REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000

ON GOODNESS-OF-FIT TESTS FOR THE NEYMAN TYPE A DISTRIBUTION

Authors: APOSTOLOS BATSIDIS ✉*
– Department of Mathematics, University of Ioannina, 45110 Ioannina, Greece (abatsidis@uoi.gr)

ARTUR J. LEMONTE
– Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Brazil (arturlemonte@gmail.com)

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- The two-parameter Neyman type A distribution is quite useful for modeling count data, since it corresponds to a simple, flexible and overdispersed discrete distribution, which is also zero-inflated. In this paper, we show that the probability generating function of the Neyman type A distribution is the only probability generating function which satisfies a certain differential equation. Based on an empirical counterpart of this specific differential equation, we propose and study a new goodness-of-fit test for this distribution. The test is consistent against fixed alternative hypotheses, while its null distribution can be consistently approximated by using parametric bootstrap. We investigate the finite sample performance of the proposed test numerically by means of Monte Carlo experiments, and comparisons with other existing goodness-of-fit tests are also considered. Empirical applications to real data are considered for illustrative purposes.

Key-Words:

- *Count data; empirical probability generating function; parametric bootstrap; probability generating function; Bell-Touchard distribution.*

AMS Subject Classification:

- 62F03, 62F40

*Corresponding author

1. INTRODUCTION

Modeling count data is an important issue in different disciplines and applied sciences such as medicine (see, for example, Joe and Zhu [25]), actuarial sciences (see, for example, Gossiaux and Lemaire [17], Lord et al. [32]), biology (see, for instance, Esnaola et al. [13]), health economics (see, for example, Zafakali and Ahmad [50]), among many others. With this aim, the one-parameter Poisson distribution and the two-parameter Negative Binomial distribution are commonly used. Nevertheless, observed count data often exhibit overdispersion (i.e., variance greater than the mean) and, therefore, the Poisson distribution is not adequate for fitting such data, since its variance is restricted to be equal to the mean. Additionally, a second usual feature of the observed count data is the presence of a high percentage of zero values (zero inflation or zero vertex). The zero-inflation index $zi = 1 + \log(p_0)/\mu$, where p_0 is the probability of zero, can be used to measure zero-inflation. Then $zi = 0$ for Poisson distribution, and $zi = 1 + log(d)/(1 - d) > 0$ for the Negative Binomial, where d denotes the Fisher dispersion index given by $d = \sigma^2/\mu$, where σ^2 and μ are the variance and mean, respectively [see 42]. Therefore, the Negative Binomial distribution is an improvement over the Poisson distribution, since it can model overdispersed and zero-inflated data.

Several other distributions have been presented in the statistical literature to handle both overdispersion and zero-inflation. In this frame, Neyman [39] developed the now well-known Neyman type A (NTA) distribution, which is overdispersed, because $d \geq 1$, and its zero-inflation index zi is always larger than the respective for the Negative Binomial for any fixed value of the dispersion index d (see Figure 1 in Puig and Valero [42]). For these reasons, the NTA distribution has been used in various disciplines such as bacteriology, ecology and entomology. The reader is referred to Johnson et al. [26, Chapter 9] and to Tripathi [49] for a list of applications of NTA distribution. Let $p_N(k;\tau,\delta)$ and $g_N(t;\tau,\delta)$ be the probability mass function (pmf) and probability generating function (pgf) of the NTA distribution, with parameters $\delta > 0$ and $\tau > 0$. We have that

$$\Pr(X = k) := p_N(k;\tau,\delta) = \frac{\tau^k \mathrm{e}^{\delta(\mathrm{e}^{-\tau}-1)}}{k!} m_k(\delta \mathrm{e}^{-\tau}), \quad k \in \mathbb{N}_0 \tag{1.1}$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\} = \{0, 1, 2, \ldots\}$, $m_k(r) = \sum_{j=0}^{k} S(j,k) r^k$ is the k-th moment about zero for the Poisson distribution with parameter $r > 0$, and $S(k,j)$ are the Stirling numbers of second kind (see, for instance, Massé and Theodorescu [33] for further details). Also, $g_N(t;\tau,\delta) = \exp[\delta(\mathrm{e}^{\tau(t-1)} - 1)], \quad |t| \leq 1$. We shall use the notation $X \sim \mathrm{NTA}(\tau,\delta)$ to refer to this distribution.

Recently, Castellares et al. [8] on the basis of a series expansion presented in Touchard [48] and Bell [4, 5], obtained a two-parameter family of distributions (named as Bell-Touchard distribution) with pmf of the form

$$\Pr(X = k) := p(k;\theta) = \frac{\mathrm{e}^{b(1-\mathrm{e}^{a})}\, a^k\, T_k(b)}{k!}, \quad k \in \mathbb{N}_0, \tag{1.2}$$

where $a > 0$ and $b > 0$, $\theta = (a,b) \in \Theta = (0,\infty) \times (0,\infty)$, and $T_k(\cdot)$ are the Touchard polynomials [48] defined by $T_k(b) = \mathrm{e}^{-b}\sum_{j=0}^{\infty} j^k\, b^j/j!$. We shall use the notation $X \sim \mathrm{BT}(a,b)$, or $X \sim \mathrm{BT}(\theta)$, to refer to the NTA distribution with this specific parameterization. If $X \sim \mathrm{BT}(a,b)$, then its pgf is given by

$$g(t;\theta) = \exp\{[b(\mathrm{e}^{ta} - \mathrm{e}^{a})]\}, \quad |t| \leq 1. \tag{1.3}$$

The Touchard polynomials $T_k(b)$ corresponds to the k-th moment of the Poisson distribution with parameter equal to b and can be obtained for different values of k. For example, $T_0(b) = 1$, $T_1(b) = b$, $T_2(b) = b^2 + b$, $T_3(b) = b^3 + 3b^2 + b$, $T_4(b) = b^4 + 6b^3 + 7b^2 + b$, $T_5(b) = b^5 + 10b^4 + 25b^3 + 15b^2 + b$, $T_6(b) = b^6 + 15b^5 + 65b^4 + 90b^3 + 31b^2 + b$, and so on.

Remark 1.1. Note that when $b = 1$ in (1.2), the pmf of the Bell distribution introduced by Castellares et al. [7] is obtained as a special case, while the $\mathrm{BT}(a,b)$ distribution corresponds to the $\mathrm{NTA}(\delta = b\mathrm{e}^{a}, \tau = a)$ distribution. So, the Bell-Touchard (BT) distribution is a reparameterization of the NTA distribution and, hence, in the whole paper the BT distribution stands for this reparameterization of the NTA distribution.

It is worth emphasizing that the two-parameter BT discrete distribution, or equivalently the NTA distribution, is very simple to deal with, since its pmf does not contain any complicated function. Tractability of the pmf may be a great advantage in computing the probabilities, as well as structural properties from that equation. The BT distribution has, among many other interesting properties the following properties: (i) it includes the one-parameter Bell distribution introduced by Castellares et al. [7] as a special case, which is also a reparameterization of the well-known NTA distribution; (ii) the Poisson distribution is not nested in the BT family, but it can be approximated for small values of a specific parameter of the BT distribution; (iii) it is a special case of a multiple Poisson process and can have a zero vertex; (iv) it is infinitely divisible; (v) it has variance larger than the mean; (vi) it is strongly unimodal for $b \geq 1$; and (vii) it has an arbitrary number of modes when $b < 1$. For a detailed description of the NTA distribution, the reader could consult Castellares et al. [8] and Johnson et al. [26, Chapter 9].

Based on the key features of the NTA distribution (or equivalently BT distribution), it can be easily justified why this distribution is a natural candidate and plays an important role in modeling count data with evidence of overdispersion and with high percentage of zero values. This implies that it is crucial to test the goodness-of-fit (gof) of this discrete distribution fitted to a given set of observations. A number of gof tests for count data are based on the pgf and the empirical pgf (epgf). To mention a few, but not limited to, we have the gof tests in Kocherlakota and Kocherlakota [29], Rueda et al. [46], Baringhaus and Henze [2], Epps [14], Rueda and O'Reilly [45], Meintanis and Bassiakos [36], Meintanis [35], Jiménez-Gamero and Alba-Fernandez [21], Batsidis et al. [3] and Milocevic et al. [37]. The motivation of using methods based on the pgf instead of the corresponding pmf when dealing with count data is, as argued by Nakamura and Perez-Abreu [38], that the pgf is usually simpler than the corresponding pmf. This is the case of the pgf of the BT distribution; compare expressions (1.2) and (1.3).

In this paper, we propose and study a consistent gof test for the two-parameter BT family of distributions; that is, based on Remark 1.1, it is equivalently to study a consistent gof test for the NTA distribution. Initially, it is shown that the pgf of the BT distribution is the only pgf satisfying a certain differential equation. Then, reasoning as Nakamura and Perez-Abreu [38] for testing Poisson distribution, Novoa-Muñoz and Jiménez-Gamero [41] for testing bivariate Poisson distribution, Jiménez-Gamero and Alba-Fernandez [21] for testing Poisson-Tweedie distribution, and Batsidis et al. [3] for testing Bell distribution, the proposed statistic is a function of the polynomial of an empirical version of the differential equation. In particular, the gof test proposed here can be considered as a generalization of the one in Batsidis et al. [3], since Bell distribution is a special case of the BT distribution. In addition, it can also be thought as a complement to the gof test for the Poisson-Tweedie distribution presented by Jiménez-Gamero and Alba-Fernandez [21], since NTA is a subset of the Poisson-Tweedie family of distributions. Additionally, for the first time, we apply some existing gof tests to the BT distribution and study their finite-sample properties from Monte Carlo simulation experiments. In particular, the numerical results reveal that two of the existing gof tests considered to the BT distribution present interesting results regarding size and power properties.

The paper is organized as follows. Section 2 contains some preliminaries related to existing gof tests. Section 3 introduces the test statistic and derives the asymptotic null distribution of the test statistic (i.e., the test statistic distribution under the null hypothesis), which depends on unknown quantities. To overcome this problem, it is shown that the parametric bootstrap consistently estimates the null distribution of the test statistic. Section 4 is devoted to study, with Monte Carlo simulation experiments, the finite sample performance of the proposed test and simultaneously to compare numerically the power of the new test with other two pgf-based tests introduced by Rueda and O'Reilly [45] and Meintanis [35]; that is, we also consider the pgf-based tests introduced by these authors to the BT distribution and study their finite sample properties in such a case. Apart from the previous gof tests, which are based on the pgf, the tests in Henze [19] and Klar [27], which are similar to that in Rueda and O'Reilly [45] but based on the distribution function and on the integrated distribution function, will also be considered in the comparison of the existing gof tests. Section 5 provides the application of the gof tests to real data sets. Section 6 closes up the paper with some concluding remarks. All technical proofs are deferred to Appendix.

Before ending this section we introduce some notation: all limits in this paper are taken when $n \to \infty$, where n denotes the sample size; $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution; $\xrightarrow{\mathcal{P}}$ denotes convergence in probability; $\xrightarrow{a.s.}$ denotes the almost sure convergence; $I(A)$ denotes the indicator function of the set A; l^2 denotes the separable Hilbert space $l^2 = \{z = (z_0, z_1, z_2, \ldots),\ z_k \in \mathbb{R},\ \sum_{k\geq 0} z_k^2 < \infty\}$ with the usual inner product $\langle z, w\rangle_2 = \sum_{k\geq 0} z_k w_k$, and $\|\cdot\|_2$ stands for the associated norm; $\mathbb{E}_\theta$ and Cov_θ denote expectation and covariance by assuming that the data come from a BT distribution with parameter vector $\theta = (a, b)$; P_*, $\mathbb{E}_*$ and Cov_* denote the conditional probability law, the conditional expectation and the conditional covariance, respectively, given the data $X_1, \ldots, X_n$.

2. PRELIMINARIES AND EXISTING GOODNESS-OF-FIT TESTS

Let $X_1, \ldots, X_n$ be n independent and identically distributed random observations from a population X taking values in $\mathbb{N}_0$, with pgf $g(t) = \mathbb{E}(t^X)$, $|t| \leq 1$. Based on the sample $X_1, \ldots, X_n$, the objective is to test the composite, in the sense that the parameter vector $\theta = (a, b)$ is unknown, null hypothesis $H_0 : X \sim \mathrm{BT}(\theta)$, for some $\theta = (a, b) \in \Theta$ against the alternative hypothesis $H_1 : X \nsim \mathrm{BT}(\theta)$, $\forall\, \theta = (a, b) \in \Theta$. Obviously, based on Remark 1.1, the previous hypothesis is equivalent in testing the null hypothesis $H_0 : X \sim \mathrm{NTA}(\delta, \tau)$, for some $(\delta, \tau) \in (0, \infty) \times (0, \infty)$, against the alternative hypothesis $H_1 : X \nsim \mathrm{NTA}(\delta, \tau)$, $\forall\, (\delta, \tau) \in (0, \infty) \times (0, \infty)$.

It is well-known that the distribution of a random variable X taking values in $\mathbb{N}_0$ is fully and uniquely determined by its pgf. Also, the pgf can be consistently estimated by the epgf given by $g_n(t) = \frac{1}{n} \sum_{i=1}^{n} t^{X_i}$. It is worth stressing that Kocherlakota and Kocherlakota [29] were the first authors who proposed to base a gof test on the so-called epgf process with estimated parameter given by $K_n(\widehat{\theta}, t) = \sqrt{n}[g_n(t) - g(t; \widehat{\theta})]$, for $0 \leq t \leq 1$, where $g(t; \theta)$ is the pgf under the law in the null hypothesis; that is, in our special case, $g(t; \theta)$ is given in relation (1.3), and $\widehat{\theta} = (\widehat{a}, \widehat{b})$ is a consistent estimator of $\theta = (a, b)$.

