three data sets with high zero counts but different tail lengths. Since this proposed distribution contains the NB, PIG and Sichel distributions as special cases, an advantage of this generalized Sichel distribution is that it can eliminate the need of piece-wise treatment of these distributions when fitting a data set with high zero counts and/or long right tail.

The organization of this paper is as follows: we present the generalized Sichel distribution in Section 2. In Section 3, the shape and properties of the generalized Sichel distribution are discussed. Statistical inference procedures are presented in Section 4 and methods for computing the estimates of the parameters are indicated. Hypothesis testing procedures are developed for testing the hypothesis $H_0 : \delta = 1$ i.e. the data follows a Sichel distribution. Section 5 contains the fitting of the proposed model to a simulated data set and two well-known data sets from the literature. Finally, some conclusion and comments are presented in Section 6.

2. The Generalized Sichel Distribution

Let *Y* be a random variable with support on nonnegative real numbers. Then the probability density function of the extended generalized inverse Gaussian (EGIG) distribution is given by

$$f(y) = \frac{1}{(2/\delta)(b/a)^{\lambda/2\delta} K_{\lambda/\delta}\left(2\sqrt{ab}\right)} y^{\lambda-1} \exp\left(-ay^{\delta} - by^{-\delta}\right), \quad y > 0 \tag{1}$$

where $K_{\nu}(z)$ is the modified Bessel function of the third kind with index ν. Here we follow the notation adopted by Gupta and Viles [9]. We adopt a similar domain of variation for the parameters to that given by Jørgensen [12], that is $\lambda \in \Re$ $(a,b,\delta) \in \Omega_{\lambda}$, where

$$\Omega_{\lambda} = \begin{cases} (a,b,\delta): a > 0, b \geq 0, \delta > 0 \text{ iff } \lambda > 0 \\ (a,b,\delta): a > 0, b > 0, \delta > 0 \text{ iff } \lambda = 0 \\ (a,b,\delta): a \geq 0, b \geq 0, \delta > 0 \text{ iff } \lambda < 0 \end{cases}.$$

When $\delta = 1$, the EGIG model reduces to the generalized inverse Gaussian (GIG) model which has been studied in detail by Jørgensen [12]. Other special and limiting cases of Eq. (1) include the inverse Gaussian distribution ($\delta = 1, \lambda = -0.5$), the gamma distribution ($\delta = 1$, b = 0 and $\delta = 1, \lambda = -0.5$), the Weibull distribution and the exponential distribution.

Definition 2.1 (*Generalized Sichel distribution*) Suppose *X* is a discrete random variable and $X \mid \Theta \sim$ Poisson(θ), where Θ is a nonnegative real valued random variable with pdf $f(\theta)$ given by Eq. (1). Then *X* has the generalized Sichel distribution with probability mass function (pmf) given by

$$P(X = k) = \frac{1}{(2/\delta)(b/a)^{\frac{\lambda}{2\delta}} K_{\lambda/\delta}(2\sqrt{ab})} \int_0^{\infty} \frac{e^{-\theta}\theta^k}{k!} \theta^{\lambda-1} \exp(-a\theta^{\delta} - b\theta^{-\delta}) d\theta, \tag{2}$$

which can be written as

$$P(X = k) = \left(\frac{1}{K_{\lambda/\delta}(2\sqrt{ab})k!}\right) \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left(\frac{b}{a}\right)^{(j+k)/2\delta} K_{(j+k+\lambda)/\delta}(2\sqrt{ab}). \tag{3}$$

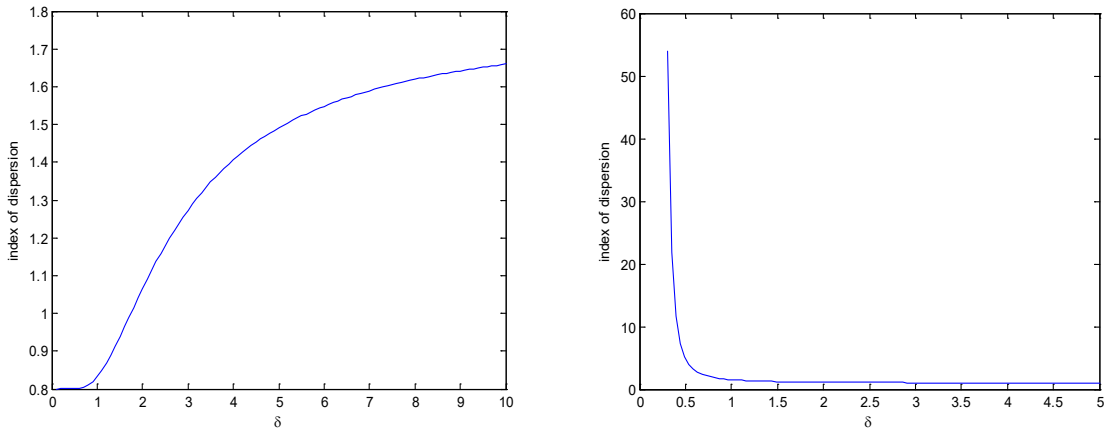
The generalized Sichel probabilities may be computed from Eq. (2) by numerical integration or by using the infinite series form in Eq. (3) where computation of the $K_{\nu}(z)$ is facilitated by the recurrence relation $K_{\nu+1}(z) = (2\nu/z)K_{\nu}(z) + K_{\nu-1}(z)$. For a discussion on issues concerning computation with recurrence formulae see Ref. 17.

