

The COM-Poisson Model for Count Data: A Survey of Methods and Applications

Kimberly F. Sellers
Department of Mathematics and Statistics
Georgetown University
Washington, DC 20057

Sharad Borle
Department of Marketing
Jones Graduate School of Business
Rice University
Houston, TX 77251

Galit Shmueli
Department of Decision, Operations & Information Technologies
Robert H. Smith School of Business
University of Maryland
College Park, MD 20742

Abstract

The Poisson distribution is a popular distribution for modeling count data, yet it is constrained by its equi-dispersion assumption, making it less than ideal for modeling real data that often exhibit over- or under-dispersion. The COM-Poisson distribution is a two-parameter generalization of the Poisson distribution that allows for a wide range of over- and under-dispersion. It not only generalizes the Poisson distribution, but also contains the Bernoulli and geometric distributions as special cases. This distribution's flexibility and special properties has prompted a fast growth of methodological and applied research in various fields. This paper surveys the different COM-Poisson models that have been published thus far, and their applications in areas including marketing, transportation, and biology, among others.

1 Introduction

With the huge growth in the collection and storage of data due to technological advances, count data have become widely available in many disciplines. While classic examples of count data are exotic in nature, such as the number of soldiers killed by horse kicks in the Prussian cavalry

[1], the number of typing errors on a page, or the number of lice on heads of Hindu male prisoners in Cannamore, South India 1937-39 [2], today's count data are as mainstream as non-count data. Examples include the number of visits to a website, the number of purchases at a brick-and-mortar or an online store, the number of calls to a call center, or the number of bids in an online auction.

The most popular distribution for modeling count data has been the Poisson distribution. Applications using the Poisson distribution for modeling count data are wide ranging. Examples include Poisson control charts for monitoring the number of non-conforming items, Poisson regression models for modeling epidemiological and transportation data, and Poisson models for the number of bidder arrivals at an online auction site [3].

Although Poisson models are very popular for modeling count data, many real data do not adhere to the assumption of equi-dispersion that underlies the Poisson distribution (namely, that the mean and variance are equal). An early result has therefore been the popularization of the negative binomial distribution, which can capture overdispersion. While initially using the negative binomial distribution posed computational challenges [4], today there is no such issue and the negative binomial distribution and regression are included in most statistical software packages. Hilbe [5] provides an extensive description of the negative binomial regression and its variants.

The negative binomial distribution provides a solution for overdispersed data, that is, when the variance is larger than the mean. Overdispersion takes place in various contexts, such as contagion between observations. The opposite case is that of underdispersion, where the variance is smaller than the mean. While the literature contains more examples of overdispersion, underdispersion is also common. Rare events, for instance, generate underdispersed counts. Examples include the number of strike outbreaks in the UK coal mining

industry during successive periods between 1948-1959 and the number of eggs per nest for a species of bird [6]. In such cases, neither the Poisson nor the negative binomial distributions provide adequate approximations. Several distributions have been proposed for modeling both over- and underdispersion. These include the weighted Poisson distributions of del Castillo and Perez-Casany [7] and the generalized Poisson distribution of Consul [8]. Both are generalizations of a Poisson distribution, where an additional parameter is added. The generalized Poisson (GP) distribution has also been developed into a regression model ([9], [10]), control charts [11], and as a model for misreporting ([12], [13]). The shortcoming of the GP model, however, is its inability to capture some levels of dispersion, because the distribution is truncated under certain conditions regarding the dispersion parameter and thus is not a true probability model [9].

In this work, we describe a growing stream of research and applications using a flexible two-parameter generalization of the Poisson distribution called the Conway-Maxwell-Poisson (COM-Poisson or CMP) distribution. The main advantage of this distribution is its flexibility in modeling a wide range of over and underdispersion with only two parameters, while possessing properties that make it methodologically appealing and useful in practice. In our opinion, these properties have led to the growing interest and development in both methodological research (theoretical and computational) and applications using the COM-Poisson, both by statisticians and by non-statisticians. While the majority of the COM-Poisson related work has been ongoing in the last 10 years, the amount of interest that it has generated among researchers in various fields and the fast rate of methodological developments warrant a survey of the current state to allow those familiar with the COM-Poisson to gauge the scope of current affairs and for those unfamiliar with the COM-Poisson to get an overview of existing and potential developments and uses.

Historically, the distribution was briefly proposed by Conway and Maxwell [14] as a model for queuing systems with state-dependent service times. To the best of our knowledge, this early form was used in the field of linguistics for fitting word lengths. The great majority of the COM-Poisson development has begun over the last decade, with the initial major publication by Shmueli et al. [15]. The motivation for developing the statistical methodology for the COM-Poisson distribution in [15] arose from an application in marketing, where the purpose was to model the number of purchases by customers at an online grocery store (one of the earlier online grocery stores), where the data exhibited different levels of dispersion when examined by different product categories. The original development started from the point of allowing the ratio of consecutive probabilities $P(Y=y-1) / P(Y=y)$ to be more flexible than a linear function in y , as dictated by a Poisson distribution.

While the COM-Poisson distribution is a two-parameter generalization of the Poisson distribution, it has special characteristics that make it especially useful and elegant. For instance, it also generalizes the Bernoulli and geometric distributions, and is a member of the exponential family in both parameters. Following the publication of [15], research involving the COM-Poisson distribution has developed in several directions by different authors and research groups. It has also been applied in various fields, including marketing, transportation, and epidemiology. The purpose of this paper is to survey the various COM-Poisson developments and applications, which have appeared in the literature in different fields, in an attempt to consolidate the accumulated knowledge and experience related to the COM-Poisson, to highlight its usefulness for solving different problems, and to propose possibilities for further methodological development needed for analyzing count data.

The paper is organized as follows. Section 2 describes the COM-Poisson distribution, its properties and estimation. In Section 3 we describe the different models that have evolved from the COM-Poisson distribution. These include regression models of various forms (e.g., constant

dispersion, group-level dispersion, and observation-level dispersion), via different approaches (GLM, Bayesian), and different estimation techniques (maximum likelihood, quasi-likelihood, MCMC) as well as other COM-Poisson based models (such as cure-rate models). Section 4 surveys a variety of applications using the COM-Poisson, including marketing and electronic commerce, transportation, biology, and disclosure limitation. Section 5 concludes the manuscript with a discussion and future directions.

2 The COM-Poisson Distribution

Conway and Maxwell [14] originally proposed what is now known as the Conway-Maxwell-Poisson (COM-Poisson) distribution as a solution to handling queuing systems with state-dependent service rates. In their article, Conway and Maxwell derived the COM-Poisson distribution from a set of differential difference equations that describe a single queue-single server system with random arrival times, with Poisson inter-arrival times with parameter λ , first-come-first-serve policy, and exponential service times that depend on the system state having mean $= n^c \mu$, where n is the number of units in the system, $1/\mu$ is the mean service time for a unit when that unit is the only one in the system, and c is the “pressure coefficient” indicating the degree to which the service rate of the system is affected by the system state. This distribution has been applied in some fields (mainly in linguistics, see Section 4). However, its statistical properties were not studied in a cohesive fashion until Shmueli et al. [15], who also named it the COM-Poisson (or CMP) distribution.

