SORT 40 (1) January-June 2016, 113-138

A goodness-of-fit test for the multivariate Poisson distribution

F. Novoa-Muñoz[1,*] and M.D. Jiménez-Gamero[2]

Abstract

Bivariate count data arise in several different disciplines and the bivariate Poisson distribution is commonly used to model them. This paper proposes and studies a computationally convenient goodness-of-fit test for this distribution, which is based on an empirical counterpart of a system of equations. The test is consistent against fixed alternatives. The null distribution of the test can be consistently approximated by a parametric bootstrap and by a weighted bootstrap. The goodness of these bootstrap estimators and the power for finite sample sizes are numerically studied. It is shown that the proposed test can be naturally extended to the multivariate Poisson distribution.

MSC: 62F03; 62F05; 62F12; 62F40.

Keywords: Bivariate Poisson distribution, goodness-of-fit, empirical probability generating function, parametric bootstrap, weighted bootstrap, multivariate Poisson distribution.

1. Introduction

Univariate count data appear in many real life situations and the univariate Poisson distribution is frequently used to model this kind of data (see for example Haight, 1967; Johnson and Kotz 1969; Sahai and Khurshid, 1993). Gürtler and Henze (2000) present a wide variety of procedures for testing goodness-of-fit (gof) for the univariate Poisson distribution.

In practice, bivariate count data appear in different areas of knowledge and the bivariate Poisson distribution (BPD), being a generalization of the Poisson distribution, plays a key role in modelling them, provided that such data present a positive correlation.

* Corresponding author: fnovoa@ubiobio.cl

[1] Departamento de Estadística, Universidad del Bío-Bío, Chile.

[2] Departamento de Estadística e I.O., Universidad de Sevilla, Spain.

Received: January 2015
Accepted: April 2016

Different authors have given a definition for the BPD (see for example Kocherlakota and Kocherlakota, 1992). In this article we will work with the one that has received more attention (see for example Holgate, 1964; Johnson, Kotz and Balakrishnan, 1997). Let

$$X_1 = Y_1 + Y_3 \quad \text{and} \quad X_2 = Y_2 + Y_3,$$

where Y_1, Y_2 and Y_3 are mutually independent Poisson random variables with means $\theta_1' = \theta_1 - \theta_3 > 0$, $\theta_2' = \theta_2 - \theta_3 > 0$ and $\theta_3 > 0$, respectively. The joint distribution of the vector (X_1, X_2) is called BPD with parameter $\theta = (\theta_1, \theta_2, \theta_3)$, $(X_1, X_2) \sim BP(\theta)$ for short. In the statistical literature on gof tests for the BPD, which is not so rich as in the univariate case, we found the following: the tests given by Crockett (1979), Loukas and Kemp (1986), Rayner and Best (1995) – these three tests are not consistent against all fixed alternatives – and, more recently, the tests in Novoa-Muñoz and Jiménez-Gamero (2014) (hereafter abbreviated to NJ).

The two tests in NJ are consistent against all fixed alternatives. The results in Janssen (2000) assert that the global power function of any nonparametric test is flat on balls of alternatives except for alternatives coming from a finite dimensional subspace. Because of this reason, it is interesting to propose new gof tests able to detect different sets of alternatives.

This paper presents a consistent gof test for the BPD. It is based on the following: since the probability generating function (pgf) of the BPD is the unique pgf satisfying certain system of partial differential equations, and the empirical probability generating function (epgf) consistently estimates the pgf, the epgf should approximately satisfy such system. The proposed test statistic is a function of the coefficients of the polynomials of an empirical version of that system. The asymptotic behaviour of the proposed test under alternatives is shared with the ones in NJ. An advantage of the test proposed in this paper over those in NJ is that its application does not entail the choice of a weight function, which is rather arbitrary.

The null distribution of the test statistic can be consistently approximated by a parametric bootstrap as well as by means of a weighted bootstrap. The finite sample performance of the proposed test is investigated by means of a simulation study, where the goodness of the proposed approximations is numerically studied and the test is compared, in terms of power, to the tests cited above. The numerical power study reveals that, as expected from the results in Janssen (2000), there is no test yielding the highest power against all considered alternatives. In most cases, the power of the proposed test is quite close to the highest one; in other cases, the proposed test is the most powerful. In addition, from a computational point of view, the test proposed in this paper is more efficient than its competitors.

The work is organized as follows. Section 2 introduces the test statistic and derives its asymptotic null distribution. Since the asymptotic null distribution does not provide a useful means of approximating the null distribution of the test statistic, Section 3 studies two bootstrap estimators. Specifically, it is shown that the parametric bootstrap and a conveniently defined weighted bootstrap estimators produce consistent null distribution estimators. This Section also studies the power of the resulting tests against fixed alternatives. Section 4 deals with the practical implementation of the bootstrap null distribution estimators as well as other related issues. Section 5 reports a summary of the results of a simulation study carried out to examine the finite sample performance of the tests and to compare them with the existing ones. All stated results are valid for $\theta_3 > 0$. Section 6 deals with the case $\theta_3 = 0$. Section 7 shows how the proposed technique can be applied to the general multivariate case. All proofs are relegated to the last section.

Hereinafter we shall use the following notation: all vectors are row vectors and $v^\top$ is the transposed of the row vector v; for any vector v, v_k denotes its kth coordinate, and $\|v\|$ its Euclidean norm; $\mathbb{N}_0 = \{0,1,2,3,\ldots\}$; $I\{A\}$ denotes the indicator function of the set A; P_θ denotes the probability law of the BPD with parameter θ; P denotes the probability law of the data; E_θ denotes expectation with respect to the probability function P_θ; E denotes expectation with respect to the true probability function of the data; P_* denote the probability law, given the data; all limits in this work are taken as $n \to \infty$; $\stackrel{L}{\longrightarrow}$ denotes convergence in distribution; $\stackrel{P}{\longrightarrow}$ denotes convergence in probability; $\stackrel{a.s.}{\longrightarrow}$ denotes almost sure (a.s.) convergence; for any function $h : S \subset \mathbb{R}^m \to \mathbb{R}$, for some fixed $m \in \mathbb{N}$, we will denote

$$D^{a_1\cdots a_m} h(u) = \frac{\partial^k}{\partial u_1^{a_1} \cdots \partial u_m^{a_m}}\, h(u),$$

$\forall a_1,\ldots,a_m \in \mathbb{N}_0$ such that $k = a_1 + \cdots + a_m$.

2. The test statistic and its asymptotic null distribution

Let $\boldsymbol{X}_1 = (X_{11},X_{12}), \boldsymbol{X}_2 = (X_{21},X_{22}),\ldots,\boldsymbol{X}_n = (X_{n1},X_{n2})$ be independent identically distributed (iid) from a random vector $\boldsymbol{X} = (X_1,X_2) \in \mathbb{N}_0^2$. Based on the sample $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$, the objective is to test the hypothesis

$$H_0 : \ (X_1,X_2) \sim BP(\theta_1,\theta_2,\theta_3), \text{ for some } (\theta_1,\theta_2,\theta_3) \in \Theta,$$

against the alternative

$$H_1 : \ (X_1,X_2) \nsim BP(\theta_1,\theta_2,\theta_3), \ \forall\, (\theta_1,\theta_2,\theta_3) \in \Theta,$$

where $\Theta = \left\{(\theta_1,\theta_2,\theta_3) \in \mathbb{R}^3 \,:\, \theta_1 > \theta_3,\ \theta_2 > \theta_3,\ \theta_3 > 0\right\}$. Since the distribution of a random vector $\boldsymbol{X} = (X_1,X_2) \in \mathbb{N}_0^2$ is determined by its pgf $g(u) = E\left(u_1^{X_1}u_2^{X_2}\right)$, $u = (u_1,u_2) \in [0,1]^2$, and the joint pgf of a random vector $\boldsymbol{X} \sim BP(\theta)$ is

$$g(u;\theta) = E_\theta(u_1^{X_1}u_2^{X_2}) = \exp\left\{\theta_1(u_1-1)+\theta_2(u_2-1)+\theta_3(u_1-1)(u_2-1)\right\}, \quad (1)$$

testing H_0 vs H_1 is equivalent to testing

$$H_0:\ g(u) = g(u;\theta),\ \forall u \in [0,1]^2,\ \text{for some } (\theta_1,\theta_2,\theta_3) \in \Theta,$$

versus

$$H_1:\ g(u) \neq g(u;\theta),\ \text{for some } u \in [0,1]^2,\ \forall\,(\theta_1,\theta_2,\theta_3) \in \Theta.$$

Proposition 2 in NJ shows that $g(u_1,u_2;\theta)$ is the only pgf in $G_2 = \{g : [0,1]^2 \to \mathbb{R}$, such that g is a pgf and $\frac{\partial}{\partial u_1}g(u_1,u_2)$ and $\frac{\partial}{\partial u_2}g(u_1,u_2)$ exist $\forall (u_1,u_2) \in [0,1]^2\}$ satisfying the following system,

$$D_i(u;\theta) = 0, \quad i = 1,2, \quad \forall u \in [0,1]^2,$$

where

$$\begin{aligned} D_1(u;\theta) &= \frac{\partial}{\partial u_1}g(u_1,u_2) - \left\{\theta_1+\theta_3(u_2-1)\right\}g(u_1,u_2), \\ D_2(u;\theta) &= \frac{\partial}{\partial u_2}g(u_1,u_2) - \left\{\theta_2+\theta_3(u_1-1)\right\}g(u_1,u_2). \end{aligned}$$

Now we consider the following empirical versions of the functions $D_i(u;\theta)$, $i = 1,2$,

$$\begin{aligned} D_{1n}(u;\hat\theta) &= \frac{\partial}{\partial u_1}g_n(u_1,u_2) - \left\{\hat\theta_1+\hat\theta_3(u_2-1)\right\} g_n(u_1,u_2), \\ D_{2n}(u;\hat\theta) &= \frac{\partial}{\partial u_2}g_n(u_1,u_2) - \left\{\hat\theta_2+\hat\theta_3(u_1-1)\right\} g_n(u_1,u_2), \end{aligned}$$

where $\hat\theta = (\hat\theta_1,\hat\theta_2,\hat\theta_3)$ is a consistent estimator of θ and $g_n(u_1,u_2)$ is the epgf associated to the data,

$$g_n(u_1,u_2) = \frac{1}{n}\sum_{i=1}^{n} u_1^{X_{i1}}u_2^{X_{i2}}.$$

Proposition 1 in NJ shows that $g(u)$ and its derivatives can be consistently estimated by the epgf and the derivatives of the epgf, respectively. Thus, if H_0 is true then $D_{1n}(u;\hat{\theta})$ and $D_{2n}(u;\hat{\theta})$ should be close to 0, $\forall u \in [0,1]^2$. This proximity to 0 can be interpreted in several ways. For example, NJ interpreted this proximity as

$$S_{n,w}(\hat{\theta}) = n\int \{D_{1n}(u;\hat{\theta})^2 + D_{2n}(u;\hat{\theta})^2\} w(u)du \approx 0, \tag{2}$$

where $w(u)$ is a non-negative function on $[0,1]^2$.

