# Computing with the COM-Poisson distribution

Thomas P. Minka Galit Shmueli Joseph B. Kadane Sharad Borle
Peter Boatwright *

**Abstract**

The Conway-Maxwell-Poisson (COM-Poisson) is a generalization of the Poisson distribution which can model both under-dispersed and over-dispersed data. However, the distribution, moments, and MLE cannot be computed in closed form. This paper describes computational schemes and handy approximations for the COM-Poisson.

## 1 The COM-Poisson probability function

Denote by $X$ a random variable from the Poisson($\lambda$) distribution, with the distribution function given by

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ for } x = 0, 1, 2, \ldots \tag{1}$$

One of the properties of the Poisson distribution is that the ratio of consecutive probabilities is linear in $x$, or

$$\frac{P(X = x-1)}{P(X = x)} = \frac{x}{\lambda}. \tag{2}$$

In some applications, this ratio may not decrease linearly in $x$, i.e. the distribution may have a thicker or thinner tail than the Poisson. Suppose instead of equation (2), we set

$$\frac{P(X = x-1)}{P(X = x)} = \frac{x^\nu}{\lambda} \tag{3}$$

for a random variable $X$. The resulting distribution for which equation (3) holds, called the Conway-Maxwell-Poisson distribution (Wimmer and Altmann, 1999), is given by

*Thomas P. Minka is Visiting Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213. Galit Shmueli is Visiting Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213. Joseph B. Kadane is Leonard J. Savage University Professor of Statistics and Social Sciences, Carnegie Mellon University, Pittsburgh, PA 15213. His research is supported by National Science Foundation grant DMS-9801401. Sharad Borle is a Ph.D. student, GSIA, Carnegie Mellon University, Pittsburgh, PA 15213. Peter Boatwright is Assistant Professor, GSIA, Carnegie Mellon University, Pittsburgh, PA 15213.

$$P(X = x \mid \lambda, \nu) = \frac{\lambda^x}{(x!)^\nu} \frac{1}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}}, \text{ for } x = 0, 1, 2, \ldots \tag{4}$$

for $\lambda > 0$ and $\nu \geq 0$ (see exception below). This satisfies the conditions for a probability function.

It can be seen that the series $\frac{\lambda^j}{(j!)^\nu}$ converges for any $\lambda > 0, \nu > 0$, since the ratio of two subsequent terms of the series:

$$\frac{\lambda}{j^\nu} \tag{5}$$

tends to 0 as $j \to \infty$.

From here on, we denote the infinite sum in the denominator by:

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} \tag{6}$$

The COM-Poisson is a generalization of some well known discrete distributions:

- For $\nu = 1$, $Z(\lambda, \nu) = e^\lambda$, and this distribution is simply the ordinary Poisson($\lambda$).
- As $\nu \to \infty$ , $Z(\lambda, \nu) \to 1 + \lambda$, and the distribution approaches a Bernoulli distribution with $P(X = 1) = \frac{\lambda}{1+\lambda}$.
- When $\nu = 0$ and $\lambda < 1$, $Z(\lambda, \nu)$ is a geometric sum:

  $$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \lambda^j = \frac{1}{1 - \lambda}, \tag{7}$$

  and the distribution itself is geometric:

  $$P(X = x \mid \lambda, \nu) = \lambda^x (1 - \lambda) \text{ for } x = 0, 1, 2, \ldots \tag{8}$$

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

The COM-Poisson can thus be thought of as a continuous bridge between the Geometric (that counts failures until the first success) ($\nu = 0$), the Poisson ($\nu = 1$), and the Bernoulli ($\nu = \infty$) distributions. Values of $\nu$ less than one correspond to flatter successive ratios (3) than the Poisson's (2) and hence to longer tails.

The sum of $n$ independent COM-Poisson random variables is also a continuous bridge between three well-known distributions:

- For $\nu = 0$ and $\lambda < 1$, this reduces to the sum of geometric variables, namely, the negative binomial distribution with parameters $n$ and $1 - \lambda$.
- For $\nu = 1$ the sum has a Poisson distribution with parameter $n\lambda$.
- For $\nu = \infty$ the distribution of the sum is Binomial with parameters $n$ and $\lambda/(1 + \lambda)$.

