REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 425–442

THE MODIFIED BOREL–TANNER (MBT) REGRESSION MODEL

Authors: Emilio Gómez-Déniz
– Department of Quantitative Methods and TIDES Institute, University of Las Palmas de Gran Canaria, Spain
emilio.gomez-deniz@ulpgc.es

Francisco J. Vázquez-Polo
– Department of Quantitative Methods and TIDES Institute, University of Las Palmas de Gran Canaria, Spain
francisco.vazquezpolo@ulpgc.es

Victoriano García
– Department of Statistics and Operational Research, University of Cádiz, Spain
victoriano.garcia@uca.es

Received: July 2014 Revised: September 2015 Accepted: September 2015

Abstract:

- A new one-parameter family of discrete distributions is presented. It has some advantages against the Poisson distribution as a suitable model for modelling data with a high frequencies of zeros and showing over-dispersion (variance larger than the mean). The distribution is obtained from a simple modification of the Borel–Tanner distribution, which has not received attention from the statistical community in the past. We also propose a generalized regression model which can be used for a count dependent variable, when the above features are observed, as an alternative to the well-known Poisson regression model, among others. Maximum likelihood estimation is investigated and illustrated with an example of interrelation between fatalities in trucks accidents on American roads and some covariates considered.

Key-Words:

- *Borel–Tanner distribution; covariate; estimation.*

AMS Subject Classification:

- 62E99, 62P25.

1. INTRODUCTION

The Borel–Tanner distribution is a discrete distribution proposed more than fifty years ago in queueing theory to model the probability distribution of the number of customers served in a queuing line with Poisson input and a constant service time, given that the length of the queue at the initial time is r. As far as we know, this distribution has not received much attention for the statistical community. The probability function of the Borel–Tanner distribution ([11]) is given by

$$\Pr(Y{=}y) \,=\, A(y,r)\, e^{-\alpha y}\, \alpha^{y-r}\,, \qquad y = r, r+1, \ldots\,, \tag{1.1}$$

where $\alpha > 0$ and r is a positive integer and where

$$A(y,r) \,=\, \frac{r}{(y-r)!}\, y^{y-r-1}\,.$$

Equivalently, [10] rewritten expression (1.1) as

$$\Pr(Y{=}y) \,=\, B(y,r)\, \frac{\alpha^{y-r}}{(1+\alpha)^{2y-r}}\,, \qquad y = r, r+1, \ldots\,, \tag{1.2}$$

where

$$B(y,r) \,=\, \frac{r}{y} \binom{2y-r-1}{y-1}\,.$$

In this paper we focus on the distribution with probability distribution given in (1.2) using a modified version of this probability distribution with support in $0, 1, \ldots$, suitable for modelling data with a high frequencies of zero and showing over-dispersion phenomena: the variance is larger than the mean.

The distribution proposed here has some advantages against some other well-known distributions as a suitable model for modelling data with a high frequencies of zeros and showing over-dispersion phenomena. We also propose a generalized regression model which can be used for a count dependent variable, when the above features are observed. Maximum likelihood estimation is investigated and illustrated with an example involving emergency room visits to hospital.

The applicability of the model is shown by fitting the number of deaths in truck accidents (fatalities) on American roads, with different explanatory covariates from real data used by [17]. The provided real data examples show that the model works reasonably well, and this assessment is confirmed by the comparison to the Poisson and negative binomial distributions.

The contents of the paper are as follows. In section 2 we present the modified version of the Borel–Tanner distribution proposed here. Some properties of the distribution are also shown, including the mean, variance and the cumulative distribution function. Some methods of estimation are developed in section 3. The regression model is developed in section 4. An application with real data is shown in section 4 and conclusions in the last section.

2. THE MODIFIED BOREL–TANNER DISTRIBUTION (MBT)

In this section we propose a modified version of the Borel–Tanner distribution given in (1.2) which has support in the positive integer numbers including the zero value. Firstly, consider $r=1$ and $X=Y-1$, then it is a simple exercise to see that the resulting shifted distribution has its probability function given by

$$\Pr(X=x) = \frac{\Gamma(2x+1)}{\Gamma(x+2)\,\Gamma(x+1)}\,\frac{\alpha^x}{(1+\alpha)^{2x+1}}\,, \qquad x=0,1,\dots\,, \tag{2.1}$$

being $0<\alpha<1$. It has to be pointed out that in the original paper of [10] any parameter value $\alpha>0$ is allowed. Nevertheless, a simple algebra shows that it is not true and the feasible set of this parameter is actually $0<\alpha<1$. This distribution can be easily written as

$$\Pr(X=x) = C_x\,\frac{\alpha^x}{(1+\alpha)^{2x+1}}\,, \qquad x=0,1,\dots\,,$$

where

$$C_x = \frac{1}{x+1}\binom{2x}{x}$$

are the Catalan numbers (see [13], p. 13 and [18]. In the sequel, when a random variable X follows the probability mass function (2.1) we will denote $X\sim\mathrm{MBT}(\alpha)$.