Kocherlakota and Kocherlakota [29] exemplified their method with the Poisson-type distributions and NTA distribution. However, their method has the disadvantage that it depends on the choice of the value of t at which the pgf is evaluated. To overcome this problem, Rueda et al. [46] suggested the use of the following Cramér-von Mises type test statistic $R_{n,0}(\widehat{\theta}) = \int_0^1 K_n(\widehat{\theta}, t)^2 dt = n \int_0^1 [g_n(t) - g(t; \widehat{\theta})]^2 dt$. In addition, Rueda and O'Reilly [45] proposed a natural generalization of the Cramér-von Mises type test statistic by introducing a suitable weight function in order to make the test more sensitive to selected alternatives; see also Baringhaus et al. [1]. In this frame, they suggested the following test statistic $R_{n,w}(\widehat{\theta}) = n \int_0^1 [g_n(t) - g(t; \widehat{\theta})]^2 w(t) dt$, where $w(t)$ is a non-negative function on $(0, 1)$ such that $\int_0^1 w(t) dt < \infty$. By straightforward algebra, we have that $R_{n,w}(\widehat{\theta}) = \frac{1}{n} \sum_{j,k=1}^{n} \{\omega(1, X_{jk}) - \omega(g(t; \widehat{\theta}), X_j) - \omega(g(t; \widehat{\theta}), X_k) + \omega(g^2(t; \widehat{\theta}), 0)\}$, where $X_{jk} = X_j + X_k$, and $\omega(f, d) = \int_0^1 t^d f(t) w(t) dt$. Note that $R_{n,w}(\widehat{\theta})$ can be equivalently expressed in the form $R_{n,w}(\widehat{\theta}) = n \sum_{r,k=0}^{\infty} \{p(r; \theta) - \widehat{p}(r)\}\{p(k; \theta) - \widehat{p}(k)\} \int_0^1 t^{r+k} w(t) dt$, where $p(k; \theta)$ is given by (1.2), and

$$\widehat{p}(k) = \frac{1}{n} \sum_{j=1}^{n} I\left(X_j = k\right), \quad k = 0, 1, \ldots. \tag{2.1}$$

Note that $\widehat{p}(k)$ corresponds to the empirical pmf for a given dataset. Hence, one rejects the null hypothesis H_0 for large values of the test statistic $R_{n,w}(\widehat{\theta})$.

After the pioneer work by Kocherlakota and Kocherlakota [29], a large number of gof tests for specific discrete distributions have been developed based on test statistics that utilize properties of the pgf of the law under the null hypothesis. In this context, Meintanis [35] presented a unified approach in testing the fit to any distribution belonging to the compound Poisson family of distributions. The compound Poisson family of distributions is defined as the distribution of $X = \sum_{j=1}^{N} Y_j$, where Y_j $(j = 1, \ldots, N)$ are independent and identically distributed with a common pgf $\psi(t;\xi)$, $\xi \in \mathbb{R}^p$ is a parameter vector, $N \sim \text{Poisson}(\lambda)$ is independent of Y_j $(j = 1, \ldots, N)$, and $\lambda > 0$. Meintanis [35] has noted that the pgf of any member of the compound Poisson family, say $\zeta(t)$, satisfies the following differential equation

$$\zeta'(t) - \lambda\psi'(t;\xi)\zeta(t) = 0, \tag{2.2}$$

where $\zeta'(t) = (d/dt)\zeta(t)$ and $\psi'(t;\xi) = (d/dt)\psi(t;\xi)$. Then, since the pgf and its derivatives can be consistently estimated by the epgf and the derivatives of the epgf (see, for example, Proposition 2 of Novoa-Muñoz and Jiménez-Gamero [40] for the uniform consistency of g_n and its derivatives), Meintanis [35] proposed the following test statistic

$$T_{n,w}(\widehat{\lambda},\widehat{\xi}) = n\int_0^1 [\zeta_n'(t) - \widehat{\lambda}\psi'(t;\widehat{\xi})\zeta_n(t)]^2 w(t)dt, \tag{2.3}$$

where $\zeta_n'(t) = (d/dt)\zeta_n(t)$, and $\zeta_n(t)$ denotes the epgf. Note that the test statistic defined in (2.3) is an integral of the squared of an empirical counterpart of equation (2.2).

The general test statistic given in (2.3) can be exemplified in the special case of the BT distribution with parameter vector $\theta = (a, b)$, once the proposition below justifies that the BT distribution belongs to the compound Poisson family of distributions. This result can be found in Feller [15] and in Castellares et al. [8].

Proposition 2.1. *Let $X \sim \mathrm{BT}(a, b)$, where $a > 0$ and $b > 0$. Then, we have that $X = \sum_{j=1}^{N} Y_j$, where Y_j $(j = 1, \ldots, N)$ are independent and identically zero-truncated Poisson distributed random variables with parameter $a > 0$ and a common pgf $\psi(t;a) = \frac{\exp(at)-1}{\exp(a)-1}$, and $N \sim$ Poisson$(b(\mathrm{e}^a - 1))$ independent of Y_j $(j = 1, \ldots, N)$.*

In terms of the notation used by Meintanis [35], it is evident that the BT distribution belongs to the compound Poisson family with $\lambda = b(\mathrm{e}^a - 1)$, $\psi(t;\xi) = \frac{\exp(\xi t)-1}{\exp(\xi)-1}$, $\psi'(t;\xi) = \frac{\xi\exp(\xi t)}{\exp(\xi)-1}$, and $\xi = a$. Therefore, based on the work of Meintanis [35], the pgf $g(t;\theta)$ of the BT distribution defined in (1.3) satisfies the following differential equation

$$g'(t) - ba\mathrm{e}^{at}g(t) = 0, \quad \forall\, t \in [0,1], \tag{2.4}$$

and so the null hypothesis H_0 is rejected for large values of the following test statistic $M_{n,w}(\widehat{\theta}) = n\int_0^1 G_n(t,\widehat{\theta})^2 w(t)dt$, where $G_n(t;\theta)$ is the empirical version of (2.4) given by

$$G_n(t;\widehat{\theta}) = g_n'(t) - \widehat{b}\widehat{a}\mathrm{e}^{\widehat{a}t}g_n(t), \tag{2.5}$$

with $g_n'(t) = (d/dt)g_n(t)$. By straightforward algebra (see also Meintanis [35, p. 753]), we have that $M_{n,w}(\widehat{\theta}) = \frac{1}{n}\sum_{j,k=1}^{n}\{X_jX_k\omega(1, X_{jk} - 2) + (\widehat{b}\widehat{a})^2\omega(\mathrm{e}^{2\widehat{a}t}, X_{jk}) -$

$\widehat{b}\widehat{a}X_{jk}\omega(\mathrm{e}^{\widehat{a}t}, X_{jk}-1)\}$. Note that $M_{n,w}(\widehat{\theta})$ can be equivalently expressed in the form $M_{n,w}(\widehat{\theta}) = n\sum_{r,k=0}^{\infty}\widehat{d}(r;\widehat{\theta})\widehat{d}(k;\widehat{\theta})\int_0^1 t^{r+k}w(t)dt$, where

$$\widehat{d}(k;\theta) = (k+1)\widehat{p}(k+1) - \sum_{u=0}^{k} coef(u;\theta)\widehat{p}(k-u), \quad k = 0, 1, \ldots, \tag{2.6}$$

and $coef(u;\theta) := coef(u;a,b) = \frac{ba^{u+1}}{u!}$ can be recursively calculated as follows: $coef(0;a,b) = ba$, and $coef(u;a,b) = coef(u-1;a,b)a/u$ for $u \geq 1$.

Remark 2.1. The asymptotic null distributions of the test statistics $R_{n,w}(\widehat{\theta})$ and $M_{n,w}(\widehat{\theta})$ are intractable (Rueda and O'Reilly [45] and Meintanis [35]) and, hence, the critical points required for the implementation of these test procedures can be determined via parametric bootstrap. It should be mentioned that the application of both tests requires the choice of a weight function. Specific choices of it, which are rather arbitrary, can lead to considerable computational simplification. In this frame, the choice of $w(t) = t^{\gamma}$, where $\gamma \geq 0$ denotes a constant, corresponds to an interesting choice. This weight function, apart from computational convenience, has the following interpretation: for large values of γ more weight is assigned to the values of $K_n(\widehat{\theta}, t)$ and $G_n(t;\widehat{\theta})$ near $t = 1$; hence, large values of γ should render the test sensitive to deviations from the moments of the hypothesized distribution; see, for instance, Gürtler and Henze [18].

Apart from the previous tests, which are based on the pgf, the tests in Henze [19] and Klar [27] denoted as H_n and W_n, which are similar to that in Rueda and O'Reilly [45] but they are based on the distribution function and on the integrated distribution function, respectively, will be also particularized for the BT distribution and will be also considered in the simulation studies of Section 4. Specifically, we consider the modified Cramér–von Mises statistic in expression (3.6) of Henze [19] given by

$$H_n = \sum_{k=0}^{X_{(n)}}[F_n(k) - F(k;\widehat{\theta})]^2[F_n(k) - F_n(k-1)], \tag{2.7}$$

where $X_{(n)} = \max_{1\leq j\leq n} X_j$, $F_n(x)$ stands for the empirical distribution function defined by $F_n(x) = n^{-1}\sum_{j=1}^{n} I(X_j \leq x)$, and $F(x;\theta)$ denotes the cumulative distribution function of the BT distribution with parameter θ. In contrast to the Cramér–von Mises statistic in expression (2.2) of Henze [19], whose practical calculation involves truncation, the calculation of H_n involves a finite sum and hence was preferred (see also Jiménez-Gamero and Alba-Fernandez [22]). Finally, following Henze [19], to perform the test based on H_n a parametric bootstrap is used and the null hypothesis is rejected for a large observed value of the test statistic H_n. We also consider the test statistic (see relation (1) in Klar [27]) $W_n = \sqrt{n}\sup_{t\geq 0}|Y_n(t) - \widehat{Y}(t)|$, where $Y(t) = \int_t^{+\infty}[1-F(x)]dx$, $Y_n(t)$ denotes its empirical counterpart and $\widehat{Y}(t)$ equals $Y(t)$ with $F(x)$ replaced by $F(x;\widehat{\theta})$. In practice (see also Jiménez-Gamero and Alba-Fernandez [22]), we consider the expression (8) in Klar [27] given by

$$W_n = \sqrt{n}\sup_{1\leq k\leq X_{(n)}}\left|\bar{X} - \mathbb{E}_{\widehat{\theta}}(X) + \sum_{j=0}^{k-1}[F_n(j) - F(j;\widehat{\theta})]\right|,$$

where $\bar{X}$ denotes the sample mean. For instance, if the moment estimator is used then the previous relation is simplified taking into account that $\mathbb{E}_{\widehat{\theta}}(X) = \bar{X}$. On the other hand, if the maximum likelihood (ML) estimator is used, then the relation is simplified taking into account that $\mathbb{E}_{\widehat{\theta}}(X) = \widehat{b}\widehat{a}\mathrm{e}^{\widehat{a}}$, where $\widehat{a}$ and $\widehat{b}$ are the ML estimates of a and b, respectively, since $\mathbb{E}(X) = ba\mathrm{e}^{a}$, when $X \sim \mathrm{BT}(a, b)$. Following Klar [27], to perform the test based on W_n a parametric bootstrap is used and hence the null hypothesis is rejected for a large value of the associated test statistic.

3. A NEW TEST STATISTIC

In this section, a new gof test statistic will be constructed based on the characterization of the BT distribution provided below and parallel with the tests discussed by Nakamura and Perez-Abreu [38] for testing Poisson distribution, Novoa-Muñoz and Jiménez-Gamero [41] for testing bivariate Poisson, Jiménez-Gamero and Alba-Fernandez [21] for testing Poisson-Tweedie, and Batsidis et al. [3] for testing Bell distribution. To be specific, the next proposition shows that the BT pgf is the unique solution of the differential equation given in (2.4).

Proposition 3.1. *Let* $G = \{g : [0, 1] \to \mathbb{R}$, *such that* g *is a pgf and* $g'(t) = (\partial/\partial t)g(t)$ *exists* $\forall\, t \in [0, 1]\}$, *which is equivalent to say that* G *is the set of probability generating functions associated with random variables taking values in* $\mathbb{N}_0$ *with finite mean. Let* $g(t; \theta)$ *be defined as in* (1.3). *Then,* $g(t; \theta)$ *is the only pgf in* G *satisfying the differential equation given in* (2.4).

Therefore, the BT pgf is the only pgf satisfying the differential equation (2.4). Also, the pgf $g(t)$ and its derivatives can be consistently estimated by the epgf and the derivatives of the epgf. Under the null hypothesis H_0, it then follows that the empirical version of (2.4) denoted by $G_n(t; \widehat{\theta})$ and given in (2.5) should be close to zero, $\forall\, t \in [0, 1]$, where $\widehat{\theta} = (\widehat{a}, \widehat{b})$ is a consistent estimator of $\theta = (a, b)$. Additionally, $G_n(t; \widehat{\theta})$ can be expressed in the form $G_n(t; \widehat{\theta}) = \sum_{k\geq 0} \widehat{d}(k; \widehat{\theta})t^k$, where $\widehat{p}(k)$ and $\widehat{d}(k; \widehat{\theta})$ are defined in (2.1) and (2.6), respectively. It implies that (under the null hypothesis) $S_n(\widehat{\theta}) = \sum_{k\geq 0} \widehat{d}(k; \widehat{\theta})^2 \approx 0$. Note that $S_n(\widehat{\theta}) = \|\widehat{d}(\cdot; \widehat{\theta})\|_2^2$, where $\widehat{d}(\cdot; \widehat{\theta}) = (\widehat{d}(0; \widehat{\theta}), \widehat{d}(1; \widehat{\theta}), \ldots)$, and $\widehat{d}(k; \widehat{\theta})$ is given in (2.6). Also, $\widehat{d}(k; \theta) = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i; k, \theta)$, where

$$\phi(X; k, \theta) = (k + 1)I(X = k + 1) - b\sum_{u=0}^{k} \frac{a^{u+1}}{u!} I(X = k - u). \tag{3.1}$$

In this paper, we propose and study a new gof test for the BT family of distributions based on the statistic $S_n(\widehat{\theta})$. In order to give a solid justification of $S_n(\widehat{\theta})$ as a test statistic for testing H_0, we derive its limit distribution in the next theorem.

Theorem 3.1. *Let* $X_1, \ldots, X_n$ *be independent and identically distributed from* X, *a random variable taking values in* $\mathbb{N}_0$ *with probability mass function* $p(k) = \Pr(X =$

k), $k \in \mathbb{N}_0$, so that $\mathbb{E}(X^2) < \infty$. Assume that $\widehat{\theta} \xrightarrow{a.s.(P)} \theta$, then $S_n(\widehat{\theta}) \xrightarrow{a.s.(P)} \eta = \|d(\cdot;\theta)\|_2^2$, where $d(\cdot;\theta) = (d(0;\theta), d(1;\theta), \ldots)$, and $d(k;\theta) = (k+1)p(k+1) - b\sum_{u=0}^{k} \frac{a^{u+1}}{u!} p(k-u), \quad k \in \mathbb{N}_0$.

It should be noted that $\eta \geq 0$ and, from Proposition 3.1, $\eta = 0$ if and only if H_0 is true. Hence, the null hypothesis H_0 should be rejected for large values of the test statistic $S_n(\widehat{\theta})$. Now, to determine what is a large value we have to obtain the distribution of the test statistic $S_n(\widehat{\theta})$ under the null hypothesis H_0, or at least an approximation to it. With this aim, we next derive its asymptotic null distribution. We will assume that the estimator $\widehat{\theta} = (\widehat{a}, \widehat{b})$ satisfies the following regularity condition.

