

University of Wollongong

Research Online

Applied Statistics Education and Research
Collaboration (ASEARC) - Conference Papers

Faculty of Engineering and Information
Sciences

2011

Smooth Tests of Fit for a Mixture of Two Poisson Distributions

John Best
University of Newcastle

John Rayner
University of Newcastle

Olivier Thas
Ghent University

Follow this and additional works at: <https://ro.uow.edu.au/asearc>

Recommended Citation

Best, John; Rayner, John; and Thas, Olivier, "Smooth Tests of Fit for a Mixture of Two Poisson Distributions" (2011). *Applied Statistics Education and Research Collaboration (ASEARC) - Conference Papers*. 20.

<https://ro.uow.edu.au/asearc/20>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Smooth Tests of Fit for a Mixture of Two Poisson Distributions

Abstract

In this note smooth tests of fit for a mixture of two Poisson distributions are derived and compared with a traditional Pearson chi-squared test. The tests are illustrated with a classic data set of deaths per day of women over 80 as recorded in the London Times for the years 1910 to 1912.

Keywords

Central moments, Deaths of London women data, Factorial moments, Orthonormal polynomials, Parametric bootstrap, Pearson X2 tests

Publication Details

Best, John; Rayner, John; and Thas, Olivier, Smooth Tests of Fit for a Mixture of Two Poisson Distributions, " Proceedings of the Fourth Annual ASEARC Conference, 17-18 February 2011, University of Western Sydney, Paramatta, Australia.

Smooth Tests of Fit for a Mixture of Two Poisson Distributions

D.J. Best, J.C.W. Rayner
The University of Newcastle, Callaghan, NSW, 2308, Australia
John.Best@newcastle.edu.au and John.Rayner@newcastle.edu.au

and O.Thas
Department of Applied Mathematics, Biometrics and Process Control, B-9000 Gent, Belgium
Olivier.Thas@Ugent.be

Abstract

In this note smooth tests of fit for a mixture of two Poisson distributions are derived and compared with a traditional Pearson chi-squared test. The tests are illustrated with a classic data set of deaths per day of women over 80 as recorded in the London Times for the years 1910 to 1912.

Keywords: Central moments, Deaths of London women data, Factorial moments, Orthonormal polynomials, Parametric bootstrap, Pearson X^2 tests

1. Introduction

A Poisson process is often used to model count data. Sometimes an underlying mechanism suggests two Poisson processes may be involved. This may be modelled by a two component Poisson mixture model. We will give some examples later. The Poisson probability function, $f(x; \theta)$ say, is given by

$$f(x; \theta) = \exp(-\theta)\theta^x/x!, \quad x = 0, 1, 2, \dots, \\ \text{in which } \theta > 0$$

and the two component Poisson mixture model has probability function

$$f^*(x; \theta_1, \theta_2, p) = p f(x; \theta_1) + (1-p) f(x; \theta_2), \\ x = 0, 1, 2, \dots, \text{ in which } \theta_1 > 0, \theta_2 > 0, \\ \theta_1 \neq \theta_2 \text{ and } 0 < p < 1.$$

A common test of fit for $f^*(x; \theta_1, \theta_2, p)$ is based on the well-known Pearson's X^2 . If there are l classes X^2 is approximately χ^2 with $l - 4$ degrees of freedom: χ_{l-4}^2 .

In section 2 we look at estimation of the parameters θ_1 , θ_2 and p . Section 2 also defines the X^2 test and some smooth tests of fit. Section 3 gives a small power comparison while section 4 considers a classic data set of deaths per day of women over 80 as recorded in the London Times for the years 1910 to 1912.

