A Note on a Test for Poisson Overdispersion

Dankmar Bohning

Biometrika, Volume 81, Issue 2 (Jun., 1994), 418-419.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28199406%2981%3A2%3C418%3AANOATF%3E2.0.CO%3B2-%2

Biometrika (1994), **81**, 2, *pp.* 418–9
Printed in Great Britain

A note on a test for Poisson overdispersion

By DANKMAR BÖHNING

Department of Epidemiology, Free University Berlin, Augustastrasse 37, 12203 Berlin, Germany

Summary

This note discusses an error occurring in a test for Poisson overdispersion suggested by Tiago de Oliveira (1965). The limiting null distribution of the suggested statistic is neither pivotal nor is it standard normal. The error lies in the computation of the asymptotic standard error of the overdispersion estimate, for which a corrected version is given. The corrected version of the test statistic becomes equivalent to the normalized version of Fisher's index of dispersion.

Some key words: Extra-Poisson variation; Fisher's dispersion index; Overdispersion test.

1. Introduction

Consider a random sample of counts of sample size $X_1, \ldots, X_n$. Let $\bar{X} = (X_1 + \ldots + X_n)/n$ be its mean and $S^2 = \{(X_1 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2\}/(n-1)$ its variance. If interest is in the hypothesis H_0 that the sample comes from a Poisson distribution with parameter λ, one naturally compares $\bar{X}$ with S^2 since, under H_0, $E(\bar{X}) = E(S^2) = \lambda$. If S^2 is much larger than $\bar{X}$, then we have found Poisson overdispersion, and $(S^2 - \bar{X})$ is its estimate. Overdispersion is a phenomenon often caused by latent heterogeneity, meaning that the sample arises from a population consisting of different subpopulations. A simple diagnostic test for overdispersion is helpful, since a lack of significance in testing overdispersion might indicate that a further investigation of latent heterogeneity might not be necessary. The test to be discussed here is based on a suggestion of Tiago de Oliveira (1965) who looked at the difference $S^2 - \bar{X}$ and argued that its variance is given by $(1 - 2\lambda^{\frac{1}{2}} + 3\lambda)/n$, which can be estimated by $(1 - 2\bar{X}^{\frac{1}{2}} + 3\bar{X})/n$ if the null hypothesis is true. The statistic of Tiago de Oliveira (1965) is thus

$$O_T = n^{\frac{1}{2}}(S^2 - \bar{X})/(1 - 2\bar{X}^{\frac{1}{2}} + 3\bar{X})^{\frac{1}{2}} \qquad (1{\cdot}1)$$

and it is claimed that the limiting null-distribution of (1·1) is standard normal. The test is also referred to by Titterington, Smith & Makov (1985, p. 152) and by Johnson, Kotz & Kemp (1992, p. 319).

The failure of the test. In a simulation study of the null distribution of (1·1) with sample size $n = 1000$, λ ranging from 1 to 25, with step size 1, and replication size 10 000, it became evident that the limiting distribution of (1·1) is neither standard normal under the null hypothesis, nor is it independent of λ.

2. The variance of $S^2 - \bar{X}$

The failure of the test is due to a false computation of the standard deviation of $S^2 - \bar{X}$. This becomes apparent by observing that $n \times \text{var}(S^2 - \bar{X})$ should go to 0 if λ becomes small which is evidently not the case for $(1 - 2\lambda^{\frac{1}{2}} + 3\lambda)$. The correct variance is provided by the following theorem.

Theorem. *Let $X_1, \ldots, X_n$ be a sample from a Poisson distribution with parameter λ. Then*

$$\text{var}(S^2 - \bar{X}) = 2\lambda^2/(n-1). \qquad (2{\cdot}1)$$

The proof is straightforward, requiring only the first four moments of a Poisson variate, e.g. Haight (1967, p. 6), and some algebra. Details of the proof and the simulation study mentioned above are available from the author.

On replacing the variance of $S^2 - \bar{X}$ by its corrected version (2·1), and estimating λ by $\bar{X}$, we obtain the test statistic

$$O_T^{\text{new}} = \{(n-1)/2\}^{\frac{1}{2}}\{(S^2/\bar{X}) - 1\},$$

essentially equivalent to $(n-1)S^2/\bar{X}$, usually referred to as Fisher's index of dispersion (Potthof & Whittinghill, 1966, p. 185). Hoel (1943) argues that the χ^2 with $n-1$ degrees of freedom gives a good approximation for this.

3. Example

The data presented in Table 1 are the daily numbers of deaths of women, with brain vessel disease (International Classification of Diseases 430–438) as cause of death, for the year 1989 in West Berlin. We calculate $n = 366$, $\bar{x} = 6{\cdot}3634$, $s^2 = 6{\cdot}8238$, indicating a slight overdispersion of $s^2 - \bar{x} = 0{\cdot}4604$, and yielding values $o_T = 2{\cdot}2706$ and $o_T^{\text{new}} = 0{\cdot}9774$, leading to different conclusions.

Table 1. *Cases of female deaths by brain vessel disease in West Berlin,* 1989

Deaths per day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Frequency	1	4	15	31	39	55	54	49	47	31	16	9	8	4	3

Acknowledgement

This research is supported by the German Research Foundation. I am grateful to the refereee for valuable suggestions leading to improved results.

References

Haight, F. A. (1967). *Handbook of the Poisson Distribution.* New York: John Wiley.

Hoel, P. G. (1943). On indices of dispersion. *Ann. Math. Statist.* **14**, 155–62.

Johnson, N. L., Kotz, S. & Kemp, A. W. (1992). *Univariate Discrete Distributions,* 2nd ed. Chichester: John Wiley.

Potthoff, R. F. & Whittinghill, M. (1966). Testing for homogeneity. II. The Poisson distribution. *Biometrika* **53**, 183–7.

Tiago de Oliveria, J. (1965). Some elementary tests of mixtures of discrete distributions. In *Classical and Contagious Discrete Distributions,* Ed. G. P. Patil, pp. 379–84. New York: Pergamon.

Titterington, D. M. Smith, A. F. M. & Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Chichester: John Wiley.

[*Received July* 1993. *Revised January* 1994]