Assumption 1. Under H_0, if $\theta = (a, b) \in \Theta$ denotes the true parameter value, then $\sqrt{n}(\widehat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell(X_i;\theta) + o_P(1)$, with $\mathbb{E}_\theta\{\ell(X_i;\theta)\} = 0$ and $J(\theta) = \mathbb{E}_\theta\{\ell(X_i;\theta)^T \ell(X_i;\theta)\} < \infty$.

Assumption 1 implies that when the null hypothesis is true and θ denotes the true parameter value, then $\sqrt{n}(\widehat{\theta} - \theta)$ is asymptotically normally distributed. This assumption is not restrictive at all since it is fulfilled by commonly used estimators such as the the ML estimator and the moment estimator (see White [51] and Jiménez-Gamero and Kim [24], among others). In Appendix B, the form of the function ℓ is provided for the aforementioned estimators under the BT family of distributions, and we show that the conditions of Assumption 1 really holds for them.

The next theorem gives the asymptotic null distribution of $S_n(\widehat{\theta})$.

Theorem 3.2. *Let $X_1, \ldots, X_n$ be independent and identically distributed from $X \sim \mathrm{BT}(\theta)$, where $\theta = (a, b) \in \Theta$. Suppose that $\widehat{\theta}$ satisfies Assumption 1. Then, $nS_n(\widehat{\theta}) \xrightarrow{\mathcal{L}} \|S(\theta)\|_2^2$, where $\{S(\theta) = (S(0;\theta), S(1;\theta), \ldots)\}$ is a centered Gaussian process in l^2 with covariance kernel $\varrho(k, r) = Cov_\theta\{Y(X;k,\theta), Y(X;r,\theta)\}$ for $k \in \mathbb{N}_0$ and $r \in \mathbb{N}_0$, $Y(X;k,\theta) = \phi(X;k,\theta) + (\mu_1(k;\theta), \mu_2(k;\theta))\ell(X;\theta)^T$, ϕ is defined in* (3.1), *$\mu_1(k;\theta) = \mathbb{E}_\theta\{(\partial/\partial a)\phi(X;k,\theta)\}$, and $\mu_2(k;\theta) = \mathbb{E}_\theta\{(\partial/\partial b)\phi(X;k,\theta)\}$.*

Remark 3.1. If someone specifies the function ℓ for a specific estimator, then the covariance kernel appeared in the statement of the previous theorem can be given explicitly since one has just to calculate an expectation. For the BT family of distributions, when the moment estimators are used, we have proved in Appendix B that the function ℓ can be obtained in a closed, but rather complicated, form. On the other hand, when the ML estimators are used, the function ℓ cannot be obtained in a closed form. For the previous reasons, we did not provide the form of the covariance kernel $\varrho(k, r)$ for the aforementioned estimators.

Note that the null distribution of $\|S(\theta)\|_2^2$ is that of $\sum_{j\geq 1} \lambda_j \, \chi_{1j}^2$, where χ_{11}^2, χ_{12}^2, $\ldots$ are independent χ^2 variates with one degree of freedom, and the set $\{\lambda_j\}$ are the positive eigenvalues of the linear operator $f \mapsto \mathcal{C}f$ on l^2 associated with the kernel ϱ

given in Theorem 3.2; that is, $(\mathcal{C}f)(k) = \sum_{r\geq 0} \varrho(r,k)f(r)$. Since these eigenvalues depend on the unknown θ, it is evident that the asymptotic null distribution of the test statistic $nS_n(\widehat{\theta})$ depends on the unknown true value of the parameter vector $\theta = (a, b)$. However, even if θ was known or replaced by an appropriate estimator $\widehat{\theta}$, to determine the eigenvalues of an operator is a quite hard problem and unfortunately we did not succeed in finding explicit expressions for such eigenvalues. For similar problems and arguments see Novoa-Muñoz and Jiménez-Gamero [41] and Jiménez-Gamero and Alba-Fernandez [22], among others. Based on the previous remarks, it is concluded that the asymptotic null distribution of $nS_n(\widehat{\theta})$ given in Theorem 3.2 does not provide a useful approximation to its null distribution. Therefore, one should find another way of approximating the null distribution of the test statistic $nS_n(\widehat{\theta})$.

A common approach is to consider a parametric bootstrap approach to estimate the null distribution of $\|S(\theta)\|_2^2$. In the sequel, the parametric bootstrap approach is defined. Given the data X_1, ..., X_n, let X_1^*, ..., X_n^* be independent and identically distributed from $X^* \sim \mathrm{BT}(\widehat{\theta})$. Let $S_n^*(\widehat{\theta}^*)$ be the bootstrap version of $S_n(\widehat{\theta})$ obtained by replacing $X_1, \ldots, X_n$ and $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n)$ with $X_1^*, \ldots, X_n^*$ and $\widehat{\theta}^* = \widehat{\theta}(X_1^*, \ldots, X_n^*)$, respectively, in the expression of $S_n(\widehat{\theta})$. Then, we approximate $P_\theta\{S_n(\widehat{\theta}) \leq x\}$ by means of its bootstrap version, i.e. $P_*\{S_n^*(\widehat{\theta}^*) \leq x\}$. In order to show that the parametric bootstrap consistently approximates the null distribution of $S_n(\widehat{\theta})$, we need the following assumption, which is a bit stronger than Assumption 1.

Assumption 2. Assumption 1 holds, and the functions $\ell(X;\theta)$ and $J(\theta)$ satisfy:

(1) $\sup_{\vartheta\in\Delta} \mathbb{E}_\vartheta \left\{\|\ell(X;\vartheta)\|^2 I\left(\|\ell(X;\vartheta)\| > \epsilon\sqrt{n}\right)\right\} \longrightarrow 0$, $\forall\, \epsilon > 0$, where $\Delta \subseteq \Theta$ is an open neighborhood of θ.

(2) $\ell(X;\vartheta)$ and $J(\vartheta)$ are continuous as functions of ϑ at $\vartheta = \theta$.

Theorem 3.3. *Let $X_1, \ldots, X_n$ be independent and identically distributed from X, a random variable taking values in $\mathbb{N}_0$. Assume that $\widehat{\theta} \overset{a.s.(P)}{\longrightarrow} \theta$, for some $\theta \in \Theta$, and that Assumption 2 holds. Then,* $\sup_{x\in\mathbb{R}} \left| P_*\{S_n^*(\widehat{\theta}^*) \leq x\} - P_\theta\{S_n(\widehat{\theta}) \leq x\}\right| \overset{a.s.(P)}{\longrightarrow} 0.$

Theorem 3.3 holds whether H_0 is true or not. It states that the conditional distribution of $S_n^*(\widehat{\theta}^*)$ and the distribution of $S_n(\widehat{\theta})$ are close when the sample is drawn from a population with $\mathrm{BT}(\theta)$ distribution, $\theta = (a, b)$ being the limit of $\widehat{\theta} = (\widehat{a}, \widehat{b})$. In particular, if the null hypothesis H_0 is true, then Theorem 3.3 states that the conditional distribution of $S_n^*(\widehat{\theta}^*)$ is close to the null distribution of $S_n(\widehat{\theta})$. Let $\alpha \in (0, 1)$. Hence, the test function

$$\Psi^* = \begin{cases} 1, & \text{if } S_n(\widehat{\theta}) \geq s_{n,\alpha}^*, \\ 0, & \text{otherwise,} \end{cases}$$

or, equivalently, the test that rejects H_0 when $p^* = P_*\{S_n^*(\widehat{\theta}^*) \geq S_{obs}\} \leq \alpha$, is asymptotically correct in the sense that when H_0 is true, $\lim_{n\to\infty} P_\theta(\Psi^* = 1) = \alpha$, where

$s^*_{n,\alpha} = \inf\{x : P_*(S^*_n(\widehat{\theta}^*) \geq x) \leq \alpha\}$ is the α upper percentile of the bootstrap distribution of $S_n(\widehat{\theta})$, and S_{obs} is the observed value of the test statistic obtained from a given dataset. An immediate consequence of Theorem 3.1 and Theorem 3.3 is that the test Ψ^* is consistent; that is, it is able to detect any fixed alternative, in the sense that $\Pr(\Psi^* = 1) \to 1$ whenever $X \nsim \mathrm{BT}(\theta)$, for any $\theta \in \Theta$.

Remark 3.2. A parametric bootstrap estimator of the null distribution of $nS_n(\widehat{\theta})$ was previously discussed. As observed before, the most important difficulty with the distribution of $\|S(\theta)\|_2^2$ is the determination of the positive eigenvalues λ_j which, however, can be consistently (a.s.) approximated following Dehling and Mikosch [11]. In this context, another solution is to approximate the null distribution of $nS_n(\widehat{\theta})$ through the conditional distribution, given $X_1, \ldots, X_n$, of $\sum_{j\geq 1} \widehat{\lambda}_j\, \chi^2_{1j}$, where $\chi^2_{11}, \chi^2_{12}, \ldots$ are independent χ^2 variates with one degree of freedom and $\widehat{\lambda}_j$ is a consistent estimator of the eigenvalue λ_j, by means of weighted bootstrap in the sense of Burke [6] (see also, for instance, Kojadinovic and Yan [30] and references therein). From a computational point of view, the weighted bootstrap is more efficient than the parametric bootstrap. On the other hand, it has the disadvantage that one needs to estimate the function ℓ (see, for instance, Jiménez-Gamero and Kim [24]). In this paper, we rely on parametric bootstrap similar to the existing gof tests described in Section 2.

Before closing this section, we have to note that the bootstrap p-value of any of the five tests, namely $S_n(\widehat{\theta})$, $R_{n,w}(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$, H_n and W_n cannot be exactly computed. In the sequel, let T denote any of the five test statistics and let T_{obs} stand for the observed value of such statistic. Then, the bootstrap p-value can be approximated as follows:

1. Calculate the observed values of the gof test statistics for the available dataset $X_1, \ldots, X_n$, say $S_{obs}(\widehat{\theta})$, $M_{obs}(\widehat{\theta})$, $R_{obs}(\widehat{\theta})$, H_{obs} and W_{obs}.
2. Generate B bootstrap samples $X_1^{*v}, \ldots, X_n^{*v}$ from $X^* \sim \mathrm{BT}(\widehat{\theta})$, for $v = 1, \ldots, B$.
3. Calculate the test statistics $S_n(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$, $R_{n,w}(\widehat{\theta})$, H_n and W_n for each bootstrap sample and denote them, respectively, by S_v^*, M_v^*, R_v^*, H_v^* and W_v^* for $v = 1, \ldots, B$.
4. Compute the p-values of the tests based on the statistics $S_n(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$, $R_{n,w}(\widehat{\theta})$, H_n and W_n by means, respectively, of the expressions

$$\widehat{p}_S = \frac{\#\{S_v^* \geq S_{obs}(\widehat{\theta})\}}{B},\ \widehat{p}_M = \frac{\#\{M_v^* \geq M_{obs}(\widehat{\theta})\}}{B},\ \widehat{p}_R = \frac{\#\{R_v^* \geq R_{obs}(\widehat{\theta})\}}{B}.$$

$$\widehat{p}_H = \frac{\#\{H_v^* \geq H_{obs}\}}{B},\ \widehat{p}_W = \frac{\#\{W_v^* \geq W_{obs}\}}{B}.$$

For a good discussion of bootstrap p-values, see Efron and Tibshirani [12, Chapter 16].

4. FINITE-SAMPLE SIZE AND POWER PROPERTIES

The properties studied in the previous section related to the test statistic $S_n(\widehat{\theta})$ are asymptotic, which means that they describe the behavior of the proposed test when the sample size is large. In this section, we empirically investigate its performance in small and moderate sample sizes through Monte Carlo simulation experiments. We also include in the Monte Carlo studies the test statistics $R_{n,w}(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$, H_n and W_n for comparison. We have not considered the test statistic $K_n(t;\widehat{\theta})$ in the Monte Carlo experiments since the question on how to select t remains unsolved and its performance depends on different values of t. It is worth stressing that the numerical results regarding the existing gof tests applied in the BT distribution are new, and so it also represents an additional contribution of the current paper in studying the performance of these specific existing gof tests for this two-parameter discrete distribution. All computations were performed by using the R language [43]. In all cases, $10{,}000$ Monte Carlo replications were considered. Without loss of generality, we consider $a = 0.8$ and 1.4, and $b = 0.6, 1.2$ and 1.8.

The computation of the test statistics $R_{n,w}(\widehat{\theta})$ and $M_{n,w}(\widehat{\theta})$ depend on the weight function $w(t)$ in their computations. Here, we consider the weight function in the form $w(t) = t^{\gamma}$, where $t \in (0,1)$ and $\gamma = 0, 1, 2, 5$ and 10. It is interesting to note that $\gamma = 0$ corresponds to the probability density function of the uniform distribution on $(0,1)$ as a weight function. The resulting tests when $w(t) = t^{\gamma}$ is used as a weight function will be denoted by $R_{n,\gamma}(\widehat{\theta})$ and $M_{n,\gamma}(\widehat{\theta})$. In particular, we have that

$$R_{n,\gamma}(\widehat{\theta}) = \sum_{r,k=0}^{\infty} \frac{\{p(r;\theta) - \widehat{p}(r)\}\{p(k;\theta) - \widehat{p}(k)\}}{r+k+\gamma+1},$$

and

$$M_{n,\gamma}(\widehat{\theta}) = \sum_{r,k=0}^{\infty} \frac{\widehat{d}(r;\widehat{\theta})\widehat{d}(k;\widehat{\theta})}{r+k+\gamma+1}.$$

It should be emphasized that the test statistics $S_n(\widehat{\theta})$, $R_{n,\gamma}(\widehat{\theta})$ and $M_{n,\gamma}(\widehat{\theta})$ are defined by means of infinite sums and, hence, these sums have to be truncated at some finite value, say M. We have noted that $M = 20$ yields sufficiently precise values of these test statistics.