2.1 The probability distribution

The COM-Poisson probability distribution function has the form

$$P(Y = y) = \frac{\lambda^y}{(y!)^{\nu} Z(\lambda, \nu)}, \quad y = 0, 1, 2, \dots, \quad \lambda > 0, \nu \geq 0$$

for a random variable Y , where $Z(\lambda, \nu) = \sum_{s=0}^{\infty} \frac{\lambda^s}{(s!)^\nu}$ is a normalizing constant; ν is considered the *dispersion parameter* such that $\nu > 1$ represents underdispersion and $\nu < 1$ overdispersion. The COM-Poisson distribution not only generalizes the Poisson distribution ($\nu=1$), but also the geometric distribution ($\nu = 0, \lambda < 1$), and the Bernoulli distribution ($\nu \rightarrow \infty$ with probability $\frac{\lambda}{1+\lambda}$). While there are no simple closed forms linking the parameters λ and ν to moments, there are several relationships that highlight the roles of each parameter and their effect on the distribution. One such formulation (derived by Ralph Snider at Monash University) is $\lambda = E(Y^\nu)$, which displays λ as the expected value of the power-transformed counts, with power ν . Other formulations for the moments include the recursive form

$$E(Y^{r+1}) = \begin{cases} \lambda[E(Y+1)]^{1-\nu} & r = 0 \\ \lambda \frac{\partial}{\partial \lambda} E(Y^r) + E(Y)E(Y^r) & r > 0, \end{cases} \quad (1)$$

and a form that presents the expected value and variance as derivatives with respect to $\log(\lambda)$:

$$\begin{aligned} E(Y) &= \frac{\partial \log Z(\lambda, \nu)}{\partial \log \lambda} \\ \text{Var}(Y) &= \frac{\partial E(Y)}{\partial \log \lambda}. \end{aligned}$$

The expected value and variance are approximated by

$$E(Y) \approx \lambda^{1/\nu} - \frac{\nu-1}{2\nu}, \quad (2)$$

[15], and $\text{Var}(Y) \approx \frac{1}{\nu} \lambda^{1/\nu}$ [16]. These approximations are accurate when $\nu \leq 1$ or $\lambda > 10^\nu$ [15].

The COM-Poisson distribution has the moment generating function, $M_Y(t) = E(e^{Yt}) = \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}$,

and probability generating function, $E(t^Y) = \frac{Z(\lambda t, \nu)}{Z(\lambda, \nu)}$.

In terms of computation, the infinite sum $Z(\lambda, \nu) = \sum_{s=0}^{\infty} \frac{\lambda^s}{(s!)^\nu}$, which is involved in computing moments and other quantities might not appear elegant computationally; however, from a practical perspective, it is easily approximated to any level of precision. Minka et al. [17]

addressed computational issues and provided useful approximations and upper bounds for $Z(\lambda, \nu)$ and related quantities. In practice, the infinite sum can be approximated by truncation. The upper bound on the truncation error from using only the first $k+1$ counts ($s=0, \dots, k$) is given by [17] as

$$\frac{\lambda^{k+1}}{(k+1)!^\nu (1 - \varepsilon_k)^\nu}$$

where $\varepsilon_k > \lambda / (j+1)^\nu$ for all $j > k$.

When $\nu > 1$ (underdispersion), the elements in the sum quickly decrease, requiring only a small number of summations and hence do not pose any computational challenge. For $\nu < 1$, where the truncation of the infinite sum must use multiple values to achieve reasonable accuracy, the following asymptotic form is useful,

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

Minka et al. [17] comment that this formula is accurate when $\lambda > 10^\nu$.

Nadarajah [18] extended the computational focus and derivation, obtaining exact expressions for integer-valued moments greater than 1 (i.e., for special cases involving underdispersion; note that the author mistakenly refers to $\nu > 1$ as overdispersion). While the above approximations are helpful, one can easily compute the exact values of COM-Poisson moments (to a high degree of accuracy), the normalizing constant, etc. by using computational packages such as **compoisson** in R.

Shmueli et al. [15] also showed that the COM-Poisson distribution belongs to the exponential family in both parameters, and determined the sufficient statistics for a COM-Poisson

distribution, namely $S_1 = \sum_{i=1}^n Y_i$ and $S_2 = \sum_{i=1}^n \log(Y_i!)$ where Y_1, \dots, Y_n denotes a random sample of n COM-Poisson random variables (see also Exercise 5 in Chapter 6 of [19]). An example of the usefulness of the sufficient statistics is given in Section 4, in the context of data disclosure.

Taking a Bayesian approach to the COM-Poisson distribution, Kadane et al. [20] used the exponential family structure of the COM-Poisson to establish a conjugate prior density of the form,

$$h(\lambda, \nu) = \lambda^{a-1} e^{-\nu b} Z^{-c} \kappa(a, b, c), \quad (3)$$

where $\lambda > 0$ and $\nu \geq 0$, and $\kappa(a, b, c)$ is the normalization constant; the posterior has the same form with $a' = a + S_1$, $b' = b + S_2$, and $c' = c + n$. They showed that for this density to be proper, a, b , and c must satisfy the condition

$$\frac{b}{c} > \log\left(\left[\frac{a}{c}\right]!\right) + \left(\frac{a}{c} - \left[\frac{a}{c}\right]\right) \log\left(\left[\frac{a}{c}\right] + 1\right).$$

The Bayesian formulation is especially useful when prior information is available. To facilitate conveying prior information more easily in terms of the conjugate prior, [20] developed an online data elicitation program where users can choose a, b, c based on the predictive distribution which is presented graphically.

2.2 COM-Poisson as a Weighted Poisson Distribution

Several authors have noted that the COM-Poisson belongs to the family of weighted Poisson distributions. In general, a random variable Y is defined to have a weighted Poisson distribution if the probability function can be written in the form

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y w_y}{w_y!}, \quad y = 0, 1, 2, \dots; \quad \lambda > 0,$$

where $W = \sum_{s=0}^{\infty} \frac{e^{-\lambda} \lambda^s w_s}{s!}$ is a normalizing constant [7]. Ridout and Besbeas [6] and Rodrigues et al. [21] note that the COM-Poisson distribution can be viewed as a weighted Poisson distribution with weight function, $w_y = (y!)^{1-\nu}$. Kokonedji et al. [22] mention that weighted Poisson distributions are widely used for modeling data with partial recording, in cases where a Poisson variable is observed or recorded with probability w_y , when the event $Y=y$ occurs. Further, presenting the COM-Poisson as a weighted Poisson distribution allows deriving knowledge regarding the types of dispersion that the model can capture. For instance, Kokonedji et al. [22] show that the COM-Poisson is closed by “pointwise duality” for all $\nu \in [0, 2]$; that is, for a given COM-Poisson distribution with $\nu_1 \in [0, 2]$, there exists another COM-Poisson distribution with $\nu_2 = 2 - \nu_1 \in [0, 2]$ which is its pointwise dual distribution. The meaning of the closed pointwise duality for all $\nu \in [0, 2]$ in the COM-Poisson case is that any value within this range is guaranteed to account for either overdispersion or underdispersion of the same magnitude.

