DATA DRIVEN EFFICIENT SCORE TESTS FOR POISSONITY

BY

TADEUSZ INGLOT (WROCLAW)

Abstract. New data driven score tests for testing goodness of fit of the Poisson distribution are proposed. They are direct applications of the general construction of data driven goodness-of-fit tests for composite hypotheses developed in Inglot et al. (1997). By a simulation study it is shown that these tests perform almost equally well as the best known solutions for standard alternatives and outperform them for more difficult alternatives.

2010 AMS Mathematics Subject Classification: Primary: 62G10; Secondary: 62G30, 65C05, 65C60.

Key words and phrases: Goodness-of-fit test, Poisson distribution, data driven test, model selection, Monte Carlo study.

1. INTRODUCTION

The Poisson distribution is often used in modelling discrete distributions. So, testing goodness of fit of the Poisson family is an important problem in statistical inference. Beginning from Fisher’s index of dispersion it has a large number of solutions and takes a constant interest in the literature. For a nice overview of existing tests we refer to Gürtler and Henze [3] and Best and Rayner [1]. Some further constructions have been proposed more recently, e.g., by Thas and Rayner [11], Meintanis and Nikitin [8], Frey [2] and Ledwina and Wlupke [7].

In the present note we propose data driven efficient score tests for testing Poissonity which are a direct application of the general construction of data driven goodness-of-fit tests for composite hypotheses studied in Inglot et al. [5]. Our construction is valid for any family of discrete distributions concentrated on nonnegative integers. We focus on the most important case of testing for Poissonity to show that this construction leads to omnibus tests being able to compete with the best existing ones. Paying a little bit of sensitivity for simple alternatives, they cover a much wider class of alternatives with stable and high power. An additional advantage is that exact critical values for moderate sample sizes practically do not depend on the nuisance parameter and therefore can be determined in advance.
2. CONSTRUCTION OF THE TEST STATISTIC

Let \( X_1, \ldots, X_n \) be a sample from a discrete distribution \( P \) on the real line taking values in the set \( \{0, 1, 2, \ldots\} \). Denote by \( P_\lambda \) the Poisson distribution with parameter \( \lambda > 0 \), i.e. \( P_\lambda(\{j\}) = \pi_j(\lambda) = e^{-\lambda} \lambda^j/j! \) for \( j = 0, 1, 2, \ldots \). The problem is to test the composite hypothesis

\[
H_0 : P \in \{P_\lambda : \lambda > 0\}.
\]

Let \( U_1, \ldots, U_n \) be independent random variables uniformly distributed over the unit interval \([0, 1]\) independent of \( X_i \)’s. Let us consider the randomized sample \( Y_1, \ldots, Y_n \), where \( Y_i = X_i + U_i \). Then \( Y_i \)'s have absolutely continuous distribution \( P \) on the half line \([0, \infty)\) with a stepwise density constant on the intervals \([j, j+1)\), \( j \geq 0 \). Now, consider a family \( \mathcal{P} \) of densities on \([0, \infty)\) defined by

\[
(2.1) \quad \mathcal{P} = \left\{ f(y, \lambda) : f(y, \lambda) = \sum_{j=0}^{\infty} \pi_j(\lambda) \mathbf{1}_{[j,j+1)}(y), \ \lambda > 0 \right\},
\]

where \( \mathbf{1}_A(y) \) denotes the indicator of a set \( A \). Since \( X_i \)'s take integer values, we can replace \( H_0 \) by the equivalent hypothesis

\[
(2.2) \quad H'_0 : \mathcal{P} \in \mathcal{P}.
\]

The cumulative distribution function of \( f(y, \lambda) \) from \( \mathcal{P} \) takes the form

\[
F(y, \lambda) = \sum_{r=0}^{j-1} \pi_r(\lambda) + (y - j) \pi_j(\lambda) \quad \text{for} \ y \in [j, j+1), \ j \geq 1,
\]

and \( F(y, \lambda) = y \pi_0(\lambda) \) for \( y \in [0, 1) \). Hence, to test \( H'_0 \) we can simply apply results of Inglot et al. [5].

To this end, let \( \psi_1(t), \psi_2(t), \ldots \) be an orthonormal system of bounded functions on \([0, 1]\) with \( \int_0^1 \psi_j(t)dt = 0 \) and such that \( \partial \log f(F^{-1}(t, \lambda), \lambda) / \partial \lambda \) is linearly independent of \( \psi_1(t), \psi_2(t), \ldots \). Let \( d(n) \) be a nondecreasing sequence of natural numbers. Consider the nested sequence \( G_k, 1 \leq k \leq d(n) \), of exponential families given by the densities

\[
(2.3) \quad g_k(y, \vartheta, \lambda) = c_k(\vartheta) \exp \left\{ \sum_{j=1}^{k} \partial_j \psi_j \left(F(y, \lambda)\right) \right\} f(y, \lambda), \quad y \in [0, \infty),
\]

where \( \vartheta = (\vartheta_1, \ldots, \vartheta_k)^T \in \mathbb{R}^k \) is a vector of parameters, \( v^T \) stands for the transposition of the vector \( v \) and \( c_k(\vartheta) \) is the normalizing constant.

