CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open-access journal indexed by

Volume 12 Number 2
December 2021
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Chilean Journal of Statistics

Volume 12, Number 2
December 2021

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Contents

Chilean Journal of Statistics
Vol. 12, No. 2, December 2021, 229–252

Statistical Modeling
Research Paper

On a weighted Poisson distribution and its associated regression model

Bernardo B. de Andrade[1,*], Raul Y. Matsushita[1], Pushpa N. Rathie[1], Luan Ozelim[2], and Sandro B. de Oliveira[3]

[1]Departamento de Estatística, Universidade de Brasília, Brasília, Brazil
[2]Departamento de Engenharia Civil e Ambiental, Universidade of Brasília, Brasília, Brazil
[3]Rede Sarah de Hospitais de Reabilitação, Brasília, Brazil

(Received: 20 November 2020 · Accepted in final form: 03 April 2021)

Abstract

Count data emerge naturally within the biomedical and economic sciences, in engineering and in industrial applications. The benchmark Poisson distribution is seldom an appropriate statistical model for counts but none of the more flexible distributions available are universally accepted as an alternative. Among such flexible models, the class of weighted Poisson distributions has recently been studied in theoretical investigations but their application is still incipient. This article investigates a particular weighted Poisson model, providing the associated statistical tools for analyzing count data. We make comparisons with other flexible models using public available datasets. For the weighted Poisson model under investigation, we have developed estimation by maximum likelihood and method of moments, random number generation, visual tools for univariate analysis and finally, regression modeling. Results indicate that weighted Poisson distributions are very flexible and capable of modeling count responses in different scenarios.

Keywords: Generalized linear models · Overdispersion · Quasi-likelihood · Touchard · Underdispersion

Mathematics Subject Classification: Primary MSC 62J12 · Secondary MSC 62Fxx

1. Introduction

The Poisson model is the default for analyzing statistically independent counts, such as number of insurance claims and days of hospitalization, and it should provide an adequate fit when data come from a population with mean equal to the variance (equidispersion). However, count data often exhibit over or underdispersion and several distributions have been proposed for modeling counts. The most documented alternatives to the Poisson distribution are the quasi-Poisson (QP) model, the negative binomial (NB) and, to a lesser extent, the generalized Poisson (GP) (Cameron and Trivedi, 1998; Hilbe, 2014). Less popular models include the Poisson-inverse Gaussian, the compound Poisson, the hyper Poisson, the Poisson-Lindley, the Conway-Maxwell-Poisson (CMP) and the weighted Poisson (WP); see

*Corresponding author. Email: bbandrade@unb.br

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Matsushita et al. (2018) and references therein. More complex models include the Poisson-Goncharov (Denuit, 1997), the Hinde-Demétrio family by Kokonendji et al. (2004) and a four-parameter extension of the CMP by Chakraborty and Imoto (2016). It is a long, yet incomplete, list and no single model dominates all others since there are multiple interrelated criteria for judging a given model. One might consider, in his own order of importance: (i) the model's ability to address varying levels in both directions of dispersion; (ii) mathematical and computational tractability; (iii) whether the model generalizes the Poisson; (iv) if the model arises naturally in some observable process like natural phenomena or in connection with a stochastic process of broad applicability; (v) if the model is a member of the exponential family (EF); (vi) availability of estimation and visualization tools for data analysis; (vii) availability of associated regression tools and interpretability of coefficients; (viii) applicability in a specific field (say, actuarial modeling of claim counts or econometric modeling of health insurance, etc.) and (ix) applicability across many disciplines.

Theoretical aspects of WP distributions, covering much of (i)–(v), have been studied by several authors (del Castillo and Pérez-Casany, 1998; del Castillo and Pérez-Casany, 2005; Kokonendji et al., 2008; Matsushita et al., 2018). However, the statistical toolbox for WP models still lacks several important tools. This paper focuses on a particular WP model which we call the Touchard model introduced by del Castillo and Pérez-Casany (1998). Ho et al. (ress) present simulation showing the effect of misspecifying the model (Poisson instead of Touchard) in the context of control charts.

In Section 2 we describe the methodological background of WP distributions including the Touchard. New results for the Touchard model are given in the Appendix A-D. Section 3 develops tools for univariate estimation, inference and visualization. Finally, regression modeling is addressed in Section 4 where we develop estimation, inference and diagnostic tools similar to those in generalized linear models (GLMs). Section 5 concludes the paper.

All computations and graphics presented here were done in the R system and all the required code will be included in the next release of the publicly available package (Andrade and Oliveira, 2019) which, so far, only deals with fixed a.

2. Weighted Poisson models

Weighted distributions date back to Fisher (1934) and have been used to adjust a given benchmark model relative to the way the data are ascertained (Rao, 1985). The adjusted distribution is used to model observational data recorded without a suitable sampling frame including situations such as size-biased sampling, damage models, nonresponse and visibility bias (Patil and Rao, 1978). Under idealized conditions, an observed value y would be a realization of a benchmark random variable Y^*. (For instance, one can think that the number of insurance claims would be Poisson distributed if there were no hunger for bonus (deliberate non-reporting of accidents to save bonus on next premium), if all drivers were insured (no selection bias), if all (insured) drivers were subjected to the same routes and same driving distances, etc. However, not only policies are not randomly drawn with a proper sampling frame but also the conditions just listed are not met. The result is a distribution of claims biased towards 0 and 1.) However, the benchmark may need to be adjusted so that its support is reweighted according to the belief that when the event $Y = y$ is realized, the probability of ascertaining it is $w(y)$. Thus, the realization y is, in fact, from a weighted version Y with probability density function $f(y) = w(y)f^*(y)/\tau$, where f^* is the benchmark density and τ is the normalization constant.

When modeling counts, the benchmark density is often Poisson, $f^*(y) = \mathrm{e}^{-\lambda}\lambda^y/y!$, leading to the general class of WP distributions which we denote by $\mathrm{WP}(\lambda; w(y))$. The model $\mathrm{WP}(\lambda; y)$ is known as the size-biased Poisson model and its distribution is simply $1 + \mathrm{Poi}(\lambda)$. Another important model is $\mathrm{WP}(\lambda; (y!)^\nu)$, for $\nu > 0$, which is the CMP distribution.

We also note that important models such as the NB and the GP are not members of the WP(λ; $w(y)$) class. Interestingly, the NB distribution is a Poisson-Gamma mixture, in which the parameter space of the Poisson is weighted, rather than its support.

A family within the WP(λ; $w(y)$) class is obtained with $w(y) = \exp[\delta t(y)]$, where t is a convex function and $\delta \in \mathrm{R}$ provides overdispersion ($\delta < 0$), Poissoness ($\delta = 0$) or underdispersion ($\delta > 0$). We denote this family by WP(λ, δ; t). Kokonendji et al. (2008) showed, among other theoretical aspects, that such WP models are pointwise dual in the sense that the entire range of δ is guaranteed to account for either over or underdispersion of the same magnitude.

An important member in the WP(λ, δ; $t(y)$) family is the one with $t(y) = \log(y + a)$, for $a \geq 0$ (del Castillo and Pérez-Casany, 1998). Also, del Castillo and Pérez-Casany (2005) fitted this model (with fixed $a = 7$) to the counts of car accidents in a year among 9461 drivers. We have fitted the WP(λ, δ; $\log(y + a)$) model to different datasets and results indicate that it is a strong competitor for both under and overdispersed cases; see Table 2.

2.1 The Touchard distribution redefined

The WP(λ, δ; $\log(y+1)$) model has been recently studied by Matsushita et al. (2018) where the choice of $a = 1$ was inspired by the Touchard polynomials, rather than by weighting schemes. Those authors have labeled it the Touchard distribution, which we redefine. Note that del Castillo and Pérez-Casany (1998) denoted the WP(λ, δ; $\log(y+a)$) model by WPD$_a$, since a is a fixed tuning parameter. The resulting two-parameter model is more tractable both mathematically and computationally but it is not justifiable for actual data analysis across different disciplines. In order to distinguish the WPD$_a$ model from the general case, which allows for choices of $w(y)$ other than $(y+a)^\delta$, and also to avoid confusion with the use of weights (in the context of regression), we hereafter call the WP(λ, δ; $\log(y + a)$) model the Touchard model and denote $Y \sim \mathrm{Tou}(\lambda, \delta, a)$. Its probability density function is given by

$$f(y, \lambda, \delta, a) = \frac{\lambda^y (y + a)^\delta}{y! \tau(\lambda, \delta, a)}, \quad y = 0, 1, \ldots, \tag{1}$$

with $a, \lambda > 0$ and $\delta \in \mathrm{R}$. Note that if $\delta \neq 0$ then $a = 0$ can be considered, in which case $\mathbb{P}(Y = 0) = 0$. We do not consider this case here.

Numerical evaluation of the normalization constant $\tau(\lambda, \delta, a)$ can be done by truncation of its defining sum. The number of terms required to reach a given precision depends on the parameter values. Even though 50 terms suffice in most cases (Matsushita et al., 2018, Table 1), τ can be computed with a pre-specified relative precision, without fixing the truncation point, using a recursive expression similar to the formula given in (10) of Matsushita et al. (2018); see also Appendix A.

The values of a and δ jointly determine the shape of the distribution. With $\delta < 0$, the smaller a is, the more the distribution is inflated at zero and the resulting model becomes an alternative to zero-inflated and hurdle models. Larger values of a provide milder zero inflation for given δ. With $\delta > 0$, the smaller a is, the more the distribution is deflated at zero relative to the Poisson. Larger values of a descrease the zero deflation, making the distribution closer to Poisson. Figure 1 illustrates these facts for some selected values of the parameters.

From now on, we define $\tau_j(\lambda, \delta, a) = \tau(\lambda, \delta + j, a)/\tau(\lambda, \delta, a)$. Statistical moments $m_k = \mathbb{E}(Y^k)$ are given by (del Castillo and Pérez-Casany, 1998)

$$m_k = \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} \tau_j(\lambda, \delta, a), \tag{2}$$

yielding the mean given by

$$\mu = \tau_1(\lambda, \delta, a) - a, \tag{3}$$

and the variance stated as

$$\sigma^2 = \tau_2(\lambda, \delta, a) - \tau_1^2(\lambda, \delta, a). \tag{4}$$

Touchard quantiles may be calculated with an initial approximation based on the Cornish-Fisher expansion (using up to m_3) followed by a search in the appropriate direction as implemented in Andrade and Oliveira (2019).