Since probability function (2.1) can be written as

$$\Pr(X=x) = C_x\exp\bigl[\lambda\cdot x - A(\lambda)\bigr]\,,$$

where

$$\lambda = \log\frac{\alpha}{(1+\alpha)^2}\,,$$

and

$$A(\lambda) = \log\left(\frac{1-\sqrt{1-4e^{\lambda}}}{2e^{\lambda}}\right) = \log(1+\alpha)\,,$$

the modified Borel–Tanner distribution proposed here is a member of the natural exponential family of distributions. Furthermore, probability function (2.1) can also be rewritten as

$$\Pr(X=x) = \frac{C_x}{1+\alpha}\left[\frac{\alpha}{(1+\alpha)^2}\right]^x.$$

Therefore, the modified Borel–Tanner distribution belongs to the class of power series distribution (see [13], p. 75) which contains for instance Bernoulli, binomial, geometric, negative binomial, Poisson and logarithmic series distributions.

On the other hand, [3] discussed discrete probability density functions $\Pr(X{=}x;\alpha)$ which obey the following relation for some functions B and D: if there exist B and D such that

$$\frac{d\Pr(X{=}x;\alpha)}{d\alpha} = B(\alpha)\left[x - D(\alpha)\right]\Pr(X{=}x;\alpha)\,, \tag{2.2}$$

then the mean μ coincides with $D(\alpha)$ and $\mu_2 = (d\mu/d\alpha)(1/B(\alpha))$ is the variance. Also, in that case

$$\mu_i = \mu_2\left[\frac{d\mu_{i-1}}{d\alpha}\,\frac{1}{d\mu/d\alpha} + (i-1)\,\mu_{i-2}\right], \qquad i = 2, 3, \ldots\,,$$

where μ_i is the i-th moment about the mean, which depends on α. Note that $\mu_0 = \mu$.

Now, observe that the MBT(α) distribution verifies (2.2) considering

$$B(\alpha) = \frac{1-\alpha}{\alpha\,(1+\alpha)}\,, \qquad D(\alpha) = \frac{\alpha}{1-\alpha}\,.$$

Then, the mean and the variance of the random variable following the probability function (2.1) are given by

$$E(X) = \frac{\alpha}{1-\alpha} \tag{2.3}$$

and

$$var(X) = \frac{\alpha\,(1+\alpha)}{(1-\alpha)^3}\,, \tag{2.4}$$

respectively. The previous expression for the mean of a MBT(α) distributed variable allows to write its probability mass function (pmf) as

$$\Pr(X{=}x) = C_x \cdot \frac{\theta^x (1+\theta)^{1+x}}{(1+2\theta)^{1+2x}}\,,$$

where $\theta = E(X) = \dfrac{\alpha}{1-\alpha}$.

Since

$$\frac{var(X)}{E(X)} = 1 + \frac{\alpha\,(3-\alpha)}{(1-\alpha)^2} > 1$$

we conclude that the distribution is overdispersed. Note that the proposed distribution is zero-inflated; that is, its proportion of 0's is greater than the proportion of 0's of a Poisson variate with the same mean. To see this we observe that the zero-inflated index (see [19]) is $z_i = 1 - \frac{1-\alpha}{\alpha}\log(1+\alpha)$, which results greater than zero.

Additionally, the probability generating function is given by

$$G_X(z) = \frac{1+\alpha}{2\,\alpha\,z}\left[1 - \frac{\sqrt{1+\alpha\,(\alpha-4\,z+2)}}{1+\alpha}\right], \qquad |z|<1\,.$$

The cumulative distribution function of a random variable following the probability function given in (2.1) is given by

$$\text{(2.5)} \qquad \Pr(X\le x) = 1 - \frac{\Gamma\!\left(x+\frac{3}{2}\right)(4\alpha)^{x+1}}{\Gamma(x+3)\,\sqrt{\pi}\,(1+\alpha)^{2x+3}}\;{}_2F_1\!\left(1,\,x+\frac{3}{2};\,x+3;\,\frac{4\alpha}{(1+\alpha)^2}\right),$$

where ${}_2F_1$ is the hypergeometric function given by

$${}_2F_1(a,b;c;z) = \sum_{k=0}^{\infty}\frac{\Gamma(a+k)\,\Gamma(b+k)\,\Gamma(c)}{\Gamma(a)\,\Gamma(b)\,\Gamma(c+k)}\,\frac{z^k}{k!}\,.$$

See the Appendix Section for details about this assert.

Some additional details about the hypergeometric function can be found in [18]. From (2.5) we get the survival function, $\Pr(X\ge x)$, and the failure or hazard rate can be easily obtained using (2.5) and (2.1).

