ORIGINAL PAPER

Analysis of discrete data by Conway–Maxwell Poisson distribution

Ramesh C. Gupta · S. Z. Sim · S. H. Ong

Abstract In this paper, we further study the Conway–Maxwell Poisson distribution having one more parameter than the Poisson distribution and compare it with the Poisson distribution with respect to some stochastic orderings used in reliability theory. Likelihood ratio test and the score test are developed to test the importance of this additional parameter. Simulation studies are carried out to examine the performance of the two tests. Two examples are presented, one showing overdispersion and the other showing underdispersion, to illustrate the procedure. It is shown that the COM-Poisson model fits better than the generalized Poisson distribution.

Keywords Overdispersion · Underdispersion · Failure rate · Stochastic comparisons · Score test

1 Introduction

It is well known that the count data show overdispersion (underdispersion) relative to the Poisson distribution for which variance equals the mean, although the phenomenon of underdispersion is uncommon. The overdispersion means that the variance is greater than the mean of a Poisson random variable having the same mean. The overdispersion can be caused due to various situations, for instance, due to having the heterogeneity

R. C. Gupta (✉)
Department of Mathematics and Statistics, University of Maine, Orono, ME 04469-5752, USA
e-mail: rcgupta@maine.edu

S. Z. Sim · S. H. Ong
Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia
e-mail: shinzhusim@gmail.com

S. H. Ong
e-mail: ongsh@um.edu.my

in the data or due to having extra zeros than produced by the model. Mullahay (1997) has demonstrated that the unobserved heterogeneity commonly assumed to be the source of overdispersion in the count data models, 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 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 (NBD) distribution. Some generalizations of this concept have been studied by applying a generalized gamma distribution resulting in a generalized form of NBD, see Gupta and Ong (2004). Another choice of the distribution of θ is taken as the inverse Gaussian or the generalized inverse Gaussian giving rise to Sichel distribution, see Ord and Whitmore (1986) and Atkinson and Yeh (1982). It is a long-tailed distribution that is suitable for highly skewed data. In addition to the choices mentioned above, various other mixing distributions have been used in the literature; see Gupta and Ong (2005) for more examples and illustrations.

Another way to analyze such data sets is to model with more general models having more than one parameter. For example, Consul (1989) had proposed a generalized Poisson distribution having two parameters and Gupta et al. (2004) have considered the zero-inflated generalized Poisson distribution with three parameters, one of which pertains to the extra zeros in the data than predicted by the generalized Poisson distribution.

More recently, the Conway–Maxwell Poisson (CMP) distribution is revived by Shmueli et al. (2005). This distribution is a two parameter extension of the Poisson distribution that generalizes some well-known discrete distributions (i.e., the binomial and the negative binomial distributions). The CMP distribution was originally proposed to handle queueing systems with state-dependent service rates.

The CMP distribution generalizes the Poisson distribution, allowing for overdispersion or underdispersion. Its probability function is given by

$$P(X = x) = \frac{\theta^x}{(x!)^\nu} \cdot \frac{1}{Z(\theta, \nu)}, \quad x = 0, 1, 2, \ldots, \tag{1}$$

where

$$Z(\theta, \nu) = \sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^\nu}, \quad \theta > 0, \nu > 0.$$

1. When $\nu = 1$, $Z(\theta, \nu) = \exp(\theta)$, an ordinary Poisson distribution.
2. As $\nu \to \infty$, $Z(\theta, \nu) \to 1 + \theta$, and the CMP distribution approaches a Bernoulli distribution with $P(X = 1) = \theta/(1 + \theta)$.
3. When $\nu = 0$ and $0 < \theta < 1$, $Z(\theta, \nu)$ is a geometric sum given by

$$Z(\theta, \nu) = \sum_{j=0}^{\infty} \theta^j = \frac{1}{1 - \theta},$$

and the distribution is geometric with

$$P(X = x) = \theta^x (1 - \theta), \quad x = 0, 1, 2, \ldots.$$

4. When $\nu = 0$, and $\theta \geq 1$, $Z(\theta, \nu)$ does not converge and the distribution is undefined.

Shmueli et al. (2005) note that this distribution is appealing from a theoretical point of view as well because it belongs to the class of two parameter power series distribution. As a result, it allows for sufficient statistics and other elegant properties. Usually, many count tables correspond to the same sufficient statistics. Kadane et al. (2006a) investigated the number of solutions which give rise to the same sufficient statistics. Rodrigues et al. (2009) develop a flexible cure rate model assuming the number of competing causes of events of interest to follow CMP distribution. The Markov Chain Monte Carlo (MCMC) methods are used by Cancho et al. (2010) to develop Bayesian procedure for the CMP model. Bayesian analysis of the CMP distribution was also studied by Kadane et al. (2006b).