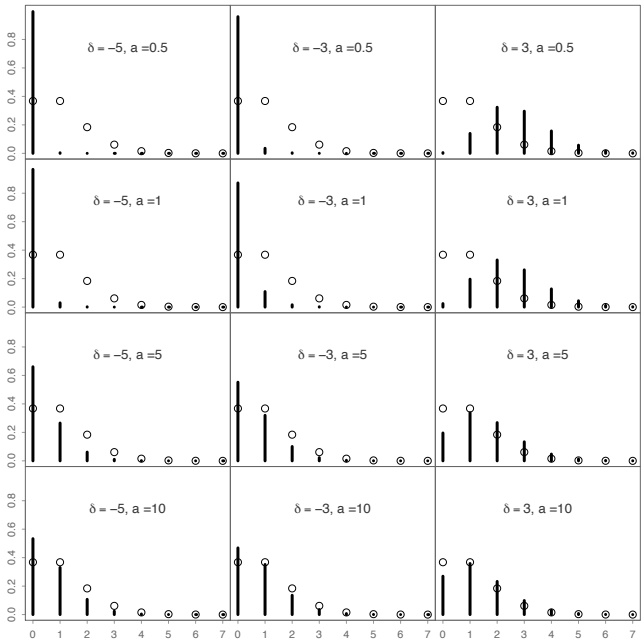

Figure 1. Examples of WP (Touchard) probabilities with $\lambda = 1$, $\delta = \pm 3$ and $a = 0.5, 1, 5$ and 10. Circles represent Poisson probabilities with $\lambda = 1$.

Exact (or approximate) results for the variance as a function of the mean are given in Table 1 for four models of interest besides the Touchard: the two usual benchmarks, QP and NB, and the CMP and GP which are well documented in the literature and which have been implemented for data analysis in the R system, including regression. The QP, NB and GP exhibit polynomial relations between their means and variances with linear, quadratic and cubic relations, respectively. As opposed to these models, the difference between the mean and the variance in the Touchard and in the CMP models are constant for large enough μ. Figure 2 illustrates the mean-variance relation for the Touchard model with different parameter values. The larger $|\delta|$, the further away the curve (μ, σ^2) is from the 45° diagonal (equidispersion). Larger values of a bring the curve closer to the diagonal and it also decreases the initial curvature. The curvature is a lot more sensible to the value of a

and the magnitude of δ in the overdispersed cases ($\delta < 0$). Therefore, it must be noted that even though the mean and variance are not explicit parameters, the Touchard model can be implicitly reparametrized by the mean and the variance (del Castillo and Pérez-Casany, 2005).

Table 1. Mean-variance relations for selected models.

	$\mathrm{QP}(\mu, \varphi)$	$\mathrm{NB}(\mu, \alpha)$	$\mathrm{GP}(\mu, \alpha)$	$\mathrm{Tou}(\lambda, \delta, a)$	$\mathrm{CMP}(\lambda, \nu)$
σ^2	$\varphi\mu$	$\mu + \alpha\mu^2$	$\mu(1+\alpha\mu)^2$	$\approx \mu - \delta\left[\frac{\lambda}{\lambda+a}\right]^2$	$\approx \frac{\mu+(\nu-1)/(2\nu)}{\nu}$

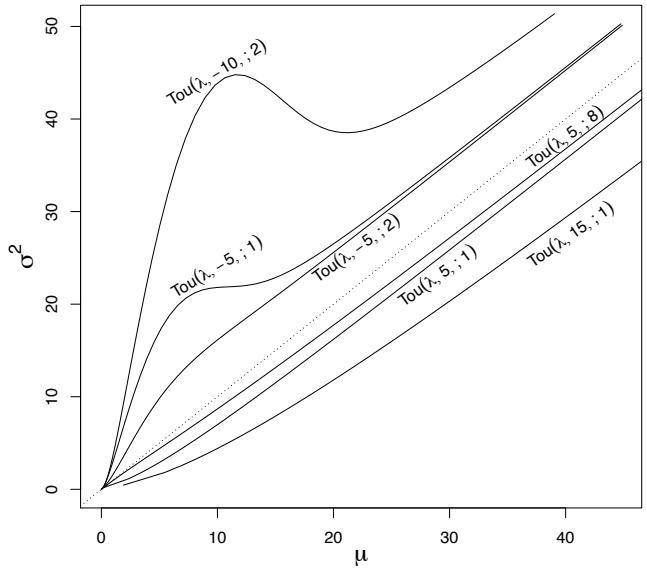

Figure 2. Variance as a function of the mean for Touchard models with arbitrary parameter values. Dotted line on diagonal represents the Poisson $\sigma^2 = \mu$.

3. Estimation, testing and visualization

Here, we describe estimation and testing for the Touchard parameters via maximum likelihood (ML) method of moment method (MM). Conditions for consistency and normal-based large-sample inference require the support to be independent of the parameters, identifiability and bounds on third derivatives (Lehmann and Casella, 1998, Sec. 6.5). We provide a proof of identifiability in Appendix B. With a fixed, the Touchard distribution is a member of the two-dimensional EF with sufficient statistics Y and $Z = \log(Y + a)$ and natural parameters $\log(\lambda)$ and δ. For instance, we already know that $\partial^3 \log(f(y))/(\partial\lambda^2\partial\delta)$ must be bounded by some $M(Y)$ with finite expectation. Therefore, we only need to check third-order derivatives of the log-likelihood involving a. These derivates involve (finite) moments of Y, Z and $(Y + a)^{-k}$, for $k \leq 3$, and therefore satisfy the necessary conditions.

3.1 Maximum likelihood estimation

Given a random sample $(Y_1, \ldots, Y_n)$ with observations $(y_1, \ldots, y_n)$, the ML estimates $\widehat{\lambda}$, $\widehat{\delta}$ and $\widehat{a}$ must satisfy the conditions stated as

$$\begin{cases} \mu(\widehat{\lambda}, \widehat{\delta}, \widehat{a}) - \bar{y} & = 0, \\ \kappa(\widehat{\lambda}, \widehat{\delta}, \widehat{a}) - \bar{z} & = 0, \\ \theta_1(\widehat{\lambda}, \widehat{\delta}, \widehat{a}) - \bar{w}_1 & = 0, \end{cases}$$

where $\kappa(\lambda, \delta, a) \equiv \mathbb{E}(Z)$, $W_j = (Y + a)^{-j}$ and $\theta_j(\lambda, \delta, a) \equiv \mathbb{E}(W_j)$.

Using results from Appendix C, it can be shown that the expected Fisher information is given by

$$\boldsymbol{I}(\lambda, \delta, a) = n \begin{pmatrix} \dfrac{\sigma^2}{\lambda^2} & \dfrac{\operatorname{Cov}(Y, Z)}{\lambda} & \dfrac{\delta \operatorname{Cov}(Y, W_1)}{\lambda} \\ & \operatorname{Var}(Z) & \delta \operatorname{Cov}(Z, W_1) \\ & & \delta^2 \operatorname{Var}(W_1) + \delta(W_2 - \theta_2) \end{pmatrix}.$$

Standard errors (SEs) of ML estimators are computed from the diagonal of $\boldsymbol{I}^{-1}(\widehat{\lambda}, \widehat{\delta}, \widehat{a})$ or from the inverse of observed Fisher information which is often produced by (Newton-type) numerical maximization routines.

3.2 Method of moments

Based on the expression given in Equation (2), the moment conditions for the MM estimators are expressed as

$$\begin{cases} \mu(\widehat{\lambda}, \widehat{\delta}, \widehat{a}) - \bar{y} & = 0, \\ m_2(\widehat{\lambda}, \widehat{\delta}, \widehat{a}) - \frac{1}{n}\sum y_i^2 & = 0, \\ m_3(\widehat{\lambda}, \widehat{\delta}, \widehat{a}) - \frac{1}{n}\sum y_i^3 & = 0. \end{cases}$$

SEs are obtained from the diagonal of a sandwich estimate of the asymptotic variance, $n\widehat{\boldsymbol{G}}^{\top}\widehat{\boldsymbol{V}}^{-1}\widehat{\boldsymbol{G}}^{-1}$, where $\widehat{\boldsymbol{G}}$ is a consistent estimator of the expected value of the gradient associated with the moment conditions and $\widehat{\boldsymbol{V}}$ consistently estimates the associated expected Hessian (Cameron and Trivedi, 1998, Sec. 2.8.4).

3.3 Score test

A score test for H_0: $\delta = 0$ is based on fitting the Poisson model, which is the distribution under the null hypothesis. Define $\nu(\lambda, \delta, a) \equiv \operatorname{Var}(Z)$ and $\rho(\lambda, \delta, a) \equiv \operatorname{Corr}(Y, Z)$. Evaluation under H_0, in which case $\widehat{\lambda} = \bar{y}$ and $\delta = 0$, is represented by a subscript zero (0). The resulting test statistic is stated as

$$S = \frac{n(\bar{z} - \kappa_0)^2}{\nu_0(1 - \rho_0^2)}, \tag{5}$$

which is asymptotically $\chi^2(1)$ distributed under H_0. We do not provide an explicit proof since this is a special case of the general result derived at the end of Section 4.1.

3.4 Visualization

Basic visualization of an observed versus a theoretical (or fitted) distribution of counts can be readily achieved by barplots or, preferably, in the form of a rootogram. The rootogram can be adapted to regression settings as seen in Section 4.3. Another visualization aimed at assessing the goodness of fit is the Touchardness plot which compares deviations between observed data and the Touchard model. These two plots are now described.