Finally, observe that the probabilities can be computed from the recursion

$$\Pr(X=x) = \frac{2\alpha}{(1+\alpha)^2}\,\frac{2x-1}{x+1}\,\Pr(X=x-1)\,, \qquad x=1,2,\ldots\,,$$

being $\Pr(X=0) = \dfrac{1}{1+\alpha}$.

Since

$$\frac{\Pr(X=x)}{\Pr(X=x-1)} - \frac{\Pr(X=x+1)}{\Pr(X=x)} = \frac{-6\alpha}{(1+\alpha)^2}\,\frac{1}{(x+1)\,(x+2)} < 0\,,$$

we have that the distribution is log-convex (infinitely divisible) and has decreasing failure rate (DFR). See [9] and [22] for details. The fact that $\Pr(X=x)/\Pr(X=x-1)$, $x=1,2,\ldots$, forms a monotone increasing sequence requires that $\Pr(X=x)$ be a decreasing sequence (see [12], p. 75). Therefore, the distribution is unimodal with modal value on zero. An overview of Figure 1 confirms this feature and that the shown plotting lines in the graph are similar to the ones corresponding to distributions of Poisson with expected value lower than 1.

Moreover, as any infinitely divisible distribution defined on non-negative integers is a compound Poisson distribution (see Proposition 9 in [15], we conclude that the probability function given in (2.1) is a compound Poisson distribution.

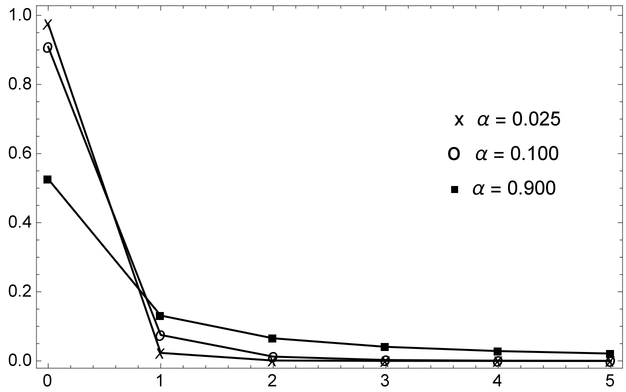

Figure 1: Some appearances (polygons) of the probability mass function for different values of the parameter α.

Furthermore, the infinitely divisible distribution plays an important role in many areas of statistics, for example, in stochastic processes and in actuarial statistics. When a distribution G is infinitely divisible then, for any integer $x \geq 2$, there exists a distribution G_x such that G is the x-fold convolution of G_x, namely, $G = G_x^{*x}$.

Since the new distribution is infinitely divisible, a lower bound for the variance can be obtained (see [12], p. 75), which is given by

$$var(X) \geq \frac{\Pr(X=1)}{\Pr(X=0)} = \frac{\alpha}{(1+\alpha)^2} .$$

3. INFERENCE FOR MBT DISTRIBUTION

In this section, different methods of estimation of the parameter of the distribution are studied.

Using (2.3) it is also simple to see that the estimator of α is given by

$$\widehat{\alpha}_1 = \frac{\bar{X}}{1+\bar{X}} , \tag{3.1}$$

where $\bar{X}$ is the sample mean.

An alternative to the method of moments is the method based on the zeros frequency. This method tends to work well only when the mode of the distribution is at zero and its proportion of zeros is relatively high ([2]). In this case we need only one equation in order to estimate the parameter of the distribution. It is straightforward obtaining an estimate for α based on the observed proportion of zeros, denoted by $\tilde{p}_0$, as

$$\widehat{\alpha}_2 = \frac{1-\tilde{p}_0}{\tilde{p}_0} \,.$$

For each of sample sizes $n = 100$ and $n = 1000$, and for $\alpha = 0.1\,[0.1]\,0.9$, 4000 samples have been simulated, obtaining the estimates mean and squared error from both methods (Table 1). In both of them, the experimental bias is higher when α takes its lower values.

Table 1: MME (equivalently, MLE) and zero proportion estimate based on n simulations from a $MBT(\alpha)$.

α		$n=100$		$n=1000$	
		$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$
.1	mean	.250	.385	.2514	.379
	S.D	.034	.087	.011	.027
.2	mean	.308	.435	.310	.429
	S.D	.041	.096	.013	.030
.3	mean	.372	.489	.376	.4826
	S.D	.048	.106	.015	.032
.4	mean	.444	.545	.449	.539
	S.D	.053	.114	.017	.036
.5	mean	.521	.607	.529	.602
	S.D	.056	.126	.018	.039
.6	mean	.606	.676	.614	.668
	S.D	.059	.139	.018	.043
.7	mean	.696	.755	.706	.741
	S.D	.059	.157	.018	.047
.8	mean	.789	.838	.801	.821
	S.D	.055	.174	.016	.052
.9	mean	.883	.924	.899	.907
	S.D	.047	.186	.013	.057

