Article

A Stochastic Condensation Mechanism for Inducing Underdispersion in Count Models

Chénangnon Frédéric Tovissodé [1] 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.)
* Correspondence: romain.glelekakai@fsa.uac.bj (R.G.K)

Abstract: It is quite easy to stochastically distort an original count variable to obtain a new count variable with relatively more variability than in the original variable. Many popular overdispersion models (variance greater than mean) can indeed be obtained by mixtures, compounding or randomly-stopped sums. There is no analogous stochastic mechanism for the construction of underdispersed count variables (variance less than mean), starting from an original count distribution of interest. This work proposes a generic method to stochastically distort an original count variable to obtain a new count variable with relatively less variability than in the original variable. The proposed mechanism, termed *condensation*, attracts probability masses from the quantiles in the tails of the original distribution and redirect them toward quantiles around the expected value. If the original distribution can be simulated, then the simulation of variates from a condensed distribution is straightforward. Moreover, condensed distributions have a simple mean-parametrization, a characteristic useful in a count regression context. An application to the negative binomial distribution resulted in a distribution allowing under, equi and overdispersion. In addition to graphical insights, fields of applications of special cases of *condensed Poisson* and *condensed negative binomial* distributions were pointed out as an indication of the potential of condensation for a flexible analysis of count data.

Keywords: Erlang process ; condensed distribution ; probability generating function ; algebraic moments ; Poisson remainder distribution; negative binomial remainder distribution

1. Introduction

The study of many physical, biological or social processes often involves some form of event analysis. However, only the total number of events occurring in some time interval is often available [1]. As a result, event analysis is generally reduced to count data analysis. The statistical analysis of count data requires the specification of an hypothesized model for the target response variable Y, *i.e.* a count distribution which describes the mechanism underlying the generation of the counts and the error structure in the target population. In practice, however, real counts often exhibit a variability, which differs from that anticipated under the hypothesized model [2]. The modelling of counts thus, often, faces overdispersion (the observed variance is greater the expected) or underdispersion (the observed variance is lower than the expected) [2,3]. These phenomena result from one or many causes including partial distortion of the observations by the method of ascertainment, lack of independence between individual item responses, contagion, clustering and heterogeneity [1–4]. In fact, an observed count x is a realization of a random variable Y which is a distorted version (the observed probability distribution) of the distribution of X (original distribution) [2].

Distortions to the original distribution often lead to overdispersion. This has motivated claims similar to "overdispersion is the polite statistician's version of Murphy's law: if something can go wrong, it will" [5]. Indeed, the distorsion of the original distribution can be interpreted as the mixture of distributions, which always induces overdispersion. This mixture is related to positive correlation between individual item responses, contagion, clustering or heterogeneity [1–4,6]. Mixture, interpreted as a distorsion mechanism, provides a simple route for the construction of overdispersed count distribution for the analysis of count data: inclusion of observation level random effect to disturb the mean of the original

check for updates

Citation: Tovissodé, C.F.; Glèlè Kakaï, R. A Stochastic Condensation Mechanism for Inducing Underdispersion in Count Models. *Preprints* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: NA.

distribution. The most popular overdispersion model (negative binomial) can indeed be obtained by multiplying the mean of a Poisson variable by a unit mean gamma variable [7]. Poisson-log normal [8] and Poisson-inverse Gaussian distributions are obtained by the same approach [9]. Randomly-stopped sum [10,11] and compounding are less known mechanisms which also leads to overdispersion [3].

Clearly, it is relatively easy to build overdispersion count models, using a stochastic mechanism to distort an original target distribution. Underdispersion models, on the other hand, have been overlooked [12], although they have recently received more attention [2,13–15]. Underdispersion is often associated with negative contagion or correlation, where the occurrence of one event makes future events somewhat less likely [6]. To the best of our knowledge, there is no analogous of stochastic processes such as mixture, compounding or randomly-stopped sum, to produce underdispersed count models.

This paper proposes a generic stochastic mechanism for the construction of an underdispersed version of any integer-valued distribution. The proposal, termed *condensation*, is a generalization of a stochastic relationship between the Morse [16]-Jewell [17] distribution [18,19], *i.e.* the asynchroneous counting distribution of the Erlang process, and the Poisson distribution. It appears that the Fisher's index of dispersion (variance-to-mean ratio) [20] of the condensed variable Y is always less than the index of dispersion of the original variable X. Our method offers a simple method to generalize existing overdispersion count models to allow full dispersion flexibility [21] in the analysis of count data.

The remaining of the paper is organized as follows. Section 2 presents preliminaries on the Erlang process, the Morse-Jewell distribution and its stochastic representation in terms of the Poisson and Bernoulli distributions. Section 3 describes the condensation mechanism and provides the general forms of the probability mass function, the probability generating function and the algebraic moments of a condensed distribution. Section 4 illustrates the proposal on the Poisson and negative binomial distributions, and Section 5 gives concluding remarks.

2. The Erlang Process and the Morse-Jewell Distribution

2.1. The Counting Distributions of the Erlang Process

The Morse-Jewell distribution [16,17] is the asynchronous counting distribution of the Erlang process. The counting distributions of the Erlang process are generalizations of the Poisson distribution, obtained by replacing the exponential inter-events time distribution by an Erlang distribution [22–24] of order $m \in \mathbb{N}^+$ (positive integers). When $m = 1$, the counting process results in the Poisson distribution with expectation $\lambda \in \mathbb{R}^+$ (positive reals) which is denoted $X \sim \mathcal{P}oi(\lambda)$ and has pmf (probability mass function) given by

$$f_P(x|\lambda) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} \tag{1}$$

for $x \in \mathbb{N}$ (non negative integers). When $m \geq 2$, two types of counting distributions can be contructed. The distribution is termed *synchronous* when the counting period starts just after an event [19,25]. The synchroneous counting distribution will be called Goodman distribution and denoted $\mathcal{G}ood(\lambda, m)$ because it is the generalized Poisson distribution of Goodman [26]. The pmf of $Z \sim \mathcal{G}ood(\lambda, m)$ is [26]

$$f_S(z|\lambda, m) = \sum_{t=0}^{m-1} f_P(mz + t|\lambda) \text{ for } z \in \mathbb{N}. \tag{2}$$

The Goodman distribution arises when every mth event are recorded in a Poisson process, starting from zero. The Goodman distribution appears for instance when gaps between points on a line follow a Pearson Type III distribution [27].

When the counting period of the Erlang process begins at an arbitrary time point, the counting distribution is termed *asynchronous* [19,25]. This distribution corresponds to the *Morse*[16]-*Jewell* [17] distribution [18,19]. A Morse-Jewell (MJ) variable Y, which will be denoted $Y \sim \mathcal{MJ}(\lambda, m)$, has pmf [19]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

$$f_A(y|\lambda, m) = \sum_{t=1-m}^{m-1} \frac{m-|t|}{m} f_P(my+t|\lambda) \text{ for } y \in \mathbb{N}. \tag{3}$$

Here, Y is also the record of every mth events in a Poisson process, but the starting point is random. As an example of application field, the MJ distribution is the most appropriate counting distribution when modelling vehicle arrivals with an Erlang process [19]. The special case $m = 2$ of the MJ distribution is the *condensed Poisson* distribution of Chatfield and Goodhardt [22] used, for instance, for inferring the true intake distribution from recall data [28].

2.2. Stochastic Representation of the Morse-Jewell Distribution

Let $\lfloor x \rfloor$ denote the integer part of any $x \in \mathbb{R}$ (reals) and $\mathcal{B}er(p)$ denote a Bernoulli distribution with success probability $p \in [0, 1]$. The following theorem stochastically relates the MJ distribution to Poisson, Goodman and Bernoulli distributions.

Theorem 1 (Stochastic relationship)**.** *Let $X \sim \mathcal{P}oi(\lambda)$ and $m \in \mathbb{N}^+$.*

1. *The count variable $Z = \lfloor \frac{X}{m} \rfloor$ follows a Goodman distribution,* i.e. *$Z \sim \mathcal{G}ood(\lambda, m)$.*
2. *Set $R = X - mZ$ and let $U|R = r \sim \mathcal{B}er(r/m)$. The count variable $Y = Z + U$ has a Morse-Jewell distribution,* i.e. *$Y \sim \mathcal{MJ}(\lambda, m)$.*

The proof of Theorem 1 is given in Appendix A.1. The stochastic relationship provides a method for generating variates from a MJ distribution. It is well known that the expectation of $Y \sim \mathcal{MJ}(\lambda, m)$ is $\mu_A(\lambda, m) = \lambda/m$ [19]. Moreover, on setting $\omega = 2\pi/m$, the variance of Y has the expression [25]

$$\sigma_A^2(\lambda, m) = \frac{\lambda}{m^2} + \frac{m^2-1}{6m^2} + E_m(\lambda), \tag{4}$$

$$E_m(\lambda) = \frac{1}{m^2} \sum_{j=1}^{m-1} \mathrm{e}^{-\lambda(1-\cos(j\omega))} \cos(\lambda \sin(j\omega)) \sum_{k=1}^{m-1} k\left(1 - \frac{k}{m}\right) \cos(jk\omega), \tag{5}$$

with $|E_m(\lambda)| \leq \frac{(m-1)(m^2-1)}{6m^2} \exp\left(-2\lambda \sin^2(\pi/m)/m\right)$ and $E_m(\lambda) \to 0$ as $\lambda \to \infty$. In particular, there is a very simple formula for the variance when $m = 2$ [19]

$$\sigma_A^2(\lambda, 2) = \frac{\lambda}{4} + \frac{1+\mathrm{e}^{-2\lambda}}{8}. \tag{6}$$

It is worthwhile noticing that for $m \geq 2$, Y is underdispersed (variance less than mean) as compared to the Poisson distribution which is equidispersed (variance equal mean). Indeed, on letting $\delta_A(\lambda, m) = \sigma_A^2(\lambda, m)/\mu_A(\lambda, m)$ denote the FID (Fisher's index of dispersion) of Y, *i.e.* the variance-to-mean ratio, we have $\frac{1}{m} < \delta_A(\lambda, m) < 1$ if $m \geq 2$.

3. The Stochastic Condensation Mechanism

The stochastic representation of the MJ distribution suggests a simple route for reducing the variability in an original discrete distribution. Athougth the stochastic representation of the Goodman distribution provides an alternative route, the MJ distribution has a simple mean value expression, a characteristic useful in a regression context [29–31].

Let us consider a family of discrete distributions $\mathcal{OD}(\lambda, \boldsymbol{\theta})$ on the set of integers $\mathbb{Z}$, indexed by the expectation λ and possibly a vector of shape parameters $\boldsymbol{\theta}$. The stochastic condensation mechanism is closely related to the remainder distribution, *i.e.* the distribution of the remainder of a count variable divided by a positive integer, which we first introduce. Let "mod" denote the modulo operator, *i.e.* "$R = X \mod m$" stands for "R is the remainder

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

of the quotient X/m" for $m \in \mathbb{N}^+$. Also set $\mathbb{S}_m = \{0, 1, \cdots, m-1\}$ for $m \in \mathbb{N}^+$ and let "$\stackrel{d}{=}$" stand for "equal in distribution to".

3.1. The Remainder Distribution

Definition 1 (Remainder distribution)**.** *A random variable R is said to follow the remainder distribution $\mathcal{RD}(\lambda, \boldsymbol{\theta}, m)$ generated by the original distribution $\mathcal{OD}(\lambda, \boldsymbol{\theta})$ with a modulus $m \in \mathbb{N}^+$, if $R \stackrel{d}{=} X \bmod m$ where $X \sim \mathcal{OD}(\lambda, \boldsymbol{\theta})$.*

The pmf $f_\rho(\cdot|\lambda, \boldsymbol{\theta}, m)$ of the remainder variable $R \sim \mathcal{RD}(\lambda, \boldsymbol{\theta}, m)$ is given by the series

$$f_\rho(r|\lambda, \boldsymbol{\theta}, m) = \sum_{z=-\infty}^{\infty} f_o(mz + r|\lambda, \boldsymbol{\theta}) \text{ for } r \in \mathbb{S}_m. \tag{7}$$

For $n \in \mathbb{R}^+$, let $\mu_\rho^{(n)}(\lambda, \boldsymbol{\theta}, m) = \sum_{r=1}^{m-1} r^n f_\rho(r|\lambda, \boldsymbol{\theta}, m)$ denote the nth algebraic moment of R. Set $\mu_\rho(\lambda, \boldsymbol{\theta}, m) = \mu_\rho^{(1)}(\lambda, \boldsymbol{\theta}, m)$ and define $\zeta_\rho(\lambda, \boldsymbol{\theta}, m) = \mathrm{E}\{\frac{R}{m}(1 - \frac{R}{m})\}$, given by

$$\zeta_\rho(\lambda, \boldsymbol{\theta}, m) = \frac{\mu_\rho(\lambda, \boldsymbol{\theta}, m)}{m} - \frac{\mu_\rho^{(2)}(\lambda, \boldsymbol{\theta}, m)}{m^2}. \tag{8}$$

The quantity $\zeta_\rho(\lambda, \boldsymbol{\theta}, m)$ will prove useful when deriving the variance of a condensed distribution. Note that $\zeta_\rho(\lambda, \boldsymbol{\theta}, m) \le min\{1/4, \mu_\rho(\lambda, \boldsymbol{\theta}, m)/m, 1 - \mu_\rho(\lambda, \boldsymbol{\theta}, m)/m\}$. Indeed, for $x \in [0, 1]$, the quantity $x(1-x)$, is less than both x and $1-x$, and is maximal at $x = 1/2$.