Ridout and Besbeas [6] compare the COM-Poisson with an alternative weighted Poisson where the weights have the form,

$$w_y = \begin{cases} e^{-\beta_1(\lambda-y)}, & y \leq \lambda \\ e^{-\beta_2(\lambda-y)}, & y > \lambda \end{cases}$$

Underdispersion exists for $\beta_1, \beta_2 > 0$, overdispersion holds for $\beta_1, \beta_2 < 0$, and equidispersion is achieved when $\beta_1 = \beta_2 = 0$. They called this the three-parameter exponentially weighted Poisson (or two-parameter exponentially weighted Poisson for $\beta_1 = \beta_2 = \beta$), and compared goodness of fit in several applications where the data display underdispersion. For one dataset, which described the number of strikes over successive periods in the UK, the two-parameter exponentially weighted Poisson was found to outperform the COM-Poisson, but the COM-Poisson still provides a good fit. In data describing clutch size (i.e., number of eggs per nest), zero-truncated versions of the distributions are considered because the dataset is severely

underdispersed with a variance-to-mean ratio of 0.10. All distributions considered performed somewhat poorly, including the zero-truncated two- or three-parameter exponentially weighted Poisson and the COM-Poisson distributions.

2.3 Parameter estimation

Shmueli et al. [15] presented three approaches for estimating the two parameters of the COM-Poisson distribution, given a set of data. The first is a weighted least squares (WLS) approach, that takes advantage of the form

$$\log P(Y=y-1)/P(Y=y) = -\log \lambda + \nu \log(y),$$

and which relies on fitting a linear relationship to the ratios of consecutive count proportions; $P(Y=y)$ is estimated by the proportion of y values in the data. WLS is needed to correct for the non-zero covariance between observations and the non-constant variance of the dependent variable. A plot of the ratios versus the counts, y , can give an initial indication of the slope and intercept, displaying the level of dispersion compared to a Poisson case (slope=1). See [15] for further details and an example.

The second estimation approach is a maximum likelihood approach. Maximum likelihood estimates are easily derived for the COM-Poisson distribution due to its membership in the exponential family. The log-likelihood function can be written as

$$\log L(y_1, \dots, y_n | \lambda, \nu) = \log \lambda \sum_{i=1}^n y_i - \nu \sum_{i=1}^n \log(y_i!) - n \log Z(\lambda, \nu), \quad (4)$$

and maximum likelihood estimation can thus be achieved by iteratively solving the set of normal equations

$$E(Y) = \bar{Y}, \text{ and}$$

$$E(\log(Y!)) = \overline{\log Y!}$$

or by directly optimizing the likelihood function using optimization software. For example, using the R software, $\hat{\lambda}$ and $\hat{\nu}$ can be obtained by maximizing the likelihood function directly via the functions **nlm** or **optim** which perform constrained optimization, under the constraint $\nu \geq 0$; alternatively, the likelihood function can be rewritten as a function of $\ln(\hat{\nu})$ so that the likelihood function can be maximized via an unconstrained optimization function such as **nlm** (in R). In either case, the WLS or even Poisson regression estimates can be used as initial estimates.

Finally, a third approach is Bayesian estimation. Because the COM-Poisson has a conjugate prior, estimation is simple and straightforward once the hyper-parameters a, b, c in Equation (3) are specified.

2.4 Further extensions

Shmueli et al. [15] discussed several extensions of the COM-Poisson distribution. Among them are zero-inflated and zero-deflated COM-Poisson distributions, which extend the COM-Poisson by including an extra parameter that captures a contaminating process that produces more or less zeros. In data with no zero counts, the authors discuss a shifted COM-Poisson distribution (also illustrated for modeling word lengths, where there are no 0-length words).

Another extension presented in [15] is the COM-Poisson-Binomial distribution. The latter arises as the conditional distribution of a COM-Poisson variable, conditional on a sum of two COM-Poisson variables with possibly different λ parameters, but same ν . The COM-Poisson-Binomial distribution generalizes the Binomial distribution, allowing for more flexible variance magnitudes compared to the Binomial variance. It can be interpreted as the sum of dependent Bernoulli variables with a specific joint distribution (see [15] for details). This idea can be further extended to a COM-Poisson-Multinomial distribution.

3 Regression models

Modeling count data as a response variable in a regression-type context is common in many applications [23]. The most common model is Poisson regression, yet it is limiting in its equidispersion assumption. Much attention has focused on the case of data overdispersion (e.g., [23], [24], and [5]), which arises in practice due to experimental design issues and/or variability within groups. Many of the proposed approaches (e.g., [25]), however, cannot be applied to address underdispersion, and/or have restrictions that make such approaches unfavorable [4]. The restricted generalized Poisson regression, for example, can effectively model data over- or underdispersion; however, the dispersion parameter is bounded in the case of underdispersion such that it is not a true probability model [9].

Recall the log-likelihood function of the COM-Poisson distribution given in Equation (4). The COM-Poisson's exponential family structure allows for various regression models to be considered to describe the relationship between the explanatory variables and the response. Because the expected value of a COM-Poisson does not have a simple form, different researchers have chosen various link functions to describe the relationship between $E(Y)$ and the set of covariates via $\mathbf{X}\boldsymbol{\beta}$ ($= \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$). The following sections discuss several proposed COM-Poisson regression models.

3.1 COM-Poisson Model with Constant Dispersion

Sellers and Shmueli [26] took a GLM approach and used the link function $\eta(E(\mathbf{Y})) = \log \lambda$ for modeling the relationship between $E(\mathbf{Y})$ and $\mathbf{X}\boldsymbol{\beta}$. This choice of link function, while an indirect function of $E(Y)$, has two advantages: first, it leads to a regression model that generalizes the common Poisson regression (with link function $\log \lambda$) as well as the logistic regression (with link function $\log \lambda = \log p/(1-p)$). Including two well-known regression models as special cases is

useful theoretically and practically. Although logistic regression is a special case of the COM-Poisson regression when $\nu \rightarrow \infty$, Sellers and Shmueli [26] showed that in practice, fitting binary response data with the COM-Poisson regression produces results identical to a logistic regression. The second advantage of using $\log \lambda$ as the link function is that it leads to elegant estimation, inference, and diagnostics. This result highlights the lesser role that the conditional mean plays when considering count distributions of a wide variety of dispersion levels. Unlike Poisson or linear regression, where the conditional mean is central to estimation and interpretation, in the COM-Poisson regression model, we must take into account the entire conditional distribution.