Finally, the MLE are easy to derive since the MBT model belongs to the exponential family. Let now $\underline{x} = (x_1, x_2, ..., x_n)$ be a random sample obtained from model (2.1), then the log-likelihood function is proportional to

$$\ell(\alpha) \;\propto\; n\Big[\bar{x}\log\alpha - (2\bar{x}+1)\log(1+\alpha)\Big] \,.$$

The likelihood equation obtained from (3) is given by

$$\frac{\partial\ell(\alpha)}{\partial\alpha} \;=\; \frac{\bar{x}}{\alpha} - \frac{2\bar{x}+1}{1+\alpha} \;=\; 0 \,,$$

from which we obtain the maximum likelihood estimator of α given again by (3.1). Therefore, as in the Poisson distribution the moment estimator coincides with the maximum likelihood estimator. Additionally, the maximum likelihood estimator $\widehat{\alpha}$ of α is unique for all n.

Proposition 3.1. *The unique maximum likelihood estimator $\widehat{\alpha}$ of α is consistent and asymptotically normal and therefore*

$$\sqrt{n}\,(\widehat{\alpha}-\alpha) \xrightarrow{d} N\big(0, I^{-1}(\alpha)\big)\,,$$

where $I(\alpha)$ is the Fisher's information about α.

Proof: See Appendix. □

By using Corollary 3.11 in [16], p. 450, we conclude that the maximum likelihood estimator of α is asymptotically efficient.

4. THE MBT REGRESSION MODEL

The Poisson regression model has been extensively used as a benchmark for the analysis of discrete count models, together with some other models such as the negative binomial regression, Poisson-inverse Gaussian regression, and some other including special functions as hypergeometric, Kummer confluent, etc. when the endogenous variable take only nonnegative integer values. In practice, count data often display over-dispersion and therefore the Poisson regression model faults to provide an appropriate fit to the data. In this section we provide a regression model based on the use of the modified Borel–Tanner distribution presented in the previous sections of this work. We shall see that the new model is simple and competitive with the traditional Poisson regression model and also with the Negative Binomial model.

For that, let Y be now a response variable and $\boldsymbol{z}$ be an associated $q\times 1$ vector of covariates. The modified Borel–Tanner regression model for Y established that given z, Y follows a modified Borel–Tanner distribution with mean $\eta(\boldsymbol{z})$, a positive-valued function. We assume that $\eta(\boldsymbol{z})$ depends on a vector $\boldsymbol{\beta}$ of unknown regression coefficients. This parameterization has the appealing property that when $\eta(\boldsymbol{z})$ takes the common log-linear form $\eta(\boldsymbol{z}) = \exp(\boldsymbol{z}'\boldsymbol{\beta})$.

Writing the likelihood in terms of θ we have

$$\Pr(Y\!=y) \,=\, C_y\Big(p(\theta)^y\,\big(1-p(\theta)\big)^{y+1}\Big)\,, \tag{4.1}$$

for $y = 0, 1, ...$, being $p(\theta) = \theta/(1+2\theta)$ and $\theta > 0$.

Therefore, we assumed that $\theta = \eta(\boldsymbol{z}) = \exp(\boldsymbol{z}'\boldsymbol{\beta})$. Let now $(y_i, \boldsymbol{z_i})$ be a random sample of size n with counts y_i and a vector $\boldsymbol{z_i}$ of covariates for $i = 1, 2..., n$. Then, the log-likelihood function, assuming model (4.1) results

$$\begin{aligned}\ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \log \Pr(Y_i = y_i \,|\, \boldsymbol{z_i}; \boldsymbol{\beta}) \\ &\propto \sum_{i=1}^{n} (1 + y_i) \log\big(1 - p(\theta_i)\big) + \sum_{i=1}^{n} y_i \log p(\theta_i) \ ,\end{aligned}$$

where

$$p(\theta_i) = \frac{\exp\big(\sum_{s=1}^{q} z_{is}\beta_s\big)}{1 + 2\exp\big(\sum_{s=1}^{q} z_{is}\beta_s\big)} \ , \qquad 1 = 1, 2, ..., n \ .$$

Some computations provide that

$$\frac{\partial p(\theta_i)}{\partial \beta_j} = z_{ij}\, p(\theta_i)\,\big(1 - 2p(\theta_i)\big) \ , \qquad i = 1, 2, ..., n; \quad j = 1, 2, ..., q \ ,$$

from which the normal equations can be written as

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{z_{ij}\big(1 - 2p(\theta_i)\big)\,\big(y_i - (1 + 2y_i)\,p(\theta_i)\big)}{1 - p(\theta_i)} = 0 \ , \qquad j = 1, 2, ..., q \ .$$

