Cumhuriyet Üniversitesi Fen Fakültesi
Fen Bilimleri Dergisi (CFD), Cilt:36, No: 3 Özel Sayı (2015)
ISSN: 1300-1949

Cumhuriyet University Faculty of Science
Science Journal (CSJ), Vol. 36, No: 3 Special Issue (2015)
ISSN: 1300-1949

A comparison of homogeneity tests in Poisson distribution based on Mont Carlo simulation

Tahereh BAHRAMI[1,*], Somayyeh CHASHIANI[1]

[1]*Khatam-al Anbia Behbahan University of Technology*

Received: 01.02.2015; Accepted: 05.05.2015

Abstract. In this article, we introduce tests for examining the hypothesis of whether the observations in a Poisson distribution with the same parameters fit with the observations of Poisson distribution with different parameters. When Poisson parameter is small, and the sample volume is large, homogeneity tests like Conditional Chi-square test, Likelihood ratio, and Nymen-Scott test are not efficient enough. Therefore, in this article, another test which is efficient enough in these conditioned is introduced, named Anscombe test. At last, based on Mont Carlo simulation we compare these tests in terms of performance and accuracy, and we illustrate use of them through a real example.

Keywords: homogeneity test, likelihood ratio test, chi-square test, Anscombe transformation, Poisson distribution

1. INTRODUCTION

Distribution of several special events with low occurrence probability exists in dependent Bernoli tests. Distributions related to number of fatal traffic accidents in a week, number of particles radiated from the sun, number of calls received in an emergency police station in a certain period of time, number of space rocks colliding into a pilot satellite, number of radioactive particle rays, particle dispersion, and other instances are examples of Poisson distribution. Different tests exist in this regard through which we can examine whether a sample of observations with Poisson distribution can fit with equal occurrence rate. Since the occurrence rate of Poisson random variable depends on the place and time span, a random sample of Poisson random variable can have different Poisson distribution with different rates because it was observed in different spans. Therefore, homogeneity tests are very important in Poisson distribution. In this article, following Brown and Jao (2005) , four tests of “Conditional Chi-square”, “Likelihood ratio”, “Nymmen-Scott test” and “Anscomb Transformation test” are introduced. These tests have rather good efficiencies when \ is low /. Only if the / is low / a test named after Anscombe (1948) and Bartlett (1947), called Anscombe Transformation test, is used.

The article structure is as follows: in the second part, the aforementioned tests for Anscombe distribution are described. In the third part, the mentioned tests are compared using Mont Carlo simulation. Also, the method of using these tests is illustrated through a real example.

2. HOMOGENEITY TESTS IN POISSON DISTRIBUTION

Suppose /are independent random variables, as x_i , $i = 1,2, \dots, n$ / we want to test $H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_n$

*Corresponding author. *Email address: Bahrami@bkatu.ac,ir*

Special Issue: The Second National Conference on Applied Research in Science and Technology

http://dergi.cumhuriyet.edu.tr/cumuscij ©2015 Faculty of Science, Cumhuriyet University

v.s

$H_1 : x_i \sim P(\lambda_i) \ \left(\sum_{i=1}^{n}(\lambda_i - \bar{\lambda})^2 > 0\right)$ (1)

We deal with a preliminary examination of different tests to answer this question.

1.1. Likelihood Ratio tests

Likelihood ratio tests is conducted based on statistic $\lambda(x) = \frac{Sup\ L(\theta;K)_{\theta\in\Theta_0}}{Sup\ L(\theta;k)_{\theta\in\Theta}}$ in which Θ is parameter space and Θ_0 is parameter space under zero hypothesis. An oft-used feature of this test is limit distribution of Chi-square of $\lambda(x)$ logarithm. Suppose $x_1, \dots, x_n$ is a random sample of family distribution with test density of supposition $\{f_\phi : \phi \in \Theta \subseteq R_K\}$, which verifies in favorable conditions, and suppose hypothesis test of $H_0: \phi_i = \phi_{i_0}\ , i = 1,2, \dots, r\ , r \leq k$ against $H_1: \phi_i \neq \phi_{i_0}$ for at least one r<k and i=1,2,…,r . Under zero hypothesis, when the volume of the sample is big, statistic $-2Ln\,\lambda(x)$ has approximate distribution of χ^2 with $r = \dim\Theta - \dim\Theta_0$ degree of freedom (Mood at all, 1997). Because under hypothesis $H_0 : \lambda_1 = \cdots = \lambda_N$ maximum likelihood is estimated $\hat{\lambda}_i = x_i\ , i = 1,2, \dots, n$ and under general parameter space $\hat{\lambda}_i = x_i\,, i = 1,2, \dots, n$, so $Sup\ L(\lambda_1, \dots, \lambda_N) = \frac{e^{-n\bar{x}}(\bar{x})}{\prod_{i=1}^{n}(x_i)!}$, $Sup\ l(\lambda_1, \dots, \lambda_n) = \frac{e(\prod x_i)^{x_i}}{\prod x_i!}$ and as a result

$$T_{LR} = -2\ Ln\ \lambda(x) = 2\sum x_i\ Ln\ (\frac{x_i}{\bar{x}}) \quad (2)$$

Also, when $n \to \infty$ we have $T_{LR} \sim \chi^2_{n-1}$. Therefore H_0 hypothesis is rejected when

$T_{LR} > \chi^2_{\alpha,(n-1)}$.

