GOODNESS-OF-FIT TESTS FOR COMPOUND POISSON DISTRIBUTIONS BASED ON THE MOMENT GENERATING FUNCTION

Simos Meintanis and I.A. Koutrouvelis

Department of Engineering Sciences, University of Patras, 261 10 Patras, GREECE
E-Mail: simosmei@upatras.gr

1. The Test Statistics

The random variable \( X \) is said to follow a Compound Poisson (Com-Poi) distribution when it can be expressed in the form

\[
X = \sum_{j=1}^{N} Y_j, \text{ if } N > 0, \text{ and } X = 0 \text{ if } N = 0.
\]

In (1) the random variables \( Y_j, j = 1, 2, ..., N \) are mutually independent and follow a common, so-called compounding, distribution, while \( N \) is a Poisson random variable with parameter \( \lambda \) which is distributed independently of the \( Y_j \)'s. The Com-Poi model is the most popular case of a random sum of random variables. The statistical significance of this model arises from its applicability in real life situations: There the researcher often observes only the total "amount" \( X \) which is composed of an unknown random number \( N \) of (unobserved) random "contributions" \( Y_j \). Despite of the statistical applications of the Com-Poi distribution, the only goodness-of-fit tests based on transform methods are confined to the special case of a Poisson distribution. See for example Kocherlakota and Kocherlakota (1986), Ruenda et al (1991), Baringhaus and Henze (1992). Let \( M(t) \) denote the moment generating function (mgf) of \( X \). Then for \( t \neq 0 \) we have

\[
\frac{\log M(t)}{m(t)} - \lambda,
\]

where \( m(t) = m(t; \theta) \) denotes the mgf of the compounding distribution with parameter \( \theta \). The above identity characterizes the Com-Poi family of distributions and constitutes a point of departure in constructing our test statistics. Let \( X_1, X_2, ..., X_n \) be independent observations on the random variable \( X \). For a specific value \( \theta_0 \) of \( \theta \), consider testing the null hypothesis \( H_0: X \) is Com-Poi with compounding mgf \( m_0(t) = m(t; \theta_0) \).

In doing so, we employ the empirical mgf

\[
M_n(t) = \frac{1}{n} \sum_{i=1}^{n} \exp(t X_i).
\]

In

\[
\delta_n = \sqrt{n} \left[ \log M_n(u) \left( \frac{m_0(u) - 1}{m_0(u) - 1} - \frac{m_0(v) - 1}{m_0(v) - 1} \right) \right], \quad u \neq v,
\]

where \( u, v \neq 0 \) are real numbers. Then under \( H_0 \) the asymptotic distribution of \( \delta_n = \delta_n(u,v) \) is a zero-mean normal with variance \( \gamma^2 = \gamma^2(u,v) \) given by

\[
\gamma^2 = \frac{1}{(m_0(u) - 1)^2} \sigma(u,u) + \frac{1}{(m_0(v) - 1)^2} \sigma(v,v) - \frac{2}{(m_0(u) - 1)(m_0(v) - 1)} \sigma(u,v).
\]

In (5), \( \sigma(\cdot,\cdot) \) denotes the asymptotic covariance structure of \( \log M_n(\cdot) \). Hence an asymptotic test of the null hypothesis can be constructed based on the chi-squared distribution with one degree
of freedom and the value of \((\delta_n / \gamma)^2\). The test statistic depends on the Poisson parameter \(\lambda\) only through \(\gamma\), but one can use a consistent estimate of \(\lambda\) which leads to equivalent large sample chi-squared statistics. The test easily generalizes to multiple pairs of points \((u_j, v_j)\), where \(u_j \neq v_j\), with \(u_j, v_j \neq 0\). We form the vector \(\Delta_n\) with \(j^\text{th}\) element \(\delta_n(u_j, v_j), j = 1, 2, \ldots, r\). Then \(\Delta_n\) is asymptotically normally distributed with zero mean and covariance matrix \(\Gamma\). Hence
\[
\Delta_n \Gamma^{-1} \Delta_n \xrightarrow{d} \chi_r^2,
\]
and an asymptotic chi-squared test of fit with \(r\) degrees of freedom can be performed.

2. SIMULATION RESULTS
In order to study the behavior of the test statistics in finite samples, we have performed simulation experiments with sample size \(n=100\), in each case with 1000 replications. Random numbers were generated from several \(\text{Com-Poi}\) models and in Table 1 (with obvious notations) level results are reported at 5% nominal value. The user-specified parameter values \(r\) and \((u_j, v_j), j = 1, 2, \ldots, r\) are selected to ensure close agreement between empirical and nominal size.

However results to be reported elsewhere, prove that this choice of parameters leads to a quite powerful test. From the figures in Table 1 we can see that the empirical level of the test is fairly close to its nominal value, lying in the great majority of cases in the interval 5%\pm1%. The test also exhibits considerable robustness in level over the parameter space.

TABLE 1: Empirical level at 5% nominal value in 1000 trials with \(n=100\)

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter(s)</th>
<th>(\lambda = 1)</th>
<th>(\lambda = 2)</th>
<th>(\lambda = 5)</th>
<th>(\lambda = 10)</th>
<th>(\lambda = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>(\mu, \sigma^2) = 0,1</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Poisson-Geometric</td>
<td>(p=0.50)</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>Poisson-Binomial</td>
<td>(v, p = 5,0.2)</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Poisson-Exponential</td>
<td>(\beta = 1)</td>
<td>0.05</td>
<td>0.06</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>Poisson-Normal</td>
<td>(\mu, \sigma^2) = 2,1</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(\mu, \sigma^2) = 5,1</td>
<td>0.05</td>
<td>0.06</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

REFERENCES


RESUMÉ

Nous proposons des tests d’adéquation pour des lois de Poisson-mélangées qui utilisent la fonction génératrice des moments. La statistique du test est prouvée de suivre, asymptotiquement, une loi de chi-carré. À l’aide d’une étude de Monte Carlo la proximité de cette loi à sa contrepartie d’échantillon fini est étudiée.