Random variates from $\mathrm{BT}(\theta)$ distribution were generated by following Proposition 9 and Remark 13 in Castellares et al. [8]. To estimate $\theta = (a,b)$, we considered the ML method. Finally, we adopted the warp-speed method [16] for evaluating the proposed resampling scheme to reduce the computational burden. On the basis of the warp-speed method, instead of computing critical points for each Monte Carlo sample, one resample is generated for each Monte Carlo sample and each test statistic, say T, is computed for that sample, obtaining say T^*. Then, the resampling critical values for T are computed from the empirical distribution determined by the resampling repetitions of T^*. It is worth mentioning that the idea behind the warp-speed bootstrap method is that taking just *one* bootstrap draw for each simulated sample is sufficient to provide a useful approximation to the statistic of interest. Applying this insight to Monte Carlo evaluation of bootstrap-based tests yields evaluation methods that work with $B = 1$ [16]. Because of the resulting dramatic computational savings, Giacomini et al. [16] called their method as "Warp-Speed" Monte Carlo method.

4.1. Size properties

First, the type I error of the gof tests based on the statistics $R_{n,\gamma}(\widehat{\theta})$, $M_{n,\gamma}(\widehat{\theta})$, H_n, W_n and $S_n(\widehat{\theta})$ are investigated. We consider the sample sizes $n = 50, 70, 90$ and 150. The nominal levels of the tests are $\alpha = 0.10$ and 0.05. We report the null rejection rates of $H_0 : X \sim \mathrm{BT}(\theta)$ for all the tests at the 10% and 5% nominal significance levels; i.e. the percentage of times that the corresponding statistics exceed the 10% and 5% upper points obtained from the reference distribution generated by parametric bootstrap. These rates estimate the type I error probability of the tests. The null rejection rates of the gof tests $R_{n,\gamma}(\widehat{\theta})$ and $M_{n,\gamma}(\widehat{\theta})$ are listed in Table 1, while Table 2 lists the null rejection rates of the gof tests $S_n(\widehat{\theta})$, H_n and W_n.

For $\gamma = 0$ (i.e., the weight function $w(t)$ corresponds to the probability density function of the uniform distribution on the unit interval), note that the gof tests based on the statistics $R_{n,0}(\widehat{\theta})$ and $M_{n,0}(\widehat{\theta})$ have not a good performance, mainly for small sample sizes and when the parameter a is less than 1 ($a < 1$). On the other hand, the performance of these gof tests improves considerably as γ increases for $a < 1$. It is also evident that values of γ greater than 5 have no effect on improving the performance of the gof tests based on the statistics $R_{n,\gamma}(\widehat{\theta})$ and $M_{n,\gamma}(\widehat{\theta})$ in such a case; compare the null rejection rates of the tests for $\gamma = 5$ and $\gamma = 10$ when $a < 1$. Hence, for $a < 1$, the weight function $w(t) = t^\gamma$ with $\gamma = 5$ seems to be a good choice for the test statistics $R_{n,\gamma}(\widehat{\theta})$ and $M_{n,\gamma}(\widehat{\theta})$ in the BT discrete distribution. It is interesting to note that the gof tests that use $R_{n,0}(\widehat{\theta})$ and $M_{n,0}(\widehat{\theta})$ as test statistics present better results when $a > 1$. However, the performance of these gof tests deteriorates as γ increases and when $a > 1$, and so the probability density function of the uniform distribution on the unit interval as weight function in such a case seems to be a good choice for these test statistics. In short, the numerical results in Table 1 reveals the difficulty of selecting the best value of γ for the gof tests based on the test statistics $R_{n,\gamma}(\widehat{\theta})$ and $M_{n,\gamma}(\widehat{\theta})$. From Table 2, note that the null rejection rates of the gof tests that use H_n and W_n as test statistics are close to the significance levels considered. It is worth stressing that the proposed gof test that uses $S_n(\widehat{\theta})$ as test statistic also presents a good performance, mainly for small sample sizes, when compared with the existing gof tests and, hence, can be an interesting alternative to these gof tests.

Table 1: Null rejection rates of the gof tests $R_{n,\gamma} := R_{n,\gamma}(\widehat{\theta})$ and $M_{n,\gamma} := M_{n,\gamma}(\widehat{\theta})$ for some weight functions $w(t)$.

		$a = 0.8$ and $b = 0.6$									
α	n	$R_{n,0}$	$R_{n,1}$	$R_{n,2}$	$R_{n,5}$	$R_{n,10}$	$M_{n,0}$	$M_{n,1}$	$M_{n,2}$	$M_{n,5}$	$M_{n,10}$
0.10	50	.066	.077	.080	.082	.083	.066	.072	.078	.080	.082
	70	.077	.087	.090	.092	.092	.081	.085	.088	.091	.092
	90	.085	.092	.095	.094	.093	.078	.089	.093	.093	.093
	150	.091	.097	.098	.098	.098	.090	.098	.098	.098	.097
0.05	50	.025	.031	.034	.038	.038	.025	.031	.034	.036	.037
	70	.035	.040	.042	.045	.045	.035	.038	.042	.044	.044
	90	.036	.039	.041	.042	.042	.035	.037	.040	.041	.041
	150	.037	.040	.042	.042	.043	.041	.041	.043	.043	.043
		$a = 0.8$ and $b = 1.2$									
α	n	$R_{n,0}$	$R_{n,1}$	$R_{n,2}$	$R_{n,5}$	$R_{n,10}$	$M_{n,0}$	$M_{n,1}$	$M_{n,2}$	$M_{n,5}$	$M_{n,10}$
0.10	50	.060	.063	.069	.078	.086	.078	.073	.074	.080	.087
	70	.066	.075	.086	.089	.092	.083	.078	.084	.089	.093
	90	.066	.075	.086	.095	.099	.085	.083	.088	.097	.101
	150	.073	.078	.085	.090	.096	.086	.084	.086	.094	.096
0.05	50	.025	.027	.030	.036	.039	.033	.030	.033	.036	.038
	70	.028	.034	.036	.042	.043	.038	.039	.040	.042	.043
	90	.025	.031	.036	.041	.044	.038	.035	.039	.042	.045
	150	.028	.035	.038	.040	.041	.040	.040	.042	.042	.041
		$a = 1.4$ and $b = 1.8$									
α	n	$R_{n,0}$	$R_{n,1}$	$R_{n,2}$	$R_{n,5}$	$R_{n,10}$	$M_{n,0}$	$M_{n,1}$	$M_{n,2}$	$M_{n,5}$	$M_{n,10}$
0.10	50	.098	.086	.085	.071	.074	.088	.082	.080	.082	.085
	70	.094	.091	.084	.073	.078	.089	.082	.078	.078	.081
	90	.096	.093	.087	.078	.081	.091	.084	.079	.078	.079
	150	.106	.100	.091	.080	.080	.098	.092	.088	.086	.090
0.05	50	.046	.041	.038	.035	.036	.042	.039	.036	.035	.035
	70	.047	.046	.041	.040	.041	.043	.039	.038	.034	.035
	90	.047	.045	.039	.035	.039	.041	.038	.037	.036	.037
	150	.050	.049	.044	.036	.038	.049	.042	.041	.041	.043

Table 2: Null rejection rates of the gof tests H_n, W_n and $S_n := S_n(\widehat{\theta})$.

		$a = 0.8$ and $b = 1.2$			$a = 0.8$ and $b = 1.2$			$a = 0.8$ and $b = 1.2$		
α	n	H_n	W_n	S_n	H_n	W_n	S_n	H_n	W_n	S_n
0.10	50	.103	.098	.079	.101	.107	.083	.095	.099	.091
	70	.095	.099	.084	.097	.099	.090	.101	.103	.086
	90	.095	.099	.082	.110	.110	.087	.105	.111	.085
	150	.101	.099	.089	.100	.098	.094	.098	.105	.090
0.05	50	.049	.045	.036	.051	.052	.040	.050	.050	.041
	70	.047	.049	.037	.050	.050	.042	.050	.052	.041
	90	.048	.045	.037	.054	.054	.040	.054	.057	.039
	150	.047	.045	.043	.049	.048	.042	.048	.052	.045

4.2. Power properties

Next, the power of the tests based on the statistics $R_{n,\gamma}(\widehat{\theta})$, $M_{n,\gamma}(\widehat{\theta})$, $S_n(\widehat{\theta})$, H_n and W_n are investigated. To compute the powers of the tests, we carried out Monte Carlo simulation experiments similar to that described above, however, the data were generated from perturbed BT distributions, and from the geometric (Geo), binomial (Bin), discrete Weibull (dWei) and negative binomial (NB) distributions. We consider two kinds of perturbations for the BT distribution. Let $X_1 \sim \mathrm{BT}(\theta)$ and X_2 be another random variable taking values on $\mathbb{N}_0$, not having a BT distribution and independent of X_1. Then, the random variables $X_1 + X_2$ and $\max\{X_1, X_2\}$ also take values on $\mathbb{N}_0$, but the corresponding distributions of these perturbed random variables do not belong to the BT family of distributions and, hence, they can be used as alternatives. In the Monte Carlo simulations, we consider X_2 as a discrete uniform random variable taking values on $\{0, 1, \ldots, k\}$, for $k = 2, 4$ and 5, being denoted as $dU2$, $dU4$ and $dU5$, respectively. Thus, we have the following alternative distributions: Alt1 $= X_1 + dU2$, Alt2 $= \max\{X_1, dU2\}$, Alt3 $= X_1 + dU4$, Alt4 $= \max\{X_1, dU4\}$, Alt5 $= X_1 + dU5$ and Alt6 $= \max\{X_1, dU5\}$.

Here, we consider $w(t) = t^\gamma$ with $\gamma = 0, 2, 5$, $n = 60, 80$, and $a = 0.8$ and $b = 0.6$. The Monte Carlo simulation results regarding the power of the gof tests $R_{n,w}(\widehat{\theta})$ and $M_{n,w}(\widehat{\theta})$ are listed in Table 3, and Table 4 lists the power results of the gof tests $S_n(\widehat{\theta})$, H_n and W_n. From Table 3, note that there is no great difference in powers when different weight functions are considered. It is interesting to note that the test based on the proposed statistic $S_n(\widehat{\theta})$ is the most powerful among the gof tests in the great majority of the cases; compare Tables 3 and 4. However, it is evident that no gof test provides the highest power against all alternatives; that is, for some alternative distributions, the new gof test exhibits the highest power, but for other ones, the existing gof tests yield greater power. In summary, there is no uniform superiority of one gof test with respect to the others, as expected from the theoretical results in [20]. As expected, as the sample size increases, the power of the tests increases. In short, the numerical results of this section reveal that the proposed gof test on the basis of the new statistic $S_n(\widehat{\theta})$ can be an interesting alternative to the existing gof tests based on the test statistics $R_{n,w}(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$, H_n and W_n. The main advantage of the test statistic $S_n(\widehat{\theta})$ in relation to the test statistics $R_{n,w}(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$ is that it is not necessary to consider a weight function for its computation. On the other hand, we have to truncate an infinite sum in a finite value to calculate the new test statistic.

Finally, we compute the powers of the gof tests by considering moment estimators. Castellares et al. [8] have provided the following moment estimators for a and b: $\widetilde{a} = \frac{s^2}{\bar{X}} - 1$, $\widetilde{b} = \frac{\bar{X}\exp(1 - s^2/\bar{X})}{s^2/\bar{X} - 1}$, where $\bar{X}$ and s^2 are the sample mean and standard deviation. Castellares et al. [8] proved that $\widetilde{a}$ and $\widetilde{b}$ are consistent estimators for a and b, respectively. The power results when using these estimators are presented in Tables 5 and 6. Note that the powers of the gof tests under the moment estimates are near the powers under the ML estimates. However, the powers under the ML estimates are in general greater than the ones under the moment estimates.

Table 3: Nonnull rejection rates of $R_{n,w}(\widehat{\theta})$ and $M_{n,w}(\widehat{\theta})$ for some weight functions $w(t)$: power.

	$n = 60$				$n = 80$			
	$R_{n,0}(\widehat{\theta})$		$M_{n,0}(\widehat{\theta})$		$R_{n,0}(\widehat{\theta})$		$M_{n,0}(\widehat{\theta})$	
Alternative	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Alt1	0.248	0.189	0.310	0.220	0.252	0.180	0.302	0.227
Alt2	0.842	0.787	0.861	0.812	0.885	0.847	0.900	0.867
Alt3	0.276	0.162	0.446	0.303	0.321	0.199	0.506	0.390
Alt4	0.947	0.931	0.970	0.952	0.971	0.959	0.988	0.976
Alt5	0.287	0.125	0.496	0.325	0.384	0.225	0.597	0.469
Alt6	0.870	0.798	0.936	0.880	0.924	0.867	0.972	0.943
Geo	0.680	0.551	0.658	0.521	0.779	0.713	0.764	0.689
Bin	0.785	0.780	0.800	0.784	0.802	0.789	0.825	0.806
dWei	0.952	0.922	0.961	0.935	0.969	0.955	0.974	0.962
NB	0.395	0.257	0.398	0.257	0.484	0.379	0.480	0.377
	$R_{n,2}(\widehat{\theta})$		$M_{n,2}(\widehat{\theta})$		$R_{n,2}(\widehat{\theta})$		$M_{n,2}(\widehat{\theta})$	
Alternative	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Alt1	0.291	0.218	0.340	0.240	0.294	0.218	0.339	0.257
Alt2	0.878	0.810	0.900	0.835	0.927	0.879	0.945	0.903
Alt3	0.344	0.178	0.493	0.308	0.406	0.254	0.567	0.428
Alt4	0.945	0.926	0.964	0.939	0.969	0.955	0.982	0.969
Alt5	0.424	0.181	0.611	0.395	0.551	0.357	0.712	0.590
Alt6	0.863	0.770	0.915	0.842	0.918	0.859	0.950	0.921
Geo	0.719	0.581	0.727	0.590	0.811	0.741	0.818	0.750
Bin	0.788	0.784	0.794	0.785	0.802	0.793	0.814	0.798
dWei	0.998	0.998	0.065	0.955	0.974	0.966	0.988	0.988
NB	0.375	0.217	0.418	0.217	0.464	0.359	0.500	0.387
	$R_{n,5}(\widehat{\theta})$		$M_{n,5}(\widehat{\theta})$		$R_{n,5}(\widehat{\theta})$		$M_{n,5}(\widehat{\theta})$	
Alternative	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Alt1	0.297	0.209	0.332	0.226	0.300	0.214	0.337	0.243
Alt2	0.881	0.803	0.897	0.821	0.933	0.876	0.946	0.894
Alt3	0.394	0.195	0.517	0.293	0.478	0.290	0.592	0.427
Alt4	0.929	0.907	0.943	0.914	0.954	0.937	0.963	0.947
Alt5	0.541	0.252	0.678	0.436	0.674	0.477	0.779	0.647
Alt6	0.892	0.783	0.932	0.839	0.944	0.883	0.966	0.931
Geo	0.725	0.585	0.739	0.600	0.816	0.741	0.822	0.752
Bin	0.779	0.775	0.781	0.775	0.792	0.782	0.797	0.784
dWei	0.999	0.998	0.998	0.998	0.999	0.981	0.999	0.994
NB	0.386	0.228	0.429	0.238	0.475	0.350	0.491	0.338

5. REAL DATA ILLUSTRATIONS

In this section, we apply the gof tests based on the test statistics $R_{n,w}(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$, $S_n(\widehat{\theta})$, H_n and W_n in some real datasets for the sake of illustration. We consider the weight function $w(t) = t^\gamma$ with $\gamma = 5$ to compute the test statistics $R_{n,w}(\widehat{\theta})$ and $M_{n,w}(\widehat{\theta})$. All computations were done using the R language [43]. The code used in the real data applications can be obtained from the authors upon request. The datasets

Table 4: Nonnull rejection rates of $S_n(\widehat{\theta})$, H_n and W_n: power.

