AN ADHOC TEST FOR GENERALIZED POISSON DISTRIBUTION

K. Muralidharan

Department of Statistics, Bhavnagar University, Bhavnagar, 364 002. India.

ABSTRACT

Let X be a random sample from a Generalized Poisson distribution with parameters θ and b. For $\beta = 0$, the model reduces to a Poisson distribution. We propose a test for $H_0: \beta = 0$ against $H_1: \beta > 0$ based on the conditional distribution of X_i given T = t, where T is the complete sufficient statistic of the nuisance parameter θ. The percentiles and powers are simulated. The performance of the test is compared with another adhoc test and the test proposed by Consul and Shenton (1973) test. We observe that, the performance of the proposed test is better for any value of $\beta > 0$. Two examples are also given to study the appropriateness of the model.

Key words: poisson distribution; conditional distribution; sufficient statistics, nuisance parameter.

1. INTRODUCTION

Let X be a discrete random variable defined over non-negative integral values and let p(x;θ,β) denote the probability that the random variable X takes the non-negative integral value x. Then the Generalized Poisson Distribution (GPD) is defined as

$$p(x;\theta,\beta) = \frac{(\theta e^{-\beta\theta})^x e^{-\theta}(1+x\beta)^{x-1}}{x!}, \quad x = 0,1,2,\ldots \tag{1.1}$$

$\theta > 0$, $\beta \geq 0$. The symbols θ and β are respectively called the first and second parameters of the GPD model. For $\beta = 0$, the model reduces to a Poisson distribution with parameter θ.

The model was first introduced by Consul and Jain (1970, 1973 a,b) as a limiting form of another model. This model provides excellent fits to various types of observed patterns which are supposed to be of Binomial, Negative Binomial or Poisson and to many other observed data where many other models fir or do not fit. Authors like Nelson (1975), Janardan and Schaeffer (1977), Janardan et.al. (1979) and Consul and Shoukri (1985) have studied the inferential problems of the parameters in the model (1.1). Consul and Shenton (1973) have proposed a test for $H_0 : \beta = \beta_0 \leq 0.5$ against $H_1 : \beta > \beta_0$ or $H_2 : \beta < \beta_0$ based on large sample approximations. Fazal (1977) has proposed likelihood ratio test for $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ based on Neyman Pearson lemma. It is also observed that, all the above tests are sensitive with respect to the nuisance parameter θ. Therefore, a test for β which doesn't depend on the nuisance parameter is of interest.

In this paper we propose an adhoc test for testing $H_0 : \beta = 0$ against $H_1 : \beta > 0$ based on conditional distribution of X_i given T = t, where T is the complete sufficient statistic for θ when β is known. In Section 2, we derive the adhoc test. The asymptotic null distribution is given in Section 3. The simulated percentile points and estimates of powers are given in Section 4. The performance of the proposed test with two other tests alongwith two numerical examples are also given at the end of the section.

2. THE ADHOC TEST

Let $X_1, X_2, \ldots, X_n$ be a random sample from (1.1). Then the joint density is given by

$$L(\underline{x};\theta,\beta) = \theta^t e^{-\theta(n+\beta t)} \prod_{i=1}^{n} \frac{(1+\beta x_i)^{x_i-1}}{x_i!},$$

is a complete family. Therefore, $T = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic for θ, having the pdf,

$$p(T=t) = \frac{n(n+\beta t)^{t-1}}{t!} \left(\theta e^{-\beta\theta}\right)^t e^{-n\theta}, t = 0,1,2,\ldots.$$

To propose the test statistic for $H_0 : \beta = 0$ against $H_1 : \beta > 0$, we consider the conditional distribution of X_i given T = t under H_0.