A clever visualization tool for count data is the Tukey hanging rootogram. It is a variation of the histogram with the vertical axis showing the square root of the frequencies to de-emphasize outlying values and right skewness (common for count data). The usual bars "hang" from the fitted values so that the discrepancies are visualized against a straight line (the axis) rather than against a curve. The bars are drawn from $\sqrt{\widehat{y}}$ to $\sqrt{\widehat{y}} - \sqrt{y}$. A generalization of the rootogram for regression is shown in Figure 4 (right) for counts of crab satellites predicted by color and weight.

Hoaglin and Tukey (2009) devised a goodness-of-fit plot for count data based on the count metameter. The key idea is to compare the observed frequency of count y, denoted hereafter by n_y, with the expected frequency from a given model. Assuming a Tou(λ, δ, a) model and n data points, the count metameter, $\varphi(n_y)$, is obtained through the equation $n_y = nf(y, \lambda, \delta, a)$ which yields the expression given by

$$\varphi(n_y) \equiv \log\left(\frac{y! n_y}{n(y+a)^{\delta}}\right) = -\log[\tau(\lambda, \delta, a)] + \log(\lambda) y. \tag{6}$$

Deviations between the observed counts and the theoretical model are assessed by examining the relation $\varphi(n_y) \times y$. If the points $(\varphi(n_y), y)$ follow a straight line with exp(slope) close to $\widehat{\lambda}$, we have indication that the Touchard model is appropriate. In practice, δ is fixed at $\widehat{\delta}$ in order to compute $\varphi(n_y)$ and the intercept and slope are obtained by least squares. In addition, Hoaglin and Tukey (2009) proposed an approximate confidence interval for the logarithm of the theoretical frequency. Figure 3 shows the Touchardness plot for data on counts of crab satellites, indicating an adequate fit.

We can obtain starting values $(\lambda^0, \delta^0, a^0)$ for numerical procedures associated with ML and MM based on Equation (6). Consider a in a grid, say $a \in \{0.1, 0.5, 1, 1.5, \ldots, a_{\max}\}$, where $a_{\max}$ is arbitrarily defined. One idea is to fit the linear model stated as

$$\log\left(\frac{y! n_y}{n}\right) = \beta_0 + \beta_1 y + \beta_2 \log(y + a)$$

and set $\lambda^0 = \exp(\widehat{\beta}_1)$, $\delta^0 = \widehat{\beta}_2$ and a^0 yielding the fit with the smallest sum of squared errors. Alternatively, one may re-interpret the model given in Equation (6) as a Poisson log-linear model with $\log(n/y!)$ as offset,

$$\log(n_y) = \beta_0 + \beta_1 y + \beta_2 \log(y + a) + \log(n/y!),$$

and again set $\lambda^0 = \exp(\widehat{\beta}_1)$, $\delta^0 = \widehat{\beta}_2$ and a^0 yielding the fit with the smallest deviance.

3.5 Illustration with selected datasets

Next, we illustrate the Touchard model and associated tools with univariate, public available, data of actuarial end economic interest. We consider the five models listed in Table 1. Except for the NB, the other models handle both under and overdispersion. The QP and the Touchard models are pointwise dual (in the sense that the entire range of δ accounts

for either over or underdispersion of the same magnitude) whereas the CMP is pointwise dual only in the region $0 < \nu < 2$ (Kokonendji et al., 2008). The QP, CMP and NB are members of the EF (thus enjoying the desirable asymptotic results for ML estimation), though only the QP is in the one-dimensional EF with dispersion. The support of the GP distribution depends on the parameter α and it violates standard conditions for consistency and asymptotic normality of ML estimators (Cameron and Trivedi, 1998).

We have fitted different count models, including the Touchard one, to 14 publicly available datasets. The datasets were not selected with any particular criteria other than being publicly available and having been used by the authors for class illustrations, mostly for audiences from Actuarial Sciences and Economics. Most cases exhibit high inflation of zeros ($\widehat{f}_0 > 0.80$), dispersion index $d = S^2/\bar{y}$ greater than one and large sample size.

Eight datasets refer to the number of traffic accidents in a year in different locations labeled by country name and year such as `Zaire74`, `Belgium58`, etc. Some have become benchmarks in the actuarial literature (Denuit, 1997). Typical of such data, the relative frequency of zeros is very high.

The `CrabSat` dataset features the highest level of overdispersion for the response among the cases studied with d close to 3. The counts are from well-known data on the number of satellites (male crabs gathered around the female attempting to fertilize her eggs) appearing in several textbooks (Agresti, 2013).

`MedVisits` consists of over 5000 counts of doctor visits in the past two weeks for a single-adult (Australian Health Survey 1977-78) and has been analyzed by several authors (see, for example, Cameron and Trivedi (1998) and references therein) in the study of health service utilization and health insurance choice. The sample variance is about twice the sample mean ($d = 2.1$). Zeros and ones correspond to 95% of the observations. These data are also used in Section 4.1 in the context of regression modeling.

The dataset `Strikes` records the number of outbreaks of strikes in the UK, in a 4-week period, during 1948-59 for the coal mining industry.

`Shells` brings the number of accidents in the manufacture of high-explosive shells in a British military factory at the time of World War I.

The dataset `Bids` contains the number of takeover bids received by 126 U.S. firms that were targets of tender offers, over a 52-week period following the initial bid.

`AZCardio` contains close to 2000 observations from the 1991 Arizona cardiovascular patient files. The counts refer to the length (days beyond the day of admission) of hospital stay (restricted to less than 9 days) for cardiovascular patients.

Table 2 summarizes the results from fitting five statistical models to the 14 datasets previously described. the Akaike information criterion (AIC) is reported along with data summaries. Qualitatively, similar results were obtained using the χ^2 metric (not shown) instead of the AIC. The Touchard model yields is among the best models all cases with the exception of `Bids` for which convergence was not achieved; these data seem to be well fit by the Poisson model, meaning that a three-parameter distribution is unnecessary.

In Section 4.1, we revisit the CrabSat data for a regression illustration. We thus provide more detail regarding its univariate mode fit. Parameter estimates under the Touchard model from ML and MM methods are shown in Table 3 and visualization of model fit by ML is depicted in Figure 3 showing an adequate fit. As expected, both the score and the likelihood ratio tests (not shown) strongly reject $\delta = 0$. The AIC values for different models are: 746 (Touchard), 771 (NB), 774 (CMP) and 782 (GP). The results place the Touchard as the primary candidate for modeling these data. We must note that some differences in AIC values are small for different models and that mere minimization of a given information criterion is not by itself a deal breaker in model selection. Several other points may be assessed including study domain knowledge and substantive interpretability of models. We remind the reader of the criteria (i)–(ix) listed in the Introduction. The main point to be

taken here is that this superficial examination of the applicability of the Touchard model across datasets from different domains strongly indicates that the Touchard model (and weighted Poisson models in general) are an important addition to the toolbox of count data models. Recall that the Touchard can address both under- and over-dispersion.

Table 2. Data summaries and AIC with selected count models fitted by ML to 14 datasets.

					AIC by model				
Dataset	n	$y_{\max}$	d	$\widehat{f}_0$	Poisson	NB	Tou	CMP	GP
`CrabSat`	173	15	2.9	0.36	990	771	746	774	782
`MedVisits`	5,190	9	2.1	0.80	7,968	7,176	7,125	7,320	7,156
`Shells`	647	5	1.5	0.69	1,236	1,189	1,189	1,189	1,189
`AZCardio`	1,982	7	1.5	0.07	9,055	8,796	8,726	8,753	8,807
`Zaire74`	4,000	5	1.4	0.93	2,494	2,372	2,371	2,419	2,371
`Belgium58`	9,461	7	1.4	0.83	10,983	10,700	10,691	10,713	10,696
`Switz61`	119,853	6	1.2	0.88	109,225	109,234	109,226	109,235	109,230
`Bids`	126	10	1.2	0.07	405	406	NA	407	406
`NewYork93`	365	8	1.2	0.06	1,450	1,449	1,461	1,449	1,449
`Belgium75`	106,974	4	1.1	0.91	72,379	72,212	72,213	72,213	72,212
`Belgium93`	63,299	4	1.1	0.90	44,303	44,133	44,130	44,134	44,131
`Belgium94`	131,182	4	1.1	0.90	90,453	90,163	90,162	90,164	90,162
`Germany60`	23,589	6	1.1	0.87	20,598	20,451	20,449	20,451	20,450
`Strikes`	156	4	0.7	0.29	386	388	381	380	382

Table 3. Estimates (with SEs) from fitting the Tou(λ, δ, a) model to the counts in the `CrabSat` dataset.

	Method	
	ML	MM
λ	8.0 (1.2)	11.4 (3.3)
δ	-2.7 (0.83)	-5.6 (2.6)
a	0.4 (0.22)	1.3 (0.80)

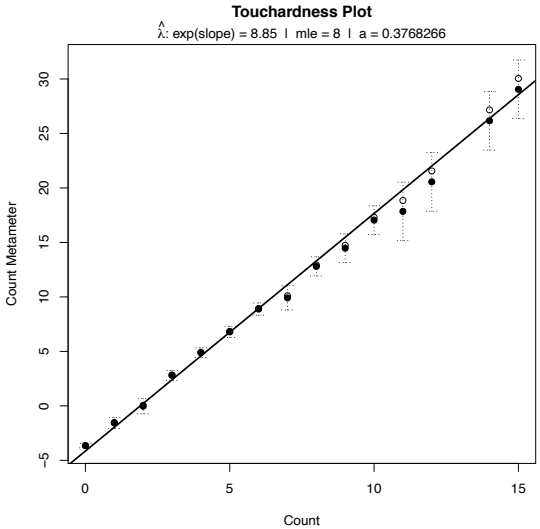

Figure 3. Goodness-of-fit (Touchardness) plots with the `CrabSat` dataset: circles are the observed count metameters; filled points show the CIs centers.

4. Regression modeling

As mentioned in the Introduction, there are several probabilistic models for count data but most have had limited applicability. Models for which regression tools are not readily available will not be considered by most data analysts. In this section we develop regression tools for the Touchard model, including ML estimation, large-sample inference, prediction, visualization and basic diagnostic tools.

