Article

On the Discretization of Continuous Probability Distributions Using a Probabilistic Rounding Mechanism

Chénangnon Frédéric Tovissodé [1] **, Sèwanou Hermann Honfo** [1,†] **, Jonas Têlé Doumatè** [1,2,†] **and Romain Glèlè Kakaï** [1,*]

1   Laboratoire de Biomathématiques et d'Estimations Forestières, Université d'Abomey-Calavi, Abomey-Calavi, Benin; chenangnon@gmail.com (C.F.T.); honfosewanou@gmail.com (S.H.H.); jonas.doumate@fast.uac.bj (J.T.D.)

2   Faculté des Sciences et Techniques, Université d'Abomey-Calavi, Abomey-Calavi, Benin

*   Correspondence: romain.glelekakai@fsa.uac.bj

†   These authors contributed equally to this work.

Abstract: Most existing flexible count distributions allow only approximate inference when used in a regression context. This work proposes a new framework to provide an exact and flexible alternative for modeling and simulating count data with various types of dispersion (equi-, under-, and over-dispersion). The new method, referred to as "balanced discretization", consists of discretizing continuous probability distributions while preserving expectations. It is easy to generate pseudo random variates from the resulting balanced discrete distribution since it has a simple stochastic representation (probabilistic rounding) in terms of the continuous distribution. For illustrative purposes, we develop the family of balanced discrete gamma distributions that can model equi-, under-, and over-dispersed count data. This family of count distributions is appropriate for building flexible count regression models because the expectation of the distribution has a simple expression in terms of the parameters of the distribution. Using the Jensen–Shannon divergence measure, we show that under the equidispersion restriction, the family of balanced discrete gamma distributions is similar to the Poisson distribution. Based on this, we conjecture that while covering all types of dispersions, a count regression model based on the balanced discrete gamma distribution will allow recovering a near Poisson distribution model fit when the data are Poisson distributed.

Keywords: flexible count models; balanced gamma distribution; Jensen–Shannon divergence; latent equidispersion

1. Introduction

The regression analysis of count responses mostly relies on the Poisson model. However, the equidispersion (variance equals mean) assumption of the Poisson distribution makes Poisson regression inappropriate in many situations where data show overdispersion (variance greater than mean) or underdispersion (variance less than mean). Moreover, it has been observed that many data analyzed using overdispersion models (e.g., negative binomial [1]), which are as popular as the Poisson regression model, may be mixtures of overdispersed and underdispersed or equidispersed counts [2]. The implication is that appropriate alternatives to the Poisson model should allow variable dispersion, i.e., full dispersion flexibility [3]. Existing count regression models associated with variable dispersion exhibit some drawbacks. The first is improperly normalized probability mass functions for underdispersion situations (quasi-Poisson [4], Consul's generalized Poisson [5], and extended Poisson–Tweedy regressions [6]), which makes inference approximate with quasi-models. Another drawback is the lack of a simple expression for the model mean value (Conway–Maxwell–Poisson [7], double Poisson [8,9], gamma count [10], semi-nonparametric Poisson polynomial [11], and discrete Weibull [12] models). The latter

drawback motivated some research works where quantities other than the mean were modeled, leading to hardly interpretable fits [11,13].

The development of discrete analogues of continuous probability distributions, which has received great attention in the last two decades, provides an attractive route for building count regression models with variable dispersion. A review by [14] offers a survey of the different methods with their specific application fields. An appealing characteristic of discrete analogues of continuous distributions is the generation of discrete pseudo random values, which only requires basic operations once the continuous distribution can be simulated. This simulation easiness is especially interesting for simulation based model evaluation [15] or parametric bootstrapping based inference [16].

Despite the various existing discretization methods, mean parameterizable approaches necessary to build easily interpretable regression models are rare. For reliability evaluation, the discretizing approach of [17] attempts to match the mean and the variance of the discrete and the related continuous variable, but it provides only an approximate solution at the cost of a tuning parameter. The proposals in [3,18,19] offer solutions for constructing count variables with a fixed mean value and variable dispersion, but they lack a physical basis, i.e., a generating mechanism to motivate their use in practice.

This work proposes a discretization procedure to start from continuous probability distributions and construct count models with (i) properly normalized probability mass functions for underdispersion, equidispersion, as well as overdispersion situations and (ii) simple expressions for the model mean values. As a result, the proposal allows full likelihood inference (as opposed to quasi-likelihood inference) for any dispersion level in observed data and is thus suited for regression analysis where the estimation of covariate effects on the mean count is of great interest.

The proposed discretization approach modifies the "discrete concentration" method, i.e., "Methodology IV" in [14], to preserve the expectation of the continuous distribution. Our proposal, referred to as "balanced discretization" is based on a probabilistic rounding mechanism, which provides a generating mechanism with a simple interpretation. Interestingly, the probabilistic rounding mechanism, expressed as a simple stochastic representation in terms of the continuous distribution, allows easily generating pseudo random variates from the resulting balanced discrete distribution.

The rest of the paper is organized as follows. Section 2 motivates and presents the balanced discretization method. The general expressions of the distribution functions and moments of balanced discrete distributions are given. The method is applied to the gamma distributions in Section 3 to produce the balanced discrete gamma distribution, which is compared to the discrete concentration of the gamma distribution and to the Poisson distribution. Concluding remarks are given in Section 4.

2. The Balanced Discretization Method

This section motivates and describes the balanced discretization method. The general expressions of the probability mass, the cumulative distribution, the survival and quantile functions, the moments, and the index of dispersion are presented. The link between balanced discretization and the mean-preserving discretization approach of [3] is also established. The proofs of the results are routine and given for completeness in Appendix A.

2.1. Notations

We denote $\mathbb{Z}$ the set of integers ($\mathbb{Z} = \{\cdots, -1, 0, 1, \cdots\}$), $\mathbb{N}$ the set of non-negative integers ($\mathbb{N} = \{0, 1, \cdots\}$), $\mathbb{N}_+$ the set of natural numbers ($\mathbb{N}_+ = \{1, 2, \cdots\}$), $\mathbb{R}$ the set of reals ($\mathbb{R} = (-\infty, \infty)$), and $\mathbb{R}_+$ the set of positive reals ($\mathbb{R}_+ = (0, \infty)$). Let $\lfloor x \rfloor$ be the integer part of any real x. Following [14], we denote continuous random variables by X and discrete random variables by Y. Accordingly, $f_X(\cdot|\boldsymbol{\theta})$, $F_X(\cdot|\boldsymbol{\theta})$, $S_X(\cdot|\boldsymbol{\theta})$, and $Q_X(\cdot|\boldsymbol{\theta})$ denote respectively the pdf (probability density function), the cdf (cumulative distribution function), the suf (survival function), and the quf (quantile function) of X, whereas $f_Y(\cdot|\boldsymbol{\theta})$, $F_Y(\cdot|\boldsymbol{\theta})$, $S_Y(\cdot|\boldsymbol{\theta})$, and $Q_Y(\cdot|\boldsymbol{\theta})$ denote respectively the pmf (probability mass function), the

cdf, the suf, and the quf of Y, all indexed by a parameter vector $\boldsymbol{\theta}$. Continuous probability distributions are assumed to be non-degenerate. A Bernoulli random variable with success probability $p \in [0,1]$ is denoted $\mathcal{BER}(p)$.

2.2. Reminders

First, we recall the discrete concentration method and the mean-preserving approach of [3]. Let $\mathcal{CD}(\boldsymbol{\theta})$ be a continuous probability distribution of interest. The discrete concentration $\mathcal{DC}(\boldsymbol{\theta})$ of $X \sim \mathcal{CD}(\boldsymbol{\theta})$ is the count variable Y with the pmf and suf:

$$
\begin{aligned}
f_Y(y|\boldsymbol{\theta}) &= F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta}) & (1)\\
S_Y(y|\boldsymbol{\theta}) &= S_X(y|\boldsymbol{\theta}) & (2)
\end{aligned}
$$

for $y \in \mathbb{Z}$, i.e., Y has cdf $F_Y(y|\boldsymbol{\theta}) = F_X(y+1|\boldsymbol{\theta})$ and quf $Q_Y(u|\boldsymbol{\theta}) = \lfloor Q_X(u|\boldsymbol{\theta}) \rfloor$ for $u \in (0,1)$. Accordingly, the nth moment about zero of Y is:

$$
\mathrm{E}[Y^n] = \sum_{y=-\infty}^{\infty} y^n [F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta})]. \tag{3}
$$

Clearly, the discrete concentration of X is simply $Y \overset{d}{=} \lfloor X \rfloor$ where $\overset{d}{=}$ means "equal in distribution to". Thus, $Y = X - U$, where U is the fractional part of X. Since $U \in (0,1)$, it satisfies $0 < \mathrm{E}[U^2] < \mathrm{E}[U] < 1$, providing bounds on the mean and the variance of the count variable: $\mathrm{E}[X] - 1 \leq \mathrm{E}[Y] \leq \mathrm{E}[X]$ and $\mathrm{Var}[X] \leq \mathrm{Var}[Y] \leq \mathrm{Var}[X] + 1/4$ [20].