		$S_n(\widehat{\theta})$		H_n		W_n	
n	Alternative	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$
60	Alt1	0.720	0.654	0.540	0.406	0.380	0.267
	Alt2	0.982	0.973	0.985	0.967	0.948	0.871
	Alt3	0.944	0.929	0.564	0.408	0.422	0.238
	Alt4	0.999	0.999	0.992	0.998	0.981	0.932
	Alt5	0.999	0.998	0.678	0.494	0.515	0.286
	Alt6	0.999	0.999	0.999	0.983	0.939	0.880
	Geo	0.919	0.826	0.557	0.426	0.648	0.512
	Bin	0.830	0.736	0.794	0.735	0.790	0.767
	dWei	0.928	0.907	0.997	0.952	0.999	0.999
	NB	0.682	0.546	0.378	0.252	0.375	0.233
80	Alt1	0.783	0.648	0.571	0.462	0.372	0.306
	Alt2	0.999	0.992	0.999	0.994	0.981	0.962
	Alt3	0.969	0.919	0.648	0.475	0.478	0.352
	Alt4	0.999	0.999	0.999	0.995	0.999	0.986
	Alt5	0.999	0.992	0.758	0.610	0.606	0.445
	Alt6	0.999	0.999	0.999	0.999	0.996	0.941
	Geo	0.939	0.890	0.646	0.497	0.774	0.674
	Bin	0.875	0.856	0.867	0.802	0.822	0.799
	dWei	0.983	0.909	0.999	0.996	0.999	0.999
	NB	0.734	0.596	0.440	0.303	0.458	0.345

we consider correspond to the number of chromatid aberrations in 24 hours [9, 10], the number of absences of workers in a particular division of a large steel corporation in an observational period of six months [47], the number of claims of automobile liability policies [28, pp. 244], and the number of hemocytometer yeast cell on European red mites on apple leaves [44]. Descriptive measures for these datasets are listed in Table 7. The ML estimates of the BT distribution parameters, asymptotic standard errors (SE), and the 90% confidence intervals (CI) for the model parameters for each dataset are presented in Table 8. Table 9 lists the bootstrap p-values (with $B = 5000$) of the gof tests on the basis of the test statistics $R_{n,w}(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$, $S_n(\widehat{\theta})$, H_n and W_n for testing gof to the BT distribution. It can be noted that the five gof tests agree that the two-parameter BT discrete distribution is not adequate for fitting the chromatid dataset, once the bootstrap p-value for all tests are < 0.01. In addition, the five gof tests agree that the BT distribution is adequate for fitting the absence data, claims data, and cell data; that is, the five tests agree that the null hypothesis cannot be rejected at any usual significance levels.

A referee reminds us that the dataset regarding the absences of workers [47] was originally fitted with the Negative Binomial (NB) distribution. From Table 9, it is evident that the BT distribution (i.e., the NTA distribution) is not rejected by any of the gof tests, and so an interesting question is: which distribution fits better this dataset, BT or NB? The pmf of the two-parameter NB distribution, specified in terms of its mean, μ say, is

Table 5: Nonnull rejection rates of $R_{n,w}(\widetilde{\theta})$ and $M_{n,w}(\widetilde{\theta})$ for some weight functions $w(t)$: power under moment estimators.

	$n=60$				$n=80$			
	$R_{n,0}(\widetilde{\theta})$		$M_{n,0}(\widetilde{\theta})$		$R_{n,0}(\widetilde{\theta})$		$M_{n,0}(\widetilde{\theta})$	
Alternative	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Alt1	0.242	0.183	0.304	0.214	0.246	0.174	0.296	0.221
Alt2	0.811	0.756	0.830	0.781	0.854	0.816	0.869	0.836
Alt3	0.257	0.143	0.427	0.284	0.302	0.180	0.487	0.371
Alt4	0.909	0.893	0.932	0.914	0.933	0.921	0.950	0.938
Alt5	0.266	0.104	0.475	0.304	0.363	0.204	0.576	0.448
Alt6	0.822	0.750	0.888	0.832	0.876	0.819	0.924	0.895
Geo	0.672	0.543	0.650	0.513	0.771	0.705	0.756	0.681
Bin	0.745	0.740	0.760	0.744	0.762	0.749	0.785	0.766
dWei	0.948	0.918	0.957	0.931	0.965	0.951	0.970	0.958
NB	0.392	0.254	0.395	0.254	0.481	0.376	0.477	0.374
	$R_{n,2}(\widetilde{\theta})$		$M_{n,2}(\widetilde{\theta})$		$R_{n,2}(\widetilde{\theta})$		$M_{n,2}(\widetilde{\theta})$	
Alternative	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Alt1	0.265	0.192	0.314	0.214	0.268	0.192	0.313	0.231
Alt2	0.831	0.763	0.853	0.788	0.880	0.832	0.898	0.856
Alt3	0.322	0.156	0.471	0.286	0.384	0.232	0.545	0.406
Alt4	0.912	0.893	0.931	0.906	0.936	0.922	0.949	0.936
Alt5	0.408	0.165	0.595	0.379	0.535	0.341	0.696	0.574
Alt6	0.859	0.766	0.911	0.838	0.914	0.855	0.946	0.917
Geo	0.693	0.555	0.701	0.564	0.785	0.715	0.792	0.724
Bin	0.743	0.739	0.749	0.740	0.757	0.748	0.769	0.753
dWei	0.978	0.961	0.983	0.970	0.984	0.975	0.988	0.982
NB	0.372	0.214	0.415	0.214	0.461	0.356	0.497	0.384
	$R_{n,5}(\widetilde{\theta})$		$M_{n,5}(\widetilde{\theta})$		$R_{n,5}(\widetilde{\theta})$		$M_{n,5}(\widetilde{\theta})$	
Alternative	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Alt1	0.287	0.199	0.322	0.216	0.290	0.204	0.327	0.233
Alt2	0.843	0.765	0.859	0.783	0.895	0.838	0.908	0.856
Alt3	0.366	0.167	0.489	0.265	0.450	0.262	0.564	0.399
Alt4	0.914	0.892	0.928	0.899	0.939	0.922	0.948	0.932
Alt5	0.505	0.216	0.642	0.400	0.638	0.441	0.743	0.611
Alt6	0.879	0.770	0.919	0.826	0.931	0.870	0.953	0.918
Geo	0.708	0.568	0.722	0.583	0.799	0.724	0.805	0.735
Bin	0.743	0.739	0.745	0.739	0.756	0.746	0.761	0.748
dWei	0.989	0.979	0.992	0.984	0.991	0.986	0.994	0.990
NB	0.382	0.224	0.425	0.234	0.471	0.346	0.487	0.334

given by

$$\Pr(Y=y)=\left(\frac{\varphi}{\varphi+\mu}\right)^{\varphi}\left(\frac{\mu}{\varphi+\mu}\right)^{y}\frac{\Gamma(y+\varphi)}{\Gamma(\varphi)\Gamma(y+1)},\quad y=0,1,2,\ldots,$$

where $\Gamma(\cdot)$ is the gamma function, and $\mu>0$ and $\varphi>0$. It can be shown that the variance can be written as $\mu+\mu^2/\varphi$ and hence the parameter φ is referred to as the "dispersion parameter". The ML estimates of μ and φ are (asymptotic SE between parentheses): $\widehat{\mu}=0.6698(0.0754)$ and $\widehat{\varphi}=0.3951(0.0752)$. The maximized log-likelihood function

Table 6: Nonnull rejection rates of $S_n(\widetilde{\theta})$, H_n and W_n: power under moment estimators.

		$S_n(\widetilde{\theta})$		H_n		W_n	
n	Alternative	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$
60	Alt1	0.706	0.605	0.490	0.392	0.331	0.217
	Alt2	0.961	0.941	0.967	0.946	0.916	0.853
	Alt3	0.942	0.901	0.552	0.406	0.394	0.226
	Alt4	0.997	0.994	0.981	0.969	0.954	0.921
	Alt5	0.980	0.960	0.634	0.475	0.477	0.242
	Alt6	0.998	0.996	0.977	0.953	0.922	0.836
	Geo	0.898	0.822	0.551	0.405	0.644	0.506
	Bin	0.794	0.700	0.769	0.699	0.754	0.742
	dWei	0.917	0.862	0.956	0.941	0.990	0.982
	NB	0.676	0.540	0.372	0.246	0.369	0.227
80	Alt1	0.733	0.634	0.522	0.412	0.358	0.257
	Alt2	0.982	0.971	0.986	0.976	0.960	0.930
	Alt3	0.957	0.917	0.620	0.463	0.476	0.324
	Alt4	0.999	0.998	0.991	0.984	0.975	0.959
	Alt5	0.987	0.973	0.720	0.566	0.587	0.407
	Alt6	0.999	0.999	0.991	0.978	0.966	0.924
	Geo	0.933	0.869	0.642	0.491	0.753	0.670
	Bin	0.850	0.820	0.831	0.777	0.786	0.763
	dWei	0.942	0.898	0.967	0.955	0.992	0.988
	NB	0.730	0.592	0.436	0.299	0.454	0.341

Table 7: Descriptive measures.

	Chromatid	Absence	Claims	Cell
n	400	318	298	80
Mean ($\bar{x}$)	0.55	0.67	1.71	1.15
Variance (s^2)	1.13	1.53	3.67	2.10
Skewness	3.12	2.19	1.72	1.27
Kurtosis	15.68	7.72	6.90	3.96
CV	1.94	1.85	1.12	1.26
ID	2.05	2.29	2.15	1.83

CV: Coefficient of variation ($= s/\bar{x}$); ID: Index of dispersion ($= s^2/\bar{x}$).

for the NB distribution is -347.95, and so the AIC is given by 699.89. The maximized log-likelihood function for the BT distribution is given by -345.60, which results in an AIC value of 695.20. On the basis of the AIC values, it seems that the two-parameter BT distribution fits better the absences of workers' data than the two-parameter NB distribution and, hence, should be preferred.

Finally, it is well-known that the NTA distribution is traditionally fitted to datasets from ecology, entomology, etc. For example, McGuire et al. [34] studied the distribution of larval populations of the European corn borer, *Pyrausta nubilalis* (Hbni.). A total of $n = 3205$ corn plants growing in an area located in Northwest Iowa were dissected and, hence, the data correspond to the number of borers per plant dissected; see Table 1

Table 8: ML estimates.

	Chromatid aberrations		
Parameter	ML estimate	SE	90% CI
a	0.6453	0.1112	(0.4630; 0.8277)
b	0.4450	0.1201	(0.2480; 0.6420)
	Absence proneness		
Parameter	ML estimate	SE	90% CI
a	1.2320	0.1589	(0.9714; 1.4926)
b	0.1586	0.0427	(0.0886; 0.2286)
	Claims of automobile		
Parameter	ML estimate	SE	90% CI
a	0.9795	0.1342	(0.7594; 1.1995)
b	0.6548	0.1728	(0.3714; 0.9382)
	Yeast cell		
Parameter	ML estimate	SE	90% CI
a	0.9340	0.2684	(0.4938; 1.3741)
b	0.4839	0.2596	(0.0582; 0.9096)

Table 9: Bootstrap p-values; $B = 5000$.

Dataset	$R_{n,w}(\widehat{\theta})$	$M_{n,w}(\widehat{\theta})$	$S_n(\widehat{\theta})$	H_n	W_n
Chromatid aberrations	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Absence proneness	0.5220	0.5290	0.1632	0.5540	0.3915
Claims of automobile	0.4614	0.3822	0.3050	0.5935	0.6100
Yeast cell	0.6804	0.6716	0.8694	0.7825	0.6355

in McGuire et al. [34, p. 74]. The ML estimates of the BT distribution parameters are (asymptotic SE between parentheses): $\widehat{a} = 0.2756(0.0325)$ and $\widehat{b} = 7.1346(1.0695)$. The bootstrap p-values (with $B = 5000$) of the gof tests on the basis of the test statistics $R_{n,w}(\widehat{\theta})$, $M_{n,w}(\widehat{\theta})$, $S_n(\widehat{\theta})$, H_n and W_n for testing gof to the BT distribution are given, respectively, by 0.082, 0.098, 0.005, 0.034 and 0.056. Note that the gof tests deliver small p-values, which indicates that the two-parameter BT discrete distribution (i.e., the NTA distribution) seems not adequate for fitting these data. In short, this empirical application illustrates that the NTA distribution, which is quite common in ecology and entomology, should be used with some caution in these areas, since for some cases, as evidenced by the gof tests, it cannot be adequate to fit such datasets. This indeed reveals the importance of gof tests to the BT distribution (i.e., the NTA distribution).

6. CONCLUSIONS

In this paper, a new gof test for the Neyman type A distribution was introduced, which is based on the interesting property that its pgf is the unique pgf satisfying a certain differential equation. The new gof test statistic is a function of the coefficients of the polynomial of the resulting equation when one replaces the pgf with the empirical pgf in the aforementioned differential equation. Also, other four related gof test statistics already introduced in the statistical literature were particularized for the two-parameter Bell-Touchard distribution for the first time, and studied by means of Monte Carlo simulations. We have that these five tests (the four already proposed and the new one) are consistent against fixed alternative hypotheses. Also, the practical computation of p-values of these tests requires a parametric bootstrap approximation to the null distribution of the corresponding test statistics. We consider Monte Carlo simulation experiments to verify the performance of the gof tests in finite samples. The Monte Carlo simulation results indicate that the null rejection rates of the five tests are, in general, close to the nominal levels. In addition, the numerical results regarding the power of the tests reveals that no test provides the highest power against all alternatives considered: for some alternatives the new test exhibits the highest power, but for other ones the competing tests yield greater power. In short, there is no uniform superiority of one test with respect to the others. Finally, it is worth emphasizing that the new test statistic $S_n(\widehat{\theta})$ has no need of choosing a weight function for its computation, unlike the test statistics $R_{n,w}(\widehat{\theta})$ and $M_{n,w}(\widehat{\theta})$, which can be a great advantage in practice. On the other hand, we have to truncate an infinite sum in a finite value to calculate the new test statistic.