Here we present another interpretation, reasoning as in Nakamura and Pérez-Abreu (1993) for the univariate case. With this aim, observe that

$$D_{in}(u;\hat{\theta}) = \sum_{r\geq 0}\sum_{s\geq 0} d_i(r,s;\hat{\theta})u_1^r u_2^s,\ \ i=1,2, \tag{3}$$

where

$$d_1(r,s;\hat{\theta}) = (r+1)p_n(r+1,s) - (\hat{\theta}_1-\hat{\theta}_3)p_n(r,s) - \hat{\theta}_3 p_n(r,s-1),$$
$$d_2(r,s;\hat{\theta}) = (s+1)p_n(r,s+1) - (\hat{\theta}_2-\hat{\theta}_3)p_n(r,s) - \hat{\theta}_3 p_n(r-1,s),$$

and

$$p_n(r,s) = \frac{1}{n}\sum_{k=1}^{n} I(X_{k1}=r, X_{k2}=s)$$

is the relative frequency of the pair (r,s). Thus, $D_{in}(u;\hat{\theta}) = 0$, $\forall u \in [0,1]^2$, $i=1,2$, if and only if the coefficient of $u_1^r u_2^s$ in the right hand side of (3) is null, $\forall r,s \geq 0$, $i=1,2$. This leads us to consider the following statistic for testing H_0,

$$W_n(\hat{\theta}) = \sum_{r\geq 0}\sum_{s\geq 0}\{d_1(r,s;\hat{\theta})^2 + d_2(r,s;\hat{\theta})^2\} = \sum_{r,s=0}^{M}\{d_1(r,s;\hat{\theta})^2 + d_2(r,s;\hat{\theta})^2\}, \tag{4}$$

where $M = \max\{X_{(n)1}, X_{(n)2}\}$, $X_{(n)k} = \max_{1\leq i\leq n} X_{ik}$, $k=1,2$.

Taking into account that

$$d_k(r,s;\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\phi_{krs}(\boldsymbol{X}_i;\hat{\theta}), \quad k=1,2,$$

with

$$\phi_{1rs}(x;\theta) = (r+1)I(x_1 = r+1, x_2 = s) - (\theta_1 - \theta_3)I(x_1 = r, x_2 = s) - \theta_3 I(x_1 = r, x_2 = s-1),$$
$$\phi_{2rs}(x;\theta) = (s+1)I(x_1 = r, x_2 = s+1) - (\theta_2 - \theta_3)I(x_1 = r, x_2 = s) - \theta_3 I(x_1 = r-1, x_2 = s),$$

where $x = (x_1, x_2)$, the statistic $W_n(\hat{\theta})$ can be expressed as follows,

$$W_n(\hat{\theta}) = \frac{1}{n^2}\sum_{i,j=1}^{n} h(\boldsymbol{X}_i, \boldsymbol{X}_j; \hat{\theta}),$$

with

$$\begin{aligned}
h(x,y;\theta) &= h_1(x,y;\theta) + h_2(x,y;\theta),\\
h_1(x,y;\theta) &= \textstyle\sum_{r\geq 0}\sum_{s\geq 0}\phi_{1rs}(x;\theta)\phi_{1rs}(y;\theta)\\
&= \{x_1^2 + (\theta_1-\theta_3)^2 + \theta_3^2\}I(x_1 = y_1, x_2 = y_2) - (\theta_1-\theta_3)x_1 I(x_1 = y_1+1, x_2 = y_2)\\
&\quad -\theta_3 x_1 I(x_1 = y_1+1, x_2 = y_2+1) + (\theta_1-\theta_3)\theta_3 I(x_1 = y_1, x_2 = y_2+1)\\
&\quad -(\theta_1-\theta_3)y_1 I(y_1 = x_1+1, y_2 = x_2) - \theta_3 y_1 I(y_1 = x_1+1, y_2 = x_2+1)\\
&\quad +(\theta_1-\theta_3)\theta_3 I(y_1 = x_1, y_2 = x_2+1),\\
h_2(x,y;\theta) &= \textstyle\sum_{r\geq 0}\sum_{s\geq 0}\phi_{2rs}(x;\theta)\phi_{2rs}(y;\theta)\\
&= \{x_2^2 + (\theta_2-\theta_3)^2 + \theta_3^2\}I(x_1 = y_1, x_2 = y_2) - (\theta_2-\theta_3)x_2 I(x_1 = y_1, x_2 = y_2+1)\\
&\quad -\theta_3 x_2 I(x_1 = y_1+1, x_2 = y_2+1) + (\theta_2-\theta_3)\theta_3 I(x_1 = y_1+1, x_2 = y_2)\\
&\quad -(\theta_2-\theta_3)y_2 I(y_1 = x_1, y_2 = x_2+1) - \theta_3 y_2 I(y_1 = x_1+1, y_2 = x_2+1)\\
&\quad +(\theta_2-\theta_3)\theta_3 I(y_1 = x_1+1, y_2 = x_2),
\end{aligned}$$

where $x = (x_1, x_2)$ and $y = (y_1, y_2)$.

In order to give a sound justification of $W_n(\hat{\theta})$ as a test statistic for testing H_0 we next derive its a.s. limit.

Theorem 1 *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be iid from $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$ with $E(X_k^2) < \infty$, $k = 1,2$. Let $p(r,s) = P(X_1 = r, X_2 = s)$. If $\hat{\theta} \xrightarrow{a.s.} \theta$, for some $\theta \in \mathbb{R}^3$, then*

$$W_n(\hat{\theta}) \xrightarrow{a.s.} \sum_{r,s\geq 0}\left\{a_1(r,s;\theta)^2 + a_2(r,s;\theta)^2\right\} = \eta(P;\theta),$$

where

$$a_1(r,s;\theta) = (r+1)p(r+1,s) - (\theta_1-\theta_3)p(r,s) - \theta_3 p(r,s-1),$$

$$a_2(r,s;\theta) = (s+1)p(r,s+1) - (\theta_2-\theta_3)p(r,s) - \theta_3 p(r-1,s).$$

Note that $\eta(P;\theta) \geq 0$ and, taking into account that

$$D_k(u;\theta) = \sum_{r\geq 0}\sum_{s\geq 0} a_k(r,s;\theta)u_1^r u_2^s, \quad k = 1,2,$$

it follows that $\eta(P;\theta) = 0$ if and only if H_0 is true. Thus, a reasonable test for testing H_0 should reject the null hypothesis for large values of $W_n(\hat{\theta})$. Now, to determine what are large values we must calculate its null distribution, or at least an approximation to it.

We first try to estimate the null distribution of $W_n(\hat{\theta})$ by means of its asymptotic null distribution. In order to derive it, it will be assumed that the estimator $\hat{\theta}$ is asymptotically linear, as expressed in the next assumption.

Assumption 1 *Under H_0, if $\theta = (\theta_1, \theta_2, \theta_3) \in \Theta$ denotes the true parameter value, then*

$$\sqrt{n}\left(\hat{\theta} - \theta\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell\left(\boldsymbol{X}_i;\theta\right) + \mathbf{o}_P(1),$$

where $\ell : \mathbb{N}_0^2 \times \Theta \longrightarrow \mathbb{R}^3$ is such that $E_\theta\left\{\ell\left(\boldsymbol{X}_1;\theta\right)\right\} = \boldsymbol{0}$ and $J(\theta) = E_\theta\left\{\ell\left(\boldsymbol{X}_1;\theta\right)^{\mathsf{T}} \ell\left(\boldsymbol{X}_1;\theta\right)\right\}$ $< \infty$.

Assumption 1 is not restrictive at all since it is fulfilled by some commonly used estimators such as the moment estimator, the maximum likelihood estimator, the double zero estimator, the even points estimator and the conditional even points estimator (see Kocherlakota and Kocherlakota, 1992, and Papageorgiou and Loukas, 1988).

The next result gives the asymptotic null distribution of $W_n(\hat{\theta})$.

Theorem 2 *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be iid from $\boldsymbol{X} = (X_1, X_2) \sim BP(\theta_1, \theta_2, \theta_3)$. Suppose that Assumption 1 holds. Then*

$$nW_n(\hat{\theta}) \stackrel{L}{\longrightarrow} \sum_{j\geq 1} \lambda_j \chi_{1j}^2,$$

where $\chi_{11}^2, \chi_{12}^2, \ldots$ are independent χ^2 variates with one degree of freedom and the set $\{\lambda_j\}$ are the non-null eigenvalues of the operator $C(\theta)$ defined on the function space $\{\tau : \mathbb{N}_0^2 \to \mathbb{R},$ such that $E_\theta\left[\tau^2(\boldsymbol{X})\right] < \infty, \forall \theta \in \Theta\}$, as follows

$$C(\theta)\tau(x) = E_\theta\{K(x, \boldsymbol{X};\theta)\tau(\boldsymbol{X})\},$$

with $K(x,y;\theta) = h(x,y;\theta) + \ell(x;\theta)\mu(y;\theta)^{\mathsf{T}} + \ell(y;\theta)\mu(x;\theta)^{\mathsf{T}} + \ell(x;\theta)S(\theta)\ell(y;\theta)^{\mathsf{T}}$, $\mu(x;\theta) = (\mu_1(x;\theta), \mu_2(x;\theta), \mu_3(x;\theta))$,

$$\begin{aligned}
\mu_1(x;\theta) &= -x_1P_\theta(x_1-1,x_2)+\theta_3P_\theta(x_1,x_2+1)+(\theta_1-\theta_3)P_\theta(x_1,x_2),\\
\mu_2(x;\theta) &= -x_2P_\theta(x_1,x_2-1)+\theta_3P_\theta(x_1+1,x_2)+(\theta_2-\theta_3)P_\theta(x_1,x_2),\\
\mu_3(x;\theta) &= -\mu_1(x;\theta)-x_1P_\theta(x_1-1,x_2-1)+\theta_3P_\theta(x_1,x_2)+(\theta_1-\theta_3)P_\theta(x_1,x_2-1)\\
&\quad -\mu_2(x;\theta)-x_2P_\theta(x_1-1,x_2-1)+\theta_3P_\theta(x_1,x_2)+(\theta_2-\theta_3)P_\theta(x_1-1,x_2),
\end{aligned}$$

$S(\theta)=\sum_{r,s\geq 0}S_{rs}(\theta)$,

$$S_{rs}(\theta)=\begin{pmatrix} a^2 & 0 & a(b-a)\\ 0 & a^2 & a(c-a)\\ a(b-a) & a(c-a) & (b-a)^2+(c-a)^2\end{pmatrix},$$

$a=P_\theta(r,s)$, $b=P_\theta(r,s-1)$, $c=P_\theta(r-1,s)$,

The asymptotic null distribution of $W_n(\hat{\theta})$ does not provide a useful approximation to its null distribution since it depends on the unknown true value of θ. Even if θ were known or replaced by an appropriate estimator, to determine the eigenvalues of an operator is a rather hard problem.