The mean-preserved discrete version Y of X is the variable with the cdf:

$$
F_Y(y|\boldsymbol{\theta}) = \int_y^{y+1} F_X(x|\boldsymbol{\theta}) dx \ \text{ for } \ y \in \mathbb{N}. \tag{4}
$$

and expectation $\mathrm{E}[Y] = \mathrm{E}[X]$ [3].

2.3. Motivating Example and Definition

Example 1 (Measuring tree diameter). *Discretization mechanisms arise when measuring any continuous quantity. Indeed, no sample can cover a whole continuum since the latter has an infinite number of points, and only a finite number of decimal places are reported in practice [14,21]. Assume for instance an operator measuring tree diameters X in a forest inventory frame, using a measurement device scaled to millimeters (mm). Since X is a continuous variable, the probability of observing $X = x$ mm is zero. When the true value x of the diameter of a tree actually falls between two consecutive graduations z and $z+1$, the operator reports either $y = z$ mm or $y = z+1$ mm, i.e., only a discretized version Y of X is observed. Beyond this example, when direct measures are taken, only the number of an arbitrary unit is actually counted. Clearly, the closer x is to z, the higher probability of reporting $y = z$, and conversely, the closer x is to $z+1$, the higher the probability of reporting $y = z+1$. Balanced discretization results from assuming that given $z \leq x < z+1$, the probability for reporting $y = z+1$ is exactly $x - z$.*

Definition 1 (Balanced discretization). *Let us consider an absolutely continuous probability distribution $\mathcal{CD}(\boldsymbol{\theta})$ of interest. A count random variable Y is said to be distributed as the balanced discrete counterpart denoted $\mathcal{BD}(\boldsymbol{\theta})$ of the continuous distribution $\mathcal{CD}(\boldsymbol{\theta})$, if it has the stochastic representation:*

$$
\begin{aligned}
Y|U = u, X = x &\overset{d}{=} z + u & (5)\\
U|X = x &\sim \mathcal{BER}(r) \\
X &\sim \mathcal{CD}(\boldsymbol{\theta})
\end{aligned}
$$

where $z = \lfloor x \rfloor$ and $r = x - z$.

Let $E_X(n, y|\boldsymbol{\theta})$ denote the nth partial moment:

$$E_X(n, y|\boldsymbol{\theta}) = \int_y^{y+1} x^n f_X(x|\boldsymbol{\theta}) dx \tag{6}$$

of X over $(y, y+1)$, and set:

$$H_X(y|\boldsymbol{\theta}) = F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta}). \tag{7}$$

The balanced discretization mechanism in Equation (5) preserves partial expectations $E_X(1, y|\boldsymbol{\theta})$ of the continuous variable as shown by Equation (10) of the following lemma.

Lemma 1. *Let X and Y be defined as in Equation (5). Then, for any $y \in \mathbb{Z}$,*

$$P(Y = y \text{ and } y \leq X < y+1) = (y+1)H_X(y|\boldsymbol{\theta}) - E_X(1, y|\boldsymbol{\theta}) \tag{8}$$
$$P(Y = y+1 \text{ and } y \leq X < y+1) = E_X(1, y|\boldsymbol{\theta}) - yH_X(y|\boldsymbol{\theta}) \tag{9}$$
$$E_Y[Y|y \leq X < y+1] = E_X(1, y|\boldsymbol{\theta}) \tag{10}$$

where $E_Y[Y|X \in \mathbb{A}]$ is the partial mean of Y for $X \in \mathbb{A}$.

2.4. Probability Mass and Distribution Functions

We derive in this section some general distributional properties of balanced discrete distributions.

Proposition 1 (Distribution function). *Let $Y \sim \mathcal{BD}(\boldsymbol{\theta})$. The pmf, the cdf, the suf, and the quf of Y are given for $y \in \mathbb{Z}$ and $0 \leq u \leq 1$ by:*

$$f_Y(y|\boldsymbol{\theta}) = (y-1)F_X(y-1|\boldsymbol{\theta}) - 2yF_X(y|\boldsymbol{\theta}) + (y+1)F_X(y+1|\boldsymbol{\theta})$$
$$+ E_X(1, y-1|\boldsymbol{\theta}) - E_X(1, y|\boldsymbol{\theta}) \tag{11}$$
$$F_Y(y|\boldsymbol{\theta}) = F_X(y|\boldsymbol{\theta}) + (y+1)H_X(y|\boldsymbol{\theta}) - E_X(1, y|\boldsymbol{\theta}) \tag{12}$$
$$S_Y(y|\boldsymbol{\theta}) = S_X(y|\boldsymbol{\theta}) - (y-1)H_X(y-1|\boldsymbol{\theta}) + E_X(1, y-1|\boldsymbol{\theta}) \tag{13}$$
$$Q_Y(u|\boldsymbol{\theta}) = \begin{cases} x_o & \text{if } u_o \geq u \\ x_o + 1 & \text{otherwise} \end{cases} \tag{14}$$

where $x_o = \lfloor Q_X(u|\boldsymbol{\theta}) \rfloor$ and $u_o = F_Y(x_o|\boldsymbol{\theta})$.

Note from Equation (11) that $\mathcal{BD}(\boldsymbol{\theta})$ assigns less probability mass to zero than the discrete concentration of $X \sim \mathcal{CD}(\boldsymbol{\theta})$ if X has support $\mathbb{R}_+$ or $(0, M)$ for $M \in \mathbb{R}_+$. Equation (13) emphasizes that the balanced discretization method does not preserve the suf of the continuous distribution, unlike the discrete concentration (see Equation (2)). Nevertheless, the balanced discrete cdf and suf satisfy the inequalities $F_X(y|\boldsymbol{\theta}) \leq F_Y(y|\boldsymbol{\theta}) \leq F_X(y+1|\boldsymbol{\theta})$ (with equalities when the support of X is upper bounded by y) and $S_X(y|\boldsymbol{\theta}) \leq S_Y(y-1|\boldsymbol{\theta}) \leq S_X(y-1|\boldsymbol{\theta})$ (with equalities when the support of X is lower bounded by y).

By Equation (14), balanced discretization somewhat preserves the median of the continuous distribution. Indeed, if X has an integral median m_X, then Y has median $m_Y = m_X - 1/2$. More generally, we have $m_Y = \lfloor m_X \rfloor + 1/2$ if $F_Y(\lfloor m_X \rfloor|\boldsymbol{\theta}) < 1/2$, $m_Y = \lfloor m_X \rfloor$ if $F_Y(\lfloor m_X \rfloor|\boldsymbol{\theta}) = 1/2$ and $m_Y = \lfloor m_X \rfloor - 1/2$ if $F_Y(\lfloor m_X \rfloor|\boldsymbol{\theta}) > 1/2$.

2.5. Moments and Index of Dispersion

This section presents expressions for the moments of balanced discrete distributions. We start with the first two moments since they are the most important in a count regression context.

Proposition 2 (Mean and variance). *Let $X \sim \mathcal{CD}(\boldsymbol{\theta})$ with mean $\mu_X(\boldsymbol{\theta})$ and variance $\sigma_X^2(\boldsymbol{\theta})$. The balanced discrete counterpart of X, $Y \sim \mathcal{BD}(\boldsymbol{\theta})$, has mean $\mu_Y(\boldsymbol{\theta}) = \mu_X(\boldsymbol{\theta})$ and variance:*

$$\sigma_Y^2(\boldsymbol{\theta}) = \sigma_X^2(\boldsymbol{\theta}) + \zeta_0(\boldsymbol{\theta}) \tag{15}$$

where $\zeta_0(\boldsymbol{\theta}) = \mathrm{E}_X[R(1-R)]$ with $R \overset{d}{=} X - \lfloor X \rfloor$. In addition, $\zeta_0(\boldsymbol{\theta})$ satisfies the inequality $0 < \zeta_0(\boldsymbol{\theta}) < min\{\mu_Y(\boldsymbol{\theta}), 1/4\}$ and is given by the sum $\zeta_0(\boldsymbol{\theta}) = \sum_{z=-\infty}^{\infty} \zeta_0(z, \boldsymbol{\theta})$ with:

$$\zeta_0(z, \boldsymbol{\theta}) = (2z+1)\mathrm{E}_X(1, z|\boldsymbol{\theta}) - \mathrm{E}_X(2, z|\boldsymbol{\theta}) - z(z+1)H_X(z|\boldsymbol{\theta}). \tag{16}$$

From Proposition 2, it appears that when the variance of a balanced discrete variable Y exists, it satisfies $\sigma_X^2(\boldsymbol{\theta}) < \sigma_Y^2(\boldsymbol{\theta}) < \sigma_X^2(\boldsymbol{\theta}) + min\{\mu_Y(\boldsymbol{\theta}), 1/4\}$. This suggests the inexpensive approximation $\hat{\sigma}_Y^2(\boldsymbol{\theta}) = \sigma_X^2(\boldsymbol{\theta}) + min\{\hat{\mu}/2, 1/8\}$ with $|\hat{\sigma}_Y^2(\boldsymbol{\theta}) - \sigma_Y^2(\boldsymbol{\theta})| < min\{\hat{\mu}/2, 1/8\}$, $\hat{\mu}$ being the mean of Y (exact or estimate). The following corollary infers the ID (index of dispersion) of a balanced discrete distribution from Proposition 2.

