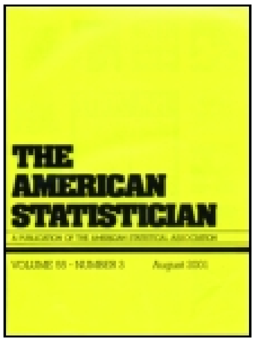

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://amstat.tandfonline.com/loi/utas20

Bridging the Gap: A Generalized Stochastic Process for Count Data

Li Zhu, Kimberly F. Sellers, Darcy Steeg Morris & Galit Shmueli

To cite this article: Li Zhu, Kimberly F. Sellers, Darcy Steeg Morris & Galit Shmueli (2017) Bridging the Gap: A Generalized Stochastic Process for Count Data, The American Statistician, 71:1, 71-80, DOI: 10.1080/00031305.2016.1234976

To link to this article: http://dx.doi.org/10.1080/00031305.2016.1234976

Accepted author version posted online: 21 Sep 2016.
Published online: 21 Sep 2016.

THE AMERICAN STATISTICIAN
2017, VOL. 71, NO. 1, 71–80
http://dx.doi.org/10.1080/00031305.2016.1234976

Taylor & Francis
Taylor & Francis Group

GENERAL

Bridging the Gap: A Generalized Stochastic Process for Count Data

Li Zhu[a,b], Kimberly F. Sellers[b,c], Darcy Steeg Morris[c], and Galit Shmueli[d]

[a]CAC Fund, Shanghai, China; [b]Department of Mathematics and Statistics, Georgetown University, Washington, DC; [c]Center for Statistical Research & Methodology, U.S. Census Bureau, Washington, DC; [d]Institute of Service Science, National Tsing Hua University, Taiwan

ABSTRACT
The Bernoulli and Poisson processes are two popular discrete count processes; however, both rely on strict assumptions. We instead propose a generalized homogenous count process (which we name the Conway–Maxwell–Poisson or COM-Poisson process) that not only includes the Bernoulli and Poisson processes as special cases, but also serves as a flexible mechanism to describe count processes that approximate data with over- or under-dispersion. We introduce the process and an associated generalized waiting time distribution with several real-data applications to illustrate its flexibility for a variety of data structures. We consider model estimation under different scenarios of data availability, and assess performance through simulated and real datasets. This new generalized process will enable analysts to better model count processes where data dispersion exists in a more accommodating and flexible manner.

ARTICLE HISTORY
Received August 2014
Revised March 2016

KEYWORDS
Bernoulli process; Conway-Maxwell-Poisson (COM-Poisson) distribution; Count process; Dispersion; Poisson process; Waiting time

1. Introduction

Throughout history, stochastic processes have been developed to model data that arise in different disciplines, with the most notable being transportation, marketing, and finance. With the rapid development of modern technology, count data have become popular in further areas. As a result, stochastic process models can play a more significant role in today's data analysis toolkit.

The simplest discrete stochastic process is the Bernoulli process, whose associated waiting times are geometric. Meanwhile, the Poisson process is the most popular and used stochastic process for count data. Often considered as the continuous-time counterpart of the Bernoulli process, its most distinguishing property is its underlying assumption of equi-dispersion (i.e., the average number of arrivals equals the variance) in a fixed time period. This assumption, however, is constraining and problematic because many real-world applications contain count data that fail to satisfy the equi-dispersion property. Barndorff-Nielsen and Yeo (1969) and Diggle and Milne (1983) considered negative binomial point processes or, more broadly, "any flexible class of distributions with a variance-to-mean ratio greater than unity" (Diggle and Milne 1983, p. 257). While such models can account for data over-dispersion (i.e., where the variance is larger than the mean), they cannot effectively model data under-dispersion (i.e., where the variance is smaller than the mean).

In this article, we use the Conway–Maxwell–Poisson (COM-Poisson) distribution to derive what we call a COM-Poisson process. The significance of the COM-Poisson distribution and, hence, the corresponding process, lies in its ability to represent a family of processes encompassing count data (including the Poisson, geometric, and Bernoulli processes), and further models a sequence of arrivals whose data display over- or under-dispersion. With the COM-Poisson process, we develop a generalized waiting time distribution that encompasses waiting time distributions associated with the Bernoulli and Poisson processes, respectively, and models the distribution of waiting times for over- or under-dispersed data. Our work develops the COM-Poisson process not only to bridge the gap between two classical count processes, but also to introduce a process that can address a wide range of data dispersion.

The remainder of the article is organized as follows. Section 2 provides background and motivation, briefly reviewing the Bernoulli and Poisson counting processes, along with their respective frameworks and associated properties. Section 3 formally introduces the COM-Poisson and sum-of-COM-Poisson (sCOM-Poisson) distributions, and uses them to develop the COM-Poisson process and study its properties; included is the derivation of the associated generalized waiting time distribution. Section 4 discusses parameter estimation and associated uncertainty quantification. Section 5 considers estimation robustness under different simulated data scenarios. Further, this section illustrates the flexibility of the COM-Poisson process when applied to real-world data, comparing this process approach with Bernoulli, Poisson, and other count processes addressing data dispersion. Finally, Section 6 provides a discussion and future directions.

2. Classical Counting Processes

Çinlar (1975) defined a Bernoulli process with success probability p as a discrete-time stochastic process of the form $\{X_n; n = 1, 2, \ldots\}$ where, for all n, $X_1, \ldots, X_n$ are independent, and X_n takes the values $\{0, 1\}$ with probabilities $P(X_n = 1) = p$ and $P(X_n = 0) = q = 1 - p$. The number of successes that have occurred through the nth trial, $N_n = X_1 + \cdots + X_n$, follows a

CONTACT Kimberly F. Sellers kfs7@georgetown.edu Center for Statistical Research, & Methodology, U.S. Census Bureau, Washington, DC 20233.

Binomial(n, p) distribution; $N_0 = 0$. For any $m, n \in \mathbb{N}$, the distribution of $N_{m+n} - N_m$ also follows a Binomial(n, p) distribution, independent of m, that is, the process has independent increments. In terms of waiting times, T_k is defined as the number of trials it takes to get the kth success. The time between successes, $T_{k+1} - T_k$ for any $k \in \mathbb{N}$, follows a geometric distribution, that is, $P(T_{k+1} - T_k = m) = pq^{m-1}, m = 1, 2, \ldots$. Further, for any m and n, the interval $T_{m+n} - T_m$ is independent of m (Çinlar 1975).

Meanwhile, a Poisson process is a popular, continuous-time, stochastic counting process used to model events such as customer arrivals, electron emissions, neuron spike activity, etc. (Kannan 1979). Let N_t denote the number of events that have occurred up to time $t \geq 0$. By definition (Durrett 2004),

- $N_0 = 0$
- $N_{s+t} - N_s \sim \text{Poisson}(\lambda t)$ [in particular, $N_{s+1} - N_s = \text{Poisson}(\lambda)$]
- N_t has independent increments, that is, for $t_0, t_1, \ldots, t_n$, the variables $N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent.
- the following relations hold:

$$P(N_{s+t} - N_s = 1) = \lambda t + o(t),$$
$$P(N_{s+t} - N_s \geq 2) = o(t),$$

where a function $f(t)$ is said to be of order $o(t)$ if $\lim_{t\to 0} \frac{f(t)}{t} = 0$ (Kannan 1979).

Thus, a homogenous Poisson process has a rate/intensity parameter λ such that the number of events to occur in the interval $(t, t+\tau]$ is Poisson $(\lambda\tau)$, that is,

$$P(N_{t+\tau} - N_t = i) = \frac{e^{-\lambda\tau}(\lambda\tau)^i}{i!}, \quad i = 0, 1, 2, \ldots,$$

and the associated waiting time between events, T_k, is exponentially distributed with parameter λ.