The probability generating function (pgf) of the generalized Sichel distribution is given by

$$G(z)=\left(\frac{1}{K_{\lambda/\delta}(2\sqrt{ab})}\right)\sum_{j=0}^{\infty}\frac{(z-1)^j}{j!}\left(\frac{b}{a}\right)^{j/2\delta}K_{(j+\lambda)/\delta}(2\sqrt{ab}),\ |z|\le 1. \tag{4}$$

The generalized Sichel distribution has mean $\mu=\dfrac{(b/a)^{1/2\delta}K_{(1+\lambda)/\delta}(2\sqrt{ab})}{K_{\lambda/\delta}(2\sqrt{ab})}$ and variance $\sigma^2=\dfrac{(b/a)^{1/\delta}}{K_{\lambda/\delta}(2\sqrt{ab})}\left[K_{(2+\lambda)/\delta}(2\sqrt{ab})-\dfrac{\left(K_{(1+\lambda)/\delta}(2\sqrt{ab})\right)^2}{K_{\lambda/\delta}(2\sqrt{ab})}\right]+\mu$. Consequently, an expression for the index of dispersion can be written as $ID_X=1+\left(\dfrac{b}{a}\right)^{1/2\delta}\left[\dfrac{K_{(2+\lambda)/\delta}(2\sqrt{ab})}{K_{(1+\lambda)/\delta}(2\sqrt{ab})}-\dfrac{K_{(1+\lambda)/\delta}(2\sqrt{ab})}{K_{\lambda/\delta}(2\sqrt{ab})}\right]$.

The special case $\delta=1$ gives rise to the Sichel distribution. Furthermore, when $\delta=1$ and $\lambda=-0.5$, we obtain the PIG distribution. The Poisson-Gamma (or NB) distribution is obtained from Eq. (2) when $\delta=1$, $b=0$ and $\lambda>0$. These special cases are derived based on the probability structure of the EGIG model. The extra parameter δ adds flexibility to the shape of the count distribution. The effect of varying the parameter δ on the index of dispersion ID_X is illustrated in Figure 1 below.



(a) (b)

Fig. 1 Plot of index of dispersion versus δ (a) $a = b = 0.95$, (b) $a = 1.0$, $b = 0.1$; λ = -0.5

3. Shape and Properties of the Generalized Sichel Distribution

The generalized Sichel distribution is a flexible model which is able to model data with zero-inflation, over dispersed and long-tailed data. Three examples of the generalized Sichel pmf plots are given in Figure 2 to illustrate the versatility of the shape of the distribution.

We examine the shape of the generalized Sichel distribution in terms of the zero-inflation index and the third central moment inflation index as defined by Puig and Valero [20]. For the generalized Sichel distribution, both the zero-inflated and the central moment indices are dependent on the mean and they are obtained using numerical computation.

3.1. *Zero-inflation index*

The zero-inflation index of a non-negative integer random variable X with mean μ and proportion of zeros p_0 is defined as $zi = 1 + \log(p_0)/\mu$ (see [20]). The Poisson random variable has a zero-inflation index of 0, and a zero-inflated random variable will have a positive zero-inflation index. It is known that any mixed Poisson random variable is zero-inflated. Thus it is of interest to know the amount of zero-inflation.

We plot the zero-inflation index versus index of dispersion for the negative binomial (NB), Poisson inverse Gaussian (PIG) and generalized Sichel (GS) distributions in Figure 3. The zero-inflation index for NB and PIG are independent of the mean of the distribution and it can be expressed as $1 + \log(ID)/(1 - ID)$ and $(ID - \sqrt{2(ID) - 1})/(ID - 1)$, respectively, where ID is the index of dispersion. We consider the cases when mean = 5 and mean = 15 for the generalized Sichel distribution, representing small and large mean, respectively. When over dispersion is small, all three distributions are similar. The zero-inflation index of the generalized Sichel distribution increases with the value of its mean. When the mean is small, the generalized Sichel's zero-inflation index is closer to that of the PIG than the NB distribution. In general, the generalized Sichel distribution has the flexibility of having a larger zero-inflation index than that of the PIG. The generalized Sichel distribution is flexible in modelling the presence of extra zeros, since it has a zero-inflation index which can be higher than the NB distribution.