Regression for count data is typically based on either the Poisson or the NB models. The (quasi) Poisson model with a dispersion parameter $\varphi \in \mathrm{R}$, $\mathrm{QP}(\mu, \varphi)$, with expectation μ and variance $\varphi\mu$ can accommodate under/overdispersed data. QP regression is a GLM with a vast literature on estimation, inference and diagnostics. In the $\mathrm{NB}(\mu, \alpha)$ model, the variance is a quadratic function of the mean, $\mu + \alpha\mu^2$, $\alpha \geq 0$, which can handle overdispersed data. The $\mathrm{GP}(\mu, \alpha)$ distribution is another model which can be explicitly parametrized in terms of its mean and an extra parameter with the variance being a cubic function of μ. The QP model is a member of the one-dimensional EF with dispersion but neither the NB nor the GP is a member of the EF. The Touchard distribution is in three-dimensional EF but cannot be formulated in terms of the one-dimensional EF with dispersion. We recall that mean-variance relations for these models are reported in Table 1.

The first question posed by any WP regression model is the choice of parametrization for the systematic component. Given observed responses $Y_i \sim \mathrm{Tou}(\lambda_i, \delta, a)$ and a vector of predictors $\boldsymbol{x}_i \in \mathrm{R}^p$, for $i = 1, \ldots, n$, we can postulate that either. (We restrict our attention to the default log link but other links for count data could be considered.) Then, we have

$$\log(\lambda_i) = \boldsymbol{x}_i^\top \boldsymbol{\alpha}, \tag{7}$$

or

$$\log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}. \tag{8}$$

We call the model given in Equation (7) is a direct regression model, which is computationally more convenient. This is the strategy adopted by Sellers and Shmueli (2010) in the context of the CMP regression model. In the GLM-type model given by Equation (8), λ_i must be treated as an implicit function of μ_i, δ and a which implies a computational cost related to solving Equation (3) for λ at every evaluation of the likelihood. We propose a quasi-likelihood approach to deal with the usual practice of linking the mean response to a linear predictor as in Equation (8) in the context of Touchard regression.

4.1 Direct regression model

In this section we assume that λ is a function of the linear predictor $\boldsymbol{x}_i\boldsymbol{\alpha}$, as in Equation (7). The model log-likelihood for independent data points is given by $\ell(\boldsymbol{\alpha}, \delta, a) = \sum_i \ell_i$, where

$$\ell_i = y_i \boldsymbol{x}_i^\top \boldsymbol{\alpha} + \delta \log(y_i + a) - \log\Big[\tau(\exp(\boldsymbol{x}_i^\top \boldsymbol{\alpha}), \delta, a)\Big] - \log(y_i!).$$

Recall that we have defined $z_i = \log(y_i + a)$, $w_{1i} = (y_i + a)^{-1}$ and $w_{2i} = (y_i + a)^{-2}$. We also define $\kappa_i = \mathbb{E}(Z_i)$, $\gamma_i = \mathrm{Cov}(Y_i, Z_i)$, $\theta_{ji} = \mathbb{E}(W_{ji})$ and $\nu_i = \mathrm{Var}(Z_i)$. Denoting by $\boldsymbol{X}_{n\times p}$ the model matrix with i-th row $\boldsymbol{x}_i^\top$, the observed responses by the n-dimensional vector $\boldsymbol{y}$ with

$\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{y})$ and $\boldsymbol{V} = \text{diag}(\sigma_i^2)$, we can write the score vector as

$$\boldsymbol{s}(\boldsymbol{\alpha}, \delta, a) = \begin{pmatrix} \boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{\mu}) \\ \sum_i (z_i - \kappa_i) \\ \delta \sum_i (w_{1i} - \theta_{1i}) \end{pmatrix} \tag{9}$$

and the Fisher information matrix is formulated as

$$\boldsymbol{F}(\boldsymbol{\alpha}, \delta, a) = -\begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{V} \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{\gamma} & \boldsymbol{X}^\top \boldsymbol{T} \boldsymbol{X} \\ & \sum_i \nu_i & \sum_i [\delta \,\text{Cov}(Z_i, W_{1i}) - (\theta_{1i} - W_{1i})] \\ & & \delta^2 \sum_i \text{Var}(W_{1i}) + \delta[\sum_i (W_{2i} - \theta_{2i})] \end{pmatrix}$$

where $\boldsymbol{T} = \text{diag}(t_i)$ with $t_i = 1 - \tau_1(\lambda_i, \delta, a)\tau_{-1}(\lambda_i, \delta, a)$. The usual inference based on the ML estimates uses asymptotic normality with

$$\widehat{\text{Var}}\left(\widehat{\boldsymbol{\alpha}}, \widehat{\delta}, \widehat{a}\right) = \widehat{\boldsymbol{F}}^{-1}.$$

The effect of the coefficients on the mean response is expressed as

$$\frac{\partial \mu_i}{\partial x_{ij}} = \sigma_i^2 \alpha_j. \tag{10}$$

Therefore, the marginal effect of α_j on μ_i is affected by the variability associated with the corresponding observation. Thus, over- or under-dispersion play a role that did not exist in the canonical Poisson model for which $\partial \mu_i / \partial x_{ij} = \mu_i \alpha_j$.

It is possible to obtain a score test for Poissoness. Calculations are facilitated by the fact that residuals and regressors are orthogonal, that is, $\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{0}$, when the score given in Equation (9) is zero. The score test allows us to test the null H_0: $\delta = 0$ without having to first fit the Touchard model.

Let $\boldsymbol{\alpha}_0$ be the coefficient estimates from H_0 (canonical Poisson GLM). We use the same notation as in Section 3.3, where the subscript 0 indicates evaluation under H_0. Here, $\boldsymbol{\lambda}_0 = \exp(\boldsymbol{X}\boldsymbol{\alpha}_0)$ and the score vector given by Equation (9) becomes

$$\boldsymbol{s}_0 = (0, \ldots, 0, K_0),$$

where $K_0 = \sum_i [z_i - \kappa_{i,0}]$. By partitioning $\boldsymbol{F}_0$ and writing its inverse (Graybill, 1983, Ch. 8), we have $F_{0,22}^{-1} = (F_{0,22} - \boldsymbol{F}_{0,21}\boldsymbol{F}_{0,11}^{-1}\boldsymbol{F}_{0,12})^{-1}$ and, after some algebra, the score statistic $\boldsymbol{S} = \boldsymbol{s}_0^\top \boldsymbol{F}_0^{-1} \boldsymbol{s}_0$ reduces to

$$\boldsymbol{S} = \frac{K_0^2}{F_{0,22}^{-1}} = \frac{K_0^2}{\boldsymbol{V}_0 - \boldsymbol{\gamma}_0^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{V}_0 \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\gamma}_0},$$

where $\boldsymbol{V}_0 = \sum_i \nu_{i,0}$. The special case with no covariates was given by Equation (5). The asymptotic distribution of S under H_0 is $\chi^2(1)$ given regularity conditions warranted by the EF.

4.2 Quasi-Poisson GLM with Touchard variance

As mentioned before, a regression model based on Equation (8) is computationally more demanding since λ_i must be treated as an implicit function of μ_i, δ and a. Derivative calculations become a lot more involved. However, the coefficients in the usual (log-mean) parametrization are more easily interpreted since, as opposed to Equation (10), we have that

$$\frac{\partial \mu_i}{\partial x_{ij}} = \mu_i \beta_j,$$

so that β_j can be interpreted as a semi-elasticity, the proportionate change in the mean when x_j changes by one unit, all else constant.

A solution is available by the well established quasi-likelihood approach. The QP(μ, φ) regression estimates, $\widehat{\boldsymbol{\beta}}$, are robust to distributional assumptions in the sense that the Poisson model is used simply to motivate the estimating equations, and that $\widehat{\boldsymbol{\beta}}$ is consistent as long as the link function and linear predictor are correctly specified (Agresti, 2015, Ch. 8). The choice of $v(\mu, \varphi) = \varphi\mu$ corresponds to assuming an EF with dispersion but other forms of conditional variance may also be entertained (Cameron and Trivedi, 1998, Ch. 3) such as with the Touchard variance $v(\mu, \delta, a)$ given by Equation (4). We denote this model by QPT(μ, δ, a). The estimated regression coefficients are simply those from a Poisson GLM, $\widehat{\boldsymbol{\beta}}$, whereas the other parameters are estimated by the MM, obtained by numerically solving the system in (δ, a) stated as

$$\begin{cases} \sum_{i=1}^{n} v(\widehat{\mu}_i, \delta, a) &= \sum_{i=1}^{n} (y_i - \widehat{\mu}_i)^2, \\ \sum_{i=1}^{n} \kappa(\widehat{\mu}_i, \delta, a) &= \sum_{i=1}^{n} z_i. \end{cases}$$

If the variance is believed to be correctly specified (in which case $\widehat{\boldsymbol{\beta}}$ is asymptotically efficient among estimators that are locally linear in $\boldsymbol{y}$) then

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}) = \frac{n}{n-p} (\boldsymbol{X}^\top \widehat{\boldsymbol{M}} \boldsymbol{X})^{-1} (\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X}) (\boldsymbol{X}^\top \widehat{\boldsymbol{M}} \boldsymbol{X})^{-1},$$

where $\widehat{\boldsymbol{M}}$ is diagonal with typical element $\widehat{\mu}_i$ and $\widehat{\boldsymbol{V}}$ is diagonal with typical element $v(\widehat{\mu}_i, \widehat{\delta}, \widehat{a})$.

4.3 Diagnostics

This section develops standard diagnostic tools to assess the fit of a Touchard regression model including an adaptation of the rootogram. We denote the predicted mean by $\widehat{\mu}$. In the case of Equation (7), then $\widehat{\mu}_i = \exp(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}})$. In the case of Equation (8), we have $\widehat{\mu}_i = \tau_1(\widehat{\lambda}_i, \widehat{\delta}, \widehat{a}) - \widehat{a}$ with $\widehat{\lambda}_i = \exp(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\alpha}})$. For the predicted λ we develop the following notation: we write $\lambda = \Lambda(\mu, \delta, a)$ for λ satisfying Equation (3) with given values of μ, δ and a. The most intuitive residual measure in regression models is the Pearson residual given by

$$r_i = \frac{y_i - \widehat{\mu}_i}{\widehat{\sigma}_i}.$$

Pearson residuals may be visualized by a Q-Q normal plot of standardized r_i with a simulated envelope (Atkinson, 1985).

