Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0332.v1

Article

On the Discretization of Continuous Probability Distributions for Flexible Count Regression

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 (Benin); romain.glelekakai@fsa.uac.bj
[2] Faculté des Sciences et Techniques, Université d'Abomey-Calavi (Benin); jonas.doumate@fast.uac.bj
* Correspondence: romain.glelekakai@fsa.uac.bj
† These authors contributed equally to this work.

Abstract: Most existing flexible count regression models allow only approximate inference. 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 overdispersion). The new method, referred as "balanced discretization", consists in 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 in terms of the continuous distribution. For illustrative purposes, we have developed the family of balanced discrete gamma distributions which can model equi-, under- and overdispersed 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 have shown that under 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 dispersion, a count regression model based on the balanced discrete gamma distribution will allow recovering a near Poisson distribution model fit when the data is Poisson distributed.

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

Citation: Tovissodé, C.F.; Honfo, H.; Doumatè, J.T.; Glèlè Kakaï, R. On the Discretization of Continuous Probability Distributions for Flexible Count Regression. *Mathematics* **2021**, *1*, 0. https://doi.org/

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 analysed 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 unproperly normalized probability mass functions for underdispersion situations (Quasi-Poisson [4], Consul's generalized Poisson [5] and Extended Poisson–Tweedy regresions [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-non parametric Poisson polynomial [11] and discrete Weibull [12] models). The later 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 turning parameter. Proposals in [3,18] and [19] offer solutions for constructing count variables with 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 describes a discretization approach which 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, balanced discretization is suited for regression analysis where estimation of covariate effects on the mean count is of the highest interest.

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 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 probability mass, cumulative distribution, 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 shall 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 $\lfloor x \rfloor$ the integer part of any real x. Following [14], we shall 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 probability density function (pdf), the cumulative distribution function (cdf), the survival function (suf) and the quantile function (quf) 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 probability mass function (pmf), the cdf, the suf and the quf of Y, all indexed by a parameter vector $\boldsymbol{\theta}$. Continuous probability distributions are assumed non degenerated. A Bernoulli random variable with success probability $\rho \in [0, 1]$ is denoted $\mathcal{BER}(\rho)$. We shall use the shorthand "Eq" for equation citations.

2.2. Remainders

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 pmf and suf

$$f_Y(y|\boldsymbol{\theta}) = F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta}) \quad (1)$$
$$S_Y(y|\boldsymbol{\theta}) = S_X(y|\boldsymbol{\theta}) \quad (2)$$

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 rth moment about zero of Y is

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

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 $0 < \mathrm{E}\left[U^2\right] < \mathrm{E}[U] < 1$, 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

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

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

2.3. Motivating Example

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 in millimeter (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 the 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 $x - z$.

2.4. Definition

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

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

where $z = \lfloor x \rfloor$ and $r = x - z$. Let $\mathrm{E}_X(r, y|\boldsymbol{\theta})$ denote the rth partial moment

$$\mathrm{E}_X(r, y|\boldsymbol{\theta}) = \int_y^{y+1} x^r f_X(x|\boldsymbol{\theta}) dx \tag{6}$$

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 Eq (5) preserves partial expectations $\mathrm{E}_X(1, y|\boldsymbol{\theta})$ of the continuous variable as shown by the Eq (10) of the following lemma.

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

$$\begin{aligned} P(Y = y \,\&\, y \le X < y+1) &= (y+1)H_X(y|\boldsymbol{\theta}) - \mathrm{E}_X(1, y|\boldsymbol{\theta}) && (8) \\ P(Y = y+1 \,\&\, y \le X < y+1) &= \mathrm{E}_X(1, y|\boldsymbol{\theta}) - yH_X(y|\boldsymbol{\theta}) && (9) \\ \mathrm{E}_Y[Y|y \le X < y+1] &= \mathrm{E}_X(1, y|\boldsymbol{\theta}) && (10) \end{aligned}$$