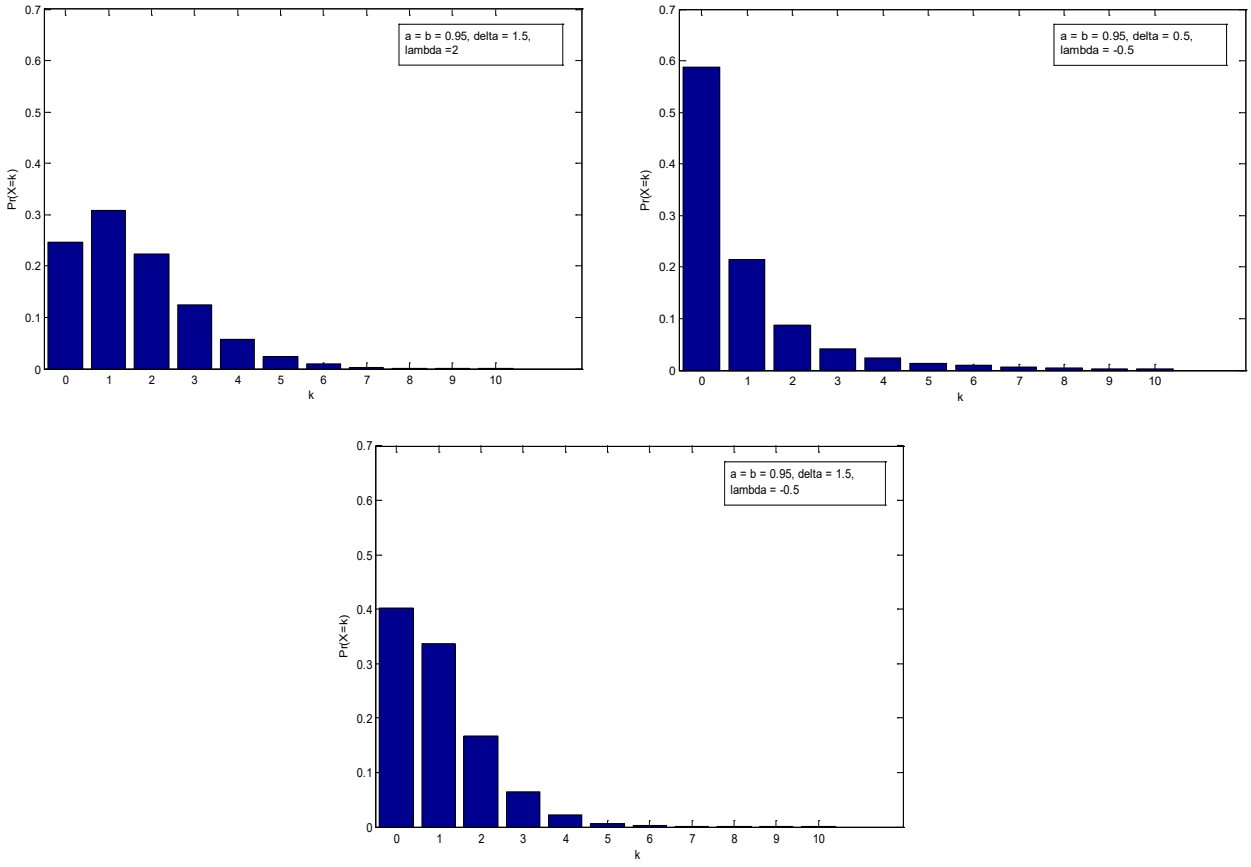


Fig. 2 Probability mass function (pmf) plots of the generalized Sichel distribution

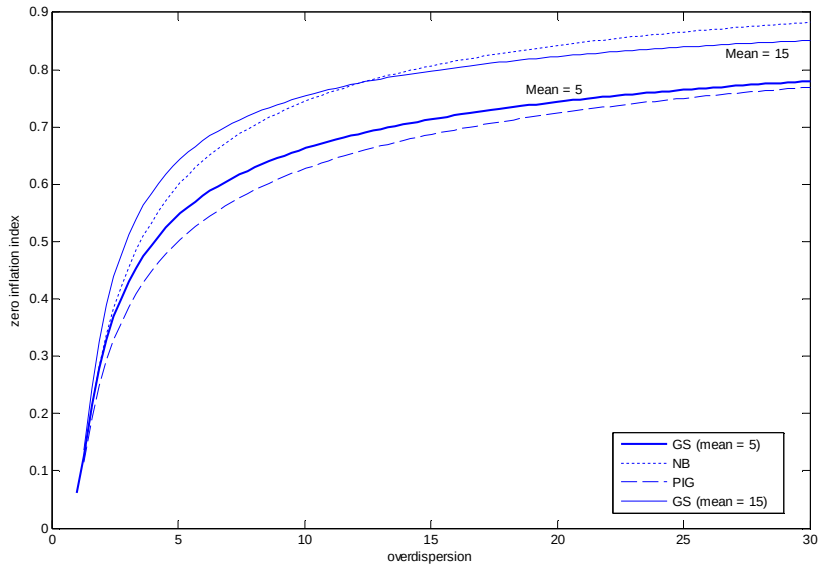


Fig. 3 Zero-inflation index versus index of dispersion

3.2. *Discriminant ratio*

Ong and Muthaloo [18] have discussed the role of the discriminant ratio which is defined as $Q(k) = P(X = k+1)/P(X = k)$, for k = 0, 1, 2, 3, …, in determining the flexibility of the distributions which they proposed for long-tailed data. The ratio has a limiting value of $Q(k) \to 1$ for long-tailed distributions. Figure 4 gives the graphs of $Q(k)$ versus k for several values of the parameter δ, holding other parameters fixed.

In Figure 4(a), we compare the graph of $Q(k)$ versus k for the PIG distribution (δ = 1, λ = -0.5) and the generalized Sichel distribution. By varying the value of δ, the discriminant ratio varies considerably especially at large values of k. The difference is most prominent for k larger than 10.

From the graphs in Figure 4(b), we note that the generalized Sichel distribution has a longer tail compared to the Sichel distribution. The trend is similar to that in Figure 4(a). As such, the parameter δ adds flexibility to the generalized Sichel distribution, enabling the distribution to model data with a very long tail.

3.3. *Third central moment inflation index*

The third central moment inflation index of a nonnegative discrete random variable X describes the skewness of the distribution and is obtained as $\kappa_3 = (\mu_3 / \mu) - 1$, where μ_3 is the third central moment of X. This index takes the value 0 for the Poisson distribution.

In Figure 5 we plot the third central moment inflation index versus over dispersion for the NB, PIG and generalized Sichel (GS) distributions. For the GS distribution, the plot is given for three different values of the mean, i.e. 5, 10, 15. The index for the NB and PIG distributions take the values $2(ID)^2$ - ID - 1 and $3(ID)^2$ – $3(ID)$, respectively. The coefficient of skewness is positive for all three distributions. For small over dispersion, all of the distributions are similar to each other. As the index of dispersion increases, the coefficient of skewness of all the distributions increases. For the GS distribution, as the mean increases, the coefficient of skewness decreases. Moreover, GS distribution with larger means are closer to the NB than the PIG. The GS distribution with a small mean has more probability mass distributed at the tail extending to the

right relative to the NB and PIG distributions. This further suggests that the generalized Sichel distribution is able to model long-tailed data better.