3.2. Integer-Coefficient Condensation of Discrete Distributions

Definition 2 (Integer-coefficient condensed distribution)**.** *A random variable Y is said to follow the condensed distribution $\mathcal{CD}(\mu, \boldsymbol{\theta}, m)$ generated by the original distribution $\mathcal{OD}(m\mu, \boldsymbol{\theta})$ with a condensation coefficient $m \in \mathbb{N}^+$, if it has the stochastic representation*

$$\begin{aligned} Y|U = u, X = x &= z + u \\ U|X = x &\sim \mathcal{B}er(r/m) \\ X &\sim \mathcal{OD}(m\mu, \boldsymbol{\theta}), \end{aligned} \tag{9}$$

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

By the definition 2, the distributions in the family $\mathcal{OD}(\lambda, \boldsymbol{\theta})$ are recovered in the class $\mathcal{CD}(\mu, \boldsymbol{\theta}, m)$ with $m = 1$. Let $\mathcal{OD}(\lambda, \boldsymbol{\theta})$ has pmf $f_o(\cdot|\lambda, \boldsymbol{\theta})$, cdf (cumulative distribution function) $F_o(\cdot|\lambda, \boldsymbol{\theta})$ and quf (quantile function) $Q_o(\cdot|\lambda, \boldsymbol{\theta})$. Also let $\lceil x \rceil = -\lfloor -x \rfloor$ for $x \in \mathbb{R}$.

Theorem 2 (Distribution function)**.** *The pmf and the cdf of $Y \sim \mathcal{CD}(\mu, \boldsymbol{\theta}, m)$ are respectively given, for $y \in \mathbb{Z}$, by*

$$f_c(y|\mu, \boldsymbol{\theta}, m) = \sum_{t=1-m}^{m-1} \frac{m - |t|}{m} f_o(my + t|m\mu, \boldsymbol{\theta}), \tag{10}$$

$$F_c(y|\mu, \boldsymbol{\theta}, m) = F_o(my - 1|m\mu, \boldsymbol{\theta}) + \sum_{t=0}^{m-1} \frac{m - t}{m} f_o(my + t|m\mu, \boldsymbol{\theta}), \tag{11}$$

and on setting $q_o = Q_o(u|m\mu, \boldsymbol{\theta})$, $y_o = \lceil \frac{q_o}{m} \rceil - 1$ and $u_o = F_c(y_o|\mu, \boldsymbol{\theta}, m)$ for $u \in (0, 1)$, the quf of Y is given by

$$Q_c(u|\mu, \boldsymbol{\theta}, m) = \begin{cases} y_o & \text{if } u_o \ge u \\ y_o + 1 & \text{otherwise} \end{cases}. \tag{12}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

The proof of Theorem 2 is given in Appendix A.2. The general form of the pgf (probability generating function) of a condensed distribution is as follows (see Appendix A.3 for a proof).

Theorem 3 (Probability generating function)**.** *For any $s = \exp(it)$, $i^2 = -1$ (imaginary unit) and $t \in \mathbb{R}$, the pgf of a condensed distribution $\mathcal{CD}(\mu, \boldsymbol{\theta}, m)$ is given by*

$$G_c(s|\mu, \boldsymbol{\theta}, m) = G_q(s|\mu, \boldsymbol{\theta}, m) + \frac{s-1}{m} \sum_{r=1}^{m-1} r \sum_{z=-\infty}^{\infty} s^z f_o(mz + r|m\mu, \boldsymbol{\theta}) \quad (13)$$

$$\text{where} \quad G_q(s|\mu, \boldsymbol{\theta}, m) = \sum_{r=0}^{m-1} \sum_{z=-\infty}^{\infty} s^z f_o(mz + r|m\mu, \boldsymbol{\theta}) \quad (14)$$

is the pgf of the quotient $Z \stackrel{d}{=} (X - R)/m$, $R = X \mod m$, $X \sim \mathcal{OD}(m\mu, \boldsymbol{\theta})$.

For $X \sim \mathcal{OD}(\lambda, \boldsymbol{\theta})$, let $\mu_{or}^{(j,k)}(\lambda, \boldsymbol{\theta}, m)$ denote the (j, k)th joint moment of X and its remainder $R = X \mod m$,

$$\mu_{or}^{(j,k)}(\lambda, \boldsymbol{\theta}, m) = \sum_{z=-\infty}^{\infty} \sum_{r=0}^{m-1} (mz + r)^j r^k f_o(mz + r|\lambda, \boldsymbol{\theta}) \quad (15)$$

and $\mu_q^{(j)}(\lambda, \boldsymbol{\theta}, m)$ denote the jth algebraic moment of the quotient $Z = (X - R)/m$,

$$\mu_q^{(j)}(\lambda, \boldsymbol{\theta}, m) = \sum_{z=-\infty}^{\infty} \sum_{r=0}^{m-1} z^j f_o(mz + r|\lambda, \boldsymbol{\theta}). \quad (16)$$

The general form of integer order algebraic moments of a condensed distribution is as follows (see Appendix A.4 for a routine proof).

Theorem 4 (Moments)**.** *The nth algebraic moment of $Y \sim \mathcal{CD}(\mu, \boldsymbol{\theta}, m)$ is given by*

$$\mu_c^{(n)}(\mu, \boldsymbol{\theta}, m) = \mu_q^{(n)}(\lambda, \boldsymbol{\theta}, m) + \frac{1}{m} \sum_{k=0}^{n-1} \binom{n}{k} \mu_{qr}^{(k,1)}(\lambda, \boldsymbol{\theta}, m) \quad (17)$$

$$\text{where} \quad \mu_{qr}^{(k,l)}(\lambda, \boldsymbol{\theta}, m) = \frac{1}{m^k} \sum_{j=0}^{k} \binom{k}{j} (-1)^j \mu_{or}^{(k-j,l+j)}(\lambda, \boldsymbol{\theta}, m). \quad (18)$$

From the definition 2, $\mathrm{E}[Y] = \mu$, by the law of iterated expectations (see (2) in [32]). Let $\sigma_o^2(\lambda, \boldsymbol{\theta})$ be the variance and $\delta_o(\lambda, \boldsymbol{\theta})$ be the FID of $\mathcal{OD}(\lambda, \boldsymbol{\theta})$. The variance $\sigma_c^2(\mu, \boldsymbol{\theta}, m)$ and the FID $\delta_c(\mu, \boldsymbol{\theta}, m)$ of $Y \sim \mathcal{CD}(\mu, \boldsymbol{\theta}, m)$ satisfy

$$\sigma_c^2(\mu, \boldsymbol{\theta}, m) = \frac{\sigma_o^2(m\mu, \boldsymbol{\theta})}{m^2} + \zeta_\rho(\mu, \boldsymbol{\theta}, m) \quad (19)$$

$$\frac{|\delta_o(m\mu, \boldsymbol{\theta})|}{m} \leq |\delta_c(\mu, \boldsymbol{\theta}, m)| \leq \frac{|\delta_o(m\mu, \boldsymbol{\theta})|}{m} + min\left\{\frac{1}{4|\mu|}, \frac{\mu_\rho(\mu, \boldsymbol{\theta}, m)}{m|\mu|}\right\}. \quad (20)$$

From Equation (20), it appears that for fixed μ, condensation allows the FID $\delta_c(\mu, \boldsymbol{\theta}, m)$ to reach smaller and smaller values as m increases. This suggests the following limiting behavior.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

Conjecture 1 (Limiting condensed distribution)**.** *Let $Y \sim \mathcal{CD}(\mu, \boldsymbol{\theta}, m)$, and set $\mu_o = \lfloor \mu \rfloor$ and $\epsilon = \mu - \mu_o$. Let $U \sim \mathcal{B}er(\epsilon)$ and set $Y_l \stackrel{d}{=} \mu_o + U$, with variance $\sigma_l^2(\mu) = \epsilon(1-\epsilon)$. As $m \to \infty$, Y tends in distribution to Y_l (a shifted Bernoulli distribution), and accordingly, $\sigma_c^2(\mu, \boldsymbol{\theta}, m) \to \sigma_l^2(\mu)$.*

Note that $\sigma_l^2(\mu)$ is the minimal variance of Y [21].

3.3. The Generalized Stochastic Condensation Mechanism

For a fixed mean value μ, the variance $\sigma_c^2(\mu, \boldsymbol{\theta}, m)$ and the FID $\delta_c(\mu, \boldsymbol{\theta}, m)$ of a condensed distribution take discrete values in terms of m. This may make condensed distributions difficult to work with, especially when one desires to estimate m along with other distribution parameters μ and $\boldsymbol{\theta}$, or in a regression context. There exists a generalization of the MJ distribution (recall this results from the condensation of the Poisson distribution) by the mean of the incomplete gamma function [19]. However, this generalization does not enjoy the simplicity of the stochastic representation in Theorem 1 and the expression of the variance in Equation (4). The following generalization to any $m \in [1, \infty)$ enjoys the simple defining stochastic representation and, if the variance of the integer-coefficient condensation has a simple expression (as for the MJ distribution), so does the generalized condensed distribution.

Definition 3 (Generalized condensed distribution)**.** *A random variable Y is said to follow the generalized condensed distribution $\mathcal{GCD}(\mu, \boldsymbol{\theta}, m)$ generated by $\mathcal{OD}(m\mu, \boldsymbol{\theta})$, with a condensation coefficient $m \in [1, \infty)$, if it has the stochastic representation*

$$\begin{aligned} Y|U = u, X = x, W = w &= z + u \\ U|X = x, W = w &\sim \mathcal{B}er(r/(m_o + w)) \\ X|W = w &\sim \mathcal{OD}((m_o + w)\mu, \boldsymbol{\theta}) \\ W &\sim \mathcal{B}er(p). \end{aligned} \tag{21}$$

where $p = (m - m_o)(m_o + 1)/m$, $m_o = \lfloor m \rfloor$, $z = \lfloor x/(m_o + w) \rfloor$ and $r = x - (m_o + w)z$.

By the definition 3, a generalized condensed distribution with condensation coefficient m is the mixture of the condensed distributions with condensation coefficients m_o and $m_o + 1$ and weights $1 - p$ and p respectively. It thus has, for $y \in \mathbb{Z}$, the pmf

$$f_g(y|\mu, \boldsymbol{\theta}, m) = (1-p) f_c(y|\mu, \boldsymbol{\theta}, m_o) + p f_c(y|\mu, \boldsymbol{\theta}, m_o + 1). \tag{22}$$

Likely, the algebraic moments of $\mathcal{GCD}(\mu, \boldsymbol{\theta}, m)$ are given by

$$\mu_g^{(n)}(\mu, \boldsymbol{\theta}, m) = (1-p)\mu_c^{(n)}(\mu, \boldsymbol{\theta}, m_o) + p\mu_c^{(n)}(\mu, \boldsymbol{\theta}, m_o + 1). \tag{23}$$

For $m \in \mathbb{N}^+$, $p = 0$ so that $\mathcal{GCD}(\mu, \boldsymbol{\theta}, m)$ is reduced to $\mathcal{CD}(\mu, \boldsymbol{\theta}, m)$. Note that the two mixed distributions have the same expectation μ so that $\mathrm{E}[Y] = \mu$ and Y has variance

$$\sigma_g^2(\mu, \boldsymbol{\theta}, m) = (1-p)\sigma_c^2(\mu, \boldsymbol{\theta}, m_o) + p\sigma_c^2(\mu, \boldsymbol{\theta}, m_o + 1). \tag{24}$$

The expression of p was chosen to yield, when the variance of $\mathcal{OD}(\lambda, \boldsymbol{\theta})$ has the form $\sigma_o^2(\lambda, \boldsymbol{\theta}) = a\lambda + b\lambda^2$ for $(a, b) \in [0, \infty)^2$ a function of $\boldsymbol{\theta}$, the simpler variance expression

$$\sigma_g^2(\mu, \boldsymbol{\theta}, m) = \frac{a\mu}{m} + b\mu^2 + (1-p)\zeta_\rho(\mu, \boldsymbol{\theta}, m_o) + p\zeta_\rho(\mu, \boldsymbol{\theta}, m_o + 1). \tag{25}$$

This holds, for instance, for the Poisson ($a = 1$, $b = 0$), negative binomial ($a = 1$, $b \in \mathbb{R}^+$) [7] and Consul's generalized Poisson ($a \in \mathbb{R}^+$, $b = 0$) [33] distributions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

3.4. Conditional Distribution of the Original Count Variable

The condensation mechanism stochastically distorts a non-observable count variable X to obtain the observable count variable Y. Interestingly, the stochastic representation Equation (21) can be used to provide insights on the original variable X. Indeed, the conditional distribution and in particular the conditional mean of the original distribution, given an observed value y of Y, may be useful for predicting X. Moreover, although this work does not target the estimation of condensed count distributions, the expectation-maximization algorithm [34] is a useful tool for maximum likelihood inference in latent class-like models, exploiting the conditional distribution of latent variables given the observed values of observables quantities. The following theorem gives some basic results useful for these purposes (see the proof in Appendix A.5).