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

2.5. 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 $0 \le u \le 1$ by*

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

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

Note from Eq (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 $M \in \mathbb{R}_+$. Eq (13) emphasies that the balanced discretization method does not preserve the suf of the continuous distribution, unlike discrete concentration (see Eq (2)). Nevertheless, the balanced discrete cdf and suf satisfy the inequalities $F_X(y|\boldsymbol{\theta}) \le F_Y(y|\boldsymbol{\theta}) \le F_X(y+1|\boldsymbol{\theta})$ (with equalities when the support of X is upper bounded by y) and $S_X(y|\boldsymbol{\theta}) \le S_Y(y-1|\boldsymbol{\theta}) \le S_X(y-1|\boldsymbol{\theta})$ (with equalities when the support of X is lower bounded by y).

By Eq (14), balanced discretization preserves somehow 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 $F_Y(\lfloor m_X \rfloor|\boldsymbol{\theta}) > 1/2$.

2.6. Moments and Index of Dispersion

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0332.v1

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 cheap 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 index of dispersion of a balanced discrete distribution from Proposition 2.

Corollary 1 (Index of dispersion)**.** *Let $Y \sim \mathcal{BD}(\boldsymbol{\theta})$ 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 $Y \sim \mathcal{BD}(\boldsymbol{\theta})$ the balanced discrete counterpart of $X \sim \mathcal{CD}(\boldsymbol{\theta})$. The rth moment of Y satisfies for $r \in \mathbb{N}_+$*

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

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

2.7. Conditional Distributions of Latent Continuous and Binary Outcomes

Although the balanced discrete distribution was motivated by the need of mean-parametrizable flexible discrete probability distributions, it may be used to model any

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0332.v1

continuous outcome measured to a few number of 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 cheap. The following result provides expressions for these purposes.

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

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

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

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

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

where $\rho_y = [f_Y(y|\boldsymbol{\theta})]^{-1} [E_X(1, y-1|\boldsymbol{\theta}) - (y-1) H_X(y|\boldsymbol{\theta})]$ 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 1 if $x \in \mathbb{A}$ and 0 otherwise.

Note from Eq (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

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

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.

2.8. Link with Mean-Preserving Discretization

Recall from Eq (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 $\int F_X(x|\boldsymbol{\theta}) dx = x F_X(x|\boldsymbol{\theta}) - \int x f_X(x|\boldsymbol{\theta}) dx$, we have $F_Y(y|\boldsymbol{\theta}) = (y+1) F_X(y+1|\boldsymbol{\theta}) - y F_X(y|\boldsymbol{\theta}) - \mathrm{E}_X(1, y|\boldsymbol{\theta})$. 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 $F_Y(y|\boldsymbol{\theta}) = F_X(y|\boldsymbol{\theta}) + (y+1)[F_X(y+1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta})] - \mathrm{E}_X(1, y|\boldsymbol{\theta})$, *i.e.* the cdf in Eq (12). Therefore,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0332.v1

balanced discretization as defined in Eq (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-square 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 $\gamma(x,a) = \int_0^x u^{a-1}\mathrm{e}^{-u}du/\Gamma(a)$ is the lower incomplete gamma ratio for $(a,x) \in \mathbb{R}_+^2$ and $\Gamma(a) = \int_0^\infty u^{a-1}\mathrm{e}^{-u}du$ is the gamma function. A random variable $X \sim \mathcal{G}(b,a)$ has expectation b/a, variance $b/(a^2)$, and rth order partial moment $E_g(r,y|b,a) = \mathrm{E}_X[X^r|y \leq X < y+1]$ given by

$$E_g(r,y|b,a) = \frac{\Gamma(b+r)}{a^r\Gamma(b)}[\gamma(a(y+1),b+r)) - \gamma(ay,b+r)]. \tag{26}$$

