South African Statistical Journal
Vol. 54, No. 2, 207–230
https://doi.org/10.37920/sasj.2020.54.2.6

AN OVERVIEW OF GOODNESS-OF-FIT TESTS FOR THE POISSON DISTRIBUTION

P. A. Mijburgh
Department of Statistics, University of Pretoria, South Africa
Department of Mathematics and Statistics, McMaster University, Canada

I. J. H. Visagie
School of Mathematical and Statistical Sciences, North-West University, South Africa
e-mail: *jaco.visagie@nwu.ac.za*

The Poisson distribution has a large number of applications and is often used as a model in both a practical and a theoretical setting. As a result, various goodness-of-fit tests have been developed for this distribution. In this paper, we compare the finite sample power performance of ten of these tests against a wide range of alternative distributions for various sample sizes. The alternatives considered include, seemingly for the first time, weighted Poisson distributions. A number of additional tests are of historical importance although their power performance is not competitive against the remaining tests. These tests are discussed, but their powers are not included in the numerical analysis.

The Monte Carlo study presented below indicates that the test with the best overall power performance is the test of Meintanis and Nikitin (2008), followed closely by the test of Rayner and Best (1990) (originally studied in Fisher, 1950).

Key words: Goodness-of-fit testing, Poisson distribution, Warp-speed bootstrap.

1. Introduction

Since its initial publication nearly 200 years ago, the Poisson distribution has been widely used to model count data in a range of different disciplines; see Poisson (1828). It is often of practical interest to test whether or not a realised dataset is compatible with the assumption of being realised from the Poisson distribution. This paper provides a review of the various tests for the Poisson distribution. Special attention is paid to the finite sample power performance of several of the tests.

The most recent reviews of tests for the Poisson distribution can be found in Gürtler and Henze (2000) and Karlis and Xekalaki (2000), both of which were published two decades ago. While Gürtler and Henze (2000) provide an in depth review of the theoretical aspects of the tests for the Poisson distribution, Karlis and Xekalaki (2000) include an extensive comparison between the power performance of the various tests considered. A number of new tests have been introduced into the statistical literature since the publication of these reviews. The current paper compares the power performance of these tests to that of the more powerful classical tests. Additionally, there exists a number of tests that are of historical importance which are not competitive in terms of power

[1] Corresponding author.
MSC2010 subject classifications. 62G10, 62E10.

performance. These tests are discussed in this paper, but omitted from the Monte Carlo study presented in Section 3.

The popularity of the Poisson distribution has motivated a large body of research relating to generalisations and modifications of this distribution; see Haight (1967), Patil and Joshi (1968) and Johnson et al. (2005) for extensive lists and discussions of these generalisations. The mentioned generalisations provide practitioners with a number of more flexible alternatives to use in cases where the Poisson assumption is rejected. A popular generalisation is the class of so-called weighted Poisson distributions (described below in Section 3.1). We are unaware of any published results relating to the power of goodness-of-fit tests against these alternatives. In the Monte Carlo study included in this paper, the powers of the various tests are compared for a variety of alternatives, including weighted Poisson alternatives.

The remainder of this paper is structured as follows. Section 2 describes the general hypothesis testing framework used throughout. Additionally, this section provides an overview of the various tests for the Poisson distribution that are available. Section 3 contains the results of a Monte Carlo power study used to compare the powers of the various tests considered. Two numerical examples are included in Section 4, while Section 5 presents the conclusions. Additional numerical results are available in the appendix.

2. Tests for the Poisson distribution

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) realisations of a random variable X taking on values in the non-negative integers. Let F denote the distribution function (df) of X. Let F_λ be the Poisson distribution function with mean $\lambda > 0$;

$$F_\lambda(x) = \mathrm{e}^{-\lambda} \sum_{j=0}^{x} \frac{\lambda^j}{j!}, \quad \text{for} \quad x \in \{0, 1, \ldots\},$$

and let f_λ be the corresponding probability mass function (pmf);

$$f_\lambda(x) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}, \quad \text{for} \quad x \in \{0, 1, \ldots\}.$$

The composite goodness-of-fit hypothesis to be tested is as follows:

$$H_0 : F(x) = F_\lambda(x), \quad \text{for} \quad x \in \{0, 1, \ldots\} \text{ and for some } \lambda > 0, \tag{1}$$

against general alternatives.

Below we consider various goodness-of-fit tests available in the literature for testing the hypothesis in (1). The tests considered are based on numerous characterisations of the Poisson distribution, including the pmf, the df, the integrated df and the probability generating function. For an overview of the role of characterisations of distributions in the construction of goodness-if-fit tests, see, for example, Meintanis (2016) and Nikitin (2019).

Several of the tests considered contain a tuning parameter. Typically, when proposing a new test, authors recommend a value for the tuning parameter in question. Another possibility is to choose the value of this parameter data-dependently; see, for example, Allison and Santana (2015). In this paper, we opt to use the values recommended in the literature for tests containing a tuning parameter.

2.1 Tests based on moments

A distribution is said to be equidispersed if the expected value and the variance of the distribution are equal. It is well-known that the Poisson class of distributions possesses this property. A number of tests have been based on the property of equidispersion. Note that this property does not characterise the Poisson distribution; meaning that tests based on this property are not consistent. In spite of this shortcoming, these tests remain popular.

Below, we provide an overview of tests based on equidispersion as well as other moment characteristics.

2.1.1 Tests based on the Fisher index (FI)

One of the first articles concerned with the discrepancy between the mean and variance of a random sample realised from a Poisson distribution is Fisher (1950). As a result, the ratio of the sample variance S^2 and the sample mean $\overline{X}$ is known as Fisher's index of dispersion or the Fisher index (FI);

$$FI = \frac{S^2}{\overline{X}}, \quad \text{with } \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, \quad S^2 = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2.$$

Note that some authors also define the FI to be

$$(n-1)\frac{S^2}{\overline{X}}.$$

For an in depth discussion of the sampling distribution of the FI, see Anderson and Siddiqui (1994).

The Fisher index is typically used as a two sided test statistic for the Poisson distribution. Due to its widespread use, the properties of the FI have been examined in a number of papers, including Selby (1965), Anderson and Siddiqui (1994), Bartko et al. (1968), Dahiya and Gurland (1969), Potthoff and Whittinghill (1966), Collings and Margolin (1985), Soo Kim and Park (1992), Perry and Mead (1979) and Kharshikar (1970). Furthermore, the power of this test against a range of alternative distributions is presented in Bateman (1950) and Darwin (1957).

A count distribution is said to be overdispersed (underdispersed) if its variance exceeds (is less than) its mean. Historically, FI has often been used as a diagnostic tool to determine whether or not a distribution is over or underdispersed. Large (Small) values of the sample FI have often been presented as evidence for overdispersion (underdispersion) in the distribution from which the data are realised. See Henze and Klar (1996) for an explanation of the erroneous conclusions that this approach may elicit. In order to discourage the use of the FI in this manner, the authors argue for the use of the following rescaled versions of the FI:

$$S^* = \frac{n^2\overline{X}\,(FI-1)^2}{\sum_{j=1}^{n}\left\{\left(X_j - \overline{X}\right)^2 - X_j\right\}},$$

and

$$U^2 = \frac{n}{2}\,(FI-1)^2, \tag{2}$$

respectively, with U^2 initially proposed in Rayner and McIntyre (1985). In both cases, the null hypothesis of the Poisson distribution is rejected for large values of the test statistic. Note that U^2 is closely related to the Neyman smooth test for Poissonity; see Rayner and Best (1990).

Other test statistics based on the Fisher index have also been proposed; see Böhning et al. (1994), Potthoff and Whittinghill (1966) as well as Zelterman and Chen (1988). Additionally, a test based on the difference between the sample mean and variance is proposed in De Oliveira (1963). A correction to the limit null distribution derived in De Oliveira (1963) can be found in Böhning et al. (1994).

2.1.2 Tests based on other moments

Pettigrew and Mohler (1967) propose a test based on the moments of the Poisson distribution. Unlike the tests mentioned above, this test can be implemented using the higher order moments of the Poisson distribution. However, Pettigrew and Mohler (1967) recommend the use of lower order moments as this is found to reduce the variability of the statistic. The proposed statistics are

$$Z_p = \frac{k_p - \overline{X}}{\sqrt{var(k_p|\overline{X})}}, \quad p = 2, 3, 4, \tag{3}$$

where k_p is the p^{th} sample cumulant and $var\left(k_p|\overline{X}\right)$ is the variance of the p^{th} cumulant given the sample mean. For $p = 2$, this statistic reduces to

$$Z_2 = \frac{S^2 - \overline{X}}{\sqrt{2n\overline{X}(n\overline{X}-1)}} n\sqrt{n-1} \ . \tag{4}$$

Another test which depends on higher order moments was proposed in Gupta et al. (1994):

$$T = \frac{1}{2}\sqrt{\frac{n}{1 + 24\overline{X} + 6\overline{X}^2}} \frac{m_2\left(m_4 - 3m_2^2\right) - m_3}{\overline{X}^2}, \tag{5}$$

where m_j denotes the j^{th} sample moment. The test statistics in (3), (4) and (5) are asymptotically standard normal and are rejected for both large and small values of the test statistics.

Finally, we mention the test proposed in Kyriakoussis et al. (1998). This test is based on the second product moment of a distribution.

2.2 Chi square tests

Below, we provide a discussion of the well-known chi-square test and its variations as these pertain to the Poisson distribution. Although this test is of historical importance, it does not provide particularly high powers against the majority of alternatives to the Poisson distribution. As a result, the powers associated with this test are not included in the power study presented below.

The chi-square test easily lends itself to testing for the discrete distributions. This test requires that the data be split into k "bins" or groups before the test statistic can be computed. The test statistic is given by

$$\chi^2 = \sum_{j=0}^{k} \frac{\left(O_j - E_j\right)^2}{E_j},$$

where E_j denotes the expected number of observations in the j^{th} category under the null hypothesis and O_j denotes the observed number of observations in this category. A natural first choice for the bins is to consider each non-negative integer as a bin. However, for the test to be asymptotically valid, the expected number of observations in each bin must not be too small. This requires that several integers, especially in the tails of the distribution, be grouped together when these bins are constructed. It is well known that the power of the test is a function of the choice of the bins used.

In order to overcome the problems associated with the choice of the bins, Nass (1959) proposed an alternative test statistic. Let n denote the sample size and let m denote the sample maximum. The proposed test statistic is

$$N = \frac{\sum_{j=0}^{m} E_j^{-1} O_j^2 - n - m + 1}{\sqrt{\frac{m-1}{m}\left(2m - \frac{(m-1)^2+2m}{n} + \sum_{j=0}^{m} E_j^{-1}\right)}} .$$

This statistics eliminates the subjectivity associated with the choice of the bins used. The asymptotic null distribution of this test statistic is standard normal and the null hypothesis is rejected in the case of small and large values of N.

A range of similar tests exist in the literature, see Neyman and Pearson (1933), Freeman and Tukey (1950) and the likelihood ratio test in Horn (1977). Read and Cressie (1988) propose a general test which includes all of these tests as special cases. Their test statistic is given by

$$I(a) = \frac{1}{a\,(a+1)} \sum_{j=0}^{m} E_j \left(\left(\frac{O_j}{E_j} \right)^{a+1} - 1 \right),$$

where $a \in \mathbb{R}$ is a tuning parameter. In the numerical results reported in Section 3, we use $a = 5$, in accordance with the recommendations made in Read and Cressie (1988).

2.3 Tests based on the empirical distribution function

We now turn our attention to several classical goodness-of-fit tests based on the empirical df. Specifically, we consider the Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling tests. Each of these tests are based on a specific distance measure between the fitted Poisson distribution function and its empirical counterpart. The fitted df is given by $F_{\widehat{\lambda}}$, which denotes the df of the Poisson with estimated parameter $\widehat{\lambda} = \overline{X}$ (note that we use this notation for the estimated value of λ throughout). The empirical df is

$$F_n\,(x) = \frac{1}{n} \sum_{j=1}^{n} \mathrm{I}(X_j \leq x),$$

where $\mathrm{I}(\cdot)$ denotes the indicator function.

2.3.1 The Kolmogorov-Smirnov test

The celebrated Kolmogorov-Smirnov test statistic is

$$KS = \sup \left| F_{\widehat{\lambda}}\left(X_{(j)}\right) - F_n\left(X_{(j)}\right) \right|, \tag{6}$$

see Kolmogorov (1933). The assumption of Poissonity is rejected for large values of KS.

The Kolmogorov-Smirnov test was originally proposed for continuous distributions, but was later extended to discrete distributions, see Conover (1972). Campbell and Oprian (1979) develop an approximation to the Kolmogorov-Smirnov test specifically in the case of the Poisson distribution. The authors also provide a series of tables which could be used in order to perform the test for various values of the parameter estimate of the Poisson distribution. Henze (1996) overcame the need for the use of these tables using critical values based on a bootstrap procedure.

2.3.2 The Cramér-von Mises test (CM)

Cramér (1928) and Von Mises (1928) proposed a non parametric goodness-of-fit test for continuous distributions. Choulakian et al. (1994), extended this test to discrete distributions. Modified versions of the test statistic, specifically for use in testing the hypothesis of Poissonity, are provided in Spinelli and Stephens (1997) as well as Henze (1996). The Cramér-von Mises test statistic is given by

$$CM = \frac{1}{n}\sum_{j=0}^{\infty}(F_{\widehat{\lambda}}(j) - F_n(j))^2 f_{\widehat{\lambda}}(j). \tag{7}$$

The critical values of CM are typically estimated using simulation; see Spinelli and Stephens (1997).

A commonly used reweighted version of the of the Cramér-von Mises test statistic, known as the Anderson-Darling statistic, is given by

$$AD = \frac{1}{n}\sum_{j=0}^{\infty}\frac{[F_{\widehat{\lambda}}(j) - F_n(j)]^2 f_{\widehat{\lambda}}(j)}{F(j)(1-F(j))}. \tag{8}$$

The Anderson-Darling test is more sensitive to deviations from the Poisson in the tails of the distribution than is the case for the Cramér-von Mises. Both of these tests reject the hypothesis that an observed dataset is realised from a Poisson distribution for large values of the test statistics.

Note that computing the test statistics in (7) and (8) require the calculation of an infinite sum. In order to practically implement these tests, the test statistics are approximated by calculating the sum for a finite number of terms. Let M denote the upper summation limit of this finite sum. In Karlis and Xekalaki (2000), the value of M is chosen such that the probability of observing a sample observation exceeding M is no more than 10^{-4}. The powers reported in Section 3 are obtained by setting $M = 100$, which corresponds to the inclusion of a substantially greater number of terms than that which is recommended in Karlis and Xekalaki (2000).

2.3.3 The test of Klar (1999)

Klar (1999) proposes the use of the sum of the absolute differences between the theoretical and empirical distribution functions as a test statistic. This test statistic can be expressed as

$$L = \sqrt{n}\sum_{j=1}^{n}\left|F\left(X_{(j)}\right) - \widehat{F}\left(X_{(j)}\right)\right|. \tag{9}$$

The null hypothesis is rejected for large values of L. Klar (1999) estimates the critical value of L using simulation.

2.4 A test based on the integrated distribution function

In addition to the test based on the df, Klar (1999) also proposes a test statistic which is similar in form to the Kolmogorov-Smirnov test. However, this test statistic is based on the supremum difference between the integrated df and its empirical counterpart. Let

$$\Psi(t) = \int_t^{\infty}(1 - F(x))\,\mathrm{d}x,$$

be the integrated df and let

$$\Psi_n(t) = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - t\right) I_{(X_j > t)},$$

denote the empirical integrated df. The proposed statistic is given by

$$ID = \sup_{t\geq 0}\sqrt{n}\left|\Psi(t) - \Psi_n(t)\right|, \tag{10}$$

which rejects the null hypothesis for large values. Klar (1999) estimate the critical value of ID using simulation. Gürtler and Henze (2000) show that this test is powerful and outperforms the Kolmogorov-Smirnov and Cramér-von Mises type tests against the majority of alternative distributions considered.

2.5 Tests based on the probability generating function (pgf)

Let

$$g(t) = E\left(t^X\right), \quad t \in [-1, 1],$$

denote the pgf of a general count distribution. The pgf of a $Poisson(\lambda)$ distribution is given by

$$g_\lambda(t) = \exp(\lambda(t-1)), \quad t \in [-1, 1].$$

The empirical counterpart of the pgf is

$$g_n(t) = \frac{1}{n}\sum_{j=1}^{n} t^{X_j}, \quad t \in [-1, 1].$$

Below, we consider several goodness-of-fit tests based on these quantities.

2.5.1 The test of Rueda et al. (1991)

Based on ideas originally proposed in Kocherlakota and Kocherlakota (1986), Rueda et al. (1991) use the squared differences between the fitted pgf and the empirical pgf, integrated over the range of the support of these functions;

$$\begin{aligned} R &= \int_0^1 \left(\sqrt{n}\left(g_n(t) - g_{\widehat{\lambda}}(t)\right)\right)^2 \mathrm{d}t \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{X_i + X_j + 1} - 2e^{-\widehat{\lambda}}\sum_{i=1}^{n} T\left(X_i, \widehat{\lambda}\right) + n\frac{1 - e^{-2\widehat{\lambda}}}{2\widehat{\lambda}}, \end{aligned} \tag{11}$$

where $T(x, \lambda) = \int_0^1 t^x e^{\lambda t}\mathrm{d}t$. The statistic in (11) rejects the hypothesis of Poissonity for large values.

2.5.2 The test of Baringhaus et al. (2000)

Baringhaus et al. (2000) further generalise the statistic in (11) via the inclusion of a weight function in the integral. The resulting statistic is given by

$$R_a = n\int_0^1 \left(g_n(t) - g_{\widehat{\lambda}}(t)\right)^2 t^a \mathrm{d}t.$$

A convenient computational form for R_a, provided in Gürtler and Henze (2000), is

$$R_a = n\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(\frac{\left(f_n(j)-f_{\widehat{\lambda}}(j)\right)\left(f_n(k)-f_{\widehat{\lambda}}(k)\right)}{j+k+a+1}\right), \tag{12}$$

where $f_n(x) = n^{-1}\sum_{j=1}^{n} \mathrm{I}(X_j = x)$. The null hypothesis is rejected for large values of R_a. The asymptotic critical values of R_a are used in order to obtain powers in Baringhaus et al. (2000). In the numerical results presented below, we use $a = 5$, which is the recommended value of the tuning parameter.

2.5.3 The test of Baringhaus and Henze (1992)

Baringhaus and Henze (1992) also propose a goodness-of-fit test based on a property of the pgf of the Poisson distribution. The Poisson distribution is characterised by the following partial differential equation:

$$\frac{\partial}{\partial t}g_\lambda(t) = \lambda g_\lambda(t). \tag{13}$$

Consequently, Baringhaus and Henze (1992) base a goodness-of-fit test on the integrated squared difference between the empirical versions of the left and right hand sides of (13). The proposed test statistic is

$$\begin{aligned} T &= n\int_0^1 \left(\widehat{\lambda}g_n(t) - g_n'(t)\right)^2 \mathrm{d}t \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\widehat{\lambda}^2}{X_i+X_j+1} + \frac{X_iX_j}{X_i+X_j-1}\right) - \widehat{\lambda}\left(n - f_{\widehat{\lambda}}(0)\right). \end{aligned} \tag{14}$$

This test rejects the null hypothesis for large values of the test statistic.

2.5.4 The test of Treutler (1995)

Treutler (1995) proposes a generalisation of test statistic in (14) via the inclusion of a weight function incorporated into the integral. The resulting test statistic is

$$\begin{aligned} T_a &= n\int_0^1 \left(\widehat{\lambda}g_n(t) - g_n'(t)\right)^2 t^a\mathrm{d}t \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\widehat{\lambda}^2}{X_i+X_j+a+1} - \frac{\widehat{\lambda}\left(X_i+X_j\right)}{X_i+X_j+a} + \frac{X_iX_j}{X_i+X_j+a-1}\right), \end{aligned} \tag{15}$$

which rejects the Poisson assumption for large values of the test statistic.

2.5.5 The test of Nakamura and Pérez-Abreu (1993)

Nakamura and Pérez-Abreu (1993) base a test on the following characterisation of the Poisson distribution;

$$\frac{\partial^2}{\partial t^2}\log\left(g_\lambda(t)\right) = 0.$$

Based on the squared coefficients of the polynomial $g_n^2(t)\frac{\partial^2}{\partial t^2}\log(g_n(t))$, the authors propose the following test statistic:

$$V = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\left(X_i\left(X_i - X_j - 1\right)X_k\left(X_k - X_l - 1\right)\mathrm{I}_{(X_i+X_j=X_k+X_l)}\right). \tag{16}$$

Let $M = \max\{X_1, X_2, ..., X_n\}$ and let $\#\{\cdot\}$ denote the cardinality of a set. Setting

$$a_k = \sum_{l=0}^{k+2} l(2l-k-3)\#\{i : X_i = l\}\#\{j : X_j = k+2-l\},$$

the test statistic in (16) can be expressed in the following computationally efficient form:

$$V = n^{-3}\sum_{k=0}^{2M-2} a_k^2.$$

Numerical evidence suggests that the limit null distribution of

$$V^* = \left(\frac{V_n}{\overline{X}^{1.45}}\right) \tag{17}$$

is approximately independent of λ. As a result, Nakamura and Pérez-Abreu (1993) recommend the use of this form of the test statistic. V as well as V^* reject the hypothesis of Poissonity for large values. Nakamura and Pérez-Abreu (1993) use asymptotic critical values in order to obtain power estimates for V^*.

2.5.6 The test of Meintanis and Nikitin (2008)

A new class of count distributions, based on a property of the pgf, is defined in Meintanis and Nikitin (2008). Consider a general count distribution with mean λ and pgf g. Consider the function

$$D(t,\lambda) = g'(t) - \lambda g(t).$$

Let Δ denote the class of all distributions such that, for all $t \in [0, 1)$ and all $\lambda \in (0, \infty)$, we have that $D(t,\lambda) = 0$ or $D(t,\lambda) < 0$ or $D(t,\lambda) > 0$. Note that, if the underlying distribution is Poisson with mean λ, then $D(t,\lambda) = 0$ for all $t \in [0, 1)$. However, from the definition of the class of distributions Δ, it follows that D will be either positive or negative for every alternative distribution contained within this class.

The proposed test statistic is based on an estimate of D obtained by substituting the sample mean in the place of λ and estimating the pgf by its empirical counterpart, g_n. The proposed test rejects the hypothesis of Poissonity for small and large values of the test statistic:

$$MN_a^* = \sqrt{n}\int_0^1 D_n(t,\overline{X}_n)t^a \mathrm{d}t, \tag{18}$$

where t^a plays the role of a weight function and $a > 0$ is a tuning parameter. A computationally tractable expression for (18) is given by

$$MN_a^* = n^{-1/2}\sum_{j=1}^{n}\left(\frac{X_j}{X_j + a} - \frac{\overline{X}}{X_j + a + 1}\right). \tag{19}$$

Asymptotically, the statistic in (19) is normal with mean 0. The asymptotic variance of MN_a^* is

$$V_a(\lambda) = \lambda^2\tilde{\epsilon}_{a+2} + \lambda(\lambda+1)\tilde{\epsilon}_{a+2} + 2\lambda^2(\epsilon_{a+2} - \epsilon_{a+1n}) - \lambda\epsilon_{a+1}^2,$$

where $\epsilon_a = E[(X+a)^{-1}]$ and $\tilde{\epsilon}_a = E[(X+a)^{-2}]$, and $\lambda\epsilon_{a+1} = 1 - a\epsilon_a$ and $\lambda\tilde{\epsilon}_{a+1} = \epsilon_a - a\tilde{\epsilon}_a$. Using the empirical versions of these quantities, the test statistic can be standardised in order to ensure that the limit null distribution of the test statistic is standard normal. This is the approach followed in Meintanis and Nikitin (2008), where critical values are obtained from the standard normal distribution.

Let $\widehat{V_a}(\widehat{\lambda})$ be the estimate of $V_a(\lambda)$ obtained by replacing λ by the sample mean and using the empirical versions of $\epsilon_\cdot$ and $\tilde{\epsilon}_\cdot$. In the numerical results presented in the current paper, we use the test statistic

$$MN_a = \frac{MN_a^*}{V_n(a)^{1/2}}, \tag{20}$$

and we obtain critical values via simulation. MN_a rejects the null hypothesis for both small and large values. Meintanis and Nikitin (2008) show that the test described above is consistent against all alternative distributions contained in Δ.

2.6 Exact tests

Lockhart et al. (2007) propose a general method of constructing so-called exact tests for distributions admitting a minimally sufficient statistic. The authors propose the use of the Gibbs sampler to generate samples from the conditional distribution of the sample, given the value of the sufficient statistic. These samples are referred to as conditional samples.

In the case of a location, scale or location-scale family of distributions, the distribution of a goodness-of-fit statistic is independent of the underlying parameters of the distribution. As a result, a single critical value can be used for a given sample size. If, on the other hand, a shape parameter is present, then the critical value is a function of the (unobserved) value of this parameter. In the latter case, exact tests are of particular interest, since the conditional distribution of the sample given the value of the sufficient statistic is independent of the shape parameter. Since λ is, in fact, a shape parameter in the case of the Poisson distribution, exact tests are of interest. The sum $T = \sum_{j=1}^n X_j$ is a minimally sufficient statistic for λ in the Poisson distribution. Given the value of T, the conditional joint distribution of $(X_1, ..., X_n)$ is $multinomial(T, (n^{-1}, ..., n^{-1}))$. As a result, we can obtain samples from the conditional distribution of $(X_1, ..., X_n)$, given the value of T. A number of so-called exact tests for the hypothesis of Poissonity have been proposed based on the conditional multinomial distribution of the sample, given the value of T. These tests are considered below.

Typically, the implementation of tests of this kind are quite computationally expensive, since they usually require either simulation or exhaustive enumeration of all possible conditional samples. For a recent simulation study comparing the powers of various exact tests for the Poisson distribution, see Beltrán-Beltrán and O'Reilly (2019). This paper also explains the details of simulating conditional samples from the Poisson distribution given an observed value of T.

The finite sample power performance of the exact tests generally fall short of the other more powerful tests available in the literature and are omitted from the Monte Carlo study presented in Section 3. For details regarding the asymptotic distribution of the test statistics of exact tests, see Lockhart (2012).

2.6.1 The test of Gonzalez-Barrios et al. (2006)

Gonzalez-Barrios et al. (2006) propose the use of the conditional likelihood function as a test statistic as follows. Given the value of T, one can calculate the conditional likelihood of a given sample under the null hypothesis of Poissonity. Since the number of possible conditional samples is finite, one can calculate the likelihood associated with each possible sample and arrange these samples in decreasing order of likelihood. Fix the nominal significance level of the test, α. If we add the likelihood function values in decreasing order until we obtain a probability as close as possible to $1-\alpha$, then the remaining samples correspond to the rejection region for the test.

Application of this test requires enumeration of all of the possible samples for a given sample size and value of T. As a result, this test is extremely computationally expensive.

2.6.2 The test of Frey (2012)

Frey (2012) propose the use of a conditional Kolmogorov-Smirnov type test. The test is based on the supremum distance between the empirical df and the expected value of this function, given the value of T. The corresponding test statistic is

$$\begin{aligned} D &= \sup_{x\in\mathbb{R}} |F_n(x) - E[F_n(x)|T]| \\ &= \max_{x\in\{0,\dots,t\}} \left| F_n(x) - \sum_{j=0}^{x} \binom{t}{j} \left(\frac{1}{n}\right)^j \left(1-\frac{1}{n}\right)^{t-j} \right|, \end{aligned}$$

which rejects the null hypothesis for large values. This test is the first test for the Poisson distribution which offers exact p-values without the need for exhaustive enumeration. This substantially reduces the computational cost of the test. However, this cost remains substantial compared to other tests. Frey (2012) includes a power study. This test does not compare favourably to existing tests for the Poisson distribution, especially in the case where the mean of the alternative distribution is large.

2.6.3 The test of Beltrán-Beltrán and O'Reilly (2019)

Beltrán-Beltrán and O'Reilly (2019) propose two new tests for the Poisson distribution. The first is a discrete version of an existing test for the skew normal distribution, while the second is closely related to the test proposed in Gonzalez-Barrios et al. (2006). The latter test differs from the test of Gonzalez-Barrios et al. (2006) in that it is based on the ratio of (and not an absolute difference between) two conditional probabilities. The numerator of this ratio is the conditional probability of the observed sample given T, while the denominator is the conditional probability of the observed sample given the order statistics. Let o_j denote the number of times that the value j occurs in the observed sample. The resulting test statistic is given by

$$\Delta = n^{-T} \frac{T!}{\prod_{j=0}^{m} (j!)^{o_j}} \frac{n!}{\prod_{j=0}^{m} o_j!}.$$

The null hypothesis is rejected for small values of Δ.

2.7 Tests based on other characteristics

Below we consider tests based on other characteristics of the Poisson distribution. In addition to the use of Charlier polynomials, we consider a test based on the average distance between a Poisson random variable and some integer. Finally, we mention the use of graphical tests.

2.7.1 Tests based on Charlier polynomials

Ledwina and Wyłupek (2017) explain that a discrete function on the non-negative integers can be expressed as

$$f_\lambda(x)\left(1+\sum_{r=1}^{\infty}\frac{c_r}{r!}C_r(x;\lambda)\right),$$

where c_r is a sequence of constants and

$$C_r(x;\lambda)=\frac{\partial^r f_\lambda(x)}{\partial\lambda^r}\frac{1}{f_\lambda(x)}.$$

C_r are referred to as Charlier or Poisson-Charlier polynomials.

Charlier polynomials have been studied in various settings and a number of goodness-of-fit tests for the Poisson distribution have been developed in the process; see, for example, Best and Rayner (1999), Rayner and Best (1988) and Ledwina and Wyłupek (2017).

2.7.2 The test of Székely and Rizzo (2004) (SR)

Székely and Rizzo (2004) propose a test for the Poisson distribution based on the mean distance between a Poisson random variable and some integer value, k. The proposed test statistic is based on the following characterisation:

Let X be a non-negative, integer valued random variable with finite mean. Denote the pmf of X by f and the corresponding df by F. X follows a $Poisson(\lambda)$ distribution if, and only if,

$$E[k-X]=2(k-\lambda)F(k-1)+2\lambda f(k-1)-(k-\lambda),$$

for every non-negative integer k.

Let $m_k = E[X-k]$. The authors show that, if $X \sim Poisson(\lambda)$, then the mass function of X can be expressed as

$$f(k)=\frac{m_{k+1}-(k+1-\lambda)(2F(k-1)-1)}{2(k-1)}. \quad (21)$$

The proposed test statistic is a Cramér-von Mises type distance based on an estimate of the df resulting from (21). This estimate, denoted below by $\widehat{F}$, is obtained upon estimating λ by the sample mean, F by the empirical df and m_k by $\widehat{m}_k = \frac{1}{n}\sum_{j=1}^{n}|k-X_j|$. The resulting test statistic is

$$SR = n\sum_{j=0}^{\infty}(\widehat{F}(j)-F_{\widehat{\lambda}}(j))^2 f_{\widehat{\lambda}}(j). \quad (22)$$

This test rejects the assumption of the Poisson distribution for large values of SR. Székely and Rizzo (2004) use simulation in order to estimate the critical values of SR.

2.7.3 Graphical tests

Many graphical methods have also been developed to detect deviations from the Poisson distribution. See Lindsay (1986) and Lindsay and Roeder (1992) for two examples of the implementation of these tests. For an extensive list of graphical tests, see Karlis and Xekalaki (2000).

3. Numerical results

In this section, we consider and compare the finite sample power performance of ten of the tests discussed above. Some of the classical tests are included as well as several of the newer tests. Our aim is to ascertain which of the tests above are powerful against the various alternative distributions considered. The tests for which numerical results are presented are as follows:

1. a test based on the Fisher Index (U^2), see (2),
2. the Kolmogorov-Smirnov test (KS), see (6),
3. the Cramér-von Mises test (CM), see (7),
4. the Cramér-von Mises type test of Klar (1999) (L), see (9),
5. the integrated df based test of Klar (1999) (ID), see (10),
6. the pgf based test of Baringhaus et al. (2000) (R_a), see (12),
7. the pgf based test of Treutler (1995) (T_a), see (15),
8. the pgf based test of Nakamura and Pérez-Abreu (1993) (V^*), see (17),
9. the mean distances test of Székely and Rizzo (2004) (SR), see (22),
10. the test of Meintanis and Nikitin (2008) (MN_a), see (20).

Below, we discuss the alternative distributions considered before turning our attention to the bootstrap methodology used in order to implement the tests. Finally, we include a comparison of the various tests considered.

3.1 Alternative distributions

We compare the power of the tests considered against a variety of alternative distributions. One class of alternatives used is the so-called weighted Poisson distributions. Early references on this class of distributions include Fisher (1934) and Rao (1965). For a more recent discussion, see Kokonendji et al. (2008).

Let w be some function such that $w(x) \geq 0$ for $x \in \{0, 1, \ldots\}$ and let $\tilde{X} \sim Poisson(\lambda)$. X is said to be a weighted Poisson random variable, with parameter λ and weight function w, if X has pmf

$$f(x) = \frac{w(x) f_\lambda(x)}{\mathrm{E}[w(\tilde{X})]}, \quad x = 0, 1, \ldots,$$

with

$$\mathrm{E}[w(\tilde{X})] = \sum_{j=0}^{\infty} w(j) f_\lambda(x) < \infty.$$

In this paper, we limit our attention to the case where the weight function is a second degree polynomial;

$$w(x) = ax^2 + bx + 1.$$

Table 1. Alternative distributions considered.

Alternative distribution	Mass function	Notation
Discrete uniform	$(b-a+1)^{-1}$	$DU(a,b)$
Binomial	$\binom{m}{x}p^x\,(1-p)^x$	$Bin(m,p)$
Negative binomial	$\binom{r+x-1}{x}p^r\,(1-p)^x$	$NB(r,p)$
Poisson mixtures	$(x!)^{-1}\left\{p\lambda_1^x\mathrm{e}^{-\lambda_1}+(1-p)\,\lambda_2^x\mathrm{e}^{-\lambda_2}\right\}$	$PM(p,\lambda_1,\lambda_2)$
Generalised Poisson	$(x!)^{-1}\lambda_1\,(\lambda_1+x\lambda_2)^{x-1}\exp(-(\lambda_1+x\lambda_2))$	$GP(\lambda_1,\lambda_2)$
Zero inflated Poisson	$\left(p\frac{x!}{e^{-\lambda}\lambda^x}\mathrm{I}\,(x=0)+1-p\right)\frac{e^{-\lambda}\lambda^x}{x!}$	$ZIP(p,\lambda)$
Weighted Poisson	$(y!)^{-1}\lambda^y\exp(-\lambda)\frac{ay^2+by+1}{a(\lambda+\lambda^2)+b\lambda+1}$	$WP(\lambda,a,b)$

The pmf of this weighted Poisson distribution is available in closed form and is provided in Table 1.

Table 1 also contains the mass functions of each of the remaining alternative distributions used in the Monte Carlo study presented below. Note that each of the distributions listed in Table 1 has support $\{0,1,\ldots\}$, with two exceptions. Using the notation used in the table, the discrete uniform and binomial distributions have supports $\{a,a+1,\ldots,b\}$ and $\{0,1,\ldots,m\}$, respectively.

3.2 Power calculations

The null distribution of the test statistics considered depend on the value of the unknown shape parameter, λ. As a result, we use a parametric bootstrap procedure in order to perform power calculations. In order to ease the computational burden associated with the parametric bootstrap, we use the *warp-speed* method proposed in Giacomini et al. (2013) in order to approximate the power of each of the tests considered against various alternative distributions. Let MC be the number of Monte Carlo replications used. The warp-speed bootstrap is implemented as follows. (The algorithm provided below is adapted from Allison et al. (2019).)

1. Obtain a sample $X_1,\ldots,X_n$ from a distribution, say F, and estimate λ by $\widehat{\lambda}=\frac{1}{n}\sum_{j=1}^{n}X_j$.

2. Calculate the test statistic, $S=S(X_1,\ldots,X_n)$, say.

3. Generate a bootstrap sample $X_1^*,\ldots,X_n^*$ by independently sampling from a $Poisson(\widehat{\lambda})$ distribution and calculate the value of the test statistic for the bootstrap values, that is, $S^*=S(X_{n,1}^*,\ldots,X_{n,n}^*)$.

4. Repeat Steps 1 to 3 MC times to obtain $S_1,\ldots,S_{MC}$ and $S_1^*,\ldots,S_{MC}^*$, where S_m denotes the value of the test statistic calculated from the mth sample data generated in Step 1, and S_m^* denotes the value of the bootstrap test statistic calculated from the single bootstrap sample obtained in the mth iteration of the Monte Carlo simulation.

5. To obtain the power approximation, reject the null hypothesis for the ith sample whenever $S_i>S^*_{(\lfloor MC\cdot(1-\alpha)\rfloor)}$, $i=1,\ldots,MC$, where $S^*_{(1)}\leq\ldots\leq S^*_{(MC)}$ are the ordered values of the statistics obtained from the bootstrap samples and $\lfloor\cdot\rfloor$ denotes the floor function.

The algorithm described above is less computationally demanding than the traditional parametric bootstrap methodology described in Gürtler and Henze (2000) in which a separate bootstrap loop is used.

Below we consider power estimates of the ten tests against the alternative distributions in Table 1 using the warp-speed bootstrap. The reported power estimates are based on 50 000 replications in each case. The sample sizes considered are $n = 30, 50, 100, 200$. As is to be expected, the powers of the tests generally increase with sample size. However, the comparative performances of the tests do not seem to depend on sample size; i.e., if a test is relatively powerful against a given alternative for a small sample, then it is also relatively powerful against that alternative for a large sample. As a result, we discuss only the results obtained for $n = 50$ in the main text. The estimated powers associated with $n = 30, 100, 200$ can be found in the Appendix. A nominal significance level of 5% is used throughout.

The entries in the tables of power estimates show the percentage of samples for which the null hypothesis is rejected, rounded to the nearest integer. The highest power against each alternative is printed in bold in order to ease comparison.

The second column of each of the power tables contains the Fisher index of the distribution. The distributions in the tables are arranged according to this index. First we consider the equidispersed distributions, including the Poisson distribution for various values of λ. The Poisson is included in order to compare the sizes of the tests. After considering the equidispersed distributions, we examine the underdispersed alternatives. Finally we turn our attention to the overdispersed alternatives.

Consider the powers presented in Table 2. Note that, under the null hypothesis, each of the tests considered keep the nominal significance level of the test closely. This phenomenon is also observed in the additional power tables presented in the Appendix.

The majority of the tests do not provide high power against the two equidispersed distributions considered. The tests providing the highest powers against these distributions are L, V and SR. When turning our attention to underdispersed alternatives, we see that, generally, the MN test is most powerful, followed by the T and SR tests. The numerical results indicate that D provides the highest power against overdispersed alternatives, followed by KS, CM and V. Although the MN test does not outperform all of its competitors against any of the overdispersed distributions considered, this test provides relatively high powers against these distributions throughout. Based on the discussion above as well as the simple asymptotic null distribution of the MN test, we recommend the use of this test in the case where high power is desired against a wide range of alternative distributions.

4. Practical applications

Below, we use the tests considered in Section 3 to test the hypothesis that each of two observed datasets are realised from a Poisson distribution. The first dataset exhibits underdispersion, while the second dataset is overdispersed. For each dataset, we use the classical parametric bootstrap procedure, detailed in Section 3 of Gürtler and Henze (2000), to estimate the p-values of the tests. Each reported p-value in this section is based on 1 million bootstrap replications. The datasets are discussed below and Table 4 provides the calculated test statistic and estimated p-value associated with each test for both examples.

The first dataset, originally published in Zar (1999) and analysed again in Gürtler and Henze

Table 2. Estimated powers for sample size $n = 50$.

Distribution	FI	U^2	KS	CM	L	ID	R_a	T_a	V^*	SR	MN_a
$Poisson(0.5)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(1)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(5)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(10)$	1.00	5	5	5	5	5	5	5	5	5	5
$DU(0,4)$	1.00	1	9	9	54	33	60	18	**74**	53	6
$WP(1,1,-1)$	1.00	4	7	7	**21**	18	16	8	15	**21**	6
$DU(0,2)$	0.67	37	29	31	52	53	59	67	**75**	52	33
$WP(1,2,1)$	0.68	40	45	46	30	13	34	41	15	30	**47**
$WP(1,1,1)$	0.75	23	26	27	19	7	21	25	8	19	**28**
$Bin(4,0.25)$	0.75	18	22	23	18	5	18	22	5	18	**25**
$Bin(20,0.25)$	0.75	21	20	19	17	7	18	23	8	17	**24**
$WP(1,1,0)$	0.83	10	11	11	13	5	12	**14**	6	13	11
$Bin(10,0.1)$	0.90	5	**7**	**7**	**7**	3	6	**7**	2	**7**	**7**
$Bin(50,0.1)$	0.90	6	6	6	6	4	6	**7**	4	**7**	**7**
$PM(0.01,1,5)$	1.03	6	**7**	**7**	5	5	6	5	6	5	6
$NB(9,0.9)$	1.11	**11**	9	9	7	10	8	8	**11**	7	9
$NB(45,0.9)$	1.11	**10**	**10**	9	7	8	9	9	**10**	6	9
$PM(0.05,1,5)$	1.16	15	**23**	**23**	8	8	13	12	12	8	20
$DU(0,5)$	1.17	11	42	41	66	48	74	40	**83**	66	34
$GPD(4,0.1)$	1.24	**23**	21	20	11	14	18	19	19	11	19
$PM(0.5,3,5)$	1.25	**24**	23	22	13	16	20	21	21	13	22
$ZIP(0.9,3)$	1.30	31	**47**	**47**	22	38	31	30	32	21	42
$NB(15,0.75)$	1.33	**36**	32	32	17	21	28	30	29	17	30
$NB(3,0.75)$	1.33	**35**	30	30	19	30	25	27	32	19	31
$DU(0,6)$	1.33	37	75	74	79	64	87	66	**91**	79	68
$NB(2,2/3)$	1.50	**54**	48	49	35	47	42	44	50	34	48
$NB(10,2/3)$	1.50	**57**	54	54	30	35	47	50	48	30	52
$ZIP(0.8,3)$	1.60	75	89	89	68	**90**	79	78	79	67	86
$PM(0.5,2,5)$	1.64	**77**	**77**	76	55	61	71	74	70	55	75
$NB(1,0.5)$	2.00	**86**	81	82	74	82	77	78	82	73	83

Table 3. Annual number of deaths due to horsekick in the Prussian army.

Count	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Frequency	1	1	2	2	1	1	2	1	3	1	0	1	2	0	1	1

Table 4. P-values of the tests under consideration.

	Sparrows		Horsekicks	
Test	Statistic	p value	Statistic	p value
U^2	1.59	0.19	7.63	0.01
KS	0.48	0.09	0.70	0.05
CM	0.24	0.03	0.14	0.10
L	1.36	0.07	5.09	0.02
I	0.68	0.05	2.48	0.01
R	0.00	0.10	0.00	0.01
T	0.02	0.09	0.21	0.01
V^*	2.30	0.22	6.05	0.03
SR	0.24	0.03	0.14	0.10
MN	0.09	0.11	−0.13	0.02

(2000), recorded the number of sparrow nests found on 40 one hectare plots. Of these, 9 areas showed no nests, 22 areas showed one, 6 areas showed two, 2 areas showed three, and the remaining area contained four nests. The sample mean and variance of the data are 1.1 and 0.81, respectively, resulting in a sample Fisher index of 0.74.

The second dataset consists of the annual number of deaths due to horsekick in the Prussian army between 1875 and 1894. These data were analysed in Bortkiewicz (1898) as well as Gürtler and Henze (2000). For this dataset, the sample mean and variance are 9.8 and 19.33 respectively, meaning that the sample Fisher index is 1.97. The observed frequencies are reported in Table 3.

Consider the p-values associated with the first example in Table 4. Only the CM and SR tests reject the assumption of Poissonity at a 5% level of significance. In contrast to the first example, the majority of the tests considered reject the null hypothesis in the second example at the 5% level.

5. Conclusions

In this paper, we consider a large number of tests for the Poisson distribution. These tests are based on various characteristics of this distribution, including the moments of the distribution, the distribution function, the integrated distribution function, and the probability generating function, the mean distance between a Poisson random variable and some integer and the joint distribution of the sample given its sum.

We compare the finite sample performance of ten different tests, using a warp-speed bootstrap methodology, against a wide range of alternatives. The list of alternatives considered include the class of weighted Poisson distributions, seemingly for the first time. The alternatives considered include several overdispersed as well as underdispersed distributions. The most powerful test against underdispersed alternatives is found to be the test of Meintanis and Nikitin (2008), while the test of Rayner and Best (1990) (originally studied in Fisher, 1950) proves to be the most powerful against overdispersed alternatives. When the aim is to achieve substantial power against a large variety of deviations from the Poisson distribution, we recommend the use of the test of Meintanis and Nikitin (2008).

Several additional tests which lack power but are of historical importance are discussed but not included in the power study. Two practical applications are also included in the paper.

6. Appendix

The estimated powers for the various tests considered are provided for sample sizes $n = 30, 100, 200$ in Tables 5, 6 and 7, respectively.

References

Allison, J. S., Betsch, S., Ebner, B., and Visagie, I. J. H. (2019). New weighted L^2-type tests for the inverse Gaussian distribution. Preprint, arXiv:1910.14119v1.

Allison, J. S. and Santana, L. (2015). On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, **85**, 3276–3288.

Anderson, J. C. and Siddiqui, M. M. (1994). The sampling distribution of the index of dispersion. *Communications in Statistics – Theory and Methods*, **23**, 897–911.

Baringhaus, L., Gürtler, N., and Henze, N. (2000). Weighted integral test statistics and components of smooth tests of fit. *Australian & New Zealand Journal of Statistics*, **42**, 179–192.

Baringhaus, L. and Henze, N. (1992). A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statistics & Probability Letters*, **13**, 269–274.

Bartko, J. J., Greenhouse, S. W., and Patlak, C. S. (1968). On expectations of some functions of Poisson variates. *Biometrics*, **24**, 97–102.

Bateman, G. I. (1950). The power of the χ^2 index of dispersion test when Neyman's contagious distribution is the alternate hypothesis. *Biometrika*, **37**, 59–63.

Beltrán-Beltrán, J. I. and O'Reilly, F. J. (2019). On goodness of fit tests for the Poisson, negative binomial and binomial distributions. *Statistical Papers*, **60**, 1–18.

Best, D. J. and Rayner, J. C. W. (1999). Goodness of fit for the Poisson distribution. *Statistics & Probability Letters*, **44**, 259–265.

Böhning, D., Dietz, E., Schaub, R., Schlattmann, P., and Lindsay, B. G. (1994). The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. *Annals of the Institute of Statistical Mathematics*, **46**, 373–388.

Bortkiewicz, L. V. (1898). *Das Gesetz der kleinen Zahlen*. Teubner, Leipzig.

Campbell, D. B. and Oprian, C. A. (1979). On the Kolmogorov–Smirnov test for the Poisson distribution with unknown mean. *Biometrical Journal*, **21**, 17–24.

Choulakian, V., Lockhart, R. A., and Stephens, M. A. (1994). Cramér-von Mises statistics for discrete distributions. *Canadian Journal of Statistics*, **22**, 125–137.

Collings, B. J. and Margolin, B. H. (1985). Testing goodness of fit for the Poisson assumption when observations are not identically distributed. *Journal of the American Statistical Association*, **80**, 411–418.

Conover, W. J. (1972). A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association*, **67**, 591–596.

Cramér, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, **1928**, 13–74.

Dahiya, R. C. and Gurland, J. (1969). 257. Note: Functions of the sample mean and sample variance of a poisson variate. *Biometrics*, **25**, 171–173.

Darwin, J. H. (1957). The power of the Poisson index of dispersion. *Biometrika*, **44**, 286–289.

De Oliveira, T. (1963). Some elementary tests for mixtures of discrete distribution. Technical report, Columbia University, New York.

Fisher, R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Human Genetics*, **6**, 13–25.

Fisher, R. A. (1950). The significance of deviations from expectation in a Poisson series. *Biometrics*, **6**, 17–24.

Freeman, M. F. and Tukey, J. W. (1950). Transformations related to the angular and the square root. *The Annals of Mathematical Statistics*, **21**, 607–611.

Frey, J. (2012). An exact Kolmogorov–Smirnov test for the Poisson distribution with unknown mean. *Journal of Statistical Computation and Simulation*, **82**, 1023–1033.

Giacomini, R., Politis, D. N., and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, **29**, 567–589.

Gonzalez-Barrios, J. M., O Reilly, F., and Rueda, R. (2006). Goodness of fit for discrete random variables using the conditional density. *Metrika*, **64**, 77–94.

Gupta, A. K., Móri, T. F., and Székely, G. J. (1994). Testing for Poissonity-normality vs. other infinite divisibility. *Statistics & Probability Letters*, **19**, 245–248.

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*. Wiley & Sons, New York.

Henze, N. (1996). Empirical-distribution-function goodness-of-fit tests for discrete models. *Canadian Journal of Statistics*, **24**, 81–93.

Henze, N. and Klar, B. (1996). Properly rescaled components of smooth tests of fit are diagnostic. *Australian Journal of Statistics*, **38**, 61–74.

Horn, S. D. (1977). Goodness-of-fit tests for discrete data: A review and an application to a health impairment scale. *Biometrics*, **33**, 237–247.

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*. Wiley & Sons, Hoboken, New Jersey.

Karlis, D. and Xekalaki, E. (2000). A simulation comparison of several procedures for testing the Poisson assumption. *Journal of the Royal Statistical Society: Series D*, **49**, 355–382.

Kharshikar, A. (1970). On expected value of $S^2/\bar{X}$. *Biometrics*, **26**, 343 - 346.

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

Kocherlakota, S. and Kocherlakota, K. (1986). Goodness of fit tests for discrete distributions. *Communications in Statistics – Theory and Methods*, **15**, 815–829.

Kokonendji, C. C., Mizere, D., and Balakrishnan, N. (2008). Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference*, **138**, 1287–1296.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 83–91.

Kyriakoussis, A., Li, G., and Papadopoulos, A. (1998). On characterization and goodness-of-fit test of some discrete distribution families. *Journal of Statistical Planning and Inference*, **74**, 215–228.

Ledwina, T. and Wyłupek, G. (2017). On Charlier polynomials in testing Poissonity. *Communications in Statistics – Simulation and Computation*, **46**, 1918–1932.

Lindsay, B. G. (1986). Exponential family mixture models (with least-squares estimators). *The Annals of Statistics*, **14**, 124–137.

Lindsay, B. G. and Roeder, K. (1992). Residual diagnostics for mixture models. *Journal of the American Statistical Association*, **87**, 785–794.

Lockhart, R. A. (2012). Conditional limit laws for goodness-of-fit tests. *Bernoulli*, **18**, 857–882.

Lockhart, R. A., O'Reilly, F. J., and Stephens, M. A. (2007). Use of the Gibbs sampler to obtain conditional tests, with applications. *Biometrika*, **94**, 992–998.

Meintanis, S. G. (2016). A review of testing procedures based on the empirical characteristic function. *The South African Statistical Journal*, **50**, 1–14.

Meintanis, S. G. and Nikitin, Y. Y. (2008). A class of count models and a new consistent test for the Poisson distribution. *Journal of Statistical Planning and Inference*, **138**, 3722–3732.

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.

Nass, C. A. G. (1959). The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism. *Biometrika*, **46**, 365–385.

Neyman, J. and Pearson, E. S. (1933). IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society, Series A*, **231**, 289–337.

Nikitin, Y. Y. (2019). Tests based on characterizations and their efficiencies: A survey. Preprint, arXiv:1707.01522v1.

Patil, G. and Joshi, S. (1968). *Dictionary and Bibliography of Discrete Distributions*. Oliver & Boyd, Edinburgh.

Perry, J. N. and Mead, R. (1979). On the power of the index of dispersion test to detect spatial pattern. *Biometrics*, **35**, 613–622.

Pettigrew, H. M. and Mohler, W. C. (1967). A rapid test for the Poisson distribution using the range. *Biometrics*, **23**, 685–692.

Poisson, S.-D. (1828). *Mémoire sur l'équilibre et mouvement des corps élastiques*. L'Académie

Royale des Sciences, Paris.

Potthoff, R. F. and Whittinghill, M. (1966). Testing for homogeneity: II. The Poisson distribution. *Biometrika*, **53**, 183–190.

Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment. *Sankhyā: The Indian Journal of Statistics, Series A*, **27**, 311–324.

Rayner, J. C. W. and Best, D. J. (1988). Smooth tests of goodness of fit for regular distributions. *Communications in Statistics – Theory and Methods*, **17**, 3235–3267.

Rayner, J. C. W. and Best, D. J. (1990). Smooth tests of goodness of fit: An overview. *International Statistical Review*, **58**, 9–17.

Rayner, J. C. W. and McIntyre, R. (1985). Use of the score statistic for testing goodness of fit of some generalised distributions. *Biometrical Journal*, **27**, 159–165.

Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-fit Statistics for Discrete Multivariate Data*. Springer, New York.

Rueda, R., O Reilly, F., and Pérez-Abreu, V. (1991). Goodness of fit for the Poisson distribution based on the probability generating function. *Communications in Statistics – Theory and Methods*, **20**, 3093–3110.

Selby, B. (1965). The index of dispersion as a test statistic. *Biometrika*, **52**, 627–629.

Soo Kim, B. and Park, C. (1992). Some remarks on testing goodness of fit for the Poisson assumption. *Communications in Statistics – Theory and Methods*, **21**, 979–995.

Spinelli, J. J. and Stephens, M. A. (1997). Cramér-von Mises tests of fit for the Poisson distribution. *Canadian Journal of Statistics*, **25**, 257–268.

Székely, G. J. and Rizzo, M. L. (2004). Mean distance test of Poisson distribution. *Statistics & Probability Letters*, **67**, 241–247.

Treutler, B. (1995). *Tests for the Poisson distribution (in German)*. Ph.D. thesis, University of Karlsruhe.

Von Mises, R. (1928). *Statistik und Wahrheit*. Springer, Vienna.

Zar, J. H. (1999). *Biostatistical Analysis*. Fourth edition. Prentice Hall, Upper Saddle River, New Jersey.

Zelterman, D. and Chen, C.-F. (1988). Homogeneity tests against central-mixture alternatives. *Journal of the American Statistical Association*, **83**, 179–182.

Manuscript received 2020-05-14, revised 2020-06-06, accepted 2020-06-07.

Table 5. Estimated powers for sample size $n = 30$.

Distribution	FI	U^2	KS	CM	L	ID	R_a	T_a	V^*	SR	MN_a
$Poisson(0.5)$	1.00	5	4	4	5	5	4	4	5	5	5
$Poisson(1)$	1.00	5	5	5	5	5	4	5	5	5	5
$Poisson(5)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(10)$	1.00	5	5	5	5	5	5	5	5	5	5
$DU(0,4)$	1.00	2	6	6	32	19	34	10	**43**	32	4
$WP(1,1,-1)$	1.00	4	6	6	**13**	11	10	6	10	**13**	5
$DU(0,2)$	0.67	18	18	18	28	29	33	**38**	29	28	19
$WP(1,2,1)$	0.68	22	27	27	20	8	20	26	6	19	**30**
$WP(1,1,1)$	0.75	12	16	17	13	5	12	16	4	13	**18**
$Bin(4,0.25)$	0.75	11	**15**	**15**	13	4	11	14	2	13	**15**
$Bin(20,0.25)$	0.75	12	12	11	12	6	11	**15**	4	11	14
$WP(1,1,0)$	0.83	6	8	8	**10**	4	8	**10**	3	10	9
$Bin(10,0.1)$	0.90	4	5	5	6	3	5	6	2	6	6
$Bin(50,0.1)$	0.90	5	5	5	**6**	4	5	**6**	3	**6**	**6**
$PM(0.01,1,5)$	1.03	6	**7**	**7**	5	5	6	5	6	5	6
$NB(9,0.9)$	1.11	**10**	8	8	6	9	7	7	**10**	6	8
$NB(45,0.9)$	1.11	**9**	**9**	8	6	7	8	7	**9**	6	7
$PM(0.05,1,5)$	1.16	12	**18**	**18**	7	7	10	10	11	7	15
$DU(0,5)$	1.17	8	26	27	41	31	47	22	**56**	40	21
$GPD(4,0.1)$	1.24	**17**	16	15	8	11	13	13	16	8	13
$PM(0.5,3,5)$	1.25	**18**	17	17	9	13	15	14	17]	10	16
$ZIP(0.9,3)$	1.30	23	**33**	**33**	14	26	21	19	24	14	29
$NB(15,0.75)$	1.33	**26**	24	23	12	16	20	20	23	12	21
$NB(3,0.75)$	1.33	**25**	20	21	13	21	18	18	**25**	14	21
$DU(0,6)$	1.33	24	54	54	54	44	62	43	**68**	53	47
$NB(2,2/3)$	1.50	**39**	34	34	23	34	28	31	**39**	23	36
$NB(10,2/3)$	1.50	**42**	39	38	20	25	34	34	36	19	35
$ZIP(0.8,3)$	1.60	55	**71**	**71**	44	70	56	55	59	44	67
$PM(1/2,2,5)$	1.64	**57**	**57**	**57**	35	41	49	52	52	35	55
$NB(1,1/2)$	2.00	**69**	62	63	53	64	56	58	67	53	65

Table 6. Estimated powers for sample size $n = 100$.

Distribution	FI	U^2	KS	CM	L	ID	R_a	T_a	V^*	SR	MN_a
$Poisson(0.5)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(1)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(5)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(10)$	1.00	5	5	5	5	5	5	5	5	5	5
$DU(0,4)$	1.00	1	18	16	88	68	94	45	**99**	88	10
$WP(1,1,-1)$	1.00	4	10	9	**40**	32	35	13	29	**40**	7
$DU(0,2)$	0.67	82	60	62	92	97	97	98	**100**	92	69
$WP(1,2,1)$	0.68	73	**78**	**78**	57	27	67	74	43	57	78
$WP(1,1,1)$	0.75	47	**51**	**51**	34	14	41	46	21	34	**51**
$Bin(4,0.25)$	0.75	45	46	46	35	11	36	43	17	35	**49**
$Bin(20,0.25)$	0.75	**45**	41	40	30	13	36	44	22	30	**45**
$WP(1,1,0)$	0.83	20	18	18	22	6	21	**24**	12	22	20
$Bin(10,0.1)$	0.90	8	**10**	**10**	8	3	8	9	3	8	**10**
$Bin(50,0.1)$	0.90	9	9	9	8	4	8	**10**	5	8	**10**
$PM(0.01,1,5)$	1.03	6	**8**	**8**	6	5	6	6	6	5	7
$NB(9,0.9)$	1.11	**15**	12	12	9	14	10	11	13	8	12
$NB(45,0.9)$	1.11	**14**	13	13	8	10	11	12	12	8	12
$PM(0.05,1,5)$	1.16	22	35	**36**	11	11	20	19	18	11	31
$DU(0,5)$	1.17	19	72	70	95	81	98	74	**99**	95	61
$GPD(4,0.1)$	1.24	**36**	32	32	17	22	29	31	27	18	31
$PM(0.5,3,5)$	1.25	**40**	37	37	22	26	33	36	31	22	36
$ZIP(0.9,3)$	1.30	51	**73**	72	38	68	55	54	52	39	68
$NB(15,0.75)$	1.33	**57**	53	52	30	35	47	50	45	30	52
$NB(3,0.75)$	1.33	**54**	49	50	35	47	41	46	47	34	50
$DU(0,6)$	1.33	63	96	95	98	93	**100**	94	**100**	98	93
$NB(2,2/3)$	1.50	**79**	74	75	60	72	68	71	71	61	75
$NB(10,2/3)$	1.50	**83**	80	80	53	58	73	77	71	54	78
$ZIP(0.8,3)$	1.60	95	99	99	94	**100**	98	98	97	94	99
$PM(0.5,2,5)$	1.64	**96**	**96**	96	86	88	94	95	92	86	96
$NB(1,0.5)$	2.00	**99**	98	98	96	98	96	97	97	96	98

Table 7. Estimated powers for sample size $n = 200$.

Distribution	FI	U^2	KS	CM	L	ID	R_a	T_a	V^*	SR	MN_a
$Poisson(0.5)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(1)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(5)$	1.00	5	5	5	5	5	5	5	5	5	5
$Poisson(10)$	1.00	5	5	5	5	5	5	5	5	5	5
$DU(0,4)$	1.00	1	37	33	**100**	99	**100**	93	**100**	**100**	20
$WP(1,1,-1)$	1.00	4	15	14	**74**	59	70	28	57	**74**	10
$DU(0,2)$	0.67	**100**	93	94	**100**	**100**	**100**	**100**	**100**	**100**	96
$WP(1,2,1)$	0.68	96	97	**98**	88	57	94	96	84	88	**98**
$WP(1,1,1)$	0.75	79	81	82	63	29	72	78	54	63	**83**
$Bin(4,0.25)$	0.75	**81**	79	79	64	26	69	75	47	64	80
$Bin(20,0.25)$	0.75	**80**	73	73	55	27	68	76	53	55	76
$WP(1,1,0)$	0.83	40	32	33	40	12	40	**44**	30	40	36
$Bin(10,0.1)$	0.90	15	16	16	12	3	13	15	6	13	**17**
$Bin(50,0.1)$	0.90	15	15	15	11	5	13	**16**	8	11	**16**
$PM(0.01,1,5)$	1.03	7	**9**	**9**	6	5	6	6	6	5	8
$NB(9,0.9)$	1.11	**21**	19	19	12	19	16	17	17	12	19
$NB(45,0.9)$	1.11	**21**	19	19	11	14	16	18	15	11	17
$PM(0.05,1,5)$	1.16	35	**56**	**56**	18	19	33	33	29	18	53
$DU(0,5)$	1.17	36	95	94	**100**	99	**100**	98	**100**	**100**	89
$GPD(4,0.1)$	1.24	**58**	53	53	32	36	48	52	43	31	53
$PM(0.5,3,5)$	1.25	**65**	61	61	40	44	56	61	51	40	61
$ZIP(0.9,3)$	1.30	79	**95**	94	70	94	85	84	81	71	93
$NB(15,0.75)$	1.33	**83**	79	79	54	57	74	77	69	53	79
$NB(3,0.75)$	1.33	**80**	77	78	61	73	69	73	69	61	78
$DU(0,6)$	1.33	91	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
$NB(2,2/3)$	1.50	**96**	95	95	89	93	92	94	92	88	**96**
$NB(10,2/3)$	1.50	**98**	97	97	84	85	95	96	93	84	97
$ZIP(0.8,3)$	1.60	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
$PM(0.5,2,5)$	1.64	**100**	**100**	**100**	99	99	**100**	**100**	**100**	99	**100**
$NB(1,0.5)$	2.00	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**