We use the usual definition of deviance in GLMs, $D(\boldsymbol{y};\boldsymbol{\mu}) = 2(\tilde{\ell} - \widehat{\ell})$, to define an approximate measure of deviance where we consider δ and a fixed (at estimated values). Here, $\widehat{\ell}$ denotes the maximized log-likelihood and $\tilde{\ell}$ is the saturated log-likelihood. Saturation is achieved by setting $\mu_i = y_i$ since $\partial \ell_i/\partial \mu_i = 0 \Leftrightarrow (y_i - \mu_i)/\sigma_i^2 = 0$. As a result, deviance residuals, d_i, are defined by the signed square root of the components of $D(\boldsymbol{y};\boldsymbol{\mu}) = \sum_1^n d_i^2$, as

$$d_i = \begin{cases} \operatorname{sign}(y_i - \widehat{\mu}_i)\sqrt{2\left[y_i \log\left(\frac{\tilde{\lambda}_i}{\widehat{\lambda}_i}\right) - \log\left(\frac{\tilde{\tau}_i}{\widehat{\tau}_i}\right)\right]}, & y_i > 0, \\ \operatorname{sign}(y_i - \widehat{\mu}_i)\sqrt{2[\log(\widehat{\tau}_i) - \delta \log(\widehat{a})]}, & y_i = 0, \end{cases}$$

where $\tilde{\lambda}_i = \Lambda(y_i, \widehat{\delta}, \widehat{a})$ and $\widehat{\lambda}_i$ is either $\exp(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\alpha}})$ or $\Lambda(\exp(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}), \widehat{\delta}, \widehat{a})$.

Observations with a large absolute value of either r_i or d_i are viewed as discrepant. Detectable patterns in the plot of residuals against the estimated linear predictor are indicative of misspecification.

Another diagnostic measure is the generalized leverage defined as the diagonal of $\boldsymbol{L}_{n\times n}(\boldsymbol{\theta}) = \partial \widehat{\boldsymbol{y}}/\partial \boldsymbol{y}^\top$. The actual computation of $\boldsymbol{L}$ is based on first derivatives of the mean vector, $\dot{\boldsymbol{\mu}}_{\boldsymbol{\theta}}$, and second derivatives of the likelihood, $\ddot{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ and $\ddot{\ell}_{\boldsymbol{\theta}\boldsymbol{y}}$, with

$$\boldsymbol{L} = \dot{\boldsymbol{\mu}}_{\boldsymbol{\theta}}(-\ddot{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}})^{-1}\ddot{\ell}_{\boldsymbol{\theta}\boldsymbol{y}},$$

evaluated at $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}, \widehat{\delta}, \widehat{a})$ or $(\widehat{\boldsymbol{\beta}}, \widehat{\delta}, \widehat{a})$ (Wei et al., 1998).

In the regression based on Equation (7), the computation of $\boldsymbol{L}$ yields the expression stated as

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{V}\boldsymbol{X} & \boldsymbol{\gamma} & \boldsymbol{\eta} \end{pmatrix} \boldsymbol{F}^{-1} \begin{pmatrix} \boldsymbol{X}^\top \\ \boldsymbol{a}_1^\top \\ \boldsymbol{a}_2^\top \end{pmatrix},$$

where the vector $\boldsymbol{\eta}$ has i-th component $\delta \operatorname{Cov}(Y_i, W_{1i})$, $\boldsymbol{a}_1$ has i-th component $1/(y_i + a)$ and $\boldsymbol{a}_2$ has i-th component $-\delta/(y_i + a)^2$. In the GLM-type model based on Equation (8), the computation of $\boldsymbol{L}$ yields the usual projection matrix of a Poisson GLM with the weights adjusted for the Touchard variance, that is, we get

$$\boldsymbol{L} = \boldsymbol{W}^{-1/2}\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{W}^{-1/2}.$$

where $\boldsymbol{W}_{n\times n}$ is diagonal with typical element μ_i^2/σ_i^2.

With the diagonal of $\boldsymbol{L}$ and the Pearson residuals, one can compute the approximate Cook distance given by

$$C_i = \frac{\boldsymbol{L}_{ii} r_i^2}{p(1 - \boldsymbol{L}_{ii})^2},$$

to measure the squared distance between $\widehat{\boldsymbol{\alpha}}$ (or $\widehat{\boldsymbol{\beta}}$) and the same estimate without i-th observation (Cook and Weisberg, 1982).

The extension of the rootogram (Section 3.4) to regression models has been proposed by Kleiber and Zeileis (2016) as a complement to residual diagnostics in order to visualize important features of count data such as dispersion, skewness, zero inflation and multimodality,

vis a vis fitted models.

Given regression estimates $\widehat{\boldsymbol{\alpha}}$ (or $\widehat{\boldsymbol{\beta}}$), $\widehat{\delta}$ and $\widehat{a}$, one obtains $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_n$ and the expected frequency of count y stated as

$$E_y = \sum_{i=1}^{n} f(y, \widehat{\lambda}_i, \widehat{\delta}, \widehat{a}), \quad j = 0, 1, \ldots$$

A rootogram can now be drawn with bars from $\sqrt{E_y} - \sqrt{n_y}$ up to $\sqrt{E_y}$, to visually assess the goodness of fit provided by the Touchard regression model. Discrepancies between the observed frequencies n_y and the E_y are visualized against a straight line; Figure 4 (right).

4.4 Example: Crab satellites

As an illustration, we fit the Touchard regression models along with other count regression models to the well-known data on the number of satellites (male crabs gathered around the female attempting to fertilize her eggs). For predicting the number of satellites y, we consider `weight` (kg) and `color` (with two categories, light and dark (baseline)). Therefore, the linear predictor is formulated as $\beta_0 + \beta_1$`weight` $+ \beta_2$`color`.

Table 4. Estimation results (with SEs) from different models fitted by ML to the `CrabSat` dataset ($n = 173$, $n_0 = 62$, $k = 2$, $y_{\max} = 15$).

	$\log(\mu)$ = linear predictor				$\log(\lambda)$ = linear predictor		
	QP	QPT	NB	GP	Tou	CMP	
intercept	-0.49	-0.49	-0.93	-1.11	1.34	-0.58	
	(0.32)	(0.29)	(0.40)	(0.48)	(0.22)	(0.08)	
weight	0.54	0.54	0.71	0.63	0.21	0.13	
	(0.12)	(0.11)	(0.16)	(0.20)	(0.05)	(0.04)	
color	0.27	0.27	0.29	0.37	0.09	0.10	
	(0.18)	(0.18)	(0.20)	(0.06)	(0.06)	(0.06)	
dispersion	$\widehat{\varphi} = 3.15$	$\widehat{\delta} = -3.99$, $\widehat{a} = 0.83$	$\widehat{\alpha} = 0.96$	$\widehat{\alpha} = 0.34$	$\widehat{\delta} = -2.05$, $\widehat{a} = 0.25$	$\widehat{\nu} = 0.08$	(*)
AIC	738*	725†	754	719	758	747	
SSR_{raw}	1,522	1,522	1,692	1,882	1,480	3,773	
SSR_{pear}	536	175	155	166	185	11,258	
SSR_{dev}	553	214	197	NA	215	602	
$\widehat{n}_0$	40	56	50	45	63	42	
AME_{wei}	1.6	1.6	2.1	2.4	1.7	0.4	
IRR_{wei}	1.7	1.7	2.0	1.9	NA	NA	

Approximate QP(μ, φ) log-likelihood (Nelder and Pregibon, 1987).
(†) Touchard log-likelihood evaluated at QPT estimates.
NA = not available.

Estimation results are shown in Table 4. The estimated coefficients from the Touchard GLM-type regression are very close to those from QP. The value of a was chosen to yield the highest likelihood. The two Touchard models yield similar results with significant improvement over the GP and CMP as seen by different sums of residuals, log-likelihood and estimated number of zeros. The Touchard models yield closer estimates of the proportion of zeros than the other models considered. The Touchard regression models provide the estimates (56 and 63) closest to $n_0 = 62$ among the models considered. The sum of deviance residuals for the NB model is the lowest despite the higher likelihood achieved by

the Touchard models. The sums of residuals estimated by the CMP is much higher than those predicted by the other models which probably explains the huge values for the raw and Pearson residuals. Rootograms are shown in Figure 4 indicating some misfitting for the Touchard but noticeable improvement over the fit provided by the NB regression model.

All models agree qualitatively in terms of lack of marginal significance of `color` and a considerable effect of `weight`. However, different models give different effect sizes for `weight`. Marginal effects in count regression are often reported either as the average marginal effect (AME) or as the relative change in the conditional mean (incidence rate ratio - IRR). For istance, the AME associated with the predictor `weight` is given by

$$\text{AME}_{\text{wei}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mu_i}{\partial x_{i1}},$$

where $x_1 \equiv$ weight. The AME_{wei} for models based on $\log(\mu_i)$ is simply $(1/n)\sum_i \widehat{\mu}_i\widehat{\beta}_1$. Therefore the AMEs for the GLM-type models are 1.59 (QP and QPT), 2.12 (NB) and 2.42 (GP). For both the direct Touchard model and the CMP model, which are based on $\log(\lambda_i)$, we have $\text{Var}(Y) = \partial\mu/\partial\log(\lambda)$. Thus, $\text{AME}_{wei} = (1/n)\sum_i \widehat{\sigma}_i^2\widehat{\alpha}_1$ and which amounts to 1.73 (Touchard) and 0.36 (CMP). Taking a unit change in `weight` as the variation of interest, the marginal effect in terms of IRR is stated as

$$\text{IRR}_{\text{wei}} = \frac{\mathbb{E}(y|\text{weight}+1, \text{color})}{\mathbb{E}(y|\text{weight}, \text{color})}.$$

For the GLM-type models, we have $\text{IRR}_{\text{wei}} = \exp(\beta_1)$ with values of 1.72 (QP and QPT), 1.65 (Tou), 2.03 (NB) and 1.88 (GP). For the direct models based on $\log(\lambda)$ the IRR does not reduce to a simple expression, varying across observations and values of `color`. Sellers and Shmueli (2010) suggest dividing the CMP coefficients by $\widehat{\nu}$ as a crude approximation for comparison with Poisson coefficients. For the coefficient of `weight` this yields $0.13/0.08 = 1.63$ which is much higher than the Poisson estimate of 0.54.