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 (BDG) distribution denoted $\mathcal{BG}(\mu,a)$ for $(\mu,a) \in \mathbb{R}_+^2$, if it is generated by the discretization mechanism in Eq (5) with $X \sim \mathcal{G}(a\mu,a)$. By Proposition 2, a $\mathcal{BG}(\mu,a)$ variable has expectation μ. Using Eq (26), 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 cdf for $y \in \mathbb{N}$, the variance and the index of dispersion of Y are respectively*

$$\begin{aligned}
f_{dg}(y|\mu,a) &= (y-1)\gamma(a(y-1),b) - 2y\gamma(ay,b) + (y+1)\gamma(a(y+1),b) \\
&\quad -\mu[\gamma(a(y-1),b+1) - 2\gamma(ay,b+1) + \gamma(a(y+1),b+1)] && (27)\\
F_{dg}(y|\mu,a) &= (y+1)\gamma(a(y+1),b) - y\gamma(ay,b) \\
&\quad -\mu[\gamma(a(y+1),b+1) - \gamma(ay,b+1)] && (28)\\
\sigma^2_{dg}(\mu,a) &= a^{-1}\mu + \zeta_{g0}(\mu,a) && (29)\\
\mathrm{ID}_{dg}(\mu,a) &= a^{-1} + \mu^{-1}\zeta_{g0}(\mu,a) \quad \text{where} && (30)\\
\zeta_{g0}(\mu,a) &= \sum_{z=0}^{\infty}\zeta_{g0}(z,\mu,a) \quad \text{with} && (31)\\
\zeta_{g0}(z,\mu,a) &= -\mu\left(\mu + a^{-1}\right)[\gamma(a(z+1),b+2) - \gamma(az,b+2)] \\
&\quad \mu(2z+1)[\gamma(a(z+1),b+1) - \gamma(az,b+1)] \\
&\quad -z(z+1)[\gamma(a(z+1),b) - \gamma(az,b)]. && (32)
\end{aligned}$$

Note that $\zeta_{g0}(\mu,a)$ can be approximated via the truncation mechanism in Eq (19). The one-parameter BDG distribution denoted $\mathcal{BG}(\mu)$ and obtained by setting $a = 1$, cor-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0332.v1

responds to a latent equidispersion mechanism and is marginally slightly overdispersed as indicated by Eq (30) 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 later corresponds to the discrete concentration of the exponential distribution [25].

Figure 1 displays the probability mass function of the BDG distributions with mean values $\mu = 2.5$ and $\mu = 5$. 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 index of dispersion which can assume any positive value, is depicted on Figure 2 for $a \geq 1$. It can be observed in accordance with Eq (30) 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 ($a < 0.5$) whatever the mean value. Very large scales induce in addition to severe underdispersion (ID < 0.2), an oscillating index of dispersion with an amplitude approaching zero as μ increases.