Corollary 1 (Index of dispersion). *Let $Y \sim \mathcal{BD}(\boldsymbol{\theta})$ be the balanced discrete counterpart of $X \sim \mathcal{CD}(\boldsymbol{\theta})$ with cdf $F_X(\cdot|\boldsymbol{\theta})$, quf $Q_X(\cdot|\boldsymbol{\theta})$, expectation $\mu_X(\boldsymbol{\theta}) \neq 0$, and index of dispersion (variance-to-mean ratio) $\mathrm{ID}_X(\boldsymbol{\theta})$. The index of dispersion $\mathrm{ID}_Y(\boldsymbol{\theta})$ of Y satisfies:*

$$\mathrm{ID}_Y(\boldsymbol{\theta}) = \mathrm{ID}_X(\boldsymbol{\theta}) + \frac{\zeta_0(\boldsymbol{\theta})}{\mu_X(\boldsymbol{\theta})} \tag{17}$$

$$|\mathrm{ID}_X(\boldsymbol{\theta})| < |\mathrm{ID}_Y(\boldsymbol{\theta})| \leq |\mathrm{ID}_X(\boldsymbol{\theta})| + \frac{1}{4|\mu_X(\boldsymbol{\theta})|}. \tag{18}$$

Furthermore, $\zeta_0(\boldsymbol{\theta})$ can be approximated with a tolerance $\alpha \in (0,1)$ by the truncated sum:

$$\widehat{\zeta}_\alpha(\boldsymbol{\theta}) = \sum_{z=z_i}^{z_f} \zeta_0(z, \boldsymbol{\theta}) \tag{19}$$

where $z_i = \lfloor Q_X(\alpha/2|\boldsymbol{\theta}) \rfloor$, $z_f = \lfloor Q_X(1 - \alpha/2|\boldsymbol{\theta}) \rfloor + 1$ and α controls the precision of $\widehat{\zeta}_\alpha(\boldsymbol{\theta})$ via $|\widehat{\zeta}_\alpha(\boldsymbol{\theta}) - \zeta_0(\boldsymbol{\theta})| < 1 - F_X\left(z_f + 1|\boldsymbol{\theta}\right) + F_X(z_i|\boldsymbol{\theta})$.

The next proposition shows the relation between moments of balanced discrete distributions and of discrete concentrations.

Proposition 3. *Let $Y \sim \mathcal{BD}(\boldsymbol{\theta})$ be the balanced discrete counterpart of $X \sim \mathcal{CD}(\boldsymbol{\theta})$. The nth moment of Y satisfies for $n \in \mathbb{N}_+$:*

$$\mu_Y^{(n)}(\boldsymbol{\theta}) = \mu_Z^{(n)}(\boldsymbol{\theta}) + \sum_{i=0}^{n-1} \binom{n}{i} \mu_{ZU}^{(i)}(\boldsymbol{\theta}) \tag{20}$$

where $\mu_Z^{(n)}(\boldsymbol{\theta})$ is the nth moment of the discrete concentration $Z = \lfloor X \rfloor$ (Equation (3)) and $\mu_{ZU}^{(i)}(\boldsymbol{\theta})$ is the expectation of the product of Z^i and U with $U|X \sim \mathcal{BER}(X - Z)$, given by $\mu_{ZU}^{(i)}(\boldsymbol{\theta}) = -\mu_Z^{(i+1)}(\boldsymbol{\theta}) + \sum_{z=-\infty}^{\infty} z^i \mathrm{E}_X(1, z|\boldsymbol{\theta})$.

2.6. Conditional Distributions of Latent Continuous and Binary Outcomes

Although the balanced discrete distribution was motivated by the need for mean-parameterizable flexible discrete probability distributions, it may be used to model any continuous outcome measured to fewer decimal places. In such instances, the conditional distribution and in particular the conditional mean of the underlying continuous distribution may be useful for predicting the continuous variable given an observed discrete value. In addition, since a balanced discrete variable is the observable feature of an underlying

continuous outcome, a useful tool for maximum likelihood inference in complex models is the expectation-maximization algorithm [22], which handles any latent class-like model. In a Bayesian inference framework, the stochastic representation of the balanced discrete distribution can also be useful for sampling the posterior distribution of model parameters when draws from the truncated form of the continuous distribution are inexpensive. The following result provides expressions for these purposes.

Proposition 4. *Let X, U, and Y be defined as in Equation* (5)*. Then, for $y \in \mathbb{Z}$ with probability mass $f_Y(y|\boldsymbol{\theta}) > 0$:*

$$f_{U|Y}(u|Y = y, \boldsymbol{\theta}) = p_y^u [1 - p_y]^{1-u} \text{ for } u \in \{0, 1\}, \tag{21}$$

$$f_{X|Y}(x|Y = y, \boldsymbol{\theta}) = \frac{f_X(x|\boldsymbol{\theta})}{f_Y(y|\boldsymbol{\theta})} \left[(1 - y + x) I_{(y-1,y)}(x) + (1 + y - x) I_{(y,y+1)}(x) \right], \tag{22}$$

and for $n \in \mathbb{R}$ such that X^n is well defined in both $(y-1, y)$ and $(y, y+1)$,

$$\begin{aligned} E_{X|Y}[X^n|Y = y, \boldsymbol{\theta}] = \ & \frac{1}{f_Y(y|\boldsymbol{\theta})} [(1-y)E_X(n, y-1|\boldsymbol{\theta}) + E_X(n+1, y-1|\boldsymbol{\theta}) \\ & + (1+y)E_X(n, y|\boldsymbol{\theta}) - E_X(n+1, y|\boldsymbol{\theta})] \end{aligned} \tag{23}$$

where $p_y = [f_Y(y|\boldsymbol{\theta})]^{-1}[E_X(1, y-1|\boldsymbol{\theta}) - (y-1)H_X(y|\boldsymbol{\theta})]$ is the conditional mean (success probability) of the Bernoulli variable U given $Y = y$ and $I_{\mathbb{A}}(x)$ is the indicator function, which equals one if $x \in \mathbb{A}$ and 0 otherwise.

Note from Equation (22) that given the continuous variable, the distribution of the discrete variable does not depend on the parameter vector $\boldsymbol{\theta}$:

$$f_{Y|X}(y|X = x, \boldsymbol{\theta}) = (1 - y + x) I_{(y-1,y)}(x) + (1 + y - x) I_{(y,y+1)}(x). \tag{24}$$

Therefore, in the expectation-maximization algorithm framework, the maximization of the joint likelihood:

$$f_{X,Y}(x, y) = f_{Y|X}(y|X = x, \boldsymbol{\theta}) f_X(x|\boldsymbol{\theta}) \tag{25}$$

of Y and X is reduced to the maximization of the likelihood $f_X(x|\boldsymbol{\theta})$ of the continuous variable X. Hence, the expectation-maximization algorithm will be appropriate for fitting a balanced discrete distribution whenever fitting the underlying continuous distribution is easy.

2.7. Link with Mean-Preserving Discretization

Recall from Equation (4) that the cdf of the mean-preserved count variable has the form:

$$F_Y(y|\boldsymbol{\theta}) = \int_y^{y+1} F_X(x|\boldsymbol{\theta}) dx.$$

From the identity $\int F_X(x|\boldsymbol{\theta})dx = xF_X(x|\boldsymbol{\theta}) - \int xf_X(x|\boldsymbol{\theta})dx$, we have:

$$F_Y(y|\boldsymbol{\theta}) = (y+1)F_X(y+1|\boldsymbol{\theta}) - yF_X(y|\boldsymbol{\theta}) - E_X(1, y|\boldsymbol{\theta}). \tag{26}$$

Then, using the identity $F_X(y + 1|\boldsymbol{\theta}) = F_X(y|\boldsymbol{\theta}) + [F_X(y + 1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta})]$ straightforwardly results in:

$$F_Y(y|\boldsymbol{\theta}) = F_X(y|\boldsymbol{\theta}) + (y+1)[F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta})] - E_X(1, y|\boldsymbol{\theta}),$$

i.e., the cdf in Equation (12). Therefore, balanced discretization as defined in Equation (5) provides a generating mechanism for the mean-preserving method of [3].

3. The Balanced Discrete Gamma Family

The class of gamma distributions is a flexible class of distributions encountered in various statistical applications. This class of distributions includes as special cases the exponential, the one-parameter gamma, and up to rescaling the chi-squared distributions [23]. Chakraborty and Chakravarty [24] studied the discrete concentration of the gamma distribution with applications in biology and socio-economics. In order to allow exact inference in flexible count regression models, we apply in this section the balanced discretization method to gamma distributions. We present the balanced discrete gamma distribution under mean parametrization convenient for regression purposes and compare the distribution to the discrete concentration of the gamma distribution and to the Poisson distribution.