Although the CMP distribution is quite well researched, there are other aspects which have not been studied. Therefore, we shall develop some structural properties of (1) and study the monotonicity of its failure rate together with stochastic comparisons with the Poisson distribution in this paper. Since the additional parameter ν controls the overdispersion or underdispersion, we develop the likelihood ratio test and the score test to test the importance of this additional parameter. The model (1) has overdispersion if $\nu < 1$ and underdispersion if $\nu > 1$. This fact can be checked using Theorem 3 of Kokonendji et al. (2008). The organization of this paper is as follows: In Sect. 2, we present some structural properties including the moments and the probability generating function. Section 3 deals with the reliability properties and some stochastic comparisons. In Sect. 4, we present the computation of moments and score equations to test the hypothesis that $\nu = 1$. Test for equidispersion and simulation study of power for the score and likelihood ratio tests are developed in Sect. 5. Two examples are presented in Sect. 6, one having overdispersion ($\nu < 1$) and the other having underdispersion ($\nu > 1$), to illustrate the procedure. In both the examples, it is shown that the CMP model fits slightly better than the generalized Poisson distribution. Finally, some conclusions and comments are presented in Sect. 7. Thus, the purpose of this paper is to present another versatile model which takes care of the overdispersion or underdispersion compared to the Poisson distribution, in analyzing discrete data.

2 Structural properties of the CMP model

2.1 Moments

To study the moments of our model, we notice that CMP distribution is a special case of the modified power series distribution introduced by Gupta (1974, 1975), as follows:

$$P(X = x) = \frac{A(x))(g(\theta))^x}{f(\theta)}, x \in B,$$

where B is a subset of the set of non-negative integers, $A(x) > 0$; $f(\theta)$ and $g(\theta)$ are positive, finite and differentiable functions of θ.

In our case, $g(\theta) = \theta$, $f(\theta) = Z(\theta, v)$ and $A(x) = [(x!)^v]^{-1}$. It can be verified that

$$E(X) = \frac{g(\theta) f'(\theta)}{f(\theta) g'(\theta)}$$

$$= \theta \cdot \frac{\partial}{\partial \theta} [\ln Z(\theta, v)].$$

2.2 Recurrence relations between the moments

Let $\mu'_r = E(X^r)$, $\mu_r = E(X - \mu)^r$ and $E(X^{[r]}) = \mu^{[r]}$, where $\mu^{[1]} = \mu'_1 = \mu$ and $\mu^{[r]} = E(X(X-1)(X-2)\ldots(X-r+1))$. Gupta (1974) has shown that

$$\mu'_{r+1} = \frac{g(\theta)}{g'(\theta)} \frac{d\mu'_r}{d\theta} + \mu'_r \mu_1, \quad r = 1, 2, 3, \ldots, \tag{2}$$

$$\mu_{r+1} = \frac{g(\theta)}{g'(\theta)} \frac{d\mu_r}{d\theta} + r\mu_2 \mu_{r-1}, \quad r = 1, 2, 3, \ldots, \tag{3}$$

$$\mu^{[r+1]} = \frac{g(\theta)}{g'(\theta)} \frac{d\mu^{[r]}}{d\theta} + \mu^{[r]} \mu^{[1]} - r\mu^{[r]}, \quad r = 1, 2, 3, \ldots. \tag{4}$$

As a special case, for our model

$$\text{variance} = \mu_2 = \theta \frac{d\mu}{d\theta}$$

$$= \theta \cdot \frac{\partial^2}{\partial \theta^2} [\ln Z(\theta, v)].$$

The recurrence relations given above can be used to obtain higher moments of the model.

2.3 Mode

To find the mode, we notice that for $k \geq 2$,

$$\frac{P(X = k)}{P(X = k - 1)} = \frac{\theta}{k^v}.$$

This means that as $k \to \infty$, the above ratio $\to 0$. It has only one mode at the point $k = 1$.

2.4 Probability generating function

$$\psi(t) = E(t^X) = \frac{Z(\theta t, \nu)}{Z(\theta, \nu)}, 0 < t \leq 1.$$

The other generating functions viz. the characteristic function, the moment generating function, the factorial moment generating function and the cumulate generating function can be obtained using the probability generating function. Apart from the usefulness of the pgf in summarizing the probabilities or moments of the distribution, and in convergence results, the pgf has been applied in statistical inference; see, for instance, Rueda and O'Reilly (1999), Sim and Ong (2010) and Ng et al. (2013) and references therein.

3 Stochastic comparisons and reliability functions

3.1 Reliability functions

Let X be a discrete random variable whose mass is concentrated on the non-negative integers. Let $p(t) = P(X = t)$. Then, the failure rate $r(t)$, the survival function $S(t)$ and the mean residual life function (MRLF) $\mu(t)$ are given by

$$r(t) = \frac{P(X = t)}{P(X \geq t)} = \frac{p(t)}{\sum_{i \geq t} p(i)}, \quad t = 0, 1, 2, \ldots,$$

$$S(t) = P(X \geq t) = \sum_{i \geq t} p(i), \quad t = 0, 1, 2, \ldots,$$

and

$$\mu(t) = E(X - t | X \geq t)$$
$$= \frac{\sum_{i \geq t} i p(i)}{\sum_{i \geq t} p(i)} - t = \frac{\sum_{i > t} S(i)}{S(t)}, \quad t = 0, 1, 2, \ldots.$$

The above functions $p(t)$, $r(t)$, $S(t)$ and $\mu(t)$ are equivalent in the sense that knowing one, others can be determined. This can be seen by the following relations.