## 2 Generating data

To sample an integer value from the COM-Poisson distribution, the inversion method is particularly simple. The COM-Poisson probabilities are summed up starting from $P(X = 0)$, until this sum exceeds the value of a simulated Uniform(0,1) variable. $X$ is then an observation from the COM-Poisson distribution. The probabilities can be computed reliably by the recursion

$$P(X+1) = P(X)\frac{\lambda}{(X+1)^\nu} \tag{9}$$

The catch in this algorithm is that you need to compute the initial probability, $P(X = 0) = Z(\lambda,\nu)^{-1}$, which is not available in closed form. The next section describes methods to approximate it.

## 3 Evaluating the normalizing constant

The compute COM-Poisson probabilities, you have to be able to compute the normalizing constant $Z(\lambda,\nu)$ given in (6). This section describes efficient ways to do so.

### 3.1 An asymptotic approximation of $Z(\lambda,\nu)$

As a first step in understanding $Z$, it helps to consider its asymptotic behavior. Defining $i = \sqrt{-1}$, we have the identity

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{e^{ia}} e^{-iaj} da \quad = \quad \frac{1}{j!} \tag{10}$$

which means

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{e^{ia}} Z(\lambda e^{-ia},\nu) da \quad = \quad \sum_{j=0}^{\infty} \frac{\lambda^j}{j!^\nu}\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{e^{ia}} e^{-iaj} da \quad = \quad Z(\lambda,\nu+1) \tag{11}$$

This represents $Z(\lambda,\nu+1)$ as an integral over $Z(\lambda,\nu)$. By applying this formula repeatedly starting from $Z(\lambda,1) = \exp(\lambda)$, we obtain a representation of $Z(\lambda,\nu)$ for integer $\nu > 0$, as a multiple integral:

$$Z(\lambda,\nu) \quad = \quad \frac{1}{(2\pi)^{\nu-1}}\int\cdots\int \exp\left(\sum_{k=1}^{\nu-1}\exp(ia_k) + \lambda\exp(-\sum_{k=1}^{\nu-1} ia_k)\right) da_1\cdots da_{\nu-1} \tag{12}$$

The behavior of this integral for large $\lambda$ can be determined by making the change of variable $ia_j = ib_j + \frac{1}{\nu}\log(\lambda)$ and applying Laplace's method (Bleistein and Handelsman, 1986, Ch 5.1). The result is:

$$Z(\lambda,\nu) \quad = \quad \frac{\exp(\nu\lambda^{1/\nu})}{\lambda^{\frac{\nu-1}{2\nu}}(2\pi)^{(\nu-1)/2}\sqrt{\nu}}(1+O(\lambda^{-1/\nu})) \tag{13}$$

This asymptotic formula has been derived for integer $\nu$, but numerical studies suggest that it holds for all real $\nu > 0$. It is accurate when $\lambda > 10^{\nu}$. This formula will be used later to understand moments of the COM-Poisson. The main message from this formula is that $Z$ grows rapidly as $\lambda$ increases or $\nu$ decreases.

### 3.2 An upper bound on $Z(\lambda, \nu)$

As shown in section 1, the series $\frac{\lambda^j}{(j!)^\nu}$ converges. In addition, $\lim \frac{\lambda^j}{(j!)^\nu} = 0$ as $j \to \infty$. Therefore there exists a value $k$ such that, for $j > k$,

$$\frac{\lambda}{j^{\nu}} < 1 \tag{14}$$

This ratio is monotonically decreasing, meaning that for $j > k$, this series converges faster than a geometric series with multiplier given by (14). Thus, $Z(\lambda, \nu)$ can be approximated by truncating the series at some $k$th term such that (14) holds, i.e.

$$Z(\lambda, \nu) = \sum_{j=0}^{k} \frac{\lambda^j}{(j!)^\nu} + R_k \tag{15}$$

where $R_k = \sum_{j=k+1}^{\infty} \frac{\lambda^j}{(j!)^\nu}$ is the absolute truncation error.