3. The Conway–Maxwell–Poisson (COM-Poisson) Process

The Conway–Maxwell–Poisson (COM-Poisson) distribution is a two-parameter generalization of the Poisson distribution introduced by Conway and Maxwell (1962) and whose statistical properties were studied by Shmueli et al. (2005); see Sellers, Shmueli, and Borle (2011) for a comprehensive discussion about this distribution and its use in various applications and statistical methods. Allowing for both data over- and under-dispersion, this distribution possesses desirable statistical properties (e.g., it belongs to the exponential family in both parameters), conceptual appeal (it generalizes the Poisson, geometric, and Bernoulli distributions), and practical use (it can be used to fit a wide range of count data). In this article, we use the COM-Poisson distribution as the basis for developing the COM-Poisson process. The latter will serve as a flexible stochastic process that generalizes the Poisson process and allows for over- and under-dispersion in counts of events because it is a special case of a weighted Poisson process as described by Balakrishnan and Kozubowski (2008).

In the following, we develop the COM-Poisson process by first introducing relevant, motivating distributions. Section 3.1 describes the COM-Poisson and sum of COM-Poisson (sCOM-Poisson) distributions, and highlights some of their statistical properties. Section 3.2 uses these distributions to derive a homogenous COM-Poisson process. Finally, Section 3.3 introduces a generalized waiting time distribution associated with the COM-Poisson process.

3.1. The COM-Poisson and sCOM-Poisson Distributions

The COM-Poisson probability mass function (pmf) takes the form

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

for a random variable X, where $\nu \geq 0$ is a dispersion parameter such that $\nu = 1$ denotes equi-dispersion, while $\nu > (<) 1$ signifies under-dispersion (over-dispersion); $Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$ is a normalizing constant (Shmueli et al. 2005) and $\lambda = E(X^\nu) > 0$. The COM-Poisson distribution includes three well-known distributions as special cases: the Poisson ($\nu = 1$), geometric ($\nu = 0$ and $\lambda < 1$), and Bernoulli ($\nu \to \infty$) distributions. The expected value and variance of the COM-Poisson distribution can be presented as derivatives with respect to $\ln(\lambda)$ (Sellers, Shmueli, and Borle 2011):

$$E(X) = \frac{\partial \ln Z(\lambda, \nu)}{\partial \ln(\lambda)} \quad \text{and} \quad \text{var}(X) = \frac{\partial^2 \ln Z(\lambda, \nu)}{\partial (\ln \lambda)^2}; \tag{1}$$

more generally, the probability and moment generating functions are $G_X(t) = E(t^X) = \frac{Z(\lambda t, \nu)}{Z(\lambda, \nu)}$ and $M_X(t) = \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}$, respectively. As a weighted Poisson distribution whose weight function is $w(x) = (x!)^{1-\nu}$ (Kokonendji, Miz, and Balakrishnan 2008), the COM-Poisson distribution belongs to both the exponential family and the two-parameter power series (Shmueli et al. 2005; Sellers, Shmueli, and Borle 2011).

The sum of n iid COM-Poisson variables leads to the sCOM-Poisson (λ, ν, n) distribution, which has the following pmf for a random variable $Y = \sum_{i=1}^{n} X_i$, where $X_i \overset{\text{iid}}{\sim}$ COM-Poisson (λ, ν):

$$P(Y = y) = \frac{\lambda^y}{(y!)^\nu [Z(\lambda, \nu)]^n} \sum_{\substack{x_1, \ldots, x_n = 0 \\ x_1 + \cdots + x_n = y}}^{y} \binom{y}{x_1 \; \ldots \; x_n}^\nu,$$
$$y = 0, 1, 2, \ldots; \quad n \in \mathbb{N},$$

where $\binom{y}{x_1 \; \ldots \; x_n}$ is a multinomial coefficient. The sCOM-Poisson (λ, ν, n) distribution encompasses the special cases of the Poisson$(n\lambda)$ distribution when $\nu = 1$, the negative binomial NB$(n, 1-\lambda)$ distribution when $\nu = 0$ and $\lambda < 1$, and the Binomial$\left(n, \frac{\lambda}{\lambda+1}\right)$ distribution as $\nu \to \infty$. The special case of the sCOM-Poisson $(\lambda, \nu, n = 1)$ distribution is a COM-Poisson (λ, ν) distribution. The sCOM-Poisson (λ, ν, n) distribution has moment generating function, $M_Y(t) = \left[\frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}\right]^n$.

The COM-Poisson and sCOM-Poisson distributions are both helpful in our development of a flexible homogenous count process to capture data over- or under-dispersion.

3.2. The Homogenous COM-Poisson Process

Let N_t denote the number of events that have occurred up to time $t \in \mathbb{N}_0$, where the number of events in a unit interval of time follows the COM-Poisson distribution, that is,

$$P(N_{t+1} - N_t = i) = \frac{\lambda^i}{(i!)^\nu Z(\lambda, \nu)}, \quad i = 0, 1, 2, \ldots, \tag{2}$$

where $\lambda > 0$ and $\nu \geq 0$. This homogenous COM-Poisson process follows a COM-Poisson distribution over the interval $(t, t+1]$ with associated parameters λ and ν. In other words, $N_{t+1} - N_t \sim$ COM-Poisson(λ, ν) and has independent increments. For the special case when $\nu = 1$, this is the homogenous Poisson process with parameter λ. By definition,

- $N_0 = 0$
- $N_{s+t} - N_s \sim$ sCOM-Poisson(λ, ν, t)
- N_t has independent and stationary increments, that is, for ordered time points $t_0, t_1, \ldots, t_n$, the variables $N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent.

Thus, a homogenous COM-Poisson process has a rate parameter λ and dispersion parameter ν so that the number of events to occur in the interval $(t, t+\tau]$ follows an sCOM-Poisson (λ, ν, τ) distribution for $\tau \in \mathbb{N}$, independent of t. Given its distributional form, the probability of a single event in the interval $(t, t+\tau]$ is given by

$$P(N_{t+\tau} - N_t = 1) = \frac{\lambda^1}{(1!)^\nu [Z(\lambda, \nu)]^\tau} \sum_{\substack{a_1,\ldots,a_\tau=0 \\ a_1+\cdots+a_\tau=1}}^{1} \begin{pmatrix} & 1 & \\ a_1 \ a_2 & \ldots & a_\tau \end{pmatrix}^\nu$$
$$= \frac{\lambda \tau}{[Z(\lambda, \nu)]^\tau}. \tag{3}$$

For the special case when $\nu = 1$, this reduces to $P(N_{t+\tau} - N_t = 1) = \lambda\tau + o(\tau)$, as noted in Section 2.

3.3. A Generalized Waiting Time Distribution

We have introduced the COM-Poisson distribution to model the number of events that occur in a period of time. We now consider the associated generalized distribution that describes the waiting time until the next occurrence.

To find the waiting time distribution, we consider the probability of no events occurring in a time interval of length τ:

$$P(N_{t+\tau} - N_t = 0) = \frac{\lambda^0}{(0!)^\nu [Z(\lambda, \nu)]^\tau}$$
$$= \frac{1}{Z[(\lambda, \nu)]^\tau} = P(T_t > \tau),$$

thus

$$P(T_t \leq \tau) = 1 - \frac{1}{[Z(\lambda, \nu)]^\tau}. \tag{4}$$

Equation (4) is the cumulative distribution function (cdf) of the COM-Poisson process waiting time. Given a continuous-time process, we differentiate Equation (4) to obtain the density function,

$$f(\tau) = [Z(\lambda, \nu)]^{-\tau} \ln(Z(\lambda, \nu)) = \ln(Z(\lambda, \nu)) e^{-\tau \ln Z(\lambda, \nu)}, \quad \tau \geq 0. \tag{5}$$

This function has the form of an exponential distribution with parameter $\ln(Z(\lambda, \nu))$; we call this a COM-exponential distribution. If we consider instead a discretized waiting time $T_t^* = \lceil T_t \rceil$ (and hence a discrete-time process), where $\lceil \cdot \rceil$ denotes the ceiling function, T_t^* has probability mass function (pmf)

$$P(T_t^* = \tau) = \frac{Z(\lambda, \nu) - 1}{[Z(\lambda, \nu)]^\tau} = \left(\frac{1}{Z(\lambda, \nu)}\right)^{\tau - 1} \left(1 - \frac{1}{Z(\lambda, \nu)}\right), \quad \tau = 1, 2, 3, \ldots. \tag{6}$$

This distribution is geometric with success parameter $1 - \frac{1}{Z(\lambda, \nu)}$; hereafter, we refer to this form as a COM-geometric distribution.