$$r(t) = \frac{p(t)}{\sum_{i \geq t} p(i)} = \frac{S(t) - S(t+1)}{S(t)} = 1 - \frac{\mu(t)}{1 + \mu(t+1)},$$

$$S(t) = \sum_{i \geq t} p(i) = \Pi_{i=0}^{t-1}(1 - r(i)) = \Pi_{i=0}^{t-1}\left[\frac{\mu(i)}{1 + \mu(i+1)}\right],$$

and

$$\mu(t) = \frac{\sum_{i \geq t} i p(i)}{\sum_{i \geq t} p(i)} - t = \sum_{i > t} \frac{S(i)}{S(t)} = \sum_{i > t} \Pi_{j=t}^{i-1}(1 - r(j)).$$

We now present the following definitions

Definition 1 A discrete life distribution has log-concave/log-convex probability mass function (pmf) if

$$\frac{p(t+2)p(t)}{[p(t+1)]^2} \le (\ge)1, \quad t \ge 0.$$

Definition 2 A discrete life distribution has increasing failure rate/decreasing failure rate (IFR/DFR) if the failure rate is non-decreasing/non-increasing.

The following result, due to Gupta et al. (1997), establishes a relation between log-concavity (log-convexity) of the pmf and IFR (DFR) distributions.

Theorem 1 *Let* $\eta(t) = 1 - P(X = t+1)/P(X = t)$ *and* $\Delta\eta(t) = \eta(t+1) - \eta(t) = [p(t+1)/p(t) - p(t+2)/p(t+1)]$. *Then,*

 (i) *If* $\Delta\eta(t) > 0$ *(log-concavity), then* $r(t)$ *is non-decreasing* (IFR).
 (ii) *If* $\Delta\eta(t) < 0$ *(log-convex), then* $r(t)$ *is non-increasing* (DFR).
(iii) *If* $\Delta\eta(t) = 0$ *for all* t, *then constant hazard rate.*

In addition to the above, the following implications hold (for proofs, see Kemp (2004) and Gupta et al (2008)).

IFR (DFR) $\Rightarrow$ DMRL (IMRL) where DMRL means decreasing mean residual life and IMRL means increasing mean residual life.

We now show that CMP distribution is log-concave.

Theorem 2 *The CMP distribution has a log-concave pmf.*

Proof In this case

$$\begin{aligned}
\Delta\eta(k) &= \frac{\theta}{(k+1)^\nu} - \frac{\theta}{(k+2)^\nu} \\
&= \theta\left[\frac{(k+2)^\nu - (k+1)^\nu}{(k+1)^\nu(k+2)^\nu}\right] > 0.
\end{aligned}$$

Thus, CMP distribution has a log-concave pmf and hence strongly unimodal, see Steutel (1985). □

Using the relationships established above, we can say that CMP distribution is IFR and DMRL.

Remark 1 Kokonendji et al. (2008) showed the log-concavity of the CMP distribution for $\nu \ge 1$ as a consequence of it being a weighted Poisson distribution where the weight function is log-concave only for $\nu \ge 1$ (refer to Theorem 5 in their paper).

3.2 Stochastic comparisons

We present some definitions for stochastic orderings in the case of discrete distributions.

Definition 3 Let X and Y be two discrete random variables with probability mass functions $f(x)$ and $g(x)$. Then,

1. X is said to be smaller than Y in the likelihood ratio order (denoted by $X \leq_{lr} Y$) if $g(x)/f(x)$ increases in x over the union of the supports of X and Y.
2. X is smaller than Y in the hazard rate order (denoted by $X \leq_{hr} Y$) if $r_X(n) \geq r_Y(n)$ for all n.
3. X is smaller than Y in the mean residual life order (denoted by $X \leq_{MRL} Y$) if $\mu_X(n) \leq \mu_Y(n)$ for all n.

See Shaked and Shanthikumar (2007) for more details and explanations.

The following theorem establishes the relationships between the above orderings.

Theorem 3 *Suppose X and Y are two discrete random variables. Then, $X \leq_{lr} Y \Rightarrow X \leq_{hr} Y \Rightarrow X \leq_{MRL} Y$*

To compare the CMP distribution with the Poisson distribution, we let Y denote the Poisson variable and X denote the CMP variable. Then,

$$\frac{P(Y = n)}{P(X = n)} = \frac{(n!)^{v-1} Z(\theta, v)}{e^\theta},$$

which is increasing in n. Thus, $X \leq_{lr} Y$. Using the above Theorem, we conclude that $X \leq_{hr} Y$ and $X \leq_{MRL} Y$

4 Computation of moments and score equations

4.1 Computation of moments

The infinite sum $Z(\theta, v)$ is the normalization constant in the CMP probability mass function given by (1). The computation of $Z(\theta, v)$ for small v can be difficult. To overcome this difficulty, Minka et al. (2003) suggested a numerical approximation by truncating the series, that is,

$$Z(\theta, v) \approx \sum_{j=0}^{k} \frac{\theta^j}{(j!)^v}.$$

Minka et al. (2003) also derived an asymptotic approximation of $Z(\theta, v)$ given as

$$Z(\theta, v) = \frac{\exp(v\theta^{1/v})}{\theta^{\frac{v-1}{2v}} (2\pi)^{\frac{v-1}{2}} \sqrt{v}} (1 + O(\theta^{-\frac{1}{v}})). \tag{5}$$