Fix \( k, 1 \leq k \leq d(n) \). We reduce \( H'_0 \) to \( H''_0 : \vartheta = 0 \) in \( G_k \) in the presence of the nuisance parameter \( \lambda \). By standard calculations we get the score vector for \( H''_0 \) in \( G_k \) of the form

\[
\ell = (\ell^T_{\vartheta}, \ell^T_{\lambda})^T \quad \text{with} \quad \ell_{\vartheta}(y) = \psi(F(y, \lambda)), \ \ell_{\lambda}(y) = \frac{\partial \log f(y, \lambda)}{\partial \lambda},
\]
where \( \psi(t) = (\psi_1(t), \ldots, \psi_k(t))^T \) is a vector of \( k \) first functions of the orthonormal system. Consequently, the effective score vector for \( H_0'' \) can be written as

\[
(2.4) \quad \ell^*(y) = \psi(F(y, \lambda)) - I_{\phi\lambda} I_{\lambda\lambda}^{-1} \ell_{\lambda}(y),
\]

where

\[
I_{\lambda\lambda} = \int_0^\infty \ell_{\lambda}^2(y) f(y, \lambda) dy = \sum_{r=0}^\infty \left( \frac{r}{\lambda} - 1 \right)^2 \pi_r(\lambda) = \frac{1}{\lambda^2} \Var X = \frac{1}{\lambda}
\]

and

\[
I_{\phi\lambda} = \int_0^\infty \psi(F(y, \lambda)) \ell_{\lambda}(y) f(y, \lambda) dy = \sum_{r=0}^\infty \int_r^{r+1} \psi(F(y, \lambda)) \left( \frac{r}{\lambda} - 1 \right) f(y, \lambda) dy
\]

\[
= \frac{1}{\lambda} \sum_{r=1}^\infty r \int_r^{r+1} \psi(F(y, \lambda)) f(y, \lambda) dy = \frac{1}{\lambda} J.
\]

Here the vector \( J = J(\lambda) \) can be expressed as

\[
J = \sum_{r=1}^\infty \frac{F(r+1, \lambda)}{F(r, \lambda)} \psi(t) dt = \sum_{r=1}^\infty \left[ \Psi(F(r+1, \lambda)) - \Psi(F(r, \lambda)) \right]
\]

\[
= -\sum_{r=0}^\infty \Psi(\pi_0(\lambda) + \ldots + \pi_r(\lambda)),
\]

where \( \Psi(t) = (\Psi_1(t), \ldots, \Psi_k(t))^T \) is a vector of the functions

\[
\Psi_j(t) = \int_0^t \psi_j(u) du, \quad t \in [0, 1].
\]

The covariance matrix of the effective score vector has the usual form

\[
I^* = I - I_{\phi\lambda} I_{\lambda\lambda}^{-1} I_{\phi\lambda}^T = I - \frac{1}{\lambda} J J^T,
\]

where \( I \) denotes the identity matrix. Its inverse can be written as (cf. formula (3.2) in Inglot et al. [5])

\[
(I^*)^{-1} = I + I_{\phi\lambda}(I_{\lambda\lambda} - I_{\phi\lambda}^T I_{\phi\lambda})^{-1} I_{\phi\lambda}^T = I + \frac{1}{\lambda - J^T J} J J^T.
\]

In consequence, the effective score statistic for testing \( H_0'' \) in \( G_k \) takes the form

\[
N_k = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^*(Y_i) \right)^T (I^*)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^*(Y_i) \right).
\]
Since the natural estimator of the parameter $\lambda$ is the sample mean $\hat{\lambda} = \bar{X}$ which is the maximum likelihood estimator in the family $\mathcal{P}$, the estimated effective score vector has a simpler form, i.e. $\hat{\psi}(y) = \psi(F(y, \hat{\lambda}))$ (cf. (3.5) in [5]).

Finally, the test statistic for $H_0''$ in $\mathcal{G}_k$ takes the form

\begin{equation}
\hat{N}_k = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(F(Y_i, \hat{\lambda})) \right)^T \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(F(Y_i, \hat{\lambda})) \right)^{-1},
\end{equation}

where

\begin{equation}
(I^*)^{-1} = I + \frac{1}{\lambda - \hat{J}^T \hat{J}} \hat{J} \hat{J}^T
\end{equation}

and $\hat{J} = -\sum_{r=0}^{\infty} \Psi(\pi_0(\hat{\lambda}) + \ldots + \pi_r(\hat{\lambda}))$. Easy calculations show that regularity conditions (R1)–(R4) in [5] for $\mathcal{P}$ are satisfied. So, when $\psi_j$’s are twice differentiable and

\begin{equation}
\sup_{t \in [0,1]} |\psi_j(t)| \leq c_j m, \quad \sup_{t \in [0,1]} (|\psi_j'(t)| + |\psi_j''(t)|) \leq c_j m^2 + 2, \quad j \geq 1,
\end{equation}

for some positive $c$ and nonnegative $m$, it follows from Theorem 3.1 in [5] that under $H_0''$, $\hat{N}_k$ converges in distribution to the chi-square distribution with $k$ degrees of freedom.