Let $\mathcal{G}(b,a)$ denote the gamma distribution with cdf $F_g(x|b,a) = \gamma(ax,b)$ for $x > 0$, where $\gamma(x,a) = \int_0^x u^{a-1}e^{-u}du/\Gamma(a)$ is the lower incomplete gamma ratio for $(a,x) \in \mathbb{R}_+^2$ and $\Gamma(a) = \int_0^\infty u^{a-1}e^{-u}du$ is the gamma function. A random variable $X \sim \mathcal{G}(b,a)$ has expectation b/a, variance $b/(a^2)$, and nth order partial moment $E_g(n,y|b,a) = E_X[X^n|y \leq X < y+1]$ given by:

$$E_g(n,y|b,a) = \frac{\Gamma(b+n)}{a^n\Gamma(b)}[\gamma(a(y+1),b+n)) - \gamma(ay,b+n)]. \tag{27}$$

The one-parameter gamma distribution is obtained for $a = 1$ (equidispersion) and is denoted $\mathcal{G}(b)$.

3.1. The Balanced Discrete Gamma Distribution

A count random variable with support $\mathbb{N}$ is said to follow a balanced discrete gamma distribution denoted $\mathcal{BG}(\mu,a)$ for $(\mu,a) \in \mathbb{R}_+^2$, if it is generated by the discretization mechanism in Equation (5) with $X \sim \mathcal{G}(a\mu,a)$. By Proposition 2, a $\mathcal{BG}(\mu,a)$ variable has expectation μ. Using Equation (27), some properties of $\mathcal{BG}(\mu,a)$ follow as in Corollary 2 hereafter.

Corollary 2 (Balanced discrete gamma distribution). *Let $Y \sim \mathcal{BG}(\mu,a)$, and set $b = a\mu$. Then, the pmf and the cdf of Y are respectively given for $y \in \mathbb{N}$ by:*

$$\begin{aligned}
f_{dg}(y|\mu,a) &= (y-1)\gamma(a(y-1),b) - 2y\gamma(ay,b) + (y+1)\gamma(a(y+1),b)\\
&\quad -\mu[\gamma(a(y-1),b+1) - 2\gamma(ay,b+1) + \gamma(a(y+1),b+1)] \tag{28}\\
F_{dg}(y|\mu,a) &= (y+1)\gamma(a(y+1),b) - y\gamma(ay,b)\\
&\quad -\mu[\gamma(a(y+1),b+1) - \gamma(ay,b+1)]; \tag{29}
\end{aligned}$$

and the variance and the index of dispersion of Y are respectively given by:

$$\sigma_{dg}^2(\mu,a) = a^{-1}\mu + \zeta_{g_0}(\mu,a) \tag{30}$$

$$\mathrm{ID}_{dg}(\mu,a) = a^{-1} + \mu^{-1}\zeta_{g_0}(\mu,a) \tag{31}$$

where:

$$\zeta_{g_0}(\mu,a) = \sum_{z=0}^\infty \zeta_{g_0}(z,\mu,a) \quad \text{with} \tag{32}$$

$$\begin{aligned}
\zeta_{g_0}(z,\mu,a) &= -\mu\left(\mu + a^{-1}\right)[\gamma(a(z+1),b+2) - \gamma(az,b+2)]\\
&\quad \mu(2z+1)[\gamma(a(z+1),b+1) - \gamma(az,b+1)]\\
&\quad -z(z+1)[\gamma(a(z+1),b) - \gamma(az,b)]. \tag{33}
\end{aligned}$$

The pmf (28) and the cdf (29) follow by Equations (11) and (12) respectively along with Equation (27). Note that the computations of the pmf and the cdf of a balanced discrete gamma distribution only require a routine for the incomplete gamma ratio $\gamma(\cdot, \cdot)$, which is available in most statistical software as the cdf of the continuous gamma distribution (e.g., pgamma in the R freeware [25] and gamcdf in MATLAB [26]). The variance (30) and the index of dispersion (31) follow by Equation (15) along with Equation (27). Note that the variance term $\zeta_{g_0}(\mu, a)$ in Equation (32) can be approximated via the truncation mechanism in Equation (19) with a tolerance $\alpha \in (0, 1)$ as:

$$\widehat{\zeta}_{g_\alpha}(\mu, a) = \sum_{z=z_i}^{z_f} \zeta_{g_0}(z, \mu, a) \tag{34}$$

where $z_i = \lfloor \gamma^{-1}(\alpha/2, b)/a \rfloor$, $z_f = \lfloor \gamma^{-1}(1 - \alpha/2, b)/a \rfloor + 1$ with $\gamma^{-1}(\cdot, b)$ the inverse function of the incomplete gamma ratio $\gamma(\cdot, b)$ (available, e.g., as qgamma in R and gaminv in MATLAB).

The one-parameter BDG (balanced discrete gamma) distribution, denoted $\mathcal{BG}(\mu)$ and obtained by setting $a = 1$, corresponds to a latent equidispersion mechanism and is marginally slightly overdispersed as indicated by Equation (31) with $a = 1$. Setting $a = \mu^{-1}$ produces the balanced discrete exponential distribution $\mathcal{BE}(\mu)$, which is close to the geometric distribution since the latter corresponds to the discrete concentration of the exponential distribution [27].

Figure 1 displays the probability mass function of the BDG distributions with mean values $\mu = 2.5$ and $\mu = 5$ (computed using Equation (28) in R). It appears that the scale parameter a controls the shape of the distribution, allowing both unimodal and reverse J shapes. It can be observed that the spread of a BDG distribution $\mathcal{BG}(\mu, a)$ decreases with a for fixed μ. The boxplots depicted in Figure 2 show that the skewness of the distribution (as measured by the difference between the mean and the median values) increases with its spread and thus decreases with a. As expected, both the length of the right tail of the distribution (Figure 1) and the probability of unusual (far from the mean and the median) observations (Figure 2) increase with the spread.

The index of dispersion, which can assume any positive value, is computed using Equation (31) in R and depicted in Figure 3 for $a \geq 1$. It can be observed in accordance with Equation (31) that for large mean values ($\mu \geq 10$), the variance-to-mean ratio is not very sensitive to μ for fixed, but not too large scale parameter values ($a < 50$). This also holds for very low scales ($a < 0.5$) for any mean value. Very large scales ($a \geq 50$) induce, in addition to severe underdispersion (ID < 0.2), an oscillating index of dispersion with an amplitude approaching zero as μ increases. Indeed, for large-scale values (as $a \to \infty$) and finite μ, the continuous gamma distribution converges to a point mass at μ. The balanced discrete gamma distribution thus converges to a binary variable, which takes the values $y = \lfloor \mu \rfloor$ with the probability $1 - (\mu - \lfloor \mu \rfloor)$ and $y = \lfloor \mu \rfloor + 1$ with the probability $(\mu - \lfloor \mu \rfloor)$ and has variance $\sigma_{d_g}^2(\mu, \infty) = (\mu - \lfloor \mu \rfloor)(1 - \mu + \lfloor \mu \rfloor)$. The oscillating pattern in Figure 3 accordingly shows that the variance is minimal when μ is an integer and maximal when μ is a half integer. When both μ and a are large, the amplitude of the oscillations of ID decreases, and ID approaches zero, in accordance with Equation (31).
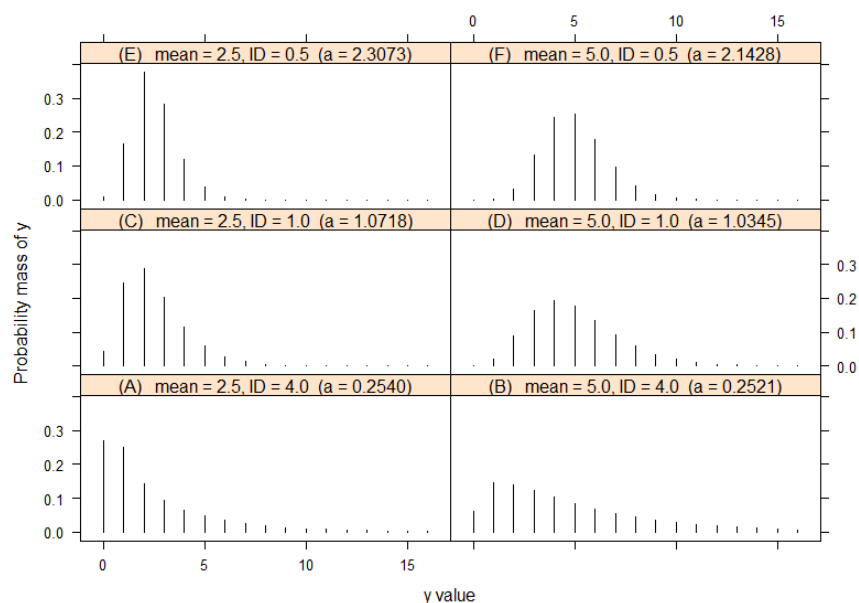
Figure 1. Probability mass plots for the balanced discrete gamma distributions with mean values $\mu = 2.5$ (left panel) and $\mu = 5$ (right panel) and scales *a* selected to yield an index of dispersion (ID, variance-to-mean ratio) of ID = 4 (bottom row), ID = 1 (central row), and ID = 0.5 (top row).
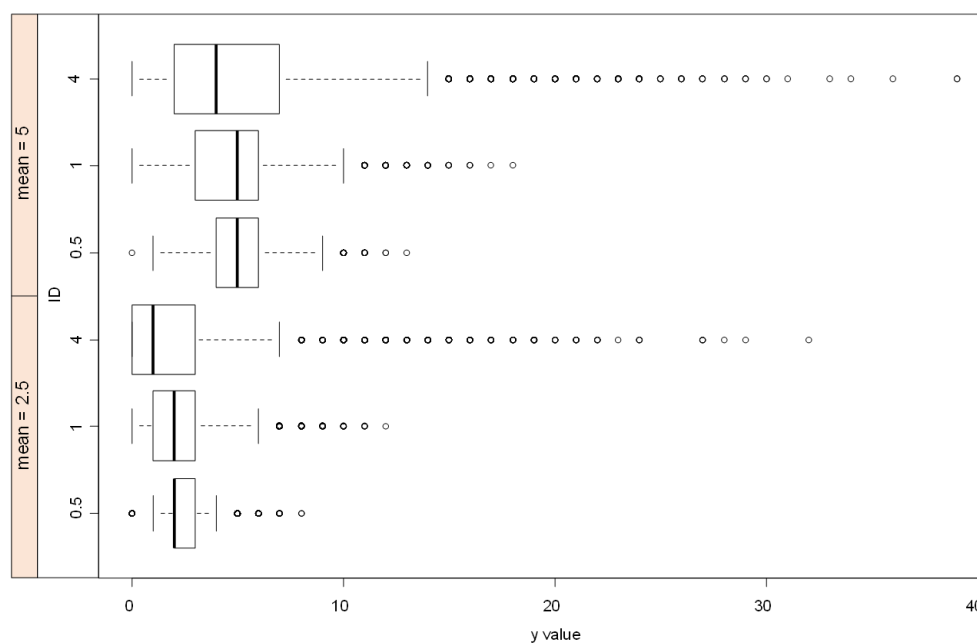