So, we next study two further ways of approximating it: a parametric bootstrap (PB) estimator and a weighted bootstrap (WB) estimator.

3. Approximating the null distribution

3.1. Parametric bootstrap

Let $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$ be iid taking values in $\mathbb{N}_0^2$ such that $\hat{\theta}=\hat{\theta}(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n)\in\Theta$. Let $\boldsymbol{X}_1^*,\boldsymbol{X}_2^*,\ldots,\boldsymbol{X}_n^*$ be iid from a population with distribution $BP(\hat{\theta})$, given $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$, and let $W_n^*(\hat{\theta}^*)$ be the bootstrap version of $W_n(\hat{\theta})$ obtained by replacing $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$ and $\hat{\theta}=\hat{\theta}(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n)$ by $\boldsymbol{X}_1^*,\boldsymbol{X}_2^*,\ldots,\boldsymbol{X}_n^*$ and $\hat{\theta}^*=\hat{\theta}(\boldsymbol{X}_1^*,\boldsymbol{X}_2^*,\ldots,\boldsymbol{X}_n^*)$, respectively, in the expression of $W_n(\hat{\theta})$. To prove that the PB can be used to consistently approximate the null distribution of $W_n(\hat{\theta})$, we will assume the following, which is a bit stronger than Assumption 1.

Assumption 2 *Assumption 1 holds and the functions ℓ and J satisfy*

(1) $\sup_{\vartheta\in\Theta_0}E_\vartheta\left[\|\ell(\boldsymbol{X};\vartheta)\|^2 I\{\|\ell(\boldsymbol{X};\vartheta)\|>\gamma\}\right]\to 0$, *as* $\gamma\to\infty$, *where* $\Theta_0\subseteq\Theta$ *is an open neighborhood of* θ.

(2) $\ell(\boldsymbol{X};\vartheta)$ *and* $J(\vartheta)$ *are continuous as functions of* ϑ *at* $\vartheta=\theta$ *and* $J(\vartheta)$ *is finite* $\forall\vartheta\in\Theta_0$.

Theorem 3 *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be iid from $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$. Suppose that Assumption 2 holds and that $\hat{\theta} \xrightarrow{a.s.} \theta$, for some $\theta \in \Theta$. Then*

$$\sup_{x \in \mathbb{R}} \left| P_* \left\{ nW_n^*(\hat{\theta}^*) \le x \right\} - P_\theta \left\{ nW_n(\hat{\theta}) \le x \right\} \right| \xrightarrow{a.s.} 0.$$

Let $w^*_{n,\alpha} = \inf\{x : P_*(W_n^*(\hat{\theta}^*) \ge x) \le \alpha\}$ be the α upper percentile of the PB distribution of $W_n(\hat{\theta})$ and let W_{obs} be the observed value of the test statistic. From Theorem 3, the test function

$$\Psi^*_{PB} = \begin{cases} 1, & \text{if } W_n(\hat{\theta}) \ge w^*_{n,\alpha}, \\ 0, & \text{otherwise}, \end{cases}$$

or equivalently, the test that rejects H_0 when $p^* = P_* \left(W_n^*(\hat{\theta}^*) \ge W_{obs} \right) \le \alpha$, is asymptotically correct in the sense that $P_\theta(\Psi^*_{PB} = 1) \to \alpha$.

3.2. Weighted bootstrap

From the proof of Theorem 2, when H_0 is true, we have that $nW_n(\hat{\theta}) = nW_{1n}(\theta) + o_P(1)$, where

$$nW_{1n}(\theta) = \frac{1}{n} \sum_{i,j=1}^{n} K(\boldsymbol{X}_i, \boldsymbol{X}_j; \theta),$$

which converges in law to $W_0 = \sum_{j \ge 1} \lambda_j \chi^2_{1j}$. As observed before, the greatest difficulty with W_0 is to determine the set $\{\lambda_j\}$. Nevertheless, Delhing and Mikosch (1994) have shown that the eigenvalues $\{\lambda_j\}$ can be consistently (a.s.) approximated by the eigenvalues of the matrix

$$H_n = \left(\frac{1}{n} K(\boldsymbol{X}_i, \boldsymbol{X}_j; \theta) \right)_{1 \le i,j \le n},$$

say $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$. Therefore, we could approximate the null distribution of $nW_{1n}(\hat{\theta})$ (and thus that of $nW_n(\hat{\theta})$) through the conditional distribution, given $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, of

$$nW^*_{1n} = \sum_{j=1}^{n} \hat{\lambda}_j \chi^2_{1j}.$$

This is tantamount to approximate the null distribution of $nW_{1n}(\hat{\theta})$ by means of the conditional distribution, given $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, of

$$W_1^* = \frac{1}{n}\sum_{i,j=1}^{n} K(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)\xi_i\xi_j,$$

where $\xi_1,\ldots,\xi_n$ are iid from a standard normal distribution, $N(0,1)$, independent of $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$, that is, by means of the WB distribution of $nW_{1n}(\hat\theta)$, in the sense of Burke (2000). The main problem with this approach is that $K(x,y;\theta)$ is unknown because it depends on θ, which is unknown, and because it also depends on $\ell(x;\theta)$, which is usually unknown. To overcome this problem we replace θ by $\hat\theta$ and $\ell(x;\theta)$ by $\hat\ell(x;\hat\theta)$ which is assumed to satisfy

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}\|\ell_1(\boldsymbol{X}_j;\theta)-\hat\ell(\boldsymbol{X}_j;\hat\theta)\|^2 \xrightarrow{P} 0,\\ &\text{with } E\{\|\ell_1(\boldsymbol{X};\theta)\|^2\}<\infty \text{ and } \ell_1(x;\theta)=\ell(x;\theta) \text{ if } H_0 \text{ is true.}\end{aligned} \tag{5}$$

So, instead of $nW_{1n}^*(\hat\theta)$ we consider

$$nW_{2n}^*(\hat\theta) = \sum_{j=1}^{n}\tilde\lambda_j\chi^2_{1j},$$

where $\tilde\lambda_1,\ldots,\tilde\lambda_n$ are the eigenvalues of the matrix

$$\hat H_n = \left(\frac{1}{n}\hat K(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)\right)_{1\le i,j\le n},$$

with $\hat K(x,y;\theta)=h(x,y;\theta)+\hat\ell(x;\theta)\mu(y;\theta)^\intercal+\hat\ell(y;\theta)\mu(x;\theta)^\intercal+\hat\ell(x;\theta)S(\theta)\hat\ell(y;\theta)^\intercal$. The next theorem gives the limit of the conditional distribution of $nW_{2n}^*(\hat\theta)$, given $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$.

Theorem 4 *Let $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$ be iid from $\boldsymbol{X}=(X_1,X_2)\in\mathbb{N}_0^2$ with $E(X_k^2)<\infty$, $k=1,2$. Suppose that $\hat\theta \xrightarrow{P} \theta$, for some $\theta\in\Theta$ and that (5) holds. Then,*

$$\sup_x \left|P_*\left\{nW_{2n}^*(\hat\theta)\le x\right\}-P\{W_1\le x\}\right| \xrightarrow{P} 0, \tag{6}$$

where $W_1=\sum_{j\ge1}\lambda_{1j}\chi^2_{1j}$, $\{\lambda_{1j}\}$ are the non-null eigenvalues of the operator $C_1(\theta)$ defined on the function space $\{\tau:\mathbb{N}_0^2\to\mathbb{R}$, such that $E\left[\tau^2(\boldsymbol{X})\right]<\infty\}$, as follows

$$C_1(\theta)\tau(x)=E\{K_1(x,\boldsymbol{X};\theta)\tau(\boldsymbol{X})\},$$

with $K_1(x,y;\theta)=h(x,y;\theta)+\ell_1(x;\theta)\mu(y;\theta)^\intercal+\ell_1(y;\theta)\mu(x;\theta)^\intercal+\ell_1(x;\theta)S(\theta)\ell_1(y;\theta)^\intercal$.

Remark 1 If in addition to the assumptions in Theorem 4 we assume that $\hat{\theta} \xrightarrow{a.s.} \theta$ and that the limit in (5) is a.s., then the convergence in (6) is a.s.

Remark 2 The result in Theorem 4 keeps on being true if instead of using the raw multipliers, $\xi_1, \ldots, \xi_n$, we use the centered multipliers, $\xi_1 - \bar{\xi}, \ldots, \xi_n - \bar{\xi}$, as suggested in Burke (2000) and Kojadinovic and Yan (2012).

Let $w^*_{2,n,\alpha} = \inf\{x : P_*(W^*_{2n}(\hat{\theta}) \geq x) \leq \alpha\}$ be the α upper percentile of the WB distribution of $W_n(\hat{\theta})$. From Theorems 2 and 4, the test function

$$\Psi^*_{WB} = \begin{cases} 1, & \text{if } W_n(\hat{\theta}) \geq w^*_{2,n,\alpha}\,, \\ 0, & \text{otherwise,} \end{cases}$$

or equivalently, the test that rejects H_0 when $p^* = P_*\left(W^*_{2n}(\hat{\theta}) \geq W_{obs}\right) \leq \alpha$, is asymptotically correct.