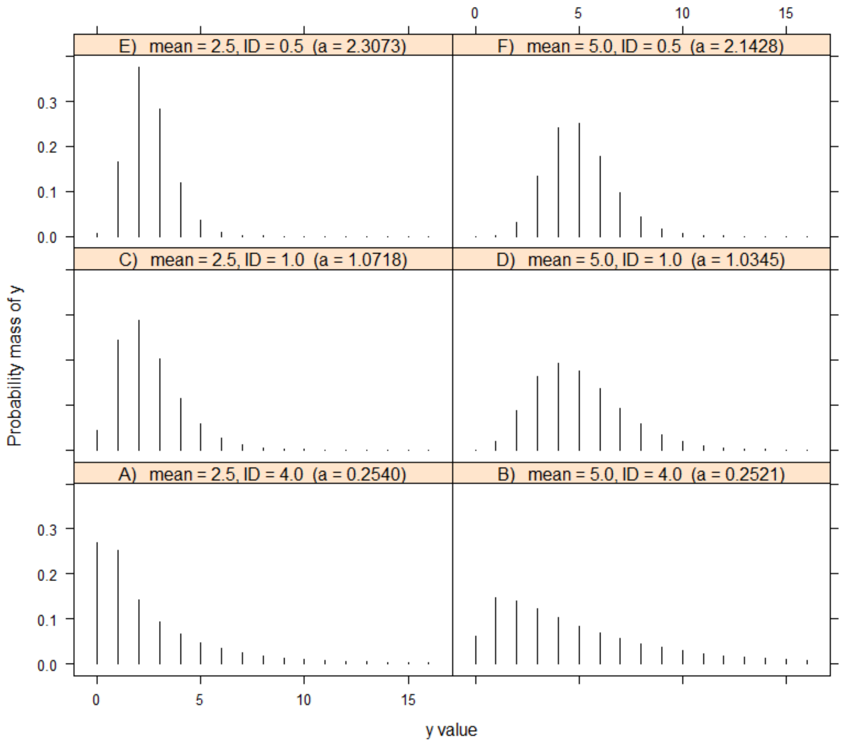

Figure 1. Probability mass plots for the balanced discrete gamma distribution with mean values $\mu = 2.5$ (left panel) and $\mu = 5$ (right panel) and scales a selected to yield Index of Dispersion (ID, variance to mean ratio) of ID = 4 (bottom row), ID = 1 (central row) and ID = 0.5 (top row).

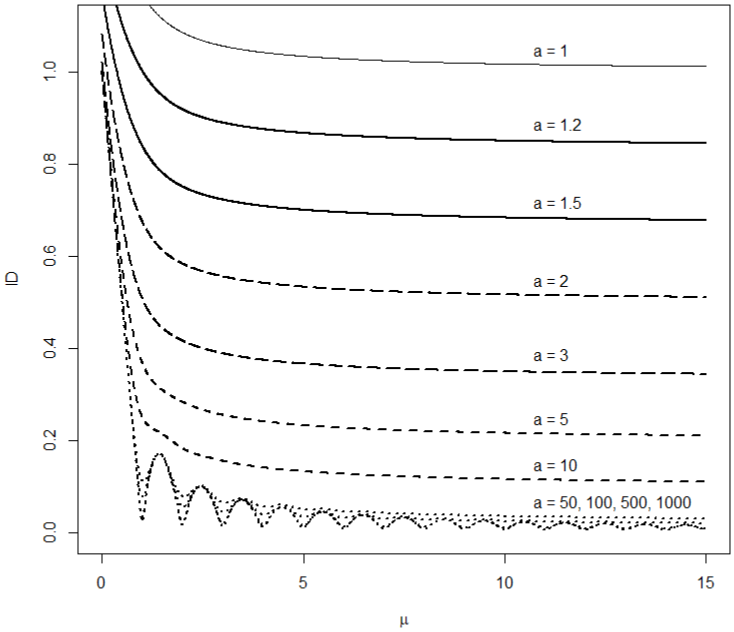

Figure 2. Index of dispersion (ID, variance to mean ratio) of the balanced discrete gamma distribution against the mean value (μ) for selected scale parameter (a) values in the range $[1, 1000]$ mostly corresponding to equidispersion and underdispersion ($0 < \text{ID} \leq 1$).

3.2. Comparison With Some Alternatives

The balanced discretization approach results from a light modification of the discrete concentration method. This section assesses on 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) [26] has many desirable properties which support its use in statistics [27]. The JSD is an information theory measure given for two pmf $p(\cdot)$ and $q(\cdot)$ by [26]

$$JSD(p,q) = K(p,q) + K(q,p) \tag{33}$$

$$\text{where } K(p,q) = \sum_{y=-\infty}^{\infty} p(y) \log_2\left(\frac{p(y)}{0.5p(y) + 0.5q(y)}\right) \tag{34}$$

with the convention $p(y)\log_2(p(y)/(0.5p(y) + 0.5q(y))) = 0$ if $p(y) = 0$. The JSD measures the discrepancy between $p(\cdot)$ and $q(\cdot)$ in *bit*, is bounded as $0 \leq JSD(p,q) \leq 2$, and is zero only if $p(y) = q(y)\ \forall\, y \in \mathbb{Z}$.

3.2.1. Balanced Discretization Versus Discrete Concentration

Figure 3 illustrates the balanced discretization method using the continuous gamma distribution with parameters $a = 1$, $b = 5$. Unlike the discrete concentration whose cdf lays above the continuous cdf, the balanced discrete distribution is constructed so that the continuous cdf interpolates the cdf of the balanced discrete distribution.