The relationship between the COM-exponential and COM-geometric waiting times is consistent with the relationship between the exponential and geometric distributions in general. The exponential and geometric distributions are known to share similar properties and be continuous and discrete analogs of each other. In particular, for an exponentially distributed random variable X with parameter μ, the random variable $X^* = \lceil X \rceil$ has a geometric distribution with parameter $p = 1 - e^{-\mu}$ on the set $\{1, 2, 3, \ldots\}$ (Stephens 2012; p. 66); letting $\mu = \ln Z(\lambda, \nu)$ for our case, this results in $p = 1 - e^{-\ln Z(\lambda, \nu)} = 1 - \frac{1}{Z(\lambda, \nu)}$. Stephens (2012) further noted that the cdfs corresponding to geometric and exponential distributions are nearly identical for small values of t, hence the same is true for our case defined above.

The waiting times associated with the Poisson and Bernoulli processes, respectively, are special cases of Equations (5) and (6), respectively. We know the Poisson process to be a special case of the continuous-time COM-Poisson process with $\nu = 1$, thus

$$f(\tau) = \ln(Z(\lambda, \nu)) e^{-\tau \ln Z(\lambda, \nu)} = \lambda e^{-\lambda \tau}, \tau \geq 0,$$

that is, an exponential distribution with parameter λ. For the Bernoulli process (i.e., $\nu \to \infty$), we use Equation (6) recognizing that T_t^* is discrete to get

$$P(T_t^* = \tau) = \frac{Z(\lambda, \nu) - 1}{[Z(\lambda, \nu)]^\tau} = \frac{(1+\lambda) - 1}{(1+\lambda)^\tau} = \frac{\lambda}{(1+\lambda)^\tau}$$
$$= \left(\frac{\lambda}{1+\lambda}\right) \left(\frac{1}{1+\lambda}\right)^{\tau - 1}, \quad \tau = 1, 2, 3, \ldots.$$

4. Parameter Estimation

In this section, we introduce the modeling procedure and associated variability study as it relates to the COM-Poisson process. For illustration ease, we discuss these procedures under the context of a discrete waiting time, hence use the COM-geometric waiting time distribution to model the time until the next event. We can similarly pursue these approaches through the continuous-time analog directly, or by refining our time structure into small enough discrete time increments to approximate continuity (Stephens 2012).

First, consider the special case of a COM-Poisson process over a single time unit. The number of events that occur in one time unit is approximated by a COM-Poisson(λ, ν) variable. The following two applications illustrate the modeling procedure for such a process.

- Method 1 (Data on Number of Events in a Single Time Unit): Suppose we have an ordered sequence of count data $x_1, \dots, x_n$, which is well approximated by a COM-Poisson distribution. To identify the respective waiting time distribution, we first apply the maximum likelihood method to estimate λ and ν by maximizing the COM-Poisson log-likelihood,

$$\ln L(\lambda, \nu | x_1, \dots, x_n) = (\ln \lambda) \sum_{i=1}^{n} x_i - \nu \sum_{i=1}^{n} \ln(x_i!) - n \ln Z(\lambda, \nu). \quad (7)$$

We can then show that the waiting time follows a geometric distribution with parameter $\hat{p} = 1 - \frac{1}{Z(\hat{\lambda}, \hat{\nu})}$, where $\hat{\lambda}, \hat{\nu}$ are the respective maximum likelihood estimates of λ, ν. Meanwhile, robustness associated with parameter estimation is often reflected through the corresponding standard errors for the stated estimates. The sampling distributions associated with λ and ν, however, are known to possess skewness (see, e.g., Sellers and Shmueli 2013); hence, we consider two approaches to address uncertainty quantification: nonparametric bootstrapping, and using the information matrix to determine standard errors. Using the latter approach, we supply the corresponding information matrix,

$$I(\lambda, \nu) = -n \cdot E \begin{pmatrix} \frac{\partial^2 \ln P(X=x)}{\partial \lambda^2} & \frac{\partial^2 \ln P(X=x)}{\partial \lambda \partial \nu} \\ \frac{\partial^2 \ln P(X=x)}{\partial \lambda \partial \nu} & \frac{\partial^2 \ln P(X=x)}{\partial \nu^2} \end{pmatrix},$$

where

$$\frac{\partial^2 \ln P(X=x)}{\partial \lambda^2} = -\frac{X}{\lambda^2} + \left(\frac{1}{[Z(\lambda,\nu)]^2}\right)\left(\frac{\partial Z(\lambda,\nu)}{\partial \lambda}\right)^2 - \left(\frac{1}{Z(\lambda,\nu)}\right)\left(\frac{\partial^2 Z(\lambda,\nu)}{\partial \lambda^2}\right),$$

$$\frac{\partial^2 \ln P(X=x)}{\partial \lambda \partial \nu} = \left(\frac{1}{[Z(\lambda,\nu)]^2}\right)\left(\frac{\partial Z(\lambda,\nu)}{\partial \lambda}\right)\left(\frac{\partial Z(\lambda,\nu)}{\partial \nu}\right) - \left(\frac{1}{Z(\lambda,\nu)}\right)\left(\frac{\partial^2 Z(\lambda,\nu)}{\partial \lambda \partial \nu}\right),$$

$$\frac{\partial^2 \ln P(X=x)}{\partial \nu^2} = \left(\frac{1}{[Z(\lambda,\nu)]^2}\right)\left(\frac{\partial Z(\lambda,\nu)}{\partial \nu}\right)^2 - \left(\frac{1}{Z(\lambda,\nu)}\right)\left(\frac{\partial^2 Z(\lambda,\nu)}{\partial \nu^2}\right),$$

and $E(X)$ is defined in Equation (1). Meanwhile, to compute parameter estimates and associated variation via nonparametric bootstrapping, we randomly draw 1000 samples with replacement from the data using the `boot` package (Canty and Ripley 2015) in R.
- Method 2 (Wait-Time Data): For discrete time, the waiting time distribution can be viewed as the pmf of a geometric distribution with parameter $p = 1 - \frac{1}{Z(\lambda,\nu)}$. For parameter estimation, the log-likelihood function of the waiting time distribution is given by

$$\ln L(\lambda, \nu | t_1, \dots, t_n) = n \ln(Z(\lambda, \nu) - 1) - \ln(Z(\lambda, \nu)) \sum_{i=1}^{n} t_i,$$

and the maximum likelihood value is achieved when

$$\frac{1}{Z(\lambda, \nu)} = 1 - \frac{1}{\bar{t}}. \quad (8)$$

Equation (8) is an underdetermined system (Soldatov 2011) because we have one equation and two unknowns (λ and ν). This implies a uniqueness issue where an infinite number of $\{\lambda, \nu\}$ pairs satisfy Equation (8) equally well. This is a powerful result as one can choose any combination of COM-Poisson variables, including Bernoulli (where $\nu \to \infty$) and Poisson variables, with appropriate $\hat{\lambda}$ such that Equation (8) holds. However, suppose instead that, along with an ordered sequence of waiting times $t_1, \dots, t_n$, we have information about the variance-to-mean ratio of the counts (i.e., the “dispersion index”). This additional information allows us to obtain a unique solution for λ, ν. Our goal remains to identify a COM-Poisson process over one time unit that best fits the waiting time data. Equation (8) allows one to estimate $Z(\lambda, \nu)$ while Equation (1) implies that the dispersion index equals

$$\frac{\text{var}(Y)}{E(Y)} = \frac{\partial^2 \ln Z(\lambda, \nu)}{\partial (\ln \lambda)^2} \Big/ \frac{\partial \ln Z(\lambda, \nu)}{\partial \ln(\lambda)}. \quad (9)$$

We can use Equations (8) and (9) to obtain the estimates, $\hat{\lambda}$ and $\hat{\nu}$. Uncertainty quantification can be determined via nonparametric bootstrapping, again resampling 1000 times using the `boot` package in R.
- Method 3 (Data on Number of Events in an s-Unit Interval): Suppose we have a random sample of event counts $y_1, \dots, y_n$ from a sCOM-Poisson(λ, ν, s) distribution with $s \geq 1$. To estimate λ and ν, we maximize the log-likelihood function of the sCOM-Poisson distribution. The associated uncertainty quantification can be determined either via the standard errors from the corresponding information matrix (which is provided in the Appendix), or using nonparametric bootstrapping.