It is well known that the choice of $k$ among $1, \ldots, d(n)$ is crucial to the performance of a test based on the score statistic $\hat{N}_k$. Therefore, we propose a data driven choice of $k$ using a Schwarz type selection rule (cf., e.g., Inglot et al. [5], Schwarz [10])

\begin{equation}
S = \min \{1 \leq k \leq d(n) : \hat{N}_k - k \log n = \max_{1 \leq j \leq d(n)} (\hat{N}_j - j \log n) \}.
\end{equation}

Taking into account promising results in Inglot and Janic [4], we consider also another, less conservative, selection rule denoted by $L$. Choose two natural numbers (not depending on $n$): a small one $1 \leq D < d(n)$ and a big one $K > D$, and set $K(n) = \min(K, d(n))$. Moreover, let $\delta_n$ be a small positive number. Define the thresholds $c_{jn}$, $j = 1, \ldots, D$, to be the solutions of the equations

\begin{equation}
1 - \Phi(c_{jn}) = \frac{1}{2} \left( \delta_n D^{-1} \left( \frac{K(n)}{j} \right)^{-1} \right)^{1/j},
\end{equation}

where $\Phi$ denotes the standard normal distribution function. Next, consider the standardized random vector $\mathcal{L} = ((I^*)^{-1})^{1/2} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(F(Y_i, \hat{\lambda})) \right)$ with $K(n)$ components, where $(I^*)^{-1}$ is given by (2.6) with $k = K(n)$ while

\begin{equation}
((I^*)^{-1})^{1/2} = I + \frac{1}{\hat{\lambda} - \hat{J}^T \hat{J} + \sqrt{\hat{\lambda} \hat{\lambda} - \hat{J}^T \hat{J}}} \hat{J} \hat{J}^T.
\end{equation}
Order the squares of its components from the smallest to the largest, obtaining $L^2_1, \ldots, L^2_{K(n)}$, and consider the event

$$E_n = \{L^2_{K(n)} \geq c^2_n\} \cup \ldots \cup \{L^2_{K(n)-D+1} \geq c^2_{Dn}\}.$$ 

Then define the data dependent penalty

$$\rho(j, n) = j(\log n \cdot 1_{E_n} + 2 \cdot 1_{E_n^c}),$$

where $1_{E_n}$ denotes the indicator of the event $E_n$ and $E_n^c$ is the complement of $E_n$, and define the corresponding selection rule $L$ as

$$(2.9) \quad L = \min \{1 \leq k \leq d(n) : \hat{N}_k - \rho(k, n) = \max_{1 \leq j \leq d(n)} (\hat{N}_j - \rho(j, n))\}.$$ 

By the definition, for $n \geq e^2$ we have $\rho(j, n) \leq j \log n$ a.s. Consequently, $L \geq S$ a.s. and $\hat{N}_L \geq \hat{N}_S$ a.s.

Note that parameters $D, K, \delta_n$ were used only to define penalty $\rho(j, n)$ for $L$. In particular, introducing an upper bound $K$ means that a choice of penalty for $L$ is based on a limited number of empirical Fourier coefficients with respect to the system $(\psi_j)$.

By using all the above considerations, $\hat{N}_S$ and $\hat{N}_L$ are applied as test statistics of upper-tailed tests for testing $H_0$ (or equivalently $H_0$).

The asymptotic behavior of $\hat{N}_S$ and $\hat{N}_L$ is established in the following theorem.

**Theorem 2.1.** Suppose $\psi_1(t), \psi_2(t), \ldots$ is an orthonormal system satisfying (2.7) and the maximal dimension $d(n)$ of $G_k$ in (2.3) satisfies the condition $d(n) = o\left((n/ \log n)^{1/(2m+4)}\right)$. Then

$$(2.10) \quad \hat{N}_S \overset{D}{\rightarrow} \chi^2_k \text{ under } H_0,$$

where $\chi^2_k$ denotes a random variable with chi-square distribution with $k$ degrees of freedom.

If, in addition, $\delta_n \rightarrow 0$, then $L - S \rightarrow 0$ in probability with respect to any null distribution, and consequently

$$(2.11) \quad \hat{N}_L \overset{D}{\rightarrow} \chi^2_1 \text{ under } H_0.$$ 

Since the assumption on $d(n)$ in Theorem 2.1 implies (D1)–(D3) in Inglot et al. [5], the assertion (2.10) follows from Theorem 4.1 in [5]. The assertion $L - S \rightarrow 0$ in probability under $H_0$ is an easy and straightforward consequence of the central limit theorem for the random vector $L$ due to the boundedness of $K(n)$. We omit details.