A. APPENDIX: Proofs

Here we prove the results provided in the previous sections.

Proof of Proposition 3.1 It can be checked that the pgf of $X \sim \mathrm{BT}(\theta)$ given in (1.3) satisfies the differential equation given in (2.4). Obviously, this part of the proof can also be obtained by the result given by Meintanis [35] since the $\mathrm{BT}(\theta)$ distribution belongs to the compound Poisson family of distributions. Next, we proof that it is the only pgf in G satisfying such differential equation. It is well-known that the solution of the linear differential equation of order one of the form $y' + p(t)y = 0$, where $y = y(t)$, $y' = (\partial/\partial t)y(t)$ and $p(t)$ is a continuous function in t, is given by $y = C\exp(-\int p(t)dt)$, where C is an arbitrary constant. Since the differential equation (2.4) is of this form, we have that $g(t) = C\exp(\int ab\mathrm{e}^{a\,t}dt) = C\exp(b\mathrm{e}^{a\,t})$. Taking into account that g is a pgf, it must satisfy $g(1) = 1$, implying that $C = \exp\left(-b\mathrm{e}^{a}\right)$ and, hence, the desired result is obtained. □

Let $\phi(x;\theta) = (\phi(x;0,\theta), \phi(x;1,\theta), \ldots)$, and $f_r(a,b) = b^r \sum_{u\geq 0}(u+r)\frac{a^u}{u!} = b^r(a+r)\mathrm{e}^{a}$. We have the following lemmas.

Lemma 1.1. *Let $X_1, \ldots, X_n$ be independent and identically distributed from X, a random variable taking values in $\mathbb{N}_0$ with probability mass function $p(k) = \Pr(X = k)$, $k \in \mathbb{N}_0$, so that $\mathbb{E}(X^2) < \infty$. Then, $\mathbb{E}(\|\phi(X;\theta)\|_2^2) \leq \mathbb{E}(X^2) + b^2 f_0^2(a,b) < \infty$, $\forall\, \theta = (a,b) \in \Theta$.*

Proof: By definition,

$$\|\phi(X;\theta)\|_2^2 = \sum_{k\geq 0}(k+1)^2 I(X=k+1) + \sum_{k\geq 0}\sum_{u=0}^{k}\frac{b^2a^{2u+2}}{(u!)^2}I(X=k-u),$$

and, thus, $\mathbb{E}(\|\phi(X;\theta)\|_2^2) = \mathbb{E}(X^2) + \sum_{k\geq 0}\sum_{u=0}^{k}\frac{b^2a^{2u+2}}{(u!)^2}p(k-u)$. To show the finiteness of $\mathbb{E}(\|\phi(X;\theta)\|_2^2)$, we must prove that $\sum_{k\geq 0}\sum_{u=0}^{k}\frac{b^2a^{2u+2}}{(u!)^2}p(k-u) < \infty$. The rest of the proof is parallel with the one in Lemma 1 of Batsidis et al. [3] and for this reason is omitted. □

Let $\frac{\partial}{\partial\theta_i}\widehat{d}(\cdot;\theta) = \left(\frac{\partial}{\partial\theta_i}\widehat{d}(0;\theta), \frac{\partial}{\partial\theta_i}\widehat{d}(1;\theta), \dots\right)$, where $i = 1,2$, and so $\theta_1 := a$ and $\theta_2 := b$.

Lemma 1.2. *Let $X_1,\dots,X_n$ be independent and identically distributed from X, a random variable taking values in $\mathbb{N}_0$. Then, $\forall\,\theta=(a,b)\in\Theta$, we have that*

(I)

$$\left\|\frac{\partial}{\partial\theta_1}\widehat{d}(\cdot;\theta)\right\|_2^2 \leqslant b^2(a+1)^2\mathrm{e}^{2a} = f_1^2(a,b) < \infty,$$

$$\left\|\frac{\partial}{\partial\theta_2}\widehat{d}(\cdot;\theta)\right\|_2^2 \leqslant a^2\mathrm{e}^{2a} = f_0^2(a,b) < \infty.$$

(II)

$$\left\|E\left\{\frac{\partial}{\partial\theta_i}\widehat{d}(\cdot;\theta)\right\}\right\|_2^2 < \infty, \quad i=1,2.$$

Proof: (I) We have that

$$\frac{\partial}{\partial a}\widehat{d}(k;\theta) = -b\sum_{u=0}^{k}\frac{(u+1)a^u}{u!}\widehat{p}(k-u). \tag{1.1}$$

Therefore,

$$\begin{aligned}\left\|\frac{\partial}{\partial a}\widehat{d}(\cdot;\theta)\right\|_2^2 &= b^2\sum_{u,v\geqslant 0}\frac{(u+1)a^u}{u!}\frac{(v+1)a^v}{v!}\sum_{k\geqslant\max\{u,v\}}\widehat{p}(k-u)\widehat{p}(k-v)\\ &\leqslant (b(a+1)\mathrm{e}^a)^2 = f_1^2(a,b) < \infty,\end{aligned}$$

once $\sum_{k\geqslant\max\{u,v\}}\widehat{p}(k-u)\widehat{p}(k-v) \leqslant \sum_{k\geqslant 0}\widehat{p}(k) = 1$ and $\sum_{l\geqslant 0}(l+1)\frac{a^l}{l!} = (a+1)\mathrm{e}^a$. Furthermore, we have that

$$\frac{\partial}{\partial b}\widehat{d}(k;\theta) = -\sum_{u=0}^{k}\frac{a^{u+1}}{u!}\widehat{p}(k-u). \tag{1.2}$$

Therefore,

$$\begin{aligned}\left\|\frac{\partial}{\partial b}\widehat{d}(\cdot;\theta)\right\|_2^2 &= \sum_{u,v\geqslant 0}\frac{a^{u+1}}{u!}\frac{a^{v+1}}{v!}\sum_{k\geqslant\max\{u,v\}}\widehat{p}(k-u)\widehat{p}(k-v)\\ &\leqslant (a\mathrm{e}^a)^2 = f_0^2(a,b)<\infty.\end{aligned}$$

(II) The result follows from part (I) by replacing $\widehat{p}(k-u)$ and $\widehat{p}(k-v)$ with $p(k-u)$ and $p(k-v)$, respectively. □

Lemma 1.3. *Let $X_1,\ldots,X_n$ be independent and identically distributed from X, a random variable taking values in $\mathbb{N}_0$. For each $k\in\mathbb{N}_0$, let $\theta_l=(a_l,b_l)$ so that $\theta_l=\gamma_l\theta+(1-\gamma_l)\widehat{\theta}$, for some $\gamma_l\in[0,1]$. Then,*

$$\sum_{k\geq 0}\left\{\frac{\partial}{\partial\theta_i}\widehat{d}(k;\theta)-\frac{\partial}{\partial\theta_i}\widehat{d}(k;\theta_l)\right\}^2\overset{a.s.(P)}{\longrightarrow}0,\quad i=1,2.$$

Proof: From relation (1.1), and after some algebra, we have that

$$\begin{aligned}\Delta_1 &= \sum_{k\geq 0}\left\{\frac{\partial}{\partial a}\widehat{d}(k;\theta)-\frac{\partial}{\partial a}\widehat{d}(k;\theta_l)\right\}^2\\ &= \sum_{u,v\geqslant 0}\frac{u+1}{u!}(b_la_l^u-ba^u)\frac{v+1}{v!}(b_la_l^v-ba^v)M_1(u,v),\end{aligned}$$

with $0\leq M_1(u,v)=\sum_{k\geq\max\{u,v\}}\widehat{p}(k-u)\widehat{p}(k-v)\leq 1$. By applying the mean value theorem, we have that $b_la_l^u=ba^u+u\widetilde{b}_u\widetilde{a}_u^{u-1}(\widetilde{a}_u-a)+\widetilde{a}_u^u(\widetilde{b}_u-b)$, $\forall u\geq 1$, where $\widetilde{\theta}_u=(\widetilde{a}_u,\widetilde{b}_u)$ with $\widetilde{\theta}_u=\gamma_u\theta_l+(1-\gamma_u)\theta$, for some $\gamma_u\in(0,1)$. Therefore, $\widetilde{a}_u-a=\gamma_u(a_l-a)$ and $\widetilde{b}_u-b=\gamma_u(b_l-a)$. Taking into further consideration that $a_u\leq\max\{a_l,a\}\leq\max\{\widehat{a},a\}:=\widetilde{a}$, $b_u\leq\max\{b_l,b\}\leq\max\{\widehat{b},b\}:=\widetilde{b}$, we have that $\mid b_la_l^u-ba^u\mid\leq u\widetilde{b}\widetilde{a}^{u-1}\mid a_l-a\mid+\widetilde{a}^u\mid b_l-b\mid\leq u\widetilde{b}\widetilde{a}^{u-1}\mid\widehat{a}-a\mid+\widetilde{a}^u\mid\widehat{b}-b\mid$, $\forall u\geq 1$. Similarly, we have that $\mid b_la_l^v-ba^v\mid\leq v\widetilde{b}\widetilde{a}^{v-1}\mid\widehat{a}-a\mid+\widetilde{a}^v\mid\widehat{b}-b\mid,\quad\forall v\geq 1$. From the above considerations. we have that $\mid\Delta_1\mid\leq(\widehat{a}-a)^2(\widetilde{b}(\widetilde{a}+2)\mathrm{e}^{\widetilde{a}})^2+2\mid\widehat{a}-a\mid\mid\widehat{b}-b\mid\widetilde{b}(\widetilde{a}+1)\mathrm{e}^{\widetilde{a}}(\widetilde{a}+2)\mathrm{e}^{\widetilde{a}}+(\widehat{b}-b)^2((\widetilde{a}+1)\mathrm{e}^{\widetilde{a}})^2$. Taking into account that in the right-hand side of the above expression all the functions are continuous functions of θ, it follows that $(\widehat{a}-a)^2(\widetilde{b}(\widetilde{a}+2)\mathrm{e}^{\widetilde{a}})^2\overset{a.s.(P)}{\longrightarrow}(a-a)^2\left(b(a+2)\mathrm{e}^a\right)^2=0$, $\mid\widehat{a}-a\mid\mid\widehat{b}-b\mid\widetilde{b}(\widetilde{a}+1)\mathrm{e}^{\widetilde{a}}(\widetilde{a}+2)\mathrm{e}^{\widetilde{a}}\overset{a.s.(P)}{\longrightarrow}\mid a-a\mid\mid b-b\mid b(a+1)\mathrm{e}^a(a+2)\mathrm{e}^a=0$, $(\widehat{b}-b)^2((\widetilde{a}+1)\mathrm{e}^{\widetilde{a}})^2\overset{a.s.(P)}{\longrightarrow}(b-b)^2\left((a+1)\mathrm{e}^a\right)^2=0$. Thus, $\Delta_1\overset{a.s.(P)}{\longrightarrow}0$.

From relation (1.2), and after some algebra, we have that

$$\begin{aligned}\Delta_2 &= \sum_{k\geq 0}\left\{\frac{\partial}{\partial b}\widehat{d}(k;\theta)-\frac{\partial}{\partial b}\widehat{d}(k;\theta_l)\right\}^2\\ &= \sum_{u,v\geqslant 0}\frac{1}{u!}(a_l^{u+1}-a^{u+1})\frac{1}{v!}(a_l^{v+1}-a^{v+1})M_1(u,v).\end{aligned}$$

By applying the mean value theorem as done when studying Δ_1 and following similar steps, we get $|\Delta_2| \leq (\widehat{a}-a)^2((\widetilde{a}+1)\mathrm{e}^{\widetilde{a}})^2$. Then, it follows that $(\widehat{a}-a)^2((\widetilde{a}+1)\mathrm{e}^{\widetilde{a}})^2 \stackrel{a.s.(P)}{\longrightarrow} (a-a)^2\left((a+1)\mathrm{e}^{a}\right)^2 = 0$, and, hence, $\Delta_2 \stackrel{a.s.(P)}{\longrightarrow} 0$. □

Lemma 1.4. *Let $X_1,\ldots,X_n$ be independent and identically distributed from X, a random variable taking values in $\mathbb{N}_0$. Assume that $\widehat{\theta} \stackrel{a.s.(P)}{\longrightarrow} \theta$, for some $\theta \in \Theta$. Given the data, let $X_1^*,\ldots,X_n^*$ be independent and identically distributed from $X^* \sim BT(\widehat{\theta})$. Let $\widehat{d}^*(k;\theta)$ be defined as $\widehat{d}(k;\theta)$ with $\widehat{p}(k)$ replaced with $\widehat{p}^*(k) = \frac{1}{n}\sum_{j=1}^n I(X_j^*=k)$, $k \geq 0$. Then, for $i=1,2$,*

$$\text{(I)} \quad \sum_{k\geq 0}\left[\frac{\partial}{\partial\theta_i}\widehat{d}^*(k;\widehat{\theta}) - \mu_i(k;\widehat{\theta})\right]^2 \stackrel{P_*}{\longrightarrow} 0, \quad \text{a.s.}(P),$$

$$\text{(II)} \quad \sum_{k\geq 0}\left[\mu_i(k;\theta) - \mu_i(k;\widehat{\theta})\right]^2 \to 0, \quad \text{a.s.}(P).$$

Proof: (I) We have that

$$\sum_{k\geq 0}\left[\frac{\partial}{\partial a}\widehat{d}^*(k;\widehat{\theta}) - \mu_1(k;\widehat{\theta})\right]^2 = \sum_{k\geq 0}\left\{-\widehat{b}\sum_{v=0}^{k}(v+1)\frac{a^v}{v!}\left[\widehat{p}^*(k-v) - p(k-v;\widehat{\theta})\right]\right\}^2$$

$$= \widehat{b}^2\sum_{u,v\geq 0}(u{+}1)\frac{\widehat{a}^u}{u!}(v{+}1)\frac{\widehat{a}^v}{v!}\sum_{k\geq\max\{u,v\}}\left\{\widehat{p}^*(k-v) - p(k-v;\widehat{\theta})\right\}\left\{\widehat{p}^*(k-u) - p(k-u;\widehat{\theta})\right\}$$

$$\leq \left[\widehat{b}(\widehat{a}+1)\mathrm{e}^{\widehat{a}}\right]^2\sum_{k\geq 0}\left\{\widehat{p}^*(k) - p(k;\widehat{\theta})\right\}^2.$$