3.3. Behaviour against alternatives

This subsections shows that, in contrast to the tests given by Crockett (1979), Loukas and Kemp (1986) and Rayner and Best (1995), the tests Ψ^*_{PB} and Ψ^*_{WB} are consistent, that is, they are able to detect any fixed alternative.

As an immediate consequence of Theorems 1 and 3 (Theorems 1 and 4) the next result gives the asymptotic power of the test Ψ^*_{PB} (Ψ^*_{WB}) against fixed alternatives.

Corollary 1 *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be iid from $\boldsymbol{X} \in \mathbb{N}_0^2$ with pgf $g(u)$. Suppose that assumptions in Theorems 1 and 3 hold. If $\eta(P;\theta) > 0$, then $P(\Psi^*_{PB} = 1) \to 1$.*

Corollary 2 *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be iid from $\boldsymbol{X} \in \mathbb{N}_0^2$ with pgf $g(u)$. Suppose that assumptions in Theorems 1 and 4 hold. If $\eta(P;\theta) > 0$, then $P(\Psi^*_{WB} = 1) \to 1$.*

It can be shown that the proposed tests are also able to detect local alternatives converging to the null at the rate $n^{-1/2}$. The statement and the proof of this result are quite similar to those of Theorem 4 in NJ, for the PB, and of Theorem 4 in Jiménez-Gamero and Kim (2015), for the WB. So, in order to save space, we omit it.

Although the tests Ψ^*_{PB} and Ψ^*_{WB} both asymptotically correct and consistent, their power for finite sample sizes differ. This point will be numerically studied by simulation in Section 5.

4. Some practical considerations

4.1. Bootstrap algorithms

In practice, the exact bootstrap estimator of the null distribution of $W_n(\hat{\theta})$ cannot be calculated. As usual, we approximate it by simulation as follows:

PB algorithm

1. Estimate θ through $\hat{\theta}$ and compute the observed value of the test statistic W_{obs}.
2. For some large integer B, repeat for every $b \in \{1,\ldots,B\}$:
   (a) Generate $\boldsymbol{X}^{*b} = (\boldsymbol{X}_1^{*b}, \boldsymbol{X}_2^{*b}, \ldots, \boldsymbol{X}_n^{*b})$, where $\boldsymbol{X}_1^{*b}, \boldsymbol{X}_2^{*b}, \ldots, \boldsymbol{X}_n^{*b}$ are iid from a $BP(\hat{\theta})$.
   (b) Calculate the test statistic evaluated at $\boldsymbol{X}^{*b}$, obtaining $W_n^{*b}(\hat{\theta}^{*b})$.
3. Approximate the p-value by $\hat{p} = \frac{1}{B}\sum_{b=1}^{B} I\{W_n^{*b}(\hat{\theta}^{*b}) > W_{obs}\}$.

In contrast to the PB distribution, the exact WB estimator of the null distribution of $W_n(\hat{\theta})$ can be calculated by using some numerical approximation method, as for example Imhof's (1961) method. Thus, to calculate the WB distribution of $W_n(\hat{\theta})$ we can proceed as follows:

WB algorithm 1

1. Estimate θ through $\hat{\theta}$ and compute the observed value of the test statistic W_{obs}.
2. Calculate $m_{ij} = \hat{K}(\boldsymbol{X}_i, \boldsymbol{X}_j; \hat{\theta})$, $1 \le i \le j \le n$. Note that $m_{ji} = m_{ij}$.
3. Calculate the eigenvalues of $\hat{H}_n$, $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$.
4. Approximate the p-value by $\hat{p} = P_*\left(\sum_{j=1}^{n} \tilde{\lambda}_j \chi_{1j}^2 > W_{obs}\right)$.

The WB estimator can be also approximated by simulation as follows:

WB algorithm 2

1. Estimate θ through $\hat{\theta}$ and compute the observed value of the test statistic W_{obs}.
2. Calculate $m_{ij} = \hat{K}(\boldsymbol{X}_i, \boldsymbol{X}_j; \hat{\theta})$, $1 \le i \le j \le n$. Note that $m_{ji} = m_{ij}$.
3. For some large integer B, repeat for every $b \in \{1,\ldots,B\}$:
   (a) Generate n iid $N(0,1)$ variates $\xi_1, \ldots, \xi_n$.
   (b) Calculate $W_{2n}^{*b}(\hat{\theta}) = \frac{1}{n^2}\sum_{i,j} \xi_i \xi_j m_{ij}$ (or $W_{2n}^{*b}(\hat{\theta}) = \frac{1}{n^2}\sum_{i,j}(\xi_i - \bar{\xi})(\xi_j - \bar{\xi}) m_{ij}$, as observed in Remark 2).
4. Approximate the p-value by $\hat{p} = \frac{1}{B}\sum_{b=1}^{B} I\{W_{2n}^{*b}(\hat{\theta}) > W_{obs}\}$.

4.2. Point estimators

All above theory assumes that the considered estimator $\hat{\theta}$ satisfies Assumption 1. Commonly used estimators such as maximum likelihood estimators (MLE) and method of moment estimators (MME) satisfy it. Lemmas 1 and 3 in Jiménez-Gamero and Kim (2015) show that the functions ℓ associated to MLEs and MMEs can be approximated by $\hat{\ell}$ satisfying (5), and give the expressions of such approximations. Specifically, if θ is estimated by means of its MLE, then a choice for $\hat{\ell} = \hat{\ell}_{ML}$ satisfying (5) is

$$\hat{\ell}_{ML}((x_1,x_2);\theta) = \left(x_1-\theta_1, x_2-\theta_2, \theta_3\left(\frac{P_\theta(x_1-1,x_2)}{P_\theta(x_1,x_2)} + \frac{P_\theta(x_1,x_2-1)}{P_\theta(x_1,x_2)} - 2\right)\right.$$
$$\left. + f(\theta)\left(\frac{P_\theta(x_1-1,x_2-1)}{P_\theta(x_1,x_2)} - \frac{P_\theta(x_1-1,x_2)}{P_\theta(x_1,x_2)} - \frac{P_\theta(x_1,x_2-1)}{P_\theta(x_1,x_2)} + 1\right)\right),$$

where

$$f(\theta) = \frac{\theta_3^2(\theta_1+\theta_2-2\theta_3)(Q-1) - \theta_3^2 + (\theta_1-2\theta_3)(\theta_2-2\theta_3)}{(\theta_1\theta_2-\theta_3^2)(Q-1)-\theta_1-\theta_2+2\theta_3},$$
$$Q = \sum_{i,j\in\mathbb{N}_0} \frac{P_\theta(i-1,j-1)^2}{P_\theta(i,j)}.$$

If θ is estimated by means of its MME, then a choice for $\hat{\ell} = \hat{\ell}_{MM}$ satisfying (5) is

$$\hat{\ell}_{MM}((x_1,x_2);\theta) = (x_1-\theta_1, x_2-\theta_2, -\theta_2(x_1-\theta_1) - \theta_1(x_2-\theta_2) + x_1x_2 - \theta_3 - \theta_1\theta_2).$$

5. Finite sample performance

The properties so far studied are asymptotic. To study the finite sample performance of the proposed tests, we conducted a simulation experiment. In this section we briefly describe it and display a summary of the results obtained. All computations in this paper were performed by using programs written in the R language (R Development Core Team, 2015).

We started by comparing the proposed approximations to the null distribution of the test statistic $W_n(\hat{\theta})$ from the point of view of the required time to get a p-value. Several values of θ_1, θ_2 and θ_3 were considered. We observed that the value of θ_3 has almost no influence in the required computation time. In contrast, the values of θ_1 and θ_2 have a high impact. We also tried two methods to estimate the parameters: maximum likelihood (ML) and the method of moments (MM), and observed that the choice of the method has little mark on the consumed time. The method used to estimate the null distribution has a high repercussion on the consumed time. In order to value some of these facts,

Table 1: *CPU time (in seconds) to get a p-value with $B = 1000$.*

ML		$n=50$			$n=100$			$n=200$		
θ_1	θ_2	PB	WB2	WB1	PB	WB2	WB1	PB	WB2	WB1
1	1	8.37	0.11	0.07	9.31	0.25	0.15	10.25	0.72	0.46
1	3	13.15	0.14	0.09	17.20	0.27	0.18	19.72	0.74	0.49
3	3	23.57	0.14	0.10	30.12	0.27	0.17	41.15	0.74	0.46
3	10	57.27	0.22	0.22	60.90	0.36	0.29	87.06	0.82	0.59
10	10	132.32	0.28	0.22	187.57	0.36	0.30	277.43	0.86	0.60
10	50	188.64	2.08	2.17	317.47	2.34	2.42	449.14	3.29	2.89
50	50	621.02	2.29	2.40	1160.52	2.48	2.43	2340.78	3.45	3.03
MM		$n=50$			$n=100$			$n=200$		
θ_1	θ_2	PB	WB2	WB1	PB	WB2	WB1	PB	WB2	WB1
1	1	7.69	0.11	0.07	9.82	0.25	0.15	10.53	0.72	0.45
1	3	11.62	0.13	0.10	14.89	0.26	0.17	21.18	0.73	0.47
3	3	25.73	0.14	0.10	31.81	0.28	0.17	43.13	0.74	0.46
3	10	69.31	0.22	0.20	60.15	0.36	0.28	79.80	0.81	0.57
10	10	88.90	0.27	0.22	195.39	0.38	0.31	278.02	0.85	0.58
10	50	174.27	2.07	2.17	280.04	2.31	2.43	462.41	3.26	2.91
50	50	717.87	2.28	2.24	1172.18	2.48	2.38	2402.07	3.43	2.89

Table 1 displays the CPU consumed time (in seconds) to get a p-value for several values of θ_1 and θ_2. The value of θ_3 was set so that the correlation coefficient between the variables, $\rho = \theta_3/\sqrt{\theta_1\theta_2}$, is equal to 0.5. To calculate the PB approximation and the approximation in WB algorithm 2 we took $B = 1000$. There is almost no difference in using the raw multipliers and the centered multipliers in WB algorithm 2. To calculate the p-value of the approximation in WB algorithm 1 we used the function `imhof` of the package `CompQuadForm` of the R language (Duchesne and Lafaye De Micheaux, 2010). From the results in this table it becomes evident that the PB is much more time consuming than the WB, specially for large values of θ_1, θ_2 and the sample size. There are small differences between WB algorithm 1 and WB algorithm 2.