Figure 2. Box plots for the balanced discrete gamma distributions with mean values $\mu = 2.5$ and $\mu = 5$ and an index of dispersion (ID, variance-to-mean ratio) of ID = 4, ID = 1, and ID = 0.5. The thick vertical bar inside the interquartile range (i.e., the rectangular box that has sides that are 25% (left side) and 75% (right side) quartiles) is the median of the distribution.
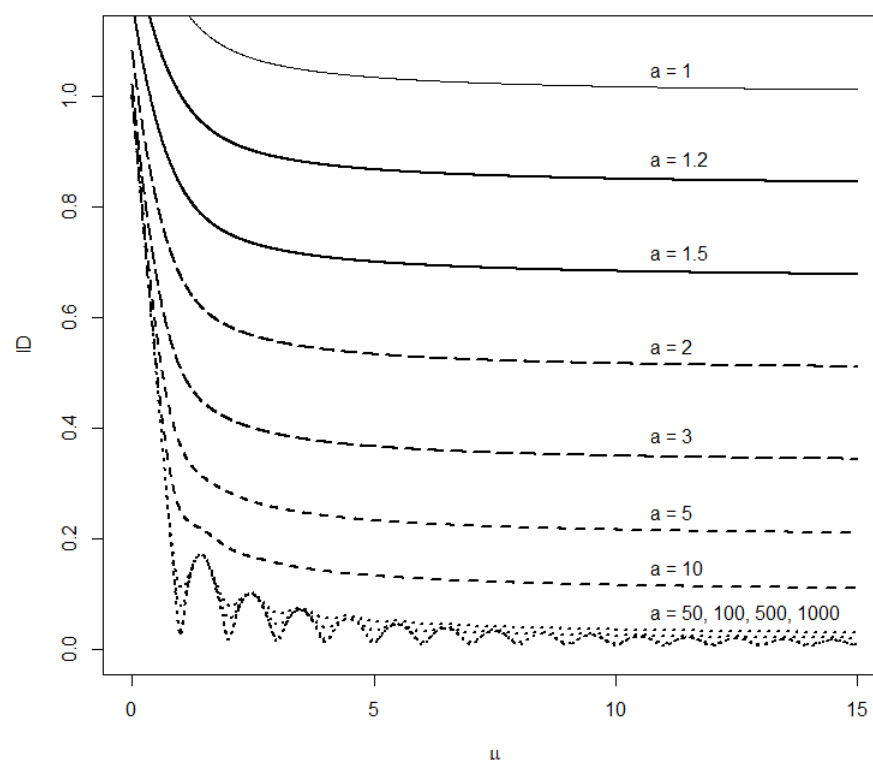
Figure 3. Index of dispersion (ID, variance-to-mean ratio) of the balanced discrete gamma distribution against the mean value (μ) for selected scale parameter (a) values in the range $[1, 1000]$ mostly corresponding to equidispersion and underdispersion ($0 < \text{ID} \leq 1$).

3.2. Comparison with Some Alternatives

The balanced discretization approach results from a light modification of the discrete concentration method. This section assesses on the one hand to what extent the two discretization approaches differ, considering the balanced discrete gamma (BDG) distribution case. On the other hand, the difference between the Poisson and the BDG distributions is evaluated under both latent and marginal equidispersion restrictions.

Among the miscellaneous measures proposed to assess the similarities between probability distributions, the Jensen–Shannon divergence (JSD) [28] has many desirable properties that support its use in statistics [29]. The JSD is an information theory measure given for two pmfs $q_1(\cdot)$ and $q_2(\cdot)$ by [28]:

$$JSD(q_1, q_2) = K(q_1, q_2) + K(q_2, q_1) \tag{35}$$

$$\text{where } K(q_1, q_2) = \sum_{y=-\infty}^{\infty} q_1(y) \log_2\left(\frac{q_1(y)}{0.5q_1(y) + 0.5q_2(y)}\right) \tag{36}$$

with the convention $q_1(y) \log_2(q_1(y)/(0.5q_1(y) + 0.5q_2(y))) = 0$ if $q_1(y) = 0$. The JSD measures the discrepancy between $q_1(\cdot)$ and $q_2(\cdot)$ in bits. It is bounded as $0 \leq JSD(q_1, q_2) \leq 2$ and is zero only if $q_1(y) = q_2(y) \, \forall \, y \in \mathbb{Z}$. The JSD values presented in this section are computed using Equation (35) using the R freeware.

3.2.1. Balanced Discretization Versus Discrete Concentration

Figure 4 illustrates the balanced discretization method using the continuous gamma distribution with parameters $a = 1$, $b = 5$. Unlike the discrete concentration whose cdf lies above the continuous cdf, the balanced discrete distribution is constructed so that the continuous cdf interpolates the cdf of the balanced discrete distribution.
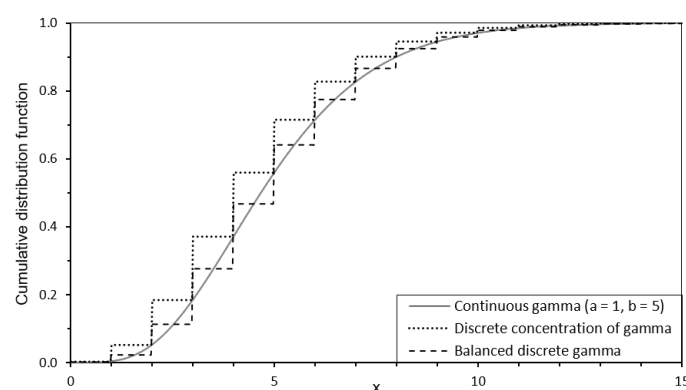
Figure 4. Comparison of the cumulative distribution functions of the balanced discrete gamma and discrete concentration of gamma distributions based on a continuous gamma distribution with scale $a = 1$ and shape $b = 5$.