The above formulation allows for estimating β and an unknown constant parameter ν via a set of normal equations. As in the case of maximum likelihood estimation of the COM-Poisson distribution (Section 2.3), the normal equations can be solved iteratively. Sellers and Shmueli [26] proposed an appropriate iterative reweighted least squares procedure to obtain the maximum likelihood estimates for β and ν . Alternatively, the maximum likelihood estimates can be obtained by using a constrained optimization function over the likelihood function (with constraint $\nu \geq 0$), or an unconstrained optimization function over the likelihood written in terms of $\log \nu$. Standard errors of the estimated parameters are derived using the Fisher Information matrix. Sellers and Shmueli [26] used this formulation to derive a dispersion test, which tests whether a COM-Poisson regression is warranted over an ordinary Poisson regression for a given set of data. In terms of inference, large sample theory allows for normal approximation when testing hypotheses regarding individual explanatory variables. For small samples, however, a parametric bootstrap procedure is proposed. The parametric bootstrap is achieved by resampling from a COM-Poisson distribution with parameters $\hat{\lambda} = x' \hat{\beta}$ and $\hat{\nu}$, where $\hat{\beta}$ and $\hat{\nu}$ are estimated from a COM-Poisson regression on the full data set. The resampled data sets

include new response values accordingly. Then, for each resampled data set, a COM-Poisson regression is fitted, thus producing new associated estimates, which can then be used for inference.

With respect to computing fitted values, [26] describes two ways of obtaining fitted values: using estimated means via the approximation, $\hat{y}|x = \hat{\lambda}^{1/\hat{\nu}} - \frac{\hat{\nu}-1}{2\hat{\nu}}$, or using estimated medians via the inverse-CDF for $\hat{y}|x$ and $\hat{\nu}$. Note that the latter is used in logistic regression to produce classifications with a cut-off of 0.5 (whereas choosing a different percentile from the median will correspond to a different cut-off value). Finally, a set of diagnostics is proposed for evaluating goodness of fit and detecting outliers. Measures of leverage, Pearson residuals and deviance residuals are described and illustrated. The *R* package, **COMPoissonReg** (available on CRAN), contains procedures for estimating the COM-Poisson regression coefficients and standard errors under the constant dispersion assumption, as well as computing diagnostics, the dispersion test, and even simulating COM-Poisson data and performing the parametric bootstrap.

A different COM-Poisson regression formulation was suggested and used by Boatwright et al. [27], as a marginal model of purchase timing, in a larger household-level model of the joint distribution of Purchase Quantity and Timing for online grocery sales. Their Bayesian specification allowed for a λ parameter with cross-sectional as well as temporal variation via a multiplicative model, $\lambda_{ij} = \lambda_i \delta_1^{x_{1ij}} \delta_2^{x_{2ij}} \dots \delta_k^{x_{kij}}$, where i denotes a household and j is the temporal index; x_{1ij} , x_{2ij} ... x_{kij} are time-varying covariates measured in logarithmic form. The estimation involves specifying appropriate independent priors over various parameters and using an MCMC framework. Additional details regarding this work are provided in Section 4.2.

Lord et al. [28] independently proposed a Bayesian COM-Poisson regression model to address data that are not equidispersed. Their formulation used an alternate link function, $\log(\lambda^{1/\nu})$, which is an approximation to the mean under certain conditions. Their choice was aimed at allowing the interpretation of the coefficients in terms of their impact on the mean. They used non-informative priors in modeling the relationship between the explanatory and response variables, and performed parameter estimation via MCMC. Comparing goodness-of-fit and out-of-sample prediction measures, [28] empirically showed the similarity in performance of the COM-Poisson and negative binomial regression for modeling overdispersed data, thereby highlighting the advantage of the COM-Poisson over the negative binomial regression in its ability to not only adequately capture overdispersion but also underdispersion and low counts.

Jowaheer & Khan [29] proposed a quasi-likelihood approach to estimate the regression coefficients, arguing that the maximum likelihood approach is computationally intensive. Their method requires the first two moments of the COM-Poisson distribution and, accordingly the authors use the approximation provided in Equation (2) (noting that [29] present this equation with a typographical error) and $\text{Var}(Y) = \frac{\lambda^{1/\nu}}{\nu}$, as well as the moments recursion provided in Equation (1) to build the iterative scheme via the Newton-Rhapson method. The resulting estimators (β^{QL}, ν^{QL}) are consistent and $\sqrt{n}((\beta^{QL}, \nu^{QL}) - (\beta, \nu))^T$ is asymptotically normal as $n \rightarrow \infty$ [29]. While the authors note only a negligible loss in efficiency, one must still be cautious in using this derivation, as the expected value representation used here is accurate under a constrained space for ν and λ ; for data structures. When $\nu > 1$ and $\lambda \leq 10^\nu$, this comparison requires further study.

3.2 COM-Poisson Model with Group-Level Dispersion

Because the COM-Poisson can accommodate a wide range of over- and underdispersion, a group-level dispersion model allows the incorporation of different dispersion levels within a single dataset. The alternative of modeling all groups with a single level of dispersion causes information loss and can lead to incorrect conclusions.

Sellers and Shmueli [30] introduced an extension of their COM-Poisson regression formulation [26], which allows for different levels of dispersion across different groups of observations. Their formulation uses the link functions,

$$\begin{aligned}\log(\lambda) &= \beta_0 + \sum_{j=1}^p \beta_j X_j \\ \log(\nu) &= \gamma_0 + \sum_{k=1}^{K-1} \gamma_k G_k,\end{aligned}$$

where G_k is a dummy variable corresponding to one of K groups in the data.

Estimating the β and γ coefficients is done by maximizing the log-likelihood, where the log-likelihood for observation i is given by

$$\log L_i(\lambda_i, \nu_i | y_i) = y_i \log(\lambda_i) - \nu_i \log(y_i!) - \log Z(\lambda_i, \nu_i),$$

where

$$\begin{aligned}\log(\lambda_i) &= \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} \doteq \mathbf{X}_i \boldsymbol{\beta}, \text{ and} \\ \log(\nu_i) &= \gamma_0 + \gamma_1 G_{i1} + \dots + \gamma_{K-1} G_{i,K-1} \doteq \mathbf{G}_i \boldsymbol{\gamma}.\end{aligned}$$

Since the COM-Poisson distribution belongs to the exponential family, the appropriate normal equations for β and γ can be derived. Using the Poisson estimates, $\beta^{(0)}$ and $\gamma^{(0)} = \mathbf{0}$, as starting values, coefficient estimation can again be achieved via an appropriate iterative reweighted least squares procedure or by using existing nonlinear optimization tools (e.g., `nlm` or `optim` in R) to directly maximize the likelihood function. The associated standard errors of the estimated coefficients are derived in an analogous manner to that described in [26].