2.2. Conditional Chi-square Test

Because under hypothesis H_0 . $\hat{\lambda}_i = \bar{x}$ and under general state $\lambda_i = x_i$, so, the relative difference between these statistics can be a criterion for testing hypothesis H_0 . The famous Chi-square is as follows:

$$T_{cc} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\bar{x}} = \frac{(n-1)}{\bar{x}} S^2 \quad (3)$$

Which can be statistic for testing H_0 hypothesis. Because under zero hypothesis, the $x_1, \dots, x_n$ distribution under the condition $\bar{x}$ is the polynomial $\left(n\bar{x}, \frac{1}{n}, \dots, \frac{1}{n}\right)$; therefore, under zero hypothesis the statistic T_{cc} has a asymptotic distribution χ^2_{n-1} . Hypothesis H_0 is rejected when $T_{cc} > \chi^2_{\alpha,(n-1)}$. This test is also known as 'Poisson Dispersion Test' or 'Variance Test'.

2.3. Nymen-Scott test

Nymen-Scott statistic is another statistic which is constructed based on relative difference between x_i and $\bar{x}$ as follows:

$$T_{Ns} = \sqrt{\frac{(n-1)}{2}}(\frac{S^2}{\bar{x}} - 1) \tag{4}$$

Nymen and Scott showed that under hypothesis H_0 , ${T_{Ns}}^2$ has a normal asymptotic distribution . Therefore, H_0 is rejected if $T_{cc} > \chi^2_{1-\alpha;n-1}$.

2.4. Anscomb Transformation Test

In 1948, Anscomb, based on a permanent transformation of second degree for random variables having Poisson distribution, proposed a test for homogeneity hypothesis in Poisson distribution. Suppose $x_1, \dots, x_n$ are independent random variables, as $x_i \sim P(\lambda_i)$ i=1, 2… n; set $Y_i = \sqrt{x_i + \frac{3}{8}}$ then under hypothesis $H_0: \lambda_1 = \dots = \lambda_n$ we have:

$$Var(Y_i) = \frac{1}{4} + O(\frac{1}{\lambda}) \tag{5}$$

In which $\lambda = \lambda_1 = \dots = \lambda_n$. So, if $\lambda \longrightarrow 0$ we have

$$2(Y_i - \upsilon(\lambda_i)) \longrightarrow N(0,1) \tag{6}$$

Therefore, under H_0 hypothesis, we have:

$$T_A = 4\sum(Y_i - \bar{Y})^2 \longrightarrow {\chi^2}_{n-1} \tag{7}$$

Also, under H_1 hypothesis, the T_A statistic inclines to non-central ${\chi^2}_{n-1}$ distribution, whose non-central parameter is $\delta = 4\sum(\upsilon(\lambda_i)_{\bar{\upsilon}})^2$ with $\bar{\upsilon}_n = \frac{1}{n}\sum_{i=1}^{n}\upsilon(\lambda_i)$. We reject H_0 hypothesis if $T_A > \chi^2_{\alpha,n-1}$. Brown et al (2001) suggest that we use transformation $\sqrt{x_i + \frac{1}{4}}$ instead of $\sqrt{x_i + \frac{3}{8}}$ because we have:

$$E_\lambda\left(\sqrt{x_i + \frac{3}{4}}\right) = \sqrt{\lambda} + O\left(\frac{1}{\lambda}\right) \tag{8}$$

3. RESULTS OF SIMULATION

In this part, using Mont Carlo simulation, we compare the aforementioned tests. To do so, at Type I level of error $\alpha = .05$ and by means of repeating sampling, we empirically calculate power amounts of the mentioned tests by statistics$T_{LR}, T_{cc}, T_{Ns}, T_A$, which is summarized in the table below.

In this part, empirical results under zero hypothesis resulted from$T_A, T_{NS}, T_{CC}, T_{LR}$ statistics are presented. The results of calculated Type I empirical error with 100000 repetition for different sample sizes($n = 5,12,20$), is summarized in table below:

Table 1. Tape I error.

n	λ	statistic	α = 0.1 ESE=0.003	α = 0.05 ESE=0.002	α = 0.01 ESE=0.001	α = 0.005 ESE=0.001
20	5	T	0.1107	0.0585	0.0132	0.0070
20	5	T_{LR}	0.1359	0.0724	0.0173	0.0089
20	5	T_{CC}	0.0977	0.0977	0.0103	0.0059
20	5	T_{NS}	0.1039	0.1039	0.0220	0.0148
12	12	T	0.1050	0.0540	0.0122	0.0065
12	12	T_{LR}	0.1102	0.0563	0.0120	0.0062
12	12	T_{CC}	0.1007	0.0505	0.0104	0.0054
12	12	T_{NS}	0.1082	0.0670	0.0260	0.0170
5	25	T	0.1008	0.0510	0.0103	0.0053
5	25	T_{LR}	0.1027	0.0517	0.0101	0.0051
5	25	T_{CC}	0.0994	0.0490	0.0095	0.0046
5	25	T_{NS}	0.1059	0.0696	0.0312	0.0231