The curves in Figure 5A show the JSD measure between the balanced discrete gamma and the corresponding discrete concentrations for mean count values $\mu \leq 30$. The selected scale values allow a wide range for the index of dispersion (ID), which roughly runs from 0.04 to 10. It can be observed that the JSD measure is relatively low (JSD < 0.80 bit) and decreases overall with the mean count (but not monotonically). In other words, the discrete concentration and balanced discretization methods produce similar discrete analogues of the considered continuous gamma distributions for large mean values. The JSD measure is especially low (JSD $\leq$ 0.10 bit) for equidispersed and overdispersed balanced discrete gamma distributions ($a \leq 1$). High discrepancy (JSD $\geq$ 0.5 bit) actually appears between the discrete analogues from the two discretization methods generally in underdispersion situations with a very low mean count ($\mu < 1$) or large scale parameter ($a > 5$, implying ID < 0.45). For very large scale values ($a \geq 30$), JSD becomes erratic, oscillating between minima right before the integer values of μ and maxima right after the integer values of μ.
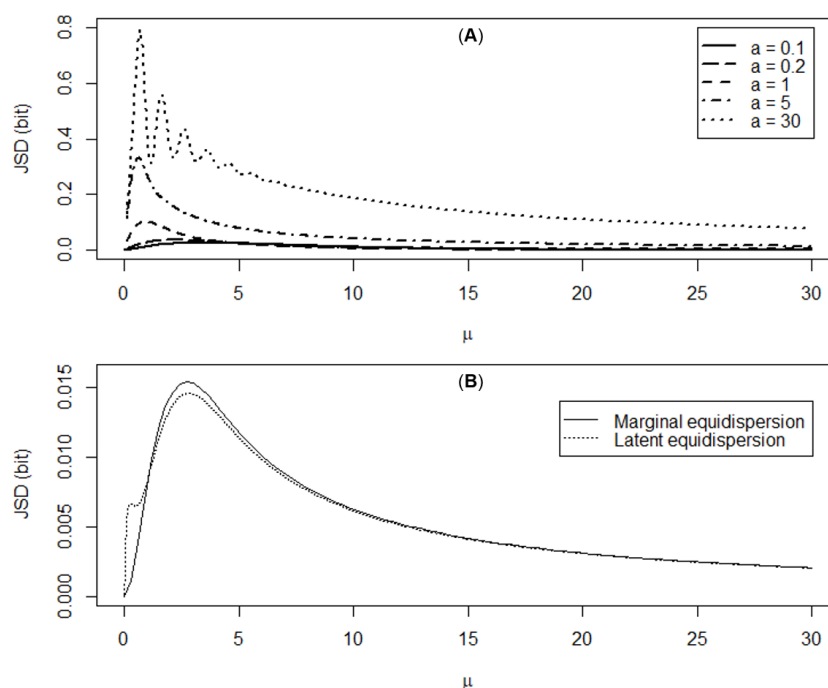

Figure 5. Jensen–Shannon divergence (JSD in bit) measured between the balanced discrete gamma (BDG) distribution and the corresponding discrete concentration (**A**) and between the Poisson and the BDG distributions under both latent equidispersion (one-parameter BDG distribution) and marginal equidispersion (unit variance-to-mean ratio) restrictions (**B**), against the mean value (μ).

3.2.2. Distance to the Poisson Distribution under Equidispersion

The Poisson regression is the most common count regression model, which is appropriate for equidispersed count data. Although the BDG distribution does not include the Poisson distribution as a special case, the distribution can be restricted to allow equidispersion. This can be achieved by solving the nonlinear equation $\sigma^2_{dg}(\mu, a) = \mu$ for a using Equation (30) (marginal equidispersion). However, the one-parameter BDG distribution $\mathcal{BG}(\mu)$ offers a conceptually insightful alternative (latent equidispersion), which is analytically tractable ($a = 1$).

In order to determine which equidispersion balanced discrete gamma model (marginal vs. latent equidispersion) is the most appropriate when seeking the parsimonious flexible count regression model, the JSD measure was computed for fixed mean count μ between the Poisson distribution and the BDG distribution under marginal, as well as latent equidispersion restrictions. The results displayed in Figure 5B against the mean count indicate that when restricted to be equidispersed, the BDG distribution becomes similar to the Poisson distribution as per the low JSD values (JSD < 0.015 bit). It can be observed that the marginally equidispersed BDG distribution is the closest to the Poisson distribution only for a very low mean count ($\mu \leq 1.1176$). For a larger mean count ($\mu > 1.1176$), the one-parameter BDG distribution is closer to the Poisson distribution than the marginally equidispersed BDG distribution, although the difference becomes unnoticeable for $\mu > 10$.

It appears that the one-parameter BDG distribution based count regression model will be an effective parsimonious (few parameters and more tractable) model [30] that can be fit to observed data to check the appropriateness of an equidispersion model. Therefore, while a BDG regression model will allow exact inference in flexible count modeling, testing for latent equidispersion will allow recovering a near Poisson regression model when supported by observed data.

4. Conclusions

With a view toward allowing exact inference in flexible count regression models, this work describes balanced discretization, a method for simulating and modeling integer valued data starting from a continuous random variable, through the use of a probabilistic rounding mechanism. Most of the existing alternatives were built to conserve a specific characteristic of the continuous variable, e.g., the failure rate [31] and the survival [14] functions for modeling reliability data. Our proposal preserves expectation and is thus appropriate for count regression. The method is very close to the discretizing approach of [17], which also preserves the mean value, but requires an a priori double truncation of the continuous variable and introduces a tuning parameter. Physical interpretation is an important selection criterion for choosing an appropriate discretization method [32]. As such, our proposal is motivated by a real-world generating mechanism and provides a physical interpretation for the mean-preserving method of [3]. Although balanced discrete distributions can model any count data, it may not be appropriate for aging data for which the integer part or the ceil is generally used [14] so that discrete concentrations are a better choice.

The flexibility of the balanced discrete gamma family developed from the continuous gamma distribution illustrates the potential of the balanced discretization method for capturing any level of dispersion in observed count data. In addition to flexibility, the balanced discrete gamma family turns out to be similar to the Poisson distribution when restricted to be equidispersed (marginal equidispersion) and when constructed using an equidispersed continuous gamma distribution (latent equidispersion). Based on this, we conjecture that while covering all types of dispersion, a flexible count regression model based on the balanced discrete gamma distribution will allow recovering a near Poisson distribution model when the data are Poisson distributed. Future research will target the use of the balanced discrete distribution in count regression analysis. The extension of balanced discretization to a multivariate setting is also considered to handle count data grouped by some sampling units and mixtures of count and continuous responses.

Abbreviations

The following abbreviations are used in this manuscript:

$\mathcal{BER}$	Bernoulli (distribution)
$\mathcal{BE}$	Balanced discrete exponential (distribution)
$\mathcal{BD}$	Balanced discrete (distribution)
$\mathcal{BG}$	Balanced discrete gamma (distribution)
BDG	Balanced discrete gamma
$\mathcal{CD}$	Continuous distribution (distribution)
$\mathcal{DC}$	Discrete concentration (distribution)
$\mathcal{G}$	Gamma (distribution)
ID	Index of dispersion
JSD	Jensen–Shannon divergence
pdf	probability density function
pmf	probability mass function
quf	quantile function
suf	survival function

Appendix A. Proofs of Lemmas and Propositions

Appendix A.1. Proof of Lemma 1

Proof of Lemma 1. By the definition in Equation (5), the probability of observing $Y = y$ given that $X = x$ with $y \leq x < y + 1$ is $1 - r = 1 - x + y$ with $r = x - y$. Thus, the probability of observing $Y = y$ and $y \leq X < y + 1$ is the integral of $[(y + 1) - x]f_X(x|\boldsymbol{\theta})$ with respect to x over $(y, y + 1)$, i.e.,

$$
\begin{aligned}
P(Y = y \text{ and } y \leq X < y + 1) &= \int_y^{y+1} [(y + 1) - x]f_X(x|\boldsymbol{\theta})dx \\
&= (y + 1)\int_y^{y+1} f_X(x|\boldsymbol{\theta})dx - \int_y^{y+1} x f_X(x|\boldsymbol{\theta})dx
\end{aligned}
$$

which proves Equation (8). Using the same argument on the probability of observing $Y = y + 1$ given that $y \leq X < y + 1$ leads to $P(Y = y + 1 \text{ and } y \leq X \leq y + 1)$ equaling the integral of $(x - y)f_X(x|\boldsymbol{\theta})$ with respect to x over $(y, y + 1)$ yields Equation (9). Next, since Y is discrete and takes one of the two values y and $y + 1$ when $y \leq X < y + 1$, the partial expectation of Y is:

$$
\begin{aligned}
\mathrm{E}_Y[Y|y \le X < y+1] &= yP(Y = y \text{ and } y \le X < y+1) \\
&\quad + (y+1)P(Y = y+1 \text{ and } y \le X < y+1) \\
&= P(Y = y+1 \text{ and } y \le X < y+1) \\
&\quad + y[P(Y = y \text{ and } y \le X < y+1) \\
&\quad + P(Y = y+1 \text{ and } y \le X < y+1)].
\end{aligned}
$$

Replacing $P(Y = y \text{ and } y \le X < y+1) + P(Y = y+1 \text{ and } y \le X < y+1)$ by the equivalent probability $P(y \le X < y+1) = F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta})$ and using Equation (9) to obtain $P(Y = y+1 \text{ and } y \le X < y+1)$ result in Equation (10). □