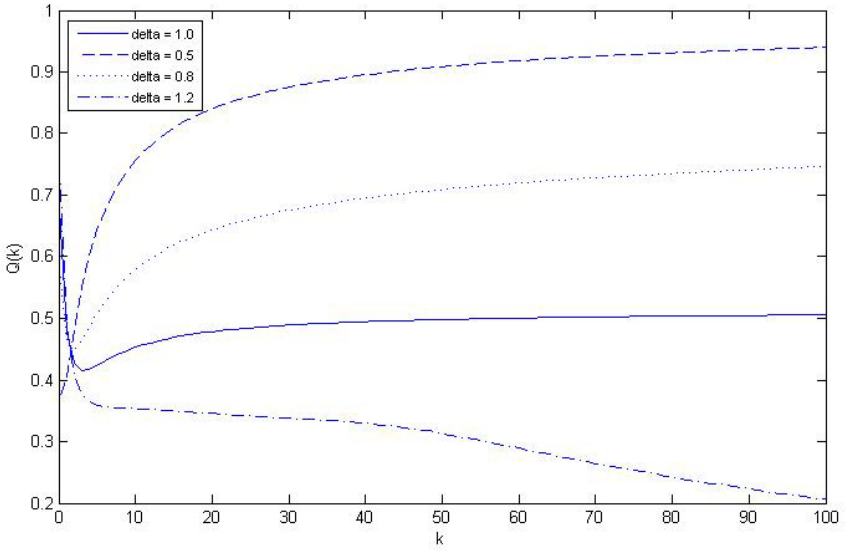


Fig. 4(a) Discriminant ratio when $a = b = 0.95$, $\lambda = -0.5$

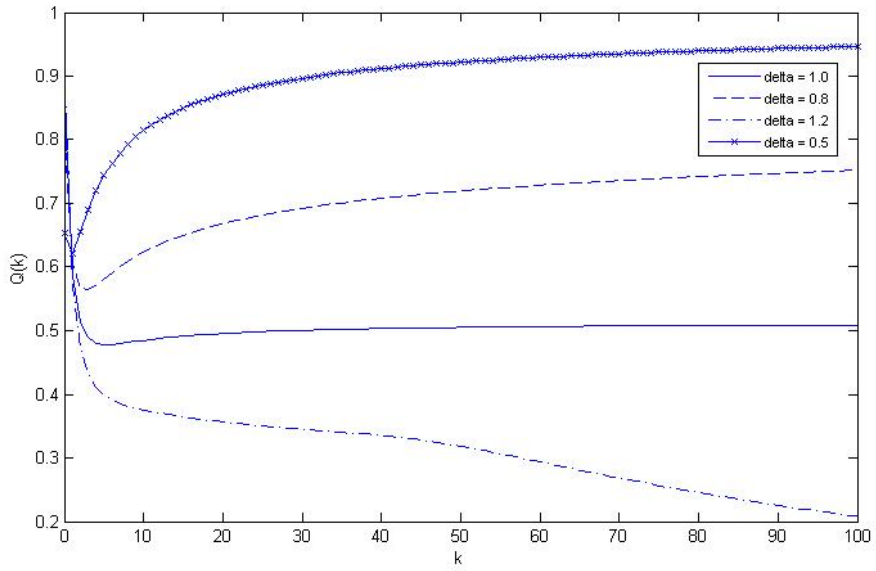


Fig. 4(b) Discriminant ratio when $a = b = 0.95$, $\lambda = 0.2$

4. Statistical Inference

4.1. *Parameter estimation*

Maximum likelihood (ML) estimation is used to estimate the unknown parameters $\boldsymbol{\omega} = (a, b, \delta, \lambda)$ of the generalized Sichel distributions, given the observations from the sample of interest. The ML estimates of the

generalized Sichel distribution is defined as $\hat{\boldsymbol{\omega}} = (\hat{a}, \hat{b}, \hat{\delta}, \hat{\lambda})^T = \arg\max_{\boldsymbol{\omega}} \log L(\boldsymbol{\omega})$, where $\log L$ is the log-likelihood function given by $\log L = \sum_{k=0}^{\infty} f_k \cdot \log[P(X=k)]$ and f_k is the observed frequency of count k in the sample.

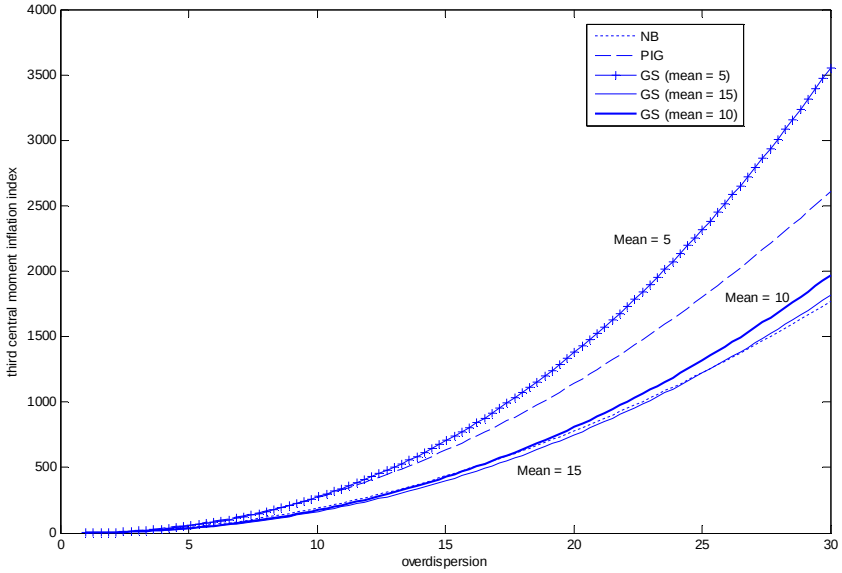


Fig. 5 Third central moment inflation index versus index of dispersion.

In order to obtain the likelihood score equations, we derive the partial derivatives of the log-likelihood function using the generalized Sichel pmf based on Eq. (2). As such,

$$\log L = \sum_{k=0}^{\infty} f_k \cdot \left[-\log\left(\frac{2}{\delta}\right) - \frac{\lambda}{2\delta}\log\left(\frac{b}{a}\right) - \log K_{\lambda/\delta}(2\sqrt{ab}) - \log k! + \log h(\theta) \right]$$

where $h(\theta) = \int_0^{\infty} e^{-\theta}\theta^{k+\lambda-1}\exp(-a\theta^{\delta} - b\theta^{-\delta}d\theta$.