An upper bound can be found, based on the fact that the series $\frac{\lambda^j}{(j!)^\nu}$ $(j = 0, 1, 2, ...)$ decreases at a faster rate than a geometric series. Thus, there exists $0 < \epsilon_k < 1$ for all $j > k$ so that

$$\frac{\lambda}{(j+1)^{\nu}} < \epsilon_k \tag{16}$$

$R_k$ is then bounded by

$$\frac{\lambda^{k+1}}{(k+1)!^{\nu}(1-\epsilon_k)} \tag{17}$$

Another computational improvement, which increases efficiency, is to bound the *relative* truncation error given by

$$\frac{R_k}{\sum_{j=0}^{k} \frac{\lambda^j}{(j!)^\nu}} \tag{18}$$

The relative truncation error can be bounded by

$$\frac{\lambda^{k+1}}{(k+1)!^{\nu}(1-\epsilon_k)} \times \frac{1}{\sum_{j=0}^{k} \frac{\lambda^j}{(j!)^\nu}} \tag{19}$$

**Bounding $Z(\lambda,\nu)^{-1}$:** Computing the inverse function, $Z(\lambda,\nu)^{-1}$, by truncating the infinite sum:

$$\hat{Z}(\lambda,\nu)^{-1} = \left(\sum_{i=0}^{k} \frac{\lambda^j}{(j!)^\nu}\right)^{-1} \tag{20}$$

leads to a relative error that is

$$\frac{\hat{Z}^{-1} - Z^{-1}}{\hat{Z}^{-1}} = \frac{\sum_{j=k}^{\infty} \frac{\lambda^j}{(j!)^\nu}}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}}, \tag{21}$$

which is smaller than (19). Thus, the relative error for computing $Z(\lambda,\nu)^{-1}$ by truncation is bounded by the same bound as that for $Z(\lambda,\nu)$.

# 4 Evaluating moments of the distribution

This section describes ways to compute moments of the COM-Poisson.

The moments are related to derivatives of the normalizing constant $Z$, therefore approximate expressions for the moments can be obtained from the asymptotic approximation of the normalizing constant (13):

$$E[X] = \lambda \frac{d \log Z(\lambda,\nu)}{d\lambda} \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \tag{22}$$

$$E[\log(X!)] = -\frac{d \log Z(\lambda,\nu)}{d\nu} \approx \frac{1}{2\nu^2} \log\lambda + \lambda^{1/\nu}\left(\frac{\log\lambda}{\nu} - 1\right) \tag{23}$$

These approximations are good for $\nu \leq 1$ or $\lambda > 10^\nu$.

To get more precise values, a truncation approach can be used as in section 3. Let the moments be written as

$$E[f(X)] = \sum_{j=0}^{\infty} f(j) \frac{\lambda^j}{j!^\nu Z(\lambda,\nu)} \tag{24}$$

The tail of this infinite sum converges faster than a geometric series. For instance, the ratio of two subsequent terms of the series in the numerator of $E(X)$ is

$$\frac{\lambda (j+1)^{1-\nu}}{j} \tag{25}$$

This ratio tends to 0 as $j \to \infty$, and in addition $\lim \frac{j\lambda^j}{(j!)^\nu} = 0$ as $j \to \infty$. Thus there exists a value $k$ such that, for $j > k$, the ratio (25) is smaller than 1. Also, note that this ratio is monotonically decreasing, meaning that for $j > k$, this series converges faster than a geometric series with a multiplier given by (25).

Similarly, the ratio of subsequent terms of the series in the numerator of $E(X^2)$ is

$$\frac{\lambda (j+1)^{2-\nu}}{j^2} \tag{26}$$

the limit of which is 0 for $\nu \geq 0$. Similar results hold for the remaining expectations.

