

# Specification Tests for the Poisson Distribution with a Stochastic Parameter

Simos G. Meintanis\*

Department of Economics, National and Kapodistrian University of Athens, Athens, Greece

**Abstract:** A new method is proposed for testing the mixed Poisson model which employs the well known identity between the probability generating function of the mixed Poisson distribution and the Laplace transform of the mixing distribution. The method utilizes a weighted L2-type statistic which may be easily calculated and the resulting test is consistent against general alternatives. A limit statistic is shown to be related to earlier methods which employ moment-based contrasts. Some simulation results are also shown.

**Keywords:** Goodness-of-fit test, Probability generating function, Laplace transform, Mixed Poisson distribution.

## 1. INTRODUCTION

The Poisson distribution typically constitutes a benchmark model in most situations involving counts. However a feature which is frequently encountered when faced with such data is the presence of overdispersion (or underdispersion) or some form of clustering, often found in biological and financial data. Hence there is a need to consider more general count models than the Poisson, which can accommodate such features and develop statistical techniques for them. The family of mixed Poisson (MP) distributions, whereby the conditional Poisson variable is coupled with a density of positive support for the (stochastic) Poisson parameter, is one approach towards generalizing the Poisson model. Members of this class are the Poisson distribution, the negative binomial distribution, as well as other less well-known models. Many researchers have considered MP laws; see for instance, Karlis and Xekalaki [1], Green and Plotkin [2], Fang [3, 4], Carriere [5], Willmot [6, 7], and Nye and Hofflander [8].

The MP class of laws is conveniently defined *via* the corresponding probability function

$$f(x) = P(X = x) = \int_0^\infty e^{-\lambda} \frac{\lambda^x}{x!} g(\lambda) d\lambda, x = 0, 1, \dots, \quad (1)$$

where  $g(\cdot)$  is called the mixing distribution. To fix notation, let  $P_X(t) = E(t^X)$  denote the probability generating function (PGF) of an arbitrary count variable  $X$ . Also write  $L_g(t) = E[e^{-\lambda(1-t)}]$  for the (modified) Laplace transform of  $g(\cdot)$ . It can be easily shown that if  $X$  is MP with mixing distribution  $g(\cdot)$ , then

$$P_X \equiv L_g. \quad (2)$$

In fact, it is well known that for MP models, the knowledge of  $g(\cdot)$  is equivalent to the knowledge of  $f(\cdot)$ , and that  $X$  is MP if and only if the PGF of  $X$  is equal to the LT of a random variable with positive support.

In what follows, we employ (2) to test the null hypothesis  $H_0$  that the law of  $X$  is MP for some specified distribution  $g(\cdot)$ , based on independent copies  $X_1, X_2, \dots, X_n$ , of  $X$ . In our approach we replace in (2), the PGF by its empirical counterpart  $P_n(t) = n^{-1} \sum_{j=1}^n t^{X_j}$ . Then the test statistics may be written as

$$T_{n,\gamma} = n \int_0^1 D_n^2(t) t^\gamma dt, \gamma > 0, \quad (3)$$

where  $D_n(t) := D(P_n(t), \hat{\theta}_n; t)$  is some distance measure incorporating  $P_n(t)$ , and  $\hat{\theta}_n$  denotes a consistent estimator of the parameter  $\theta$  (possibly vector-valued) which is involved in the mixing distribution  $g(\cdot)$  (under the null hypothesis  $H_0$ ). Note that the test statistic in equation (3) is formulated as a weighted L2-type statistic, and in this connection the function  $D_n(t)$  plays the role of an empirical contrast which should be identically, i.e. for all  $0 < t < 1$ , equal to zero under  $H_0$ , as  $n \rightarrow \infty$ . For more information on these type of statistics the reader is referred to Meintanis and Swanepoel [9].

## 2. TEST STATISTICS

Typical examples for the mixing distribution  $g(\cdot)$  of  $\lambda$  are the Gamma distribution and the uniform distribution. A less well known, but nevertheless interesting, example is when stochastic behavior of  $\lambda$  is driven by a strictly positive stable law. These possibilities are considered below.