Since $[\widehat{b}(\widehat{a}{+}1)\mathrm{e}^{\widehat{a}}]^2$ is a continuous function of $\widehat{\theta} = (\widehat{a},\widehat{b})$, we have that $\widehat{b}(\widehat{a}{+}1)\mathrm{e}^{\widehat{a}}]^2 \stackrel{a.s.(P)}{\longrightarrow} [b(a+1)\mathrm{e}^{a}]^2 < \infty, \quad \forall\,\theta\in\Theta$. We also have that (see proof of Lemma 4 in Batsidis et al. [3]) $\sum_{k\geq 0}\{\widehat{p}^*(k) - p(k;\widehat{\theta})\}^2 \stackrel{P_*}{\longrightarrow} 0$, and it follows that

$$\sum_{k\geq 0}\left[\frac{\partial}{\partial\theta_1}\widehat{d}^*(k;\widehat{\theta}) - \mu_1(k;\widehat{\theta})\right]^2 \stackrel{P_*}{\longrightarrow} 0, \quad a.s.(P).$$

Also, we have that

$$\sum_{k\geq 0}\left[\frac{\partial}{\partial b}\widehat{d}^*(k;\widehat{\theta}) - \mu_2(k;\widehat{\theta})\right]^2 = \sum_{k\geq 0}\left\{\sum_{v=0}^{k}\frac{a^{v+1}}{v!}[\widehat{p}^*(k-v) - p(k-v;\widehat{\theta})]\right\}^2$$

$$= \sum_{u,v\geq 0}\frac{\widehat{a}^{u+1}}{u!}\frac{\widehat{a}^{v+1}}{v!}\sum_{k\geq\max\{u,v\}}\{\widehat{p}^*(k-v) - p(k-v;\widehat{\theta})\}\{\widehat{p}^*(k-u) - p(k-u;\widehat{\theta})\}$$

$$\leq \left(\widehat{a}\mathrm{e}^{\widehat{a}}\right)^2 \sum_{k\geq 0}\{\widehat{p}^*(k) - p(k;\widehat{\theta})\}^2.$$

Using similar arguments as above, we have $(\widehat{a}\mathrm{e}^{\widehat{a}})^2 \xrightarrow{a.s.(P)} (a\mathrm{e}^a)^2 < \infty, \quad \forall\,\theta \in \Theta$. Then, taking into account that $\sum_{k\geq 0}\{\widehat{p}^*(k) - p(k;\widehat{\theta})\}^2 \xrightarrow{P_*} 0$, we obtain

$$\sum_{k\geq 0}\left[\frac{\partial}{\partial\theta_2}\widehat{d}^*(k;\widehat{\theta}) - \mu_2(k;\widehat{\theta})\right]^2 \xrightarrow{P_*} 0, \quad a.s.(P).$$

(II) We have that $\sum_{k\geq 0}[\mu_1(k;\theta) - \mu_1(k;\widehat{\theta})]^2 = \Delta_{11} + 2\Delta_{12} + \Delta_{13}$, where

$$\Delta_{11} = \sum_{k\geq 0}\sum_{u,v=0}^{k}(u{+}1)\frac{\widehat{b}\widehat{a}^u}{u!}(v{+}1)\frac{\widehat{b}\widehat{a}^v}{v!}\{p(k{-}u;\widehat{\theta}){-}p(k{-}u;\theta)\}\{p(k{-}v;\widehat{\theta}){-}p(k{-}v;\theta)\},$$

$$\Delta_{12} = \sum_{k\geq 0}\sum_{u,v=0}^{k}(u+1)\frac{\widehat{b}\widehat{a}^u}{u!}\frac{v+1}{v!}\{p(k-u;\widehat{\theta}) - p(k-u;\theta)\}p(k-v;\theta)\{\widehat{b}\widehat{a}^v - ba^v\},$$

$$\Delta_{13} = \sum_{k\geq 0}\sum_{u,v=0}^{k}\frac{u+1}{u!}\frac{v+1}{v!}p(k-u;\theta)p(k-v;\theta)\{\widehat{b}\widehat{a}^u - ba^u\}\{\widehat{b}\widehat{a}^v - ba^v\}.$$

It follows that

$$\Delta_{11} \leq (\widehat{b}(\widehat{a}+1)\mathrm{e}^{\widehat{a}})^2\sum_{k\geq 0}\{p(k;\widehat{\theta}) - p(k;\theta)\}^2.$$

Since $(\widehat{b}(\widehat{a}+1)\mathrm{e}^{\widehat{a}})^2 \xrightarrow{a.s.(P)} (b(a+1)\mathrm{e}^a)^2$, it suffices to show that

$$\sum_{k\geq 0}\{p(k;\widehat{\theta}) - p(k;\theta)\}^2 \xrightarrow{a.s.(P)} 0,$$

then, $\Delta_{11} \xrightarrow{a.s.(P)} 0$. Taking into account that

$$\sum_{k\geq 0}\{p(k;\widehat{\theta}) - p(k;\theta)\}^2 \leq \sum_{k\geq 0}k^2\{p(k;\widehat{\theta}) - p(k;\theta)\}^2,$$

and that $\mathbb{E}_\theta(X^2) = (ba\mathrm{e}^a)^2 + ba\mathrm{e}^a(1+a), \forall\theta\in\Theta$, the rest of the proof is parallel with the proof of Lemma 4 II given in Jiménez-Gamero and Alba-Fernandez [21] and, hence, it is omitted.

We now deal with Δ_{12}. After some algebra and by applying the mean value theorem as in the proof of Lemma 9, we have that $|\Delta_{12}| \leq \widehat{b}(\widehat{a}+1)\mathrm{e}^{\widehat{a}}\widetilde{b}(\widehat{a}-a)(\widetilde{a}+2)\mathrm{e}^{\widetilde{a}} + \widehat{b}(\widehat{a}+1)\mathrm{e}^{\widehat{a}}(\widehat{b}-b)(\widetilde{a}+1)\mathrm{e}^{\widetilde{a}}$. Thus, $\Delta_{12} \xrightarrow{a.s.(P)} 0$.

Related to $|\Delta_{13}|$, note that after some algebra and following similar arguments as above, we have that

$$|\Delta_{13}| \leq \sum_{u\geq 1}\frac{u+1}{u!}\{u\widetilde{b}\widetilde{a}^{u-1}|\widehat{a}-a| + \widetilde{a}^u|\widehat{b}-b|\}$$
$$\times\sum_{v\geq 1}\frac{v+1}{v!}\{v\widetilde{b}\widetilde{a}^{v-1}|\widehat{a}-a| + \widetilde{a}^v|\widehat{b}-b|\},$$

or

$$\begin{aligned}|\Delta_{13}| \;\leq\; & (\widehat{a}-a)^2\widetilde{b}\sum_{u,v\geq 1}\frac{u+1}{u!}uv\widetilde{a}^{u-1}\widetilde{a}^{v-1}\\ &+\;(\widehat{b}-b)^2\sum_{u,v\geq 1}\frac{u+1}{u!}\widetilde{a}^{u}\widetilde{a}^{v}\\ &+\;2|\widehat{a}-a||\widehat{b}-b|\widetilde{b}\sum_{u,v\geq 1}\frac{u+1}{u!}u\widetilde{a}^{u-1}\widetilde{a}^{v}.\end{aligned}$$

Also, we have that $\sum_{k\geq 0}[\mu_2(k;\theta)-\mu_2(k;\widehat{\theta})]^2=\Delta_{21}+2\Delta_{22}+\Delta_{23}$, where

$$\Delta_{21}=\sum_{k\geq 0}\sum_{u,v=0}^{k}\frac{\widehat{a}^{u+1}}{u!}\frac{\widehat{a}^{v+1}}{v!}\{p(k-u;\widehat{\theta})-p(k-u;\theta)\}\{p(k-v;\widehat{\theta})-p(k-v;\theta)\},$$

$$\Delta_{22}=\sum_{k\geq 0}\sum_{u,v=0}^{k}\frac{\widehat{a}^{u+1}}{u!}\frac{1}{v!}\{p(k-u;\widehat{\theta})-p(k-u;\theta)\}p(k-v;\theta)\{\widehat{a}^{v+1}-a^{v+1}\},$$

$$\Delta_{23}=\sum_{k\geq 0}\sum_{u,v=0}^{k}\frac{1}{u!}\frac{1}{v!}p(k-u;\theta)p(k-v;\theta)\{\widehat{a}^{u+1}-a^{u+1}\}\{\widehat{a}^{v+1}-a^{v+1}\}.$$

Similarly, $\Delta_{21}\;\leq\;(\widehat{a}\mathrm{e}^{\widehat{a}})^2\sum_{k\geq 0}\{p(k;\widehat{\theta})-p(k;\theta)\}^2$. Since $(\widehat{a}\mathrm{e}^{\widehat{a}})^2\overset{a.s.(P)}{\longrightarrow}(a\mathrm{e}^{a})^2$ and $\sum_{k\geq 0}\{p(k;\widehat{\theta})-p(k;\theta)\}^2\overset{a.s.(P)}{\longrightarrow}0$, we have that $\Delta_{21}\overset{a.s.(P)}{\longrightarrow}0$. Also,

$$\begin{aligned}\mid\Delta_{22}\mid\;\leq\;&\mid\widehat{a}-a\mid\sum_{u\geq 0}\frac{\widehat{a}^{u+1}}{u!}\sum_{v\geq 0}\frac{v+1}{v!}\widetilde{a}^{v}\\ =\;&\mid\widehat{a}-a\mid\widehat{a}\mathrm{e}^{\widehat{a}}(\widetilde{a}+1)\mathrm{e}^{\widetilde{a}}.\end{aligned}$$

Since $\mid\widehat{a}-a\mid\widehat{a}\mathrm{e}^{\widehat{a}}(\widetilde{a}+1)\mathrm{e}^{\widetilde{a}}\overset{a.s.(P)}{\longrightarrow}0$, it follows that $\Delta_{22}\overset{a.s.(P)}{\longrightarrow}0$.

Finally, it holds that

$$\begin{aligned}\mid\Delta_{23}\mid\;\leq\;&\sum_{u,v\geq 0}\frac{\widehat{a}^{u+1}-a^{u+1}}{u!}\frac{\widehat{a}^{v+1}-a^{v+1}}{u!}\\ \leq\;&(\widehat{a}-a)^2\left((\widetilde{a}+1)\mathrm{e}^{\widetilde{a}}\right)^2,\end{aligned}$$

and since $(\widehat{a}-a)^2\left(\widetilde{a}+1)\mathrm{e}^{\widetilde{a}}\right)^2\overset{a.s.(P)}{\longrightarrow}0$, it follows that $\Delta_{23}\overset{a.s.(P)}{\longrightarrow}0$. □

Proof of Theorem 3.1 By applying the mean value theorem, we get, for each $k\in\mathbb{N}_0$, that

$$\widehat{d}(k;\widehat{\theta})=\widehat{d}(k;\theta)+\left\{\frac{\partial}{\partial\theta}\widehat{d}(k;\theta)\right\}(\widehat{\theta}-\theta)^T+\left\{\frac{\partial}{\partial\theta}\widehat{d}(k;\theta_l)-\frac{\partial}{\partial\theta}\widehat{d}(k;\theta)\right\}(\widehat{\theta}-\theta)^T, \tag{1.3}$$

with $\theta_l = \gamma_l\theta + (1-\gamma_l)\widehat{\theta}$, for some $\gamma_l \in (0,1)$. From Lemma 1.1, $\mathbb{E}(\|\phi(X;\theta)\|_2^2) < \infty$ and thus by the strong law of large number (SLLN) in Hilbert spaces and the continuous mapping theorem, it follows that

$$\|\widehat{d}(k;\theta)\|_2^2 \xrightarrow{a.s.} \|E\{\phi(X;\theta)\}\|_2^2 = \eta < \infty. \tag{1.4}$$

Finally, the result follows from (1.3), (1.4) and Lemmas 1.2 and 1.3. □

Proof of Theorem 3.2 From expansion (1.3), Assumption 1 and Lemmas 1.2 and 1.3, it follows that

$$\sqrt{n}\widehat{d}(\cdot;\widehat{\theta}) = \sqrt{n}\widehat{d}(\cdot;\theta) + \left\{\frac{\partial}{\partial\theta}\widehat{d}(\cdot;\theta)\right\}\sqrt{n}(\widehat{\theta}-\theta)^T + r_1, \tag{1.5}$$

with $\|r_1\|_2 = o_P(1)$. Now, by applying the SLLN in Hilbert spaces and Assumption 1, we get

$$\sqrt{n}\widehat{d}(\cdot;\theta) + \left\{\frac{\partial}{\partial\theta}\widehat{d}(\cdot;\theta)\right\}\sqrt{n}(\widehat{\theta}-\theta)^T = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y(X_i;\cdot,\theta) + r_2, \tag{1.6}$$

with $\|r_2\|_2 = o_P(1)$. By the central limit theorem in Hilbert spaces,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y(X_i;\cdot,\theta) \xrightarrow{\mathcal{L}} S(\theta), \tag{1.7}$$

where $Y(X;\cdot,\theta) = (Y(X;0,\theta), Y(X;1,\theta),\ldots)$. The result follows from (1.5)–(1.7) and the continuous mapping theorem. □

Proof of Theorem 3.3 Proceeding as in the proof of Theorem 3.2, we have that

$$\sqrt{n}\widehat{d}^*(\cdot;\widehat{\theta}^*) = \sqrt{n}\widehat{d}^*(\cdot;\theta) + \left\{\frac{\partial}{\partial\theta}\widehat{d}^*(\cdot;\widehat{\theta})\right\}\sqrt{n}(\widehat{\theta}^*-\widehat{\theta})^T + r_1^*,$$

with $\|r_1^*\|_2 = o_{P_*}(1)$ a.s.(P). Let $Y_n^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y(X_i^*;\cdot,\widehat{\theta})$. By applying Lemma 1.4 and Assumption 2, we get

$$\sqrt{n}\widehat{d}^*(\cdot;\theta) + \left\{\frac{\partial}{\partial\theta}\widehat{d}^*(\cdot;\widehat{\theta})\right\}\sqrt{n}(\widehat{\theta}^*-\widehat{\theta})^T = Y_n^* + r_2^*,$$

with $\|r_2^*\|_2 = o_{P_*}(1)\ a.s.(P)$. To prove the result we derive the asymptotic distribution of Y_n^*, showing that it coincides with the asymptotic distribution of $S_n(\widehat{\theta})$ when the data come from $X \sim \mathrm{BT}(\theta)$. With this aim, we apply Theorem 1.1 in Kundu et al. [31]. So, we will show that conditions (i)–(iii) in that theorem hold. This can be done in a similar way with the proof of Theorem 3 in Jiménez-Gamero and Alba-Fernandez [21].