Then the partial derivatives are

$$\frac{\partial \log L}{\partial a} = \sum_{k=0}^{\infty} f_k \left[\frac{\lambda}{2a\delta} - \frac{\partial}{\partial a}\log K_{\lambda/\delta}(2\sqrt{ab}) + \frac{\partial}{\partial a}\log h(\theta) \right] \tag{5}$$

$$\frac{\partial \log L}{\partial b} = \sum_{k=0}^{\infty} f_k \left[-\frac{\lambda}{2b\delta} - \frac{\partial}{\partial b}\log K_{\lambda/\delta}(2\sqrt{ab}) + \frac{\partial}{\partial b}\log h(\theta) \right] \tag{6}$$

$$\frac{\partial \log L}{\partial \delta} = \sum_{k=0}^{\infty} f_k \left[\frac{1}{\delta} + \frac{\lambda}{2\delta^2}\log\left(\frac{b}{a}\right) - \frac{\partial}{\partial \delta}\log K_{\lambda/\delta}(2\sqrt{ab}) + \frac{\partial}{\partial \delta}\log h(\theta) \right] \tag{7}$$

$$\frac{\partial \log L}{\partial \lambda} = \sum_{k=0}^{\infty} f_k \left[-\frac{1}{2\delta} \log\left(\frac{b}{a}\right) - \frac{\partial}{\partial \lambda} \log K_{\lambda/\delta}(2\sqrt{ab}) + \frac{\partial}{\partial \lambda} \log h(\theta) \right] \tag{8}$$

where

$\frac{\partial}{\partial a}\{h(\theta)\} = \int_0^{\infty} (-\theta^{\delta}) e^{-\theta} \theta^{k+\lambda-1} \exp(-a\theta^{\delta} - b\theta^{-\delta}) d\theta$,

$\frac{\partial}{\partial b}\{h(\theta)\} = \int_0^{\infty} (-\theta^{-\delta}) e^{-\theta} \theta^{k+\lambda-1} \exp(-a\theta^{\delta} - b\theta^{-\delta}) d\theta$,

$\frac{\partial}{\partial \delta}\{h(\theta)\} = \int_0^{\infty} (\log\theta)(-a\theta^{\delta} + b\theta^{-\delta}) e^{-\theta} \theta^{k+\lambda-1} \exp(-a\theta^{\delta} - b\theta^{-\delta}) d\theta$,

$\frac{\partial}{\partial \lambda}\{h(\theta)\} = \int_0^{\infty} (\log\theta) e^{-\theta} \theta^{k+\lambda-1} \exp(-a\theta^{\delta} - b\theta^{-\delta}) d\theta$, and the derivatives of the modified Bessel function of the third kind is obtained by differentiating its integral representation

$$K_{\lambda/\delta}(2\sqrt{ab}) = \int_0^{\infty} \exp(-2\sqrt{ab}\cosh t) \cosh\left(\frac{\lambda}{\delta} t\right) dt.$$

Since the pmf of the generalized Sichel distribution is complicated, ML estimation can be done using numerical optimization methods such as the simulated annealing algorithm discussed by Goffe *et al*. [4]. Simulated annealing is a stochastic-type global optimization algorithm which is able to work with functions which are not smooth or having many local maxima or minima.

4.2. *Akaike Information Criterion*

The Akaike Information Criterion (AIC) is a model selection criterion to choose from several competing models for a particular data set. It is calculated as AIC = -2 log L + 2p, where log L in the formula is the maximized log-likelihood value and p is the number of parameters. The AIC penalizes the model with more parameters. Based on this criteria, the model which the smallest AIC value is selected as the best model.

4.3. *Hypothesis testing*

The generalized Sichel distribution nests the Sichel distribution. As such, of particular interest would be hypothesis test for the additional parameter, which we shall name as the Sichel test. For the Sichel test, the hypotheses can be written as

$$H_0 : \delta = 1 \quad \text{vs} \quad H_A : \delta \neq 1 \tag{9}$$

Hypothesis testing procedures such as the likelihood ratio test, score test and Wald test can be performed based on the likelihood function. Under the null hypothesis, the likelihood ratio test and score test are asymptotically equivalent. We employ the score test which has the advantage of being simpler to compute since it requires only the restricted maximum likelihood estimates, which corresponds to those of the Sichel model for the Sichel test in Eq. (9). The score test statistic is $T = U(\boldsymbol{\omega}_0) J^{-1}(\boldsymbol{\omega}_0) U(\boldsymbol{\omega}_0)$, where the score

vector $U(\cdot) = (\partial \log L / \partial a, \partial \log L / \partial b, \partial \log L / \partial \delta, \partial \log L / \partial \lambda)$ is being evaluated at the null hypothesis, i.e. $\boldsymbol{\omega}_0 = (\hat{a}, \hat{b}, \mathrm{l}, \hat{\lambda})^T$ the restricted maximum likelihood estimates. The partial derivatives $\partial \log L / \partial a$, $\partial \log L / \partial b$, $\partial \log L / \partial \delta$ and $\partial \log L / \partial \lambda$ are given by Eqs. (5), (6), (7) and (8), respectively. In general, $J(\boldsymbol{\omega}_0)$ is either the expected or observed information matrix, also evaluated at the null hypothesis. In our case for the generalized Sichel distribution, we use the observed information matrix because the expected information matrix is intractable. The observed information matrix is the matrix of second partial derivatives of the log-likelihood function and its elements are obtained as $I(\cdot) = \left[i_{rs}(\boldsymbol{\omega}) \right] = \left[-\frac{\partial^2 \log L}{\partial \omega_r \partial \omega_s} \right]$, where $i_{rs}(\boldsymbol{\omega})$ is the element in the *r*-th row and *s*-th column and ω_p is the *p*-th element in the parameter vector $\boldsymbol{\omega} = (a, b, \delta, \lambda)$. The score test statistic T has an asymptotic chi-square distribution with one degree of freedom.