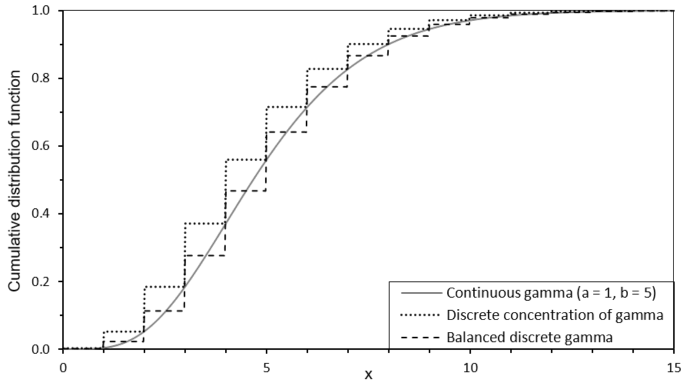

Figure 3. Comparison of the cumulative distribution functions of the balanced discrete gamma and discrete concentration of gamma distributions based on a continuous gamma distribution with scale $a = 1$ and shape $b = 5$.

Curves on Figure 4 (A) 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 ≤ 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 very low mean count ($\mu < 1$) or large scale parameter ($a > 5$, implying ID < 0.45). For very large scale ($a \geq 30$), the JSD becomes erratic, oscillating between minima right before integer values of μ and maxima right after integer values of μ.

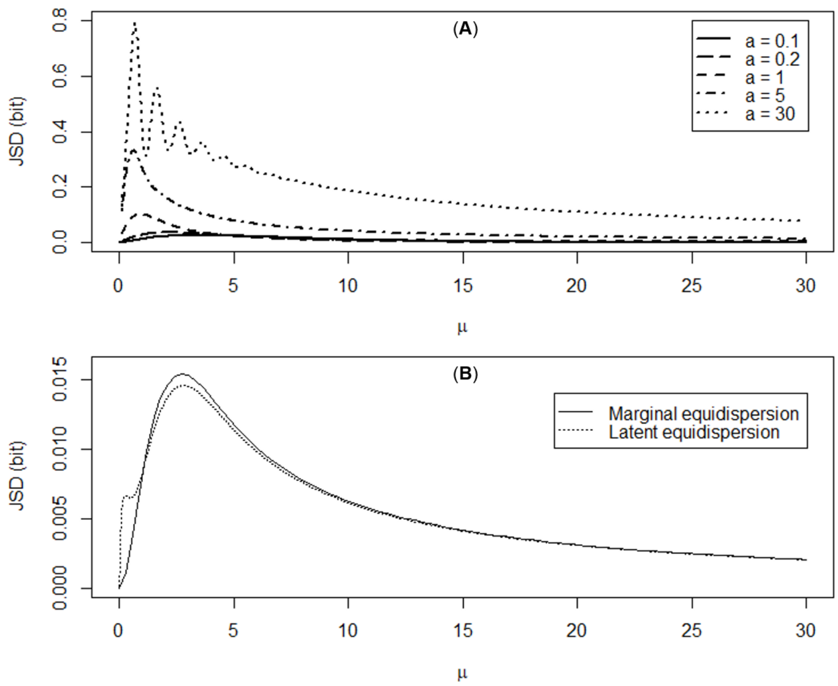

Figure 4. Jensen Shannon divergence (JSD in *bit*) measure between the balanced discrete gamma (BDG) distribution and the corresponding discrete concentration (**A**) and between the Poisson and the BDG distributions under both latent equidispersion (one-parameter BDG distribution) and marginal equidispersion (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 non linear equation $\sigma^2_{dg}(\mu, a) = \mu$ for a using Eq (29) (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 for 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 on Figure 4 (B) 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 very low mean count ($\mu \leq 1.1176$). For 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 [28] that can be fitted 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. Conclusion

With a view to allow 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 existing alternatives were built to conserve a specific characteristic of the continuous variable *e.g.* the failure rate [29] 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 turning parameter. Physical interpretation is an important selection criterion for choosing an appropriate discretization method [30]. As such, our proposal was motivated by a real world generating mechanism and provides a physical interpretation for the mean-preserving method of [3]. Although balanced discret distributions can model any count data, it may not be appropriate for ageing 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 exact inference in flexible count regression analisis. In addition to flexibility, the balanced discrete gamma family turns to be similiar 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 is Poisson distributed. Future research will target the use of 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.