Theorem 5 (Conditional distributions)**.** *Let X, W and Y be defined as in Equation (21). Then, for $y \in \mathbb{Z}$ with probability mass $f_g(y|\mu, \boldsymbol{\theta}, m) > 0$:*

$$f_{X|W,Y}(x|W = w, Y = y, \mu, \boldsymbol{\theta}, m) = \frac{f_o(x|(m_o + w)\mu, \boldsymbol{\theta})}{f_c(y|\mu, \boldsymbol{\theta}, m_o + w)} \eta(x, y, m_o + w) \quad (26)$$

where

$$\eta(x, y, m) = \left(1 + y - \frac{x}{m}\right) I_{\mathbb{S}_m}(x - my) + \left(1 - y + \frac{x}{m}\right) I_{\mathbb{S}^*_m}(my - x), \quad (27)$$

$\mathbb{S}_m = \{0, \cdots, m-1\}$ for $m \in \mathbb{N}^+$, $\mathbb{S}^*_m = \{1, \cdots, m-1\}$ if $m \geq 2$ and $\mathbb{S}^*_m = \{\}$ otherwise, $I_{\mathbb{A}}(x)$ is the indicator function, which equals one if $x \in \mathbb{A}$ and zero otherwise. Moreover,

$$f_{W|Y}(w|Y = y, \mu, \boldsymbol{\theta}, m) = p_y^w [1 - p_y]^{1-w} \text{ for } w \in \{0, 1\}, \quad (28)$$

$$f_{X|Y}(x|Y = y, \mu, \boldsymbol{\theta}, m) = (1 - p) \frac{f_o(x|m_o\mu, \boldsymbol{\theta})}{f_c(y|\mu, \boldsymbol{\theta}, m_o)} \eta(x, y, m_o) + p \frac{f_o(x|(m_o + 1)\mu, \boldsymbol{\theta})}{f_c(y|\mu, \boldsymbol{\theta}, m_o + 1)} \eta(x, y, m_o + 1) \quad (29)$$

where $p_y = [(1 - p) f_c(y|\mu, \boldsymbol{\theta}, m_o) + p f_c(y|\mu, \boldsymbol{\theta}, m_o + 1)]^{-1} p f_c(y|\mu, \boldsymbol{\theta}, m_o + 1)$ is the success probability of the Bernoulli variable W given that $Y = y$, $p = (m - m_o)(m_o + 1)/m$, $m_o = \lfloor m \rfloor$. Furthermore, for $n \in \mathbb{N}$, the nth moment of X conditional on $Y = y$ is

$$E_{X|Y}[X^n|Y = y, \mu, \boldsymbol{\theta}, m] = (1 - p) \frac{E_X(n, y|\mu, \boldsymbol{\theta}, m_o)}{f_c(y|\mu, \boldsymbol{\theta}, m_o)} + p \frac{E_X(n, y|\mu, \boldsymbol{\theta}, m_o + 1)}{f_c(y|\mu, \boldsymbol{\theta}, m_o + 1)} \quad (30)$$

where, for $m \in \mathbb{N}^+$, $E_X(n, y|\mu, \boldsymbol{\theta}, m)$ denotes the nth partial moment of X for $X \in \mathbb{S}_{m,y}$, $\mathbb{S}_{m,y} = \{my + (1 - m), \cdots, my, \cdots, my + (m - 1)\}$:

$$E_X(n, y|\mu, \boldsymbol{\theta}, m) = \sum_{t=1-m}^{m-1} (my + t)^n \frac{m - |t|}{m} f_o(my + t|m\mu, \boldsymbol{\theta}). \quad (31)$$

Remark 1. *Given the original variable X and the Bernoulli variable W, the distribution of the distorted variable Y does not depend on the parameters μ and $\boldsymbol{\theta}$:*

$$f_{Y|W,X}(y|W = w, X = x, \mu, \boldsymbol{\theta}, m) = \eta(x, y, m_o + w). \quad (32)$$

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

$$f_{W,X,Y}(w,x,y) = p^w(1-p)^{1-w} f_o(x|(m_o+w)\mu, \boldsymbol{\theta})\eta(x,y,m_o+w) \tag{33}$$

of W, X and Y is reduced, for a fixed m, to the maximization of the conditional likelihood of X given W:

$$f_{X|W}(x|W=w,\mu,\boldsymbol{\theta},m) = f_o((m_o+w)y+t|(m_o+w)\mu,\boldsymbol{\theta}). \tag{34}$$

The expectation-conditional-maximization algorithm [35] allows breaking the maximization step of the expectation-maximization algorithm in many but computationally simpler conditional-maximization sub-steps, while retaining the appealing properties of the basic expectation-maximization algorithm. Each conditional-maximization sub-step updates only a subset of the model parameters conditional on the remaining parameters. It stems from Equation (34) that if the maximum likelihood estimation of the original distribution parameters μ and $\boldsymbol{\theta}$ is easy, then the expectation-conditional-maximization algorithm will be appropriate for fitting the corresponding condensed distribution.

4. Applications

We illustrate the use of the condensation mechanism on two popular count models. Apart from the equidispersed Poisson model, another appealing candidate for the application of the condensation method is the negative binomial distribution, which is the most popular overdispersion model. In fact, there already exists a *condensed negative binomial* distribution [22,36–38] which, as for the *condensed Poisson* distribution [22], corresponds to the condensation of the negative binomial distribution with condensation coefficient $m = 2$. This special condensed negative binomial distribution has been used in consumer-purchasing models [22,36,37] and alcohol consumption modelling [38]. We obtain here the general form of the distributions with $m \in [1, \infty)$ by applying condensation.

4.1. The Generalized Condensed Poisson Distribution

Definition 4 (Generalized condensed Poisson distribution)**.** *A random variable Y is said to follow the generalized condensed Poisson distribution $\mathcal{GCP}(\mu, m)$ with $\mu \in \mathbb{R}^+$ and $m \in [1, \infty)$, if it is generated by the generalized condensation mechanism (21) with the Poisson distribution $\mathcal{P}oi(m\mu)$ as generator, and accordingly has, for $y \in \mathbb{N}$, the pmf*

$$f_Y(y|\mu,m) = (1-p)f_A(y|m_o\mu, m_o) + pf_A(y|(m_o+1)\mu, m_o+1) \tag{35}$$

where $p = (m-m_o)(m_o+1)/m$, $m_o = \lfloor m \rfloor$ and $f_A(\cdot|\lambda, m)$ is given in Equation (3).

The generalized condensed Poisson (GCP) distribution has the Morse-Jewell (MJ) distribution introduced in section 2 as a special case, when the condensation coefficient m is a positive integer. The condensed Poisson distribution of [22] is the special case when $m = 2$. The general form of the moments of the MJ distribution has for long evaded discovery, apart from the first two moments [25]. We obtain these moments using the representation as a condensed distribution.

4.1.1. The Poisson Remainder Distribution

As defined in section 3, condensation is closely related to the remainder distribution of the original distribution to be condensed. We introduce in this section the Poisson remainder (PR) distribution, *i.e.* for $m \in \mathbb{N}^+$, the distribution of $R = X \mod m$ when $X \sim \mathcal{P}oi(\lambda)$. The probability that X is even is given by $P(R=0) = \left(1+\mathrm{e}^{-2\lambda}\right)/2$, in the instance $m = 2$. This special case is well known [39] and has attention in informal discussions on probability theory, see *e.g.* https://math.stackexchange.com/questions/2007238/what-is-the-probability-of-getting-an-even-number-from-a-poisson-random-draw and https://stefanengineering.com/2020/05/15/probability-of-even-generating-functions (accessed on

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

March 09, 2021). A generalization of this result (*i.e.* $P(R=0)$) for any integer m has been given by Serfling [25] who indicated that this was already well known before the 1970s. As a useful device for obtaining many distributional properties of the generalized condensed Poisson distribution, we derive here $P(R=r)$ for any r in the sample space $\mathbb{S}_m = \{0, 1, \cdots, m-1\}$ of R.

Definition 5 (Poisson remainder distribution)**.** *A random variable R is said to follow the Poisson remainder distribution $\mathcal{PR}(\lambda, m)$ with $\lambda \in \mathbb{R}^+$ and $m \in \mathbb{N}^+$, if $R \overset{d}{=} X \bmod m$ where $X \sim \mathcal{P}oi(\lambda)$.*

Let us consider the series defined for $z > 0$, $m \in \mathbb{N}^+$ and $t \in \mathbb{Z}$ by

$$f_{m,t}(z) = \sum_{j=0}^{\infty} \frac{z^{mj+t}}{(mj+t)!} \tag{36}$$

with the convention $z^{mj+t}/(mj+t)! = 0$ if $mj+t < 0$. The function $f_{m,t}$ and the following related lemma will prove to be central to the study of both the PR and the GCP distributions.

Lemma 1. *Set $f_{m,t}^{(k)}(z) = \partial^k f_{m,t}(z)/\partial z^k$. Let $r \in \{0, 1, \cdots, m-1\}$, and for $t \in \mathbb{Z}$, let $v = \lfloor t/m \rfloor$ and $s = t - mv$. Then, on setting $\omega = 2\pi/m$,*

$$f_{m,t}(z) = f_{m,s}(z) - \sum_{j=0}^{v-1} \frac{z^{mj+s}}{(mj+s)!} \quad if \quad v \geq 1 \tag{37}$$

$$f_{m,t}(z) = f_{m,s}(z) \quad if \quad v \leq 0 \tag{38}$$

$$f_{m,r}^{(k)}(z) = f_{m,r-k}(z) \tag{39}$$

$$f_{m,r}(z) = \frac{1}{m} \sum_{j=0}^{m-1} \mathrm{e}^{z\cos(j\omega)} \cos(z \sin(j\omega) - jr\omega). \tag{40}$$

The proof of Lemma 1 is given in Appendix A.6. Equation (40) generalizes *Lemma 3.2* in [25]. Using Equation (40), the pmf and the cdf of a PR variable are as follows (see a routine proof in Appendix A.7).

Theorem 6 (Poisson remainder distribution function)**.** *Let $R \sim \mathcal{PR}(\lambda, m)$. Then, the pmf and the cdf of R are respectively given, for $r \in \{0, 1, \cdots, m-1\}$, by*

$$f_R(r|\lambda, m) = \frac{1}{m} \sum_{k=0}^{m-1} \mathrm{e}^{\lambda(\cos(k\omega)-1)} \cos(\lambda \sin(k\omega) - kr\omega) \tag{41}$$

$$F_R(r|\lambda, m) = \frac{1}{m} \left\{ r + 1 + \sum_{k=1}^{m-1} \sin\left(k\omega \frac{r+1}{2}\right) \frac{\cos(\lambda \sin(k\omega) + k\omega r/2)}{\sin(k\omega/2)\mathrm{e}^{\lambda(1-\cos(k\omega))}} \right\}. \tag{42}$$

The nth moment of $\mathcal{PR}(\lambda, m)$ is denoted $\mu_R(r|\lambda, m)$ and given by

$$\mu_R^{(n)}(\lambda, m) = \sum_{r=0}^{m-1} r^n f_R(r|\lambda, m). \tag{43}$$

4.1.2. Probability Generating Function and Moments

On using Equations (13) and (37), the probability generating function (pgf) of a GCP variable is obtained as follows.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

Corollary 1 (Probability generating function of a generalized condensed Poisson variable)**.** *Let $Y \sim \mathcal{GCP}(\lambda, m)$. The pgf of Y is given for any $s = \exp(it)$, $i^2 = -1$, $t \in \mathbb{R}$, by*

$$G_Y(s|\mu, m) = (1-p)G_A(s|m\mu, m_o) + pG_A(s|m\mu, m_o+1) \quad (44)$$

where

$$G_A(s|\lambda, m) = G_S(s|\lambda, m) + \frac{s-1}{m}\mathrm{e}^{-\lambda}\sum_{r=1}^{m-1} rs^{-r/m} f_{m,r}(\lambda s^{1/m}) \quad (45)$$

is the pgf of a Morse-Jewell distribution $\mathcal{MJ}(\lambda, m)$ which has the pmf (3) and

$$G_S(s|\lambda, m) = \mathrm{e}^{-\lambda}\sum_{r=0}^{m-1} s^{-r/m} f_{m,r}(\lambda s^{1/m}) \quad (46)$$

is the pgf of a Goodman distribution $\mathcal{G}ood(\lambda, m)$ which has the pmf (2).

Equation (46) is equivalent to formula (3) in [27]. The expression (46) is simpler in the sens it uses the more common exponential, sinus and cosinus functions (through Equation (40)), instead of the raw infinite sum in [27]. By the definition 4, $\mathcal{GCP}(\mu, m)$ has expectation μ. The algebraic moments of a GCP distribution are as follows (see the proof in Appendix A.8).

Theorem 7 (Moments of a generalized condensed Poisson variable)**.** *Let $Y \sim \mathcal{GCP}(\mu, m)$. The nth moment of Y is given, for $n \in \mathbb{N}^+$, by*

$$\mu_Y^{(n)}(\mu, m) = (1-p)\mu_A^{(n)}(m_o\mu, m_o) + p\mu_A^{(n)}((m_o+1)\mu, m_o+1) \quad (47)$$

where $\mu_A^{(n)}(\lambda, m)$ is the nth moment of the Morse-Jewell distribution $\mathcal{MJ}(\lambda, m)$

$$\mu_A^{(n)}(\lambda, m) = \mu_S^{(n)}(\lambda, m) + \frac{1}{m}\sum_{k=0}^{n-1}\binom{n}{k}\frac{1}{m^k}\sum_{l=0}^{k}\binom{k}{l}(-1)^l \mu_{XR}^{(k-l,l+1)}(\lambda, m) \quad (48)$$

where $\mu_S^{(n)}(\lambda, m)$ is the nth moment of the Goodman distribution $\mathcal{G}ood(\lambda, m)$

$$\mu_S^{(n)}(\lambda, m) = \frac{(-1)^n}{m^n}\left\{\mu_{\mathrm{R}}^{(n)}(\lambda, m) + \sum_{k=1}^{n}\binom{n}{k}(-1)^k \mu_{XR}^{(k,n-k)}(\lambda, m)\right\}, \quad (49)$$

$\mu_{XR}^{(k,l)}(\lambda, m)$ is the (k,l)th joint moment of $X \sim \mathcal{P}oi(\lambda)$ and $R = X \mod m$

$$\mu_{XR}^{(k,l)}(\lambda, m) = \sum_{j=1}^{k} S(k,j)\lambda^j \mu_{RC}^{(l,j)}(\lambda, m), \quad (50)$$

$S(k,j)$ is the Stirling number of the second kind [40,41]

$$S(k,j) = \frac{1}{j!}\sum_{l=0}^{j}(-1)^l\binom{j}{l}(j-l)^k \quad (51)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

and $\mu_{RC}^{(n,j)}(\lambda, m)$ *are probability cycled (PC) moments of a PR variable* $\mathcal{PR}(\lambda, m)$*, defined as*

$$\mu_{RC}^{(n,j)}(\lambda, m) = \mathrm{e}^{-\lambda} \sum_{r=0}^{m-1} r^n f_{m,r-j}(\lambda), \tag{52}$$

and given, for any $v \in \mathbb{N}$ *and* $s \in \{1, 2, \cdots, m-1\}$*, by*

$$\mu_{RC}^{(0,v)}(\lambda, m) = 1 \tag{53}$$

$$\mu_{RC}^{(n,mv)}(\lambda, m) = \mu_R^{(n)}(\lambda, m) \tag{54}$$

$$\mu_{RC}^{(n,mv+s)}(\lambda, m) = \sum_{r=0}^{m-s-1} (r+s)^n f_R(r|\lambda, m) + \sum_{r=m-s}^{m-1} (r+s-m)^n f_R(r|\lambda, m). \tag{55}$$

Note that, as expected, Equation (47) is reduced to the nth moment of a Poisson distribution $\mathcal{P}oi(\mu)$ when $m = 1$ [42]:

$$\mu_Y^{(n)}(\mu, 1) = \sum_{j=1}^{n} S(n, j)\mu^j. \tag{56}$$

On setting $\zeta_R(\lambda, m) = \mu_R^{(1)}(\lambda, m)/m - \mu_R^{(2)}(\lambda, m)/m^2$, the variance of $Y \sim \mathcal{GCP}(\mu, m)$ reads

$$\sigma_Y^2(\mu, m) = \frac{\mu}{m} + (1-p)\zeta_R(m_o\mu, m_o) + p\zeta_R((m_o+1)\mu, m_o+1). \tag{57}$$

and the FID of Y is given by

$$\delta_Y(\mu, m) = \frac{1}{m} + \frac{1}{\mu}[(1-p)\zeta_R(m_o\mu, m_o) + p\zeta_R((m_o+1)\mu, m_o+1)]. \tag{58}$$

4.1.3. Illustrations

In order to point out the potential of the GCP distribution to accomodate equidispersion and underdispersion situations, we display the pmf (probability mass function) and the FID (Fisher's index of dispersion) of the GCP distribution. The pmf and FID values were computed in the R freeware [43] using Equations (35) and (58) respectively.