5. Applications

To examine the suitability of the model for zero-inflated, over dispersed and long-tailed data sets, we fit one simulated data set and two well-known data sets from the literature with our proposed model and compare it with related mixed Poisson distributions and zero-inflated Poisson distribution. In terms of mixed Poisson distributions, we compare the model fitting of the generalized Sichel distribution with the negative binomial (NB), Poisson-inverse Gaussian (PIG) and Poisson-generalized inverse Gaussian (Sichel) (PGIG) distributions. The NB, PIG and PGIG distributions can be derived as a special case of the generalized Sichel distribution. The pmf of the NB, PIG and PGIG distributions used in this model fitting are given below.

Mixed Poisson Distribution	**Probability mass function**	**Parameter Domain**
Negative binomial (NB)	$P(X=k) = \binom{a+k-1}{a-1} \left(\frac{\beta}{\beta+1} \right)^k \left(\frac{1}{\beta+1} \right)^\alpha$	$\alpha > 0$ $\beta > 0$
Poisson-inverse Gaussian (PIG)	$P(X=k) = \sqrt{\frac{2\alpha}{\pi}} \frac{\exp(\alpha\sqrt{1-\theta}) \left(\frac{\alpha\theta}{2} \right)}{k!} K_{k-\frac{1}{2}}(\alpha)$	$\alpha > 0$ $0 < \theta < 1$
Poisson-generalized inverse Gaussian (PGIG)	$P(X=k) = \frac{(1-\theta)^{\frac{\gamma}{2}} \left(\frac{\alpha\theta}{2} \right)^k}{k! K_\gamma(\alpha\sqrt{1-\theta})} K_{k+\gamma}(\alpha)$	$\alpha > 0$ $0 < \theta < 1$ $-\infty < \gamma < \infty$

Besides the mixed Poisson distributions, we also fitted the two real data sets to the zero-inflated Poisson (ZIP) distribution since the high proportion of zeros in both data sets suggests that some of these may be

structural zeros. The pmf of the ZIP distribution is defined as $P(X=0) = p + (1-p)e^{-\lambda}$ and for k = 1, 2, 3, ..., $P(X=k) = (1-p)\frac{e^{-\lambda}\lambda^k}{k!}$.

The ML estimates together with their maximized log-likelihoods are presented in Table 1. The standard error for the parameters of the generalized Sichel distribution are obtained from the observed information matrix defined in Section 4.3. The observed frequency and the fitted distributions are presented in Tables 2, 3 and 4, together with the degrees of freedom, χ^2-statistic, p-values and AIC values. The degree of freedom is equal to $(t-p-1)$ where t = number of classes and p = number of parameters.

Table 1: Maximum Likelihood Estimates and Log-likelihood Function Values

Data Set	Maximum Likelihood Estimates and Log-likelihood Values				
	NB	PIG	PGIG	Generalized Sichel (standard error)	Zero-inflated Poisson
Simulated data	$\hat{\alpha}$ = 0.2989 $\hat{\beta}$ = 33.9736 L = -13697.53	$\hat{\alpha}$ = 1.3198 $\hat{\theta}$ = 0.9983 L = -13836.28	$\hat{\alpha}$ = 0.3576 $\hat{\theta}$ = 0.9816 $\hat{\gamma}$ = -0.0774 L = -13654.18	$\hat{a}$ = 0.2058 (0.0000) $\hat{b}$ = 0.0624 (0.0305) $\hat{\delta}$ = 0.5750 (0.0057) $\hat{\lambda}$ = 0.2632 (0.0385) L = -13650.23	N/A
Tröbliger's data (1961) on number of claims	$\hat{\alpha}$ = 1.1514 $\hat{\beta}$ = 0.1246 L = -10180.29	$\hat{\alpha}$ = 1.2443 $\hat{\theta}$ = 0.2055 L = -10178.42	$\hat{\alpha}$ = 1.3138 $\hat{\theta}$ = 0.4060 $\hat{\gamma}$ = -1.8177 L = -10177.62	$\hat{a}$ = 1.1407 (1.3912) $\hat{b}$ = 0.1514 (0.2447) $\hat{\delta}$ = 1.0560 (0.4198) $\hat{\lambda}$ = -1.9458 (0.8237) L = -10177.60	$\hat{\lambda}$ = 0.2538 $\hat{p}$ = 0.4348 L = -10190.58
Accident injuries data	$\hat{\alpha}$ = 2.0361 $\hat{\beta}$ = 0.3474 L = -11485.48	$\hat{\alpha}$ = 2.4604 $\hat{\theta}$ = 0.4330 L = -11466.82	$\hat{\alpha}$ = 2.7035 $\hat{\theta}$ = 0.9671 $\hat{\gamma}$ = -3.4594 L = -11454.22	$\hat{a}$ = 0.0027 (0.0067) $\hat{b}$ = 0.0715 (0.1170) $\hat{\delta}$ = 3.0000 (1.3142) $\hat{\lambda}$ = -2.1791 (0.2218) L = -11450.78	$\hat{\lambda}$ = 0.9135 $\hat{p}$ = 0.2257 L = -11613.88

5.1. *Simulated data*

In this section, we illustrate the application of the generalized Sichel distribution with a simulated data set with very long tail. The Malayan butterfly data is a well-known example in the literature on long-tailed data, see Ref. 6. However, the frequencies after k = 25 are grouped hence the individual observations at the tail is lost. We simulate a long-tailed data using the estimated parameters of the Malayan butterfly data and compare the model fit of the NB, PIG, PGIG and generalized Sichel distributions. The mean and variance of the simulated data set are 10.6990 and 22.7307, respectively. The minimum value of the data set is 0, whilst the maximum is 224. A plot of the simulated data is given in Figure 6. The data set has a high zero count and a very long tail. During the model fitting, frequencies after k = 50 are grouped. The model fitting results are presented in Table 2. For presentation purposes, observations after 20 have been displayed as groups.

From the table, the generalized Sichel distribution gives the best fit to the data in terms of both chi-square goodness-of-fit statistic and AIC value. The generalized Sichel distribution fits well on not only the observations at the tail but also the zero counts.