Author Contributions: "Conceptualization, C.F.T. and R.G.; methodology, C.F.T.; software, C.F.T.; validation, C.F.T., H.H.S., J.T.D. and R.G.; formal analysis, C.F.T.; resources, R.G.; writing—original draft preparation, C.F.T.; writing—review and editing, C.F.T., H.H.S., J.T.D. and R.G.; visualization, C.F.T.; supervision, R.G.; project administration, R.G.; funding acquisition, R.G. All authors have read and agreed to the published version of the manuscript."

Funding: "This research received no external funding"

Acknowledgments: The first author gratefully acknowledges financial support from the Centre d'Excellence d'Afrique en Sciences Mathematiques, Informatique et Applications (CEA-SMIA).

Conflicts of Interest: "The authors declare no conflict of interest." "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

Abbreviations

The following abbreviations are used in this manuscript:

 doi:10.20944/preprints202101.0332.v1

$\mathcal{BER}$	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 Eq (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.*

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

which proves Eq (8). Using the same argument on the probability of observing $Y = y + 1$ given that $y \leq X < y + 1$ leads to $P(Y = y + 1 \,\&\, y \leq X \leq y + 1)$ equaling the integral of $(x - y) f_X(x|\boldsymbol{\theta})$ with respect to x over $(y, y + 1)$ which yields Eq (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

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

Replacing $P(Y = y \,\&\, y \leq X < y + 1) + P(Y = y + 1 \,\&\, y \leq X < y + 1)$ by the equivalent probability $P(y \leq X < y + 1) = F_X(y + 1|\boldsymbol{\theta}) - F_X(y|\boldsymbol{\theta})$ and using Eq (9) to obtain $P(Y = y + 1 \,\&\, y \leq X < y + 1)$ results in Eq (10). □

Appendix A.2 Proof of Proposition 1

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0332.v1

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

where the second equality follows from replacing y by $y-1$ in Eq (9) to compute the probability $P(Y=y \,\&\, y-1 \le X < y)$ and using Eq (8) to obtain $P(Y=y \,\&\, y \le X < y+1)$. 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 Eq (11). Again using the defining mechanism in Eq (5), it follows that $Y \le y$ is equivalent to $X < y$ or $\{Y = y$ and $y \le X < y+1\}$. Since the two instances are mutually exclusive, this results on using Eq (8) in

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

which proves Eq (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 $S_Y(y|\boldsymbol{\theta}) = P(X \ge y) + P(Y=y \,\&\, y-1 \le X < y)$ from the definition $S_Y(y|\boldsymbol{\theta}) = P(Y \ge y)$, what straightforwardly results in Eq (13) on replacing $P(X \ge y) = S_X(y|\boldsymbol{\theta})$ and using Eq (9) properly to compute $P(Y=y \,\&\, y-1 \le X < y)$. 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 $q_o = Q_X(u|\boldsymbol{\theta})$ and set $x_o = \lfloor q_o \rfloor$ and $u_o = F_Y(x_o|\boldsymbol{\theta})$. By the inequality (a), we have on one hand (b) if $u = F_Y(y-1|\boldsymbol{\theta})$ then $q_o \in (y-1, y)$ and on the other hand, (c) if $u = F_Y(y|\boldsymbol{\theta})$ then $q_o \in (y, y+1)$. Since $F_Y(\cdot|\boldsymbol{\theta})$ is increasing, (b) and (c) result in $q_o \in (y-1, y+1)$ and thus $x_o \in \{y-1, y\}$ or equivalently $y \in \{x_o, x_o+1\}$. Hence $Q_Y(u|\boldsymbol{\theta}) = x_o$ if $u_o \ge u$ and $Q_Y(u|\boldsymbol{\theta}) = x_o + 1$ otherwise. □