**Bounding the ratios** Geometric series can be used to specify upper and lower bounds for expectations of the form(24). Write the expectation in the form:

$$E[f(X)] = \frac{\sum_{i=1}^{\infty} a_i}{\sum_{i=1}^{\infty} b_i} = \frac{\sum_{i=1}^{k} a_i + \sum_{i=k+1}^{\infty} a_i}{\sum_{i=1}^{k} b_i + \sum_{i=k+1}^{\infty} b_i} \tag{27}$$

Then an upper bound is given by

$$U_k = \frac{\sum_{i=1}^{k} a_i + \frac{a_{k+1}}{1-a_{k+2}/a_{k+1}}}{\sum_{i=1}^{k} b_i} \tag{28}$$

and a lower bound is given by

$$L_k = \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} b_i + \frac{b_{k+1}}{1-b_{k+2}/b_{k+1}}} \tag{29}$$

One may then choose $k$ so that $U_k - L_k < \epsilon$, for any desired value of $\epsilon$, and then approximate $E[f(X)]$ by

$$E[f(X)] \approx \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} b_i} \tag{30}$$

Alternatively, the relative error can be bounded, setting

$$(U_k - L_k)\frac{\sum_{i=1}^{k} b_i}{\sum_{i=1}^{k} a_i} < \epsilon \tag{31}$$

To avoid reaching machine limits for computing (28) and (29), the ratio can be normalized each time the sum reaches machine limits. This enables a numerical approximation with high precision even when each of the sums exceeds machine limits.

# 5 Maximum-likelihood estimation

The COM-Poisson log-likelihood function for a data set $(X_1, ..., X_n)$ is:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{32}$$

$$\overline{\log(X!)} = \frac{1}{n}\sum_{i=1}^{n} \log(X_i!) \tag{33}$$

$$L(\lambda, \nu) = n\overline{X}\log(\lambda) - n\overline{\log(X!)}\nu - n\log Z(\lambda, \nu) \tag{34}$$

This likelihood is only a function of the sufficient statistics $\overline{X}$ and $\overline{\log(X!)}$, and the maximum-likelihood COM-Poisson will have moments matching these statistics. The likelihood is concave in the parameter $\boldsymbol{\theta} = (\log(\lambda), \nu)$, which means it can be rapidly and reliably maximized by Newton's method.

The gradient of the log-likelihood is

$$\nabla L(\boldsymbol{\theta}) = n \begin{bmatrix} \overline{X} - E[X] \\ -(\overline{\log(X!)} - E[\log(X!)]) \end{bmatrix} \tag{35}$$

and the second derivative matrix is

$$\nabla^2 L(\boldsymbol{\theta}) = n \begin{bmatrix} -\mathrm{var}(X) & \mathrm{cov}(X, \log(X!)) \\ \mathrm{cov}(X, \log(X!)) & -\mathrm{var}(\log(X!)) \end{bmatrix} \tag{36}$$

This matrix has positive determinant, because

$$\mathrm{var}(X)\mathrm{var}(\log(X!)) > \mathrm{cov}(X, \log(X!))^2 \tag{37}$$

However, it also has negative trace, which implies both of its eigenvalues are negative. This proves that the likelihood is concave. The moments needed in these formulas can be computed as in section 4. The Newton update is

$$\boldsymbol{\theta}^{new} = \boldsymbol{\theta} - (\nabla^2 L(\boldsymbol{\theta}))^{-1} \nabla L(\boldsymbol{\theta}) \tag{38}$$

A reasonable starting point for the iteration is the ordinary Poisson MLE: $(\lambda = \overline{X}, \nu = 1)$ so that $\boldsymbol{\theta} = (\log(\overline{X}), 1)$.

## 6 Discussion

The infinite sum, $Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!^\nu}$ is a generalization of several well known infinite sums (such as $e^\lambda$ and $\frac{1}{1-\lambda}$). This function plays a major role in different computations that are needed in the COM-Poisson context. By truncating this sum and bounding the error, the difficulty of computing an infinite sum is overcome for all practical purposes. The numerical approximation does not require specialized software, and can be programmed easily in any language. (We used C++ and S-Plus programs for our application and illustrations).

## Bibliography


Bleistein, N. and Handelsman, R. (1986), *Asymptotic Expansions of Integrals*, Dover, NY.

Wimmer, G. and Altmann, G. (1999), *Thesaurus of univariate discrete probability distributions*, Essen: Stamm. `www.stamm.de/thesaurus/`