Table 2: Fit of simulated data set

k	Observed frequency	Expected frequency			
		Generalized Sichel	NB	PIG	PGIG
0	1643	1643.24	1727.76	1409.66	1637.31
1	625	626.66	501.73	928.69	655.59
2	395	374.13	316.54	537.70	371.88
3	227	264.38	235.63	335.58	256.15
4	198	202.79	188.78	228.84	194.40
5	168	163.26	157.67	167.20	156.01
6	142	135.72	135.26	128.51	129.80
7	117	115.43	118.24	102.53	110.73
8	111	99.87	104.79	84.16	96.23
9	72	87.58	93.87	70.64	84.83
10	101	77.64	84.79	60.35	75.62
11	66	69.44	77.12	52.31	68.03
12	70	62.57	70.53	45.90	61.66
13	63	56.74	64.82	40.68	56.25
14	61	51.73	59.82	36.37	51.59
15	34	47.39	55.39	32.76	47.53
16	46	43.60	51.45	29.71	43.98
17	38	40.27	47.92	27.09	40.83
18	33	37.31	44.73	24.84	38.03
19	39	34.67	41.85	22.87	35.52
20	32	32.31	39.23	21.15	33.26
21 to 30	223	232.85	287.42	149.43	244.71
31 to 40	132	137.61	169.78	88.51	148.87
41 to 49	80	82.03	97.87	54.62	89.89
50 or more	284	280.77	227.02	319.91	271.30
Total	5000	5000	5000	5000	5000
Number of classes	51	51	51	51	51
Chi-square		54.4979	144.4311	447.4297	60.8280
Degree of freedom		46	48	48	47
p-value		0.1827	0.0000	0.0000	0.0848
AIC		27308.47	27399.05	27676.57	27314.35

5.2. *Real data*

We present in Table 3 the fit for Tröbliger's data which has been published in 1961 (as cited by Gathy and Lefèvr [3]) on the frequency of the number of claims. This data set has an 87% proportion of zeros. It has a mean of 0.1434 with standard deviation 0.4031, thus giving a dispersion index of 1.1328. For this data, the generalized Sichel distribution provides a good fit amongst the four mixed Poisson distributions based on the p-value of the chi-square goodness-of-fit test. We note that the PIG and PGIG (Sichel) distributions, which are simpler, also give a good fit for this over dispersed data set with small counts. In this case, fitting the generalized Sichel distribution eliminates the need for piece-wise treatment in the empirical modelling of the data.

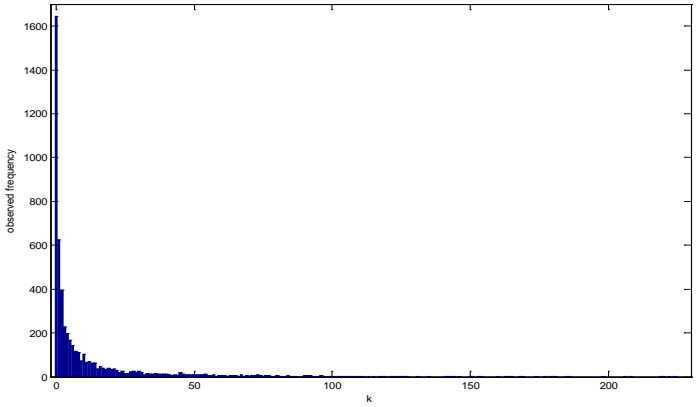


Fig. 6. A plot of the frequency distribution of the simulated data

Table 3: Fit of Trobliger's data

Number of Claims	Observed Frequency	Expected Frequency				
		Generalized Sichel	NB	PIG	PGIG	ZIP
0	20592	20593.33	20597.11	20597.30	20593.00	20592.01
1	2651	2647.44	2627.56	2633.40	2647.52	2624.01
2	287	291.39	313.16	303.63	291.69	332.94
3	41	38.52	36.45	38.37	38.50	28.16
4	7	6.52	4.19	5.34	6.50	1.79
5	0	1.35	0.48	0.80	1.35	0.09
6	1	0.32	0.05	0.13	0.32	0.00
7 or more	0	0.12	0.01	0.02	0.12	0.00
Total	23579	23579	23579	23579	23579	23579
Number of classes		8	8	8	8	8
Chi-square		3.0364	21.7883	8.6005	3.1599	286.5225
Degree of freedom		3	5	5	4	5
p-value		0.3860	0.0006	0.1261	0.5314	0.0000
AIC		20363.21	20364.57	20360.83	20361.25	20385.16

The fit for data on number of injuries sustained in 10,000 accidents in the United States in 2001 (as cited in Kadane *et al*. [13]) is presented in Table 4. It has a mean of 0.7073, standard deviation 1.0020 thus yielding a dispersion index of 1.4194. Its proportion of zeros is at 54%. The generalized Sichel distribution gives a significantly better fit on this data in terms of its AIC values and chi-square goodness-of-fit statistic, compared to the other mixed Poisson distributions considered here.

All of the data sets cited here do not fit well to the zero-inflated Poisson distribution (the last column of Tables 3 and 4). This is due to a poor fit on the counts at the right tail of the data although it fits the zero counts very well. We also attempted to fit the data sets with the zero-inflated negative binomial (ZINB) distribution. However, the iterative method used to estimate the ZINB parameters failed to converge. This convergence failure is a common problem with the ZINB and it was also noted by Famoye and Singh [2]. Moreover, we observe that for all of the data sets cited here, the negative binomial predicted a higher

frequency of zeros than which is observed, hence it may not be necessary to fit the ZINB model at all. As such, the generalized Sichel model can serve as an alternative to model zero-inflated count data.

5.2.1 Hypothesis testing

The hypothesis testing results for the Sichel test are presented in Table 5. At a significance level of $\alpha = 0.05$, the null hypothesis is not rejected for the Trobliger's data but is rejected for the accident injuries data. This conclusion corroborates the analysis of our model fitting results discussed in the preceding section.