The approximation expression for the moments, obtained from the asymptotic approximation of (5), is given by

$$E[X] = \theta \frac{\partial \log Z(\theta, \nu)}{\partial \theta} \approx \theta^{\frac{1}{\nu}} - \frac{\nu - 1}{2\nu}, \tag{6}$$

$$E[\log(X!)] = -\frac{\partial \log Z(\theta, \nu)}{\partial \nu} \approx \frac{1}{2\nu^2} \log \theta + \theta^{\frac{1}{\nu}} \left(\frac{\log \theta}{\nu} - 1 \right). \tag{7}$$

According to Minka et al. (2003), these approximations are good for $\nu \leq 1$ or $\theta > 10^\nu$, and they suggested a truncation approach to get more precise values. The accuracy of (7) can be improved by adding extra terms to get

$$E[\log(X!)] \approx \frac{1}{2\nu^2} \log \theta + \theta^{\frac{1}{\nu}} \left(\frac{\log \theta}{\nu} - 1 \right) + \frac{1}{2\nu} + \frac{\log 2\pi}{2}. \tag{8}$$

As no comparison of accuracy has been done, we examine the accuracy in the computation of (5), (7) and (8), with and without using the asymptotic approximation of $Z(\theta, \nu)$. This is presented in Table 1. The differences between the two quantities are given in italics. The partial derivatives of the infinite series $Z(\theta, \nu)$ are as follows:

$$\theta \frac{\partial \log Z(\theta, \nu)}{\partial \theta} = \theta \frac{\sum_{j=1}^{\infty} \frac{j\theta^{j-1}}{(j!)^\nu}}{\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^\nu}}, \tag{9}$$

$$-\frac{\partial \log Z(\theta, \nu)}{\partial \nu} = -\frac{\sum_{j=2}^{\infty} \frac{-\theta^j}{(j!)^\nu} \ln(j!)}{\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^\nu}}. \tag{10}$$

Table 1 illustrates the discrepancy between the asymptotic approximation and infinite series for $Z(\theta, \nu)$. For $E[X]$, the difference decreases when ν increases for any θ. When $\nu = 1$, Eqs. (5) and (8) achieve the same value. On the other hand, for $E[\log(X!)]$ Eq. (8) showed great improvement in accuracy over Eq. (7). Since the extra two terms in (8) are easily computed, the use of (8) over (7) is recommended.

4.2 Score equations

The log-likelihood function is given by

$$L = \sum_{x=0}^{\infty} \pi_x \ln P(X = x),$$

where π_x = observed frequency and $\ln P(X = x) = x \ln \theta - \nu \ln(x!) - \ln \sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^\nu}$.

Table 1 Comparison between moments obtained from asymptotic approximation (Asym) and infinite series (InfS) of $Z(\theta, \nu)$

ν	0.7	0.8	0.9	1.0
10^ν	5.0119	6.3096	7.9433	10.0000
$\theta = 5$				
$E(X)$				
Asym = Eq. (6)	10.1805	7.6017	6.0346	5.0000
InfS = Eq. (9)	10.1858	7.6058	6.0368	5.0000
(Asym-InfS)	−0.0054	−0.0041	−0.0022	0.0000
$E[\log(X!)]$				
Asym = Eq. (7)	14.5903	8.8223	5.7066	3.8519
Asym* = Eq. (8)	16.2235	10.3663	7.1811	5.2708
InfS = Eq. (10)	16.2213	10.3579	7.1669	5.2516
(Asym-InfS)	−1.6310	−1.5356	−1.4603	−1.3997
(Asym*-InfS)	0.0022	0.0084	0.0142	0.0192
$\theta = 10$				
$E[X]$				
Asym	27.0412	17.9078	12.9711	10.0000
InfS	27.0430	17.9092	12.9719	10.0000
(Asym-InfS)	−0.0018	−0.0014	−0.0008	0.0000
$E[\log(X!)]$				
Asym	63.7674	35.1991	21.5492	14.1771
Asym*	65.4006	36.7430	23.0237	15.5961
InfS	65.4014	36.7408	23.0182	15.5873
(Asym-InfS)	−1.6340	−1.5417	−1.4690	−1.4102
(Asym*-InfS)	−0.0008	0.0022	0.0055	0.0088
$\theta = 20$				
$E[X]$				
Asym	72.4271	42.4199	27.9546	20.0000
InfS	72.4277	42.4204	27.9549	20.0000
(Asym-InfS)	−0.0006	−0.0005	−0.0003	0.0000
$E[\log(X!)]$				
Asym	239.8873	118.4256	66.8146	41.4125
Asym*	241.5205	119.9696	68.2891	42.8315
InfS	241.5214	119.9691	68.2869	42.8272
(Asym-InfS)	−1.6341	−1.5435	−1.4723	−1.4147
(Asym*-InfS)	−0.0009	0.0005	0.0022	0.0043