Figures 1 and 2 show the pmf of the GCP distribution for selected mean (μ) and condensation coefficient (m) values. It can be observed that the probability masses of quantiles around the mean value increase with m, whereas the probability masses of quantiles far from the mean value decrease with m. In other words, the spread of the distribution decreases with m.

Figure 3 shows the FID of the GCP distribution for mean values μ in the range $(0, 20]$ and selected condensation coefficients m in the range $[1, 500]$. It appears that the FID is close to $1/m$ for large mean values ($\mu > 5$) and not too large condensation coefficient ($m \leq 10$). For very large condensation coefficients ($m > 10$), we observe, in addition to severe underdispersion (FID < 0.2 for $\mu > 1$), an oscillating index of dispersion with an amplitude approaching zero as m increases. The observed limiting behavior is in accordance with the Conjecture 1, which implies maxima of the limiting variance at half integer mean values and minima at integer mean values.

4.2. The Generalized Condensed Negative Binomial Distribution

By allowing underdispersion, the generalized condensed Poisson distribution is more flexible than the basic Poisson distribution. However, a phenomenon displaying very low variability at micro-scale usually becomes more versatile at the aggregate level. For instance, when the number of purchases in a fixed time interval is Poisson distributed, but the

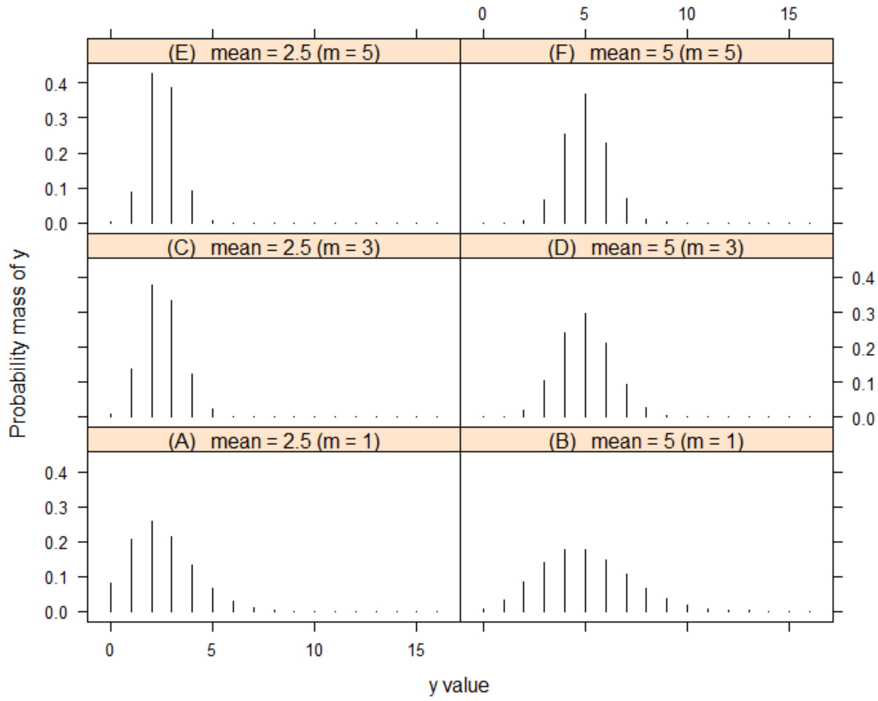

Figure 1. Probability mass of the generalized condensed Poisson distribution for mean values of 2.5 (left panel) and 5 (right panel) and condensation coefficients $m = 1, 3$ and 5.

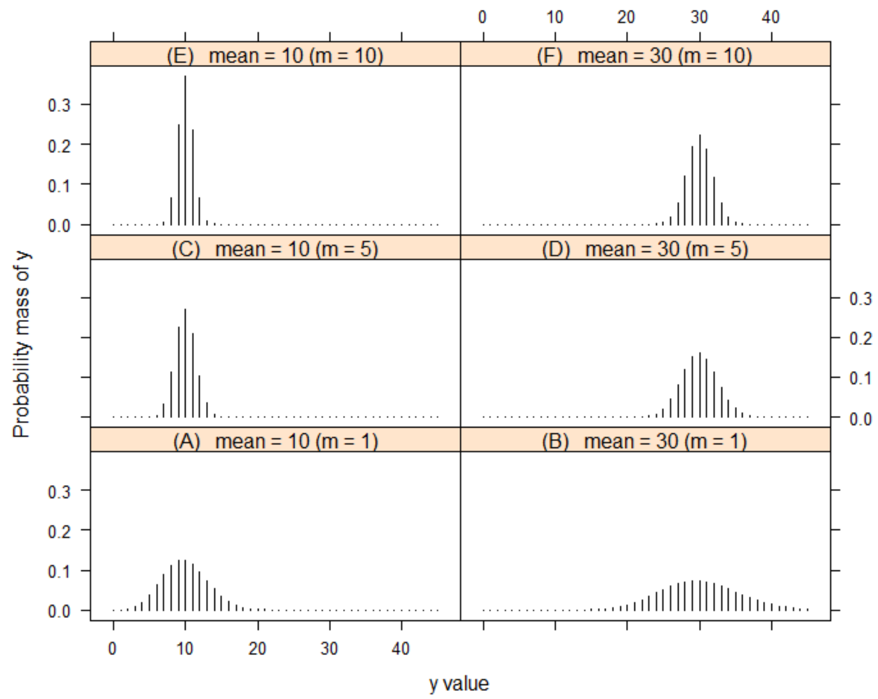

Figure 2. Probability mass of the generalized condensed Poisson distribution for mean values of 10 and 30 and condensation coefficients $m = 1, 5$ and 10.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

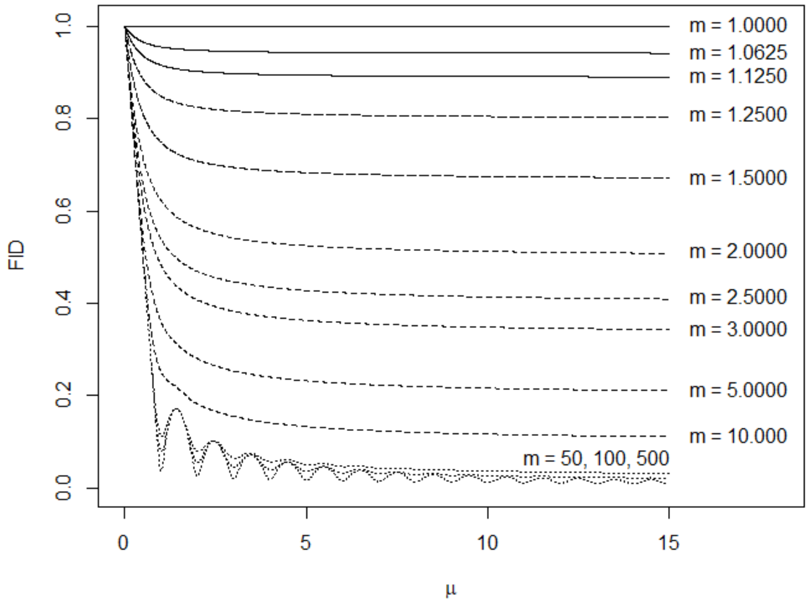

Figure 3. Fisher's index of dispersion (FID), *i.e.* variance-to-mean ratio, of the generalized condensed Poisson distribution for mean values μ in the range $[0, 20]$ and selected condensation coefficients m in the range $[1, 500]$.

purchasing rate is gamma distributed for individuals in a target population, the distribution of purchase events for persons randomly chosen from the population is negative binomial [44]. If the individual level distribution is condensed Poisson, then the distribution is condensed negative binomial at the aggregate level [37]. We develop in this section the generalized condensed negative binomial distribution which allows full dispersion flexibility [21]. Let $X \sim \mathcal{NB}(\lambda, \phi)$, *i.e.* X follows a negative binomial distribution with expectation $\lambda \in \mathbb{R}^+$ and dispersion parameter $\phi \in \mathbb{R}^+$. For $y \in \mathbb{N}$, the pmf of X is given by

$$f_N(x|\lambda, \phi) = \frac{\Gamma(\phi + x)}{\Gamma(\phi)\Gamma(x+1)} \left(\frac{\phi}{\lambda + \phi}\right)^{\phi} \left(\frac{\lambda}{\lambda + \phi}\right)^{x}. \tag{59}$$

A negative binomial variable $X \sim \mathcal{NB}(\lambda, \phi)$ has mean λ and variance $\lambda + \lambda^2/\phi$.

Definition 6 (Generalized condensed negative binomial distribution)**.** *A random variable Y is said to follow the generalized condensed negative binomial distribution $\mathcal{GCNB}(\mu, \phi, m)$ with $\mu \in \mathbb{R}^+$, $\phi \in \mathbb{R}^+$ and $m \in [1, \infty)$, if it is generated by the generalized condensation mechanism* (21) *with the negative binomial distribution $\mathcal{NB}(m\mu, \phi)$ as generator, and accordingly has, for $y \in \mathbb{N}$, the pmf*

$$f_Y(y|\mu, m) = (1-p) f_{CN}(y|m_o\mu, \phi, m_o) + p f_{CN}(y|(m_o+1)\mu, \phi, m_o+1) \tag{60}$$

where $p = (m - m_o)(m_o + 1)/m$, $m_o = \lfloor m \rfloor$ and

$$f_{CN}(y|\mu, \phi, m) = \sum_{t=1-m}^{m-1} \frac{m - |t|}{m} f_N(my + t|m\mu, \phi). \tag{61}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

The expectation of $X \sim \mathcal{GCNB}(\mu, \phi, m)$ is μ. The moment generating function [45] and the moments [37] of the GCNB distribution have been derived for the special case $m = 2$. In order to obtain the moments of a GCNB variable for arbitrary m values, we first derive the distribution of the remainder of the negative binomial distribution. Interestingly, the negative binomial remainder distribution can be obtained from the Poisson remainder distribution, using the Poisson-gamma mixture property of the negative binomial distribution [7].

4.2.1. The Negative Binomial Remainder Distribution

Definition 7 (Negative binomial remainder distribution)**.** *A random variable R is said to follow the negative binomial remainder distribution $\mathcal{NR}(\lambda, \phi, m)$ with $\lambda \in \mathbb{R}^+$, $\phi \in \mathbb{R}^+$ and $m \in \mathbb{N}^+$, if $R \overset{d}{=} X \bmod m$ where $X \sim \mathcal{NB}(\lambda, \phi)$.*

By using the representation of the negative binomial distribution as the Poisson-gamma mixture [7], the distribution function of the negative binomial remainder distribution is obtained as follows (see the proof in Appendix A.9).

Theorem 8 (Negative binomial remainder distribution function)**.** *Let $R \sim \mathcal{NR}(\lambda, \phi, m)$. Then, the pmf and the cdf of R are respectively given, for $r \in \{0, 1, \cdots, m-1\}$, by*

$$f_{\bar{R}}(r|\lambda, \phi, m) = \frac{1}{m}\sum_{k=0}^{m-1} \alpha_k^{\phi}(\lambda)\cos(\phi\beta_k(\lambda) - kr\omega) \quad (62)$$

$$F_{\bar{R}}(r|\lambda, \phi, m) = \frac{1}{m}\left\{ r + 1 + \sum_{k=1}^{m-1} \alpha_k^{\phi}(\lambda)\sin\left(k\omega\frac{r+1}{2}\right)\frac{\cos(\phi\beta_k(\lambda) + k\omega r/2)}{\sin(k\omega/2)} \right\} \quad (63)$$

where $\omega = 2\pi/m$, $\alpha_k(z) = \frac{\phi}{\sqrt{(\lambda+\phi)^2 - 2z(\lambda+\phi)\cos(k\omega) + z^2}}$ and $\beta_k(z) = \arctan\left(\frac{z\sin(k\omega)}{\phi+\lambda-z\cos(k\omega)}\right)$.