3.3 COM-Poisson Model with Observation-Level Dispersion

An even more flexible model in terms of dispersion is to allow dispersion to differ for the different observations. [16] suggested that such a model would be useful in modeling power outages.

[31] independently proposed the model that allows for the dispersion parameter, ν_i , to vary with observation i , and considered a relationship between ν_i and the explanatory variables in the $(p+1)$ -dimensional row vector, \mathbf{X}_i . Accordingly, the log-likelihood for observation i is given as

$$\log L_i(\lambda_i, \nu_i | y_i) = y_i \log \lambda_i - \nu_i \log(y_i!) - \log Z(\lambda_i, \nu_i),$$

where

$$\log \lambda_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \doteq \mathbf{X}_i \boldsymbol{\beta}, \text{ and}$$

$$\log \nu_i = \gamma_0 + \gamma_1 x_{i1} + \dots + \gamma_p x_{ip} \doteq \mathbf{X}_i \boldsymbol{\gamma}.$$

As in previous models, appropriate normal equations can be derived for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and coefficient estimation can be achieved via an appropriate iterative reweighted least squares procedure or by using existing nonlinear optimization tools.

3.4 Other Models

COM-Poisson Cure-Rate Model

Rodrigues et al. [21] used the COM-Poisson distribution to establish a cure-rate model. Letting Y denote the minimum time-to-occurrence among all competing causes, the survival function is defined as

$$S_p(Y) = P(Y \geq y) = \sum_{m=0}^{\infty} P(M = m) [S(y)]^m,$$

where $S(y) = 1 - F(y)$ is the i.i.d. survival function for the j competing causes, W_j ($j=1, 2, \dots, m$), and M (i.e., the number of competing causes) is a weighted Poisson random variable as defined in Section 1.2. Thus, letting M follow a COM-Poisson distribution implies the survival

function $S_p(y) = \frac{Z(\eta S(y), \nu)}{Z(\eta, \nu)}$, where $\eta = \exp(\theta)$. Analogous to the flexibility of the COM-Poisson

distribution, the COM-Poisson cure-rate model allows for flexibility with regard to dispersion. In particular, it encompasses the promotion time cure model ($\nu = 1$), and the mixture cure model ($\nu \rightarrow \infty$).

COM-Poisson Model with Censoring

Censored count data arise in applications such as surveys, where the possible answers to a question are, e.g., 0,1,2,3,4+. Extensions of the Poisson and negative binomial regression models exist for the case of censoring, as is a generalized Poisson model [10].

Sellers and Shmueli [32] introduced a COM-Poisson model that allows for right-censored count data. To incorporate censoring, an indicator variable δ_i is used to denote a censored observation; an observation is either completely observed ($\delta_i=0$), or censored ($\delta_i=1$). The likelihood function is therefore given by

$$\begin{aligned} \log L &= \sum_{i=1}^n (1 - \delta_i) \log P(Y_i = y_i) + \delta_i \log P(Y_i \geq y_i) \\ &= \sum_{i=1}^n (1 - \delta_i) [y_i \log \lambda_i - \nu \log(y_i!) - \log Z(\lambda_i, \nu)] + \delta_i \log P(Y_i \geq y_i), \end{aligned}$$

where $P(Y_i \geq y_i) = 1 - \sum_{s=0}^{y_i-1} \frac{\lambda_i^s}{(s!)^\nu Z(\lambda_i, \nu)}$. One can analogously modify the above equations to

allow for left- or interval-censoring. The authors compare the predictive power (i.e., the performance of the model in terms of predicting new observations) of the censored COM-Poisson model with other censored regressions applied to an underdispersed dataset, and find that the censored COM-Poisson and censored GP models perform comparably well, with the COM-Poisson model obtaining the best predictive scores in a majority of cases.

4 Applications

4.1 Linguistics

Early applications of the COM-Poisson distribution have been in linguistics for the purpose of fitting word lengths. Theory in linguistics suggests that the most “elementary” form of a word length distribution follows the difference equation [33],

$$P(Y = y + 1) = \frac{a}{(y + 1)^b} P(Y = y);$$

that is, words of successive length (e.g., lengths y and $y+1$) are represented in texts in a proportional way [34]. Hence, the COM-Poisson distribution, which adheres to this form, has been used for modeling word length. Wimmer et al. [33] used it to model the number of syllables in the Hungarian dictionary and in Slovak poems. Best [34] notes that the COM-Poisson distribution best models the German alphabetical lexicon (Viëtor) as well as a German frequency lexicon (Kaeding). Among 135 texts, in all but four texts the COM-Poisson was the best fitting model among a set of models (including the hyper-Poisson and hyper-Pascal).

4.2 Marketing and eCommerce

The Poisson and the negative binomial distributions remain popular tools when it comes to applications involving count data in Marketing and eCommerce. Since the revival of COM-Poisson distribution, however, there have been several applications in Marketing and eCommerce using the COM-Poisson. These applications have benefited from the property of COM-Poisson that allows for added flexibility to model under- as well as overdispersed data.

Boatwright et al. [27] used the COM-Poisson distribution to develop a joint (quantity and timing) model for grocery sales at an online retailer. The number of weeks between purchases was modeled as a COM-Poisson variable, namely

$$w_{ij} \sim \begin{cases} (1 - \pi_{ij})\text{COM} - \text{Poisson}(\lambda_{ij}, \nu), & \text{for } w_{ij} > 0 \\ \pi_{ij}, & \text{for } w_{ij} = 0, \end{cases}$$

where $w_{ij} = 0, 1, \dots$ measures the inter-purchase time in weeks rounded off to the nearest week, i indexes the household and j indexes the temporal purchases; π_{ij} is the probability that $w_{ij} = 0$. A multiplicative form for λ_{ij} was used, $\lambda_{ij} = \lambda_i \delta_1^{x_{1ij}} \delta_2^{x_{2ij}} \dots \delta_k^{x_{kij}}$, where $x_{1ij}, x_{2ij} \dots x_{kij}$ were time-varying covariates measured in logarithmic form. Household heterogeneity in the expected inter-purchase time was incorporated by allowing a hierarchy over $\lambda_i \sim \text{gamma}(a, b)$. Independent gamma priors were specified for the δ, ν , and a parameters, and an inverse gamma prior was specified for the b parameter. The model was estimated using an MCMC sampler. Borle et al. [35] used a similar joint model to study the impact of a large scale reduction in assortment at an online grocery retailer. They found that the decline in shopping frequency resulted in a greater loss than did the reduction in purchase quantities, and that the impact of assortment cut varied widely by category, with less-frequently purchased categories more adversely affected.

In another application of the COM-Poisson, [36] used the distribution to model quarterly T-shirt sales across 196 stores of a large retailer. The data consisted of quarterly sales of T-shirts (in eight colors and four sizes) across these stores along with information on the extent of stockouts in these shirts. The authors built a demand model (in the presence of stockouts) and studied the impact of each SKU on sales. An imputation procedure was used to impute demand when stockouts were observed. The results showed many items (T-shirts) affected category sales over and above their own sales volume. After deconstructing the role of a stockout of individual items into three effects (namely, lost own sales, substitution to other items, and the category sales impact), they found that the category impact has the largest magnitude. Interestingly, the disproportionate impact of individual items on category sales was not restricted to top selling items, for almost every single item affected category sales.