The likelihood score equations of θ and ν are found to be

$$\frac{\partial \ln P(X = x)}{\partial \theta} = \frac{x}{\theta} - \frac{\sum_{j=1}^{\infty} \frac{j\theta^{j-1}}{(j!)^\nu}}{\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^\nu}}, \tag{11}$$

$$\frac{\partial \ln P(X = x)}{\partial \nu} = -\ln(x!) - \frac{\sum_{j=2}^{\infty} \frac{-\theta^j}{(j!)^{\nu}} \ln(j!)}{\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^{\nu}}}. \tag{12}$$

The infinite sum under the CMP distribution, $Z(\theta, \nu)$ is calculated recursively with double precision and truncation of the series, that is, $Z(\theta, \nu) \leq 1 \times 10^{50}$. The recursive computation adopts the approach given in Lee et al. (2001). The simulated annealing (SA) algorithm (Metropolis et al. 1953) is used in the numerical optimization to obtain the maximum likelihood estimates required in the log-likelihood ratio test.

5 Test for dispersion

The CMP distribution reduces to the ordinary Poisson distribution with parameter θ when $\nu = 1$. Since the parameter controls the under, equi and overdispersion of the distribution, we derive the Rao's score test and the likelihood ratio test (LRT) to test the null hypothesis $H_0: \nu = 1$ against the alternative hypothesis $H_1 : \nu \neq 1$. The study of the statistical power of these two tests is developed and presented in this section. The brief introduction to Rao's score test and LRT is given in Appendix.

In the simulation study of the power of the score and likelihood ratio tests, we consider the significance level α at 5 and 10 % and samples of $N = 100$ (small), 500 (moderate) and 1,000 (large). The effect size ($|\nu - 1|$) which serves as the index of departure from the null hypothesis, is set at 0.2, 0.5, 1.0, 3.0 and 4.0. It is found that 1,000 simulation runs give results of sufficient accuracy.

The results of the simulation study are presented in Tables 3, 4, 5 for $\theta = 5$ (short-tailed data), 10 (moderate-tailed data) and 20 (long-tailed data). Furthermore, the estimated empirical level for $\theta = 1, 5, 7, 10$ and 20 is studied and the results are presented in Table 6. In the tables, the power is number of rejection divided by number of repetitions.

The results in Tables 2, 3 and 4 are similar. The powers of the Rao's score test and LRT are very close to each other when the sample size N is large enough ($N \geq 500$), for overdispersion ($\nu < 1$) and underdispersion ($\nu > 1$). For the case of equidispersion (Table 5, $\nu = 1$), both tests have estimated empirical levels close to the specified significance levels of 5 and 10 %.

The statistical power as shown in Tables 2, 3, 4 and 5 greatly depends upon the sample size and the effect size. As expected, the larger the sample size, the higher is the statistical power and the power increases with the deviation from $\nu = 1$. For overdispersion data, when the effect size is 0.5, a 100 % detection is achieved even for a small sample size of 100. When the sample size increases ($N \geq 500$), an effect size of 0.2 can be easily detected with a power close to 1. However, for underdispersion, a higher value of effect size and sample sizes is needed to achieve 100% detection. When $N = 100$, an effect size larger than 1.0 is required to detect the deviation from $\nu = 1$.

Table 2 Simulated power of Rao's score and LR tests ($\theta = 5$)

θ			5	5	5	5	5	5	5
ν			0.5	0.8	1.2	1.5	2	4	5
Effect size			0.5	0.2	0.2	0.5	1.0	3.0	4.0
N	α	Method				Power			
100	0.05	Score	1.00	0.35	0.17	0.64	0.98	1.00	1.00
		LR	1.00	0.30	0.21	0.70	0.99	1.00	1.00
	0.10	Score	1.00	0.45	0.28	0.76	1.00	1.00	1.00
		LR	1.00	0.42	0.33	0.79	1.00	1.00	1.00
500	0.05	Score	1.00	0.92	0.69	1.00	1.00	1.00	1.00
		LR	1.00	0.91	0.71	1.00	1.00	1.00	1.00
	0.10	Score	1.00	0.94	0.81	1.00	1.00	1.00	1.00
		LR	1.00	0.94	0.83	1.00	1.00	1.00	1.00
1000	0.05	Score	1.00	1.00	0.95	1.00	1.00	1.00	1.00
		LR	1.00	1.00	0.95	1.00	1.00	1.00	1.00
	0.10	Score	1.00	1.00	0.97	1.00	1.00	1.00	1.00
		LR	1.00	1.00	0.97	1.00	1.00	1.00	1.00

Table 3 Simulated power of Rao's score and LR tests ($\theta = 10$)