From Theorems 2.6, 4.2 and 4.3 in [5] we immediately obtain a consistency result for the tests based on $\hat{N}_S$ and $\hat{N}_L$.
Theorem 2.2. Let $d(n) \to \infty$ and the conditions of Theorem 2.1 be satisfied. Then for any alternative discrete distribution $P$ concentrated on nonnegative integers with probability mass function $p_r$, $r \geq 0$, and the expected value $\lambda > 0$ such that for some $j \geq 1$

$$
(2.12) \quad \sum_{r=0}^{\infty} p_r \int \frac{r+1}{r} \psi_j(F(y, \lambda)) dy \neq 0
$$

we have $\hat{N}_S \xrightarrow{P} \infty$ and $\hat{N}_L \xrightarrow{P} \infty$. Consequently, the tests based on $\hat{N}_S$ and $\hat{N}_L$ are consistent against any $P$ satisfying (2.12).

Remark 2.1. When $\psi_1, \psi_2, \ldots$ form a complete orthonormal system of bounded functions, then the assumption (2.12) is a weak one and is satisfied for a large class of alternatives. For example, if $p_r \leq C \pi_r(\lambda)$ for all $r \geq 0$ and some positive constant $C$, where $\lambda = \sum_{r=1}^{\infty} r p_r$, then (2.12) holds. In particular, (2.12) is satisfied for any distribution with finite support or with $p_r = \pi_r(\lambda)$ except finitely many $r$.

Now, we discuss two key choices needed for an implementation tests based on $\hat{N}_S$ and $\hat{N}_L$ in Section 3.

Firstly, we discuss a choice of an orthonormal system $\{\psi_j\}$. The most popular is the Legendre system on $[0, 1]$, we shall denote by $\{b_j\}$. It satisfies (2.7) with $m = 1/2$. For our particular family $P$ of discontinuous densities this is rather not an optimal choice. In spite of this we do apply it in our implementation. However, for $P_\lambda$ with small $\lambda$, high variation of $b_j$’s near zero has nothing to do with large values of the first few probabilities of $P_\lambda$ and results in a less sensitive test. To overcome simply this problem we define another orthonormal system $\{h_j\}$ with $h_j(t) = b_{2j}((1 + t)/2)$. The functions $h_j$ are smooth on the left end of $[0, 1]$ with high variation only on the right end of the unit interval. So, we shall use the system $\{h_j\}$ for small $\lambda$, say for $\lambda \leq \lambda_0$, and the system $\{b_j\}$ otherwise. Since $b_1$ and $h_1$ are strongly correlated with $\ell_{h}(F^{-1}(t, \lambda), \hat{\lambda})$, we remove them from the system. Data driven tests usually attain the highest power for alternatives for which the second empirical Fourier coefficient (under an actually applied orthonormal system) is the largest one. For most typical alternatives to the Poisson family the largest Fourier coefficient corresponds to $b_2$ or $h_2$. So, we order the both systems as follows: $b_3, b_2, b_4, b_5, \ldots$ and $h_3, h_2, h_4, h_5, \ldots$ Since $\lambda$ is unknown, we shall use the estimator $\hat{\lambda}$ to decide which orthonormal system will be applied in the test statistic.

In effect, we define the orthonormal system $\{\psi_j\}$ as follows:

$$
\psi_1(t) = h_3(t) \text{ if } \hat{\lambda} \leq \lambda_0 \quad \text{or} \quad \psi_1(t) = b_3(t) \text{ otherwise,}
$$

$$
\psi_2(t) = h_2(t) \text{ if } \hat{\lambda} \leq \lambda_0 \quad \text{or} \quad \psi_2(t) = b_2(t) \text{ otherwise,}
$$

$$
\psi_j(t) = h_{j+1}(t) \text{ if } \hat{\lambda} \leq \lambda_0 \quad \text{or} \quad \psi_j(t) = b_{j+1}(t) \text{ otherwise, } j = 3, 4, \ldots
$$
The test statistics $\tilde{N}_S$ and $\tilde{N}_L$ for such a switched over orthonormal system will be denoted by $M_S$ and $M_L$, respectively. The tests based on these statistics are examples of data driven score tests with an orthonormal system depending on the data. Since the system $(h_j)$ satisfies (2.7), Theorem 2.1 applies to $(h_j)$ and to $M_S$ and $M_L$ as well. Also, after some small obvious reformulation, the statement of Theorem 2.2 remains valid for both $M_S$ and $M_L$.

Secondly, let us discuss a choice of the maximal dimension $d(n)$. The assumption of Theorem 2.1 as well as properties of our particular family $P$ of the Poisson distributions suggest to take a slowly increasing sequence $d(n) = \lfloor cn^{r} \rfloor$ with $r < 1/5$. When $M_L$ is applied, the relation $c \geq D$ seems to be reasonable. Moreover, for moderate sample sizes $d(n) < K$ (if $K$ is not too small). Therefore, a choice of $K$ has practically no influence on the selection rule $L$ for moderate sample sizes. But, it allows for simplifying assumptions in Theorems 2.1 and 2.2.

The above specifications can be thought only as reasonable recommendations. For example, at a cost of some loss in power for alternatives with small expectations $\lambda$, one can consider simply tests based on $b_N$ or $b_N^L$ for the Legendre system $(b_j)$ without switching the orthonormal system.