5. Fitting The COM-Poisson Process to Data

5.1. Data Simulations

We estimate the COM-Poisson process model from simulated Poisson count data to illustrate the three parameter estimation procedures described in Section 4. First, we simulate a random sample of size 50 from a Poisson ($\lambda = 2$) distribution. We model the count data assuming a sCOM-Poisson ($\lambda, \nu, n = 2$) distribution and obtain the associated parameter estimates. Then, we repeat the simulation 500 times and inspect the corresponding parameter estimates obtained from each simulation. Based on the theory, we expect $\lambda = 2/n = 2/2 = 1$, and $\nu = 1$. The respective distributions for their estimates ($\hat{\lambda}$ and $\hat{\nu}$) from the 500 simulations are provided in Figure 1. We see that the

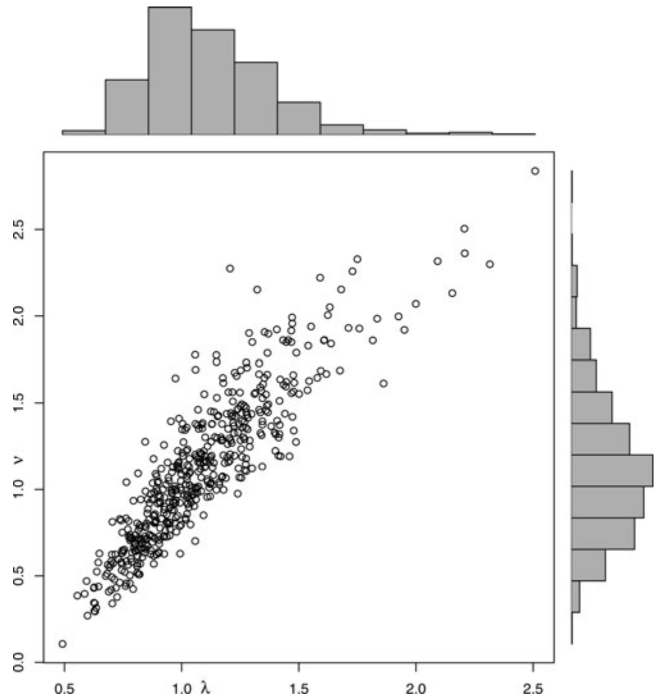

Figure 1. $(\hat{\lambda}, \hat{\nu})$ pairs and corresponding marginal distributions from 500 simulations of Poisson(2) data with sample size = 50. Theoretically, both $\hat{\lambda}, \hat{\nu}$ have expected values equal to 1.

marginal distributions are both right-skewed where the respective medians of the sampling distributions provide reasonable approximations to the theoretical parameter values. Meanwhile, jointly, there appears to be a relative linear association between λ and ν; see Figure 1.

We also conduct a simulation study to compare the two approaches for quantifying uncertainty; see Table 1. We randomly generate 500 data values from a COM-Poisson($\lambda = 1.5, \nu$) distribution with $\nu \in \{0.5, 1, 2, 10\}$, and use the resulting data to estimate $\hat{\lambda}$ and $\hat{\nu}$ via maximum likelihood estimation. To quantify uncertainty via the information matrix approach, we obtain the standard errors associated with the maximum likelihood estimates. Meanwhile, for the nonparametric bootstrapping procedure, we consider the percentile-based 95% confidence interval; see Table 1. Both approaches produce similar results for ν values close to 1. For $\nu >> 1$, however, we see that the bootstrap method succumbs to nuances of varying data resamples that display different dispersion levels; because the process samples with replacement, resulting datasets may contain over-, equi-, or under-dispersion, such that the sampling distribution for ν can vary far more widely than that for λ (Sellers and Shmueli 2013). In this example, the percentile-based 95% confidence interval suggests that bootstrapped resamples generally range in their level of under-dispersion from moderate (7.016) to the extreme Bernoulli case (with an estimated dispersion, 34.139). The associated bias (8.164) indicates, however, that the dispersion level can achieve any potential type of dispersion.

Table 1. Uncertainty quantification comparisons for $\hat{\lambda}$ and $\hat{\nu}$ obtained via the information matrix and nonparametric bootstrapping, respectively. Bootstrap bias associated with estimate $\hat{\theta}$ is bias$(\hat{\theta}) = \frac{1}{1000}\sum_{b=1}^{1000}\hat{\theta}_b - \hat{\theta}$, where b is the bootstrap index . True values are $\lambda = 1.5$ and $\nu \in \{0.5, 1, 2, 10\}$, respectively.

				Nonpar. Boot.	
	Truth	MLE	Info. Mat. SE	Bias	Percentile-based 95% CI
λ	1.5	1.470	0.107	0.006	(1.293,1.710)
ν	0.5	0.489	0.054	0.002	(0.400,0.595)
λ	1.5	1.637	0.149	0.007	(1.391,1.960)
ν	1.0	1.073	0.104	0.005	(0.898,1.268)
λ	1.5	1.508	0.144	0.006	(1.270,1.832)
ν	2.0	1.907	0.173	0.004	(1.594,2.298)
λ	1.5	1.508	0.138	0.002	(1.242,1.819)
ν	10.0	8.821	1.456	8.164	(7.016,34.139)

Table 2. Coverage probability analysis: 500 datasets of 500 data values each are generated from a COM-Poisson($\lambda = 1.5, \nu$) distribution with $\nu \in \{0.5, 1, 2, 10\}$, and the true value is compared with the respective 95% nominal coverage confidence intervals obtained via the information matrix (i.e., the confidence interval obtained by the maximum likelihood estimate ±1.96SE) and nonparametric bootstrapping (i.e., the percentile-based 95% confidence interval).

		Coverage probabilities	
	Truth	Info. Mat.	Nonpar. Boot
λ	1.5	0.952	0.944
ν	0.5	0.954	0.946
λ	1.5	0.960	0.962
ν	1.0	0.968	0.962
λ	1.5	0.950	0.948
ν	2.0	0.954	0.938
λ	1.5	0.934	0.942
ν	10.0	0.938	0.382

Investigating this matter further, Table 2 displays the respective coverage probabilities corresponding to each of the previously considered simulated examples. To determine the respective coverage probabilities, we generate 500 datasets of 500 data values from a COM-Poisson($\lambda = 1.5, \nu$) distribution with $\nu \in \{0.5, 1, 2, 10\}$, and compare the true value to the respective 95% nominal coverage confidence intervals obtained via the information matrix (i.e., the confidence interval obtained by the maximum likelihood estimate ±1.96SE) and nonparametric bootstrapping (i.e., the percentile-based 95% confidence interval) approaches; see Table 2 for details.

Again, we find that both approaches perform comparably and reasonably for ν close to 1. For $\nu >> 1$, however, we see that the nonparametric bootstrap coverage probability is severely underestimated (coverage probability for $\nu = 10$ is 0.382). In this example, the simulated datasets each contain many 0s and 1s, and relatively few 2s. Accordingly, there are many cases where the bootstrap procedure sampling with replacement produces datasets with a binary structure. Under such conditions, the estimation procedure recognizes the data as Bernoulli outcomes and estimates the dispersion parameter as 30 or more. In fact, all of the resulting bootstrapped CIs that did not contain the true value ($\nu = 10$) consistently estimated the dispersion parameter to be at least 30. Thus, while the low coverage probability eludes to potential over-estimation of the dispersion parameter, we are actually seeing an artifact of the bootstrap variation and its impact on the estimation procedure.

5.2. *Real-Data Examples*

In this section, we use two sets of real count data to illustrate the flexibility and predictive power of the COM-Poisson process and compare its performance to other established count models, namely, the Bernoulli and Poisson processes, as well as modeling via a negative binomial or a condensed Poisson (whose pmf for some random variable X is $P(X=0)=e^{-\lambda t}(1+\frac{\lambda t}{2})$; $P(X=x)=e^{-\lambda t}(\frac{(\lambda t)^{2x-1}}{2(2x-1)!}+\frac{(\lambda t)^{2x}}{(2x)!}+\frac{(\lambda t)^{2x+1}}{2(2x+1)!})$, $x=1,2,\ldots$; see Chatfield and Goodhardt (1973) and Johnson, Kemp, and Kotz (2005)) to address over- or under-dispersed data, respectively.