Appendix A.3 Proof of Proposition 2

Proof of Proposition 2. Applying the law of iterated expectations [31] (Eq (2)) to the representation in Eq (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 Eq (3) in [31], we have $\sigma_Y^2(\boldsymbol{\theta}) = \mathrm{Var}_X\left[\mathrm{E}_{U|X}[Y]\right] + \mathrm{E}_X\left[\mathrm{Var}_{U|X}[Y]\right]$. Eq (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 $0 \le R < 1$ and $\mathrm{Var}_{U|X}[Y] > 0$ so that $0 < R(1-R) < 1/4$ but also $R(1-R) < R \le X$ 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,

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2021 doi:10.20944/preprints202101.0332.v1

hence Eq (16) follows. □

Appendix A.4 Proof of Proposition 3

Proof of Proposition 3. Using Eq (5), Y^r can be represented as $Y^r = (Z+U)^r$ which expands as $Y^r = \sum_{i=0}^{r}\binom{r}{i} Z^i U^{n-i}$ giving $Y^r = Z^r + \sum_{i=0}^{r-1}\binom{r}{i} Z^i U$ since $U^j = U$ for $j \in \mathbb{N}_+$, and Eq (20) follows. Next, using the law of iterated expectations, we have $\mu_{ZU}^{(i)}(\boldsymbol{\theta}) = \mathrm{E}_Z\left[Z^i \mathrm{E}_{U|Z}[U]\right]$. 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 Eq (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 Eq (1)). □

Appendix A.5 Proof of Proposition 4

Proof of Proposition 4. Given $Y = y$, U remains Bernoulli distributed. Moreover, $Y = y$ and $U = 1$ is equivalent to $Y = y$ and $y - 1 \leq X < y$. The success probability of U given $Y = y$ is thus $\rho_y = [f_Y(y|\boldsymbol{\theta})]^{-1}[P(Y = y \,\&\, y-1 \leq X < y)]$ by Bayes's rule. Note the identity $P(Y = y \,\&\, y-1 \leq X < y) = \mathrm{E}_X(1, y-1|\boldsymbol{\theta}) - (y-1)[F_X(y|\boldsymbol{\theta}) - F_X(y-1|\boldsymbol{\theta})]$ which follows by Eq (9) on using $y - 1$ instead of y. The expression of ρ_y then follows as given in Eq (21). From Eq (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

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

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

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

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

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

□

References

1. Hilbe, J.M. *Negative binomial regression*; Cambridge University Press, 2011; p. 573.

2. Sellers, K.F.; Shmueli, G. Data dispersion: now you see it . . . now you don't. *Communications in Statistics-Theory and Methods* **2013**, *42*, 3134–3147. doi:https://doi.org/10.1080/03610926.2011.621575.
3. Hagmark, P.E. On construction and simulation of count data models. *Mathematics and Computers in Simulation* **2008**, *77*, 72–80. doi:https://doi.org/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:https://doi.org/10.1093/biomet/61.3.439.
5. Consul, P.; Famoye, F. Generalized Poisson regression model. *Communications in Statistics-Theory and Methods* **1992**, *21*, 89–109. doi:https://doi.org/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. *Statistical Modelling* **2018**, *18*, 24–49. doi:https://doi.org/10.1177/1471082X17715718.
7. Conway, R.W.; Maxwell, W.L. A queuing model with state dependent service rates. *Journal of Industrial Engineering* **1962**, *12*, 132–136.
8. Efron, B. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **1986**, *81*, 709–721. doi:https://doi.org/10.1080/01621459.1986.10478327.
9. Zou, Y.; Geedipally, S.R.; Lord, D. Evaluating the double Poisson generalized linear model. *Accident Analysis & Prevention* **2013**, *59*, 497–505. doi:https://doi.org/10.1016/j.aap.2013.07.017.
10. Winkelmann, R. A count data model for gamma waiting times. *Statistical Papers* **1996**, *37*, 177–187. doi:https://doi.org/10.1007/BF02926581.
11. Cameron, A.C.; Johansson, P. Count data regression using series expansions: with applications. *Journal of Applied Econometrics* **1997**, *12*, 203–223. doi:https://doi.org/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:https://doi.org/10.3390/e20020142.
13. Zeviani, W.M.; Ribeiro Jr, P.J.; Bonat, W.H.; Shimakura, S.E.; Muniz, J.A. The Gamma-count distribution in the analysis of experimental underdispersed data. *Journal of Applied Statistics* **2014**, *41*, 2616–2626. doi:https://doi.org/10.1080/02664763.2014.922168.
14. Chakraborty, S. Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. *Journal of Statistical Distributions and Applications* **2015**, *2*, 30. doi:https://doi.org/10.1186/s40488-015-0028-6.
15. Plan, E.L. Modeling and simulation of count data. *CPT: pharmacometrics & systems pharmacology* **2014**, *3*, 1–12. doi:https://doi.org/10.1038/p
16. Veraart, A.E. Modeling, simulation and inference for multivariate time series of counts using trawl processes. *Journal of Multivariate Analysis* **2019**, *169*, 110–129. doi:https://doi.org/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 Transactions on Reliability* **2001**, *50*, 145–150. doi:https://doi.org/10.1109/24.963121.
18. Hagmark, P.E. A new concept for count distributions. *Statistics & probability letters* **2009**, *79*, 1120–1124. doi:https://doi.org/10.1016/j.spl.200
19. Hagmark, P.E.; others. An Exceptional Generalization of the Poisson Distribution. *Open Journal of Statistics* **2012**, *2*, 313–318. doi:https://doi.org/10.4236/ojs.2012.23039.
20. Chakraborty, S. A New Discrete Distribution Related to Generalized Gamma Distribution and Its Properties. *Communications in Statistics - Theory and Methods* **2013**, *44*, 1691–1705. doi:https://doi.org/10.1080/03610926.2013.781635.
21. Holland, B.S. Some Results on the discretization of continuous probability distributions. *Technometrics* **1975**, *17*, 333–339. doi:https://doi.org/10.1080/00401706.1975.10489337.
22. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)* **1977**, *39*, 1–38. doi:https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.
23. Lawless, J.F. *Statistical Models and Methods for Lifetime Data*, 2 ed.; Wiley Series in Probability and Statistics, Wiley-Interscience, 2002.
24. Chakraborty, S.; Chakravarty, D. Discrete Gamma Distributions: Properties and Parameter Estimations. *Communications in Statistics - Theory and Methods* **2012**, *41*, 3301–3324. doi:https://doi.org/10.1080/03610926.2011.563014.
25. Roy, D. Reliability measures in the discrete bivariate set-up and related characterization results for a bivariate geometric distribution. *Journal of Multivariate Analysis* **1993**, *46*, 362–373. doi:https://doi.org/10.1006/jmva.1993.1065.
26. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory* **1991**, *37*, 145–151. doi:https://doi.org/10.1109/18.61115.
27. Zhang, X.; Delpha, C.; Diallo, D. Performance of Jensen Shannon Divergence in Incipient Fault Detection and Estimation. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 2742–2746.
28. 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 Journal on Transportation and Logistics* **2012**, *1*, 47–65.
29. Roy, D.; Ghosh, T. A new discretization approach with application in reliability estimation. *IEEE Transactions on reliability* **2009**, *58*, 456–461. doi:https://doi.org/10.1109/TR.2009.2028093.
30. Bracquemond, C.; Gaudoin, O. A survey on discrete lifetime distributions. *International Journal of Reliability, Quality and Safety Engineering* **2003**, *10*, 69–98. doi:https://doi.org/10.1142/S0218539303001007.
31. Kaehler, J. Laws of iterated expectations for higher order central moments. *Statistical Papers* **1990**, *31*, 295–299.