2. Estimation and Test Statistics

The two most common approaches for estimating θ_1 , θ_2 and p are based on moments (MOM) and maximum likelihood (ML). If we have n data points x_1, x_2, \dots, x_n and $\bar{x} = \sum_{i=1}^n x_i/n$ and $m_t = \sum_{i=1}^n (x_i - \bar{x})^t/n$, $t = 2, 3, \dots$ the MOM estimators satisfy

$$\tilde{p} = (\bar{x} - \tilde{\theta}_2)/(\tilde{\theta}_1 - \tilde{\theta}_2), \quad \tilde{\theta}_1 = (A - D)/2, \\ \text{and } \tilde{\theta}_2 = (A + D)/2$$

in which

$$A = 2\bar{x} + (m_3 - 3m_2 + 2\bar{x})/(m_2 - \bar{x})$$

$$\text{and } D^2 = A^2 - 4A\bar{x} + 4(m_2 + \bar{x}^2 - \bar{x}).$$

This method clearly fails if $D^2 < 0$, if any of $\tilde{\theta}_1$, $\tilde{\theta}_2$ and \tilde{p} are outside their specified bounds, or if $m_2 = \bar{x}$.

Iteration is needed to find the ML estimates and given the speed of modern computers an EM type algorithm is satisfactory. This will always converge to $\hat{\theta}_1$, $\hat{\theta}_2$ and \hat{p} within the specified bounds if the initial estimates are also within these bounds. However convergence can be slow – occasionally more than 1,000 iterations – and a local, but not universal, maximum may be found. A grid of initial values is often worth examining. This was not done for the calculations in Table 1 because all the sizes were 0.05 suggesting universal maxima were indeed found. To check on the possibility of a local stationary point it is also useful to examine contour plots of the likelihood surface. This was done for the *Deaths of London Women* example in section 4. The following estimation equations are needed:

$$\hat{p}_k = \frac{\hat{p}_{k-1} \sum_{i=1}^n f(x_i, \hat{\theta}_{1,k-1})}{nf^*(x_i; \hat{\theta}_{1,k-1}, \hat{\theta}_{2,k-1}, \hat{p}_{k-1})} \text{ and}$$

$$\hat{\theta}_{r,k} = \frac{\sum_{i=1}^n x_i f(x_i, \hat{\theta}_{r,k-1})}{nf^*(x_i; \hat{\theta}_{1,k-1}, \hat{\theta}_{2,k-1}, \hat{p}_{k-1})}, r = 1, 2,$$

where $\hat{\theta}_{r,k}$ is the estimate of θ_r at the k th iteration, \hat{p}_k is the estimate of p at the k th iteration and $(\hat{\theta}_{1,0}, \hat{\theta}_{2,0}, \hat{p}_0) = (\tilde{\theta}_1, \tilde{\theta}_2, \tilde{p})$ may be an admissible initial value. See, for example, Everitt and Hand (1981, p.97). Newton's method will sometimes converge to the correct values and when it does the convergence is much quicker than the above estimating equations. However, Newton's method doesn't always converge and may give estimates outside the specified bounds.

Now suppose that for the n data points O_j values are equal to j , $j = 0, 1, 2, \dots$. Let $E_j = nf^*(j; \hat{\theta}_1, \hat{\theta}_2, \hat{p})$. Often classes are pooled in the tail until the greatest l is found such that the expectation of the classes from the l th on is at least 5. Then the Pearson test of fit statistic is

$$X^2 = \sum_{j=1}^l (O_j - E_j)^2 / E_j$$

and X^2 is taken to have the χ_{l-4}^2 distribution.

Smooth test components V_s can be defined as

$$\hat{V}_s = \sum_{i=1}^n g_s(x_i; \hat{\theta}_1, \hat{\theta}_2, \hat{p}) / \sqrt{n}, s = 2, 3, \dots$$

Here $\{g_s(\cdot)\}$ is the set of orthonormal functions on the null distribution. We give formulae, in terms of the population moments μ, μ_2, \dots, μ_6 for the first four orthonormal functions and \hat{V}_2 and \hat{V}_3 in Appendix A. For the mixture of two Poissons these moments can be calculated from the population factorial moments $\mu_{[r]} = p\theta_1^r + (1-p)\theta_2^r$. Smooth tests of fit are discussed in detail in Rayner et al. (2009).