For model comparison using Akaike's information criterion (AIC), Burnham and Anderson (2002) suggest considering $\Delta_i = \text{AIC}_i - \text{AIC}_{\min}$, where $\text{AIC}_{\min}$ is the minimum of the different AIC values, thus inferring that the best model has $\Delta = 0$ and the other models have $\Delta > 0$. Accordingly, analysts can compare models as being a best approximating model via these difference measures in that "models having $\Delta_i \le 2$ have substantial support (evidence), those in which $4 \le \Delta_i \le 7$ have considerably less support, and models having $\Delta_i > 10$ have essentially no support" (Burnham and Anderson 2002; pp. 70–71). We will apply this approach for model comparison accordingly.

Example 1: Fetal lamb movements.
Figure 2 shows the number of movements by a fetal lamb observed by ultrasound and counted in successive 5 sec intervals (Guttorp 1995). The data indicate over-dispersion, with $\widehat{\text{var}(Y)}/\widehat{E(Y)} = 1.810$. Using the procedure described for counts of events (Method 1), we obtain $\hat{\lambda} = 0.277$ [95% CI = (0.222,0.337)] and $\hat{\nu} = 0$ [95% CI = (0.000,0.444)], and the associated waiting time distribution is geometric with parameter 0.277 [95% CI = (0.222,0.333)], as given by Equation (6). Note the difference between the confidence intervals for λ and the waiting time parameter, even though they share the same estimated values (which is expected, since $\hat{\nu} = 0$). This difference in bounds is due to the nonparametric bootstrapping procedure that is performed to determine the percentile-based confidence interval for the waiting time parameter. The procedure resamples the data, from which λ, ν are estimated and those resulting estimates are used to determine the waiting time parameter estimate. Some of the resampled datasets produced $\hat{\nu} > 0$, hence those estimates are not consistent with the special case of a geometric (i.e., the COM-Poisson distribution where $\nu = 0$) distribution. Nonetheless, we see that the resulting confidence intervals are very similar.

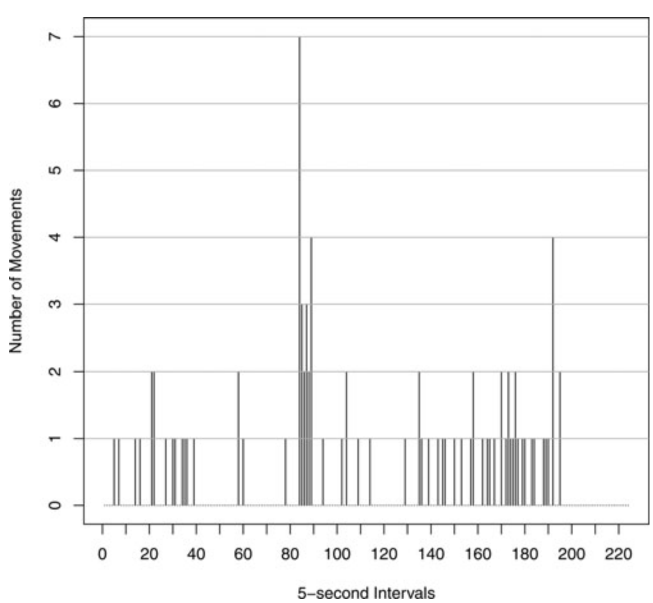

Figure 2. The number of movements by a fetal lamb observed by ultrasound and counted in successive 5 sec intervals (Guttorp 1995).

Meanwhile, assume that we are given only the data dispersion index (1.810) and the waiting times associated with Figure 2 ($\bar{t} = 3.33$). Using Equations (8) and (9) (Method 2), we obtain $\hat{\lambda} = 0.382$ [95% CI = (0.357,0.431)] and $\hat{\nu} = 0$ [95% CI = (0.000,0.000)], and the associated waiting time distribution is geometric with parameter 0.382 [95% CI = (0.356,0.431)]. As expected, since $\hat{\nu} = 0$, we obtain a geometric waiting time with parameters that equal the respective $\hat{\lambda}$ under each method. While we obtain equal estimates for ν under either approach, the estimates for λ vary, due to the lesser information provided in Method 2. By only having waiting time data, analysts know only that at least one event has occurred by the stated time—not how many events occurred by that time. Knowledge of the dispersion level allows us to uniquely determine estimates for λ and ν, yet lack of knowledge persists in that multiple data constructs can produce the same dispersion level.

Given the discussion above, it is interesting that the nonparametric bootstrap percentile-based 95% confidence intervals derived under Method 2 are smaller than those for Method 1. Under Method 2, we assume the dispersion index is known and thus can only resample the waiting times between events. In actuality, however, bootstrapping the real data would introduce variability in both the dispersion index and waiting time values, where both components impact the estimation of λ, ν. Thus, bootstrapping for Method 2 is performed in a restrictive manner, constraining the amount of quantifiable variation.

Figure 3 displays the bootstrap distribution of $\hat{\lambda}$ and $\hat{\nu}$ for Methods 1 and 2. We observe comparable results for the respective distributional shape of both parameters under both methods. We see via the histograms for $\hat{\nu}$ in Figure 3 (and, more prevalently, in Figure 4) the greater skewness associated with the sampling distribution for $\hat{\nu}$, compared to that for $\hat{\lambda}$. Figure 4 demonstrates the joint dependence of λ and ν in maximizing the log-likelihood function described in Equation (7).

Table 3 provides the process comparisons (based on AIC) of the COM-Poisson process when compared with Poisson, negative binomial, Bernoulli, and condensed Poisson processes used to fit the fetal lamb data. These results show that the negative binomial process performs best with the COM-Poisson a close second in performance, outperforming the other distributions considered. This confirms the theoretical results shown in Section 3.2 where the negative binomial distribution is a special case of the sCOM-Poisson distribution for $\nu = 0$. Thus, because we obtain the maximum likelihood estimate $\hat{\nu} = 0$ for this dataset, it implies that the negative binomial process is optimal here. Constraining ourselves to COM-Poisson process consideration, we find that the geometric process best models the data structure. Noting that the geometric distribution is a special case of a negative binomial distribution implies that the negative binomial process will outperform the geometric distribution because of its flexibility. The small difference in AIC