Estimated marginal effects are thus higher under the NB and GP models and strikingly lower under the CMP model.

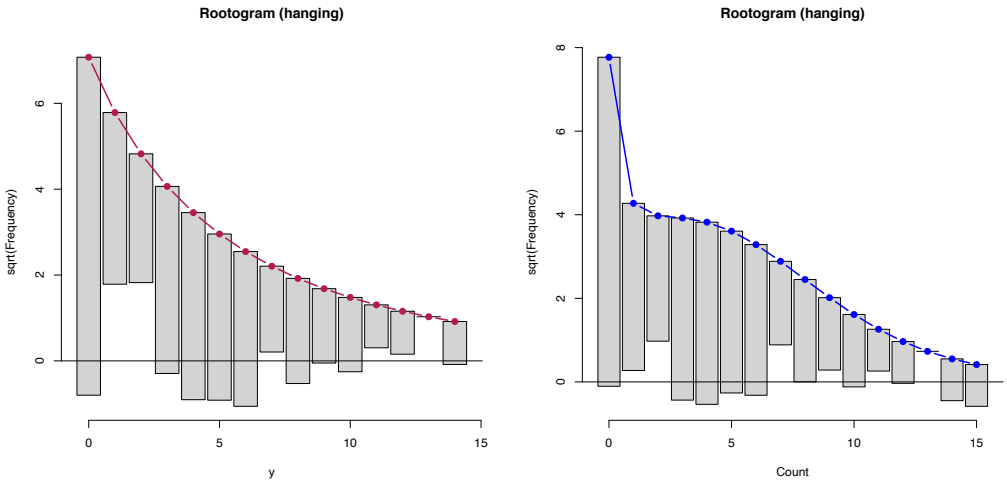

Figure 4. Rootgram associated with NB regression model (left) and Touchard GLM-type model (right) with `CrabSat` dataset.

5. Concluding remarks

We have provided several tools for analyzing count data with a flexible weighted Poisson distribution (Touchard) including regression modeling. We have concluded that the Touchard is a viable and flexible alternative to model over or underdispersed count data. Data analyses presented here and in Matsushita et al. (2018) show that the Touchard is a competitive alternative to traditional models within the exponential family. The statistical tools developed here are based on classical methods including maximum likelihood, method of moments and quasi-likelihood. In terms of tractability, elegance and numerical implementation, the Touchard model is more flexible than the negative binomial and the Conway-Maxwell-Poisson, besides many other count models for which similar tools are not yet available. A major advantage of using the weighted Poisson with $(\lambda, \delta; \log(y+a))$ model is that one does not need to switch between highly different models for different datasets.

Regression modeling has been presented in terms of two variants and in the form of quasi-Poisson estimation with the Touchard variance. Direct modeling of $\log(\lambda)$ in terms of a linear predictor has been developed. When mean prediction and interpretability are wanted, regression based on $\log(\mu)$ is available in the form of a GLM-type model. The QPT methodology overcomes the computational burden associated with the GLM-type model and is a viable choice for large datasets and when directly modeling is either too slow computationally or the related maximization is unstable. Future work may investigate quantile-based regression and modeling of δ in terms of covariates. Bayesian modeling based on the Touchard model is also open for research.

Appendix A. Further results on the normalizing constant

Before providing new results, we state, for mere completion, the following result proved by del Castillo and Pérez-Casany (1998).

Theorem A.1 The series $\tau(\lambda, \delta, a) = \sum_{i=0}^{y} [\lambda^y (y+a)^\delta]/y!$ converges for $\lambda, a > 0$ and $\delta \in \mathrm{R}$.

We now provide asymptotic expressions for τ and for the first two moments of $Y \sim \mathrm{Tou}(\lambda, \delta, a)$.

Theorem A.2 To first-order, the following approximations hold:

$$\tau \approx \exp(\lambda)(\lambda + a)^\delta,$$

$$\mu \approx \lambda + \frac{\lambda\delta}{\lambda + a}, \tag{A1}$$

and

$$\sigma^2 \approx \lambda\left(1 + \frac{a\delta}{(\lambda + a)^2}\right). \tag{A2}$$

Proof The first approximation may be obtained by replacing $(y+a)^\delta$ by $(\lambda+a)^\delta + (\lambda + a)^{\delta-1}(y-\lambda) + o(|(y-\lambda|)$ into the series defining $\tau(\lambda, \delta, a)$. Alternatively, it is easy to see

that $\tau = \exp(\lambda)\mathbb{E}(Y^* + a)^\delta$, where $Y^* \sim \text{Poi}(\lambda)$. Therefore, by the Delta method, we have

$$\begin{aligned}\tau &\approx \exp(\lambda)[\mathbb{E}(Y^*) + a]^\delta \\ &= \exp(\lambda)(\lambda + a)^\delta,\end{aligned}$$

Using the above approximation for τ and the fact that $\mu = (\lambda/\tau)(\partial\tau/\partial\lambda)$ (see Appendix C) we obtain the approximation stated in Equation (A1) for μ. Since $\sigma^2 = \lambda(\partial\mu/\partial\lambda)$, we obtain approximation stated in Equation (A2) for σ^2. ■

We observe that: (i) the above approximations are exact for $\delta = 0$ and $\delta = 1$; (ii) they are better the larger λ is relative to $|\delta|$ and the larger a is; (iii) for large λ, $\mu \to \lambda + \delta$ and $\sigma^2 \to \mu - \delta$; see Figure 2.

Having a compact notation for τ in terms of other special functions is of interest to study further properties and analytical characteristics. This has been explored, for example, by Castellares and Lemonte (2019), were a previous diverging series was re-derived in terms of the integro-exponential function to provide a correct converging result for the moments of the generalized Gompertz distribution. The following theorem provides a representation of τ in terms of the generalized hypergeometric function.

THEOREM A.3 The function $\tau(\lambda, \delta, a)$ can be represented in terms of the I generalized hypergeometric function as

$$\tau(\lambda, \delta, a) = I_{1,2}^{1,1}\left[-\lambda \,\middle|\, \begin{matrix}(-a, 1, \delta) \\ (0, 1, 1), (1 - a, 1, \delta)\end{matrix}\right], \tag{A3}$$

in which the I-function (Rathie, 1997) is defined as an contour complex integral which contain powers of Gamma functions in their integrands by

$$\begin{aligned}&I_{p,q}^{m,n}\left[z \,\middle|\, \begin{matrix}(a_1, \alpha_1, A_1), \ldots, (a_n, \alpha_n, A_n), (a_{n+1}, \alpha_{n+1}, A_{n+1}), \ldots, (a_p, \alpha_p, A_p) \\ (b_1, \beta_1, B_1), \ldots, (b_m, \beta_m, B_m), (b_{m+1}, \beta_{m+1}, B_{m+1}), \ldots, (b_q, \beta_q, B_q)\end{matrix}\right] \\ &\quad = \frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m}\Gamma^{B_j}(b_j - \beta_j s)\prod_{j=1}^{n}\Gamma^{A_j}(1 - a_j + \alpha_j s)}{\prod_{j=m+1}^{q}\Gamma^{B_j}(1 - b_j + \beta_j s)\prod_{j=n+1}^{p}\Gamma^{A_j}(a_j - \alpha_j s)} z^s \mathrm{d}s,\end{aligned} \tag{A4}$$

in which α_j, A_j, β_j and B_j are assumed to be positive quantities and all the a_j and b_j are complex such that no singularity of $\Gamma^{B_j}(b_j - \beta_j s)$ coincides with any singularities of $\Gamma^{A_j}(1 - a_j + \alpha_j s)$. In general, these singularities are not poles.

There are three different contours L of integration stated as:

- L goes from from $\sigma - i\infty$ to $\sigma + i\infty$ (σ real) such that all the sigularities of $\Gamma^{B_j}(b_j - \beta_j s)$, $j = 1, \ldots, m$ lie to the right of L and all the singularities of $\Gamma^{A_j}(1 - a_j + \alpha_j s)$, for $j = 1, \ldots, n$, lie to the left of L.
- L is a loop beginning and ending at $+\infty$ and encircling all the singularities of $\Gamma^{B_j}(b_j - \beta_j s)$, for $j = 1, \ldots, m$, once in the clock-wise direction, but none of the singularities of $\Gamma^{A_j}(1 - a_j + \alpha_j s)$, $j = 1, \ldots, n$.
- L is a loop beginning and ending at $-\infty$ and encircling all the singularities of $\Gamma^{A_j}(1 - a_j + \alpha_j s)$, for $j = 1, \ldots, n$, once in the anti-clockwise direction, but none of the singularities of $\Gamma^{B_j}(b_j - \beta_j s)$, for $j = 1, \ldots, m$.