θ			10	10	10	10	10	10	10
ν			0.5	0.8	1.2	1.5	2	4	5
Effect size			0.5	0.2	0.2	0.5	1.0	3.0	4.0
N	α	Method				Power			
100	0.05	Score	1.00	0.38	0.18	0.72	0.99	1.00	1.00
		LR	1.00	0.33	0.24	0.77	1.00	1.00	1.00
	0.10	Score	1.00	0.48	0.31	0.83	1.00	1.00	1.00
		LR	1.00	0.44	0.35	0.86	1.00	1.00	1.00
500	0.05	Score	1.00	0.93	0.76	1.00	1.00	1.00	1.00
		LR	1.00	0.92	0.77	1.00	1.00	1.00	1.00
	0.10	Score	1.00	0.96	0.86	1.00	1.00	1.00	1.00
		LR	1.00	0.95	0.86	1.00	1.00	1.00	1.00
1000	0.05	Score	1.00	1.00	0.97	1.00	1.00	1.00	1.00
		LR	1.00	1.00	0.97	1.00	1.00	1.00	1.00
	0.10	Score	1.00	1.00	0.99	1.00	1.00	1.00	1.00
		LR	1.00	1.00	0.99	1.00	1.00	1.00	1.00

6 Application

To illustrate the application of the CMP distribution to data modeling, we consider the goodness-of-fit to data sets exhibiting under and overdispersion.

Table 4 Simulated power of Rao's score and LR tests ($\theta=20$)

θ			20	20	20	20	20	20	20
ν			0.5	0.8	1.2	1.5	2	4	5
Effect size			0.5	0.2	0.2	0.5	1.0	3.0	4.0
N	α	Method				Power			
100	0.05	Score	1.00	0.39	0.19	0.73	1.00	1.00	1.00
		LR	1.00	0.71	0.25	0.79	1.00	1.00	1.00
	0.10	Score	1.00	0.48	0.32	0.85	1.00	1.00	1.00
		LR	1.00	0.75	0.36	0.88	1.00	1.00	1.00
500	0.05	Score	1.00	0.94	0.77	1.00	1.00	1.00	1.00
		LR	1.00	1.00	0.79	1.00	1.00	1.00	1.00
	0.10	Score	1.00	0.96	0.87	1.00	1.00	1.00	1.00
		LR	1.00	1.00	0.88	1.00	1.00	1.00	1.00
1000	0.05	Score	1.00	1.00	0.97	1.00	1.00	1.00	1.00
		LR	1.00	0.99	0.97	1.00	1.00	1.00	1.00
	0.10	Score	1.00	1.00	0.99	1.00	1.00	1.00	1.00
		LR	1.00	0.99	0.99	1.00	1.00	1.00	1.00

Table 5 Estimated Empirical level of Rao's score and LR tests (effect size = 0)

θ			1	5	7	10	20	30
ν			1	1	1	1	1	1
Number of repetition			1000	1000	1000	1000	1000	1000
N	α	Method			Power			
100	0.05	Score	0.04	0.06	0.05	0.06	0.06	0.06
		LR	0.05	0.07	0.06	0.06	0.05	0.04
	0.10	Score	0.10	0.11	0.11	0.11	0.12	0.12
		LR	0.10	0.11	0.11	0.12	0.10	0.10
500	0.05	Score	0.05	0.05	0.05	0.05	0.05	0.06
		LR	0.04	0.06	0.05	0.06	0.05	0.04
	0.10	Score	0.10	0.11	0.10	0.11	0.12	0.13
		LR	0.09	0.11	0.10	0.11	0.12	0.11
1000	0.05	Score	0.04	0.05	0.06	0.06	0.06	0.06
		LR	0.04	0.05	0.05	0.05	0.06	0.06
	0.10	Score	0.09	0.09	0.10	0.10	0.10	0.10
		LR	0.09	0.09	0.10	0.10	0.10	0.97

Example 1 (death notice data of London times)

The data consist of the number of death notices of women 80 years of age and older, appearing in the London Times on each day for three consecutive years. Hasselblad (1969) analyzed this data by mixture of two Poisson distributions. The counts are given below

Observed	0	1	2	3	4	5	6	7	8	9
Frequency	162	267	271	185	111	61	27	8	3	1

The above data set was also analyzed by Gupta et al. (1996) by adjusting Poisson distribution for extra zeros. We have analyzed this data by CMP distribution and compared with the generalized Poisson distribution of Consul and Jain (1973). The results are given in the following Table 6. As can be seen, the index of dispersion is 1.21 and the estimated value of ν using CMP distribution is 0.75 showing overdispersion relative to the Poisson distribution. The fit by the CMP distribution is much better than the generalized Poisson distribution in terms of Chi-square values. Using the CMP model, the hypothesis is rejected both by the LR test and the score test.

Example 2

(Consul 1989, p. 131, Table 5.13) considers the number of discentrics per cell for 8 different doses and fitted GPD. We consider the following data (Dose 1200).

Observed	0	1	2	3	4	5	6	7	8	9
Frequency	0	4	5	23	24	38	21	10	1	4

Table 6 Frequency distribution of death notice data of London times.