Appendix A.2. Proof of Proposition 1

Proof of Proposition 1. It follows from the defining mechanism in Equation (5) that the unique ways to obtain $Y = y$ are $(U = 1 \text{ and } y-1 \le X < y)$ and $(U = 0 \text{ and } y \le X < y+1)$. In other words, $Y = y$ is equivalent to $y-1 \le X < y$ or $y \le X < y+1$. Since the two instances are mutually exclusive, this gives:

$$
\begin{aligned}
f_Y(y|\boldsymbol{\theta}) &= P(Y = y \text{ and } y-1 \le X < y) + P(Y = y \text{ and } y \le X < y+1) \\
&= \mathrm{E}_X(1, y-1|\boldsymbol{\theta}) - (y-1)[F_X(y|\boldsymbol{\theta}) - F_X(y-1|\boldsymbol{\theta})] \\
&\quad + (y+1)[F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta})] - \mathrm{E}_X(1, y|\boldsymbol{\theta})
\end{aligned}
$$

where the second equality follows from replacing y by $y-1$ in Equation (9) to compute the probability $P(Y = y \text{ and } y-1 \le X < y)$ and using Equation (8) to obtain $P(Y = y \text{ and } y \le X < y+1)$. Rearranging the right-hand side of the last equation as $f_Y(y|\boldsymbol{\theta}) = (y-1)F_X(y-1|\boldsymbol{\theta}) + [-(y-1) - (y+1)]F_X(y|\boldsymbol{\theta}) + (y+1)F_X(y+1|\boldsymbol{\theta}) + \mathrm{E}_X(1, y-1|\boldsymbol{\theta}) - \mathrm{E}_X(1, y|\boldsymbol{\theta})$ yields Equation (11). Again, using the defining mechanism in Equation (5), it follows that $Y \le y$ is equivalent to $X < y$ or $\{Y = y \text{ and } y \le X < y+1\}$. Since the two instances are mutually exclusive, this results in using Equation (8) in:

$$
\begin{aligned}
F_Y(y|\boldsymbol{\theta}) &= P(X < y) + P(Y = y \text{ and } y \le X < y+1) \\
&= F_X(y|\boldsymbol{\theta}) + (y+1)[F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta})] - \mathrm{E}_X(1, y|\boldsymbol{\theta})
\end{aligned}
$$

which proves Equation (12) and implies that (a) $F_X(y|\boldsymbol{\theta}) < F_Y(y|\boldsymbol{\theta}) < F_X(y+1|\boldsymbol{\theta})$. The suf is obtained as $S_Y(y|\boldsymbol{\theta}) = P(X \ge y) + P(Y = y \text{ and } y-1 \le X < y)$ from the definition $S_Y(y|\boldsymbol{\theta}) = P(Y \ge y)$, which straightforwardly results in Equation (13) on replacing $P(X \ge y) = S_X(y|\boldsymbol{\theta})$ and using Equation (9) properly to compute $P(Y = y \text{ and } y-1 \le X < y)$. From the definition of the quantile function for $0 \le u \le 1$ as $Q_Y(u|\boldsymbol{\theta}) = \inf\{y \in \mathbb{Z}|F_Y(y|\boldsymbol{\theta}) \ge u\}$, $y = Q_Y(u|\boldsymbol{\theta})$ implies the inequality $F_Y(y-1|\boldsymbol{\theta}) < u \le F_Y(y|\boldsymbol{\theta})$. Let $q_\mathrm{o} = Q_X(u|\boldsymbol{\theta})$, and set $x_\mathrm{o} = \lfloor q_\mathrm{o} \rfloor$ and $u_\mathrm{o} = F_Y(x_\mathrm{o}|\boldsymbol{\theta})$. By the inequality (a), we have on the one hand, (b) if $u = F_Y(y-1|\boldsymbol{\theta})$, then $q_\mathrm{o} \in (y-1, y)$, and on the other hand, (c) if $u = F_Y(y|\boldsymbol{\theta})$, then $q_\mathrm{o} \in (y, y+1)$. Since $F_Y(\cdot|\boldsymbol{\theta})$ is increasing, (b) and (c) result in $q_\mathrm{o} \in (y-1, y+1)$, and thus, $x_\mathrm{o} \in \{y-1, y\}$ or equivalently $y \in \{x_\mathrm{o}, x_\mathrm{o}+1\}$. Hence, $Q_Y(u|\boldsymbol{\theta}) = x_\mathrm{o}$ if $u_\mathrm{o} \ge u$ and $Q_Y(u|\boldsymbol{\theta}) = x_\mathrm{o}+1$ otherwise. □

Appendix A.3. Proof of Proposition 2

Proof of Proposition 2. Applying the law of iterated expectations [33] (Equation (2)) to the representation in Equation (5), we have $\mu_Y(\boldsymbol{\theta}) = \mathrm{E}_X\left[\mathrm{E}_{U|X}[Y]\right]$. However, $\mathrm{E}_{U|X}[Y] = \mathrm{E}_{U|X}[Z + U] = Z + \mathrm{E}_{U|X}[U]$ with $Z = \lfloor X \rfloor$. Then, from $\mathrm{E}_{U|X}[U] = R = X - Z$, we get $\mathrm{E}_{U|X}[Y] = X$, which results in $\mu_Y(\boldsymbol{\theta}) = \mathrm{E}_X[X]$ and proves that $\mu_Y(\boldsymbol{\theta}) = \mu_X(\boldsymbol{\theta})$. Using Equation (3) in [33], we have $\sigma_Y^2(\boldsymbol{\theta}) = \mathrm{Var}_X\left[\mathrm{E}_{U|X}[Y]\right] + \mathrm{E}_X\left[\mathrm{Var}_{U|X}[Y]\right]$. Equation (15) then follows from using $\mathrm{Var}_X\left[\mathrm{E}_{U|X}[Y]\right] = \mathrm{Var}_X[X]$ and $\mathrm{Var}_{U|X}[Y] = R(1 - R)$. Moreover, R

satisfies $0 \leq R < 1$ and $\mathrm{Var}_{U|X}[Y] > 0$ so that $0 < R(1-R) < 1/4$, but also $R(1-R) < R \leq X$; hence, $0 < \zeta_0(\boldsymbol{\theta}) < 1/4$ and $\zeta_0(\boldsymbol{\theta}) < \mathrm{E}[X]$, and $0 < \zeta_0(\boldsymbol{\theta}) < min\{\mathrm{E}[X], 1/4\}$ follows. Using $R = X - Z$ gives $R(1-R) = R - R^2 = (2Z+1)X - X^2 - Z(Z+1)$. Then, with $f_X(\cdot|\boldsymbol{\theta})$, the pdf of X,

$$
\begin{aligned}
\zeta_0(\boldsymbol{\theta}) &= \mathrm{E}_X[R(1-R)] \\
&= \int_{-\infty}^{\infty} \Big[(2z+1)x - x^2 - z(z+1)\Big] f_X(x|\boldsymbol{\theta}) dx \ \text{ with } \ z = \lfloor x \rfloor \\
&= \sum_{z=-\infty}^{\infty} \int_{z}^{z+1} \Big[(2z+1)x - x^2 - z(z+1)\Big] f_X(x|\boldsymbol{\theta}) dx \\
&= \sum_{z=-\infty}^{\infty} \bigg\{ (2z+1) \int_{z}^{z+1} x f_X(x|\boldsymbol{\theta}) dx - \int_{z}^{z+1} x^2 f_X(x|\boldsymbol{\theta}) dx \\
&\qquad\qquad - z(z+1) \int_{z}^{z+1} f_X(x|\boldsymbol{\theta}) dx \bigg\};
\end{aligned}
$$

hence, Equation (16) follows. $\square$

Appendix A.4. Proof of Proposition 3

Proof of Proposition 3. Using Equation (5), Y^n can be represented as $Y^n = (Z+U)^n$, which expands as $Y^n = \sum_{i=0}^{n} \binom{n}{i} Z^i U^{n-i}$ giving $Y^n = Z^n + \sum_{i=0}^{n-1} \binom{n}{i} Z^i U$ since $U^j = U$ for $j \in \mathbb{N}_+$, and Equation (20) follows. Next, using the law of iterated expectations, we have $\mu_{ZU}^{(i)}(\boldsymbol{\theta}) = \mathrm{E}_Z\Big[Z^i \mathrm{E}_{U|Z}[U] \Big]$. Since $Z = \lfloor X \rfloor$, $Z = z$ is equivalent to $z \leq X < z+1$, hence $\mathrm{E}_{ZU}[U|Z = z] = \mathrm{E}_{X,U}[U|z \leq X < z+1]$. However, we have by Equation (9) the identity $\mathrm{E}_{X,U}[U|z \leq X < z+1] = -z[F_X(z+1|\boldsymbol{\theta}) - F_X(z|\boldsymbol{\theta})] + \mathrm{E}_X(1, z|\boldsymbol{\theta})$ so that we get the identity $Z^i \mathrm{E}_{Z,U}[U|Z = z] = -z^{i+1}[F_X(z+1|\boldsymbol{\theta}) - F_X(z|\boldsymbol{\theta})] + z^i \mathrm{E}_X(1, z|\boldsymbol{\theta})$. Summing the latter partial expectations for $z \in \mathbb{Z}$ yields $\mu_{ZU}^{(i)}(\boldsymbol{\theta}) = -\mu_Z^{(i+1)}(\boldsymbol{\theta}) + \sum_{z=-\infty}^{\infty} z^i \mathrm{E}_X(1, z|\boldsymbol{\theta})$ since $F_X(z+1|\boldsymbol{\theta}) - F_X(z|\boldsymbol{\theta})$ is the probability mass of the discrete concentration of X (see Equation (1)). $\square$