B. APPENDIX: Function ℓ

Here, the form of the function ℓ, appeared in Assumption 1, associated with the ML estimators, and the moment estimators are provided. Moreover, it is proved that the conditions given in Assumption 1 really hold for the aforementioned estimators. For details about the existence of the ML estimators, and ways of computing them in practice, we refer to Section 4.2 in Castellares et al. [8].

In this context, when the ML estimators of the BT distribution are used, particularized for this special distribution, the general relation given in the the proof of Theorem 3.2 in White [51] (see also Jiménez-Gamero and Kim [24]), the ℓ function is given by $\ell(x;\theta) = -A(\theta)^{-1}\nabla \log f(x;\theta)$, with

$$A(\theta) = -\begin{pmatrix} ba^{-1}(1+a)\mathrm{e}^{a} & \mathrm{e}^{a} \\ \mathrm{e}^{a} & K_{bb} \end{pmatrix},$$

where K_{bb} cannot be obtained in closed-form and is provided in Castellares et al. [8, p. 4846], and $\nabla \log f(x;\theta) = \left(-b\mathrm{e}^{-a} + \frac{x}{a}, (1-\mathrm{e}^{a}) + \frac{\partial}{\partial b}\log T_x(b)\right)^T$. Note that $-A(\theta) = K(\theta)$ is the unit (per observation) expected Fisher information matrix. Despite the fact that $K(\theta)$ cannot be obtained in closed-form, we have from Castellares et al. [8, p. 4846] that $K_{bb} \leq \mathrm{e}^{a}b^{-1}$ and $\det(K(\theta)) < \infty$. This implies that the inverse of this matrix exists. Furthermore, we have from Castellares et al. [8] that $\mathbb{E}_\theta(\frac{\partial}{\partial\theta_1}\log f(x;\theta)) = 0$ and $\mathbb{E}_\theta(\frac{\partial}{\partial\theta_2} log f(x;\theta)) = 0$. Therefore, the relation $\mathbb{E}_\theta\{\ell(X_i;\theta)\} = 0$ is fulfilled when the ML estimator is used. Finally, we have that $J(\theta) = \mathbb{E}_\theta\{\ell(X_i;\theta)^T\ell(X_i;\theta)\} = tr((K(\theta))^{-1}K(\theta)^{-1}\Sigma_1) = tr(K(\theta)^{-1}) < \infty$, where $tr(A)$ denotes the trace of the matrix A, and $\Sigma_1 = Cov_\theta(\nabla \log f(X;\theta)) = K(\theta)$.

Now, we consider the moment estimators of the BT distribution parameters to find the expression ℓ and to confirm that the conditions given in Assumption 1 are satisfied. Initially, note that from Remark 12 in Castellares et al. [8], we have after some algebra that $(a,b)^T = (g_1(\mu_1,\mu_2), g_2(\mu_1,\mu_2))^T$, where

$$g_1(\mu_1,\mu_2) = \frac{\mu_2 - (\mu_1)^2}{\mu_1} - 1, \quad g_2(\mu_1,\mu_2) = \frac{\mu_1 \exp(1 - \frac{\mu_2 - \mu_1^2}{\mu_1})}{\frac{\mu_2 - (\mu_1)^2}{\mu_1} - 1},$$

with $\mu_k = \mathbb{E}(X^k)$, given in Remark 12 by Castellares et al. [8]. Therefore, since $g = (g_1, g_2)^T$ is continuously differential at $(\mu_1,\mu_2)^T$ and $\mathbb{E}(||X||^4) < \infty$, we have that (see for instance Jiménez-Gamero and Kim [24]) $\ell(x;\theta) = (\ell_1(x;\theta), \ell_2(x;\theta))^T$, and

$$\ell_1(x;\theta) = \left(\frac{\partial}{\partial\mu_1}g_1(\mu_1,\mu_2), \frac{\partial}{\partial\mu_2}g_1(\mu_1,\mu_2)\right)(x-\mu_1, x^2-\mu_2)^T,$$

$$\ell_2(x;\theta) = \left(\frac{\partial}{\partial\mu_1}g_2(\mu_1,\mu_2), \frac{\partial}{\partial\mu_2}g_2(\mu_1,\mu_2)\right)(x-\mu_1, x^2-\mu_2)^T.$$

Obviously, $\mathbb{E}_\theta\{\ell(X_i;\theta)\} = 0$ since $\mathbb{E}_\theta(X-\mu_1) = \mathbb{E}_\theta(X^2-\mu_2) = 0$. Therefore, the condition $\mathbb{E}_\theta\{\ell(X_i;\theta)\} = 0$ is fulfilled when the moment estimator is used. In the sequel, let us denote by $K_1(\theta)$ the 2×2 matrix with (i,j) element ($i,j = 1,2$) equal to $\frac{\partial}{\partial \mu_j} g_i(\mu_1,\mu_2)$. The elements of the matrix $K_1(\theta)$, which depend only on μ_1 and μ_2, are omitted here, however, they are available upon request and can be given in closed-form. Finally, we have that $J(\theta) = E_\theta\{\ell(X_i;\theta)^T \ell(X_i;\theta)\} = tr(K_1(\theta)^T K_1(\theta)\Sigma_2)$, where

$$\Sigma_2 = Cov_\theta \left(X - \mu_1, X^2 - \mu_2\right)^T = \begin{pmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix}.$$

Therefore, $J(\theta) < \infty$ since $tr((K_1(\theta))^T K_1(\theta)\Sigma_2) < \infty$.

ACKNOWLEDGMENTS

Artur Lemonte acknowledges the financial support of the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant 304776/2019–0). Apostolos Batsidis would like to thank M.D. Jiménez-Gamero, Bojana Milosevic and Jochen Einbeck for valuable discussion during the preparation of this paper. We are very grateful to two anonymous referees for the valuable comments and suggestions which have improved the first version of the paper.

References

[1] BARINGHAUS, L., GÜRTLER, N., HENZE, N. (2000). Theory and methods: Weighted integral test statistics and components of smooth tests of fit. *Australian and New Zealand Journal of Statistics*, **42**, 179–192.

[2] BARINGHAUS, L., HENZE, N. (1992). A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statistics and Probability Letters*, **13**, 269–274.

[3] BATSIDIS, A., JIMÉNEZ-GAMERO, M.D., LEMONTE, A.J. (2020). On goodness-of-fit tests for the Bell distribution. *Metrika*, **83**, 297–319.

[4] BELL, E.T. (1934a). Exponential polynomials. *Annals of Mathematical*, **35**, 258–277.

[5] BELL, E.T. (1934b). Exponential numbers. *The American Mathematical Monthly*, **41**, 411–419.

[6] BURKE, M. (2000). Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. *Statistics and Probability Letters*, **46**, 13–20.

[7] CASTELLARES, F., FERRARI, S.L.P., LEMONTE, A.J. (2018). On the Bell distribution and its associated regression model for count data. *Applied Mathematical Modelling*, **56**, 172–185.

[8] CASTELLARES, F., LEMONTE, A.J., MORENO-ARENAS, G. (2020). On the two-parameter Bell-Touchard discrete distribution. *Communications in Statistics - Theory and Methods*, **4**, 4834–4852.

[9] CATCHESIDE, D.G., LEA, D.E., THODAY, J.M. (1946a). Types of chromosome structural change induced by the irradiation of Tradescantia microspores. *Journal of Genetics*, **47**, 113–136.

[10] CATCHESIDE, D.G., LEA, D.E., THODAY, J.M. (1946b). The production of chromosome structural changes in Tradescantia microspores in relation to dosage, intensity and temperature. *Journal of Genetics*, **47**, 137–149.

[11] DEHLING, H., MIKOSCH, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *Journal of Multivariate Analysis*, **51**, 392–413.

[12] EFRON, B., TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

[13] ESNAOLA M., PUIG, P., GONZALEZ, D., CASTELO, R., GONZALEZ, J.R. (2013). A flexible count data model to fit the wide diversity of expression profiles arising from extensively replicated RNA-seq experiments. *BMC Bioinformatics*, **14**, 254.

[14] EPPS, T.W. (1995). A test of fit for lattice distributions. *Communications in Statistics - Theory and Methods*, **24**, 1455–1479.

[15] FELLER, W. (1943). On a General Class of Contagious Distributions. *The Annals of Mathematical Statistics*, **14**, 389–400.

[16] GIACOMINI, R., POLITIS, D.N., WHITE, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, **29**, 567–589.

[17] GOSSIAUX, A., LEMAIRE, J. (1981). Methodes d'ajustement de distributions de sinistres. *Bulletin of the Association of Swiss Actuaries*, **81**, 87–95.

[18] GÜRTLER N., HENZE, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution. *Journal of Statistical Planning and Inference*, **90**, 207-225.

[19] HENZE, N. (1996). Empirical-Distribution-Function Goodness-of-Fit Tests for Discrete Models. *The Canadian Journal of Statistics*, **24**, 81–93.

[20] JANSSEN, A. (2000). Global power functions of goodness of fit tests. *Annals Statistics*, **28**, 239–253.

[21] JIMÉNEZ-GAMERO, M.D., ALBA-FERNÁNDEZ, M.V. (2019). Testing for the Poisson–Tweedie distribution. *Mathematics and Computers in Simulation*, **164**, 146–162.

[22] JIMÉNEZ-GAMERO, M.D., ALBA-FERNÁNDEZ, M.V. (2021). A test for the geometric distribution based on linear regression of order statistics. *Mathematics and Computers in Simulation*, **186**, 103-123.

[23] JIMÉNEZ-GAMERO, M.D., BATSIDIS, A. (2017). Minimum distance estimators for count data based on the probability generating function with applications. *Metrika*, **80**, 503–545.

[24] JIMÉNEZ-GAMERO, M.D., KIM, H.-M. (2015). Fast goodness-of-fit tests based on the characteristic function. *Computational Statistics and Data Analysis*, **89**, 172–191.

[25] JOE H., ZHU R. (2005). Generalized poisson distribution: the property of mixture of poisson and comparison with Negative Binomial distribution. *Biometrical Journal*, **47**, 219–229.

[26] JOHNSON, N. L., KOTZ, S., KEMP, A. (1992). *Univariate discrete distributions*. 2nd edition, Wiley, New York.

[27] KLAR, B. (1999). Goodness-of-fit tests for discrete models based on the integrated distribution function. *Metrika*, **49**, 53–69.

[28] KLUGMAN, S., PANJER, H., WILLMOT, G. (1998). *Loss Models. From Data to Decisions*. John Wiley and Sons, New York.

[29] KOCHERLAKOTA, S., KOCHERLAKOTA, K. (1986). Goodness of fit test for discrete distributions. *Communications in Statistics - Theory and Methods*, **15**, 815–829.

[30] KOJADINOVIC, I., YAN, J. (2012). Goodness-of-fit testing based on a weighted bootstrap: A fast large sample alternative to the parametric bootstrap. *The Canadian Journal of Statistics*, **40**, 480–500.

[31] KUNDU, S., MAJUMDAR, S. MUKHERJEE, K. (2000). Central Limit Theorems revisited, *Statistics and Probability Letters*, **47**, 265–275.

[32] LORD, D., WASHINGTON S.P., IVAN, J.N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*, **37**, 35–46.

[33] MASSÉ, J., THEODORESCU, R. (2005). Neyman type A distribution revisited. *Statistica Neerlandica*, **59**, 206–213.

[34] MCGUIRE, J.U, BRINDLEY, T.A., BANCROFT, T.A. (1957). The distribution of European corn borer larvae Pyrausta nubilalis (Hbn.), in field corn. *Biometrics*, **13**, 65–78.

[35] MEINTANIS, S. (2008). New inference procedures for generalized Poisson distributions. *Journal of Applied Statistics*, **35**, 751–762.

[36] MEINTANIS, S., BASSIAKOS, Y. (2005). Goodness-of-fit test for additively closed count models with an application to the generalized Hermite distribution. *Sankhya*, **67**, 538–552.

[37] MILOSEVIC, B., JIMÉNEZ-GAMERO, M.D., ALBA-FERNANDEZ, M.V. (2021). Quantifying the ratio-plot for the geometric distribution, *Journal of Statistical Computation and Simulation*, DOI: 10.1080/00949655.2021.1887185.

[38] NAKAMURA, M., PEREZ-ABREU, V. (1993). Empirical probability generating function. An overview. *Insurance: Mathematics and Economics*, **12**, 287–295.

[39] NEYMAN, J. (1939), On a new class of contagious distributions applicable in entomologyand bacteriology. *Annals of Mathematical Statistics*, **10**, 35–57.

[40] NOVOA-MUÑOZ, F., JIMÉNEZ-GAMERO, M.D. (2014). Testing for the bivariate Poisson distribution. *Metrika*, **77**, 771–793.

[41] NOVOA-MUÑOZ, F., JIMÉNEZ-GAMERO, M.D. (2016). A goodness-of-fit test for the multivariate Poisson distribution. *Sort*, **40**, 1–26.

[42] PUIG, P. AMD VALERO, J. (2006). Count Data Distributions: Some Characterizations with Applications. *Journal of the American Statistical Association*, **101:473**, 332–340.

[43] R CORE TEAM (2020). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[44] RASHID, A., AHMAD, Z., JAN, T.R. (2016). A new count data model with application in genetics and ecology. *Electronic Journal of Applied Statistical Analysis*, **9**, 213–226.

[45] RUEDA, R., O'REILLY, F. (1999). Tests of fit for discrete distributions based on the probability generating function. *Communications in Statistics - Simulation and Computation*, **28**, 259–274.

[46] RUEDA, R., PEREZ-ABREU, V., O'REILLY, F. (1991). Goodness of fit for the Poisson distribution based on the probability generating function. *Communications in Statistics - Theory and Methods*, **20**, 3093–3110.

[47] SICHEL, H.S. (1951). The estimation of the parameters of a Negative Binomial distribution with special reference to psychological data. *Psychometrika*, **16**, 107–127.

[48] TOUCHARD, J. (1933). Propriétés arihtmétiques de certains nombres récurrents. *Annales de la Société Scientifique de Bruxelles*, **53**, 21–31.

[49] TRIPATHI, R.C. (2004). Neyman type A, B, and C Distributions. In Encyclopedia of Statistical Sciences (eds S. Kotz, C.B. Read, N. Balakrishnan, B. Vidakovic and N.L. Johnson).

[50] ZAFAKALI, N.S., AHMAD, W.M.A.W. (2013). Modeling and handling overdispersion health science data with zero-inflated Poisson model. *Journal of Modern Applied Statistical Methods*, **12**, Article 28.

[51] WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.