Table 1. 100×powers based on 10,000 Monte Carlo samples for $n = 100$ and $\alpha = 0.05$ for a null Poisson mixture with $p = 0.5$, $\theta_1 = 2$ and $\theta_2 = 5$.

Alternative	\hat{V}_2^2	\hat{V}_3^2	\hat{V}_4^2	X^2
Null	5	5	5	5
NB(2, 0.4)	45	39	40	41
NB(3, 0.5)	18	20	20	18
NB(4, 0.5)	19	20	24	27
NTA(1, 2)	79	69	51	54
0.5 × NB(2, 0.4) + 0.5 × NB(2, 0.5)	33	28	30	31
0.5 × NB(2, 0.3) + 0.5 × NB(3, 0.5)	64	48	59	65
NTA(2, 2)	88	66	55	81
NTA(2, 1)	26	26	22	16
NTA(1, 3)	98	94	72	92
P(4)	37	14	13	4
P(6)	33	13	5	10

3. Indicative Size and Power Study

We consider the case $\alpha = 0.05$, $p = 0.5$, $\theta_1 = 2$, $\theta_2 = 5$. Based on 25,000 Monte Carlo samples the critical values of \hat{V}_2^2 , \hat{V}_3^2 and \hat{V}_4^2 are 0.31, 0.91 and 0.56 respectively. We note that $\hat{V}_1 \equiv 0$ as shown in Appendix B. We use 17.5 as the X^2 critical value. Table 1 gives some powers for

- negative binomial alternatives with probability function $\binom{m+x-1}{x} \pi^m (1-\pi)^x$ for $x = 0, 1, 2, \dots$ with $m > 0$, denoted by NB(m, π),

- Neyman Type A alternatives with probability function $\frac{e^{-\lambda_1} \lambda_2^x}{x!} \sum_{j=0}^{\infty} \frac{j^x}{j!} (\lambda_1 e^{-\lambda_2})^j$ for $x = 0, 1, 2, \dots$ with $\lambda_1 > 0$ and $\lambda_2 > 0$, denoted by $\text{NTA}(\lambda_1, \lambda_2)$ and
- Poisson alternatives $f(x; \theta)$ for $x = 0, 1, 2, \dots$ with $\theta > 0$, denoted by $\text{P}(\theta)$.

In Table 1 no one test dominates but overall perhaps that based on \hat{V}_2^2 does best. Double precision arithmetic was used in the Table 1 calculations. In a few cases no estimate was obtained after 10,000 iterations and these cases were discarded.

4. Example: Deaths of London Women During 1910 to 1912

A classic data set considered by a number of authors starting with Whitaker (1914) considers deaths per day of women over 80 in London during the years 1910, 1911 and 1912 as recorded in the Times newspaper. Table 2 shows the data and expected counts for $(\hat{\theta}_1, \hat{\theta}_2, \hat{p}) = (1.254, 2.661, 0.358)$. Using ten classes $X^2 = 1.29$ with six degrees of freedom and χ^2 p-value 0.65. Also $\hat{V}_2^2 = (-0.077)^2$, $\hat{V}_3^2 = (-0.314)^2$ and $\hat{V}_4^2 = (-0.429)^2$, with bootstrap p-values 0.70, 0.46 and 0.55 respectively. Possibly due to different death rates in summer and winter, all tests indicate a good fit by a Poisson mixture. If a single Poisson is used to describe the data then $X^2 = 27.01$ with eight degrees of freedom and a χ^2 p-value is 0.001.