Kalyanam et al. [36] also performed some robustness checks to validate their model; these tests were primarily a comparison of results under alternative model specifications. The alternative models considered were a Poisson model, a geometric model, a COM-Poisson model without imputation (using a censored likelihood), a normal regression model and a “Rule of thumb” approach. In terms of log-likelihood measures, the COM-Poisson models (with and without imputation) performed best. In terms of predictive ability (as measured by the MAD statistic), the COM-Poisson and Poisson models were somewhat similar in performance.

In yet another interesting application of the COM-Poisson model, [37] used it to study customer behaviors at a US automotive services firm. The study was carried out to evaluate the effects of participation in a satisfaction survey and examine the role of customer characteristics and store-specific variables in moderating the effects of participation. The data for this study came from a longitudinal field study of customer satisfaction conducted by the US automotive services firm. The data contained two groups of customers, namely, one group that was surveyed for customer satisfaction, while the other group was not surveyed by the firm. Four customer behaviors were studied: (1) the number of promotions redeemed by a customer on each visit, (2) the number of automotive services purchased on each visit, (3) the time since the last visit in days (i.e., inter-purchase time), and (4) the amount spent during each visit. Results revealed a substantial positive relationship between satisfaction survey participation and the four customer behaviors studied. Two of the four behaviors (number of services bought per visit, and the number of promotions redeemed per visit) were modeled using a COM-Poisson distribution. In particular, the number of services bought was modeled as a COM-Poisson variable with a one-unit location shift, i.e., a shifted COM-Poisson variable. Figure 1 below shows bar charts of these variables across the entire data.

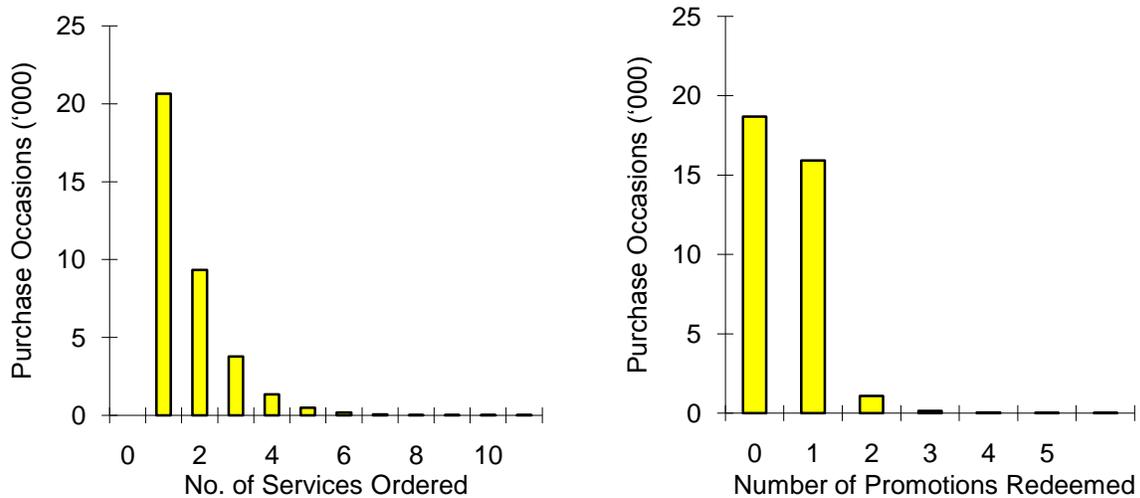


Figure 1: Distribution of the number of services bought per visit (left panel) and the number of promotions redeemed per-visit (right panel). Frequencies are in thousands.

As seen from the bar charts, there is overdispersion in the number of services bought, while the number of promotions redeemed is underdispersed. This is also borne out by the estimated ν parameters for these two variables ($\nu_{\text{services}} = 0.61$, $\nu_{\text{promotions}} = 4.69$, respectively). This is an example where flexibility of the COM-Poisson in accounting for *over-* as well as *under-*dispersion in the data is very useful. The same count distribution can be used to model these two variables. An alternative count distribution such as the negative binomial may have performed equally well in modeling ‘number of services bought’, however it would have been a poor choice in modeling ‘number of promotions redeemed’.

An e-commerce application of COM-Poisson is the study of the extent of multiple and late bidding in eBay online auctions. Borle et al. [38] empirically estimated the distribution of bid timings and the extent of multiple bidding in a large set of eBay auctions, using bidder experience as a mediating variable. The extent of multiple bidding (the number of times a bidder changes his/her bids in a particular auction) was modeled as a COM-Poisson distribution. The

data consisted of over 10,000 auctions from 15 consumer product categories. The two estimated metrics (extent of late bidding, and the extent of multiple bidding) allowed the authors to place these product categories along a continuum of these metrics. The analysis distinguished most of the product categories from one another with respect to these metrics, implying that product categories, after controlling for bidder experience, differ in the extent of multiple bidding and late bidding observed in them.

Apart from these applications, there have also been applications of the COM-Poisson in another important area of marketing, namely that of customer lifetime value estimation (the monetary worth of a customer to a firm); see [39]. The customer lifetime value estimation becomes important because it is used as a metric in many marketing decisions that a firm makes, hence any improvements in its estimation has direct benefits to these decisions. In the marketing literature, the 'lifetime of a customer' in these models is typically measured either as a continuous time or in terms of 'number of lifetime purchases'. Accordingly, a continuous or a count distribution is used to model lifetime respectively. Singh et al. [39] propose a modeling framework using data augmentation; the framework allows for a multitude of lifetime value models to be proposed and estimated. As a demonstration, the authors estimate two extant models in the literature and also propose and estimate three new models. One of the proposed models uses a COM-Poisson distribution to model the lifetime purchases. Compared to the other similar model which uses the Beta-Geometric distribution as the count distribution), the proposed COM-Poisson model improved the customer lifetime value predictions across a sample of 5000 customers by about 42% (a MAD statistic of \$62.46 as compared to \$107.72 using the Beta-Geometric).

4.3 Transportation

[28] and [40] used a COM-Poisson regression model to analyze motor vehicle crash data in

order to link the number of crashes to the entering flows at intersections or on segments. [28] analyzed two datasets via several negative binomial and COM-Poisson generalized linear models: one containing crash data from signalized four-legged intersections in Toronto, Ontario; the second dataset contained information from rural four-lane divided and undivided highways in Texas. The results of this study found that the COM-Poisson models performed as well as the negative binomial models with regard to goodness-of-fit and predictive performance. The authors have thus argued in favor of the COM-Poisson model given its flexibility in handling underdispersed data as well. [40] analyzed an underdispersed crash dataset from 162 railway-highway crossings in South Korea during 1998-2002. COM-Poisson models were found to produce better statistical performance than Poisson and gamma models.