The elements of the expected Fisher information matrix $\boldsymbol{I} = (I_{jl})$, $j, l = 1, ..., q$, about β_j, $j = 1, ..., q$, are given by

$$\begin{aligned}I_{jj} &= \sum_{i=1}^{n} z_{ij}^2 \,\frac{(1 + 2\theta_i)\,(1 + \theta_i)}{2\theta_i^2 + 4\theta_i + 3} \ , \qquad j = 1, ...q \ , \\ I_{jl} &= \sum_{i=1}^{n} z_{ij}\, z_{il}\, \frac{(1 + 2\theta_i)\,(1 + \theta_i)}{2\theta_i^2 + 4\theta_i + 3} \ , \qquad j, l = 1, ..., q, \quad j \neq l \ .\end{aligned}$$

The residuals can now be used to identify discrepancies between models and data, so the computation of the individual residuals from each observation can be useful to evaluate the model-fitting.

The common Pearson residuals are obtained by dividing the raw residuals by their scaled standard deviation, according to the model

$$\epsilon_i^P = \frac{y_i - \widehat{\theta}_i}{\sqrt{var\big(Y_i, \widehat{\theta}_i\big)}} \ , \qquad i = 1, 2, ..., n \ .$$

Here, $var(Y_i; \widehat{\theta}_i)$ is the variance of Y_i as a function of θ and $\widehat{\theta}_i$ is the maximum likelihood estimate of the i-th mean as fitted to the regression model.

With the aim of comparison between models, we consider as alternative options that the conditional distribution of the response variable can be described by Poisson, negative binomial or MBT distributions. This way, we obtain the corresponding Pearson residuals for each model:

a) Poisson: $\epsilon_i^P = \dfrac{y_i - \widehat{\theta}_i}{\sqrt{\widehat{\theta}_i}}, \quad i = 1, 2, ..., n.$

b) MBT: $\epsilon_i^P = \dfrac{y_i - \widehat{\theta}_i}{\sqrt{\widehat{\theta}_i\,(1+\widehat{\theta}_i)\,(1+2\widehat{\theta}_i)}}, \quad i = 1, 2, ..., n.$

c) Negative binomial: $\epsilon_i^P = \sqrt{r}\,\dfrac{y_i - \widehat{\theta}_i}{\sqrt{\widehat{\theta}_i\,(r+\widehat{\theta}_i)}}, \quad i = 1, 2, ..., n.$

Another common choice of residuals is the signed square root of the contribution to the deviance goodness-of-fit statistic. This is given by $D = \sum_{i=1}^n d_i$, where

$$d_i \;=\; \operatorname{sgn}(\widehat{\theta}_i - y_i)\,\sqrt{2\left(\ell(y_i) - \ell(\widehat{\theta}_i)\right)}\;, \qquad i = 1, 2, ..., n\;,$$

where sgn is the function that returns the sign (plus or minus) of the argument. The $\ell(y_i)$ term is the value of the log likelihood when the mean of the conditional distribution for the i-th individual is the individual's actual score of the dependent variable. The $\ell(\widehat{\theta}_i)$ is the log likelihood when the conditional mean is substituted in the log likelihood. Usually the deviance divided by its degree of freedom is examined taking into account that a value much greater than one indicates a poorly fitting model. See for example [14].

It is well-known that for the Poisson distribution with parameter θ_i the deviance residuals are given by (see [8])

$$d_i \;=\; \operatorname{sgn}\bigl(y_i - \widehat{\theta}_i\bigr)\left[2\left(y_i \log\!\left(\frac{y_i}{\widehat{\theta}_i}\right) - \bigl(y_i - \widehat{\theta}_i\bigr)\right)\right]^{1/2}, \qquad i = 1, 2, ..., n\;.$$

For the MBT distribution proposed here the deviance residual are obtained as follows for each $i = 1, ..., n$:

$$d_i \;=\; \operatorname{sgn}\bigl(y_i - \widehat{\theta}_i\bigr)\left[2\left(\bigl(1+y_i\bigr)\,\log\!\left(\frac{1-p(y_i)}{1-p(\widehat{\theta}_i)}\right) + y_i \log\!\left(\frac{p(y_i)}{p(\widehat{\theta}_i)}\right)\right)\right]^{1/2}.$$

For the negative binomial distribution, an expression for the deviance residuals can be found in [14]:

$$d_i \;=\; \operatorname{sgn}\bigl(y_i - \widehat{\theta}_i\bigr)\left[2\left(y_i \log\!\left(\frac{y_i}{\widehat{\theta}_i}\right) - \bigl(y_i + r\bigr)\,\log\!\left(\frac{y_i + r}{\widehat{\theta}_i + r}\right)\right)\right]^{1/2}.$$

In the three above considered cases we assume $y_i \neq 0$ for all i.

5. NUMERICAL ILLUSTRATION

In this section, we examine an application of the MBT regression model proposed here in order to analyse the number of deaths in truck's accidents ([17]).