We end this section by a brief comment how a randomization of a given sample $X = (X_1, \ldots, X_n)$ by noise values $U_1, \ldots, U_n$ may influence the acceptance or rejection decision of the null hypothesis. One may expect that when the sample $X$, at hand, is “typical”, then the randomization will change the decision rarely. We have roughly verified this statement by some empirical study taking $n = 50$ and $\alpha = 0.05$. We have observed that the distribution of conditional power $\beta(n|\bar{X})$ for our tests is strongly polarized, i.e. strongly concentrated near 0 and 1. Under null distributions, $\beta(n|\bar{X})$ belong to the interval $[0, 0.1]$ for ca. 85–95% of samples $\bar{X}$ while $\beta(n|\bar{X})$ do not exceed the nominal significance level 0.05 for ca. 75–85% of samples $\bar{X}$. Under alternatives, $\beta(n|\bar{X})$ lie outside the interval $(0.1, 0.9)$ for ca. 70% of samples $\bar{X}$ and divide between $[0, 0.1]$ and $[0.9, 1]$ approximately proportionally to the (unconditional) power $\beta(n)$ attained by the test.

3. SIMULATION STUDY

The aim of this section is to study how our new tests based on $M_S$ and $M_L$, described in Section 2, perform empirically in comparison with some known tests for Poissonity which proved to be powerful, particularly with the test $\bar{T} = \bar{T}_n$ of Klar [6].

We restrict our attention to a typical sample size $n = 50$ and standard significance level $\alpha = 0.05$. We take the orthonormal system $(\psi_j)$ defined in (2.13) and the maximal dimension of the exponential model $d(n) = 5$ for $n = 50$ which roughly corresponds to the formula $d(n) = \lfloor 3n^{1/7} \rfloor$. For the selection rule $L$ we took $D = 3$, $K = 20$ and $\delta_n = 0.05$. An analysis of values $\pi_r(\lambda)$ for small $r$ and different $\lambda$ suggests to choose $\lambda_0$ nearby 2. We took $\lambda_0 = 1.8$. 

Probability and Mathematical Statistics 39, z. 1, 2019
© for this edition by CNS
The behavior of all compared tests for other sample sizes is similar to that for \( n = 50 \), so we do not report it here.

### 3.1. Critical values of \( M_S \) and \( M_L \)

Nowadays it is a strong evidence that, for data driven tests, the convergence of \( M_S \) and \( M_L \) to the limiting null chi-square distribution with one degree of freedom is slow. So, we have determined the critical values empirically. The results are shown in Table 1. They practically do not depend on the nuisance parameter \( \lambda \). This is not surprising since our tests are asymptotically distribution free. The results from Table 1 fully justify that average simulated critical values 5.910 for \( M_S \) and 7.092 for \( M_L \) may be used as fixed critical values for the sample size \( n = 50 \).

#### Table 1. Empirical critical values of \( M_S \) and \( M_L \) for several values of \( \lambda, n = 50, \alpha = 0.05 \), 30,000 MC.

| \( \lambda \) | 0.2 | 0.5 | 1   | 2   | 5   | 10  | 30  | Average |
|--------------|-----|-----|-----|-----|-----|-----|-----|---------|
| \( M_S \)    | 5.874 | 5.877 | 5.892 | 5.964 | 5.818 | 5.907 | 6.036 | 5.910   |
| \( M_L \)    | 7.095 | 6.994 | 6.997 | 7.258 | 6.878 | 7.259 | 7.165 | 7.092   |

Under the same \( d(n), D, K, \delta_n, \lambda_0 \) and \( \alpha \), average critical values for the sample size \( n = 25 \) equal 6.842 for \( M_S \) and 7.592 for \( M_L \); and similarly 5.416 and 6.789 for \( n = 100 \); 4.967 and 6.550 for \( n = 200 \); 4.288 and 6.342 for \( n = 500 \); respectively.

### 3.2. Power comparisons

For easier comparisons with the results available in the literature, we start our study with the list of 20 alternatives considered in the recent paper by Ledwina and Wyłupek [7]. In most cases we apply the notation from that paper or from Gürtler and Henze [3]. For completeness, we present description of all alternatives in the Appendix. We choose three tests for comparison which proved to be powerful: the test \( T \) (proposed by Klar [4] and in his notation), the test \( I \) (proposed by Klar [3] in the notation of Gürtler and Henze [3]) and the test \( V^* \) of Nakamura and Pérez-Abreu [9]. Detailed description of these tests is also provided in the Appendix. For \( M_S \) and \( M_L \) we take critical values determined in Section 3.1. The results are shown in Table 2.

In the upper part of Table 2 we collect ‘smooth’ alternatives, i.e. those for which only one or two first Fourier coefficients (with respect to \( (\psi_j) \) given in (2.13)) are significant, and in the lower part we put alternatives with wider spectrum. To some extent, ‘smoothness’ is related to a number and positions of changes of sign for differences \( p_r - \pi_r(\lambda) \), where \( \lambda = \sum_{r \geq 1} r p_r \) is the expectation of an alternative. Since \( M_L \) has been designed to be more sensitive for ‘less smooth’ alternatives, it can be seen that in the upper part of Table 2 \( M_S \) is better than \( M_L \) while in the lower part an opposite relation occurs. Powers of \( T, I \) and \( V^* \) have been taken from Ledwina and Wyłupek [2].