We then studied the goodness of the proposed bootstrap approximations to the null distribution of the test statistic for finite sample sizes. With this aim, we generated 1000 samples of size $n = 50, 100, 200, 300$ from a $BP(\theta_1, \theta_2, \theta_3)$, for several values of θ_1 and θ_2, with θ_3 such that ρ equals to 0.25 and 0.75, in order to examine the approximations for low and high correlated data, respectively, when $\theta_1 = \theta_2$, and $\rho = 0.25$ for $\theta_1 \neq \theta_2$ ($\rho = 0.75$ was not considered because it gives values of θ_3 out of the parametric space for the tried values of $\theta_1 \neq \theta_2$). Because of the results in Table 1, for $\theta_1 = \theta_2 = 50$, the PB was only tried for $n = 50, 100$. For $\theta_1 = \theta_2 = 50$ the WB was also tried for greater sample sizes. For each sample, the p-values were calculated with $B = 500$. The p-values obtained with the WB approximation calculated by means of simulation (WB algorithm 2 with raw and centered multipliers) and numerical approximation (WB

Table 2: *Simulation results for the type I error probability for nominal levels $\alpha = 5\%$ and 10%.*

		Ψ^*_{PB}				Ψ^*_{WB}			
		ML		MM		ML		MM	
	n	5%	10%	5%	10%	5%	10%	5%	10%
$\theta_1 = 1$	50	3.1	7.5	3.5	7.6	3.3	7.8	3.0	7.8
$\theta_2 = 1$	100	3.9	8.6	4.0	9.2	4.2	9.1	3.9	8.8
$\rho = 0.25$	200	4.4	8.4	4.5	8.6	4.8	8.8	4.8	8.6
	300	5.0	10.3	5.5	10.7	4.9	10.3	5.3	10.5
$\theta_1 = 1$	50	3.2	7	3.7	8.5	4.3	8.3	4.6	9.6
$\theta_2 = 1$	100	4.5	8.4	4.1	8.9	4.9	9.1	4.5	9.8
$\rho = 0.75$	200	4.3	8.7	4.7	9.1	4.5	9.2	4.4	9.3
	300	4.7	9.4	4.8	9.6	5.1	10.1	4.4	9.4
$\theta_1 = 1$	50	3.4	7.7	3.6	7.5	2.9	6.6	2.9	6.7
$\theta_2 = 3$	100	3.4	8.3	3.8	8.3	2.8	8.3	2.7	8.3
$\rho = 0.25$	200	4.2	9.1	4.5	9.1	4.4	9.0	4.5	9.1
	300	4.5	9.8	4.6	9.3	4.2	9.7	4.0	9.9
$\theta_1 = 3$	50	3.4	7.7	3.4	7.2	1.6	5.6	1.8	5.3
$\theta_2 = 3$	100	3.3	8.0	3.2	7.7	2.3	6.0	2.3	5.9
$\rho = 0.25$	200	4.4	7.7	4.2	8.0	3.5	7.6	3.2	7.3
	300	4.2	9.6	4.5	9.3	3.7	8.4	3.8	8.6
$\theta_1 = 3$	50	3.7	7.4	4.0	8.8	3.1	7.1	2.9	6.7
$\theta_2 = 3$	100	3.8	8.1	4.0	8.9	3.3	7.6	2.8	7.2
$\rho = 0.75$	200	5.5	10.2	5.5	9.9	5.4	10.5	5.2	10.3
	300	5.3	9.8	5.1	9.7	4.8	10.0	4.4	9.6
$\theta_1 = 3$	50	4.7	9.4	4.3	9.1	2.3	5.8	2.3	5.2
$\theta_2 = 10$	100	5.1	9.1	4.8	8.6	3.0	6.3	3.0	6.4
$\rho = 0.25$	200	5.4	9.9	5.9	9.9	4.3	9.0	4.3	8.6
	300	5.5	10.6	5.6	10.7	4.3	9.7	4.2	10.0
$\theta_1 = 10$	50	3.6	7.3	3.3	6.8	1.0	3.1	1.0	3.0
$\theta_2 = 10$	100	3.9	7.7	4.0	7.3	1.1	3.8	1.1	3.9
$\rho = 0.25$	200	4.4	9.6	4.7	9.2	2.3	6.7	2.3	6.5
	300	4.8	9.5	4.7	8.9	3.2	6.7	3.2	6.9

		Ψ^*_{PB}				Ψ^*_{WB}			
		ML		MM		ML		MM	
	n	5%	10%	5%	10%	5%	10%	5%	10%
$\theta_1 = 10$	50	3.8	7.9	3.9	8.2	1.7	4.2	1.5	3.9
$\theta_2 = 10$	100	3.7	7.9	3.7	8.5	2.0	5.5	1.9	5.3
$\rho = 0.75$	200	4.0	8.3	3.8	8.9	2.6	7.1	2.6	7.0
	300	5.5	9.8	5.3	9.5	4.7	8.7	4.6	8.7
$\theta_1 = 10$	50	3.8	8.4	2.9	7.9	0.0	0.7	0.0	0.6
$\theta_2 = 50$	100	3.9	8.5	3.7	8.0	0.4	2.1	0.4	2.0
$\rho = 0.25$	200	4.2	9.4	3.8	9.0	0.8	3.9	0.9	3.9
	300	5.9	10.0	5.8	10.1	2.3	7.0	2.5	7.0
$\theta_1 = 50$	50	1.1	3.9	1.1	3.6	0.0	0.1	0.0	0.1
$\theta_2 = 50$	100	1.2	4.0	1.2	3.5	0.1	0.2	0.1	0.2
$\rho = 0.25$	200					0.2	1.5	0.4	1.4
	300					0.2	2.2	0.2	2.2
	500					0.5	2.2	0.5	2.2
	1000					2.8	6.2	2.8	6.2
	2000					3.7	6.9	3.8	6.9
	3000					3.7	7.6	3.7	7.7
	4000					4.9	10.0	4.9	10.1
$\theta_1 = 50$	50	3.2	7.0	1.2	3.5	0.1	0.2	0.1	0.2
$\theta_2 = 50$	100	3.8	8.5	3.7	8.3	0.1	0.9	0.1	0.8
$\rho = 0.75$	200					0.0	0.9	0.0	1.0
	300					0.3	1.7	0.3	1.7
	500					1.5	5.5	1.5	5.6
	1000					2.3	7.2	2.3	7.3
	2000					4.0	8.5	3.9	8.3
	3000					5.1	9.1	5.1	9.2
	4000					4.7	9.3	4.6	9.2

algorithm 1 with Imhof's method) were, as expected, quite close. As for raw multipliers versus centered multipliers, a bit better results are obtained when using the centered multipliers. Table 2 displays the fraction of estimated p-values less than or equal to 0.05 and 0.10, which are the estimated type I error probabilities for $\alpha = 0.05$ and 0.10, respectively by using PB and WB with centered multipliers. From the results in this table it can be concluded that both approximations give rise to conservative tests for small sample sizes. As the values of θ_1 and θ_2 increase, the tests become more conservative, specially the one based on the WB approximation. For example, when $\theta_1 = \theta_2 = 50$ and $\rho = 0.25$, the sample size required to get empirical levels close to the nominal values is $n = 4000$. For $\theta_1 = \theta_2 = 50$ and $\rho = 0.75$, $n = 3000$ is enough. In general, better results (in the sense of closeness to the nominal values) are obtained for $\rho = 0.75$ than for $\rho = 0.25$. Finally, it is also observed a bit better results when the parameter is estimated by the maximum likelihood estimator.

To study the power we repeated the above experiment for samples with size $n =$ 50 from the following alternatives: bivariate binomial distribution $BB(m; p_1, p_2, p_3)$, where $p_1 + p_2 - p_3 \leq 1$, $p_1 \geq p_3, p_2 \geq p_3$ and $p_3 > 0$; bivariate Hermite distribution $BH(\mu, \sigma^2; \lambda_1, \lambda_2, \lambda_3)$, where $\mu > \sigma^2(\lambda_1 + \lambda_2 + \lambda_3)$; bivariate logarithmic series distribution $BLS(\lambda_1, \lambda_2, \lambda_3)$, where $0 < \lambda_1 + \lambda_2 + \lambda_3 < 1$; bivariate Neyman type A distribution $BNTA(\lambda; \lambda_1, \lambda_2, \lambda_3)$, where $0 < \lambda_1 + \lambda_2 + \lambda_3 \leq 1$; bivariate Poisson distribution mixtures of the form $pBP(\theta) + (1-p)BP(\lambda), 0 < p < 1$, denoted by $BPP(p; \theta, \lambda)$; and (X_1, X_2) with $X_1 = \max\{Y_1, Y_3\}$ and $X_2 = |Y_1 - Y_3|$ (type 1), $X_1 = \max\{Y_2, Y_3\}$ and $X_2 = |Y_2 - Y_3|$ (type 2), $X_1 = \max\{Y_1, Y_3\}$ and $X_2 = \min\{Y_2, Y_3\}$ (type 3), $X_1 = \max\{Y_2, Y_3\}$ and $X_2 = \min\{Y_1, Y_3\}$ (type 4), $X_1 = \max\{Y_1, Y_3\}$ and $X_2 = \max\{Y_2, Y_3\}$ (type 5), where Y_1, Y_2, Y_3 are independent variables taking values in $\mathbb{N}_0$ whose distribution are binomial $B(m; p)$, negative binomial $BN(m; p)$, Poisson $P(\lambda)$ and uniform on $1, 2, \ldots, m$, $U(m)$. The values of the parameters were chosen so that the expectations $E(X_1)$ and $E(X_2)$ are small for the PB and the WB not to be excessively conservative. In this part of the simulation experiment we only considered the maximum likelihood estimator of the parameter.