In the present study, we model the number of deaths in the accident as the dependent variable. The explanatory variables are as follows: (1) the number of occupants; (2) a dummy variable for seat belt usage; (3) a set of dummy variables for rain, snow and fog, respectively; (4) a dummy variable for dark; (5) a dummy variable for weekdays; (6) a dummy variable for the first driver being drunk; (7) Dummy for the second driver being drunk and (8) a dummy variable for the first driver to be under 21 and finally, (9) a dummy variable for the first driver to be over 60. Due to the fact that the dependent variable is a count variable, data analysis including covariates would be a more appropriate method (see e.g. [7]; [6]; [5]; among others). Table 2 presents the estimates of the MBT, Poisson and Negative Binomial regression models, respectively.

Only for comparative purposes, we fit the MBT, Poisson and Negative Binomial distributions to this data set (see Table 3). We used the value of the log-likelihood function, the Akaike Information Criterion (AIC) (see [1]), the Bayesian Information Criterion (BIC) (see [20] and the Consistent Akaike Information Criterion (CAIC) (see [4]) to compare the estimated models.

Table 3 shows that the MBT model performs very well in fitting the distribution, compared to other uniparametric models Poisson, and provides a fit as good as that of the biparametric Negative Binomial model. Based on the BIC and CAIC, the MBT distribution fits the data better than NB, and NB distribution is better than Poisson. Furthermore, the MBT model presented is somewhat simpler than the NB and therefore it might appear to be preferable as a less complex model, taking into account the Ockham's razor principle (Jaynes, 1994).

The comparative study of Pearson residuals, deviance, log-likelihood and information criteria are also collected in Table 3. Note that the MBT model obtains a better result than the Negative Binomial when the Pearson statistic is the comparison criterion. Furthermore, graphical models diagnostics is now developed using the above residual expressions (see Figure 2).

In addition, one can be interested in testing the null that models are equally close to the actual model, against the alternative that one model is closer ([21]). The z-statistic is

$$Z = \frac{1}{\omega\sqrt{n}}\left(\ell_f(\widehat{\theta}_1) - \ell_g(\widehat{\theta}_2)\right),$$

Table 2: Parameter estimates for data in Li (23012) under the models considered. The response variable is Number of deaths in the accident. Variables statistically significant (at level < 0.05) in boldface.

Variable	MBT model			
	Estimate	S.D.	$\|t\|$-value	Pr($>\|t\|$)
Intercept	-6.50341	0.23392	27.8019	< 0.001
Number of occupants	0.47966	0.04522	10.6070	< 0.001
Seat belt usage	-1.16920	0.16188	7.2224	< 0.001
Rain	-0.78854	0.34854	2.2623	0.0236
Snow	-0.25206	0.59300	0.4250	0.6707
Fog	-7.02683	37.2207	0.1887	0.8502
Dark	0.71680	0.16366	4.3798	< 0.001
Weekday	-0.11630	0.16636	0.6991	0.4844
First driver drunk	1.37529	0.23324	5.8962	< 0.001
Second driver drunk	1.06292	0.22884	4.6446	< 0.001
First driver drug	0.70307	0.20635	3.4071	< 0.001
Second driver drug	0.65230	0.20220	3.2258	0.0012
Age driver < 21	0.09295	0.24036	0.3867	0.6989
Age driver > 60	1.03063	0.21773	4.7334	< 0.001
	Poisson model			
Intercept	-6.41786	0.21997	29.1749	< 0.001
Number of occupants	0.44529	0.03434	12.9656	< 0.001
Seat belt usage	-1.15769	0.15749	7.3507	< 0.001
Rain	-0.80440	0.34272	2.3470	0.0189
Snow	-0.28794	0.58335	0.4936	0.6215
Fog	-5.02453	13.6193	0.3689	0.7121
Dark	0.73782	0.15848	4.6556	< 0.001
Weekday	-0.14283	0.15969	0.8944	0.3710
First driver drunk	1.33410	0.21764	6.1297	< 0.001
Second driver drunk	1.03833	0.21365	4.8599	< 0.001
First driver drug	0.66979	0.19880	3.3691	< 0.001
Second driver drug	0.64306	0.19391	3.3162	< 0.001
Age driver < 21	0.11134	0.22843	0.4874	0.6259
Age driver > 60	1.02884	0.21108	4.8740	< 0.001
	Negative Binomial model			
r	0.12527	0.04463	2.8067	0.0050
Intercept	-6.54213	0.24326	26.8932	< 0.001
Number of occupants	0.49415	0.05028	9.8269	< 0.001
Seat belt usage	-1.19020	0.16753	7.1042	< 0.001
Rain	-0.74537	0.35349	2.1085	0.0349
Snow	-0.19223	0.60245	0.3190	0.7496
Fog	-2.24947	3.45645	0.6508	0.5151
Dark	0.70502	0.16911	4.1689	< 0.001
Weekday	-0.09364	0.17342	0.5399	0.5892
First driver drunk	1.39773	0.24756	5.6458	< 0.001
Second driver drunk	1.09946	0.24494	4.4885	< 0.001
First driver drug	0.72924	0.21385	3.4099	< 0.001
Second driver drug	0.63340	0.21050	3.0090	0.0026
Age driver < 21	0.08526	0.24970	0.3414	0.7327
Age driver > 60	1.03876	0.22568	4.6026	< 0.001

Table 3: Summaries of fitting measures results for the models considered.