From Equation (62), the nth moment of $\mathcal{NR}(\lambda, \phi, m)$ can be obtained as

$$\mu_{\bar{R}}^{(n)}(\lambda, \phi, m) = \sum_{r=0}^{m-1} r^n f_{\bar{R}}(r|\lambda, \phi, m) \quad (64)$$

and the quantity (8) is given by $\zeta_{\bar{R}}(\lambda, \phi, m) = \mu_{\bar{R}}^{(1)}(\lambda, \phi, m)/m - \mu_{\bar{R}}^{(2)}(\lambda, \phi, m)/m^2$.

4.2.2. Probability Generating Function and Moments

The probability generating function and the moments of the generalized condensed negative binomial distribution are as follows.

Theorem 9 (Probability generating function of a condensed negative binomial variable)**.** *Let $Y \sim \mathcal{GCNB}(\lambda, m)$. The pgf of Y is given for any $s = \exp(it)$, $i^2 = -1$, $t \in \mathbb{R}$, by*

$$G_Y(s|\mu, \phi, m) = (1-p)G_{N_c}(s|\mu, \phi, m_o) + pG_{N_c}(s|\mu, \phi, m_o + 1) \quad (65)$$

where

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

$$
\begin{aligned}
G_{N_c}(s|\mu,\phi,m_o) &= G_{N_s}(s|\mu,\phi,m) + G_{N_u}(s|\mu,\phi,m) && (66)\\
G_{N_s}(s|\mu,\phi,m) &= \frac{1}{m}\sum_{r=0}^{m-1} s^{-r/m}\sum_{k=0}^{m-1}\alpha_k^{\phi}\left(\lambda s^{1/m}\right)\cos\left(\phi\beta_k(\lambda s^{1/m}) - kr\omega\right) && (67)\\
G_{N_u}(s|\mu,\phi,m) &= \frac{s-1}{m^2}\sum_{r=1}^{m-1} rs^{-r/m}\sum_{k=0}^{m-1}\alpha_k^{\phi}\left(\lambda s^{1/m}\right)\cos\left(\phi\beta_k(\lambda s^{1/m}) - kr\omega\right). && (68)
\end{aligned}
$$

Theorem 10 (Moments of a condensed negative binomial variable)**.** *Let us consider a count variable $Y \sim \mathcal{GCNB}(\mu,\phi,m)$. The nth moment of Y is given, for $n \in \mathbb{N}^+$, by*

$$
\mu_Y^{(n)}(\mu,\phi,m) = (1-p)\mu_{CN}^{(n)}(m_o\mu,\phi,m_o) + p\mu_{CN}^{(n)}((m_o+1)\mu,\phi,m_o+1) \quad (69)
$$

where $\mu_{CN}^{(n)}(\lambda,\phi,m)$ is the nth moment of a condensed negative binomial distribution when $m \in \mathbb{N}^+$

$$
\mu_{CN}^{(n)}(\lambda,\phi,m) = \mu_Z^{(n)}(\lambda,\phi,m) + \frac{1}{m}\sum_{k=0}^{n-1}\binom{n}{k}\frac{1}{m^k}\sum_{l=0}^{k}\binom{k}{l}(-1)^l\mu_{X\bar{R}}^{(k-l,l+1)}(\lambda,m), \quad (70)
$$

$\mu_Z^{(n)}(\lambda,\phi,m)$ is the nth moment of $Z = \lfloor \frac{X}{m} \rfloor$, $X \sim \mathcal{NB}(\lambda,\phi)$:

$$
\mu_Z^{(n)}(\lambda,\phi,m) = \frac{(-1)^n}{m^n}\left\{\mu_{\bar{\mathrm{R}}}^{(n)}(\lambda,\phi,m) + \sum_{k=1}^{n}\binom{n}{k}(-1)^k\mu_{X\bar{R}}^{(k,n-k)}(\lambda,\phi,m)\right\}, \quad (71)
$$

$\mu_{X\bar{R}}^{(k,l)}(\lambda,\phi,m)$ is the (k,l)th joint moment of $X \sim \mathcal{NB}(\lambda,\phi)$ and $R = X \mod m$

$$
\mu_{X\bar{R}}^{(k,l)}(\lambda,\phi,m) = \sum_{j=1}^{k} S(k,j)\frac{\lambda^j\Gamma(\phi+j)}{\phi^j\Gamma(\phi)}\sum_{r=0}^{m-1} r^l\nu_j(r), \quad (72)
$$

$\nu_j(r) = \frac{1}{m}\sum_{k=0}^{m-1}\alpha_k^{\phi+j}(\lambda)\cos((\phi+j)\beta_k(\lambda) - k\omega s)$, $s = s_o$ if $s_o \geq 0$ and $s = m + s_o$ if $s_o < 0$ with $s_o = (r-j) \mod m$.

Theorems 9 and 10 are straightforwardly obtained starting from Theorems 1 and 7 respectively, and using the Poisson-gamma mixture property of the negative binomial distribution (see *e.g.* the proof of Theorem 8 in Appendix A.9). Their proofs are thus omitted. The variance of $Y \sim \mathcal{GCNB}(\mu,\phi,m)$ is given by

$$
\sigma_Y^2(\mu,\phi,m) = \frac{\mu}{m} + \frac{\mu^2}{\phi} + (1-p)\zeta_{\bar{R}}(m_o\mu,\phi,m_o) + p\zeta_{\bar{R}}((m_o+1)\mu,\phi,m_o+1). \quad (73)
$$

and the FID of Y is given by

$$
\delta_Y(\mu,\phi,m) = \frac{1}{m} + \frac{\mu}{\phi} + \frac{1}{\mu}[(1-p)\zeta_{\bar{R}}(m_o\mu,\phi,m_o) + p\zeta_{\bar{R}}((m_o+1)\mu,\phi,m_o+1)]. \quad (74)
$$

4.2.3. Illustrations

To illustrate the full dispersion flexibility property of the generalized condensed negative binomial distribution, we display its pmf (probability mass function) and FID (Fisher's

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

index of dispersion). The pmf and FID values were computed in R using Equations (60) and (74) respectively.

Figures 4 and 5 show the pmf of the GCNB distribution for $\phi = 0.25$ and selected mean (μ) and condensation coefficient (m) values. It appears that the probability masses are more and more attracted and condensed around the mean value, for increasing m value. In addition, condensation reduces the skewness of the distorted distribution ($m > 1$) as compared to the original distribution ($m = 1$).

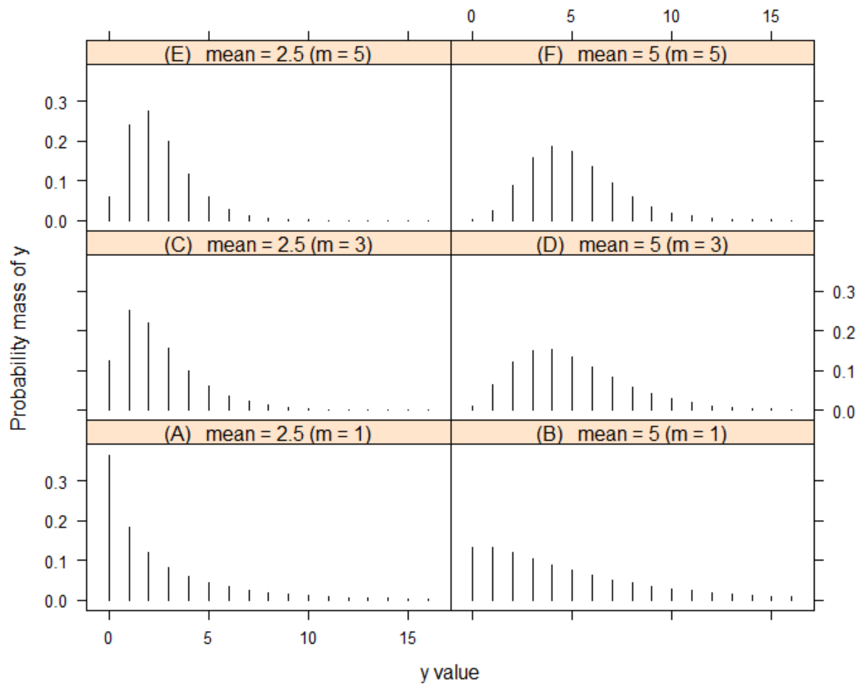

Figure 4. Probability mass of the generalized condensed negative binomial distribution with $\phi = 0.25$ for mean values of 2.5 and 5 and condensation coefficients $m = 1, 3$ and 5.

It appears from Equation (74) that the FID essentially decreases with both m and ϕ. In order to avoid redundancy, the distribution can be reparametrized in terms of μ, m and $\tau \in \mathbb{R}$, by setting $\phi = (m+1)^{\tau}$. Under this parameterization, and with $\mu > 1$, the distribution is overdispersed for any $\tau \leq 0$. However, when $\tau > 0$, the distribution allows full dispersion flexibility. Figure 6 shows the FID of the GCNB distribution for mean values μ in the range $(0, 20]$ and selected condensation coefficients m in the range $[1, 500]$ with $\tau = 2$. It can be observed that the distribution can be under-, equi- or overdispersed, with FID values in the range $(0, 5)$. For instance, for $\mu = 5$, the distribution is overdispersed when $m < 2.1406$, equidispersion if $m = 2.1406$, and underdispersed when $m > 2.1406$. In accordance with Equation (74), the FID increases linearly for large mean values ($\mu > 5$) and not too large condensation coefficient ($m \leq 10$). For very large condensation coefficients ($m > 10$), we observe, in addition to severe underdispersion (FID < 0.25 for $\mu > 1$), an oscillating index of dispersion with an amplitude approaching zero as m increases.

5. Conclusion

This paper introduces condensation, a generic stochastic mechanism for the construction of an underdispersed version of an original integer-valued distribution of interest. The method can be described as a probabilistic rounding mechanism operating on the original count variable divided by a positive integer. In this respect, the method is similar to the recent balanced discretization method [31], but operates on count distributions whereas the later starts with continuous probability distributions. An important ingredient in condensation is the distribution of a variable following the original count distribution

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

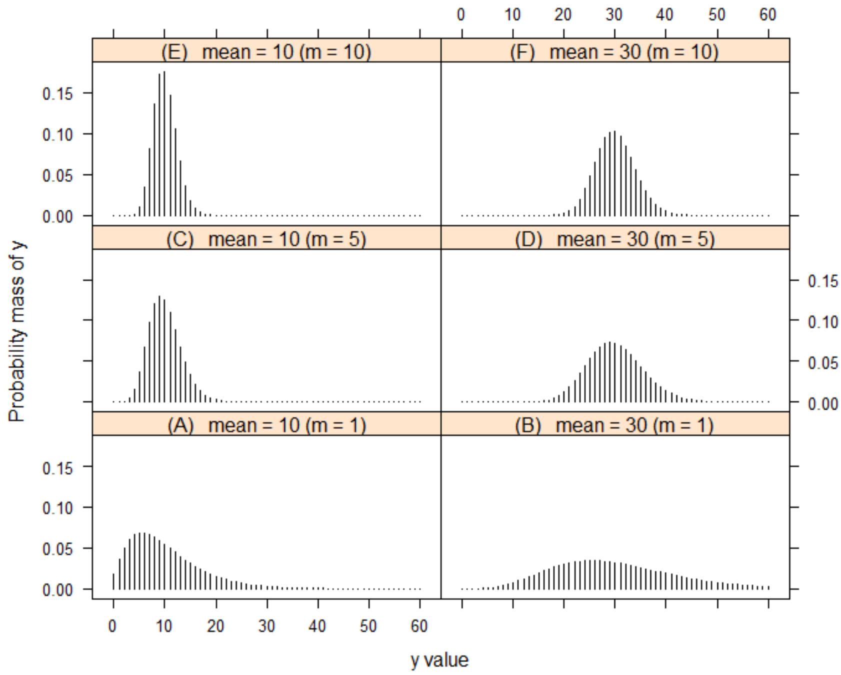

Figure 5. Probability mass of the generalized condensed negative binomial distribution with $\phi = 0.25$ for mean values of 10 and 30 and condensation coefficients $m = 1, 5$ and 10.

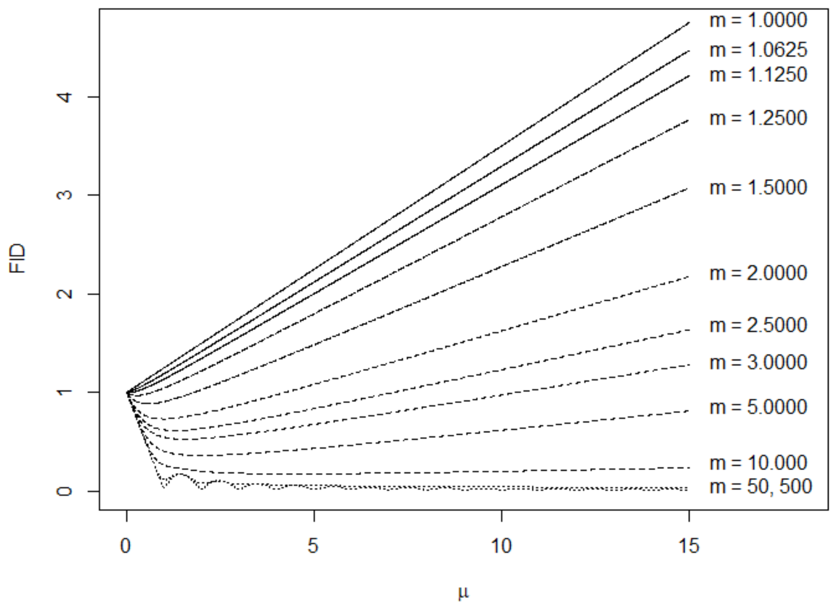

Figure 6. Fisher's index of dispersion (FID), *i.e.* variance-to-mean ratio, of the generalized condensed negative binomial distribution with $\phi = (m + 1)^2$ for mean values μ in the range $[0, 20]$ and selected condensation coefficients m in the range $[1, 500]$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

modulo a positive integer, useful for establishing basic distributional properties (generating function and higher order moments) of a condensed distribution. The general form of the probability mass function of a condensed distribution indicates that the stochastic condensation mechanism attracts probability masses from the quantiles in the tails of the original distribution and redirect them toward quantiles close to the expected value. As a result, the condensed distribution becomes more symmetrical than the original distribution. Interestingly, if the original distribution can be simulated, then the simulation of variates from a condensed distribution is straightforward. Furthermore, condensed distributions have a simple mean-parametrization, a characteristic useful in a regression context [29–31].