4.4 Biology

Ridout and Besbeas [6] considered various forms of a weighted Poisson distribution (of which, they note, the COM-Poisson distribution is one) to model underdispersed data; To illustrate this, they modeled the clutch size (i.e., number of eggs per nest) for a species of bird. The clutch size data were collected for the linnnet in the United Kingdom between 1939-1999. The data exhibited strong underdispersion where the variance-to-mean ratio was 0.10; see [6] for details. The COM-Poisson model reportedly “performed poorly” with regard to goodness-of-fit (obtaining a chi-squared statistic equaling 4222.8 with 3 d.f.), yet did considerably better than many other weighted Poisson distributions that were considered by the authors.

4.5 Disclosure Limitation

Count data arise in various organizational settings. When the release of such data is sensitive, organizations need information-disclosure policies that protect data confidentiality while still providing data access. Kadane et al. [41] used the COM-Poisson as a tool for disclosure limitation. They showed that, by disclosing only the sufficient statistics (S_1 and S_2) and sample

size n of a COM-Poisson distribution fitted to a confidential one-way count table, the exact cell counts $(f_j, j=0, 1, 2, \dots, J)$ as well as the table size (J) is sufficiently masked. The masking is obtained via two results: (1) usually many count tables correspond to the disclosed sufficient statistics; and (2) the various possible count tables are equally likely to be the undisclosed table. Moreover, finding the various solutions requires solving a system of linear equations, which are underdetermined for tables with more than three cells, and can be computationally prohibitive for even small tables. They illustrated the proposed policy with two examples. The “small” example consisted of a one-way table with counts of 0-5, which when given only the sufficient statistics produced 14 possible solutions. The larger and more realistic example consisted of a one-way table with 0-11 counts (the number of injuries in 10,000 car accidents in 2001), which when given only the sufficient statistics produced more than 80,000,000 possible solutions. The proposed policy is the first to deal with count data by releasing sufficient statistics.

5 Discussion

The COM-Poisson distribution is a flexible distribution that generalizes several classical distributions (namely, the Poisson, geometric, and Bernoulli) via its dispersion parameter. As a result, it bridges data distributions that demonstrate under-, equi-, and overdispersion. Because it generalizes three well-known distributions, and some regression formulations generalize two popular models (logistic regression and Poisson regression), the COM-Poisson offers more than just a new model for count data. Its ability to handle different dispersion types and levels makes it useful in applications where the level of dispersion might vary, yet a single analytical framework is preferred.

In terms of fit, the COM-Poisson appears to fit data at least as well as competing two-parameter models. For instance, [15] reports that while the generalized Poisson of [8] is flexible in terms of

modeling overdispersion and has simple expressions for the normalizing constant and moments, it cannot handle underdispersion and is not in the exponential family, which makes analysis more difficult. Numerical studies show that for every COM-Poisson distribution with $0.75 < \nu < 1$ (or so) there is a generalized Poisson distribution with very similar form. For $\nu < 0.75$, however, the two families differ markedly. In the linguistics applications, [34] has compared the COM-Poisson with several alternatives and found it as a better overall alternative for modeling word lengths. [28] shows that it performs at least as well as a negative binomial for over-dispersed data.

From a theoretical point of view, the COM-Poisson has multiple properties that make it favorable for methodological development. These properties have likely led to the fast growth of papers in disparate areas developing COM-Poisson based models.

As to computational considerations, while the COM-Poisson does not offer simple closed form formulas for moments, and includes the infinite sum as a normalizing constant, in practice these only rarely are a cause for concern. Computationally, the infinite sum can be truncated while achieving precision to any pre-specified level. In cases of overdispersion where the sum might require many more terms to reach some accuracy, approximations are available. Software packages such as the *R* packages **compoisson** and **COMpoissonReg** include Z-function calculations and offer users an easy way to estimate parameters and compute moments numerically. The unavailability of a simple formula for the mean does make interpreting coefficients in a regression model more complicated than in linear or multiplicative models where the mean and variance/dispersion are either independent or identical (such as linear or Poisson regression). As with other popular and useful regression models, the solution is to use marginal analysis and interpret coefficients by examining the conditional mean or median (see [26]).

As illustrated by the breadth of work surveyed here, the COM-Poisson distribution has swiftly grown in interest, both with regard to methodological and applied work. Yet, there remain numerous opportunities for continued research and investigation with regard to this flexible distribution. One open question that has not been thoroughly studied regards the predictive power of the COM-Poisson model. While the COM-Poisson has been consistently shown to perform well in terms of fit, the few papers that report predictive power results using the COM-Poisson do not convey a consistent picture. In some cases, the predictive power of the COM-Poisson is shown to be similar to that of a Poisson distribution or the generalized Poisson distribution, while in others the COM Poisson outperforms these distributions. Predicting individual values requires defining point predictions (e.g., means or medians) as well as predictive intervals. Evaluating the predictive performance in the count data context is also non-trivial [42] (e.g., how is a prediction error defined?).

Meanwhile, in almost every application where the Poisson distribution plays an important role, there is an opportunity to expand the toolkit to consider over- and underdispersion, which are frequently the case with real data. Examples of such areas include the fields of reliability, data tracking and process monitoring, and time series analysis of count data. Even within the context of queuing (or more generally, stochastic processes) where the COM-Poisson originated, there has been very little in the way of further developing and applying the COM-Poisson distribution. In addition to the many fields where Poisson distributions are assumed, there is a growing number of fields where more and more count data are becoming available. Thus, there remain many opportunities for significant contribution to expanding the scope and usefulness of statistical modeling of count data, taking the parsimonious COM-Poisson approach.

Acknowledgements

The authors thank Jay Kadane from Carnegie Mellon University for his insightful comments and discussion regarding this manuscript.

References

- [1] Bortkewicz L von. *Das Gesetz der Kleinen Zahlen*. Leipzig: Teubner, 1898.
- [2] Bliss CI, Fisher RA. Fitting the negative binomial distribution to biological data. *Biometrics* 1953; **9**:176-199.
- [3] Etzion H, Pinker E, Seidmann A. Analyzing the Simultaneous Use of Auctions and Posted Prices for Online Selling. *Manufacturing and Service Operations Management* 2006; **8**: 68-91. DOI: 10.1287/msom.1060.0101
- [4] McCullagh P, Nelder, J. *Generalized Linear Models, 2nd edition*. Chapman & Hall/CRC: New York, NY, 1997.
- [5] Hilbe JM. *Negative Binomial Regression*, Cambridge University Press: New York, NY, 2008.
- [6] Ridout MS, Besbeas P. An empirical model for underdispersed count data. *Statistical Modelling* 2004; 4: 77-89. DOI: 10.1191/1471082X04st064oa
- [7] del Castillo J, Pérez-Casany M. Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference* 2005; **134**: 486-500. DOI: 10.1016/j.jspi.2004.04.019

[8] Consul PC. *Generalized Poisson Distributions: Properties and Applications*. Marcel Dekker: New York, NY, 1989.

