Article

An Over and Underdispersed Biparametric Extension of the Waring Distribution

Valentina Cueva-López [1,‡] 0000-0001-9485-9737, María José Olmo-Jiménez [2,‡] 0000-0003-3527-3239 and José Rodríguez-Avi [3,*‡] 0000-0002-1673-9876

[1] Department of Statistics and Operations Research, University of Jaén (Spain); vcueva@ujaen.es
[2] Department of Statistics and Operations Research, University of Jaén (Spain); mjolmo@ujaen.es
[3] Department of Statistics and Operations Research, University of Jaén (Spain); jravi@ujaen.es
* Correspondence: jravi@ujaen.es; Tel.: +34953212207
‡ These authors contributed equally to this work.

Abstract: A new discrete distribution for count data called extended biparametric Waring (EBW) distribution is developed. Its name is related to the fact that, in a specific configuration of its parameters, it can be seen as a biparametric version of the univariate generalized Waring (UGW) distribution, a well-known model for the variance decomposition into three components: randomness, liability and proneness. Unlike the UGW distribution, the EBW can model both overdispersed and underdispersed data sets. In fact, the EBW distribution is a particular case of a UWG distribution when its first parameter is positive; otherwise, it is a particular case of a Complex Triparametric Pearson (CTP) distribution. Hence, this new model inherits most of their properties and, moreover, it helps to solve the identification problem in the variance components of the UGW model. We compare the EBW with the UGW by a simulation study, but also with other over and underdispersed distributions through the Kullback-Leibler divergence. Additionally, we have carried out a simulation study in order to analyse the properties of the maximum likelihood parameter estimates. Finally, some application examples are included which show that the proposed model provides similar or even better results than other models, but with fewer parameters.

Keywords: count data distribution; goodness of fit; overdispersion ; underdispersion

1. Introduction

The univariate generalized Waring (UGW) is a triparametric distribution for overdispersed count data that has been studied by [1–4], among others. The interest of the UGW distribution lies in the decomposition of its variance into three components, randomness, liability and proneness, which allows us to get a deeper knowledge of the nature of data variability, that is, how and why data vary. Whereas the Poisson distribution provides the simplest answer to this issue (pure chance), any one-step Poisson mixture distributions assume that there are only two sources of variability (for example, the negative binomial or NB distribution which is a Poisson-Gamma mixture).

For this reason, the UGW distribution and the related regression model [5–7] have been widely applied for modelling overdispersed count data sets in different fields, such as lexicology [8], the number of authors in scientific articles [9], the evolution of the number of links in the World Wide Web [10], accident theory [11], clustered data [12], sources of variance in motor vehicle crash analysis [13], completeness errors in geographic data sets [14] or agriculture [15].

However, the UGW distribution has a serious drawback related to the variance decomposition. Since its first two parameters are interchangeable in the expression of the probability mass function (pmf), it is difficult to determine which component refers to liability or proneness. There are in the literature some

suggestions available to avoid this problem. [1] recommends choosing the values of liability and proneness according to the researcher experience; [11] proposes the calculus of a bivariate version of the Waring distribution and [5] solves the problem using additional information provided by covariates through a regression model. In all cases the solution of the indetermination needs external information that is not always available.

Several extensions have been developed, such as the extended Waring distribution or $GHDI$ [4], the Stuttering generalized Waring distribution [16] and the bivariate generalized Waring distribution [11], but they do not manage to solve the identification problem aforementioned.

In this paper we study a specific biparametric distribution within the Gaussian hypergeometric distributions (GHD) family [17] and we propose it as an extension of the UGW distribution but with only two parameters. The proposed model does not only perform similar to the UGW distribution for overdispersed data sets but also solves the identification problem of the variance components. Moreover, the way in which the extension is carried out also allows for modelling underdispersed count datasets, since it can be seen as a particular case of a complex triparametric Pearson (CTP) distribution [18,19] although, in this case, the result of decomposition of the variance is not verified because the model cannot be expressed as a Poisson mixture. Thus, this extension - that we will call extended biparametric Waring (henceforward EBW) distribution - inherits the good properties of the UGW and CTP distributions.

The rest of the paper is laid out as follows. Section 2 is devoted to defining the EBW distribution and to exploring its main probabilistic properties. In Section 3 some estimation methods are described and the properties of the maximum likelihood estimators are analised by a simulation study. In Section 4 we compare the EBW distribution with some other biparametric over- and underdispersed distributions. Some examples of application to real over- and underdispersed data that illustrate the versatility of the proposed model are included in Section 5. Finally, in Section 6, some conclusions of the current research are presented.

2. The Extended bivariate Waring distribution

2.1. Definition

The GHD family, generated by the ${}_2F_1(\alpha,\beta;\gamma,\lambda)$ function

$$ {}_2F_1(\alpha,\beta;\gamma;\lambda)=\sum_{x=0}^{\infty}\frac{(\alpha)_x(\beta)_x}{(\gamma)_x}\frac{\lambda^x}{x!}, \qquad ,x=0,1,2\ldots$$

with $(\gamma)_x=\frac{\Gamma(\gamma+x)}{\Gamma(\gamma)}$, $\alpha,\beta\in\mathbb{C}$ and $\gamma,\lambda\in\mathbb{R}$, arises as a solution of the difference equation

$$G(x)f(x+1)-L(x)f(x)=0, \quad x=0,1,2,\ldots \tag{1}$$

where G and L are quadratic polynomials with real coefficients, $G(x)=(\gamma+x)(x+1)$ and $L(x)=\lambda(x-\alpha)(x-\beta)$ [20]. When convergence, positivity and normalization conditions are verified, the solution $f(x)$ is the pmf of as a discrete distribution, that is

$$f(x)=P(X=x)=\frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta){}_2F_1(\alpha,\beta;\gamma;\lambda)}\frac{\Gamma(\alpha+x)\Gamma(\beta+x)}{\Gamma(\gamma+x)}\frac{\lambda^x}{x!} \tag{2}$$

and the probability generating function (pgf)

$$G(t)={}_2F_1(\alpha,\beta;\gamma;\lambda t)/{}_2F_1(\alpha,\beta;\gamma;\lambda), \quad t\in\mathbb{R}.$$

It is important to point out that the first three parameters of the GHD are the roots of the polynomials $L(x)$ and $G(x)$ (except the sign of γ).

A thorough classification of the GHD family in terms of the parameters can be seen in [4]. In the aforementioned paper, a detailed study of the GHD when α, β and γ are positive real numbers and $0 < \lambda \leq 1$ (denoted by type I) is made. The case when α and β are conjugate complex numbers, $\gamma > 0$ and $0 < \lambda \leq 1$ (denoted by type II distributions) has been studied in [18,19,21]. Type VII distributions, a finite case which may be seen as a generalization of the beta-binomial model, have been addressed by [22]. Likewise, the case when $\lambda = 1$ has been analysed by [23,24], among others.

In this paper we focus on the case in which both GHD of type I and type II with $\lambda = 1$ converge. Thus, $L(x)$ in (1) has a real double root, that is, $\alpha = \beta$. Then, the solution of (1) is given in terms of a ${}_2F_1(\alpha, \alpha; \gamma; 1)$ function leading to a highly versatile biparametric discrete distribution with infinite range which is formalized in the following definition. From now on we will call it EBW, the acronym of *Extended Bivariate Waring*, distribution. Later on we will explain the nomenclature chosen for this new distribution.

Definition 1. *A random variable X following a $EBW(\alpha, \gamma)$ distribution is defined by the following pmf*

$$P(X = x) = \frac{\Gamma(\gamma - \alpha)^2}{\Gamma(\alpha)^2 \Gamma(\gamma - 2\alpha)} \frac{\Gamma(\alpha + x)^2}{\Gamma(\gamma + x)} \frac{1}{x!}, \quad x = 0, 1, \ldots \tag{3}$$

where $\alpha \in \mathbb{R}$ and $\gamma > \max(0, 2\alpha)$.

The mean, μ, and variance, σ^2, of X are

$$\mu = \frac{\alpha^2}{\gamma - 2\alpha - 1}, \quad \sigma^2 = \frac{\alpha^2(\gamma - \alpha - 1)^2}{(\gamma - 2\alpha - 1)^2(\gamma - 2\alpha - 2)} = \mu \frac{\mu + \gamma - 1}{\gamma - 2\alpha - 2} \tag{4}$$

so it is necessary that $\gamma > 2\alpha + 1$ and $\gamma > 2\alpha + 2$ to guarantee the existence of μ and σ^2, respectively. In general, it can be proved that $\gamma > 2\alpha + m$ to guarantee the existence of the m-th raw moment.

2.2. Properties

To study the properties of the EBW distribution we will distinguish among $\alpha > 0$ and $\alpha < 0$ but $\alpha \notin \mathbb{Z}^-$.

2.2.1. $\alpha > 0$

It is necessary that $\gamma > 2\alpha$, so we consider another parametrization of the distribution in terms of α and $\rho = \gamma - 2\alpha > 0$. Then, the expression of the pmf given in (3) is now

$$P(X = x) = \frac{\Gamma(\alpha + \rho)^2}{\Gamma(\alpha)^2 \Gamma(\rho)} \frac{\Gamma(\alpha + x)^2}{\Gamma(2\alpha + \rho + x)} \frac{1}{x!}, \quad x = 0, 1, \ldots \tag{5}$$

and the expressions in (4) reduce to

$$\mu = \frac{\alpha^2}{\rho - 1}, \quad \sigma^2 = \frac{\alpha^2(\alpha + \rho - 1)^2}{(\rho - 1)^2(\rho - 2)} = \mu \frac{\mu + 2\alpha + \rho - 1}{\rho - 2}. \tag{6}$$

To guarantee the existence of the $m-$th raw moment it is necessary that $\rho > m$.

Theorem 1. *The $EBW(\alpha, \rho)$ distribution with $\alpha, \rho > 0$ is a $UGW(\alpha, \alpha, \rho)$ distribution.*

Proof. Considering $\alpha = \beta > 0$ and $\lambda = 1$ in (2) and applying that

$$ {}_2F_1(\alpha, \beta; \gamma; 1) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}, $$

it is easy to see that the pmf given in (5) coincides with that of a $UGW(\alpha, \alpha, \rho)$ distribution. □

Hence, our model may be seen as a biparametric case of a UGW distribution when $\alpha > 0$. As a consequence, it inherits the properties of the UGW distribution which are listed below:

1. It can be obtained from a two-step Poisson mixture:
   - $X|\lambda \sim \mathcal{P}(\lambda)$
   - $\lambda|\alpha, v \sim Gamma(\alpha, v)$ with density

     $$ f(\lambda|\alpha, v) = \frac{1}{\Gamma(\alpha)v^\alpha}\lambda^{\alpha-1}e^{-\lambda/v}, \quad \lambda > 0, \quad \alpha, v > 0 $$

     Therefore, $X|\alpha, v \sim NB(\alpha, v)$ with pmf

     $$ f(x|\alpha, v) = \frac{1}{x!}\frac{\Gamma(x+\alpha)}{\Gamma(\alpha)}\left(\frac{1}{1+v}\right)^\alpha\left(\frac{v}{1+v}\right)^x, \quad x = 0, 1, \ldots $$

   - $v|\alpha, \rho \sim Beta(\alpha, \rho)$ with density

     $$ f(v|\alpha, \rho) = \frac{\Gamma(\alpha+\rho)}{\Gamma(\alpha)\Gamma(\rho)}v^{\alpha-1}(1+v)^{\alpha+\rho}, \quad v > 0, \quad \alpha, \rho > 0 $$

2. Since the EBW distribution with $\alpha > 0$ is a Poisson mixture, it is always overdispersed.
3. It converges to $\mathcal{P}(\mu)$ when ρ and $\alpha^2 \to \infty$ with the same order of convergence.
4. As a consequence of the mixture, the variance of X can be split into three components known as randomness, liability and proneness, respectively:

   $$ \sigma^2 = \frac{\alpha^2}{\rho - 1} + \frac{\alpha^2(\alpha+1)}{(\rho-1)(\rho-2)} + \frac{\alpha^3(\alpha+\rho-1)}{(\rho-1)^2(\rho-2)}. \tag{7} $$

   Since we have got rid of one of the first two parameters of the UGW distribution, the indetermination problem with regard to the components of the variance [4] disappears in the biparametric model and, therefore, it is not necessary to provide additional information when determining the partition of the variance.

In order to know the effect of each parameter on the variance components of the EBW model we consider the proportion of variance explained by each one, that is:

$$ 1 = \frac{(\rho-1)(\rho-2)}{(\alpha+\rho-1)^2} + \frac{(\alpha+1)(\rho-1)}{(\alpha+\rho-1)^2} + \frac{\alpha}{(\alpha+\rho-1)} $$

Figure 1 shows the evolution of the variance partition percentages for each component considering α fixed and ρ variable (low and high values) and then, ρ fixed and α variable (low and high values). In the first column (ρ fixed), we can observe that the greater α is, the more important is the proneness. In the second column (α fixed), the greater ρ is, the more important is the randomness. Otherwise, if α and ρ increase with the same convergence order, the proneness has a lower limit in 50% of the variance, whereas the other two parts tend to 25% each one.

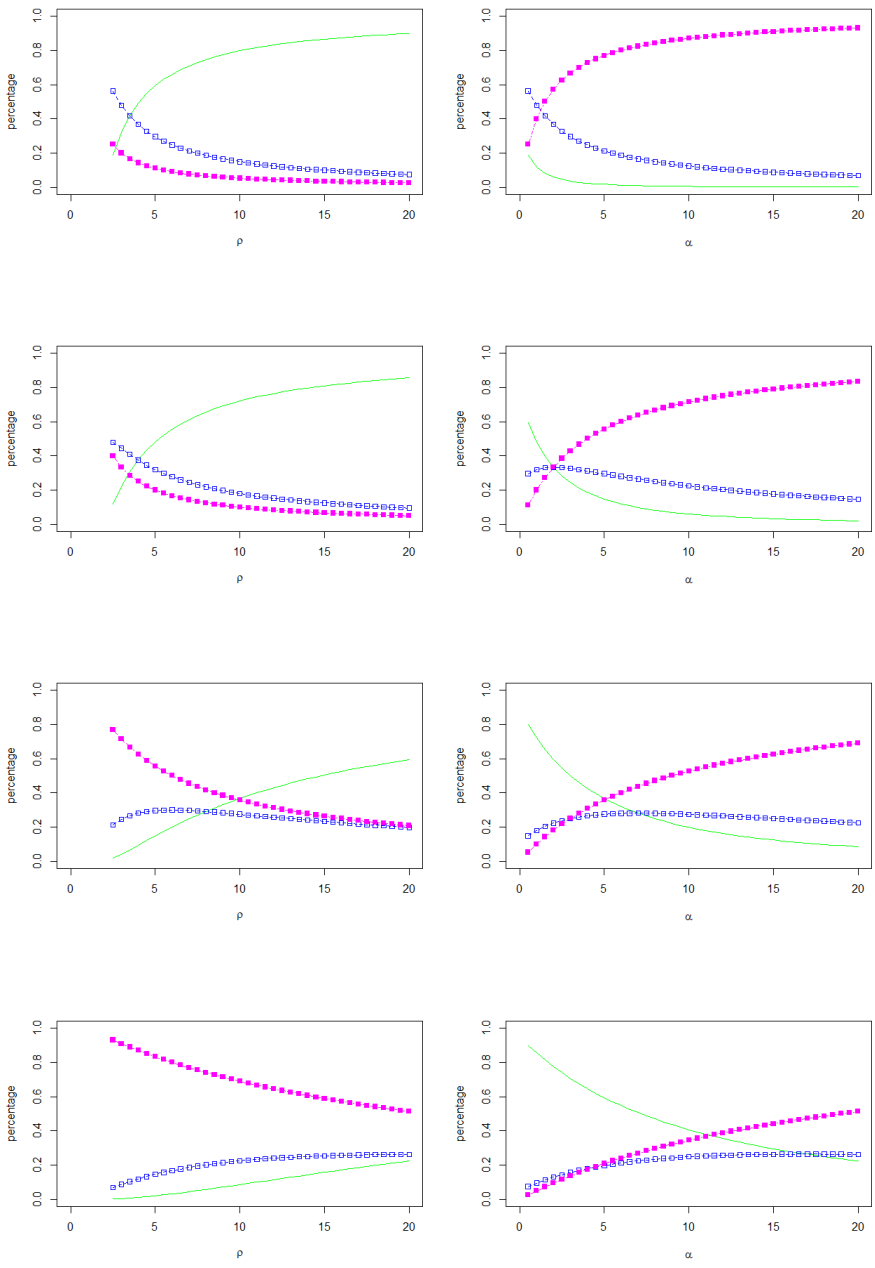

Figure 1. Percentages of the *EBW* variance components: randomness (green solid line), liability (blue dashed line) and proneness (purple dotted line). Column 1 has values of $\alpha = 0.5, 1, 5$ and 20, respectively, for ρ from 2.1 to 20. Column 2 has values of $\rho = 2.5, 5, 10$ and 20, respectively, for α from 0.1 to 20

Due to the structure of the UGW distribution in which the first two parameters are interchangeable and appear in a multiplicative form in the pmf, moments and decomposition of the variance, the maximum likelihood estimates of its first two parameters are usually almost equal. In those cases the EBW type I distribution has the property of providing a similar fit but with one more degree of freedom. In general, the EBW distribution is able to provide acceptable fits for data simulated from a UGW distribution. This implies that, in most cases, there exist a EBW model reasonably similar to the UGW, but with the advantage of having fewer parameters. To show this fact we have simulated $M = 1000$ samples of size $N = 100, 300$ and 500 from a UGW distribution with several values of its parameters and, for each sample, we have obtained the corresponding EBW and UGW fits. All the estimates have been computed by the maximum likelihood method. We have implemented our own functions in R [25] using the `optim` function of the `stats` package and considering as initial values the estimates provided by the method of moments(see Section 3 for more details).

Table 1. Percentage of: (1) EBW fits achieved for UGW generated data; (2) EBW fits with less AIC value than the corresponding UGW fit; (3) samples that come from a EBW model at 5% significance level according to the χ^2-goodness of fit test

	Achieved EBW fits			$< AIC$			$p-value > 0.05$		
	N			N			N		
$UGW(a,k,\rho)$	100	300	500	100	300	500	100	300	500
$(0.5,1,2.5)$	94.4	93.4	95.9	92.6	89.3	86.4	95.4	94.1	93.3
$(0.5,10,2.5)$	99.8	100	100	53.4	25.1	9.9	86.1	86.8	90.5
$(0.5,10,20)$	95.2	93.8	94.4	99.1	95.5	89	95.4	95.3	93.3
$(1.5,3,2.5)$	100	100	100	88.3	88.8	87.2	91.2	87.5	84.3
$(1.5,3,25)$	97.9	99	98.5	100	100	99.9	95.9	94.3	93.7
$(1.5,20,25)$	100	100	100	96.9	82.3	74.3	95.5	94.3	92.9
$(4,6,2.5)$	99.9	100	100	90	89.8	91.1	82.8	77	73.9
$(4,6,10)$	100	100	100	96.3	91.1	92.3	93.7	91.4	91.4
$(4,6,50)$	95.8	96.5	97.6	100	99.9	99.8	94	91	91.8

For each group of 1000 samples we have computed the percentage of EBW fits achieved as well as the percentage of these fits which are better than the corresponding UGW fit in two senses: the AIC value and the χ^2-goodness of fit test. Specifically, for the EBW fits achieved we have computed the percentage of them whose AIC value is less than that of the UGW fit and the percentage of $p-$values in the χ^2-goodness of fit test greater than 0.05 (that is, the null hypothesis *data comes from a EBW model* cannot be rejected). These results appear in Table 1.

2.2.2. Case $\alpha < 0$ but $\alpha \notin \mathbb{Z}^-$

It can be seen as a particular case of a CTP distribution [18,21], which arises when the polynomial $L(x)$ in (1) has conjugate complex roots $\alpha = a + ib$ and $\beta = a - ib$; specifically, we have the following result.

Theorem 2. *If $\alpha < 0$ the $EBW(\alpha,\gamma)$ distribution with $\gamma > 0$ is a $CTP(\alpha,0,\gamma)$ distribution.*

Proof. The proof is straightforward since the pmf of the $CTP(\alpha,0,\gamma)$ with $\alpha \in \mathbb{R}$ and $\gamma > 0$ [see for instance 21] coincides with the pmf of the $EBW(\alpha,\gamma)$ given in (3). □

This result is also true when $\alpha > 0$. So, a $UGW(\alpha,\alpha,\rho) \equiv CTP(\alpha,0,2\alpha+\rho)$.

At this point we can justify the name chosen for the model proposed. On the one hand, when $\alpha > 0$ the model may be seen as a biparametric case of the UGW distribution, which is always overdispersed and that may replace it with fewer parameters; on the other hand, when $\alpha < 0$ the model can be underdispersed, so it may be considered as an underdispersed extension of a biparametric UGW distribution.

Once again, the proposed distribution inherits the properties of another distribution, in this case of the CTP distribution that we next summarize:

1. If $\frac{(\alpha-1)^2}{\gamma-2\alpha+1} \in \mathbb{Z}$, the distribution has two consecutive modes in the values

$$\frac{(\alpha-1)^2}{\gamma-2\alpha+1} - 1 = \frac{\alpha^2-\gamma}{\gamma-2\alpha+1} \quad \text{and} \quad \frac{(\alpha-1)^2}{\gamma-2\alpha+1}.$$

   Otherwise the distribution is unimodal with mode in 0 if $\alpha^2 < \gamma$ or in $\left[\frac{(\alpha-1)^2}{\gamma-2\alpha+1}\right]$, where $[\cdot]$ symbolises the integer part. Hence, the pmf is J-shaped or bell-shaped.
2. It may be underdispersed, equidispersed or overdispersed. Specifically:
   - It is underdispersed when $\alpha \leq -1$ or when $-1 < \alpha < -0.5$ and $\gamma > \frac{3\alpha^2+4\alpha+1}{2\alpha+1}$.
   - It is equidispersed when $-1 < \alpha < -0.5$ and $\gamma = \frac{3\alpha^2+4\alpha+1}{2\alpha+1}$.
   - It is overdispersed when $\alpha \geq -0.5$ or when $-1 < \alpha < -0.5$ and $\gamma < \frac{3\alpha^2+4\alpha+1}{2\alpha+1}$.
3. A sufficient condition to be infinitely divisible (i.d.) is that $\alpha > -0.5$ and $\gamma > \alpha^2/(1+2\alpha)$. So, if $\alpha \leq -0.5$ the EBW distribution is not i.d. As a consequence, an underdispersed EBW cannot be i.d. since a necessary condition to be underdispersed is $\alpha < -0.5$.
4. It converges to the:
   - $\mathcal{P}(\mu)$ when γ and $\alpha^2 \to \infty$ with the same order of convergence.
   - Normal distribution, $\mathcal{N}(\mu, \sigma)$, when γ and $|\alpha|$ have the same order of convergence.

The CTP distribution cannot be expressed as a mixture, so in the EBW with $\alpha < 0$ there is no a result of variance decomposition.

3. Estimation

3.1. Methods for obtaining estimators

We can estimate the two parameters of the EBW distribution using the method of moments and the maximum likelihood estimation method.

To apply the method of moments, we first solve the equations given in (4). To this end we substitute $\gamma - 2\alpha - 1 = \alpha^2/\mu$ in the equation of σ^2. Then,

$$\sigma^2 = \mu \frac{\mu^2 + \alpha^2 + 2\alpha\mu}{\alpha^2 - \mu},$$

which is equivalent to $\alpha^2(\sigma^2-\mu)-2\mu^2\alpha-\mu(\mu^2+\sigma^2)=0$. Then, replacing μ and σ^2 by their sample counterparts, $\overline{x}$ and s^2, and solving the equation there are two possible estimates for α by the method of moments:

$$\begin{aligned}\widehat{\alpha}_1 &= \frac{\overline{x}^2+\sqrt{\overline{x}^4+\overline{x}(s^2-\overline{x})(\overline{x}^2+s^2)}}{s^2-\overline{x}}\\ \widehat{\alpha}_2 &= \frac{\overline{x}^2-\sqrt{\overline{x}^4+\overline{x}(s^2-\overline{x})(\overline{x}^2+s^2)}}{s^2-\overline{x}}\end{aligned}$$

It is clear that if data exhibit overdispersion, then $\widehat{\alpha}_1>0$ and $\widehat{\alpha}_2<0$. On the other hand, if data are underdispersed both $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ are negative. Estimated α, the estimate of γ is calculated as $\widehat{\gamma}=\widehat{\alpha}^2/\overline{x}+2\widehat{\alpha}+1$. Hence, there are also two possible estimates for γ with the only restriction of being positive, which it is true when:

- $0<\overline{x}<1$ or
- $\overline{x}>1$ and $\widehat{\alpha}<-\overline{x}-\sqrt{\overline{x}(\overline{x}-1)}$ or $\widehat{\alpha}>-\overline{x}+\sqrt{\overline{x}(\overline{x}-1)}$.

Using the MLE method we have to maximize the log-likelihood function. Thus, if $\mathbf{x}'=(x_1,\ldots,x_n)$ is a sample of size n, the expression of the log-likelihood function is:

$$\begin{aligned}\ln L_{x_1,\ldots,x_n}(\alpha,\rho)=&\sum_{i=1}^{n}\left[2\ln\Gamma(\alpha+x_i)-\ln\Gamma(\rho+2\alpha+x_i)\right]\\ &-n\left[2\ln\Gamma(\alpha)-2\ln\Gamma(\rho+\alpha)+\ln\Gamma(\rho)\right],\end{aligned}\tag{8}$$

when $\alpha>0$, using the parametrization given in Section 2.2.1, or

$$\begin{aligned}\ln L_{x_1,\ldots,x_n}(\alpha,\gamma)=&\sum_{i=1}^{n}\left[2\ln\Gamma(\alpha+x_i)-\ln\Gamma(\gamma+x_i)\right]\\ &-n\left[2\ln\Gamma(\alpha)-2\ln\Gamma(\gamma-\alpha)+\ln\Gamma(\gamma-2\alpha)\right],\end{aligned}\tag{9}$$

in another case. Both expressions can be maximized using numerical methods. In particular, we have used the $L-BFGS-B$ method implemented in the `optim` function of the `MASS` package in `R`. This method allows box constraints on the parametric space, so we can impose $\rho>0$ or $\gamma>0$ in (8) and (9), respectively. We consider the estimates obtained by the method of moments as initial values, in such a way that we maximize (8) if $\widehat{\alpha}>0$ or (9) in another case.

3.2. Properties of the estimators

We have carried out a simulation study in order to analyse the performance of the estimates of the model parameters. Specifically, we have simulated $M=1000$ samples of size $N=500$ of the EBW distribution and we have fitted the EBW model for each sample using the MLE method described in the previous section.

We have considered two scenarios: $\alpha>0$, in which case the EBW distribution is always overdispersed, and $\alpha<0$, in which case the EBW distribution can be under- and overdispersed. In all cases the values of the parameters satisfy the conditions for the existence of μ and σ^2.

Results of the simulation procedure are shown in Table 2. Thus, Column 1 contains the mean bias and the s.d., in brackets (* indicates a significant bias at 5% level based on a normal 95% confidence interval, given that there are 1000 observations). Column 2 shows the average of the mean square error (MSE) of the parameter estimates and Column 3 the percentage of simulations in which the parameter estimate does

Table 2. Mean bias and s.d. in brackets (* indicates a statistically significant bias at 5% level), average MSE and coverage for EBW fits

$\alpha > 0$				$\alpha < 0$			
Parameters	Bias (s.d.)	MSE	Coverage	Parameters	Bias (s.d.)	MSE	Coverage
$\alpha = 0.5$	0.02(.10)*	0.02	96.3	$\alpha = -0.75$	-0.01(0.05)*	0.00	95.1
$\rho = 2.1$	0.23(.80)*	1.30	94.8	$\gamma = 0.75$	0.02(0.11)*	0.02	95.2
$\alpha = 1$	0.01(.10)*	0.02	95.5	$\alpha = -0.75$	-0.01(0.08)*	0.01	97.1
$\rho = 2.1$	0.07(.36)*	0.27	96.3	$\gamma = 1.5$	0.07(0.35)*	0.26	95.3
$\alpha = 1.5$	0.01(.13)*	0.03	94.2	$\alpha = -0.75$	-0.02(0.14)*	0.05	96.8
$\rho = 2.1$	0.04(.29)*	0.16	94.5	$\gamma = 3$	0.29(1.34)*	5.01	94
$\alpha = 1.5$	0.01(.14)*	0.04	94.4	$\alpha = -1.5$	0.01(0.09)	0.01	94.2
$\rho = 2.5$	0.06(.39)*	0.30	94.4	$\gamma = 0.75$	0.00(0.13)	0.03	94.1
$\alpha = 1.5$	0.03(.29)*	0.06	96.1	$\alpha = -2.5$	-0.00(0.10)	0.02	94.5
$\rho = 3.5$	0.14(.70)*	1.00	96.6	$\gamma = 0.75$	0.01(0.20)	0.07	94.2

not differ significantly at 5% from the true value, known as coverage. We have only included low values of ρ and γ because the higher these values are compared with α, the lower the mean and the variance are. In fact, if these parameters tend to infinity, holding α fixed, the EBW distribution degenerates into 0. In addition, if both α and ρ (or γ) are high, the EBW is similar to the Poisson or the Normal distribution.

In general, we can deduce that:

- If $\alpha > 0$ the estimates are biased to the right, but the bias decreases as α increases, holding ρ fixed. The opposite happens with the bias of ρ, which increases as ρ increases.
- If $\alpha < 0$ the estimates are also biased, those for α to the left and for γ to the right, but the bias disappears as α decreases ($\alpha < -1$). Holding γ fixed, the bias decreases as α decreases and the same happens for γ.
- The average MSE is low for both parameter estimates, although this measure increases as ρ (or γ) increases since the estimates accuracy and precision decrease.
- Regarding the coverage, it approaches 95%, the confidence level considered, so it shows the validity of the inference made.

4. Comparison with other count data distributions

Next we study the differences and similarities between the EBW and other well-known biparametric discrete distributions for count data using the Kullback-Leibler (KL) divergence [for more details see 26]. Specifically, we consider the distributions NB, Complex Biparametric Pearson or CBP [19], which is a particular case of the CTP distribution, Generalized Poisson or GP [27,28], COM-Poisson or CMP [29,30] and Hyper-Poisson or HP [31]. The first two are suitable only for overdispersed data, whereas the other three can cope with both underdispersed and overdispersed data, although the GP has finite range in the underdispersed case.

We focus on the overdispersed scenario since the underdispersed one, for being the EBW distribution a particular case of the CTP distribution, was already carried out by [21].

To compute the KL divergence between the EBW distribution and the above-mentioned distributions (and vice versa), we have considered several values of μ and σ^2, with $\sigma^2 > \mu$, and then we have obtained the corresponding values of the parameters of each distribution (see A). For the CMP and HP distributions it should be taken into account that not all the combinations of μ and σ^2 are possible; empirically there seems to be an upper limit for σ^2 in $\mu(\mu+1)$. Thus, the values of the KL divergence are shown in Figure 2.

In general, we can observe that in an overdispersed scenario the most distant models from the EBW distribution are the CBP and HP distributions and the closest ones to the EBW distribution are the GP

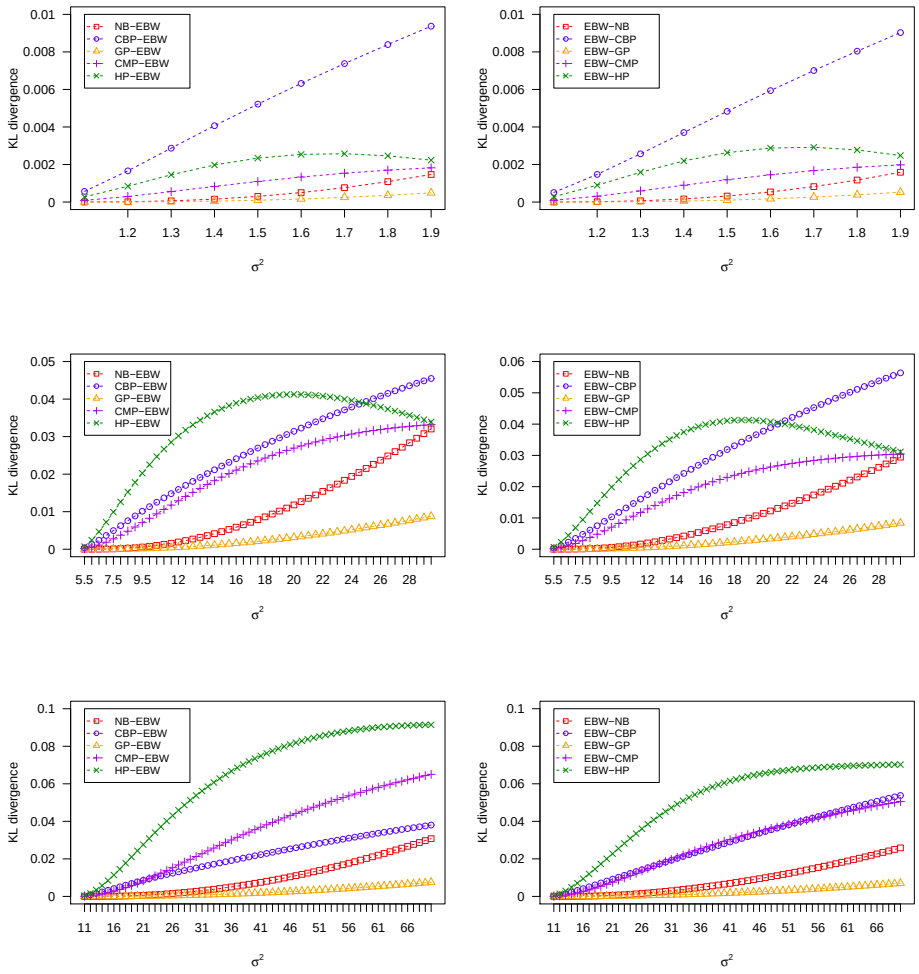

Figure 2. Kullback Leibler divergence between the NB, CBP, GP, CMP, HP and EBW distributions (and vice versa) in an overdispersed scenario. Rows 1-3 have $\mu = 1, \mu = 5$ and $\mu = 10$, respectively

and NB distributions. On the other hand, in an underdispersed scenario the HP distribution, which is very similar to the CMP distribution, is the closest one [21]. Nevertheless, these distances in relation to the EBW distribution are really small, which implies that the performance of these distributions is very similar.

5. Examples

In this section we use the EBW distribution to fit both over- and underdispersed real data and we compare this fit with those obtained from other discrete distributions.

5.1. Overdispersed data: Fire outbreaks by municipality in Andalusia (Spain)

We consider the variable X: number of fire outbreaks by municipality in the region of Andalusia (Spain). Data have been obtained from the Nature Databank in the Ministry of the Environment (Spain) and count the number of fire outbreaks from 2001 to 2014. A fire outbreak is defined as a wildfire whose

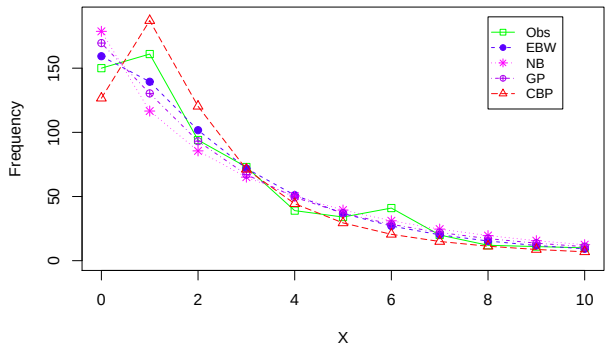

Figure 3. Observed and expected frequencies for data about fire outbreaks

total area is less than 1 hectare. Moreover, municipalities whose forest land is 0 have been removed from data. A description of these data appears in Table 3, which contains the mean, variance, AI, quartiles and maximum.

Table 3. Descriptive statistics for data in examples

	$\overline{x}$	s^2	AI	Q_1	Q_2	Q_3	Max
Fire outbreaks	3.53	28.97	8.21	1	2	4.75	56
Syllables -1	1.58	1.17	0.74	1	2	2	5

We will model these data by the following distributions: EBW, NB, GP, CBP, UGW, HP and CMP. AIC values, statistics and $p-$values corresponding to the χ^2-goodness of fit test are shown in Table 4. We can see that the best fit is that provided by the EWB distribution. The Wald test supports this statement since the null hypothesis $a = k$ cannot be rejected: the statistic value is $-2.1 \cdot 10^{-5}$ and the corresponding $p-$value is 1. With the likelihood ratio test (LRT) we come to the same conclusion ($LRT = 3.8 \cdot 10^{-9}$ and p-value $\simeq 1$).

Table 4. AIC values and χ^2-goodness of fit test for data about fire outbreaks

	EBW	NB	GP	CBP	UGW	HP	CMP
AIC	**3252.2**	3292.7	3263.3	3278.4	3254.2	3303.3	3303.0
χ^2	**24.06**	45.29	44.28	60.06	24.06	60.25	60.04
d.f.	15	15	15	16	14	14	14
$p-$value	**0.06**	0.00	0.00	0.00	0.04	0.00	0.00

Table 5 includes the observed and expected frequencies for the EWB, NB, GP and CBP fits and Figure 3 shows graphically the first of these frequencies. We can see that the only fitted distribution which has the modal value in 1 is the CBP, although the expected frequencies for values 0, 1 and 2 are further from the observed ones. From the EBW fit, the proportion of data variability due to randomness, liability and proneness is 10.2%, 33.6% and 56.2%, respectively. We can observe that randomness does not play a very important role with respect to the total variability of data and that the most important component is proneness, which refers to the particular conditions of each municipality in relation to the number of fire outbreaks, instead of other shared conditions.

Table 5. Observed and expected frequencies for data about fire outbreaks

		Expected			
X	*Observed*	*EBW*	*NB*	*GP*	*CBP*
0	150	159.31	178.69	169.56	126.62
1	161	139.42	116.60	130.34	187.08
2	94	101.69	85.56	93.07	120.30
3	73	71.87	65.12	67.13	71.24
4	39	51.04	50.41	49.41	44.41
5	34	36.79	39.45	37.09	29.43
6	41	27.00	31.08	28.32	20.55
7	20	20.17	24.61	21.94	14.97
8	12	15.33	19.56	17.21	11.28
9	11	11.83	15.59	13.64	8.74
10	10	9.26	12.45	10.91	6.93
11	4	7.34	9.96	8.80	5.60
12	6	5.89	7.99	7.14	8.44
13	4	8.68	6.41	5.83	
14	2		5.15	8.74	6.00
15	2	5.92	7.47		
16	8			6.00	6.24
17	7				
18	0	5.79	6.59		
19	1			5.77	
20	0				5.37
21	0	5.30			
22	1				
23-24	1		7.35	9.07	5.41
25-29	5				
30-44	1	7.36			5.12
45-56	3				6.276
		$\widehat{\alpha} = 2.749$ (0.162)	$\widehat{\theta} = 0.801$ (0.055)	$\widehat{\lambda} = 1.403$ (0.057)	$\widehat{b} = 1.486$ (0.088)
		$\widehat{\rho} = 3.139$ (0.317)	$\widehat{\mu} = 3.527$ (0.166)	$\widehat{\theta} = 0.602$ (0.019)	$\widehat{\gamma} = 1.495$ (0.122)

5.2. Underdispersed data: Turkish poem

We consider data about the word length (in terms of number of syllables) in the turkish poem *Gidisat* by Ercüment Behzat Lâv available in [32]. Following these authors, the count for 1 is treated as a count for 0, and in general the count for the response variable X is treated as $X - 1$, as though the data are generated by adding 1 to the distribution. These data exhibit underdispersion with a variance-mean ratio of 0.74 (see Table 3). Table 6 contains the parameter estimates, their standard errors (in parenthesis), the *AIC*, the observed and expected frequencies and the corresponding Pearson χ^2 test for each one of the models that copes with underdispersion, that is, *EBW*, *CTP*, *CMP* and *HP* (the *GP* distribution has been excluded because it is of finite range).

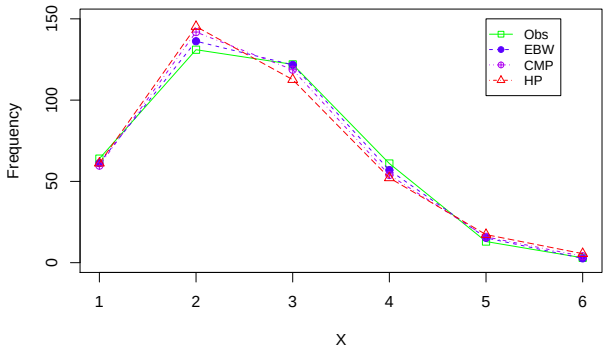

Figure 4. Observed and expected frequencies for data about the number of syllables of a Turkish poem

Table 6. Parameter estimates, standard errors (in parenthesis), observed and expected frequencies, AIC and χ^2 test for fits to data about the word length of a Turkish poem

		Expected			
X	*Observed*	*EBW*	*CTP*	*CMP*	*HP*
1	64	61.24	61.24	59.69	61.20
2	131	136.23	136.23	141.87	145.24
3	122	121.68	121.68	118.70	112.60
4	61	56.93	56.93	53.92	52.17
5	13	15.27	15.27	15.88	17.27
≥ 6	3	2.66	2.66	3.94	5.55
		$\widehat{\alpha} = -10.530$ (2.144)	$\widehat{a} = -10.530$ (2.158)	$\widehat{\lambda} = 2.377$ (0.276)	$\widehat{\gamma} = 0.485$ (0.099)
		$\widehat{\gamma} = 49.843$ (24.257)	$\widehat{b} = 0.001$ (14.254)	$\widehat{v} = 1.506$ (0.137)	$\widehat{\lambda} = 1.151$ (0.104)
			$\widehat{\gamma} = 49.843$ (24.416)		
	AIC	**1158.3**	1160.3	1160.7	1164.4
	$\chi^2 - statistic$	**1.000**	1.000	2.914	6.014
	$p - value$	**0.801**	0.606	0.405	0.111

CTP and *EBW* fits provide practically the same results. In fact, $b = 0$ using the Wald test ($z_{exp} = 2.3 \cdot 10^{-5}$ and $p - value \approx 1$) and the LRT ($\chi^2_{exp} = 0$ and $p - value = 1$). Observed and expected frequencies for each fit are represented in Figure 4 (the *CTP* distribution has been suppressed). Although the three fits are very similar and really good, the *EBW* distribution fit is the most accurate taking into account the expected frequencies.

6. Conclusions

The *EBW* distribution is a very flexible biparametric discrete distribution that allows for modelling a wide variety of over and underdispersed count datasets. There are other biparametric distributions that can also cope with over and underdispersion such as the *GP*, *CMP* or *HP* distributions, but the *EBW* distribution is more general since it presents fewer restrictions in the aggregation index range that it may describe and, moreover, its pmf and moments can be explicitly obtained in terms of the parameters.

In addition, when the first parameter of the *EBW* distribution is positive, the model is always overdispersed and the variance can be split into three components (randomness, liability and proneness) in

such a way that all the components are now well defined: the *EBW* distribution solves the indetermination problem of the *UGW* parameters. Taking into account this property, the *EBW* distribution is more adequate than other biparametric discrete distributions for modelling overdispersed data in which the variability is due more to individual internal factors than to external ones.

Furthermore, when the first parameter is a negative integer the *EBW* distribution has finite range and it is underdispersed. Something similar happens with the *GP* distribution that also has finite range but only in the underdispersed case, whereas the *EBW* distribution can also be underdispersed with infinite range.

Appendix. Obtaining the parameters in terms of μ and σ^2

For the *EBW* distribution there is a pair of solutions for α and γ from (4):

$$\alpha_1 = \frac{\mu^2 + \sqrt{\mu^4 + \mu(\sigma^2 - \mu)(\mu^2 + \sigma^2)}}{\sigma^2 - \mu}, \quad \gamma_1 = \frac{\alpha_1}{\mu} + 2\alpha_1 + 1 \tag{A1a}$$

$$\alpha_2 = \frac{\mu^2 - \sqrt{\mu^4 + \mu(\sigma^2 - \mu)(\mu^2 + \sigma^2)}}{\sigma^2 - \mu}, \quad \gamma_2 = \frac{\alpha_2}{\mu} + 2\alpha_2 + 1 \tag{A1b}$$

It can be shown that if the *EBW* distribution is overdispersed, $\alpha_1, \gamma_1 > 0$ and $\alpha_2 < 0$, but $\gamma_2 > 0$ if $\mu < 1$. If the *EBW* distribution is underdispersed, both α_1 and α_2 are negative, but:

- $\gamma_1 > 0$ when $\mu < 1$ and $\sigma^2 > \mu(1-\mu)$ or when $\mu \geq 1$ and $\sigma^2 > \frac{\mu - \mu^2 + \sqrt{\mu^3(\mu-1)}}{2}$
- $\gamma_2 > 0$ when $\mu < 1$ and $\sigma^2 > \mu(1-\mu)$.

As a consequence, if $\mu \geq 1$, the only possible solution is that given in (A1a) for both cases (over- and underdispersed).

Regarding the rest of the models, the expressions of their parameters in terms of μ and σ^2 can be seen in [21].

References

1. Irwin, J.O. The generalized Waring distribution applied to accident theory. *J. Royal Stat. Soc. Series A* **1968**, *131*, 205–225.
2. Xelakaki, E. Infinite divisibility, completeness and regression properties of the univariate generalized Waring distribution. *Ann. Inst. Stat. Math.* **1983**, *35*, 279–289.
3. Xelakaki, E. The univariate generalized Waring distribution in relation to accident theory: proneness, spells or contagion? *Biometrics* **1983**, *39*, 887–895. doi:10.2307/2531324.
4. Rodríguez-Avi, J.; Conde-Sánchez, A.; Sáez-Castillo, A.J.; Olmo-Jiménez, M.J. A new generalization of the Waring distribution. *Comput. Stat. Data Anal.* **2007**, *51*, 6138–6150. doi:10.1016/j.csda.2006.12.029.
5. Rodríguez-Avi, J.; Conde-Sánchez, A.; Sáez-Castillo, A.J.; Olmo-Jiménez, M.J.; Martínez-Rodríguez, A.M. A generalized Waring regression model for count data. *Comput. Stat. Data Anal.* **2009**, *53*, 3717–3725. doi:http://dx.doi.org/10.1016/j.csda.2009.03.013.
6. Hilbe, J.M. *Negative Binomial Regression*; Cambridge University Press, 2011.
7. Vílchez-López, S.; Sáez-Castillo, A.J.; Olmo-Jiménez, M.J. GWRM: An R Package for Identifying Sources of Variation in Overdispersed Count Data. *PLoS ONE 11(12): e0167570* **2016**, *11*.
8. Tesitelova, H. On the role of nouns in the lexical statistics. *Prague Stud. Math. Linguist.* **1967**, *2*, 121–131.
9. Ajiferuke, I. A probabilistic model for the distribution of authorships. *J. Amer. Soc. Inform. Sci.* **1991**, *42*, 279–289.
10. Levene, M.; Fenner, T.; Loizou, G.; Wheeldon, R. A stochastic model for the evolution of the web. *Comput. Netw.* **2002**, *39*, 277–287.

11. Xekalaki, E. The bivariate generalized Waring distribution and its application to Accident Theory. *J. Royal Stat. Soc. Series A* **1984**, *147*, 488–498. doi:10.2307/2981580.
12. Grunwaldm, G.K.; Bruce, S.L.; Jiang, L.; Strand, M.; Rabinovitch, N. A statistical model for under- or overdispersed clustered and longitudinal count data. *Biom. J.* **2011**, *53*, 578–594. doi:10.1002/bimj.201000076.
13. Peng, Y.; Lord, D.; Zou, Y. Applying the generalized Waring model for investigating sources of variance in motor vehicle crash analysis. *Accid. Anal. Prev.* **2014**, *73*, 20–26. doi:10.1016/j.aap.2014.07.031.
14. Ariza-López, F.J.; Rodríguez-Avi, J. Estimating the count of completeness errors in geographic data sets by means of a generalized Waring regression model. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 1394–1418. doi:10.1080/13658816.2015.1010536.
15. Huete-Morales María Dolores, M.D.; Marmolejo-Martín, J.A. The Waring Distribution as a Low-Frequency Prediction Model: A Study of Organic Livestock Farms in Andalusia. *Mathematics , Volume 8; doi:* **2020**, *8*, 2025. doi:10.3390/math8112025.
16. Panaretos, J.; Xekalaki, E. Extension of the Application of Conway-Maxwell-Poisson Models: Analyzing Traffic Crash Data Exhibiting Underdispersion. *Risk Analysis* **1986**, *4*, 313–318.
17. Johnson, N.L.; Kemp, A.W.; Kotz, S. *Univariate discrete distributions*, 3rd ed.; Wiley: New York, 2005.
18. Rodríguez-Avi, J.; Conde-Sánchez, A.; Sáez-Castillo, A.J.; Olmo-Jiménez, M.J. A triparametric discrete distribution with complex parameters. *Stat. Pap.* **2004**, *45*, 81–95. doi:10.1007/BF02778271.
19. Rodríguez-Avi, J.; Olmo-Jiménez, M.J. A regression model for overdispersed data without too many zeros. *Stat. Pap.* **2017**, *58*, 749–773. doi:10.1007/s00362-015-0724-9.
20. Jordan, C. *Calculus on finite differences*; Chelsea Publishing Company, 1965.
21. Olmo-Jiménez, M.J.; Rodríguez-Avi, J.; Cueva-López, V. A review of the CTP distribution: a comparison with other over- and underdispersed count data models. *J. Stat. Comput. Sim.* **2018**, *88*, 2684–2706. doi:10.1080/00949655.2018.1482897.
22. Rodríguez-Avi, J.; Conde-Sánchez, A.; Sáez-Castillo, A.J.; Olmo-Jiménez, M.J. A generalization of the Beta-Binomial distribution. *J. Roy. Statist. Soc. Ser. C* **2007**, *56*, 51–61. doi:10.1111/j.1467-9876.2007.00564.x.
23. Sibuya, M. Generalized hypergeometric, digamma and trigamma distributions. *Ann. Inst. Statist. Math.* **1979**, *31*, 373–390. doi:10.1007/BF02480295.
24. Sibuya, M.; Shimizu, R. Classification of the generalized hypergeometric family of distributions. *Keio Sci. Tech. Rep.* **1981**, *34*, 1–38.
25. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
26. Burnham, K.P.; Anderson, D.R. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*, 2nd ed.; Springer-Verlag: New York, 2002.
27. Consul, P.C. *Generalized Poisson Distributions: Properties and Applications*; Marcel Dekker: New York, 1989.
28. Joe, H.; Zhu, R. Generalized Poisson Distribution: the Property of Mixture of Poisson and Comparison with Negative Binomial Distribution. *Biom. J.* **2005**, *45*, 219–229.
29. Conway, R.W.; Maxwell, W.L. A queuing model with state dependent service rates. *J. Ind. Eng.* **1962**, *12*, 132–136.
30. Sellers, K.F.; Borle, S.; Shmueli, G. The COM-Poisson model for count data: a survey of methods and applications. *Appl. Stoch. Models Bus. Ind.* **2012**, *28*, 104–116. doi:10.1002/asmb.918.
31. Bardwell, G.E.; Crow, E.L. A two parameter family of hyper-Poisson distributions. *J. Am. Stat. Assoc.* **1964**, *54*, 133–141. doi:10.1080/01621459.1964.10480706.
32. Wimmer, G.; Köhler, R.; Grotjahn, R.; Altmann, G. Towards a Theory of Word Length Distribution. *J. Quant. Linguist.* **1994**, *1*, 98–106. doi:10.1080/09296179408590003.