The following functions can we calculate each of four tests for n = 5, 12, 20

Table 2. Power function for n=20.

	statistic	α = .1	α = .05	α = .01	α = .005
$\lambda_1 = \cdots = \lambda_{10} = 5$	T_A	.23	.14	.04	.024
	T_{NS}	.23	.13	.036	.02
$\lambda_{11} = \cdots = \lambda_{20} = 10$	T_{CC}	.23	.13	.04	.024
	T_{LR}	.23	.13	.04	.024
$\lambda_1 = \cdots = \lambda_{10} = 5$	T_A	.4	.28	.11	.073
	T_{NS}	.4	.29	.11	.07
$\lambda_{11} = \cdots = \lambda_{20} = 8$	T_{CC}	.4	.28	.11	.07
	T_{LR}	.4	.28	.11	.07
$\lambda_1 = \cdots = \lambda_7 = 6$	T_A	.43	.3	.132	.082
$\lambda_8 = \cdots = \lambda_{14} = 9$	T_{NS}	.43	.28	.11	.07
$\lambda_{15} = \cdots = \lambda_{20}$	T_{CC}	.4	.3	.11	.073
	T_{LR}	.4	.3	.11	.073

Table 3. Power function for n=12.

λ_i	statistic	$\alpha = .1$	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$
$\lambda_1 = \cdots = \lambda_6 = 5$	T_A	.66	.54	.3	.23
$\lambda_7 = \cdots = \lambda_{12} = 10$	T_{NS}	.78	.65	.3	.23
	T_{CC}	.65	.52	.29	.22
	T_{LR}	.65	.52	.29	.22

$\lambda_1 = \cdots = \lambda_6 = 1$ $\lambda_7 = \cdots = \lambda_{12} = 2$	T_A T_{NS} T_{CC} T_{LR}	.2 .19 .2 .2	.12 .11 .12 .12	.03 .028 .03 .03	.02 .015 .019 .019
$\lambda_1 = \cdots = \lambda_4 = 1$ $\lambda_5 = \cdots = \lambda_8 = 5$ $\lambda_9 = \cdots = \lambda_{12} = 9$	T_A T_{NS} T_{CC} T_{LR}	.995 .99 .98 .98	.99 .99 .95 .95	.95 .95 .86 .86	.93 .99 .81 .81

Table 4. Power function for n=5.

λ_i	Statistic	$\alpha = .1$	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$
$\lambda_1 = \lambda_2 = 20$ $\lambda_3 = \cdots = \lambda_5 = 25$	T_A T_{NS} T_{CC} T_{LR}	.21 .19 .12 .12	.127 .11 .125 .125	.0374 .028 .035 .035	.022 .015 .021 .021
$\lambda_1 = \lambda_2 = 5$ $\lambda_3 = \cdots = \lambda_5 = 8$	T_A T_{NS} T_{CC} T_{LR}	.52 .57 .49 .4	.39 .4 .36 .36	.18 .18 .17 .17	.13 .13 .12 .12
$\lambda_1 = \lambda_2 = 15$ $\lambda_3 = \lambda_4 = 16$ $\lambda_5 = 19$	T_A T_{NS} T_{CC} T_{LR}	.15 .14 .15 .15	.084 .076 .084 .084	.022 .017 .022 .022	.012 .009 .012 .012

As you can see, in the case of the Poisson parameter to increase the sample size is small, the test may be higher Anscomb

REFERENCES

[1] Parsian, A. (2009). Mathematical statistics (2nd edition), Isfahan Industrial University Publications.
[2] Anscombe F.J. (1948). The transformation of Poisson, binomial and negativebinomial data, Biometrika.35, 246-254.
[3] Bartlett, M.S. (1947). The use of transformations, Biometrics, 3, 39-52.
[4] Brown, L.D. and Zhao, L.H.(2002). A test for the Poisson disterbution,Sankhya: 64, Series A, Pt.3,611-625.
[5] Cochran, W.G. (1954). Some method of string thing the common χ^2Test, Biometrics, 10, 417-451.
[6] Brown, L.D., Zhang, R. and Zhao L.(2001). Root un-root methodologhy for non-parametric density estimation and Poisson random effects models. Technical report. www.stat.wharton.upenn.edu/1brown.
[7] Mood, A.M., Graybill, F.A.andBoes, D, C..(1997). "Introduction to the Theory of Statistic" 3 ED., Mc. Graw-Hill Book Company, New York.
[8] Rice, J. (1995). Mathematical Statistic and Data Analysis. Second edition. Duxbury press, CA.