In addition to the tests proposed in this paper, Ψ^*_{PB} and Ψ^*_{WB}, we also considered the tests given in Crockett (1979) (denoted by T), Loukas and Kemp (1986) (denoted by I_B), Rayner and Best (1995) (denoted by NI_B) and NJ (denoted by R_n and S_n, with weight function $w(u) = 1$). Table 3 displays the alternatives considered and the estimated power for nominal significance level $\alpha = 0.05$. Looking at this table we conclude that the tests Ψ^*_{PB}, Ψ^*_{WB}, R_n and S_n are able to detect all considered alternatives while, as expected, the other tests cannot, specially the tests based on I_B and NI_B. For the alternatives in the first half of Table 3 we see that the powers of the new tests, R_n and S_n are quite close; while for the other alternatives the tests proposed in this paper are more powerful than R_n and S_n. We also compared these tests from a computational point of view. From the results in Table 1 we saw that, in this respect, Ψ^*_{WB} is more efficient than Ψ^*_{PB}. Since R_n and S_n are both based on a PB, for the comparisons to be fair, we compared Ψ^*_{PB}, R_n and S_n. Table 4 reports the ratio of the average CPU to get a p-value. Clearly, regarding the required computing time, Ψ^*_{PB} is more efficient than R_n and S_n.

Table 3: *Simulation results for the power.*

Alternative	$E(X_1)$	$\frac{var(X_1)}{E(X_1)}$	$E(X_2)$	$\frac{var(X_2)}{E(X_2)}$	ρ	R_n	S_n	Ψ^*_{PB}	Ψ^*_{WB}	T	I_B	NI_B
$BB(1,0.45,0.02,0.01)$	0.450	0.550	0.020	0.980	0.014	0.953	0.961	0.944	1.000	0.263	0.000	0.000
$BB(1,0.55,0.03,0.02)$	0.550	0.450	0.030	0.970	0.041	0.998	0.999	0.998	1.000	0.768	0.000	0.000
$BB(2,0.71,0.04,0.03)$	1.420	0.290	0.080	0.960	0.018	0.997	0.993	1.000	1.000	0.999	0.000	0.000
$BH(0.99,1,0.66,0.10,0.10)$	0.752	1.768	0.198	1.202	0.446	0.963	0.985	0.996	0.982	0.747	0.809	0.842
$BH(1.40,1,1.00,0.26,0.12)$	1.568	1.800	0.532	1.271	0.430	0.967	0.983	0.999	0.999	0.795	0.849	0.899
$BH(1.50,1,1.00,0.38,0.10)$	1.650	1.733	0.720	1.320	0.411	0.939	0.971	0.989	0.968	0.771	0.849	0.887
$BLS(0.30,0.01,0.11)$	1.298	0.409	0.380	0.827	0.303	1.000	1.000	1.000	1.000	0.830	0.012	0.006
$BLS(0.40,0.01,0.02)$	1.311	0.426	0.094	0.959	0.039	1.000	1.000	1.000	1.000	0.763	0.008	0.007
$BLS(0.50,0.01,0.02)$	1.465	0.641	0.085	0.083	0.093	1.000	1.000	1.000	1.000	0.446	0.027	0.025
$BNTA(0.1,0.01,0.01,0.93)$	0.094	1.940	0.094	1.940	0.995	0.797	0.793	0.860	0.225	0.565	0.007	0.599
$BNTA(0.1,0.01,0.01,0.92)$	0.093	1.930	0.093	1.930	0.994	0.810	0.810	0.869	0.222	0.580	0.005	0.617
$BNTA(0.1,0.01,0.01,0.95)$	0.096	1.960	0.096	1.960	0.995	0.835	0.833	0.884	0.188	0.604	0.003	0.629
$BPP(0.30;(0.2,0.2,0.1),(0.9,0.9,0.5))$	0.690	1.149	0.690	1.149	0.665	0.846	0.844	0.623	0.729	0.548	0.003	0.001
$BPP(0.31;(0.2,0.2,0.1),(1.0,1.2,0.9))$	0.752	1.182	0.890	1.240	0.909	0.726	0.652	0.673	0.755	0.527	0.004	0.001
$BPP(0.35;(0.2,0.2,0.1),(0.9,0.9,0.6))$	0.655	1.170	0.655	1.170	0.778	0.805	0.799	0.618	0.722	0.516	0.001	0.000
$B\_B\_P(2,0.1;2,0.1;0.9)$, type 1	0.983	0.854	0.870	0.935	0.937	0.229	0.486	0.992	0.938	0.054	0.045	0.058
$B\_P\_P(2,0.1;0.9;0.9)$, type 1	0.981	0.850	0.867	0.930	0.936	0.224	0.481	0.991	0.926	0.081	0.06	0.058
$B\_U\_U(2,0.1;2;2)$, type 1	1.070	0.566	0.941	0.643	0.902	0.066	0.265	1.000	1.000	0.483	0.000	0.000
$B\_B\_P(2,0.1;2,0.1;2.1)$, type 2	2.126	0.950	1.953	1.063	0.962	0.329	0.455	0.965	0.773	0.060	0.075	0.158
$B\_B\_U(2,0.1;2,0.1;2)$, type 2	1.069	0.566	0.940	0.641	0.901	0.073	0.289	1.000	1.000	0.491	0.000	0.000
$B\_P\_U(2,0.1;0.1;2)$, type 2	1.032	0.618	0.966	0.659	0.949	0.095	0.448	1.000	1.000	0.350	0.000	0.001
$B\_BN\_P(2,0.1;2,0.1;2.1)$, type 3	2.129	0.950	2.032	0.997	0.951	0.370	0.516	1.000	0.995	0.028	0.080	0.134
$P\_BN\_B(0.1;2,0.1;2,0.1)$, type 3	0.281	0.825	0.196	0.906	0.796	0.120	0.350	0.838	0.851	0.007	0.051	0.018
$P\_BN\_P(0.6;2,0.1;0.6)$, type 3	0.976	0.725	0.586	1.003	0.644	0.123	0.205	0.848	0.825	0.177	0.066	0.026
$BN\_B\_B(2,0.1;2,0.1;2,0.15)$, type 4	0.446	0.695	0.296	0.848	0.750	0.033	0.146	0.855	0.914	0.063	0.006	0.001
$BN\_B\_P(2,0.1;2,0.1;0.4)$, type 4	0.537	0.790	0.394	0.993	0.833	0.176	0.355	0.927	0.850	0.085	0.087	0.053
$BN\_B\_P(2,0.1;2,0.1;0.5)$, type 4	0.623	0.809	0.493	1.002	0.872	0.180	0.407	0.971	0.899	0.114	0.087	0.053
$B\_U\_B(2,0.1;1;2,0.15)$, type 5	0.445	0.700	0.662	0.406	0.428	0.994	0.996	0.982	0.995	0.940	0.000	0.000
$B\_U\_U(2,0.1;1;1)$, type 5	0.605	0.427	0.752	0.248	0.462	1.000	1.000	1.000	1.000	1.000	0.000	0.000
$B\_U\_U(2,0.1;3;3)$, type 5	1.557	0.741	2.126	0.404	0.597	0.300	0.187	0.953	0.952	0.580	0.000	0.000

Table 4: *Ratio of average CPU time (in seconds).*

	$n=30$	$n=50$	$n=70$	$n=100$	$n=200$	$n=300$
R_n/Ψ^*_{PB}	73.50	75.71	77.44	80.19	79.69	79.20
S_n/Ψ^*_{PB}	5.01	11.07	20.20	43.73	145.28	303.92

Table 5: *Results for the real data sets.*

	Plants	Health
$R_{n,(0,0)}$	0.003	0.000
$R_{n,(1,0)}$	0.005	0.000
$R_{n,(0,1)}$	0.010	0.000
$S_{n,(0,0)}$	0.005	0.002
$S_{n,(1,0)}$	0.009	0.000
$S_{n,(0,1)}$	0.011	0.000
Ψ^*_{PB}	0.049	0.000
$\hat{\theta}_n$	(0.64000, 0.94000, 0.19852)	(0.30173, 1.21830, 0.12518)

To end this section, Ψ^*_{PB} is applied to two real data sets. The first one were first given and analysed by Holgate (1966), and refers to the number of plants of the species *Lacistema aggregatum* and *Protium guianense* in each of 100 contiguous quadrats. Crockett (1979), Loukas and Kemp (1986), Rayner and Best (1995) and NJ tested the data for agreement with the bivariate Poisson model, they all concluded the data were not well modelled by a BPD. The second data set were analysed in Karlis and Tsiamyrtzis (2008), who used two variables, the number of consultations with a doctor or a specialist (X_1) and the total number of prescribed and non-prescribed medications used in past 2 days (X_2), from the Australian Health survey for 1977–1978. The sample size was quite large ($n = 5190$). These authors assumed that (X_1, X_2) has a BPD. NJ tested these data sets for agreement with the bivariate Poisson model, concluding that they were not well modelled by a BPD. The p-values obtained by applying the test proposed in this paper to these two real data sets are 0.049 and 0.000, respectively, in agreement with the previous analyses.

Table 6: *Simulations results for the type I error probabilities when $\theta_3 = 0$.*

	$\theta_1=\theta_2=1$				$\theta_1=\theta_2=3$				$\theta_1=\theta_2=10$			
	ML		MM		ML		MM		ML		MM	
n	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
100	3.4	7.4	3.5	7.9	3.1	7.3	3.0	7.3	1.0	3.4	0.9	3.4
200	4.2	8.0	4.3	9.1	3.4	8.0	3.3	7.9	2.2	6.6	2.3	6.6
300	4.4	8.7	4.6	9.4	3.7	8.5	3.7	8.5	3.7	7.9	3.7	8.0

6. Case $\theta_3 = 0$

The case $\theta_3 = 0$ has been excluded from H_0 because it is a boundary point. It is well-known (see, for example Andrews, 1999, Self and Liang, 1987, and the references therein) that in such a case the MLE is not asymptotically normally distributed and thus Assumption 1 is not satisfied. Moreover, Andrews (2000) have proven that the bootstrap does not provides a consistent estimator of the distribution. Therefore, the theory so far developed is not valid for $\theta_3 = 0$.

Next we give two possible ways of dealing with this case. A first way consist in applying the method in Feng and McCulloch (1992), which proposed to enlarge the parametric space to $\theta \in \mathbb{R}^3$, so that negative values for $\hat{\theta}_3$ are allowed. With this approach all required assumptions in our theory are satisfied. The only problem with this solution is how to apply in practice the PB approximation because it implies the generation of samples from a $BP(\theta_1, \theta_2, \theta_3)$ distribution with $\theta_3 < 0$. Nevertheless, the WB approximation can be applied. Table 6 gives the result of a small simulation that studies the goodness of this solution. Observe that the results are quite close to those obtained for $\theta_3 > 0$.