Proof Let one consider the I-function on the right-hand side of Equation (A3) and its

contour integral representation given by Equation (A4) as

$$I_{1,2}^{1,1}\left[-\lambda \,\middle|\, \begin{matrix}(-a,1,\delta)\\(0,1,1),(1-a,1,\delta)\end{matrix}\right] = \frac{1}{2\pi i}\int_L \frac{\Gamma(-s)\Gamma^\delta(1+a+s)}{\Gamma^\delta(a+s)}(-\lambda)^s \mathrm{d}s$$

Since none of the singularities of $\Gamma^\delta(1+a+s)$ coincide with the poles of $\Gamma(-s)$, the simple application of the residue theorem (Springer, 1979) to the poles of the latter imply that

$$\begin{aligned} I_{1,2}^{1,1}\left[-\lambda \,\middle|\, \begin{matrix}(-a,1,\delta)\\(0,1,1),(1-a,1,\delta)\end{matrix}\right] &= \sum_{r=0}^{\infty} \lim_{s\to r} \frac{(-s+r)\Gamma(-s)\Gamma^\delta(1+a+s)}{\Gamma^\delta(a+s)}(-\lambda)^s \\ &= \sum_{r=0}^{\infty} \frac{\Gamma^\delta(1+a+r)(-\lambda)^r}{\Gamma^\delta(a+r)r!(-1)^r} \\ &= \sum_{r=0}^{\infty} \frac{(a+r)^\delta \lambda^r}{r!}, \end{aligned}$$

as desired. ■

Appendix B. Identifiability

By definition (Lehmann and Casella, 1998, Sec. 1.5), for a given statistical model $\mathcal{P} = \{P_\zeta : \zeta \in \mathcal{Z}\}$, where $\mathcal{Z}$ denotes the parameter space, we say $\mathcal{P}$ is identifiable if

$$P_{\zeta_1} = P_{\zeta_2} \implies \zeta_1 = \zeta_2 \quad \forall \zeta_1, \zeta_2 \in \mathcal{Z}.$$

With a fixed, the two-parameter Tou(λ, δ, a) model is clearly identifiable since the statistics Y and $\log(Y+a)$ are linearly independent (Lehmann and Casella, 1998, Sec. 1.5). A more general result with all parameters free is developed next. We begin by defining $\zeta = (\lambda, \delta, a)$ and $f_\zeta(y)$ in place of Equation (1). Thus, we define

$$\mathcal{P} = \left\{ f_\zeta(y) = \frac{\lambda^y (y+a)^\delta}{y!\tau(\lambda,\delta,a)} : \lambda, a > 0, \delta \in \mathrm{R} \right\}.$$

In order to prove identifiability, we must set $f_{\zeta_1}(y) = f_{\zeta_2}(y), \forall y$. However, we can avoid $\tau(\lambda, \delta, a)$ by working instead with

$$\frac{f_{\zeta_1}(y+1)}{f_{\zeta_1}(y)} = \frac{f_{\zeta_2}(y+1)}{f_{\zeta_2}(y)},$$

which reduces to

$$\frac{\lambda_1}{(y+1)}\left(\frac{y+1+a_1}{y+a_1}\right)^{\delta_1} = \frac{\lambda_2}{(y+1)}\left(\frac{y+1+a_2}{y+a_2}\right)^{\delta_2}.$$

In order to compare both sides of Equation (B), let us consider the series representation of the function $h(\delta, a, x) = ((x+a)/(x+1+a))^\delta$, which is can be seen as the product of the functions $h_a(\delta, a, x) = (x+a)^\delta$ and $h_b(\delta, a, x) = (x+1+a)^{-\delta}$. Thus, the MacLaurin Series

of $h(\delta, a, x)$ in terms of x is obtained as

$$h(\delta, a, x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \frac{\partial^n}{\partial x^n} (h_a(\delta, a, x) h_b(\delta, a, x)) \Big|_{x=0}. \tag{B1}$$

The generalized Leibiniz rule states that

$$\frac{\partial^n}{\partial x^n} h_a(\delta, a, x) h_b(\delta, a, x) = \sum_{y=0}^{n} \frac{n!}{(n-y)!k!} \frac{\partial^{n-y}}{\partial x^{n-y}} h_a(\delta, a, x) \frac{\partial^y}{\partial x^k} h_b(\delta, a, x). \tag{B2}$$

The derivatives on the right hand side of Equation (B2) are quite simple and result in

$$\frac{\partial^{n-y}}{\partial x^{n-y}} h_a(\delta, a, x) = \frac{\Gamma(\delta+1)}{\Gamma(\delta-n+y+1)} (x+a)^{\delta-n+y} \tag{B3}$$

and

$$\frac{\partial^y}{\partial x^y} h_b(\delta, a, x) = (-1)^y \frac{\Gamma(\delta+y)}{\Gamma(\delta)} (x+1+a)^{-\delta-y}. \tag{B4}$$

By combining Equations (B2)), (B3) and (B4), after some algebra we arrive at

$$\frac{\partial^n}{\partial x^n} h_a(\delta, a, x) h_b(\delta, a, x) = \\ \frac{\Gamma(\delta+1)(x+a)^{\delta-n}(x+1+a)^{-\delta}}{\Gamma(\delta-n+1)} \sum_{y=0}^{n} \binom{n}{y} \frac{(-1)^k \Gamma(\delta+y)\Gamma(\delta-n+1)\left(\frac{x+a}{x+a+1}\right)^y}{\Gamma(\delta)\Gamma(\delta-n+y+1)}. \tag{B5}$$

By definition, the right hand side of Equation (B5) may be expressed in terms of the hypergeometric function ${}_2F_1$ and by means of Equation (B1), we obtain

$$h(\delta, a, x) = \sum_{n=0}^{\infty} \frac{\Gamma(\delta+1) a^{\delta-n}(1+a)^{-\delta}}{n!\Gamma(\delta-n+1)} {}_2F_1\left(-n, \delta; 1-n+\delta; \frac{a}{a+1}\right) x^n. \tag{B6}$$

By applying the ratio test to the series, we get

$$R = \frac{x(a)^{-1}(\delta-n)}{(n+1)} \frac{{}_2F_1\left(-n-1, \delta; -n+\delta; \frac{a}{1+a}\right)}{{}_2F_1\left(-n, \delta; 1-n+\delta; \frac{a}{1+a}\right)}. \tag{B7}$$

In general, the hypergeometric function in Equation (B7)) vanishes for finite n and, in the limiting case, the continued fraction representation for the ratio of hypergeometric functions given by

$$\frac{{}_2F_1(a+1, b; c+1; z)}{{}_2F_1(a, b; c; z)} = \cfrac{1}{1 + \cfrac{\frac{(a-c)b}{c(c+1)} z}{1 + \cfrac{\frac{(b-c-1)(a+1)}{(c+1)(c+2)} z}{1 + \ddots}}} \tag{B8}$$

can be used. By noticing that whenever $(a-c-j)(b+j)$, $\forall j \in \{0,1,2,\dots\}$ is in the numerator of the ratio which multiplies z in Equation (B8), this ratio vanishes to 0 as $n \to \infty$, we get

$$R \xrightarrow[n\to\infty]{} \frac{-x}{(a)}.$$

Thus, Equation (B6) is valid for $x < a$. To account for other values of x, we can define

$$g(\delta, a, \tilde{x}) = \left(\frac{1+a\tilde{x}}{1+(1+a)\tilde{x}}\right)^{\delta},$$

where $\tilde{x} = 1/x$ and back to Equation (B) by following a similar procedure as above, the series representation for $g(\delta, a, \tilde{x})$ can be obtained as

$$g(\delta, a, x^{-1}) = \sum_{n=0}^{\infty} \frac{\Gamma(\delta+1)(1+a)^n}{n!\Gamma(\delta-n+1)} {}_2F_1\left(-n, \delta; 1-n+\delta; \frac{a}{a+1}\right) x^{-n}. \tag{B9}$$

Equation (B9) is valid for $x > a+1$ and it remains to be addressed the case $a < x < 1+a$. For real values of a, since x is a positive integer in our case, the only possible value for x which falls into this interval is $x = \lceil a \rceil$. This special case is treated in the last paragraph. Now, without loss of generality, let $a_1 \leq a_2$. From Equation (B), we conclude that the identifiability problem reduces to the expression given by

$$\begin{cases} \lambda_1 h(\delta_2, a_2, y) = \lambda_2 h(\delta_1, a_1, y), & \text{for } 0 \leq y < a_1, \\ \lambda_1 h(\delta_2, a_2, y) = B_1, & \text{for } a_1 \leq y < \min(a_1+1, a_2), \\ B_2 = B_3, & \text{for } \min(a_1+1, a_2) < y < \max(a_1+1, a_2), \\ B_4 = \lambda_1 g(\delta_1, a_1, 1/y), & \text{for } \max(a_1+1, a_2) < y < a_2+1, \\ \lambda_1 g(\delta_1, a_1, 1/y) = \lambda_2 g(\delta_2, a_2, 1/y), & \text{for } y > a_2+1, \end{cases} \tag{B10}$$

where the functions B_i, $i = 1, 2, 3, 4$, depend on the max and min functions applications. For instance, if $\min(1+a_1, a_2) = 1+a_1$ then $B_1 = \lambda_2((\lceil a_1 \rceil + a_1)/(1+(\lceil a_1 \rceil + a_1)))^{\delta}$. Regarding the first expression in Equation (B10), by performing a term-by-term matching procedure, the first three terms of the series in Equation (B6) indicate that

$$\begin{array}{ll} \text{(i)} & \lambda_1 a_2^{\delta_2}(1+a_2)^{-\delta_2} = \lambda_2 a_1^{\delta_1}(1+a_1)^{-\delta_1}, \\ \text{(ii)} & \lambda_1 \delta_2 a_2^{\delta_2-1}(1+a_2)^{-\delta_2-1} = \lambda_2 \delta_1 a_1^{\delta_1-1}(1+a_1)^{-\delta_1-1}, \\ \text{((iii)} & \frac{\lambda_1}{2}\delta_2 a_2^{\delta_2-2}(1+a_2)^{-\delta_2-2}(-1+\delta_2-2a_2) = \\ & \qquad \frac{\lambda_2}{2}\delta_1 a_1^{\delta_1-2}(1+a_1)^{-\delta_1-2}(-1+\delta_1-2a_1). \end{array} \tag{B11}$$

The quotients $(ii)/(i)$ and $(iii)/(ii)$ imply

$$\begin{array}{ll} \text{(iv)} & \frac{\delta_1}{a_1(1+a_1)} = \frac{\delta_2}{a_2(1+a_2)}, \\ \text{(v)} & \frac{-1+\delta_1-2a_1}{2a_1(1+a_1)} = \frac{-1+\delta_2-2a_2}{2a_2(1+a_2)}. \end{array} \tag{B12}$$

By solving the system in Equation (B12), the two possible solutions for a_1 are $a_1 = a_2$ or $a_1 = -(1+a_2)(1+2a_2)^{-1}$. Obviously, since $a_i \geq 0$, for $i = 1, 2$, the only possible solution is $a_1 = a_2$.