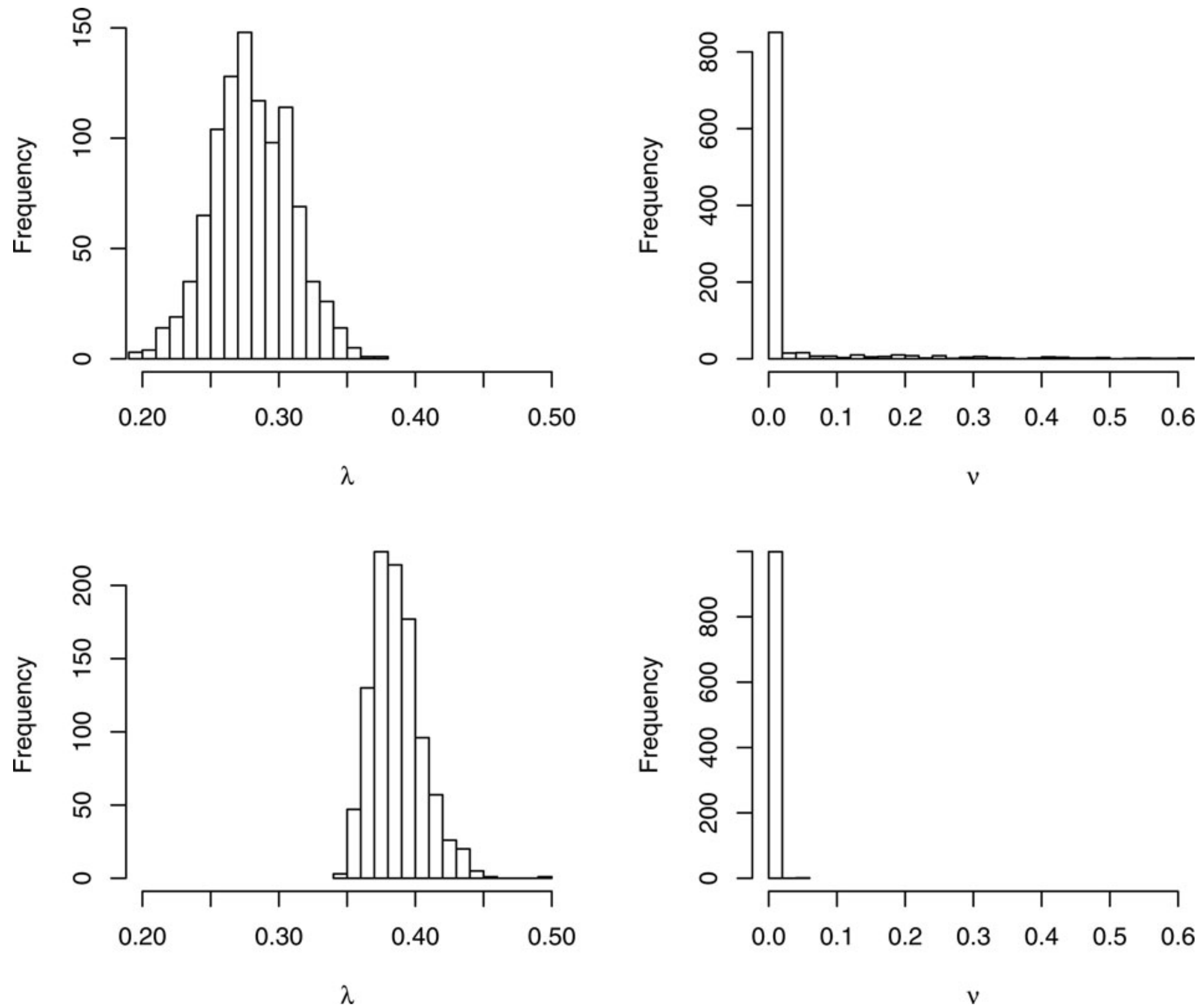

Figure 3. Histogram representations of sampling distributions for $\hat{\lambda}$ (first column) and $\hat{\nu}$ (second column), respectively, under Methods 1 (first row) and 2 (second row). Sampling distributions obtained by conducting nonparametric bootstrapping on the fetal lamb movement dataset.

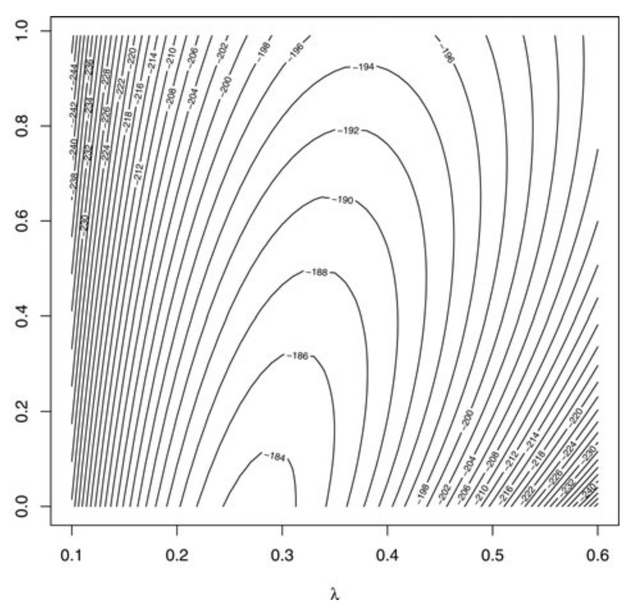

Figure 4. Contour plot of log-likelihood values (where the log-likelihood is defined in Equation (7)) for various values of λ and ν when modeling the fetal lamb movement dataset as a COM-Poisson(λ, ν) process.

between the negative binomial and COM-Poisson processes ($\Delta_i = 1.9$) nonetheless demonstrates substantial support that the COM-Poisson process is likewise a Kullback–Leibler best model (Burnham and Anderson 2002).

Example 2: Flooding on the Rio Negro River.
Figure 5 shows the years of major floods between 1892 and 1992 (inclusive) on the Rio Negro River in Brazil (Brillinger 1995; Guttorp 1995); the data are under-dispersed with $\widehat{\text{var}(Y)}/\widehat{E(Y)} = 0.800$. Using the method described for events data (Method 1), we obtain the maximum likelihood estimates, $\hat{\lambda} = 0.262$ [95% CI = (0.148,0.423)] and $\hat{\nu} = 30.197$ [95% CI = (28.78,30.94)], and the associated waiting time distribution is therefore geometric with parameter 0.208 [95% CI = (0.129,0.297)]; see Equation (6). Instead, suppose we are

Table 3. Model comparison for the fetal lamb movement data; $\Delta_i = \text{AIC}_i - \text{AIC}_{\min}$ as described in Burnham and Anderson (2002).

Model	Poisson	Neg-Bin	Bernoulli	Con. Poisson	COM-Poisson
AIC	392.2	368.2	392.5	457.2	370.1
Δ_i	24.0	0.0	24.3	89.0	1.9

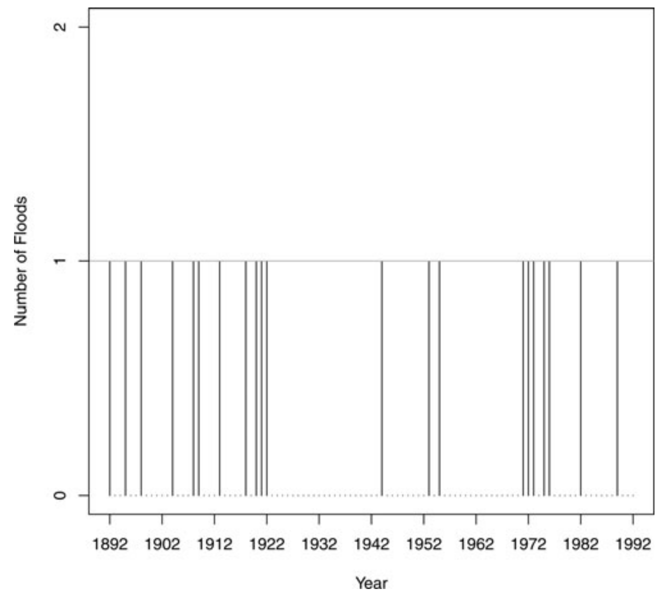

Figure 5. The number of floods on the Rio Negro River.

only given the data dispersion index (0.800), and the interflood durations ($\bar{t} = 4.85$). Using Equations (8) and (9) (Method 2), we obtain the maximum likelihood estimates, $\hat{\lambda} = 0.259$ [95% CI $=$ (0.187,0.513)] and $\hat{\nu} = 6.221$ [95% CI $=$ (2.22,41.04)], implying a geometric wait time distribution with parameter 0.206 [95% CI $=$ (0.157,0.364)]. While the estimates for ν differ considerably, the resulting estimated wait time distributions under either approach are very similar.

The above results demonstrate the relative robustness of $\hat{\lambda}$ while $\hat{\nu}$ shows broad variation in its sampling distribution. The bootstrap distributions of $\hat{\lambda}$ and $\hat{\nu}$ in Figure 6 show a skewed distribution for $\hat{\lambda}$ under either method, yet the range of possible values for $\hat{\lambda}$ remains small in both cases. Meanwhile, the sampling distribution for $\hat{\nu}$ differs vastly depending on the method of bootstrapping. Bootstrapping in the usual way for Method 1 produces a skewed distribution for $\hat{\nu}$ consistent with values expected from a Bernoulli process, as seen empirically in Sellers and Shmueli (2010) and Sellers (2012). Conducting a bootstrapping procedure in accordance with Method 2, however, displays a bimodal (almost U-shaped) distribution where values for $\hat{\nu}$ can either range between [0, 10] (with much of the frequency falling in [0,5]) or [25, 45] (with much of the frequency lying in [35,40]). Figure 7 displays the contour plot of the log-likelihood function as provided in Equation (7) associated with this dataset; in this figure, we see the large range of possible $\hat{\nu}$ values versus the small range in values for $\hat{\lambda}$ such that $\{\hat{\lambda}, \hat{\nu}\}$ pairs produce a maximum log-likelihood value.

Table 4 compares the COM-Poisson process with Poisson, negative binomial, Bernoulli, and condensed Poisson processes

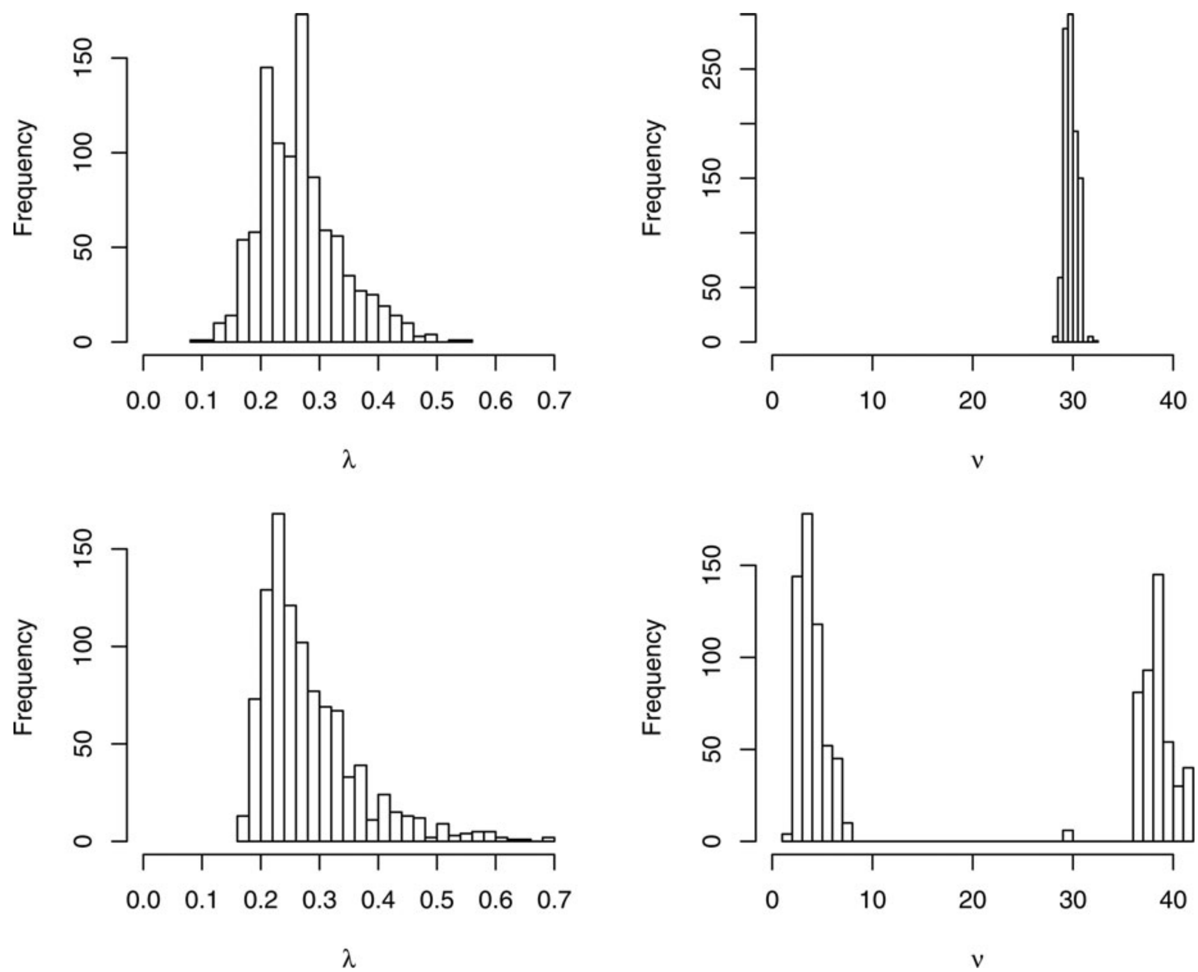

Figure 6. Histogram representations of sampling distributions for $\hat{\lambda}$ (first column) and $\hat{\nu}$ (second column), respectively, under Methods 1 (first row) and 2 (second row). Sampling distributions obtained by conducting nonparametric bootstrapping on the Rio Negro flood dataset.

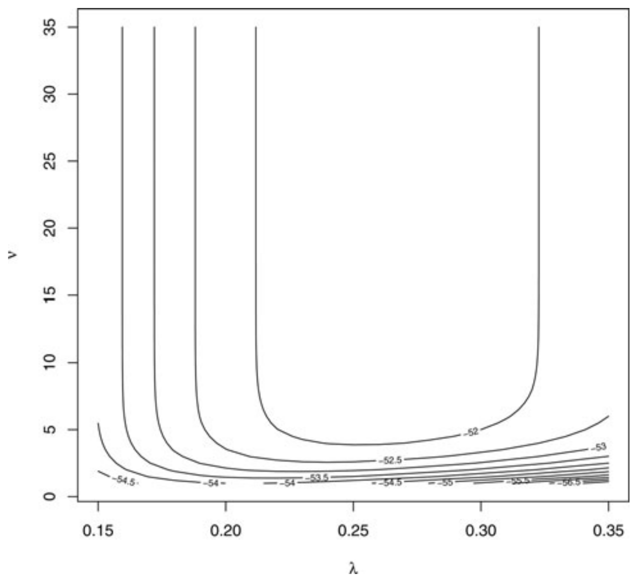

Figure 7. Contour plot of log-likelihood (as described in Equation (7)) values for various values of λ and ν when modeling the Rio Negro flood dataset as a COM-Poisson(λ, ν) process.

used to fit the flood data. These results show that the condensed Poisson process performs best, with the COM-Poisson process ranking second in performance (based on AIC). The small difference in AIC between the condensed Poisson and COM-Poisson processes ($\Delta_i = 1.0$) demonstrate substantial support that the COM-Poisson process is likewise a Kullback–Leibler best model (Burnham and Anderson 2002).

Example 3: Fetal Lamb Movements Over 15 sec Intervals.
We restructure the fetal lamb movement data in Figure 2 to count the number of fetal lamb movements over successive 15 sec intervals. According to Method 3, we obtain the maximum likelihood estimates $\hat{\lambda} = 0.277$ [95% CI = (0.205,0.350)] and $\hat{\nu} = 0$ [95% CI = (0.000,0.123)], and infer a geometric waiting time distribution with parameter 0.276 [95% CI = (0.205,0.350)].

When we increase the level of temporal aggregation for the COM-Poisson process, we lose precision relative to the disaggregated counts. Yet, we notice that the maximum likelihood methods used in Examples 1 and 3 yield similar parameter estimates. This result illustrates the robustness of the COM-Poisson modeling procedures to the choice of temporal aggregation.

6. Discussion

This work develops the COM-Poisson process to serve as a flexible stochastic process for count data containing equi-, over-, or under-dispersion. The significance of the COM-Poisson distribution and, hence, the corresponding process lies in its ability to represent a family of processes encompassing count data, including the Poisson, geometric, and Bernoulli processes. Thus, the COM-Poisson process not only serves as a flexible model for count data expressing a wide range of dispersion, but it can also aid in model selection through insights regarding the estimated dispersion parameter, for cases where the estimated dispersion parameter takes a value associated with one of the special case distributions. With this flexible stochastic process is a likewise flexible waiting time distribution, which can be used to represent the discrete or continuous waiting time associated with the count process. As demonstrated through the examples, the COM-Poisson process and COM-geometric distribution can be used to model and forecast rare events and small counts, without concern regarding the existence and type of data dispersion.

We developed various estimation approaches based on the potentially different data structures made available to the analyst: Methods 1 and 3, under the more common scenario where count data are provided over a given time frame; and Method 2, which considers the less likely scenario of only being provided with waiting time data and dispersion information. With regard to Method 2, we exclude the final waiting time, which is censored by the end of the data collection period. Excluding this censored waiting time is conceptually more appealing because its inclusion would imply that a “final event” had occurred (which could be incorrect). As we see from the examples considered, count occurrences are measured over relative time intervals. Because we do not know the point at which counting occurs, we can only presume the recording process to start at some random point within the interevent time interval. This approach is consistent with assumptions made by Chatfield and Goodhardt (1973) for the condensed Poisson model. Future research can consider how to instead incorporate this censored information into the waiting-time approach, and the associated impact of this alternative. Additional future directions include considering penalized maximum likelihood estimation for parameter estimation for various goals such as predicting future waiting times. Nonetheless, as illustrated in Section 5, all of the methods have shown consistent results, demonstrating the robustness of the proposed process framework. The COM-Poisson process thus serves as a viable, flexible method for modeling count data where significant data dispersion is a possibility.

Accompanying *R* functions were developed for this project and are available upon request. Future work regarding the statistical computing aspects of this work includes the development of an *R* package containing these codes so that interested analysts may access and use these functions on CRAN in their applied work.

Table 4. Model comparison for Rio Negro River flood data, where $\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\mathrm{min}}$ as described in Burnham and Anderson (2002).

Model	Poisson	Neg-Bin	Bernoulli	Con. Poisson	COM-Poisson
AIC	110.0	112.0	109.9	106.3	107.3
Δ_i	3.7	5.7	3.6	0.0	1.0

Appendix: Information Matrix for Method 3

The information matrix for Method 3 is described as follows. Given a random sample of event counts $y_1, \ldots, y_n$ from a sCOM-Poisson(λ, ν, s) distribution, the corresponding information matrix associated with the parameters, λ and ν, is

$$I_{M3}(\lambda,\nu) = -n \cdot E \begin{pmatrix} \frac{\partial^2 \ln P(Y=y)}{\partial\lambda^2} & \frac{\partial^2 \ln P(Y=y)}{\partial\lambda\partial\nu} \\ \frac{\partial^2 \ln P(Y=y)}{\partial\lambda\partial\nu} & \frac{\partial^2 \ln P(Y=y)}{\partial\nu^2} \end{pmatrix},$$

where

$$\begin{aligned} \frac{\partial^2 \ln P(Y=y)}{\partial\lambda^2} &= -\frac{Y}{\lambda^2} + \left(\frac{s}{[Z(\lambda,\nu)]^2}\right)\left(\frac{\partial Z(\lambda,\nu)}{\partial\lambda}\right)^2 \\ &\quad - \left(\frac{s}{Z(\lambda,\nu)}\right)\left(\frac{\partial^2 Z(\lambda,\nu)}{\partial\lambda^2}\right), \\ \frac{\partial^2 \ln P(Y=y)}{\partial\lambda\partial\nu} &= \left(\frac{s}{[Z(\lambda,\nu)]^2}\right)\left(\frac{\partial Z(\lambda,\nu)}{\partial\lambda}\right)\left(\frac{\partial Z(\lambda,\nu)}{\partial\nu}\right) \\ &\quad - \left(\frac{s}{Z(\lambda,\nu)}\right)\left(\frac{\partial^2 Z(\lambda,\nu)}{\partial\lambda\partial\nu}\right), \\ \frac{\partial^2 \ln P(Y=y)}{\partial\nu^2} &= \left(\frac{s}{[Z(\lambda,\nu)]^2}\right)\left(\frac{\partial Z(\lambda,\nu)}{\partial\nu}\right)^2 \\ &\quad - \left(\frac{s}{Z(\lambda,\nu)}\right)\left(\frac{\partial^2 Z(\lambda,\nu)}{\partial\nu^2}\right) \\ &\quad - \frac{1}{[C(\nu)]^2}\left(\frac{\partial C(\nu)}{\partial\nu}\right)^2 + \frac{1}{C(\nu)}\left(\frac{\partial^2 C(\nu)}{\partial\nu^2}\right), \end{aligned}$$

where

$$C(\nu) = \sum_{\substack{x_1,\ldots,x_s=0 \\ x_1+\cdots+x_s=y}}^{y} \begin{pmatrix} & y & \\ x_1 & \ldots & x_s \end{pmatrix}^{\nu},$$

$$\frac{\partial^k C(\nu)}{\partial\nu^k} = \sum_{\substack{x_1,\ldots,x_s=0 \\ x_1+\cdots+x_s=y}}^{y} \left[\left\{\ln\begin{pmatrix} & y & \\ x_1 & \ldots & x_s \end{pmatrix}\right\}^k \begin{pmatrix} & y & \\ x_1 & \ldots & x_s \end{pmatrix}^{\nu}\right],$$

$$k = 1, 2, 3, \ldots,$$

and $E(Y)$ equals s times the COM-Poisson expectation defined in Equation (1).

Acknowledgments

This article is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. The authors thank Ralph Sneider (Monash University) for insightful discussion.

Funding

Support for Li Zhu was provided by the Georgetown Undergraduate Research Opportunities Program (GUROP). Partial funding support for Kimberly Sellers is provided through the ASA/NSF/Census Fellowship Program (U. S. Census Bureau Contract YA1323-14-SE-0122). Galit Shmueli was supported in part by grant 104-2410-H-007-001-MY2 from the Ministry of Science and Technology in Taiwan.

References

Balakrishnan, N., and Kozubowski, T. J. (2008), "A Class of Weighted Poisson Processes," *Statistics and Probability Letters*, 78, 2346–2352. [72]

Barndorff-Nielsen, O., and Yeo, G. F. (1969), "Negative Binomial Processes," *Journal of Applied Probability*, 6, 633–647. [71]

Brillinger, D. R. (1995), "Trend Analysis: Binary-Valued and Point Cases," *Stochastic Hydrology and Hydraulics*, 9, 207–213. [77]

Burnham, K. P., and Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.), New York: Springer. [76]

Canty, A., and Ripley, B. (2015), `boot`: Bootstrap Functions, Version 1.3-15. Available at *http://cran.r-project.org/web/packages/boot/index.html* [74]

Chatfield, C., and Goodhardt, G. J. (1973), "A Consumer Purchasing Model with Erlang Inter-Purchase Time," *Journal of the American Statistical Association*, 68, 828–835. [76,79]

Çinlar, E. (1975), *Introduction to Stochastic Processes*, New York: Prentice-Hall, Inc. [71]

Conway, R. W., and Maxwell, W. L. (1962), "A Queuing Model with State Dependent Service Rates," *Journal of Industrial Engineering*, 12, 132–136. [72]

Diggle, P. J., and Milne, R. K. (1983), "Binomial Quadrat Counts and Point Processes," *Scandinavian Journal of Statistics*, 10, 257–267. [71]

Durrett, R. (2004), *Essentials of Stochastic Processes*, New York: Springer. [72]

Guttorp, P. (1995), *Stochastic Modeling of Scientific Data*, Boca Raton, FL: Chapman & Hall/CRC. [77]

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005), *Univariate Discrete Distributions* (3rd ed.), New York: Wiley. [76]

Kannan, D. (1979), *An Introduction to Stochastic Processes*, New York: Elsevier North Holland. [72]

Kokonendji, C. C., Mizère, D., and Balakrishnan, N. (2008), "Connections of the Poisson Weight Function to Overdispersion and Underdispersion," *Journal of Statistical Planning and Inference*, 138, 1287–1296. [72]

Sellers, K. F. (2012), "A Generalized Statistical Control Chart for Over- or Under-Dispersed Data," *Quality and Reliability Engineering International*, 28, 59–65. [78]

Sellers, K. F., Shmueli, G., and Borle, S. (2011), "The COM-Poisson Model for Count Data: A Survey of Methods and Applications," *Applied Stochastic Models in Business and Industry*, 28, 104–116. [72]

Sellers, K. F., and Shmueli, G. (2010), "A Flexible Regression Model for Count Data," *Annals of Applied Statistics*, 4, 943–961. [78]

——— (2013), "Data Dispersion: Now You See it … Now You Don't," *Communications in Statistics—Theory and Methods*, 42, 3134–3147. [74,75]

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. (2005), "A Useful Distribution for Fitting Discrete Data: Revival of the Conway-Maxwell-Poisson Distribution," *Applied Statistics—Journal of the Royal Statistical Society*, Series C, 54, 127–142. [72]

Soldatov, A. P. (2011), *Underdetermined system, Encyclopedia of Mathematics* Available at *http://www.encyclopediaofmath.org/index.php?title=Underdetermined_system&oldid=16734*. [74]

Stephens, K. S. (2012), *Reliability Data Analysis with Excel and Minitab*, Milwaukee, WI: ASQ Quality Press. [73]