Criterion	Model		
	Poisson	Neg. Bin.	MBT
$\left(\epsilon_i^P\right)^2$	55196.5	54638.5	55225.8
Deviance	492.792	387.865	421.177
Deviance/df	0.00825	0.00649	0.00705
$\ell_{\max}$	-1090.49	-1074.91	-1077.42
AIC	2208.97	2179.81	2182.85
BIC	2334.94	2314.77	2308.81
CAIC	2348.94	2329.77	2322.81

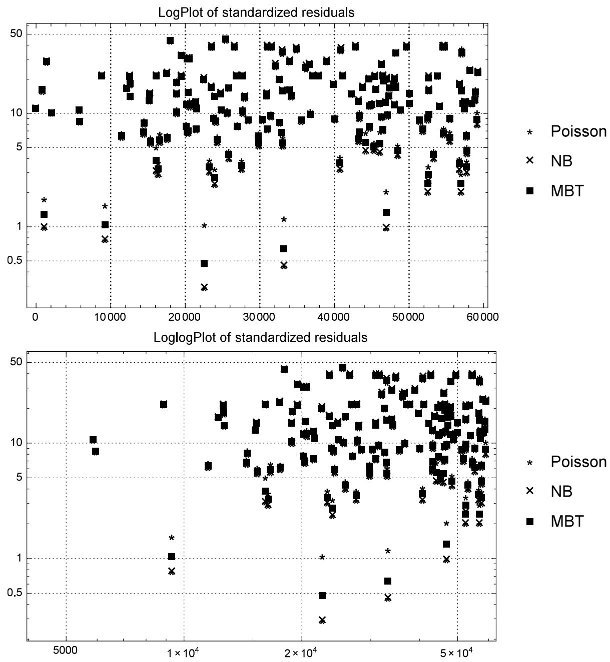

Figure 2: LogPlot and LoglogPlot of standardized residuals for the models considered.

where

$$\omega^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{f(x_i\,|\,\widehat{\theta}_1)}{g(x_i\,|\,\widehat{\theta}_2)}\right)\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f(x_i\,|\,\widehat{\theta}_1)}{g(x_i\,|\,\widehat{\theta}_2)}\right)\right]^2$$

and f and g represent here the MBT and the alternative distributions, respectively.

Due to the asymptotic normal behavior of the Z statistic under the null, rejection of the test in favor of f happens, with significance level α, when $Z > z_{1-\alpha}$ being $z_{1-\alpha}$ the $(1-\alpha)$ quantile of the standard normal distribution.

Table 4 shows the results obtained for each comparison by means of the Voung test. The MBT model is preferred to the Poisson model and we cannot reject the null that the models, Negative binomial and MBT, are statistically the same.

Table 4: Voung test results.

	Z-score	p-value
MBT vs Poisson	2.29834	0.01
MBT vs Neg.bin.	−0.87924	0.81

6. CONCLUSIONS

This paper has introduced a modified version of the Borel–Tanner distribution which takes its values from the non-negative integers, in contrast with the original Borel–Tanner distribution which is restricted to the positive integers.

We obtain an over-dispersed distribution (its variance is larger than its mean) depending on just one parameter, which is also unimodal with mode at zero. Furthermore the distribution is infinitely divisible (log-convex) and therefore it may be considered as a compound Poisson distribution. Some other properties based on results in [3] are also verified.

In addition, a simple reparameterization of the MBT distribution allows to incorporate in an easy way covariates into the model.

In this paper, a numerical application is provided, where both the Poisson and the negative binomial model-fitting are compared to the MBT. The practical use of the modified Borel–Tanner distribution here proposed does not only bring a significant improvement relative to the Poisson distribution but also a wider flexibility due to its main properties, as for instance its over-dispersion. The MBT distribution is found to be a better model to describe the data used in this paper than the Poisson and the negative binomial, according to their BIC and CAIC values.

APPENDIX

In this Appendix we provide a proof for the cumulative distribution function of the distribution and Proposition 3.1.