With a view to illustrate the proposal, we have developed the generalized condensed Poisson and negative binomial distributions. We have derived the Poisson and negative binomial remainder distributions in order to obtain the generating functions and the algebraic moments of the two developed classes of condensed distributions. We observed that some special cases of condensed Poisson and negative binomial distributions have been used in applications which demonstrate the potential of the proposal to allow full dispersion flexibility [21] in the analysis of real count data. The Consul's generalized Poisson distribution [33] is a popular overdispersion model aside the negative binomial model. For underdispersion situations, the distribution requires a sample dependent truncation of the distribution support, and the model violates a probability axiom (probabilities do not sum up to one) [15,31]. A generalized condensed version of the Consul's generalized Poisson distribution will endow the distribution with full dispersion flexibility.

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

Funding: This research received no external funding.

Appendix A. Proofs of Lemmas and Theorems

Appendix A.1. Proof of Theorem 1

Proof. Point *1.* $(Z = z) \Leftrightarrow (X = mz + r | r \in \{0, 1, \cdots, m-1\})$. Applying the law of total probability straightforwardly gives the pmf (2) and the result follows.

Point *2.* Since $u \in \{0, 1\}$ are the elementary events in the sample space of U, the probability to obtain $Y = y$ is by the law of total probability (a): $P_Y(Y = y) = \sum_{u=0}^{1} P_{Y,U}(Y = y, U = u)$. Moreover, $P_{Y,U}(Y = y, U = u) = P_{Z,U}(Z = y - u, U = u)$. Since $r \in \{0, 1, \cdots, m-1\}$ are the elementary events in the sample space of R, we have by the law of total probability $P_{Z,U}(Z = y - u, U = u) = \sum_{r=0}^{m-1} P_{X,U}(X = m(y-u) + r, U = u)$. We therefore have $P_{Y,U}(Y = y, U = u) = \sum_{r=0}^{m-1} P_{U|X}(U = u | X = m(y-u) + r) P_X(X = m(y-u) + r)$. Note that $P_X(X = m(y-u) + r) = f_P(my + r - mu | \lambda)$. $P_{U|X}(U = u | X = m(y-u) + r) = 1 - r/m$ for $u = 0$ hence (b): $P_{Y,U}(Y = y, U = 0) = \sum_{r=0}^{m-1} \frac{m-r}{m} f_P(my + r | \lambda)$. For $u = 1$, we have $P_{U|X}(U = u | X = m(y-u) + r) = \frac{r}{m}$, so that $P_{Y,U}(Y = y, U = 1) = \sum_{r=1}^{m-1} \frac{r}{m} f_P(my + r - m | \lambda)$. The change of variable $t = r - m$ yields on using $r = m + t = m - |t|$ (since $t < 0$) (c): $P_{Y,U}(Y = y, U = 1) = \sum_{t=-1}^{1-m} \frac{m-|t|}{m} f_P(my + t | \lambda)$. Finally, (b) and (c) in (a) yield Equation (3). □

Appendix A.2. Proof of Theorem 2

Proof. A proof of Equation (10) follows the same path as the proof given for Theorem 1, replacing $X \sim \mathcal{P}oi(\lambda)$ by $X \sim \mathcal{OD}(m\mu, \boldsymbol{\theta})$. Equation (11) is trivial for $m = 1$. For $j \in \mathbb{N}$, let us define $H_y(j) = \sum_{k=y-j}^{y} f_c(k | \mu, \boldsymbol{\theta}, m)$, and note that $H_y(j) \to F_c(y | \mu, \boldsymbol{\theta}, m)$ as $j \to \infty$. For $m \geq 2$, let

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

$$
\begin{aligned}
(P): H_y(j) &= \sum_{t=1-m}^{-1} \frac{m+t}{m} f_o(m(y-j)+t|m\mu,\boldsymbol{\theta}) + \sum_{t=m(y-j)}^{my} f_o(t|m\mu,\boldsymbol{\theta}) \\
&+ \sum_{t=1}^{m-1} \frac{m-t}{m} f_o(my+t|m\mu,\boldsymbol{\theta}).
\end{aligned}
$$

The proposition (P) can be proved by induction on j as follows. For $j=0$, we have $H_y(y) = f_c(y|\mu,\boldsymbol{\theta},m)$, and this is consistent with (P) by Equation (10). Next, notice from the definition of $H_y(j)$ that $H_y(k+1) = f_c(y-k-1|\mu,\boldsymbol{\theta},m) + H_y(k)$. Assuming that (P) holds for $j=k$, routine algebra give for $j=k+1$:

$$
\begin{aligned}
H_y(k+1) &= \sum_{t=1-m}^{-1} \frac{m+t}{m} f_o(m(y-(k+1))+t|m\mu,\boldsymbol{\theta}) + \sum_{t=m(y-(k+1))}^{my} f_o(t|m\mu,\boldsymbol{\theta}) \\
&+ \sum_{t=1}^{m-1} \frac{m-t}{m} f_o(my+t|m\mu,\boldsymbol{\theta}).
\end{aligned}
$$

Therefore, (P) holds for all $j \in \mathbb{N}$. As $j \to \infty$, $f_o(m(y-j)+t|m\mu,\boldsymbol{\theta}) \to 0$. Moreover, we have $\sum_{t=m(y-j)}^{my} f_o(t|m\mu,\boldsymbol{\theta}) \to F_o(my|m\mu,\boldsymbol{\theta})$ as $j \to \infty$. As a result, as $j \to \infty$, we have $H_y(j) \to F_o(my|m\mu,\boldsymbol{\theta}) + \sum_{t=1}^{m-1} \frac{m-t}{m} f_o(my+t|m\mu,\boldsymbol{\theta})$, and Equation (11) follows. From Equation (11), we have $u = F_o(my|m\mu,\boldsymbol{\theta}) + \sum_{t=1}^{m-1} \frac{m-t}{m} f_o(my+t|m\mu,\boldsymbol{\theta})$ which implies that $u > F_o(my|m\mu,\boldsymbol{\theta})$ and thus $q_o \geq my$. Next, note that Equation (11) can be rewritten:

$$
\begin{aligned}
F_c(y|\mu,m) &= F_o(my|m\mu,\boldsymbol{\theta}) + \sum_{t=1}^{m-1} \frac{m-t}{m} f_o(my+t|m\mu,\boldsymbol{\theta}) + \sum_{t=1}^{m} \frac{t}{m} f_o(my+t|m\mu,\boldsymbol{\theta}) \\
&- \sum_{t=1}^{m} \frac{t}{m} f_o(my+t|m\mu,\boldsymbol{\theta}) \\
&= F_o(my|m\mu,\boldsymbol{\theta}) + \sum_{t=1}^{m-1} f_o(my+t|m\mu,\boldsymbol{\theta}) + f_o(m(y+1)|m\mu,\boldsymbol{\theta}) \\
&- \sum_{t=1}^{m} \frac{t}{m} f_o(my+t|m\mu,\boldsymbol{\theta}) \\
&= F_o(m(y+1)|m\mu,\boldsymbol{\theta}) - \sum_{t=1}^{m} \frac{t}{m} f_o(my+t|m\mu,\boldsymbol{\theta}),
\end{aligned}
$$

which implies that $u < F_o(m(y+1)|m\mu,\boldsymbol{\theta})$ and thus $q_o \leq m(y+1)$. Overall, we have $q_o - m \leq my \leq q_o$, *i.e.* $\frac{q_o}{m} - 1 \leq y \leq \frac{q_o}{m}$. If $(q_o/m) \in \mathbb{Z}$, $y_o = (q_o/m) - 1$. In this case, if $u \leq u_o$, then $y = y_o$. Otherwise, $y = y_o + 1$. If, on the contrary, $(q_o/m) \notin \mathbb{Z}$, then $y = \lfloor q_o/m \rfloor = y_o$ (here $y_o = \lceil q_o/m \rceil - 1$) in accordance with Equation (12) (since $u = F_c(y_o|\mu,\boldsymbol{\theta},m)$). $\square$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

Appendix A.3. Proof of Theorem 3

Proof. A direct calculation using Equation (11) is straightforward:

$$\begin{aligned}
G_c(s|\mu,\boldsymbol{\theta},m) &= \sum_{z=-\infty}^{\infty} s^z \sum_{r=0}^{m-1} \frac{m-r}{m} f_o(mz+r|m\mu,\boldsymbol{\theta}) \\
&+ \sum_{z=-\infty}^{\infty} s^z \sum_{t=1-m}^{-1} \frac{m+t}{m} f_o(mz+t|m\mu,\boldsymbol{\theta}) \\
&= \sum_{z=-\infty}^{\infty} s^z \sum_{r=0}^{m-1} f_o(mz+r|m\mu,\boldsymbol{\theta}) - \sum_{z=-\infty}^{\infty} s^z \sum_{r=1}^{m-1} \frac{r}{m} f_o(mz+r|m\mu,\boldsymbol{\theta}) \\
&+ \sum_{z=-\infty}^{\infty} s^z \sum_{r=1}^{m-1} \frac{r}{m} f_o(m(z-1)+r|m\mu,\boldsymbol{\theta}) \\
&= G_q(s|\mu,\boldsymbol{\theta},m) - \frac{1}{m} \sum_{z=-\infty}^{\infty} s^z \sum_{r=1}^{m-1} r f_o(mz+r|m\mu,\boldsymbol{\theta}) \\
&+ \frac{s}{m} \sum_{z=-\infty}^{\infty} s^z \sum_{r=1}^{m-1} r f_o(mz+r|m\mu,\boldsymbol{\theta}).
\end{aligned}$$

□

Appendix A.4. Proof of Theorem 4

Proof. By the representation (9), we have $Y^n = (Z+U)^n = Z^n + \sum_{k=0}^{n-1} \binom{n}{k} Z^k U$ since $U^{n-k} = U$ for $k \neq n$. Next, using the law of iterated expectations (see (2) in [32]), we have $\mu_c^{(n)}(\mu,\boldsymbol{\theta},m) = \mathrm{E}_X\{\mathrm{E}_U\{Y^n|X = mZ+R\}\} = \mathrm{E}_X\{Z^n + \sum_{k=0}^{n-1} \binom{n}{k} Z^k R/m\}$ (since $\mathrm{E}_U\{U|X = mZ+R\} = R/m$). Setting $\mu_{qr}^{(k,l)}(\lambda,\boldsymbol{\theta},m) = \mathrm{E}_X\{Z^k R^l\}$ yields Equation (17) and replacing $Z = (X-R)/m$ gives Equation (18). □

Appendix A.5. Proof of Theorem 5

Proof. The joint pmf of W, X, U and Y follows from Equation (21) by Bayes' rule as

$$\begin{aligned}
f_{W,X,U,Y}(w,x,u,y) &= p^w(1-p)^{1-w} f_{X|W}(x|W=w,\mu,\boldsymbol{\theta},m) \\
&\quad \times \left(\frac{r}{m_o+w}\right)^u \left(1-\frac{r}{m_o+w}\right)^{1-u} I_{\{y\}}(z+u) \\
&= p^w(1-p)^{1-w} f_{X|W}(x|W=w,\mu,\boldsymbol{\theta},m) \\
&\quad \times \left(\frac{x}{m_o+w}-z\right)^u \left(1-\frac{x}{m_o+w}+z\right)^{1-u} I_{\{y\}}(z+u).
\end{aligned}$$

where $z = \lfloor x/(m_o+w) \rfloor$, $r = x-(m_o+w)z$. Then, replacing $f_{X|W}(x|W=w,\mu,\boldsymbol{\theta},m)$ by $f_o(x|(m_o+w)\mu,\boldsymbol{\theta})$, $z = y-u$, $I_{\{y\}}(z+u) = I_{\mathbb{S}_{m_o+w}}(r)$ if $u=0$ and $I_{\{y\}}(z+u) = I_{\mathbb{S}^*_{m_o+w}}(-r)$ if $u=1$ with $r = x-(m_o+w)(y-u)$ and summing over $u \in \{0,1\}$, we obtain

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

$$
\begin{aligned}
f_{W,X,Y}(w,x,y) &= \sum_{u=0}^{1} f_{W,X,U,Y}(w,x,u,y) \\
&= p^w(1-p)^{1-w} f_o(x|(m_o+w)\mu,\boldsymbol{\theta}) \\
&\quad \times\left[\left(1+y-\frac{x}{m_o+w}\right) I_{\mathbb{S}_{m_o+w}}(x-(m_o+w)y) \right.\\
&\quad \left. +\left(1-y+\frac{x}{m_o+w}\right) I_{\mathbb{S}^*_{m_o+w}}((m_o+w)y-x)\right].
\end{aligned}
$$

The reason for using $I_{\mathbb{S}^*_{m_o+w}}((m_o+w)y-x)$ instead of $I_{\mathbb{S}_{m_o+w}}((m_o+w)y-x)$ when $u=1$ is that $r=0$ implies $u=0$, so that the case $r=0$ must be ruled out if $u=1$. Next, by summing $f_{W,X,Y}(w,x,y)$ over x, we get

$$
f_{W,Y}(w,y) = p^w(1-p)^{1-w} \sum_{x=-\infty}^{\infty} f_o(x|(m_o+w)\mu,\boldsymbol{\theta})\eta(x,y,m_o+w).
$$