[9] Famoye F. Restricted generalized Poisson regression model. *Communications in Statistics – Theory and Methods* 1993; **22**: 1335-1354. DOI: 10.1080/03610929308831089

[10] Famoye F, Wang WR. Censored generalized Poisson regression model. *Computational Statistics and Data Analysis* 2004; **46**: 547-560. DOI: 10.1016/j.csda.2003.08.007

[11] Famoye F. Statistical control charts for shifted generalized Poisson distribution. *Statistical Methods and Applications* 2007; **3**: 339-354. DOI: 10.1007/BF02589023

[12] Neubauer G, Djuras G. A generalized Poisson model for underreporting. In *Proceedings of the 23rd International Workshop on Statistical Modelling*, Eilers PHC (ed). Universiteit Utrecht: Netherlands, 2008; 368-373.

[13] Pararai M, Famoye F, Lee C. Generalized Poisson-Poisson mixture model for misreported counts with an application to smoking data. *Journal of Data Science* 2010; **8**: 607-617.

[14] Conway RW, Maxwell WL. A queuing model with state dependent service rates. *Journal of Industrial Engineering* 1962; **12**:132-136.

[15] Shmueli G, Minka TP, Kadane JB, Borle S, Boatwright P. A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Applied Statistics* 2005; **54**:127-142. DOI: 10.1111/j.1467-9876.2005.00474.x

- [16] Guikema SD, Coffelt JP. A flexible count data regression model for risk analysis. *Risk Analysis* 2008; **28**: 213-223. DOI: 10.1111/j.1539-6924.2008.01014.x
- [17] Minka TP, Shmueli G, Kadane JB, Borle S, Boatwright P. Computing with the COM-Poisson distribution. Carnegie Mellon University Department of Statistics Technical Report Series: Pennsylvania, 2003; 776: 1-7.
- [18] Nadarajah S. Useful moment and CDF formulations for the COM-Poisson distribution. *Statistical Papers* 2009; **50**: 617-622. DOI: 10.1007/s00362-007-0089-9
- [19] Faraway JJ. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Models*, Chapman & Hall / CRC: New York, NY, 2005.
- [20] Kadane JB, Shmueli G, Minka TP, Borle S, Boatwright P. Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Analysis* 2005; **1**:363-374. DOI:10.1214/06-BA113
- [21] Rodrigues J, de Castro M, Cancho VG, Balakrishnan N. COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference* 2009; **139**: 3605-3611. DOI: 10.1016/j.jspi.2009.04.014
- [22] Kokonedji CC, Mizere D, Balakrishnan N. Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference* 2008; **138**:1287-1296. DOI: 10.1016/j.jspi.2007.05.028

- [23] Barron D. The analysis of count data: overdispersion and autocorrelation. *Sociological Methodology* 1992; **22**: 179-220.
- [24] Breslow N. Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association* 1990; **85**: 565-571.
- [25] Ismail N, Jemain A. Handling overdispersion with negative binomial and generalized Poisson regression models. *Casualty Actuarial Society Forum* 2007; 103-158.
- [26] Sellers K, Shmueli G. A flexible regression model for count data. *Annals of Applied Statistics* 2010; **4**: 943-961. DOI: 10.1214/09-AOAS306
- [27] Boatwright P, Borle S, Kadane JB. A model of the joint distribution of purchase quantity and timing. *Journal of the American Statistical Association* 2003; **98**:564-572. DOI: 10.1198/016214503000000404
- [28] Lord D, Guikema SD, Geedipally SR. Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis & Prevention* 2008; **40**:1123-1134.
- [29] Jowaheer V, Khan NAM. Estimating regression effects in COM-Poisson generalized linear model. *World Academy of Science, Engineering and Technology* 2009; **53**: 1046-1050.
- [30] Sellers K, Shmueli G. Data dispersion: Now you see it... now you don't. Robert H. Smith School Research Paper No. RHS 06-122. <http://ssrn.com/abstract=1612755> [21 May 2010]

[31] Sellers K, Shmueli G. A regression model for count data with observation-level dispersion. In *Proceedings of the 24th International Workshop on Statistical Modelling*, Booth JG (ed). Cornell University: New York, 2009; 337-344.

[32] Sellers K, Shmueli G. Predicting Censored Count Data with COM-Poisson Regression. Robert H. Smith School Research Paper No. RHS-06-129. <http://ssrn.com/abstract=1702845> [October 29, 2010]

[33] Wimmer G, Kohler R, Grotjahn R, Altmann G. Toward a theory of word length distributions. *J. Quant. Linguistics* 1994; **1**: 98–106. DOI: 10.1080/09296179408590003

[34] Best K-H. Probability distributions of language entities. *Journal of Quantitative Linguistics* 2001; **8**: 1-11. DOI: 10.1076/jqul.8.1.1.4091

[35] Borle S, Boatwright P, Kadane JB, Nunes J, Shmueli G. Effect of Product Assortment Changes on Consumer Retention. *Marketing Science* 2005; **24**: 616-622. DOI: 10.1287/mksc.1050.0121

[36] Kalyanam K, Borle S, Boatwright P. Deconstructing each item's category contribution. *Marketing Science* 2007; **26**: 327-341. DOI: 10.1287/mksc.1070.0270

[37] Borle S, Dholakia U, Singh S, Westbrook R. The impact of survey participation on subsequent behavior: An empirical investigation. *Marketing Science* 2007; **26**: 711-726. DOI: 10.1287/mksc.1070.0268

[38] Borle S, Boatwright P, Kadane JB. The timing of bid placement and extent of multiple bidding: An empirical investigation using ebay online auctions. *Statistical Science* 2006; **21**: 194-205. DOI: 10.1214/088342306000000123

[39] Singh S, Borle S, Jain D. A Generalized Framework for Estimating Customer Lifetime Value When Customer Lifetimes Are Not Observed. *Quantitative Marketing and Economics* 2009; **7**: 181-205. DOI: 10.1007/s11129-009-9065-0

[40] Lord D, Geedipally SR, Guikema SD. Extension of the application of Conway-Maxwell-Poisson models: analyzing traffic crash data exhibiting underdispersion. *Risk Analysis* 2010; **30**: 1268–1276. DOI: 10.1111/j.1539-6924.2010.01417.x

[41] Kadane JB, Krishnan R, Shmueli G. A Data Disclosure Policy for Count Data Based on the COM-Poisson Distribution. *Management Science* 2006; **52**: 1610-1617. DOI: 10.1287/mnsc.1060.0562

[42] Czado C, Gneiting T, Held L. Predictive model assessment for count data. *Biometrics* 2009; **65**: 1254-1261. DOI: 10.1111/j.1541-0420.2009.01191.x