Observed death count	Observed frequency	Expected frequency	
		GPD	CMP
0	162	155.03	161.62
1	267	276.21	268.33
2	271	269.53	264.93
3	185	190.93	192.99
4	111	109.89	113.31
5	61	54.55	56.28
6	27	24.23	24.38
7	8	9.87	9.41
8	3	3.75	3.28
9	1	2.02	1.47
Total	1096		
χ^2		2.92	1.59
		$\hat{\theta} = 1.96$	$\hat{\lambda} = 1.66$
		$\hat{\lambda} = 0.09$	$\hat{\nu} = 0.75$
ID		1.21	
df		7	7
p value		0.89	0.98
LR test		Reject H_0 with 22.51	
Score test		Reject H_0 with 24.90	

Table 7 Frequency distribution of dicentrics for dose 1200

Number of dicentrics per cell	Observed frequency	Expected frequency	
	Dose 1200	GPD	CMP
0	0	0.27	0.09
1	4	2.3	1.69
2	5	8.75	8.61
3	23	19.71	21.09
4	24	29.26	30.65
5	38	30.15	29.71
6	21	22.1	20.7
7	10	11.64	10.9
8	1	4.39	4.51
9	4	1.42	2.04
Total	130		
χ^2		14.29	13.40
		$\hat{\theta} = 6.17$	$\hat{\lambda} = 17.95$
		$\hat{\lambda} = -0.32$	$\hat{\nu} = 1.81$
ID	0.57		
df		7	7
p value		0.05	0.06
LR test		Reject H_0 with 17.05	
Score test		Reject H_0 with 16.06	

We have analyzed this data by the CMP and GPD distributions. The results are given in Table 7. As can be seen, the index of dispersion is 0.57 showing the underdispersion and the estimated value of for the CMP distribution is 1.813. Based upon the Chi-square values, the fit by the CMP distribution is better than the GPD distribution. Using the CMP model, the hypothesis is rejected both by the LR test and the score test.

7 Conclusion and comments

This paper deals with the problem of overdispersion (underdispersion) relative to Poisson distribution in analyzing discrete data. There are various ways of modeling such data sets including models having more than one parameter (for example, generalized Poisson distribution) or mixture models (negative binomial and their generalized forms). In this paper, we have presented another alternative, a CMP distribution which has one additional parameter ν, allowing for overdispersion ($\nu < 1$) and underdispersion ($\nu > 1$). Likelihood ratio test and score test are developed for testing $H: \nu = 1$. Simulation studies are carried out to examine the performance of these tests. Two examples are presented, one showing overdispersion and the other showing underdispersion. We hope that our investigation will be helpful to researchers modeling discrete data.

Appendix

The Rao's score test statistic is given by

$$T = VI^{-1}V^T,$$

where V is the score vector and I is the information matrix. The score vector and the information matrix, obtained by evaluating the derivative of the log-likelihood function, $\ln L$ under the null hypothesis, are given by

$$V = \left(\frac{\partial \ln L}{\partial v}, \frac{\partial \ln L}{\partial \theta} \right),$$

$$I = - \begin{bmatrix} E\left[\frac{\partial^2 \ln L}{\partial v^2}\right] & E\left[\frac{\partial^2 \ln L}{\partial \theta \partial v}\right] \\ E\left[\frac{\partial^2 \ln L}{\partial \theta \partial v}\right] & E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right] \end{bmatrix}.$$

The likelihood score functions, $\frac{\partial \ln L}{\partial \theta}$ and $\frac{\partial \ln L}{\partial v}$ are given by (11) and (12). The second-order partial derivatives of the probability mass function are

$$\frac{\partial^2 \ln P(X=x)}{\partial \theta^2} = -\frac{x}{\theta^2} + \frac{\left(\sum_{j=1}^{\infty} \frac{j\theta^{j-1}}{(j!)^v}\right)^2}{\left(\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^v}\right)^2} - \frac{\left(\sum_{j=2}^{\infty} \frac{(j-1)j\theta^{j-2}}{(j!)^v}\right)}{\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^v}},$$

$$\frac{\partial^2 \ln P(X=x)}{\partial v^2} = \frac{\left(\sum_{j=2}^{\infty} \frac{-\theta^j}{(j!)^v} \ln(j!)\right)^2 - \left(\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^v}\right) \sum_{j=2}^{\infty} \frac{\theta^j}{(j!)^v} \ln(j!)^2}{\left(\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^v}\right)^2},$$

$$\frac{\partial^2 \ln P(X=x)}{\partial v \partial \theta} = \frac{1}{\left(\sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^v}\right)^2} \left\{ \sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^v} \left(\sum_{j=2}^{\infty} \frac{j\theta^{j-1}}{(j!)^v} \ln(j!) \right) \right.$$
$$\left. - \left(\sum_{j=1}^{\infty} \frac{j\theta^{j-1}}{(j!)^v} \right) \left(\sum_{j=2}^{\infty} \frac{\theta^j}{(j!)^v} \ln(j!) \right) \right\}.$$

The LRT requires estimation of the models under the null and alternative hypotheses. By comparing the log-likelihood scores under the null and alternative hypotheses,

the LRT gives evidence whether the deviation of one model from the other is statistically significant. The LR test statistic is

$$\text{LR} = -2\ln\left(\frac{L(\hat{\beta}^*; x)}{L(\hat{\beta}; x)}\right),$$

where $\hat{\beta}^*$ is the restricted ML estimator (null) and $\hat{\beta}$ is the unrestricted ML estimator (alternative).