Under H_0,

$$f(x_i|t) = \binom{t}{x}\left(\frac{1}{n}\right)^x \left(1-\frac{1}{n}\right)^{t-x}, \quad x = 0,1,2,\ldots,t. \tag{2.1}$$

and

$$f(x_i, x_j|t) = \frac{t!}{x_i!x_2!(t - x_2 - x_2)!}\left(\frac{1}{n}\right)^{x_1+x_2}\left(1 - \frac{2}{n}\right)^{t-x_1-x_2}, \; x_1 + x_2 \le t. \quad (2.2)$$

Since the density (2.1) is free from the nuisance parameter θ, we shall derive the expression for the test statistic Q for testing $H_0 : \beta = 0$ against $H_1 > \beta = 0$, considering that the data ($X_1, X_2, ..., X_n$) have come from (2.1) for known values for T = t. For this, we define $Q = (\underline{X} - \underline{\mu})' \Sigma^{-1}(\underline{X} - \underline{\mu})$, where $\mu = E_{H_0}(X_i/T = t)$ and $\Sigma = ((\sigma_{ij}))$ is the variance-covariance matrix of $(X_i \mid t)$ computed under H_0. Thus,

$$\mu = E_{H_0}(X_i/T = t) = t/n \quad (2.3)$$

$$\sigma_{ii} = V_{H_0}(X_i/T = t) = (n-1)t/n^2 \quad (2.4)$$

and

$$\mathrm{Cov}_{H_0}(X_i, X_j/T = t) = -t/n^2. \quad (2.5)$$

Therefore using (2.4) and (2.5), the variance-covariance matrix can be written as,

$$\Sigma = \begin{bmatrix} \frac{(n-1)t}{n^2} & \frac{-t}{n^2} & \cdots & \frac{-t}{n^2} \\ \frac{-t}{n^2} & \frac{(n-1)t}{n^2} & \cdots & \frac{-t}{n^2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{-t}{n^2} & \frac{-t}{n^2} & \cdots & \frac{(n-1)t}{n^2} \end{bmatrix}$$

$$= \frac{t}{n}\left[I_n - \frac{1}{n}E_{nn}\right],$$

and the inverse of Σ is given by

$$\Sigma^{-1} = \frac{n}{t}\left[I_n + \frac{1}{n}E_n\right],$$

where I_n and E_n are respectively the identity matrix and unit matrix of order n. Thus the test statistic is given by

$$Q = \left(\underline{X} - \underline{\mu}\right)' \Sigma^{-1} \left(\underline{X} - \underline{\mu}\right)$$

$$= n \frac{\sum_{i-1}^{n} x_i^2}{t} - t.$$

Replacing T by its random variable, we define the test statistic as

$$Q^* = n \frac{\sum_{i-1}^{n} X_i^2}{\sum_{i-1}^{n} X_i} - \sum_{i-1}^{n} X_i. \tag{2.6}$$

Under H_0, $E(Q^*) = n - 1$. Since, $E(Q^*)$ under the alternative is difficult to obtain, we simulate $E(Q^*)$ under H_1 for different values of n and β and is presented in Table.1.

Tab. 1: E(Q*) for different value of n and β.

n	β = 0.0	β = 0.1	β = 0.2	β = 0.3	β = 0.4	β = 0.5
10	8.9856	10.5925	11.8936	17.7769	21.7730	24.2370
20	18.8898	22.1031	27.9343	34.1708	43.5823	57.1884
30	28.9756	34.5907	43.3993	55.0048	78.5211	94.8128
40	38.9649	50.6549	59.4567	74.7442	93.1032	125.6614
50	48.8865	60.0650	76.1897	93.8896	123.6256	164.2045

From the table it is observed that $E_{H_1}(Q^*) \geq E_{H_0}(Q^*)$ for $\beta > 0$. Hence the test procedure is to reject H_0 for large values of Q^*.