Setting $\mathbb{S}_{m,y}=\{my+(1-m),\cdots,my,\cdots,my+(m-1)\}$, notice that $\eta(x,y,m_o+w)\neq 0$ only for $x\in\mathbb{S}_{m_o+w,y}$, *i.e.* for $x=(m_o+w)y+t$ with $t\in\{1-(m_o+w),\cdots,0,\cdots,(m_o+w)-1\}$. In addition, note that $1+y-\frac{x}{m_o+w}=1-t/m$ (with $t\geq 0$) and $1-y+\frac{x}{m_o+w}=1+t/(m_o+w)$ (with $t<0$) can both be written $1-|t|/(m_o+w)$. We thus have

$$
\begin{aligned}
f_{W,Y}(w,y) &= p^w(1-p)^{1-w} \sum_{x=1-(m_o+w)}^{(m_o+w)-1} \frac{m_o+w-|t|}{m_o+w} f_o((m_o+w)y+t|(m_o+w)\mu,\boldsymbol{\theta}) \\
&= p^w(1-p)^{1-w} f_c(y|\mu,\boldsymbol{\theta},m_o+w).
\end{aligned}
$$

Note that the last equation is consistent with the definition of generalized condensation as the mixture of condensed distributions with $m=m_o$ and $m=m_o+1$ and weights $1-p$ and p respectively. Equation (26) follows by Bayes' rule as $f_{X|W,Y}(x|W=w,Y=y,\mu,\boldsymbol{\theta},m)=f_{W,X,Y}(w,x,y)/f_{W,Y}(w,y)$. Likely, Equation (28) follows by Bayes' rule and on using $f_Y(y)=f_g(y|\mu,\boldsymbol{\theta},m)$ as $f_{W|Y}(w|Y=y,\mu,\boldsymbol{\theta},m)=f_{W,Y}(w,y)/f_Y(y)$, which leads to $p_y=f_{W,Y}(1,y)/f_Y(y)$. Moreover, Equation (29) follows by first using Bayes's rule to obtain $f_{W,X|Y}(w,x|Y=y)=f_{X|W,Y}(x|W=w,Y=y)\times f_W(w)$ with $f_W(w)=p^w(1-p)^{1-w}$ and then summing the result over $w\in\{0,1\}$. Finally, Equation (30) is obtanied by directly summing $x^n f_{X|Y}(x|Y=y)$ over $x\in\mathbb{S}_{m_o+w,y}$ for $w\in\{0,1\}$ (since $x\in\mathbb{S}_{m_o+w,y}$ if $f_{X|W,Y}(x|W=w,Y=y)>0$). □

Appendix A.6. Proof of Lemma 1

Proof. From $t=mv+s$, we have $mj+t=m(v+j)+s$ so that $f_{m,t}(z)=\sum_{i=v}^{\infty}\frac{z^{mi+s}}{(mi+s)!}$. For $v\geq 1$, this gives $f_{m,t}(z)=\sum_{i=0}^{\infty}\frac{z^{mi+s}}{(mi+s)!}-\sum_{i=0}^{v-1}\frac{z^{mi+s}}{(mi+s)!}$ and yields Equation (37). For $v\leq 0$, because $i<0 \implies mi+s<0$, Equation (38) follows by the convention $z^j/j!=0$ if $j<0$. Next, the derivatives $f_{m,r}^{(k)}(z)=\sum_{j=0}^{\infty}\frac{z^{mj+r-k}}{(mj+r-k)!}$ are straightforwardly obtained, resulting in Equation (39). On using then Equation (38), we get $\sum_{k=0}^{m-1}f_{m,r}^{(k)}(z)=\sum_{r=0}^{m-1}f_{m,r}(z)=\sum_{j=0}^{\infty}\frac{z^k}{k!}$. Hence $f_{m,r}$ satisfies the $m-1$ order linear differential Equation (E_1): $\sum_{k=0}^{m-1}f_{m,r}^{(k)}(z)=\mathrm{e}^z$. On setting $\omega_o=\mathrm{e}^{\omega i}$ the mth root of unity with i the imaginary unit ($i^2=-1$), the general solution of the homogeneous Equation has the form $y(z)=\sum_{j=1}^{m-1}c_j\mathrm{e}^{z\omega_o^j}$ where c_j are complex coefficients [46]. A particular solution of the nonhomogeneous Equation is $y^*(z)=\frac{1}{m}\mathrm{e}^z$ so that the general solution of the non-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

homogeneous Equation (E_1) has the form $f_{m,r}(z) = \frac{1}{m}\mathrm{e}^z + \sum_{j=1}^{m-1} c_j \mathrm{e}^{z\omega_o^j}$ [25]. Evaluating the limits $f_{m,r}^{(k)}(0)$ gives the boundary conditions required for finding the coefficients c_j as

$$f_{m,r}^{(k)}(0) = \begin{cases} 1 & \text{if } k = r \\ 0 & \text{otherwise} \end{cases}.$$

From the general form of $f_{m,r}$, we have $f_{m,r}^{(k)}(z) = \frac{1}{m}\mathrm{e}^z + \sum_{j=1}^{m-1} c_j \omega_o^{jk} \mathrm{e}^{z\omega_o^j}$. This especially results in $f_{m,r}^{(r)}(0) = \frac{1}{m} + \sum_{j=1}^{m-1} c_j \omega_o^{jr} = 1$ leading to $c_j = \frac{1}{m}\omega_o{}^{-jr}$. The series is thus

$$f_{m,r}(z) = \frac{1}{m} \sum_{j=0}^{m-1} \omega_o^{-jr} \mathrm{e}^{z\omega_o^j}. \tag{A1}$$

Because $f_{m,r}$ is real valued, the sum over j of the imaginary parts of $\omega_o^{-jr}\mathrm{e}^{z\omega_o^j}$ is zero. From $\omega_o^{-jr}\mathrm{e}^{z\omega_o^j} = \exp(-jr\omega i + z\cos(j\omega) + iz\sin(j\omega)) = \exp(z\cos(j\omega) + i(z\sin(j\omega) - jr\omega))$, the real part of $\omega_o^{-jr}\mathrm{e}^{z\omega_o^j}$ is $Re\left(\omega_o^{-jr}\mathrm{e}^{z\omega_o^j}\right) = \mathrm{e}^{z\cos(j\omega)}\cos(z\sin(j\omega) - jr\omega)$ and Equation (40) follows. □

Appendix A.7. Proof of Theorem 6

Proof. From the definition 5, $f_R(r|\lambda, m) = P_X(X = mj + r, j \in \mathbb{N})$, which by Equation (36), reads $f_P(r|\lambda, m) = \mathrm{e}^{-\lambda} f_{m,r}(\lambda)$, so that Equation (41) follows by Equation (40). Next, for $r \in \{0, 1, \cdots, m-1\}$, the cdf is

$$\begin{aligned} F_R(r|\lambda, m) &= \sum_{t=0}^{r} \left\{ \frac{1}{m} + \frac{1}{m} \sum_{k=1}^{m-1} \mathrm{e}^{\lambda(\cos(k\omega)-1)} \cos(\lambda\sin(k\omega) - kt\omega) \right\} \\ &= \frac{1}{m} \left\{ r + 1 + \sum_{k=1}^{m-1} \mathrm{e}^{\lambda(\cos(k\omega)-1)} \sum_{t=0}^{r} \cos(\lambda\sin(k\omega) - kt\omega) \right\} \end{aligned}$$

and Equation (42) follows by the second equation of the Theorem in [47]. □

Appendix A.8. Proof of Theorem 7

To prove Theorem 7, we shall make use of the following lemma.

Lemma A1. *Let us consider the function $f_{m,r}$ given in Equation (40). For $k \in \mathbb{N}^+$ and $t \in \mathbb{R}$,*

$$\frac{\partial^k}{\partial t^k}\left(f_{m,r}\left(\lambda \mathrm{e}^{t/m}\right)\right) = \frac{1}{m^k} \sum_{j=1}^{k} S(k, j)\ \lambda^j\ \mathrm{e}^{jt/m} f_{m,r-j}\left(\lambda \mathrm{e}^{t/m}\right) \tag{A2}$$

where $S(k, j)$ is the Stirling number of the second kind [40,41] given by Equation (51).

Proof. We prove Equation (A2) by induction on k. The expression is obvious for $k = 1$ because $\frac{\partial}{\partial t}\left(f_{m,r}\left(\lambda \mathrm{e}^{t/m}\right)\right) = \frac{\lambda}{m}\mathrm{e}^{t/m} f_{m,r}^{(1)}\left(\lambda \mathrm{e}^{t/m}\right) = \frac{\lambda}{m}\mathrm{e}^{t/m} f_{m,r-1}\left(\lambda \mathrm{e}^{t/m}\right)$ by Equation (39) and $S(1, 1) = 1$. Assuming that Equation (A2) holds for $k \in \mathbb{N}^+$, we have

$$
\begin{aligned}
\frac{\partial^{k+1}}{\partial t^{k+1}}\left(f_{m,r}\left(\lambda e^{t/m}\right)\right) &= \frac{1}{m^k}\sum_{j=1}^{k} S(k,j)\lambda^j \frac{j}{m} e^{jt/m} f_{m,r-j}\left(\lambda e^{t/m}\right) \\
&\quad + \frac{1}{m^k}\sum_{j=1}^{k} S(k,j)\lambda^j \frac{\lambda}{m} e^{(j+1)t/m} f_{m,r-(j+1)}\left(\lambda e^{t/m}\right) \\
&= \frac{1}{m^{k+1}}\sum_{j=1}^{k} jS(k,j)\lambda^j e^{jt/m} f_{m,r-j}\left(\lambda e^{t/m}\right) \\
&\quad + \frac{1}{m^{k+1}}\sum_{j=2}^{k+1} S(k,j-1)\lambda^j\, e^{jt/m} f_{m,r-j}\left(\lambda e^{t/m}\right) \\
&= \frac{1}{m^{(k+1)}} S(k,1)\lambda e^{t/m} f_{m,r-1}\left(\lambda e^{t/m}\right) \\
&\quad + \frac{1}{m^{k+1}}\left\{\sum_{j=2}^{k} [jS(k,j)+S(k,j-1)] \frac{\lambda^j}{e^{jt/m} f_{m,r-j}}\left(\lambda e^{t/m}\right)\right\} \\
&\quad + \frac{1}{m^{k+1}} S(k,k)\lambda^{k+1}\, e^{(k+1)t/m} f_{m,r-(k+1)}\left(\lambda e^{t/m}\right).
\end{aligned}
$$

Next, from the identities $S(k,1) = S(k+1,1) = 1$ and $S(k,k) = S(k+1,k+1) = 1$ for all $k \in \mathbb{N}^+$, and the recurence relation $S(k+1,j) = S(k,j-1) + jS(k,j)$ [40], we get

$$
\begin{aligned}
\frac{\partial^{k+1}}{\partial t^{k+1}}\left(f_{m,r}\left(\lambda e^{t/m}\right)\right) &= \frac{1}{m^{k+1}} S(k+1,1)\ \lambda\ e^{t/m} f_{m,r-1}\left(\lambda e^{t/m}\right) \\
&\quad + \frac{1}{m^{k+1}}\left\{\sum_{j=2}^{k} S(k+1,j)\ \lambda^j\ e^{jt/m} f_{m,r-j}\left(\lambda e^{t/m}\right)\right\} \\
&\quad + \frac{1}{m^{k+1}} S(k+1,k+1)\ \lambda^{k+1}\ e^{(k+1)t/m} f_{m,r-(k+1)}\left(\lambda e^{t/m}\right) \\
&= \frac{1}{m^{k+1}}\sum_{j=1}^{k+1} S(k+1,j)\ \lambda^j\ e^{jt/m} f_{m,r-j}\left(\lambda e^{t/m}\right).
\end{aligned}
$$

□

We now give the proof of Theorem 7.

Proof. Equations (47) and (48) result from a direct application of Equations (23) and (17) respectively. From Equation (46), the moment generating function of $Z \sim \mathcal{G}ood(\lambda, m)$ is given, for $t \in \mathbb{R}$, by

$$
M_S(t) = e^{-\lambda}\sum_{r=0}^{m-1} e^{-\frac{rt}{m}} f_{m,r}(\lambda e^{t/m}).
$$

Set $M_S^{(n)}(t) = \partial^n M_S(t)/\partial t^n$. Since each term of $M_S(t)$ is the product of $e^{-rt/m}$ and $f_{m,r}(\lambda e^{t/m})$, the nth derivative of M_S with respect to t is by the general Leibniz's rule [48]

$$
M_S^{(n)}(t) = e^{-\lambda}\sum_{r=0}^{m-1}\sum_{k=0}^{n}\binom{n}{k}\frac{\partial^k}{\partial t^k}\left(f_{m,r}\left(\lambda e^{t/m}\right)\right)\frac{\partial^{n-k}}{\partial t^{n-k}}\left(e^{-rt/m}\right).
$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

It is easly seen that $\frac{\partial^{n-k}}{\partial t^{n-k}}\left(\mathrm{e}^{-rt/m}\right) = \left(\frac{-r}{m}\right)^{n-k}\mathrm{e}^{-rt/m}$. Furthermore, by Lemma A1, we have, for $k \geq 1$, $\frac{\partial^k}{\partial t^k}\left(f_{m,r}\left(\lambda \mathrm{e}^{t/m}\right)\right) = \frac{1}{m^k}\sum_{j=1}^{k} S(k,j)\lambda^j \mathrm{e}^{jt/m} f_{m,r-j}\left(\lambda \mathrm{e}^{t/m}\right)$. Therefore,

$$\begin{aligned}
M_S^{(n)}(t) &= \frac{(-1)^n}{m^n}\mathrm{e}^{-\lambda}\sum_{r=0}^{m-1} r^n\, \mathrm{e}^{-\frac{rt}{m}} f_{m,r}\left(\lambda \mathrm{e}^{t/m}\right) \\
&\quad + \frac{\mathrm{e}^{-\lambda}}{m^n}\sum_{r=0}^{m-1}\sum_{k=1}^{n}\binom{n}{k}(-r)^{n-k}\, \mathrm{e}^{-rt/m}\sum_{j=1}^{k} S(k,j)\lambda^j \mathrm{e}^{jt/m} f_{m,r-j}\left(\lambda \mathrm{e}^{t/m}\right) \\
&= \frac{(-1)^n}{m^n}\sum_{r=0}^{m-1} r^n\, \mathrm{e}^{-\frac{rt}{m}-\lambda} f_{m,r}\left(\lambda \mathrm{e}^{t/m}\right) \\
&\quad + \frac{1}{m^n}\sum_{k=1}^{n}\binom{n}{k}(-1)^{n-k}\sum_{j=1}^{k} S(k,j)\lambda^j\sum_{r=0}^{m-1} r^{n-k}\mathrm{e}^{-\frac{t}{m}(r-j)-\lambda} f_{m,r-j}\left(\lambda \mathrm{e}^{t/m}\right).
\end{aligned}$$