Appendix A.5. Proof of Proposition 4

Proof of Proposition 4. Given $Y = y$, U remains Bernoulli distributed. Moreover, $Y = y$ and $U = 1$ are equivalent to $Y = y$ and $y - 1 \leq X < y$. The success probability of U given $Y = y$ is thus $p_y = [f_Y(y|\boldsymbol{\theta})]^{-1}[P(Y = y \text{ and } y - 1 \leq X < y)]$ by Bayes's rule. Note the identity $P(Y = y \text{ and } y - 1 \leq X < y) = \mathrm{E}_X(1, y-1|\boldsymbol{\theta}) - (y-1)[F_X(y|\boldsymbol{\theta}) - F_X(y-1|\boldsymbol{\theta})]$, which follows by Equation (9) on using $y - 1$ instead of y. The expression of p_y then follows as given in Equation (21). From Equation (5), the conditional density of Y given $X = z$ and $U = u$ is $f_{Y|X,U}(y|X = x, U = u) = I_{(z,z+1)}(x)$ with $z = y - u$. The likelihood (joint density and probability mass) of X, U, and Y is thus:

$$
\begin{aligned}
f_{X,U,Y}(x, u, y) &= f_{U|X}(u|X = x) f_X(x|\boldsymbol{\theta}) I_{(y-u, y-u+1)}(x) \\
&= (x - \lfloor x \rfloor)^u (1 + \lfloor x \rfloor - x)^{1-u} f_X(x|\boldsymbol{\theta}) I_{(y-u, y-u+1)}(x) \\
&= (x - y + u)^u (1 + y - u - x)^{1-u} f_X(x|\boldsymbol{\theta}) I_{(y-u, y-u+1)}(x)
\end{aligned}
$$

where the first line follows by Bayes' rule, and the last line follows on using $y = \lfloor x \rfloor + u$. Summing $f_{X,U,Y}(x, u, y)$ over $u \in \{0, 1\}$, we get:

$$
f_{X,Y}(x, y) = (1 - y + x) f_X(x|\boldsymbol{\theta}) I_{(y-1, y)}(x) + (1 + y - x) f_X(x|\boldsymbol{\theta}) I_{(y, y+1)}(x)
$$

The pdf in Equation (22) then follows by Bayes' rule as $f_{X|Y}(x|Y = y, \boldsymbol{\theta}) = f_{X,Y}(x,y) / f_Y(y|\boldsymbol{\theta})$. Finally, Equation (23) follows from a direct integration as:

$$
\begin{aligned}
E_{X|Y}[X^n|Y = y, \boldsymbol{\theta}] &= \frac{1}{f_Y(y|\boldsymbol{\theta})} \left[\int_{y-1}^{y} x^n (1 - y + x) f_X(x|\boldsymbol{\theta}) dx \right. \\
&\qquad + \left. \int_{y}^{y+1} x^n (1 + y - x) f_X(x|\boldsymbol{\theta}) dx \right] \\
&= \frac{1}{f_Y(y|\boldsymbol{\theta})} \left[(1 - y) \int_{y-1}^{y} x^n f_X(x|\boldsymbol{\theta}) dx + \int_{y-1}^{y} x^{n+1} f_X(x|\boldsymbol{\theta}) dx \right. \\
&\qquad + \left. (1 + y) \int_{y}^{y+1} x^n f_X(x|\boldsymbol{\theta}) dx - \int_{y}^{y+1} x^{n+1} f_X(x|\boldsymbol{\theta}) dx \right].
\end{aligned}
$$

$\square$

References

1. Hilbe, J.M. *Negative Binomial Regression*; Cambridge University Press: Cambridge, UK, 2011; p. 573.
2. Sellers, K.F.; Shmueli, G. Data dispersion: now you see it . . . now you don't. *Commun. Stat. Theory Methods* **2013**, *42*, 3134–3147, doi:10.1080/03610926.2011.621575.
3. Hagmark, P.E. On construction and simulation of count data models. *Math. Comput. Simul.* **2008**, *77*, 72–80, doi:10.1016/j.matcom.2007.01.037.
4. Wedderburn, R.W. Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. *Biometrika* **1974**, *61*, 439–447, doi:10.1093/biomet/61.3.439.
5. Consul, P.; Famoye, F. Generalized Poisson regression model. *Commun. Stat. Theory Methods* **1992**, *21*, 89–109, doi:10.1080/03610929208830766.
6. Bonat, W.H.; Jørgensen, B.; Kokonendji, C.C.; Hinde, J.; Demétrio, C.G. Extended Poisson–Tweedie: Properties and regression models for count data. *Stat. Model.* **2018**, *18*, 24–49, doi:10.1177/1471082X17715718.
7. Conway, R.W.; Maxwell, W.L. A queuing model with state dependent service rates. *J. Ind. Eng.* **1962**, *12*, 132–136.
8. Efron, B. Double exponential families and their use in generalized linear regression. *J. Am. Stat. Assoc.* **1986**, *81*, 709–721, doi:10.1080/01621459.1986.10478327.
9. Zou, Y.; Geedipally, S.R.; Lord, D. Evaluating the double Poisson generalized linear model. *Accid. Anal. Prev.* **2013**, *59*, 497–505, doi:10.1016/j.aap.2013.07.017.
10. Winkelmann, R. A count data model for gamma waiting times. *Stat. Pap.* **1996**, *37*, 177–187, doi:10.1007/BF02926581.
11. Cameron, A.C.; Johansson, P. Count data regression using series expansions: With applications. *J. Appl. Econom.* **1997**, *12*, 203–223, doi:10.1002/(SICI)1099-1255(199705)12:3<203::AID-JAE446>3.0.CO;2-2.
12. Klakattawi, H.; Vinciotti, V.; Yu, K. A simple and adaptive dispersion regression model for count data. *Entropy* **2018**, *20*, 142, doi:10.3390/e20020142.
13. Zeviani, W.M.; Ribeiro, P.J., Jr.; Bonat, W.H.; Shimakura, S.E.; Muniz, J.A. The Gamma-count distribution in the analysis of experimental underdispersed data. *J. Appl. Stat.* **2014**, *41*, 2616–2626, doi:10.1080/02664763.2014.922168.
14. Chakraborty, S. Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. *J. Stat. Distrib. Appl.* **2015**, *2*, 30, doi:10.1186/s40488-015-0028-6.
15. Plan, E.L. Modeling and simulation of count data. *CPT Pharmacomet. Syst. Pharmacol.* **2014**, *3*, 1–12, doi:10.1038/psp.2014.27.
16. Veraart, A.E. Modeling, simulation and inference for multivariate time series of counts using trawl processes. *J. Multivar. Anal.* **2019**, *169*, 110–129, doi:10.1016/j.jmva.2018.08.012.
17. Roy, D.; Dasgupta, T.A. A discretizing approach for evaluating reliability of complex systems under stress-strength model. *IEEE Trans. Reliab.* **2001**, *50*, 145–150, doi:10.1109/24.963121.
18. Hagmark, P.E. A new concept for count distributions. *Stat. Probab. Lett.* **2009**, *79*, 1120–1124, doi:10.1016/j.spl.2009.01.006.
19. Hagmark, P.E. An Exceptional Generalization of the Poisson Distribution. *Open J. Stat.* **2012**, *2*, 313–318, doi:10.4236/ojs.2012.23039.
20. Chakraborty, S. A New Discrete Distribution Related to Generalized Gamma Distribution and Its Properties. *Commun. Stat. Theory Methods* **2013**, *44*, 1691–1705, doi:10.1080/03610926.2013.781635.
21. Holland, B.S. Some Results on the discretization of continuous probability distributions. *Technometrics* **1975**, *17*, 333–339, doi:10.1080/00401706.1975.10489337.
22. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* **1977**, *39*, 1–38, doi:10.1111/j.2517-6161.1977.tb01600.x.
23. Lawless, J.F. *Statistical Models and Methods for Lifetime Data*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley-Interscience: New York, NY, USA, 2002.
24. Chakraborty, S.; Chakravarty, D. Discrete Gamma Distributions: Properties and Parameter Estimations. *Commun. Stat. Theory Methods* **2012**, *41*, 3301–3324, doi:10.1080/03610926.2011.563014.

25. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
26. MATLAB. *Version 9.0.0 (R2016a)*; The MathWorks Inc.: Natick, MA, USA, 2016.
27. Roy, D. Reliability measures in the discrete bivariate set-up and related characterization results for a bivariate geometric distribution. *J. Multivar. Anal.* **1993**, *46*, 362–373, doi:10.1006/jmva.1993.1065.
28. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151, doi:10.1109/18.61115.
29. Zhang, X.; Delpha, C.; Diallo, D. Performance of Jensen Shannon Divergence in Incipient Fault Detection and Estimation. In Proceedings of the ICASSP 2019, 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2742–2746.
30. Daganzo, C.F.; Gayah, V.V.; Gonzales, E.J. The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions. *EURO J. Transp. Logist.* **2012**, *1*, 47–65.
31. Roy, D.; Ghosh, T. A new discretization approach with application in reliability estimation. *IEEE Trans. Reliab.* **2009**, *58*, 456–461, doi:10.1109/TR.2009.2028093.
32. Bracquemond, C.; Gaudoin, O. A survey on discrete lifetime distributions. *Int. J. Reliab. Qual. Saf. Eng.* **2003**, *10*, 69–98, doi:10.1142/S0218539303001007.
33. Kaehler, J. Laws of iterated expectations for higher order central moments. *Stat. Pap.* **1990**, *31*, 295–299.