Another possible way of dealing with this case is to adapt the alternatives to the usual bootstrap proposed in Andrews (2000). Two of them consists in subsampling, while the other two are based on testing if the parameter is in the boundary. For the later methods we could calculate a confidence interval for θ_3 and look if it contains 0 by applying, for example, the method in Feng and McCulloch (1992) but, as recognized by the authors, it requires rather large sample sizes. Note that testing for $\theta_3 = 0$ is tantamount to having two independent Poisson variables. Another way of investigating the independence of the marginal distributions is by applying the classical χ^2-test. Nevertheless, such test requires the data to be grouped in classes, and the decision could depend on the grouping. In our view, there is a need of a test for independence of variables taking values on $\mathbb{N}_0$, which will be the topic of a future research.

If it can be reasonably assumed that the variables are independent, then by using Raikov's theorem (which states that the sum of two independent non-negative random variables has a Poisson distribution if and only if both random variables have the Poisson distribution), testing gof for an independent Poisson model is equivalent to testing gof to the sum of the components to a univariate Poisson model. In the statistical literature there is a variety of test for testing gof to a univariate Poisson model (see, for example, the review in Gürtler and Henze, 2000).

7. The general *m*-variate case

This section shows that the proposed test can be extended to the general m-variate case, for any $m \geq 2$. Let

$$X_1 = Y_1 + Y_{m+1}, \;\; X_2 = Y_2 + Y_{m+1}, \;\; \ldots, \;\; X_m = Y_m + Y_{m+1},$$

where $Y_1, Y_2, \ldots, Y_{m+1}$ are mutually independent Poisson random variables with means $\theta_1' = \theta_1 - \theta_{m+1} > 0, \ldots, \theta_m' = \theta_m - \theta_{m+1} > 0$ and $\theta_{m+1} > 0$, respectively. The joint distribution of the vector $(X_1, X_2, \ldots, X_m)$ is called a m-variate Poisson distribution with parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_{m+1})$ (see Johnson, Kotz and Balakrishnan, 1997). The joint pgf of $(X_1, X_2, \ldots, X_m)$ is

$$g(u;\theta) = \exp\left\{\sum_{i=1}^{m} \theta_i (u_i - 1) + \theta_{m+1}\left(\prod_{i=1}^{m} u_i - \sum_{i=1}^{m} u_i + m - 1\right)\right\}, \;\; \forall u \in \mathbb{R}^m. \qquad (7)$$

Now, the objective is to test the hypothesis

$$H_{0m} : (X_1, X_2, \ldots, X_m) \text{ has a } m\text{-variate Poisson distribution.}$$

In order to extend the proposed test to the general m-variate case we will use the following result in Proposition 3 in NJ which states that $g(u;\theta)$ is the only pgf in $G_m = \{g : [0,1]^m \to \mathbb{R}$, such that g is a pgf and $\frac{\partial}{\partial u_i} g(u_1, u_2, \ldots, u_m)$ exists $\forall u \in [0,1]^m$, $1 \leq i \leq m\}$ satisfying the following system,

$$D_i(u;\theta) = 0, \;\; 1 \leq i \leq m, \qquad (8)$$

$\forall u \in [0,1]^m$, where $D_i(u;\theta) = \dfrac{\partial}{\partial u_i} g(u) - \left\{\theta_i + \theta_{m+1}\left(\prod_{j \neq i} u_j - 1\right)\right\} g(u), \;\; 1 \leq i \leq m.$

Let $(X_1, X_2, \ldots, X_m) \in \mathbb{N}_0^m$ be a random vector and let $g(u_1, u_2, \ldots, u_m) = E\left(u_1^{X_1} u_2^{X_2} \cdots u_m^{X_m}\right)$ its pgf. Then, taking into account that

$$g(u) = \sum_{r_1, r_2, \ldots, r_m \geq 0} u_1^{r_1} u_2^{r_2} \cdots u_m^{r_m} p(r_1, r_2, \ldots, r_m),$$

where $p(r_1, r_2, \ldots, r_m) = P(X_1 = r_1, X_2 = r_2, \ldots, X_m = r_m)$, we can write

$$D_i(u;\theta) = \sum_{r_1, r_2, \ldots, r_m \geq 0} \Big\{(r_i + 1) p(r_1, \ldots, r_{i-1}, r_i + 1, r_{i+1}, \ldots, r_m) - (\theta_i - \theta_{m+1}) p(r_1, r_2, \ldots, r_m)$$
$$- \theta_{m+1} p(r_1 - 1, \ldots, r_{i-1} - 1, r_i, r_{i+1} - 1, \ldots, r_m - 1)\Big\} u_1^{r_1} u_2^{r_2} \cdots u_m^{r_m}, \;\; 1 \leq i \leq m.$$

Let $D_{in}(u;\hat{\theta})$ denote the empirical counterpart of $D_i(u;\theta)$ obtained by replacing the pgf g by the epgf g_n and θ by a consistent estimator $\hat{\theta}$, $1 \le i \le m$. If H_{0m} is true then the functions $D_{in}(u;\hat{\theta})$, $1 \le i \le m$, should be close to 0, $\forall u \in [0,1]^m$. This proximity to zero can be interpreted as we did in Section 2, for the bivariate case. Observe that

$$D_{in}(u;\hat{\theta}) = \sum_{r_1,r_2,\ldots,r_m \ge 0} d_i(r_1,r_2,\ldots,r_m;\hat{\theta}) u_1^{r_1} u_2^{r_2} \cdots u_m^{r_m}, \;\; 1 \le i \le m,$$

where

$$\begin{aligned} d_i(r_1,r_2,\ldots,r_m;\hat{\theta}) &= (r_i+1)p_n(r_1,\ldots,r_{i-1},r_i+1,r_{i+1},\ldots,r_m) \\ &\quad - (\hat{\theta}_i - \hat{\theta}_{m+1}) p_n(r_1,r_2,\ldots,r_m) \\ &\quad - \hat{\theta}_{m+1}\, p_n(r_1-1,\ldots,r_{i-1}-1,r_i,r_{i+1}-1,\ldots,r_m-1), \;\; 1 \le i \le m, \end{aligned}$$

and $p_n(r_1,r_2,\ldots,r_m) = \frac{1}{n}\sum_{k=1}^n I(X_{k1}=r_1, X_{k2}=r_2,\ldots,X_{km}=r_m)$ is the relative frequency of $(r_1,r_2,\ldots,r_m)$. Therefore, $D_{in}(u;\hat{\theta}) = 0$, $\forall u \in [0,1]^m$, $1 \le i \le m$, if and only if the coefficients of $u_1^{r_1} u_2^{r_2} \cdots u_m^{r_m}$ in the previous expansions are null, $\forall r_1,r_2,\ldots,r_m \ge 0$. This leads us to consider the following statistic for testing H_{0m},

$$W_{m,n}(\hat{\theta}) = \sum_{r_1,r_2,\ldots,r_m \ge 0} \left\{ \sum_{i=1}^m d_i(r_1,r_2,\ldots,r_m;\hat{\theta})^2 \right\} = \sum_{r_1,r_2,\ldots,r_m = 0}^{M} \left\{ \sum_{i=1}^m d_i(r_1,r_2,\ldots,r_m;\hat{\theta})^2 \right\},$$

where $M = \max\{X_{(n)1}, X_{(n)2},\ldots,X_{(n)m}\}$, $X_{(n)k} = \max_{1\le i\le n} X_{ik}$, $1 \le k \le m$. Similar results to those stated in Sections 2 and 3 for the bivariate case can be established for $W_{m,n}(\hat{\theta})$.

8. Proofs

Here we give a sketch of the proofs of the results in Sections 2 and 3. A detailed derivation of the results can be obtained from the authors upon request.

Proof of Theorem 1 Observe that

$$d_1(r,s;\hat{\theta}) = d_1(r,s;\theta) - (\hat{\theta}_1 - \theta_1) p_n(r,s) + (\hat{\theta}_3 - \theta_3)\{p_n(r,s) - p_n(r,s-1)\}$$

and

$$\sum_{r,s\ge 0} d_1(r,s;\theta)^2 = \frac{1}{n^2}\sum_{i\ne j} h_1(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta) + \frac{1}{n^2}\sum_{i=1}^n h_1(\boldsymbol{X}_i,\boldsymbol{X}_i;\theta).$$

By the SLLN,

$$\frac{1}{n}\sum_{i=1}^{n} h_1(\boldsymbol{X}_i,\boldsymbol{X}_i;\theta) \xrightarrow{a.s.} E\left\{\sum_{r,s\geq 0}\phi_{1rs}(\boldsymbol{X}_1;\theta)^2\right\} < \infty.$$

By the SLLN for U-statistics (Theorem 5.4 in Serfling, 1980),

$$\frac{1}{n^2}\sum_{i\neq j} h_1(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta) \xrightarrow{a.s.} E\{h_1(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\} = \sum_{r,s\geq 0} a_1(r,s;\theta)^2.$$

Therefore,

$$\sum_{r,s\geq 0} d_1(r,s;\theta)^2 \xrightarrow{a.s.} \sum_{r,s\geq 0} a_1(r,s;\theta)^2.$$

Since $p_n(r,s)^2 \leq p_n(r,s)$, $\forall r,s \geq 0$, and $\sum_{r,s\geq 0} p_n(r,s) = 1$, we have

$$(\hat{\theta}_1-\theta_1)^2\sum_{r,s\geq 0} p_n(r,s)^2 \leq (\hat{\theta}_1-\theta_1)^2 = o(1),$$

and analogously,

$$(\hat{\theta}_3-\theta_3)^2\sum_{r,s\geq 0}\{p_n(r,s)-p_n(r,s-1)\}^2 = o(1).$$

Thus,

$$\sum_{r,s\geq 0} d_1(r,s;\hat{\theta})^2 \xrightarrow{a.s.} \sum_{r,s\geq 0} a_1(r,s;\theta)^2. \tag{9}$$

Following similar steps we get

$$\sum_{r,s\geq 0} d_2(r,s;\hat{\theta})^2 \xrightarrow{a.s.} \sum_{r,s\geq 0} a_2(r,s;\theta)^2. \tag{10}$$

Finally, the result is obtained from (9) and (10). ∎

Proof of Theorem 2 Let us consider the separable Hilbert space of functions $\mathscr{H} = \{g : \mathbb{N}_0 \to \mathbb{R}$, so that $\|g\|_{\mathscr{H}}^2 = \sum_{r\geq 0}\sum_{s\geq 0} g(r,s)^2 < \infty\}$. We have that

$$\sqrt{n}d_k(r,s;\hat{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{krs}(\boldsymbol{X}_i;\theta) + \sqrt{n}(\hat{\theta}-\theta)\hat{v}_k(r,s)^{\intercal}, \quad k=1,2,$$

with $\hat{v}_1(r,s) = (-p_n(r,s), 0, p_n(r,s) - p_n(r,s-1))$ and $\hat{v}_2(r,s) = (0, -p_n(r,s), p_n(r,s) - p_n(r-1,s))$. From Assumption 1 and the SLLN, we get that

$$\sqrt{n}d_k(r,s;\hat{\theta}) = \sqrt{n}d_{1k}(r,s;\theta) + R_k(r,s), \quad k=1,2,$$

with

$$d_{1k}(r,s;\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\phi_{krs}(\boldsymbol{X}_i;\theta) + \ell(\boldsymbol{X}_i;\theta)v_k(r,s;\theta)^{\intercal}\right\}, \quad k=1,2,$$

$$v_1(r,s;\theta) = (-P_\theta(r,s), 0, P_\theta(r,s) - P_\theta(r,s-1)),$$

$$v_2(r,s;\theta) = (0, -P_\theta(r,s), P_\theta(r,s) - P_\theta(r-1,s)),$$

and $\|R_k\|_{\mathscr{H}} = o_P(1)$, $k=1,2$. From the CLT in Hilbert spaces (see, for example, van der Vaart and Wellner, 1996, pp. 50–51), it follows that $\|\sqrt{n}d_{1k}\|^2_{\mathscr{H}} = O_P(1)$, $k=1,2$, and therefore

$$nW_n(\hat{\theta}) = \|\sqrt{n}d_{1k}\|^2_{\mathscr{H}} + \|\sqrt{n}d_{12}\|^2_{\mathscr{H}} + o_P(1).$$

Routine calculations show that

$$\|\sqrt{n}d_{1k}\|^2_{\mathscr{H}} + \|\sqrt{n}d_{12}\|^2_{\mathscr{H}} = \frac{1}{n}\sum_{i,j=1}^{n}K(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta).$$

The result is achieved by applying Theorem 6.4.1.B in Serfling (1980) to $\frac{1}{n}\sum_{i,j=1}^{n}K(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)$. ■

Proof of Theorem 3 Following similar steps to those given in the proof of Theorem 2 but instead of applying the CLT for iid random elements taking values in $\mathscr{H}$, we apply a CLT for triangular arrays, such as Theorem 1.1 in Kundu et al. (2000). ■

Proof of Theorem 4 $nW^*_{2n}(\hat{\theta})$ can be expressed as $nW^*_{2n}(\hat{\theta}) = W^*_1 + W^*_2 + 2W^*_3 + W^*_4$, where

$$W^*_1 = \frac{1}{n}\sum_{i,j=1}^{n}K(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)\xi_i\xi_j,$$

$$W_2^* = \frac{1}{n}\sum_{i,j=1}^{n}\{h(\boldsymbol{X}_i,\boldsymbol{X}_j;\hat{\theta}) - h(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)\}\xi_i\xi_j,$$

$$W_3^* = \frac{1}{n}\sum_{i,j=1}^{n}\{\hat{\ell}(\boldsymbol{X}_i;\hat{\theta})\mu(\boldsymbol{X}_j;\hat{\theta})^\intercal - \ell_1(\boldsymbol{X}_i;\theta)\mu(\boldsymbol{X}_j;\theta)^\intercal\}\xi_i\xi_j,$$

$$W_4^* = \frac{1}{n}\sum_{i,j=1}^{n}\{\hat{\ell}(\boldsymbol{X}_i;\hat{\theta})S(\hat{\theta})\hat{\ell}(\boldsymbol{X}_j;\hat{\theta})^\intercal - \ell_1(\boldsymbol{X}_i;\theta)S(\theta)\ell_1(\boldsymbol{X}_j;\theta)^\intercal\}\xi_i\xi_j.$$

From the results in Delhing and Mikosch (1994),

$$\sup_x |P_*\{W_1^* \leq x\} - P\{W_1 \leq x\}| \xrightarrow{a.s.} 0.$$

Thus, to show the result it suffices to see that $W_k^* = o_{P_*}(1)$ in probability, $k = 2,3,4$. We first deal with W_2^*. Observe that

$$E_*(W_2^{*2}) \leq M\frac{1}{n^2}\sum_{i,j=1}^{n}\{h(\boldsymbol{X}_i,\boldsymbol{X}_j;\hat{\theta}) - h(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)\}^2,$$

for some positive $M > 0$. From the assumptions made, the right-hand side of the above expression is $o_P(1)$. Therefore, $W_2^* = o_{P_*}(1)$ in probability. As for W_3^*, we have that $W_3^* = W_{31}^* W_{32}^{*\top} + W_{33}^* W_{34}^{*\top}$, with

$$W_{31}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\hat{\ell}(\boldsymbol{X}_i;\hat{\theta}) - \ell_1(\boldsymbol{X}_i;\theta)\}\xi_i,$$

$$W_{32}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mu(\boldsymbol{X}_i;\hat{\theta})\xi_i,$$

$$W_{33}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell_1(\boldsymbol{X}_i;\theta)\xi_i,$$

$$W_{34}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\mu(\boldsymbol{X}_i;\hat{\theta}) - \mu(\boldsymbol{X}_i;\theta)\}\xi_i.$$

From the assumptions made, $E_*(W_{31}^{*2}) = o_P(1)$, $E_*(W_{32}^{*2})$ is bounded in probability and $E_*(W_{33}^{*2})$ is bounded a.s.. Now taking into account that

$$\frac{\partial}{\partial\theta_1}P_\theta(r,s) = P_\theta(r-1,s) - P_\theta(r,s),$$

$$\frac{\partial}{\partial\theta_2}P_\theta(r,s) = P_\theta(r,s-1) - P_\theta(r,s),$$

$$\frac{\partial}{\partial\theta_3}P_\theta(r,s) = P_\theta(r-1,s-1) - P_\theta(r-1,s) - P_\theta(r,s-1) + P_\theta(r,s),$$

it follows that

$$\sup_{r,s\in\mathbb{N}_0} |P_{\hat{\theta}}(r,s) - P_\theta(r,s)| \leq M\|\hat{\theta} - \theta\|, \tag{11}$$

for some positive $M > 0$. This implies that $E_*(W_{34}^{*2}) = o_P(1)$. Therefore, $W_3^* = o_{P_*}(1)$ in probability. By using (11) and the assumptions made, it readily follows that $W_4^* = o_{P_*}(1)$ in probability. This concludes the proof. ∎

Acknowledgements

The authors thank the anonymous referees for their valuable time and careful comments, which improved the quality of this paper. F. Novoa-Muñoz wishes to thank his institution, University of Bío-Bío (Chile) and the scholarship given by the Chilean Ministry of Education through the Superior Education MECESUP Program 2, which make his doctorate studies possible. M.D. Jiménez-Gamero acknowledges financial support from grant MTM2014-55966-P of the Spanish Ministry of Economy and Competitiveness.

References

Andrews, D.W.K. (1999) Estimation when a parameter is on a boundary. *Econometrica*, 67, 1341–1383.

Andrews, D.W.K. (2000) Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68, 399–405.

Burke, M.D. (2000). Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. *Statistics & Probability Letters*, 46, 13–20.

Crockett, N. G. (1979). A quick test of fit of a bivariate distribution. In *Interactive Statistics*, D. McNeil (ed.), 185–191. Amsterdam: North-Holland.

Delhing, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *Journal of Multivariate Analysis*, 51, 392–413.

Duchesne, P. and Lafaye De Micheaux, P. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics and Data Analysis*, 54, 858–862.

Feng, Z., McCulloch, C.E. (1992). Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space. *Statistics & Probability Letters*, 13, 325–332.

Gürtler, N. and Henze, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution. *Journal of Statistical Planning and Inference*, 90, 207–225.

Haight, F. A. (1967). *Handbook of the Poisson Distribution*. New York: John Wiley & Sons.

Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika*, 51, 241–245.

Holgate, P. (1966). Bivariate generalizations of Neyman's type a distribution. *Biometrika*, 53, 241–245.

Janssen, A. (2000). Global power functions of goodness of fit tests. *The Annals of Statistics*, 28, 239–253.

Imhof, J.P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419–426.

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.

Johnson, N. L. and Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions*. Wiley, New York.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York.

Karlis, D. and Tsiamyrtzis, P. (2008). Exact Bayesian modelling for bivariate Poisson data and extensions. *Statistics and Computing*, 18, 27–40.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. Marcel Dekker, Inc, New York.

Kojadinovic, I. and Yan, J. (2012) Goodness-of-fit testing based on a weighted bootstrap: A fast large sample alternative to the parametric bootstrap, *Canadian Journal of Statistics*, 40, 2012, 480–500.

Kundu, S., Majumdar, S. and Mukherjee, K. (2000). Central limits theorems revisited. *Statistics & Probability Letters*, 47, 265–275.

Loukas, S. and Kemp, C.D. (1986). The index of dispersion test for the bivariate Poisson distribution. *Biometrics*, 42, 941–948.

Nakamura, M. and Pérez-Abreu, V. (1993). Use of an empirical probability generating function for testing a Poisson model. *Canadian Journal of Statistics*, 21, 149–156.

Novoa-Muñoz, F. and Jiménez-Gamero, M. D. (2014). Testing for the bivariate Poisson distribution. *Metrika*, 77, 771–793.

Papageorgiou, H. and Loukas, S. (1988). Conditional even point estimation for bivariate discrete distributions. *Communications in Statistics–Theory and Methods*, 17, 3403–3412.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Rayner, J.C. W. and Best, D.J. (1995). Smooth Tests for the bivariate Poisson distribution. *Australian & New Zealand Journal of Statistics*, 37, 233–245.

Sahai, H. and Khurshid, A. (1993). Confidence intervals for the mean of a Poisson distribution: A review. *Biometrical Journal*, 35(7), 857–867.

Self, S.G. and Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82, 605–610.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.