Then, on using the identity $\mu_S^{(n)}(\lambda,m) = M_S^{(n)}(0)$ and Equation (43), we obtain

$$\begin{aligned}
\mu_S^{(n)}(\lambda,m) &= \frac{(-1)^n \mu_{\mathrm{R}}^{(n)}(\lambda,m)}{m^n} \\
&\quad + \frac{(-1)^n}{m^n}\sum_{k=1}^{n}\binom{n}{k}(-1)^k\sum_{j=1}^{k} S(k,j)\lambda^j \mathrm{e}^{-\lambda}\sum_{r=0}^{m-1} r^{n-k} f_{m,r-j}(\lambda)
\end{aligned}$$

which yields Equation (49) on setting $\mu_{RC}^{(n,j)}(\lambda,m) = \mathrm{e}^{-\lambda}\sum_{r=0}^{m-1} r^n f_{m,r-j}(\lambda)$. For any $j \in \mathbb{N}$, let $s = j \mod m$ so that j has the form $j = mv + s$. Then $f_{m,r-j}(\lambda) = f_{m,r-s}(\lambda)$ by Equation (38), and $\sum_{r=0}^{m-1}\mathrm{e}^{-\lambda} f_{m,r-j}(\lambda) = \sum_{r=0}^{m-1}\mathrm{e}^{-\lambda} f_{m,r-s}(\lambda) = \sum_{l=0}^{m-1}\mathrm{e}^{-\lambda} f_{m,l}(\lambda) = \sum_{r=0}^{m-1} f_R(r|\lambda,m)$, hence Equation (53) holds. If $s = 0$, we then have $\mu_{RC}^{(n,j)}(\lambda,m) = \mathrm{e}^{-\lambda}\sum_{r=0}^{m-1} r^n f_{m,r}(\lambda)$ by Equation (38), hence Equation (54) follows. For $s > 0$,

$$\begin{aligned}
\mu_{RC}^{(n,j)}(\lambda,m) &= \mathrm{e}^{-\lambda}\sum_{r=0}^{m-1} r^{n-k} f_{m,r-s}(\lambda) \\
&= \mathrm{e}^{-\lambda}\sum_{r=0}^{s-1} r^{n-k} f_{m,m-(r-s)}(\lambda) + \mathrm{e}^{-\lambda}\sum_{r=s}^{m-1} r^{n-k} f_{m,r-s}(\lambda) \\
&= \mathrm{e}^{-\lambda}\sum_{r=m-s}^{m-1}(r+s-m)^{n-k} f_{m,r}(\lambda) + \mathrm{e}^{-\lambda}\sum_{r=0}^{m-s-1}(r+s)^{n-k} f_{m,r}(\lambda)
\end{aligned}$$

and Equation (55) follows. In order to obtain $\mu_{XR}^{(k,l)}(\lambda,m)$ by identification, we compute directly $\mu_S^{(n)}(\lambda,m) = \mathrm{E}\{Z^n\}$ (where $Z = (X-R)/m$, $R = X \mod m$ and $X \sim \mathcal{P}oi(\lambda)$) as:

$$\begin{aligned}
\mu_S^{(n)}(\lambda,m) &= \frac{1}{m^n}\sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k}\mathrm{E}\{\mathrm{X}^k \mathrm{R}^{n-k}\} \\
&= \frac{(-1)^n}{m^n}\left\{\mathrm{E}[\mathrm{R}^n] + \sum_{k=1}^{n}\binom{n}{k}(-1)^k \mathrm{E}\left[\mathrm{X}^k \mathrm{R}^{n-k}\right]\right\}.
\end{aligned}$$

Equation (50) then follows by the identification of $\mathbb{E}\left[\mathrm{X}^k \mathrm{R}^{n-k}\right]$ in the expression (49) of $\mu_S^{(n)}(\lambda,m)$. □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

Appendix A.9. Proof of Theorem 8

Proof. Note that $f_N(x|\lambda,\phi) = \int_0^\infty f_P(x|v\lambda) f_G(v|\phi,\phi)dv$ where $f_G(v|\phi,\lambda) = \frac{\lambda^\phi v^{\phi-1}}{\Gamma(\phi)} \mathrm{e}^{-\lambda v}$ is, for $v \in \mathbb{R}^+$, the probability density function (pdf) of a gamma distribution with mean ϕ/λ. Thus, $f_{\bar{R}}(r|\lambda,\phi,m) = \sum_{z=0}^\infty f_N(mz+r|\lambda,\phi) = \sum_{z=0}^\infty \int_0^\infty f_P(mz+r|v\lambda) f_G(v|\phi,\phi)dv$. On switching the integration and summation order (since all the terms are positive) and then using Equation (A1), we have

$$\begin{aligned} f_{\bar{R}}(r|\lambda,\phi,m) &= \int_0^\infty \mathrm{e}^{-\lambda v} f_{m,r}(\lambda v) f_G(v|\phi,\phi)dv \\ &= \frac{1}{m}\sum_{j=0}^{m-1} \int_0^\infty \mathrm{e}^{-\lambda v}\mathrm{e}^{-jr\omega i}\mathrm{e}^{\lambda v \omega_o^j} f_G(v|\phi,\phi)dv \\ &= \frac{1}{m}\sum_{j=0}^{m-1} \phi^\phi \mathrm{e}^{-jr\omega i} \int_0^\infty \frac{v^{\phi-1}}{\Gamma(\phi)} \mathrm{e}^{-(\lambda+\phi-\lambda \mathrm{e}^{j\omega i})v} dv \\ &= \frac{1}{m}\sum_{j=0}^{m-1} \frac{\phi^\phi \mathrm{e}^{-jr\omega i}}{(\lambda+\phi-\lambda\mathrm{e}^{j\omega i})^\phi}. \end{aligned}$$

Then, writing $\lambda+\phi-\lambda\mathrm{e}^{j\omega i} = \sqrt{((\lambda+\phi)^2 - 2\lambda(\lambda+\phi)\cos(j\omega) + \lambda^2)}\mathrm{e}^{i\arctan\left(\frac{\lambda\sin(j\omega)}{\phi+\lambda-\lambda\cos(j\omega)}\right)}$ and extracting and keeping only the real parts of the terms in the sum result in Equation (62). Equation (63) then follows by the second Equation of the Theorem in [47]. □

References

1. Barron, D.N. The analysis of count data: Overdispersion and autocorrelation. *Sociological methodology* **1992**, pp. 179–220.
2. Xekalaki, E. Under-and overdispersion. *Wiley StatsRef: Statistics Reference Online* **2014**, pp. 1–9.
3. Kokonendji, C.C. Over-and underdispersion models. *Methods and applications of statistics in clinical Trials: Planning, analysis, and inferential methods* **2014**, *2*, 506–526.
4. Dean, C.B. Overdispersion. *Wiley StatsRef: Statistics Reference Online* **2014**.
5. Crawley, M. *The R book*; Wiley Online Library, 2007; pp. 1–951.
6. King, G. Variance specification in event count models: From restrictive assumptions to a generalized estimator. *American Journal of Political Science* **1989**, pp. 762–784.
7. Hilbe, J.M. *Negative binomial regression*; Cambridge University Press, 2011; p. 573.
8. Harrison, X.A. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* **2014**, *2*, e616.
9. Zha, L.; Lord, D.; Zou, Y. The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data. *Journal of Transportation Safety & Security* **2016**, *8*, 18–35.
10. Smithson, M.; Shou, Y. Randomly stopped sums: models and psychological applications. *Frontiers in psychology* **2014**, *5*, 1279.
11. Victor, H.; Govindarajulu, Z. A note on second moment of a randomly stopped sum of independent variables. *Statistics & Probability Letters* **1992**, *14*, 275–281.
12. Sellers, K.F.; Morris, D.S. Underdispersion models: Models that are "under the radar". *Communications in Statistics-Theory and Methods* **2017**, *46*, 12075–12086.
13. Astor, T.; Strengbom, J.; Berg, M.P.; Lenoir, L.; Marteinsdóttir, B.; Bengtsson, J. Underdispersion and overdispersion of traits in terrestrial snail communities on islands. *Ecology and Evolution* **2014**, *4*, 2090–2102.
14. Ridout, M.S.; Besbeas, P. An empirical model for underdispersed count data. *Statistical Modelling* **2004**, *4*, 77–89.
15. Weiß, C.H. Integer-valued autoregressive models for counts showing underdispersion. *Journal of Applied Statistics* **2013**, *40*, 1931–1948.
16. Morse, P.M. *Queues, Inventories, and Maintenance*; John Wiley & Sons, 1958.
17. Jewell, W.S. The properties of recurrent-event processes. *Operations Research* **1960**, *8*, 446–472.
18. Whittlesey, J.R. Incomplete Gamma functions for evaluating Erlang process probabilities. *Mathematics of Computation* **1963**, pp. 11–17.
19. Whittlesey, J.R.; Haight, F.A. Counting distributions for an Erlang process. *Annals of the Institute of Statistical Mathematics* **1961**, *13*, 91–103.
20. Fisher, R.A. The effect of methods of ascertainment upon the estimation of frequencies. *Annals of eugenics* **1934**, *6*, 13–25.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 March 2021 doi:10.20944/preprints202103.0570.v1

21. Hagmark, P.E. On construction and simulation of count data models. *Mathematics and Computers in Simulation* **2008**, *77*, 72–80.
22. Chatfield, C.; Goodhardt, G.J. A consumer purchasing model with Erlang inter-purchase times. *Journal of the American Statistical Association* **1973**, *68*, 828–835.
23. Cox, D. *Renewal Theory*, 1 ed.; Methuen & Co., 1962.
24. Frederick S. Hillier, G.J.L. *Introduction to operations research*, 3rd ed ed.; McGraw-Hill series in industrial engineering and management science, McGraw-Hill, 1980.
25. Serfling, R. The variance function of the Erlang process. *Annals of the Institute of Statistical Mathematics* **1970**, *22*, 327–337.
26. Goodman, L.A. On the Poisson-gamma distribution problem. *Annals of the Institute of Statistical Mathematics* **1951**, *3*, 123–125.
27. Haight, F.A. The generalized Poisson distribution. *Annals of the Institute of Statistical Mathematics* **1959**, *11*, 101–105.
28. Alanko, T. Modelling the true intake distribution from recall data with measure- ment/response errors. The Bulletin of the International Statistical Institute Proceedings of the 58th World Statistics Congress, 21st-26th August 2011; International Statistical Institute, , 2012; Vol. 1.
29. Huang, A. Mean-parametrized Conway–Maxwell–Poisson regression models for dispersed counts. *Statistical Modelling* **2017**, *17*, 359–380.
30. Sáez-Castillo, A.; Conde-Sánchez, A. A hyper-Poisson regression model for overdispersed and underdispersed count data. *Computational Statistics & Data Analysis* **2013**, *61*, 148–157.
31. Tovissodé, C.F.; Honfo, S.H.; Doumatè, J.T.; Glèlè Kakaï, R. On the Discretization of Continuous Probability Distributions Using a Probabilistic Rounding Mechanism. *Mathematics* **2021**, *9*, 555.
32. Kaehler, J. Laws of iterated expectations for higher order central moments. *Statistical Papers* **1990**, *31*, 295–299.
33. Consul, P.C.; Jain, G.C. A generalization of the Poisson distribution. *Technometrics* **1973**, *15*, 791–799.
34. 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.
35. Meng, X.L.; Rubin, D.B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **1993**, *80*, 267–278.
36. Morrison, D.G.; Schmittlein, D.C. Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *Journal of Business & Economic Statistics* **1988**, *6*, 145–159.
37. Schmittlein, D.G.; Morrison, D.G. Prediction of future random events with the condensed negative binomial distribution. *Journal of the American Statistical Association* **1983**, *78*, 449–456.
38. Tan, F.E. Statistical Models for Estimating the Distribution Function of Alcohol Consumption. A parametric approach. *Addiction* **1999**, *94*, 927.
39. S., R. *Stochastic Processes*, second edition ed.; John Wiley & Sons Inc, 1996.
40. Abramowitz, M.; Stegun, I. Stirling Numbers of the Second Kind. In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; Abramowitz, M.; Stegun, I., Eds.; 1972; chapter 24, pp. 824–825.
41. Riordan, J. Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions. *The Annals of Mathematical Statistics* **1937**, *8*, 103–111.
42. Philipson, C. A note on moments of a Poisson probability distribution. *Scandinavian Actuarial Journal* **1963**, *1963*, 243–244.
43. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
44. Ehrenberg, A.S. The pattern of consumer purchases. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **1959**, *8*, 26–41.
45. Zufryden, F.S. A composite heterogeneous model of brand choice and purchase timing behavior. *Management Science* **1977**, *24*, 121–136.
46. Kaplan, W. *Ordinary Differential Equations*; Addison-Wesley, Reading, Massachusetts, 1958.
47. Knapp, M.P. Sines and cosines of angles in arithmetic progression. *Mathematics magazine* **2009**, *82*, 371–372.
48. Brand, L. *Advanced Calculus - An Introduction to Classical Analysis*, 3rd ed ed.; John Wiley & Sons, 1955; p. 111.