Table 2. Deaths per day of London women over 80 during 1910 to 1912

# deaths	0	1	2	3	4
Count	162	267	271	185	111
Mixture expected	161	271	262	191	114
Poisson expected	127	273	295	212	114

# deaths	5	6	7	8	9
Count	61	27	8	3	1
Mixture expected	58	25	9	3	1
Poisson expected	49	18	5	1	0

A plot of likelihood contours indicated the likelihood has a maximum at $(\hat{\theta}_1, \hat{\theta}_2)$ and that there are no other stationary points nearby. As $\hat{V}_1 \equiv 0$ we can give \hat{p} in terms of $\hat{\theta}_1$ and $\hat{\theta}_2$ and so \hat{p} does not need to be included in any likelihood contour plot.

References

- [1] Everitt, B.S. and Hand, D.J.C. (1981). *Finite Mixture Distributions*. London: Chapman and Hall.
- [2] Rayner, J.C.W., Thas, O. and Best, D.J. (2009). *Smooth Tests of Goodness of Fit: Using R* (2nd ed.). Singapore: Wiley.
- [3] Whitaker, L. (1914). On the Poisson law of small numbers. *Biometrika*, 10(1), 36-71.

Appendix A: Orthonormal Polynomials for a Poisson Mixture

Let μ be the mean and μ_t for $t = 2, 3, \dots$ the central moments, assumed to exist, of some distribution of interest. Then the first four orthonormal polynomials are, for $x = 0, 1, 2, \dots$

$$g_0(x) = 1, g_1(x) = (x - \mu)/\sqrt{\mu_2},$$

$$g_2(x) = \{(x - \mu)^2 - \mu_2(x - \mu)/\mu_2 - \mu_2\}/\sqrt{d}$$

and $g_3(x) = \{(x - \mu)^3 - a(x - \mu)^2 - b(x - \mu) - c\}/\sqrt{e}$

where

$$d = \mu_4 - \mu_3^2 / \mu_2 - \mu_2^2 \text{ and } e = \mu_6 - 2a\mu_5$$

$$+ (a^2 - 2b)\mu_4 + 2(ab - c)\mu_3 + (b^2 + 2ac)\mu_2 + c^2$$

in which

$$a = (\mu_5 - \mu_3\mu_4 / \mu_2 - \mu_2\mu_3) / d$$

$$b = (\mu_4^2 / \mu_2 - \mu_2\mu_4 - \mu_3\mu_5 / \mu_2 + \mu_3^2) / d$$

$$c = (2\mu_3\mu_4 - \mu_3^3 / \mu_2 - \mu_2\mu_5) / d.$$

Again assuming they exist, for $t = 2, 3, \dots$ write $\mu'_{[t]}$ for the t th factorial moment. It now follows routinely that

$$\mu_2 = \mu'_{[2]} + \mu - \mu^2$$

$$\mu_3 = \mu'_{[3]} + 3\mu'_{[2]} + \mu - 3\mu(\mu'_{[2]} + \mu) \mu_2 + 2\mu^3$$

$$\mu_4 = \mu'_{[4]} + 6\mu'_{[3]} + 7\mu'_{[2]} + \mu - 4\mu(\mu'_{[3]} + 3\mu'_{[2]} + \mu) + 6\mu^2(\mu'_{[2]} + \mu) - 3\mu^4$$

$$\mu_5 = \mu'_{[5]} + 10\mu'_{[4]} + 25\mu'_{[3]} + 15\mu'_{[2]} + \mu$$

$$- 5\mu(\mu'_{[4]} + 6\mu'_{[3]} + 7\mu'_{[2]} + \mu)$$

$$\begin{aligned}
& + 10\mu^2(\mu'_{[3]} + 3\mu'_{[2]} + \mu) - 10\mu^3(\mu'_{[2]} + \mu) + 4\mu^5 \\
\mu_6 = & \mu'_{[6]} + 15\mu'_{[5]} + 65\mu'_{[4]} + 90\mu'_{[3]} + 31\mu'_{[2]} + \mu - \\
& 6\mu(\mu'_{[5]} + 10\mu'_{[4]} + 25\mu'_{[3]} + 15\mu'_{[2]} + \mu) + 15\mu^2(\\
& \mu'_{[4]} + 6\mu'_{[3]} + 7\mu'_{[2]} + \mu) - 20\mu^3(\mu'_{[3]} + 3\mu'_{[2]} + \mu) \\
& + 15\mu^4(\mu'_{[2]} + \mu) - 5\mu^6.
\end{aligned}$$