References

Atkinson, A.C., Yeh, L.: Inference for Sichel's compound Poisson distribution. J. Am. Stat. Assoc. **77**, 153–158 (1982)

Cancho, V.G., Castro, M., Rodrigues, J.: A Bayesian analysis of the Conway–Maxwell–Poisson cure rate model. Stat. Pap. **53**(1), 165–176 (2012)

Consul, P.C., Jain, G.C.: On some interesting properties of the generalized Poisson distribution. Biometrische Z **15**, 495–500 (1973)

Consul, P.C.: Generalized Poisson Distributions. Marcel Dekker, New York (1989)

Gupta, P.L., Gupta, R.C., Tripathi, R.C.: On the monotonic properties of discrete failure rates. J. Stat. Plan. Inference **65**, 255–268 (1997)

Gupta, P.L., Gupta, R.C., Ong, S.H., Srivastava, H.M.: A study of Hurwitz–Lerch-Zeta distribution with application. Appl. Math. Comput. **196**(2), 521–531 (2008)

Gupta, P.L., Gupta, R.C., Tripathi, R.C.: Score test for zero inflated generalized Poisson regression model. Commun. Stat. Theory Methods **33**(1), 47–64 (2004)

Gupta, P.L., Gupta, R.C., Tripathi, R.C.: Analysis of zero adjusted count data. Comput. Stat. Data Anal. **23**, 207–218 (1996)

Gupta, R.C.: Modified power series distributions and some of its applications. Sankhya Ser. B **36**, 288–298 (1974)

Gupta, R.C.: Maximum likelihood estimation of a modified power series distribution and some of its applications. Commun. Stat. Theory Methods **4**, 689–697 (1975)

Gupta, R.C., Ong, S.H.: A new generalization of the negative binomial distribution. Comput. Stat. Data Anal. **45**(2), 287–300 (2004)

Gupta, R.C., Ong, S.H.: Analyses of long-tailed count data by Poisson mixtures. Commun. Stat. Theory Methods **34**(3), 557–573 (2005)

Hasselblad, V.: Estimation of finite mixtures of distributions from the exponential family. J. Am. Stat. Assoc. **64**, 1459–1471 (1969)

Kadane, J.B., Krishan, R., Shmueli, G.: A data disclosure policy for count data based on the COM-Poisson distribution. Manag. Sci. **52**(10), 1610–1617 (2006a)

Kadane, J.B., Shmeuli, G., Munka, T.P., Borle, S., Boatwright, P.: Conjugate analysis of the Conway–Maxwell–Poisson distribution. Bayesian Anal. **2**(2), 363–374 (2006b)

Kokonendji, C.C., Mizere, D., Balakrishnan, N.: Connections of the Poisson weight function to overdispersion and underdispersion. J. Stat. Plan. Inference **138**, 1287–1296 (2008)

Kemp, A.W.: Classes of discrete lifetime distributions. Commun. Stat. Theory Methods **33**, 3069–3093 (2004)

Lee, P.A., Ong, S.H., Srivastava, H.M.: Some integrals of products of Laguerre polynomials. Int. J. Comput. Math. **78**, 303–321 (2001)

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)

Minka, T.P., Shmueli, G, Kadane, J.B., Borle, S., Boatwright, P.: Computing with the COM-Poisson distribution. Technical Report #776, Department of Statistics, Carnegie Mellon University. http://www.stat.cmu.edu/tr/tr776/tr776.html (2003)

Mullahay, J.: Heterogeneity, excess zeros, and the structure of count data models. J. Appl. Econ. **12**, 337–350 (1997)

Ng, C.M., Ong, S.H., Srivastava, H.M.: Parameter estimation by Hellinger type distance for multivariate distributions based on probability generating functions. Appl. Math. Model. **37**, 7374–7385 (2013)

Ord, J.K., Whitmore, G.A.: The Poisson-inverse gaussian distribution as a model for species abundance. Commun. Stat. Theory Methods **15**, 853–871 (1986)

Rodrigues, J., Castro, M., Cancho, V.G., Balakrishnan, N.: COM Poisson cure rate survival models and an application to a cutaneous melanoma data. J. Stat. Plan. Inference **139**, 3605–3611 (2009)

Rueda, R., O'Reilly, F.: Tests of fit for discrete distributions based on the probability generating function. Commun. Stat. Simul. Comput. **28**(1), 259–274 (1999)

Shaked, M., Shanthikumar, J.G.: Stochastic Orders. Springer, Berlin, Heidelberg, New York (2007)

Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright, S.: A useful distribution for fitting discrete data-revival of the Conway–Maxwell–Poisson distribution. Appl. Stat. **54**(Part 1), 127–142 (2005)

Sim, S.Z., Ong, S.H.: Parameter estimation for discrete distributions by generalized Hellinger-type divergence based on probability generating function. Commun. Stat. Simul. Comput. **39**, 305–314 (2010)

Steutel, F.W.: Logconcave and log-convex distributions. In: Kotz, S., Johonson, N.L., Read, C.B. (eds.) Encyclopedia of Statistical Sciences, vol. 5, pp. 116–117. Wiley, New york (1985)