The test \( I \) is unstable. For some cases it attains extremely high power (cf. \( b(10, 0.5) \)) but for some others, poor power (e.g. \( U(5; 15) \) or \( TG(0.45) \)). The test \( T \)
performs more stable than $V^*$ and outperforms it in average. Data driven tests $M_S$ and particularly $M_L$ are very stable and both perform equally well. For alternatives presented in Table 2, $T$ outperforms $M_S$ and $M_L$ ca. 7% in average.

| Alternative                  | $M_S$ | $M_L$ | $T$  | $I$  | $V^*$ | $\lambda$ |
|------------------------------|-------|-------|------|------|-------|----------|
| $U(0; 1)$                    | 79    | 76    | 98   | 99   | 94    | 0.5      |
| $U(0; 2)$                    | 57    | 57    | 64   | 68   | 73    | 1        |
| $b(10, 0.5)$                 | 75    | 68    | 81   | 88   | 60    | 5        |
| $b(20, 0.35)$                | 33    | 27    | 37   | 46   | 19    | 7        |
| $nb(2, 2/3)$                 | 28    | 28    | 42   | 45   | 48    | 1        |
| $P\delta(9, 3)$             | 38    | 38    | 33   | 45   | 48    | 2.7      |
| $P\delta(0.7, 1.5)$         | 45    | 42    | 59   | 59   | 52    | 1.05     |
| $GP(3, -0.24)$               | 30    | 24    | 38   | 46   | 19    | 2.42     |
| $GP\delta(4.59, -0.33, 0.025)$ | 41   | 35    | 49   | 53   | 33    | 3.36     |
| $GH(0.5, 0.25)$             | 42    | 41    | 57   | 58   | 53    | 1        |
| $PS\delta(1, 0.75)$         | 56    | 55    | 77   | 78   | 74    | 0.75     |
| $TG(0.45)$                  | 83    | 85    | 86   | 47   | 73    | 2.22     |
| **Average**                 | 50.6  | 48.0  | 60.1 | 61.0 | 53.8  |          |
| $U(0; 4)$                    | 44    | 51    | 60   | 16   | 73    | 2        |
| $U(5; 15)$                   | 35    | 47    | 39   | 7    | 53    | 10       |
| $PP(0.1, 1.1, 6.9)$          | 55    | 60    | 56   | 54   | 55    | 6.3      |
| $PP(0.1, 1.1, 6.1)$          | 46    | 50    | 45   | 44   | 43    | 5.6      |
| $GP\delta(4.59, -0.33, 0.127)$ | 63   | 64    | 55   | 21   | 56    | 3        |
| $PBM(0.55, 10, 0.97)$        | 95    | 96    | 97   | 94   | 93    | 9.7      |
| $BB(5, 1.6, 0.67)$           | 94    | 95    | 99   | 96   | 99    | 3.5      |
| $BB(5, 1, 0.67)$             | 83    | 88    | 93   | 48   | 96    | 3        |
| **Average**                 | 64.4  | 68.9  | 68.0 | 47.5 | 71.0  |          |
| **Total average**           | 56.1  | 56.4  | 63.3 | 55.6 | 60.7  |          |

The alternatives considered in Table 2 are well-known families of discrete distributions and not necessarily represent typical departures from the Poisson family. More realistically one may expect small changes of several probabilities $\pi_j$ of $P_\lambda$. To see how our new tests are able to detect such contaminated Poisson distributions, we introduce three additional families of alternatives.

The first alternative is a modification of $P_{0.5}$ and preserves its expectation 0.5. We replace four first probabilities $\pi_0(0.5), \pi_1(0.5), \pi_2(0.5)$ and $\pi_3(0.5)$ by $\pi_0(0.5) - u + a + 2b, \pi_1(0.5) + v - 2a - 3b, a$ and $b$, respectively, with $u = \pi_2(0.5) + 2\pi_3(0.5)$ and $v = 2u - \pi_3(0.5)$, and keep the remaining probabilities unchanged. Parameters $a, b$ are nonnegative with $2a + 3b$ not exceeding $\pi_1(0.5) + v$. We shall denote this alternative by $A_{1.5}^{3.5}(a, b)$. 
The second alternative modifies $P_\lambda$ with any $\lambda$ and preserves its expectation $\lambda$. We replace four probabilities $\pi_j(\lambda), \pi_{j+1}(\lambda), \pi_{j+2}(\lambda)$ and $\pi_{j+3}(\lambda)$ by $\pi_j(\lambda) + c, \pi_{j+1}(\lambda) - c, \pi_{j+2}(\lambda) - c, \pi_{j+3}(\lambda) + c$, respectively, and keep the remaining probabilities unchanged. The parameter $j$ is a nonnegative integer while $c$ can take positive or negative values in such a way that all four obtained numbers are nonnegative. We shall denote this alternative by $A_4(\lambda, j, c)$.

The third alternative modifies $P_\lambda$ with any $\lambda$ and also preserves its expectation $\lambda$. We replace eight probabilities $\pi_j(\lambda), \ldots, \pi_{j+7}(\lambda)$ by $\pi_j(\lambda) + a + b + c, \pi_{j+1}(\lambda) - a - b - c, \pi_{j+2}(\lambda) - a, \pi_{j+3}(\lambda) + a, \pi_{j+4}(\lambda) - b, \pi_{j+5}(\lambda) + b, \pi_{j+6}(\lambda) - c, \pi_{j+7}(\lambda) + c$, respectively, and keep the remaining probabilities unchanged. The parameter $j$ is a nonnegative integer while $a, b, c$ are such that all resulting numbers are nonnegative. We shall denote this alternative by $A_8(\lambda, j, a, b, c)$.

In Table 3 we show, by typical examples of new alternatives, empirical powers of $M_S$ and $M_L$ compared with $\bar{T}$, the leader in Table 2. It can be observed that for more difficult alternatives the new test $M_L$ performs essentially better than $\bar{T}$.

| Alternative | \( M_S \) | \( M_L \) | \( \bar{T} \) |
|-------------|----------|----------|----------|
| \( A_4(0.05, 0.05) \) | 22 | 24 | 33 |
| \( A_4(0.07) \) | 33 | 45 | 39 |
| \( A_4(5, 1, 0.08) \) | 49 | 62 | 28 |
| \( A_4(7, 7, 0.10) \) | 27 | 45 | 27 |
| \( A_8(3, 0, 0.04, -0.09, 0.05) \) | 33 | 46 | 38 |
| \( A_8(5, 3, 0.10, -0.05, 0.036) \) | 54 | 71 | 49 |
| \( A_8(9, 9, 0.097, -0.02, 0.019) \) | 18 | 30 | 23 |
| Average | 33.7 | 47.1 | 33.9 |

Table 4. Powers of $M_S, M_L$ and $\bar{T}$ for three selected alternatives and different sample sizes. $\alpha = 0.05$, 10,000 MC for $M_S$ and $M_L$, and 2000 MC for $\bar{T}$.

| Alternative | \( M_S \) | \( M_L \) | \( \bar{T} \) |
|-------------|----------|----------|----------|
| \( P(0.7, 1.5) \) | 23 | 45 | 76 | 98 |
| The first group in Table 2 | 22 | 42 | 74 | 97 |
| \( \bar{T} \) | 34 | 59 | 90 | 100 |
| \( P(0.1, 1.1, 0.1) \) | 31 | 46 | 69 | 92 |
| The second group in Table 2 | 32 | 50 | 76 | 96 |
| \( \bar{T} \) | 30 | 46 | 72 | 95 |
| \( A_4(5, 1, 0.08) \) | 29 | 49 | 81 | 99 |
| A new alternative from Table 3 | 33 | 62 | 94 | 100 |
| \( \bar{T} \) | 17 | 28 | 54 | 89 |
To illustrate power curves of compared tests under increasing sample size we have selected three alternatives, one from each group considered above. For all sample sizes we took the same parameters \( d(n), D, K, \delta_n, \alpha, \lambda_0 \) as for \( n = 50 \). The results are shown in Table 4.

Concluding, one can say that the new tests perform comparable sensitivity to \( T \) as well as to \( I \) and \( V^* \) while \( M_L \) preserves stable and high sensitivity for a much wider class of various types of alternatives. So, if ‘less smooth’ departure from Poissonity is expected, then the test \( M_L \) may be recommended.

4. APPENDIX

Description of alternatives. Probability mass functions of distributions different from \( P_\lambda \) will be denoted by \( p_j \) for integer \( j \geq 0 \).

- \( U(m;l), m < l \) – the uniform distribution on the set \( \{ m, m+1, \ldots, l \} \).
- \( b(m,p) \) – the binomial distribution with parameters \( m \in \mathbb{N} \) and \( p \in (0,1) \).
- \( nb(m,p) \) – the negative binomial distribution with parameters \( m \in \mathbb{N} \) and \( p \in (0,1) \).
- \( P\delta(\varepsilon,\lambda) \) – the mixture \( \varepsilon P_\lambda + (1 - \varepsilon)\delta_0 \) of the Poisson distribution and the Dirac delta in zero.
- \( PP(\varepsilon,\lambda_1,\lambda_2) \) – the mixture \( \varepsilon P_{\lambda_1} + (1 - \varepsilon)P_{\lambda_2} \) of two Poisson distributions.
- \( GP(\lambda,\vartheta) \) – the generalized Poisson distribution with parameters \( \lambda > 0, -\lambda < \vartheta < 0, \vartheta > -1 \) and probability mass function given by
  \[
  p_j = \frac{\lambda (\lambda + \vartheta j)^{j-1} e^{-\lambda - \vartheta j}}{j!}, \quad j = 0, 1, \ldots, \lfloor -\lambda / \vartheta \rfloor.
  \]
- \( GP\delta(\lambda,\vartheta,\varepsilon) \) – the mixture \( (1 - \varepsilon)GP(\lambda,\vartheta) + \varepsilon \delta_0 \) of the generalized Poisson distribution and the Dirac delta in zero.
- \( PBM(\varepsilon,m,p) \) – the mixture \( \varepsilon P_{mp} + (1 - \varepsilon)b(m,p) \) of the Poisson distribution and the binomial distribution with the same mean.
- \( GH(\lambda_1,\lambda_2) \) – the generalized Hermite distribution, i.e. the distribution of \( Y_1 + 2Y_2 \), where \( Y_1, Y_2 \) are independent random variables with Poisson distributions \( P_{\lambda_1}, P_{\lambda_2} \).
- \( TG(p) \) – the geometric distribution with parameter \( p \), i.e. with probability mass function \( p_j = p(1-p)^{j-1}, \quad j = 1, 2, \ldots \)
- \( BB(m,p,q) \) – the beta-binomial distribution with parameters \( m \in \mathbb{N}, \ p, q > 0, \) i.e. with probability mass function given by
  \[
  p_j = \frac{1}{\binom{m}{j}} x^j (1-x)^{m-j} B_{pq}(x) dx, \quad j = 0, 1, \ldots, m,
  \]
  where \( B_{pq}(x) \) is the density of the beta distribution with parameters \( p, q \).
- \( PSS(\lambda_1,\lambda_2) \) – the Poisson-stopped-sum distribution, i.e. the distribution of the sum \( \sum_{i=1}^{N} Y_i \), where \( N \) has \( P_{\lambda_1} \) distribution while \( Y_i \) are i.i.d. with Poisson \( P_{\lambda_2} \) distribution and \( N \) is independent of \( Y_i \)’s.
Tests for comparison. All tests described below reject the null hypothesis for large values of the corresponding statistics.

- \( T \) defined by the statistic
  \[
  T = \sqrt{n} \left[ \sum_{j=0}^{m} \left( |F_{n}(j) - F(j, \bar{X})| + F(j, \bar{X}) \right) \right] + \sqrt{n} (\bar{X} - m - 1),
  \]
  where \( m = \max_{1 \leq i \leq n} X_i \), \( F_n \) is the empirical distribution function and \( F(k, \lambda) \) is the distribution function of \( P_{\lambda} \) (the notation as in Klar [6]).

- \( I \) defined by the statistic
  \[
  I = \sqrt{n} \max_{1 \leq k \leq m} \sum_{j=0}^{k} \left| F_{n}(j) - F(j, \bar{X}) \right|,
  \]
  where \( m, F_n \) and \( F(k, \lambda) \) are as above (cf. Klar [6], the notation after Gürtler and Henze [3]).

- \( V^* \) defined by the statistic
  \[
  V^* = n^{-\frac{3}{2}} \sum_{j=0}^{2m-2} a_j^2,
  \]
  where
  \[
  a_j = \sum_{l=0}^{j+2} l(2l - j - 3)N_lN_{j+2-l}
  \]
  while \( N_j \) is the number of observations equal to \( j \) (cf. Nakamura and Pérez-Abreu [9]).

R code for tests proposed in the present paper is available at:
http://prac.im.pwr.wroc.pl/inglot/badania.html

REFERENCES

[1] D. J. Best and J. C. W. Rayner, Goodness of fit for the Poisson distribution, Statist. Probab. Lett. 44 (1999), pp. 259–265.
[2] J. Frey, An exact Kolmogorov–Smirnov test for the Poisson distribution with unknown mean, J. Stat. Comput. Simul. 82 (2012), pp. 1023–1033.
[3] N. Gürtler and N. Henze, Recent and classical goodness-of-fit tests for the Poisson distribution, J. Statist. Plann. Inference 90 (2000), pp. 207–225.
[4] T. Inglot and A. Janic, How powerful are data driven tests for uniformity, Appl. Math. 36 (2009), pp. 375–395.
[5] T. Inglot, W. C. M. Kallenberg, and T. Ledwina, Data driven smooth test for composite hypotheses, Ann. Statist. 25 (1997), pp. 1222–1250.
[6] B. Klar, Goodness-of-fit tests for discrete models based on the integrated distribution function, Metrika 49 (1999), pp. 53–69.
[7] T. Ledwina and G. Wypułpek, On Charlier polynomials in testing Poissonity, Commun. Statist. Simul. Comput. 46 (2017), pp. 1918–1932.
[8] S. G. Meintanis and Ya. Yu. Nikitin, A class of count models and a new consistent test for the Poisson distribution, J. Statist. Plann. Inference 138 (2008), pp. 3722–3732.
[9] M. Nakamura and V. Pérez-Abreu, Use of an empirical probability generating function for testing a Poisson model, Canad. J. Statist. 21 (1993), pp. 149–156.
[10] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978), pp. 461–464.
[11] O. Thas and J. C. W. Rayner, Smooth tests for the zero-inflated Poisson distribution, Biometrics 61 (2005), pp. 808–815.

Tadeusz Inglot
Faculty of Pure and Applied Mathematics
Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
E-mail: Tadeusz.Inglot@pwr.edu.pl

Received on 18.10.2016;
revised version on 3.10.2017