Table 4: Accident Injuries Data

Injuries	Observed Frequency	Expected Frequency				
		Generalized Sichel	NB	PIG	PGIG	ZIP
0	5363	5389.30	5449.36	5446.28	5408.54	5363.00
1	3091	3025.28	2860.61	2900.77	2984.26	2837.12
2	1008	1059.84	1119.59	1086.47	1072.46	1295.85
3	348	332.94	388.34	372.34	345.44	394.59
4	105	111.96	126.06	126.44	114.36	90.11
5	46	43.16	39.23	43.60	41.47	16.46
6	19	18.89	11.86	15.35	16.81	2.51
7	9	9.04	3.51	5.52	7.59	0.33
8	7	4.56	1.02	2.02	3.76	0.04
9	2	2.37	0.29	0.75	2.02	0.00
10	1	1.25	0.08	0.28	1.15	0.00
11 or more	1	1.39	0.03	0.18	2.13	0.00
Total	10000	10000	10000	10000	10000	10000
Number of classes		12	12	12	12	9
Chi-square		6.9124	136.2590	47.9200	13.3075	3703.5929
Degree of freedom		7	9	9	8	6
p-value		0.4381	0.0000	0.0000	0.1017	0.0000
AIC		22909.56	22974.97	22937.64	22914.45	23231.76

Table 5 Score test results

Data Set	Score Test Statistic	p-value
Tröbliger's data (1961) on number of claims	0.0566	0.8120
Accident injuries data	4.2257	0.0398

6. Some Conclusion and Comments

The generalized Sichel distribution presented in this paper, is an extension of the Sichel (PGIG) distribution. It has one additional parameter, δ, which makes it more flexible for modelling count data sets having zero-inflation as well as overdispersion. The proposed distribution has an advantage of eliminating a piecewise treatment when fitting a data set to the NB, PIG and PGIG distributions since these are its special cases. We hope that the model, proposed in this paper, will provide a viable alternative to analyze count data sets which exhibit similar characteristics.

Acknowledgements

The authors would like to thank the reviewers for their insightful comments and suggestions which have vastly improved this paper. Y.C. Low and S. H. Ong are supported by the Fundamental Research Grant Scheme, Ministry of Higher Education, Malaysia [FP045-2015A].

References

[1] A.C. Atkinson and L. Yeh, Inference for Sichel's compound Poisson distribution, *Journal of the American Statistical Association* **77** (1982) 153-158.
[2] F. Famoye and K.P. Singh, Zero-inflated generalized Poisson regression model with an application to domestic violence data, *Journal of Data Science* **4** (2006) 117-130.
[3] M. Gathy and C. Lefèvre, On the Lagrangian Katz family of distributions as a claim frequency model, *Insurance: Mathematics and Economics* **47** (2010) 76-83.
[4] W.L. Goffe, G.D. Ferrier and J. Rogers, Global optimization of statistical functions with simulated annealing, *Journal of Econometrics* **60** (1994) 65-99.
[5] M. Greenwood and G.K. Yule, 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** (1920) 255-279.
[6] R.C. Gupta and S.H. Ong, A new generalization of the negative binomial distribution, *Computational Statistics & Data Analysis* **45** (2004) 287-300.
[7] R.C. Gupta and S.H. Ong, Analysis of long-tailed count data by Poisson mixtures, *Communications in Statistics—Theory and Methods* **34** (2005) 557-573.
[8] R.C. Gupta and W. Viles, Roller-coaster failure rates and mean residual life functions with application to the extended generalized inverse Gaussian model, *Probability in the Engineering and Informational Sciences* **25** (2011) 103-118.
[9] R.C. Gupta and W. Viles, Statistical inference for the extended generalized inverse Gaussian model, *Journal of Statistical Computation and Simulation* **82** (2011) 1855-1872.
[10] M.S. Holla, On a Poisson-inverse Gaussian distribution, *Metrika* **11** (1966) 115-121.
[11] N.L. Johnson, A.W. Kemp and S. Kotz, *Univariate Discrete Distributions,* 3rd edn. (John Wiley & Sons, New York, 2005).
[12] B. Jørgensen, *Statistical properties of the generalized inverse Gaussian distribution*, *Lecture Notes in Statistics,* Vol. 9, (Springer Verlag, New York, 1982)
[13] J.B. Kadane, R. Krishnan and G. Shmueli, A data disclosure policy for count data based on the COM-Poisson distribution, *Management Science* **52** (2006) 1610-1617.
[14] D. Karlis and E. Xekalaki, Mixed Poisson distributions, *International Statistical Review* **73** (2005) 35-58.
[15] J. Mullahy, Heterogeneity, excess zeros, and structure of count data model, *Journal of Applied Econometrics* **12** (1997) 337-350.
[16] A.K. Nikoloulopoulos and D. Karlis, On modeling count data: a comparison of some well-known discrete distributions, *Journal of Statistical Computation and Simulation* **78** (2008) 437-457.
[17] S.H. Ong, Computation of probabilities of a generalized log-series and related distributions, *Communications in Statistics-Theory and Methods* **24** (1995) 253-271.
[18] S.H. Ong and S. Muthaloo, A class of discrete distributions suited to fitting very long-tailed data, *Communications in Statistics-Simulation and Computation* **24** (1995) 929-945.
[19] J.K. Ord and G.A. Whitmore, The Poisson-inverse Gaussian distribution as a model for species abundance, *Communications in Statistics-Theory and Methods* **15** (1986) 853-871.
[20] P. Puig and J. Valero, Count data distributions: some characterizations with applications, *Journal of the American Statistical Association* **101** (2006) 332-340.
[21] R.A. Rigby, D.M. Stasinopoulos and C. Akantziliotou, A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution, *Computational Statistics & Data Analysis* **53** (2008) 381-393.
[22] M. Shaked, On mixtures from exponential families, *Journal of the Royal Statistical Society. Series B (Methodological)* **42** (1980) 192-198.
[23] G.Z. Stein, W. Zucchini and J.M. Juritz, Parameter estimation for the Sichel distribution and its multivariate extension, *Journal of the American Statistical Association* **82** (1987) 938-944.