(i) Gamma mixing: Then the density of  $\lambda$  is given by  $g(\lambda) = (\vartheta^\alpha \Gamma(\alpha))^{-1} \lambda^{\alpha-1} \exp(-\lambda/\vartheta)$ ,  $\lambda > 0$ . Consequently

\*Address correspondence to this author at the Department of Economics, National and Kapodistrian University of Athens, Athens, Greece; Tel: ++302103689814; E-mail: simosmei@econ.uoa.gr

one gets from (1) the well known fact that the distribution of  $X$  is negative binomial with probability function

$$f(x) = \frac{1}{x! \Gamma(\alpha)} \frac{\vartheta^x}{(\vartheta + 1)^{\alpha+x}} \Gamma(\alpha + x).$$

Since the LT of  $g(\cdot)$  is given by  $L_g(t) = [1 + \vartheta(1-t)]^{-\alpha}$  with  $L'_g(t) = \alpha\vartheta[1 + \vartheta(1-t)]^{-(\alpha+1)}$ , an appropriate distance function is

$$D_n(t) = [1 + \vartheta(1-t)]P'_n(t) - \alpha\vartheta P_n(t). \tag{4}$$

By straightforward calculation we arrive from (3) at the following representation of the test statistic,

$$\begin{aligned} T_{n,\gamma} &= \frac{1}{n} \sum_{j,k=1}^n X_j X_k \\ &\left\{ (1 + \vartheta)^2 I(X_{jk} + \gamma - 2) + \vartheta^2 I(X_{jk} + \gamma) - (2\vartheta^2 + 2\vartheta) I(X_{jk} + \gamma - 1) \right\} \\ &+ \alpha^2 \vartheta^2 \frac{1}{n} \sum_{j,k=1}^n I(X_{jk} + \gamma) - 2\alpha\vartheta \frac{1}{n} \sum_{j,k=1}^n X_j \\ &\left\{ (1 + \vartheta) I(X_{jk} + \gamma - 1) - \vartheta I(X_{jk} + \gamma) \right\}, \end{aligned}$$

where  $X_{jk} = X_j + X_k$  and  $I(x) = \int_0^1 t^x dt = (1+x)^{-1}$ .

(ii) Uniform mixing: Then the density of  $\lambda$  is given by  $g(\lambda) = 1/\alpha, 0 < \lambda < \alpha$ . Consequently one gets from (1) the probability function of  $X$  as

$$f(x) = \frac{1}{\alpha} \left[ 1 - \frac{\Gamma(1+x, \alpha)}{x!} \right],$$

where  $\Gamma(x_1, x_2) = \int_{x_2}^{\infty} t^{x_1-1} e^{-t} dt$  is the complimentary incomplete Gamma function. Since the LT of  $g(\cdot)$  is given by  $L_g(t) = \alpha^{-1} [1 - e^{-\alpha(1-t)}] (1-t)^{-1}$ , an appropriate distance function is

$$D_n(t) = \alpha(1-t)P'_n(t) + e^{\alpha(1-t)} - 1.$$

By straightforward calculation we arrive from (3) at the following representation of the test statistic,

$$\begin{aligned} T_{n,\gamma} &= \alpha^2 \frac{1}{n} \sum_{j,k=1}^n \left\{ I(X_{jk} + \gamma) + I(X_{jk} + \gamma + 2) - 2I(X_{jk} + \gamma + 1) \right\} \\ &+ 2\alpha e^\alpha \sum_{j,k=1}^n \left\{ J(\alpha, X_j + \gamma) - J(\alpha, X_j + \gamma + 1) \right\} \\ &- 2\alpha \sum_{j=1}^n \left\{ I(X_j + \gamma) - I(X_j + \gamma + 1) \right\} \\ &+ n \left[ e^{2\alpha} J(2\alpha, \gamma) - 2e^\alpha J(\alpha, \gamma) \right] \end{aligned}$$

where  $J(x_1, x_2) = x_1^{-1-x_2} [\Gamma(1+x_2) - \Gamma(1+x_2, x_1)]$ .

(iii) Stable mixing: Strictly positive stable laws are conveniently defined via the LT  $L_g(t) = e^{-c\alpha(1-t)^\alpha}$ ; see for

instance Meintanis [10]. Scale is denoted by  $c > 0$ , while  $0 < \alpha < 1$ , often termed the characteristic exponent, denotes a shape parameter regulating the tail behavior of the underlying density. For  $\alpha = 1/2$ , the density of  $\lambda$  is given by  $g(\lambda) = (1/2)\sqrt{c(\pi\lambda^3)^{-1}} e^{-c/4\lambda}, \lambda > 0$ . Consequently one gets from (1) the probability function of  $X$  as

$$f(x) = \frac{1}{x! \sqrt{\pi}} 2^{\frac{1-x}{2}} c^{\frac{1}{2}(x-1)} K\left(\frac{1}{2} - x, \sqrt{c}\right),$$

where  $K(\cdot, \cdot)$  denotes the modified Bessel function of the second kind. When  $\alpha \neq 1/2$ , closed-form expression for  $g(\lambda)$  (and of course for  $f(x)$ ) are not available. Since the LT of  $g(\cdot)$  satisfies  $L'_g(t) = \alpha c^\alpha (1-t)^{\alpha-1} L_g(t)$ , an appropriate distance function is

$$D_n(t) = P'_n(t) - \alpha c^\alpha (1-t)^{\alpha-1} P_n(t).$$

By straightforward calculation we arrive from (3) at the following representation of the test statistic,

$$\begin{aligned} T_{n,\gamma} &= \frac{1}{n} \left[ \sum_{j,k=1}^n X_j X_k I(X_{jk} + \gamma - 2) + \alpha^2 c^{2\alpha} \sum_{j,k=1}^n B(X_{jk} + \gamma + 1, 2\alpha - 1) \right] \\ &- \frac{2}{n} \alpha c^\alpha \sum_{j,k=1}^n X_j B(X_{jk} + \gamma, \alpha), \end{aligned}$$

where  $B(x_1, x_2) = \int_0^1 t^{x_1-1} (1-t)^{x_2-1} dt$  is the Beta function.

### 3. CONSISTENCY

It is transparent from the expressions of  $T_{n,\gamma}$  above that in order to actually perform the test, estimates of the parameters involved are required. Write  $\hat{T}_{n,\gamma}$  for the test statistic resulting from  $T_{n,\gamma}$  when parameter are replaced by corresponding estimates. Because (2) implies that the moments of  $X$  can be analytically obtained from the moments of  $\lambda$ , a standard and straightforward choice is to employ classical moment estimation. As moment estimation is non-parametric we will assume that the estimator  $\hat{\theta}_n$  attains a limit value not only under the null hypothesis  $H_0$ , but also under alternatives, that is  $\hat{\theta}_n \rightarrow \theta$ . In fact we assume that  $\hat{\theta}_n$  (taken to be univariate for simplicity) is a regular estimator of  $\theta$  in the sense that there is an expansion of the form

$$\hat{\theta}_n = \theta + \int \psi(x; \theta) dF_n(x) + \varepsilon_n,$$

where (i)  $F_n$  denotes the empirical distribution function of  $X_j, j = 1, \dots, n$ , (ii)  $\varepsilon_n \rightarrow 0$ , almost surely, and (iii)  $\psi$  is a measurable function satisfying  $E[\psi(X; \theta)] = 0$ , and  $E[\psi^2(X; \theta)] < \infty$ . Under this assumption and since  $P'_n(t) = \int t^x dF_n(x) \rightarrow P'_X(t)$ , almost surely and uniformly in  $t$

(see Rémillard and Theodorescu, [11]), the distance function  $D_n(t) := D(P_n(t), \hat{\theta}_n; t)$  may be written as

$$D_n(t) = \int d(x, \theta; t) dF_n(x) + u_n(t), \tag{5}$$

for some measurable function  $d(x, \theta; t)$  which satisfies  $E[d(X, \theta; t)] = 0$ , identically under  $H_0$ , where  $\int u_n^2(t) t^\gamma dt \rightarrow 0$ , almost surely. We now consider the asymptotic behavior of  $\hat{T}_{n,\gamma}$  in a nonparametric, setting. The following result implies the consistency of a goodness-of-fit test based on  $\hat{T}_{n,\gamma}$  against general alternatives.

**Theorem 3.1** Under the standing assumptions,

$$\liminf_{n \rightarrow \infty} \frac{\hat{T}_{n,\gamma}}{n} \geq \int_0^\infty (E[d(X, \theta; t)])^2 t^\gamma dt \tag{6}$$

almost surely.

Proof. From (5) we have that as  $n \rightarrow \infty$ ,  $D_n(t) \rightarrow E[d(X, \theta; t)]$ , almost surely. Then by Fatou's lemma,

$$\liminf_{n \rightarrow \infty} \frac{\hat{T}_{n,\gamma}}{n} \geq \Delta := \int_0^1 (E[d(X, \theta; t)])^2 t^\gamma dt$$

almost surely, which finishes the proof of the theorem.

Assume that  $E[d(X, \theta; t)] = 0$ , holds identically only under  $H_0$ . Then  $\Delta$  is positive if  $H_0$  is not true. Consequently, a level  $\alpha$ -test that rejects  $H_0$  for large values of  $\hat{T}_{n,\gamma}$  is consistent against each alternative satisfying the conditions of Theorem 3.1.

#### 4. LIMIT STATISTICS

A user-specified choice is that of  $\gamma$  in the weight function  $t^\gamma$ . In fact, there is an underlying connection between the choice of  $\gamma$  in the test statistic (3) and moment estimation. By way of example we will consider this connection in the case of Gamma mixing. Make the substitution  $t = e^{-u}$  and use (4) in (3) to write the test statistic as

$$T_{n,\gamma} = \int_0^\infty \phi(u) e^{-(\gamma+1)u} du, \tag{7}$$

where  $\phi(u) = nD_n^2(e^{-u})$ , and

$$D_n(e^{-u}) = [1 + \vartheta(1 - e^{-u})] \frac{1}{n} \sum_{j=1}^n X_j e^{-u(X_j-1)} - \alpha\vartheta \frac{1}{n} \sum_{j=1}^n e^{-uX_j}.$$

Simple Taylor approximation of the exponential functions followed by some algebra yields,

$$D_n(e^{-u}; \alpha, \vartheta) = (\bar{X}_n - \alpha\vartheta) + \left( \vartheta \bar{X}_n + \alpha\vartheta \bar{X}_n + \bar{X}_n - \overline{X^{(2)}} \right) u + M_n(\alpha, \vartheta) u^2, \tag{8}$$

$u \rightarrow 0$ , where  $\bar{X}_n = n^{-1} \sum_{j=1}^n X_j$ ,  $\overline{X^{(2)}} = n^{-1} \sum_{j=1}^n X_j^2$ , and

$$M_n(\alpha, \vartheta) := M_n(X_1, \dots, X_n; \alpha, \vartheta) = \left( \frac{1}{2} \left[ \overline{X^{(12)}} + \vartheta \bar{X}_n - \alpha\vartheta \overline{X^{(2)}} \right] - \vartheta \overline{X^{(2)}} \right),$$

with  $\overline{X^{(12)}} = n^{-1} \sum_{j=1}^n X_j (X_j - 1)^2$ .

Let  $(\hat{\alpha}_n, \hat{\vartheta}_n)$  denote the moment estimator of  $(\alpha, \vartheta)$ . Then the first two terms in (8) vanish, giving

$$D_n^2(e^{-u}; \hat{\alpha}_n, \hat{\vartheta}_n) = M_n^2(\hat{\alpha}_n, \hat{\vartheta}_n) u^4 + o(u^4). \tag{9}$$

From (7) and (9) and an Abelian theorem for Laplace transforms (see Zayed [12], § 5.11), we conclude that

$$\lim_{\gamma \rightarrow \infty} \gamma^5 \hat{T}_{n,\gamma} = 4! n M_n^2(\hat{\alpha}_n, \hat{\vartheta}_n).$$

Hence when the test is implemented via moment estimation, the resulting test statistic (properly scaled) approaches a limit value as  $\gamma \rightarrow \infty$ , this limit value involving a complicated moment equation. Moreover since moment estimators are typically consistent, and provided that  $E(X^3) < \infty$ , we have by the law of large numbers

$$M_n(X_1, \dots, X_n; \hat{\alpha}_n, \hat{\vartheta}_n) \xrightarrow{P} M(X; \alpha, \vartheta),$$

where

$$M(X; \alpha, \vartheta) = \frac{1}{2} \left[ E\{X(X-1)^2\} + \vartheta E(X) - \alpha\vartheta E(X^2) \right] - \vartheta E(X^2),$$

and it is easy to show after some further computation of  $E(X^m)$ ,  $m = 1, 2, 3$ , that  $M(X; \alpha, \vartheta) \equiv 0$  under Gamma mixing. Thus the limit statistic, incorporating an equation involving the first three empirical moments, is asymptotically a contrast under Gamma mixing. Consequently our test statistics appear as generalizations of moment--type procedures such as those employed in [5] and [3].

#### 5. SIMULATIONS

We have run a small simulation study for the test statistic in the case of Gamma mixing with moment estimation of the parameters. The sample size is  $n = 100$  and the number of replications was set to 1,000. Since the finite--sample as well as the asymptotic distribution of the test statistic under  $H_0$  is, for all practical purposes, intractable, we employ the following parametric bootstrap procedure in order to actually carry out the test: (i) Conditionally on the observations  $x_j, j = 1, \dots, n$ , compute the estimate  $\hat{\theta}_n$  (ii) generate a bootstrap sample of size  $n$  from the postulated negative binomial distribution with parameter  $\hat{\theta}_n$ , (iii) compute the estimator  $\hat{\theta}_n^*$  from the bootstrap sample, and (iv) compute the value of the test statistic corresponding to the bootstrap sample. If steps (ii)--(iv) are repeated a number of times, say  $B$ , the bootstrap distribution of the test statistic is reproduced, and we obtain the critical value of an  $\alpha$ --size test as the  $(1-\alpha) \times 100\%$  quantile of this distribution.

Table 1 shows percentage of rejection (rounded to the nearest integer) from this bootstrap procedure with bootstrap size  $B = 200$ . The alternatives considered are the Poisson

**Table 1. Rejection Rate for 1,000 Monte Carlo Samples for  $n = 100$ ,  $\alpha = 0.10$**

$\gamma =$	2	3	5	10	2	3	5	10
$\lambda = 0.5$	60	60	60	60	92	92	92	92
$\lambda = 1$	60	60	60	60	82	81	80	79
$\lambda = 2$	60	60	59	59	78	75	74	69

distribution with parameter  $\lambda$  (left part of the table), and a fifty--fifty mixture of a Poisson distribution with parameter  $\lambda$ , and a negative binomial distribution with parameters with parameters  $\vartheta = \alpha = 1$  (right part of the table). The results show that the test has considerable power and at the same time it is not so much affected by the value of the weight parameter  $Y$ .

**REFERENCES**

[1] D. Karlis, and E. Xekalaki, "Mixed poisson distributions", *Inter. Stat. Rev.*, vol. 73, pp. 35-58, 2005.

[2] J.L. Green, and J.B. Plotkin, "A statistical theory for sampling species abundances", *Ecol. Lett.*, vol. 10, pp. 1037-1045, 2007.

[3] Y. Fang, "Semi parametric specification tests for discrete probability models", *J. Risk Insur.*, vol. 70, pp. 75-86, 2003.

[4] Y. Fang, "Semi parametric specification tests for mixing distributions", *Comput. Stat. Data Anal.*, vol. 52, pp. 2829-2839, 2008.

[5] Carriere, J, "Nonparametric tests for mixed Poisson distributions", *Insurance: Math. Econ.*, vol. 12, pp. 3-8, 1993.

[6] G.E. Willmot, "Mixed compound Poisson distributions", *ASTIN Bull.*, vol. 16, pp. S59-S79, 1986.

[7] G.E. Willmot, "The Poisson inverse gaussian distribution as an alternative to the negative binomial", *Scand. Actuar. J.*, vol. 34, pp. 113-127, 1987.

[8] B.F. Nye, and A.E. Hofflander. "Experience rating in medical professional liability insurance", *J. Risk. Insur.*, vol. 55, pp. 150-157, 1988.

[9] S.G. Meintanis, and J. Swanepoel, "Bootstrap goodness of fit tests with estimated parameters based on empirical transforms", *Stat. Probab. Lett.*, vol. 77, pp. 1003-1014, 2007.

[10] S.G. Meintanis, "Omnibus tests for strictly positive stable laws based on the empirical Laplace transform", *Math. Methods Stat.*, vol. 4, pp. 468-478, 2005.

[11] B. Rémillard, and R. Theodorescu, "Inference based on the empirical probability generating function for mixtures of Poisson distributions", *Stat. Decis.*, vol. 18, pp. 349-366, 2000.

[12] A.I. Zayed, *Handbook of Function and Generalized Function Transformations*, CRC Press: New York, 1996.

Received: August 25, 2011

Revised: September 28, 2011

Accepted: October 10, 2011

© Simos G. Meintanis; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.