3. ASYMPTOTIC NULL DISTRIBUTION OF Q*.

If we define $U_1 = \sum_{i-1}^{n} X_i/n$ and $U_2 = \sum_{i-1}^{n} X_i^2/n$, then $Q^* = nU_2/U_1 - nU_1 =$ $g(U_1/U_2)$. If $\underline{U} = (U_1/U_2)'$, then $E_0(\underline{U}) = \underline{\mu}_0 = (\theta, \theta + \theta^2)'$ and

$$D_0(\underline{U}) = \begin{bmatrix} \frac{\theta}{n} & \frac{2\theta^2+\theta}{n} \\ \frac{2\theta^2+\theta}{n} & \frac{4\theta^3+6\theta^2+\theta}{n} \end{bmatrix}.$$ Now since $\sqrt{n}\left(\underline{U}-\underline{\mu_0}\right) \to N_2\left(\underline{0}, nD_0\right)$,

using 61.2 (I) page no. 387 of Rao (1974), we have the asymptotic distribution of Q^* is standard normal. Thus, for large values of n, one my use the tables of normal distribution for the test procedure.

4. SIMULATION STUDY

We obtain the percentile points of Q^* as well as the power of the test by using Monte-Carlo experiment. The simulations are carried out for 2000 times for different samples of size n=10(10)50. Table 2 presents the percentile points Q^* for various values of n.

Tab. 2: Percentile points of Q^*.

n	1%	5%	95%	99%
10	2.0305	3.5882	16.5556	20.8182
20	8.5238	10.7778	29.1429	37.0952
30	14.7143	18.0032	42.7143	50.5005
40	21.7778	26.3077	55.1111	62.5385
50	29.0755	34.2308	66.5833	74.8532

Table 3 presents the powers of the tests Q^*, another adhoc test based on $\overline{X}$, say Q_1 and the test proposed by Consul and Shenton (1973), say CS. To test $H_0: \beta = 0$ against $H_1 > \beta = 0$, the test based on $\overline{X}$ rejects H_0 for large values of say Q_1. Since the CS test is restricted for values of $\beta \leq 0.5$, under the alternative, we compute the powers of Q^*, Q_1 and CS test for values of $\beta \leq 0.5$ only. Further, the power of CS test depends on the choice of θ under H_0. Therefore, we assume $\beta = 0$ and $\theta = 1$ under H_0 for the CS test.

It is observed that, the performance of the adhoc test is better than any other test. For large β and n the performance of all the tests are at par.

Tab. 3: Power of the test for β > 0.

n	Test	β = 0.1		β = 0.2		β = 0.3		β = 0.4		β = 0.5	
		1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
10	Q^*	.04	.12	.12	.23	.24	.39	.41	.56	.57	.70
	Q_1	.01	.11	.08	.21	.21	.38	.44	.55	.41	.60
	CS	.07	.12	.18	.26	.21	.27	.53	.61	.51	.54
20	Q^*	.05	.19	.18	.42	.40	.62	.65	.84	.82	.91
	Q_1	.03	.12	.15	.34	.34	.53	.50	.66	.80	.90
	CS	.09	.16	.26	.33	.39	.52	.60	.63	.74	.80
30	Q^*	.07	.21	.29	.50	.56	.75	.81	.91	.92	.98
	Q_1	.04	.14	.16	.36	.38	.57	.65	.71	.83	.86
	CS	.11	.16	.31	.35	.46	.56	.69	.75	.89	.91
40	Q^*	.11	.24	.38	.56	.70	.83	.89	.96	.98	.99
	Q_1	.05	.15	.18	.43	.43	.74	.72	.86	.91	.97
	CS	.14	.20	.39	.50	.60	.68	.79	.83	.92	.96
50	Q^*	.12	.28	.45	.65	.79	.91	.96	.98	.99	1.000
	Q_1	.07	.18	.24	.45	.68	.76	.81	.87	.93	.98
	CS	.26	.30	.49	.56	.72	.79	.82	.88	.93	.98

As an application, we consider two examples from Consul (1989). The first example corresponds to the accidents of 122 experienced shunting men for the years 1937-1947. The data alongwith the computed value of Q^* are as follows:

Number of accidents	0	1	2	3	4	5	≥ 6
Frequency	21	31	26	19	7	9	9

n = 122 $\quad \sum_{i-1}^{n} x_i = 267 \quad \sum_{i-1}^{n} x_i^2 = 967 \quad Q^* = 174.85$

The second example is based on the distribution of 402 sow bugs and the data alongwith the computed value of Q^* are:

Number per board	0	1	2	3	4	5	6	7	8
Frequency	28	28	14	11	8	11	2	3	3

n = 122 $\quad \sum_{i-1}^{n} x_i = 402 \quad \sum_{i-1}^{n} x_i^2 = 2990 \quad Q^* = 505.42$

The table value of Q^* for n = 122 are 151.518 and 145.5954 correspond to upper 1% and 5% significance levels. Hence the computed value of Q^* for both the above examples are greater than the tabulated value and hence the null hypothesis is rejected. Which shows that a GPD model is a best alternative for identifying the distribution. Thus the advantages of the proposed adhoc test can be stated as follows:

- the test is not sensitive with respect to the nuisance parameter.
- easy to compute and draw conclusions.
- for any value of $\beta > 0$, the test performs well.

ACKNOWLEDGEMENT

The author is pleased to thank the referee and the Editor for their valuable comments.

REFERENCES

CONSUL P.C., (1989), Generalized Poisson distributions, Marcel Dekker Inc., New York.

CONSUL P.C. and JAIN G.C., (1970, 1973,a), A generalization of the Poisson distribution, Abstract, *Ann. Math. Statist.* 41(4); *Technometrics*, 15, 791-799.

CONSUL P. C. and JAIN G.C., (1973,b), On some interesting properties of the generalized Poisson distribution, *Biometrische, Z,* 15, 495-500.

CONSUL P.C. and SHENTON L.R., (1973), Some interesting properties of lagrange distribution. *Comm.Statist.- Theory & Methods*, 2(3), 263-272.

CONSUL P.C. and SHOUKRI M.M., (1985), The generalized Poisson distribution when the Sample mean is larger than the sample variance, *Comm. Statist. Simulation Comp.* 14(3), 667-681.

FAZAL S.S., (1977), A test for a generalized Poisson distribution, *Biometrical J.,* 19(4), 245-251.

LAWLESS J.F., (1982), Statistical models and Methods for Lifetime data., John Wiley & Sons.

NELSON D.L., (1975), Some remarks on generalizations of the negative binomial and Poisson distributions, *Technometrics*, 17(1), 135-136.

JANARDAN K.G. and SCHAEFFER D.J., (1977), Models for the analysis of chromosomal aberrations in human leukocytes, *Biometrical J.* 19(8), 599-612.

JANARDAN K.G., KERSTER H.W. and SCHAEFFER D.J., (1979), Biological applications of The Lagrangian Poisson distribution, *Bioscience*, 29(10), 599-602.

JANI P.N., SHANUBHOGUE A. and MURALIDHARAN K., (2000), A method of constructing test Based on the quadratic form. *Metron,* Vol. LVIII- N. 1-2.

RAO C.R., (1974), Linear Statistical inference and its applications. (Second edition), Wiley Eastern Ltd.

UN TEST ADHOC PER LA DISTRIBUZIONE DI POISSON GENERALIZZATA

Riassunto

Sia X un campione casuale proveniente da una Distribuzione Generalizzatta di Poisson con parametri θ *e b. Per* $\beta = 0$ *il modello considerato si può ricondurre ad una distribuzione di Poisson. In questo lavoro si propone un test per* $H_0 : \beta = 0$ *(contro* $H_1 : \beta > 0$*) basato sulla distribuzione condizionata di* x_i *dato* $T = t$*, dove T è una statistica sufficiente per* θ*. Due esempi sono riportati per dimostrare la bontà del test.*