By using such solution back on (iv) of Equation (B12), $\delta_1 = \delta_2$. Thus, (i) of Equation (B11) implies that $\lambda_1 = \lambda_2$.

Now, for the last expression of Equation (B10), the term-by-term series comparison provides

$$\begin{aligned}&\text{(vi)} \quad \lambda_1 = \lambda_2,\\ &\text{(vii)} \quad \lambda_1\delta_1 = \lambda_2\delta_2,\\ &\text{(viii)} \quad \tfrac{\lambda_1}{2}\delta_1(-1+\delta_1-2a_1) = \tfrac{\lambda_2}{2}\delta_2(-1+\delta_2-2a_2).\end{aligned}$$

The system solution is quite straightforward, implying that: $\lambda_1 = \lambda_2$, $\delta_1 = \delta_2$ and $a_1 = a_2$. The remaining cases to be discussed are the other equations which have not been addressed yet in Equation (B10). Let S denote the support of the probability density function of the Touchard distribution. So far, it has been shown that the function is identifiable over $S \setminus \{[a_1, a_2+1)\}$. In order to prove the identifiability over the whole support, it is sufficient to check if the conditions found for $S \setminus \{[a_1, a_2+1)\}$ also work when $y \in [a_1, a_2+1)$. This easily follows by noticing that the identifiability of a probability density function boils down to a system of equations. Therefore, all the equations must be simultaneously satisfied in order to exist a solution. Thus, it has been shown that for $[f_{\zeta_1}(y) = f_{\zeta_2}(y)], \forall y \implies \zeta_1 = \zeta_2$, which proves identifiability when all three parameters are free.

Appendix C. Useful Derivatives

The following expressions were used to obtain several formulas involving derivatives presented in the article. Recall that we have defined $Z = \log(Y+a)$, $W_1 = (Y+a)^{-1}$, $\mu = \mathbb{E}(Y)$, $\kappa = \mathbb{E}(Z)$, $\sigma^2 = \mathrm{Var}(Y)$, $\nu = \mathrm{Var}(Z)$ and $\gamma = \mathrm{Cov}(Y, Z)$. We also define $\kappa_2 = \mathbb{E}[Z^2]$ and $\kappa_3 = \mathbb{E}[YZ]$. We start with first and second derivatives of $\tau(\lambda, \delta)$ given by

$$\frac{\partial\tau}{\partial\lambda} = \frac{\tau\mu}{\lambda}; \quad \frac{\partial\tau}{\partial\delta} = \tau\kappa; \quad \frac{\partial\tau}{\partial a} = \tau(\lambda, \delta-1, a); \frac{\partial^2\tau}{\partial\lambda^2} = \frac{\tau[m_2-\mu]}{\lambda^2}; \quad \frac{\partial^2\tau}{\partial\delta^2} = \tau\kappa_2 \quad \text{and} \quad \frac{\partial^2\tau}{\partial\lambda\partial\delta} = \frac{\tau\kappa_3}{\lambda}.$$

Next, we list some partial derivatives of the expectations $\mu(\lambda, \delta)$ and $\kappa(\lambda, \delta)$ stated as

$$\frac{\partial\mu}{\partial\lambda} = \frac{\sigma^2}{\lambda}; \quad \frac{\partial\mu}{\partial\delta} = \gamma; \quad \frac{\partial\mu}{\partial a} = \delta\,\mathrm{Cov}(Y, W_1), \frac{\partial\kappa}{\partial\lambda} = \frac{\gamma}{\lambda}, \frac{\partial\kappa}{\partial\delta} = \nu.$$

Appendix D. Random number generation (RNG) by the inverse transformation

In order to generate y from $Y \sim \mathrm{Tou}(\lambda, \delta, a)$, we take $F(y) = \sum_{x=0}^{y} f(x)$ and $Y = \min\{y\colon F(y-1) < U \leq F(y), U \sim \mathrm{Unif}(0,1)\}$. We note that the RNG can avoid costly computation of cumulative probabilities and factorial terms by using the ratio $c(y) = f(y)/f(y-1)$.

Pseudo-code for the proposed RNG is shown in Algorithm D and it is implemented for the `R` system (Andrade and Oliveira, 2019).

The expected number of iterations in the while loop in Algorithm D is $1 + \mathbb{E}(Y)$ which, according to results from previous section, is approximately $1 + \lambda[1 + \delta/(\lambda + a)]$ or simply $1 + \lambda + \delta$ for $\lambda \gg \delta$. The generation time increases with λ and δ with δ being a stronger factor than λ (Table D1). The value of a is of least importance for running time. Generation of highly overdispersed data is faster. Efficient generators use multiple schemes taking into consideration the parameter values (Fishman, 2013) and this is still open for research in the context of Touchard RNG.

Pseudo-code for Touchard RNG.
Input: $p_0 = \tau(\lambda, \delta, a)^{-1}$.
Output: $Y \sim \text{Tou}(\lambda, \delta, a)$.
Define: $p_k := f(k; \lambda, \delta, a)$ and $c(k) := \dfrac{p_k}{p_{k-1}},\ k = 1, 2, \ldots$.

`initialization`
$p \leftarrow p_0; q \leftarrow p_0; k \leftarrow 0$
$U \sim \text{Unif}(0, 1)$

`While:` $U > q$
$k \leftarrow k + 1$
$p \leftarrow p \cdot c(k)$
$q \leftarrow q + p$

Return: $Y = k$

Table D1. Time to generate 10^4 values using Algorithm D (in plain C) for different combinations of parameter values with $a = 1$. Times reported in 1/1000 seconds (Intel Pentium dual core i5 1200MHz running Linux Debian).

		δ					
		-10	-5	-1	1	5	10
λ	0.5	2	1	2	2	10	15
	2	2	1	2	3	19	27
	10	1	9	8	10	54	66
	20	1	55	15	18	92	109

Author contributions Conceptualization, investigation, methodology, software, writing-original draft preparation, writing review and editing, writing-original: all the authors participated equally. The authors have read and agreed to the published version of the manuscript.

Acknowledgements The author thanks Editors and Referees for this suggestions which allowed us to improve the presentation of this work.

Funding Not applicable

Conflicts of interest The authors declare no conflict of interest.

References

Agresti, A., 2013. Categorical Data Analysis. Wiley, New York, US.

Agresti, A., 2015. Foundations of Linear and Generalized Linear Models. Wiley, New York, US.

Andrade, B. and Oliveira, S., 2019. touchard: Touchard Model and Regression. R package version 2.0.0.

Atkinson, A., 1985. Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Clarendon Press, Oxford, UK.

Cameron, A.C. and Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, MA, US.

Castellares, F. and Lemonte, A.J., 2019. On the moments of the generalized gompertz distribution. Applied Mathematical Modelling, 72, 420 – 424.

Chakraborty, S. and Imoto, T., 2016. Extended Conway-Maxwell-Poisson distribution and its properties and applications. Journal of Statistical Distributions and Applications, 3, 1–19.

Cook, R.D. and Weisberg, S., 1982. Residuals and Influence in Regression. Chapman and Hall, New York, US.

del Castillo, J. and Pérez-Casany, M., 1998. Weighted poisson distributions for overdispersion and underdispersion situations. Annals of the Institute of Statistical Mathematics, 50, 567–585.

del Castillo, J. and Pérez-Casany, M., 2005. Overdispersed and underdispersed poisson generalizations. Journal of Statistical Planning and Inference, 134, 486–500.

Denuit, M., 1997. A new distribution of poisson-type for the number of claims. ASTIN Bulletin, 27, 229–242.

Fisher, R.A., 1934. The effect of methods of ascertainment upon the estimation of frequencies. Annals of Eugenics, 6, 13–25.

Fishman, G., 2013. Monte Carlo: Concepts, Algorithms, and Applications. Springer, New York, US.

Graybill, F.A., 1983. Matrices with applications in statistics. Wadsworth International Group, London, Uk.

Hilbe, J.M., 2014. Modeling Count Data. Cambridge University Press, Cambridge, MA, US.

Ho, L.L., Andrade, B., Bourguignon, M., and Fernandes, F.H., 2021. Monitoring count data with Shewhart control charts based on the Touchard model. Quality and Reliability Engineering International, 37, 1875–1893.

Hoaglin, D.C. and Tukey, J.W., 2009. Checking the Shape of Discrete Distributions, Chapter 9, 345–416. Wiley, New York, US.

Kleiber, C. and Zeileis, A., 2016. Visualizing count data regressions using rootograms. The American Statistician, 70, 296–303.

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.

Kokonendji, C.C., Simplice, D.G., and Demétrio, C.G.B., 2004. Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. SORT, 28, 201–214.

Lehmann, E.L. and Casella, G., 1998. Theory of Point Estimation. Springer, New York, US.

Matsushita, R., Pianto, D., Andrade, B., and Cançado, A., 2018. The touchard distribution and process. Communications in Statistics: Theory and Methods.

Nelder, J.A. and Pregibon, D., 1987. An extended quasi-likelihood function. Biometrika, 74, 221–232.

Patil, G.P. and Rao, C.R., 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics, 34, 179–189.

Rao, C.R., 1985. Weighted distributions arising out of methods of ascertainment: What population does a sample represent? In Atkinson, A.C. and Fienberg, S.E. (eds.) A Celebration of Statistics, pp. 543–569, Springer, New York, US.

Rathie, A.K., 1997. A new generalization of generalized hypergeometric functions. Le Mathematiche, 12, 297–310.

Sellers, K.F. and Shmueli, G., 2010. A flexible regression model for count data. The Annals of Applied Statistics, 4, 943–961.

Springer, M.D., 1979. The Algebra of Random Variables. Wiley, New York, US.
Wei, B.C., Hu, Y.Q., and Fung, W.K., 1998. Generalized leverage and its applications. Scandinavian Journal of Statistics, 25, 25–37.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://soche.cl/chjs/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright