Journal of Modern Applied Statistical Methods

Volume 23 | Issue 2 Article 3

The Performance of the Maximum Likelihood Estimator for the Bell Distribution for Count Data

David E. Giles
Professor Emeritus, Department of Economics, University of Victoria, Canada

Received: 29-10-2023; Revised: 30-11-2023; Accepted: 28-12-2023; Published: 06-01-2024.

Recommended Citation

David E. Giles (2024). The Performance of the Maximum Likelihood Estimator for the Bell Distribution for Count Data. Journal of Modern Applied Statistical Methods, 23(2), https://doi.org/10.56801/Jmasm.V23.i2.3

Journal of Modern Applied Statistical Methods
2024, Vol. 23, No. 2,
Doi: 10.56801/Jmasm.V23.i2.3

ISSN 1538 – 9472

The Performance of the Maximum Likelihood Estimator for the Bell Distribution for Count Data

David E. Giles
Department of Economics,
University of Victoria, Canada

The single-parameter "Bell distribution" for discrete data allows for over-dispersion in the data. The maximum likelihood estimator for its parameter is downward-biased in finite samples. We consider various methods for reducing this bias. A simulation study shows that these are effective and also lead to a small improvement in the mean squared error of the estimator. The Cox-Snell correction is the recommended. choice among the options that are considered.

Keywords: Discrete data, over-dispersion, maximum likelihood estimation and bias reduction.

1. Introduction

There are many statistical distributions that can be used to model integer-valued "count" data. The most commonly encountered one is the Poisson distribution, which has both advantages and disadvantages in practice. While the distribution enjoys the numerical simplicity of a single parameter, it is limited by the fact that this parameter is both the mean and variance. This renders the Poisson distribution inappropriate when the data are either over- or under-dispersed. Alternatives, such as the negative binomial distribution can deal with over-dispersion, but this is achieved at the "cost" of the probability mass function (p.m.f.) involving one or more additional parameters.

Recently, Castellares *et al.* (2018) introduced a new single-parameter discrete probability distribution based on the Bell numbers (Bell, 1934a, 1934b), which exhibits over-dispersion for all values of its parameter. Several authors have explored the use of this "Bell distribution", and its zero-inflated counterpart, in the context of regression analysis. For example, see Castellares *et al.* (2018), Lemonte *et al.* (2020), Abduljabbar and Algamal (2022), Abduljabbar *et al.* (2022), Shewa and Ugwowo (2022), and Ertan *et al.* (2023). While this application of the distribution is not our primary concern here, we present some results that relate to it. Our primary focus is on the Bell distribution itself.

The maximum likelihood (ML) estimator of the Bell parameter is readily obtained, although it cannot be expressed in a simple closed form. The ML estimator is also the method of moments estimator in this case. The sampling properties of this estimator in finite samples have not been examined systematically to date, and this is the objective of this paper. In the next section we derive the second-order bias of the ML estimator in question, and consider various bias-reduction techniques. We also show that the ML estimator of the distribution's mean is unbiased for all sample sizes, with implications for regression modelling. The results of a simulation experiment, described in section 3, illustrate the effectiveness of the various bias-correction methods under consideration.

2. Estimation issues

2.1 Maximum likelihood estimation

The p.m.f. for the Bell distribution is

$$p(y) = Pr.[Y = y] = \theta^y B_y exp\left(-e^{\theta} + 1\right)/y! \quad ; \quad y = 0, 1, 2, \ldots.. \quad ; \quad \theta > 0 \qquad (1)$$

where the Bell numbers are given by

$$B_j = \sum_{k=0}^{\infty} k^j \,/j!\ . \qquad (2)$$

The latter satisfy the recurrence relationship:

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k \quad ; \quad n = 1, 2, \ldots \ldots.. \qquad (3)$$

with $B_0 = B_1 = 1$. The Bell p.m.f. is uni-modal, with $E[Y] = \theta e^{\theta}$ and $Var.[Y] = \theta(1+\theta)e^{\theta} = (1+\theta)E[Y]$, implying that the distribution allows for a particular form of over-dispersion. Figure 1 illustrates the p.m.f. for various values of θ.

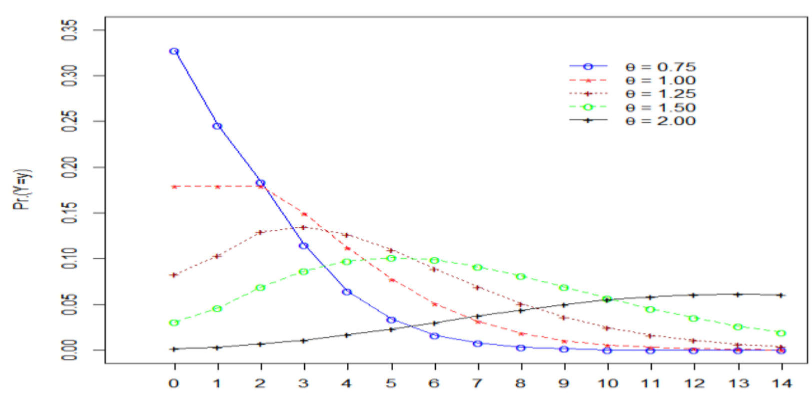

Figure 1. Bell Distribution p.m.f.

The Bell distribution is a member of the one-parameter exponential family. Given a sample of *n* independent observations, the associated log-likelihood function is

$$l(\theta) = n\bar{y}\log(\theta) - ne^{\theta} + c, \tag{4}$$

where *c* is a constant, and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Noting that

$$\frac{\partial l}{\partial \theta} = (n\bar{y}/\theta) - ne^{\theta}, \tag{5}$$

the ML estimator of θ is obtained as $\tilde{\theta} = W_0(\bar{y})$, where $W_0(.)$ is the principal branch of the Lambert W-function, which does not have a closed-form expression. It is simply the inverse function of $f(W) = We^{W}$, and is computed numerically. For further details, see Corless *et al*. (1996), for example.

Equating $E[Y]$ and $\bar{y}$ and solving for θ, we see immediately that $\tilde{\theta}$ is also the method of moments estimator of θ. Although $\tilde{\theta}$ is a consistent estimator of θ, it is biased, as we see below.

2.2 Bias reduction

First, consider the "corrective" approach to bias reduction developed by Cox and Snell (1968). Adopting the notation of Cordeiro and Klein (1994), to obtain an analytic expression for the bias of $\tilde{\theta}$ to to $O(n^{-1})$ we use the cumulants of the log-likelihood function. The second and third derivatives of the log-likelihood function are

$$\frac{\partial^2 l}{\partial \theta^2} = -\left(\frac{n\bar{y}}{\theta^2}\right) - ne^{\theta} \tag{6}$$

$$\frac{\partial^3 l}{\partial \theta^3} = (2n\bar{y}/\theta^3) - ne^{\theta} \quad , \tag{7}$$

and their expected values are

$$\kappa_{11} = E\left[\frac{\partial^2 l}{\partial \theta^2}\right] = -n(1+\theta)e^{\theta}/\theta \tag{8}$$

$$\kappa_{111} = E\left[\frac{\partial^3 l}{\partial \theta^3}\right] = ne^{\theta}(2-\theta^2/\theta^2. \tag{9}$$

Although the ML estimator does not have a closed form expression, its bias (to $O(n^{-1})$) can be obtained in terms of the following quantities:

$$\kappa_{11}^{(1)} = \frac{\partial}{\partial \theta}(\kappa_{11}) = -ne^{\theta}(\theta^2+\theta+1)/\theta^2$$

and

$$a_{11} = \kappa_{11}^{(1)} - \frac{\kappa_{111}}{2} = -ne^{\theta}(\theta+2)/(2\theta).$$

The results of Cordeiro and Klein (1994) imply that the bias of $\tilde{\theta}$ is

$$B(\tilde{\theta}) = \frac{a_{11}}{(\kappa_{11})^2} = -\frac{\theta(\theta+2)}{[2ne^{\theta}(1+\theta)^2]} + O(n^{-2}) \quad , \quad (10)$$

which is negative for all θ, and is plotted (for $n = 10$) in Figure 2. Equating the derivative of (10) with respect to θ to zero, and using the 'uniroot' function in R to solve the equation,

$$\theta^3 + 3\theta^2 + 2\theta - 2 = 0 \, , \quad (11)$$

we can locate the minimum of the bias function at $\theta = 0.5218$.

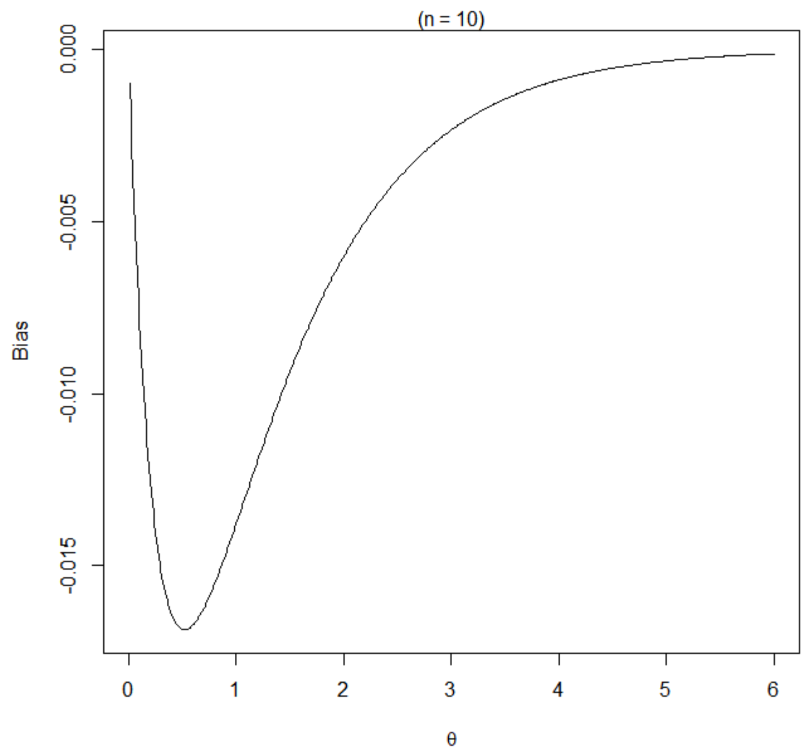

Figure 2. First-Order Bias of $\hat{\theta}$

Using the expression in (10), the traditional Cox-Snell "corrective" method of obtaining an estimator of θ that is unbiased to $O(n^{-2})$, is to construct the estimator,

$$\hat{\theta} = \tilde{\theta} - \tilde{B}(\tilde{\theta}) \, , \quad (12)$$

where $\tilde{B}(\tilde{\theta})$ is the bias expression in (10), evaluated at the original ML estimator, $\tilde{\theta}$. There are numerous applications of this methodology in the literature. For example, see the extensive references in Cordeiro and Cribari-Neto (2014).

Although the absolute value of the bias in Figure 2 is very small (and indeed, negligible for $\theta > 5$), the shape of the bias function warrants further consideration. As is noted by MacKinnon and Smith (1998) a corrective estimator of the form given in (12) is appropriate if the bias function is "flat" with respect to the parameter being estimated, and this is clearly not the case here. Godwin and Giles (2019) suggest a modified corrective estimator, in which the bias is evaluated (implicitly) using the bias-corrected estimator rather than the (biased) ML estimator. That is, they propose the estimator, θ^*, that is the solution to the expression.

$$\theta^* = \tilde{\theta} - B^*(\tilde{\theta}) \quad , \tag{13}$$

where $B^*(\tilde{\theta})$ is the bias expression in (10), evaluated at θ^*. Godwin and Giles provide several examples that justify this modified corrective approach.

In the case of the Bell distribution, this involves solving

$$\theta^* - \tilde{\theta} - \frac{\theta^*(\theta^*+2)}{[2ne^{\theta^*}(1+\theta^*)^2]} = 0 \tag{14}$$

for θ^*, which is easily done numerically.

An alternative to using a corrective procedure is to use the “preventive” approach to bias reduction proposed by Firth (1993) for ML estimators This involves adjusting the score function, and instead obtaining the solution to the equation.

$$\frac{\partial l}{\partial \theta} + \frac{a_{11}}{\kappa_{11}} = 0. \tag{15}$$

In the case of the Bell distribution, we need to find the estimator, $\breve{\theta}$, that solves

$$\frac{(n\bar{y} - n\theta e^{\theta})}{\theta} - \frac{\theta(\theta+2)e^{\theta}}{[2(\bar{y}+\theta^2 e^{\theta})]} = 0 \quad . \tag{16}$$

Again, this is easily achieved numerically.

2.3 Estimation of the mean

If interest centers on the mean of the distribution, rather than on the parameter θ itself, then it is easily established that the associated ML estimator is just the sample mean, which is unbiased. To see this, we invoke the invariance property of ML estimators, and use the result that $\mu = E[Y] = \theta e^{\theta}$.

So, the ML estimator of μ is

$$\tilde{\mu} = \tilde{\theta} e^{\tilde{\theta}} = W_0(\bar{y}) \exp\,(W_0(\bar{y})). \tag{17}$$

That is, $\tilde{\mu} = xe^{x}$, which implies that $x = W_0(\tilde{\mu})$. However, from (17), $x = W_0(\bar{y})$, and so $\tilde{\mu} = \bar{y}$.

This result can also be obtained by re-parameterizing the p.m.f. (and hence the log-likelihood function) in terms of μ, yielding

$$l(\mu) = n[1 - \exp(W_0(\mu)] + n\bar{y}\log\big(W_0(\mu)\big) + c', \tag{18}$$

where c' is a constant. Noting that $\frac{\partial W_0(\mu)}{\partial \mu} = W_0(\mu)/[\mu(1 + W_0(\mu)]$, the score function is

$$\frac{\partial l}{\partial \mu} = n[\bar{y} - W_0(\mu)\exp(W_0(\mu)]\,/[\mu(1 + W_0(\mu))]. \tag{19}$$

Equating (19) to zero and solving for μ confirms that $\tilde{\mu} = \bar{y}$.

The unbiasedness of the ML estimator for the Bell distribution's mean implies that the ML estimator for the coefficient vector in the Bell regression model proposed by Castellares *et al.* (2018) is also unbiased if a linear link function is used. This result need not hold for other choices of the link function. However, the simulation results reported by Castellares *et al*. (2018, p.178) suggest that the bias of the ML estimator is quite small (even for n = 50) when a logarithmic link function is used.

3. Simulation analysis

While bias reduction is of considerable interest, it is well known that in many situations this is achieved at the cost of an increase in the estimator's variance, or even its mean squared error (MSE). Accordingly, we have conducted a Monte Carlo simulation experiment to systematically explore both the (relative) bias and (relative) MSE of the ML estimator, and the various "bias-reduced" estimators, for the Bell distribution introduced in section 2. Various values of θ and sample sizes, n, have been considered.

The experiment involves N = 50,000 replications, with the Bell-distributed variates being generated using the 'rbell' function in the 'bellreg' package for R (Demarqui, *et al*., 2022). The 'lambertW0' function in the R package 'lamW' (Adler, 2023) was used to compute $\tilde{\theta}$; and the 'uniroot' function in R was used to solve the non-linear equations to obtain $\theta^*$and $\breve{\theta}$. The R code used in the experiment is available at https://github.com/DaveGiles1949/My-Documents.

Many of the other studies that have evaluated the Cox-Snell and Firth estimators for various distributions have also considered the use of the bootstrap as an alternative way of reducing the bias of the ML estimator. For example, see Cribari-Neto and Vasconcellos (2002), Xiao and Giles (2014), among others. The bootstrap bias-corrected estimator of θ is obtained as $\ddot{\theta} = 2\tilde{\theta} - (\frac{1}{N_B})\left[\sum_{j=1}^{N_B} \tilde{\theta}_{(j)}\right]$, where $\tilde{\theta}_{(j)}$ is the ML estimator of θ obtained from the j^{th} of the N_B (= 999) bootstrap samples. See Efron (1982, p.33). This estimator is also unbiased to $O(n^{-2})$, but in practice it this may come at the expense of increased variance. Moreover, the bootstrap frequently over-corrects the first-order bias of ML estimators, as is found by Cribari-Neto and Vasconcellos (2002) and Schwartz *et al*. (2013), for example. In overall terms, Giles *et al*. (2013, 2016) find that the parametric bootstrap estimator is inferior to the Cox-Snell estimator for the two-parameter Lomax model and for the generalized Pareto distribution; and Xiao and Giles (2014) come to a similar conclusion in the case of the generalize Rayleigh family of distributions.

In our experiment, for each (θ, n) combination the simulated percentage bias and percentage MSE were computed for each estimator, as follows:

$$\%Bias(\tilde{\theta}) = 100 * \left[\frac{1}{N}\sum_{j=1}^{N}(\tilde{\theta}_j) - \theta\right]/\theta$$

$$\%MSE(\tilde{\theta}) = 100 * \left[\frac{1}{N}\sum_{j=1}^{N}(\tilde{\theta}_j - \theta)^2\right]/\theta^2 ,$$

and similarly for $\hat{\theta}$, θ^*, $\breve{\theta}$, and $\ddot{\theta}$.

The simulation results appear in Table 1, where the percentage MSE values are in parentheses below the corresponding percentage biases. We see that the percentage biases and MSE's decline as the sample size increases, reflecting the consistency of al of the estimators. As expected from Figure 2, the percentage bias of the (uncorrected) ML estimator becomes negligible, and the corresponding percentage MSE values decrease, as the population value of θ increases. (Although not shown in Table 1, when $\theta = 4$ and n = 10, the % bias and % MSE of $\tilde{\theta}$ are -0.017 and 0.009 respectively. The corresponding values for each of the bias-adjusted estimators are 0.005% and 0.09%.)

Table 1. Simulated %Bias (%MSE) for estimators of θ

$\boldsymbol{\theta = 0.75}$					
$\boldsymbol{n}$	$\tilde{\boldsymbol{\theta}}$	$\hat{\boldsymbol{\theta}}$	$\breve{\boldsymbol{\theta}}$	$\boldsymbol{\theta}^*$	$\ddot{\boldsymbol{\theta}}$
10	-2.194	-0.089	-0.074	-0.101	-0.138
	(3.779)	(3.693)	(3.689)	(3.690)	(3.686)
15	-1.484	-0.080	-0.070	-0.082	-0.150
	(2.477)	(2.436)	(2.435)	(2.435)	(2.424)
25	-0.859	-0.012	-0.010	-0.015	-0.024
	(1.455)	(1.440)	(1.440)	(1.440)	(1.443)
50	-0.419	0.005	0.005	0.005	-0.082
	(0.723)	(0.720)	(0.719)	(0.719)	(0.731)
75	-0.279	0.003	0.004	0.004	-0.015
	(0.482)	(0.481)	(0.481)	(0.481)	(0.481)
100	-0.190	0.022	0.023	0.023	-0.010
	(0.359)	(0.358)	(0.358)	(0.358)	(0.360)
150	-0.114	0.027	0.027	0.027	-0.014
	(0.239)	(0.239)	(0.239)	(0.239)	(0.239)
250	-0.085	-0.001	-0.000	-0.001	0.003
	(0.144)	(0.144)	(0.144)	(0.144)	(0.144)
500	-0.025	0.017	0.017	0.017	-0.006
	(0.072)	(0.072)	(0.072)	(0.072)	(0.072)

$\boldsymbol{\theta = 1.0}$					
$\boldsymbol{n}$	$\tilde{\boldsymbol{\theta}}$	$\hat{\boldsymbol{\theta}}$	$\breve{\boldsymbol{\theta}}$	$\boldsymbol{\theta}^*$	$\ddot{\boldsymbol{\theta}}$
10	-1.494	-0.105	-0.102	-0.116	-0.168
	(1.955)	(1.899)	(1.898)	(1.899)	(1.890)

15	-0.919	0.005	0.006	-0.000	-0.129
	(1.268)	(1.245)	(1.245)	(1.245)	(1.249)
25	-0.552	0.002	0.002	0.001	-0.009
	0.750)	(0.741)	(0.741)	(0.741)	(0.745)
50	-0.265	0.012	0.011	0.011	-0.034
	(0.372)	(0.370)	(0.370)	(0.370)	(0.372)
75	-0.157	0.027	0.027	0.027	-0.028
	(0.246)	(0.245)	(0.245)	(0.245)	(0.246)
100	-0.109	0.029	0.029	0.029	0.008
	0.184)	(0.184)	(0.184)	(0.184)	0.184)
150	-0.079	0.013	0.012	0.013	-0.006
	(0.123)	(0.122)	(0.122)	(0.122)	(0.124)
250	-0.057	-0.002	-0.002	-0.002	-0.005
	(0.073)	(0.073)	(0.073)	(0.073)	(0.073)
500	-0.015	0.013	0.012	0.012	-0.017
	(0.037)	(0.036)	(0.036)	(0.036)	(0.037)

$\boldsymbol{\theta = 1.25}$					
$\boldsymbol{n}$	$\boldsymbol{\tilde{\theta}}$	$\boldsymbol{\hat{\theta}}$	$\boldsymbol{\breve{\theta}}$	$\boldsymbol{\theta^*}$	$\boldsymbol{\ddot{\theta}}$
10	-0.951	-0.021	-0.022	-0.028	-0.124
	(1.060)	(1.032)	(1.032)	(1.032)	(1.040)
15	-0.623	-0.005	-0.006	-0.008	-0.032
	(0.695)	(0.683)	(0.683)	(0.683)	(0.688)
25	-0.367	0.003	0.003	0.002	0.055
	(0.412)	(0.408)	(0.408)	(0.408)	(0.411)
50	-0.152	0.032	0.032	0.032	-0.009
	(0.203)	(0.202)	(0.202)	(0.202)	(0.204)
75	-0.088	0.034	0.034	0.034	-0.029
	(0.136)	(0.135)	(0.135)	(0.135)	(0.136)
100	-0.063	0.029	0.029	0.029	0.011
	(0.102)	(0.102)	(0.102)	(0.102)	(0.102)
150	-0.046	0.016	0.016	0.016	0.010
	(0.067)	(0.067)	(0.067)	(0.067)	(0.068)
250	-0.023	0.013	0.013	0.013	0.004
	(0.041)	(0.040)	(0.040)	(0.040)	(0.041)
500	-0.007	0.011	0.011	0.011	0.005
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)

$\boldsymbol{\theta = 1.5}$					
$\boldsymbol{n}$	$\boldsymbol{\tilde{\theta}}$	$\boldsymbol{\hat{\theta}}$	$\boldsymbol{\breve{\theta}}$	$\boldsymbol{\theta^*}$	$\boldsymbol{\ddot{\theta}}$
10	-0.640	-0.008	-0.009	-0.012	-0.020
	(0.616)	(0.602)	(0.602)	(0.602)	(0.610)
15	-0.417	0.003	0.002	0.001	-0.028
	(0.405)	(0.399)	(0.399)	(0.399)	(0.405)
25	-0.241	0.010	0.009	0.009	-0.002
	(0.240)	(0.238)	(0.238)	(0.238)	(0.239)

50	-0.117	0.009	0.008	0.009	-0.004
	(0.119)	(0.119)	(0.119)	(0.119)	(0.118)
75	-0.073	0.010	0.009	0.010	0.004
	(0.079)	(0.079)	(0.079)	(0.079)	(0.080)
100	-0.006	0.002	0.001	0.002	0.007
	(0.060)	(0.059)	(0.059)	(0.059)	(0.060)
150	-0.038	0.004	0.003	0.004	-0.000
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
250	-0.014	0.011	0.011	0.011	-0.010
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
500	0.003	0.009	0.008	0.009	0.005
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)

$\boldsymbol{\theta = 2.0}$					
$\boldsymbol{n}$	$\tilde{\boldsymbol{\theta}}$	$\hat{\boldsymbol{\theta}}$	$\breve{\boldsymbol{\theta}}$	$\boldsymbol{\theta}^*$	$\ddot{\boldsymbol{\theta}}$
10	-0.292	0.012	0.010	0.010	-0.011
	(0.278)	(0.224)	(0.224)	(0.224)	(0.230)
15	-0.185	0.017	0.016	0.016	-0.025
	(0.149)	(0.148)	(0.148)	(0.148)	(0.151)
25	-0.110	0.011	0.009	0.010	-0.003
	(0.090)	(0.089)	(0.089)	(0.089)	(0.089)
50	-0.038	-0.008	-0.007	-0.008	-0.015
	(0.019)	(0.019)	(0.019)	(0.019)	(0.045)
75	-0.045	-0.005	-0.006	-0.005	-0.010
	(0.030)	(0.030)	(0.030)	(0.030)	(0030)
100	-0.016	-0.001	-0.000	-0.001	-0.010
	(0.009)	(0.009)	(0.009)	(0.009)	(0.023)
150	-0.004	0.008	0.007	0.008	0.009
	(0.009)	(0.009)	(0.009)	(0.009)	(0.015)
250	-0.004	0.008	0.007	0.008	-0.002
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)

$\boldsymbol{\theta = 3.0}$					
$\boldsymbol{n}$	$\tilde{\boldsymbol{\theta}}$	$\hat{\boldsymbol{\theta}}$	$\breve{\boldsymbol{\theta}}$	$\boldsymbol{\theta}^*$	$\ddot{\boldsymbol{\theta}}$
10	-0.086	-0.009	-0.009	-0.009	-0.008
	(0.043)	(0.042)	0.042)	(0.042)	(0.042)
15	-0.046	0.006	0.006	0.006	-0.003
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
25	-0.030	0.001	0.001	0.001	0.004
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
50	-0.013	0.002	0.002	0.002	0.003
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
75	-0.006	0.004	0.004	0.004	0.001
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)

100	-0.005	0.002	0.002	0.002	0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
150	-0.005	-0.000	-0.000	-0.000	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
250	-0.004	-0.001	-0.001	-0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

In terms of bias, the performance of the bootstrap bias-corrected estimator is generally inferior to that of the other modified ML estimators for small *n*, especially when the true value of the parameter is also small. For moderate sample sizes, its performance in terms of bias is “mixed”. In all cases considered, the percentage MSE of the boostrap estimator is almost the same as that of its analytical competitors.

A key feature of the results is that the three analytical bias-reduction methods essentially perform equally well, and in addition they all reduce the percentage MSE of the original ML estimator to a similar (slight) degree. As the Cox-Snell corrective procedure has an obvious computational advantage over the methods proposed by Firth and by Godwin and Giles (and the bootstrap estimator) it may be preferred in practice.

4. Conclusions

The Bell distribution for discrete “count” data has the advantage of allowing for over-dispersion while being based on just a sole parameter. This sets it apart from competitors such as the negative binomial distribution. We have shown that while the ML estimator for the Bell parameter is downward-biased in small samples, this bias can be reduced substantially by using the simple Cox-Snell “corrective approach, while simultaneously reducing the MSE of the estimator slightly. Other analytical and bootstrap approaches to reducing the bias are also very effective, but are slightly more burdensome, computationally.

Our results support the use of the Bell distribution, even for relatively small samples. Moreover, as the ML estimator of the distribution’s mean is exactly unbiased, this adds support for the Bell regression that has been discussed by Castellares *et al*. (2018) and others. if a linear link function is employed.

References

Abduljabbar, L. A. & Z. Y. Algamal, 2022. Jackknifed K-L estimator in Bell regression model. Mathematical Statistician and Engineering Applications, 71, 267-278.

Adler, A., 2023. Package ‘lamW’, https://github.com/aadler/lamW.

Algamal,Z. Y. A. F. Lukman, M. R. Abonazel, & F. A. Awwad, 2022. Performance of the ridge and Liu estimators on the zero-inflated Bell regression model. Journal of Mathematics, Article ID 9503460.

Ali, E., M. L. Diop, & A. Diop, 2022. Statistical inference in a zero-inflated Bell regression model. Mathematical Methods of Statistics, 31, 91-104.

Bell, E. T., 1934a. Exponential polynomials. Annals of Mathematics, 35, 258-277.

Bell, E. T., 1934b. Exponential numbers. American Mathematical Monthly, 41, 411-419.

Castellares, F., S. L. P. Ferrari, & A. J. Lemonte, 2018. On the Bell distribution and its associated regression model for count data. Applied Mathematical Modelling, 56, 172-185.

Cordeiro, G. M. & F. Cribari-Neto, 2014. An Introduction to Bartlett Correction and Bias Reduction. Springer, Berlin.

Cordeiro, G. M. & R. Klein, 1994. Bias correction in ARMA models. Statistics and Probability Letters, 19, 169-176.

Corless, R. M., G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, & D. E. Knuth, 1996. On the Lambert W function. Advances in Computational Mathematics, 5. 329-359.

Cox, D. R. & E. J. Snell, 1968, A general definition of residuals. Journal of the Royal Statistical Society, B, 30, 248-275.

Cribari-Neto, F. & K. L. P. Vasconcellos, 2002. Nearly unbiased maximum likelihood estimation for the beta distribution. Journal of Statistical Computation and Simulation, 72, 107-118.

Demarqui, F., M. Prates, & F. Caseres, 2022. Package 'bellreg', https://github.com/fndemarqui/bellreg .

Efron, B., 1982. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Ertan, E. Z., Y. Algamal, A. Erkoç, & K U. Akay, 2023. A new improvement Liu-type estimator for the Bell regression model. Communications in Statistics – Simulation and Computation, in press.

Firth, D.,1993, Bias reduction for maximum likelihood estimates. Biometrika, 80, 27-38.

Giles, D. E., 2012. Bias reduction for the maximum likelihood estimator of the parameters in the half-logistic distribution. Communications in Statistics – Theory & Methods, 41, 212-222.

Giles, D. E., H. Feng, & R. T. Godwin, 2013. On the bias of the maximum likelihood estimator for the two- parameter Lomax distribution. Communications in Statistics – Theory and Methods, 42, 1934-1950.

Giles, D. E. H. Feng, & R. T. Godwin, 2016. Bias-corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. Communications in Statistics – Theory and Methods, 45, 2465-2483.

Godwin, R. F. & D. E. Giles, 2019. Improved analytic bias correction for maximum likelihood estimators. Communications in Statistics – Simulation and Computation, 48, 15-26.

Lemonte, A. J., G. Moreno-Arenas, & F. Castellares, 2020. Zero-inflated Bell regression models for count data. Journal of Applied Statistics, 47, 265-286.

MacKinnon, J. G. & Smith, A. A., 1998. Approximate bias correction in econometrics. Journal of Econometrics, 85, 205–230.

Majid. A, M. Amin, & M. N. Akram, 2022. On the Liu estimation of Bell regression model in the presence of multicollinearity. Journal of Statistical Computation and Simulation, 92, 262-282.

Schwartz, J., R. T. Godwin, & D. E. Giles, 2013. Improved maximum likelihood estimation of the shape parameter in the Nakagami distribution. Journal of Statistical Computation and Simulation, 83, 434-445.

Shewa, G. A. & F. I. Ugwowo, 2022. Combating the multicollinearity in Bell regression model: Simulation and application. Journal of the Nigerian Society of Physical Sciences, 4, 713.

Xiao, L. & D. E. Giles, 2014. Bias reduction for the maximum likelihood estimator of the generalized Rayleigh family of distributions. Communications in Statistics – Theory and Methods, 43, 1778-1792.