Proof of the cdf of the distribution: We have that

$$F(x) = 1 - \sum_{j=x+1}^{\infty} \frac{\Gamma(2j+1)}{\Gamma(j+2)\,\Gamma(j+1)}\,\frac{\alpha^j}{(1+\alpha)^{2j+1}}\ .$$

Now by putting $k = j - x - 1$ and using the identity

$$\Gamma(2m) = \frac{1}{\sqrt{2\pi}}\, 2^{2m-1}\,\Gamma(m)\,\Gamma(m+1/2)\ ,$$

which appears in [13], p. 7, we obtain the result after some computations. □

Proof of Proposition 3.1: The discrete distribution with probability function given in (2.1) satisfies the regularity conditions (see [16], p. 449)) under which the unique maximum likelihood estimator $\widehat{\alpha}$ of α is consistent and asymptotically normal. They are simply verified in the following way. Firstly, the parameter space $(0,1)$ is a subset of the real line and the range of x is independent of α. By using expression (3) it is easy to show that $E\big(\frac{\partial \log \Pr(X=x;\alpha)}{\partial\alpha}\big) = 0$. Now, since $\frac{\partial^2 \ell(\alpha)}{\partial\alpha^2}\big|_{\alpha=\widehat{\alpha}} < 0$, the Fisher's information is positive. Finally, by taking $M(x) = 2x/\alpha^3$ we have that

$$\left|\frac{\partial^3 \log \Pr(X=x;\alpha)}{\partial\alpha^3}\right| = \left|\frac{2x}{\alpha^3} - \frac{2(2x+1)}{(1+\alpha)^3}\right| \le M(x)\ ,$$

with $E(M(X)) = 2/(\alpha(1-\alpha)) < \infty$. Hence the proposition. □

ACKNOWLEDGMENTS

The authors would like to express their gratitude to two anonymous referees for their relevant and useful comments. Authors are funded by grant ECO2013–47092 (MINECO, Spain).

REFERENCES

[1] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

[2] Alanko, T. and Duffy, J. (1996). Compound binomial distributions for modelling consumption data, *The Statistician*, **45**(3), 269–286.

[3] Bardwell, E. (1960). On Certain Characteristics of Some Discrete Distributions, *Biometrika*, **47**(3/4), 473–475.

[4] Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, **52**(3), 345–379.

[5] Cartwright, W.; Hu, S.T. and Huang, L.-F. (1992). Medigap insurance coverage on the use of medical services of the elderly, *Applied Economics*, **2**(4), 529–539.

[6] Christensen, S; Long, S. and Rodgers, J. (1987). Acute health care costs for the aged Medicare population: overview and policy options, *Millbank Quarterly*, **65**(3), 397–425.

[7] Duan, N.; Manning, W.; Morris, C. and Newhouse, J. (1983). A comparison of alternative models for the demand for medical care, *Journal of Business and Economic Statistics*, **1**, 115–126.

[8] Dunteman, G.H. and Ho, M.-H.R. (1994). *Probability Theory: The Logic of Science*, `http://omega.math.albany.edu:8008/JaynesBook.html`.

[9] Gupta, P.L.; Gupta, R.C. and Tripathi, R.C. (1997). On the monotonic properties of discrete failure rates, *Journal of Statistical Planning and Inference*, **65**, 255–268.

[10] Haight, F.A. (1961). A distribution analogous to the Borel–Tanner, *Biometrika*, **48**(1–2), 167–168.

[11] Haight, F.A. and Breuer, M.A. (1960). The Borel–Tanner distribution, *Biometrika*, **47**(1–2), 143–150.

[12] Johnson, N.L. and Kotz, S. (1982). Developments in Discrete Distributions, 1969–1980, *International Statistical Review*, **50**(1), 71–101.

[13] Johnson, N.L.; Kemp, A.W. and Kotz, S. (2005). *Univariate Discrete Distributions, Third Edition*, John Wiley, New York.

[14] Jong, P.D. and Heller, G.Z. (2008). *Generalized Linear Models for Insurance Data*, Cambridge University Press, Cambridge.

[15] Karlis, D. and Xekalali, E. (2005). Mixed Poisson Distributions, *International Statistical Review*, **73**(1), 35–58.

[16] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, Springer, New York.

[17] Li, S. (2012). Traffic safety and vehicle choice: quantifying the effects of the "Arms Race" on American roads, *Journal of Applied Econometrics*, **27**(1), 34–62.

[18] Olver, F.W.J.; Lozier, D.W.; Boisvert, R.F. and Clark, C.W.; Eds. (2010). *NIST Handbook of Mathematical Functions*, Cambridge University Press. http://dlmf.nist.gov/

[19] Puig, P. and Valero, J. (2006). Count data distributions: Some characterizations with applications, *Journal of the American Statistical Association*, **101**(437), 332–340.

[20] Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**(2), 461–464.

[21] Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, **57**(2), 307–333.

[22] Warde, W.D. and Katti, S.K. (1971). Infinite divisibility of discrete distributions II, *The Annals of Mathematical Statistics*, **42**(3), 1088–1090.