For a Poisson mixture the t th factorial moment is $\mu'_{[t]} = p\theta_1^t + (1-p)\theta_2^t$ so that, for example, $\mu = p\theta_1 + (1-p)\theta_2$. Using the ML estimators $\hat{\theta}_1$, $\hat{\theta}_2$ and \hat{p} and the above formulae for μ , ..., μ_6 we can calculate \hat{V}_2 and \hat{V}_3 where $\hat{V}_s = \sum_{i=1}^n g_s(x_i; \hat{\theta}_1, \hat{\theta}_2, \hat{p}) / \sqrt{n}$, $s = 2, 3$.

Appendix B: Proof That $\hat{V}_1 \equiv 0$

Given $g_1(x)$ from Appendix A above, the first smooth component $\sum_{i=1}^n g_1(x_i; \hat{\theta}_1, \hat{\theta}_2, \hat{p}) / \sqrt{n}$ is proportional to $\bar{X} - \hat{\mu}$, where $\hat{\mu} = \hat{p}\hat{\theta}_1 + (1-\hat{p})\hat{\theta}_2$ is the ML estimator of $\mu = E[X]$. For notational convenience arguments involving θ_1 , θ_2 and p are henceforth suppressed. To obtain the ML estimators of θ_1 , θ_2 and p note that the likelihood is $L = \prod_{i=1}^n f^*(x_i; \theta_1, \theta_2, p)$. Taking logarithms and differentiating gives

$$\begin{aligned}
\frac{\partial \log L}{\partial \theta_1} &= \sum_{i=1}^n \{pf_1(x_i)[-1 + \frac{x_i}{\theta_1}]\} / f^*(x_i), \\
\frac{\partial \log L}{\partial \theta_2} &= \sum_{i=1}^n \{pf_2(x_i)[-1 + \frac{x_i}{\theta_2}]\} / f^*(x_i) \\
&\text{and} \\
\frac{\partial \log L}{\partial p} &= \sum_{i=1}^n \{f_1(x_i) - f_2(x_i) / f^*(x_i)\}.
\end{aligned}$$

From $\partial \log L / \partial \theta_r = 0$ for $r = 1$ and 2 we obtain

$$\hat{\theta}_r = \frac{\sum_{i=1}^n x_i f_r(x_i) / f^*(x_i)}{\sum_{i=1}^n f_r(x_i) / f^*(x_i)}$$

and from $\partial \log L / \partial p = 0$ we obtain

$$\sum_{i=1}^n f_1(x_i) / f^*(x_i) = \sum_{i=1}^n f_2(x_i) / f^*(x_i).$$

Using $f^*(x) = pf_1(x) + (1-p)f_2(x)$ and the equation immediately above shows that $\sum_{i=1}^n f_r(x_i) / f^*(x_i) = n$ for $r = 1$ and 2 . It now follows that

$$\begin{aligned}
\hat{\theta}_r &= \sum_{i=1}^n x_i f_r(x_i) / \{n f^*(x_i)\} \text{ and} \\
\hat{\mu} &= \hat{p}\hat{\theta}_1 + (1-\hat{p})\hat{\theta}_2 = \\
&\hat{p} \sum_{i=1}^n x_i f_1(x_i) / f^*(x_i) / n + \\
&(1-\hat{p}) \sum_{i=1}^n x_i f_2(x_i) / f^*(x_i) / n = \\
&\sum_{i=1}^n x_i \{ \hat{p}f_1(x_i) + (1-\hat{p})f_2(x_i) \} / \{n f^*(x_i)\} = \\
&\sum_{i=1}^n x_i / n = \bar{x}.
\end{aligned}$$

It thus follows that $\hat{V}_1 \equiv 0$.