Journal of AgriSearch 1(3): 183-187
Peer Reviewed Quarterly
ISSN : 2348-8808 (Print), 2348-8867 (Online)

Methods of Estimation of Generalized Negative Binomial Distribution

ABHAY KUMAR*, RC BHARATI, SK SINGH[1], A MISHRA[1] AND KM SINGH

ICAR Research Complex for Eastern Region, Patna, Bihar (India)

ARTICLE INFO

Received on	: 03.05.2014
Revised received on	: 14.06.2014
Accepted on	: 12.07.2014
Published online	: 07.09.2014

ABSTRACT

The negative binomial distribution was perhaps the first probability distribution, considered in statistics, whose variance is larger than its mean. On account of wide variety of available discrete distributions, the research workers in applied fields have begun to wonder which distribution would be most suitable one in a particular case and how to choose it. Generalized Negative Binomial Distribution (GNBD) reduces the binomial or the negative binomial distribution as particular cases and converges to a Poisson-type distribution in which the variance may be more than, equal to or less than the mean, depending upon the value of the parameter. A number of methods for estimation of parameters of GNBD, like weighted discrepancies method, minimum chi-square method etc. are available but these methods produce such equations which are not simple to be solved directly and hence some iterations has to be applied to find the solution. An alternative estimator has been suggested here, which is capable of giving more or less as good results as given by the moment estimators. Although, the values of P (χ^2), the probability of the observed value of χ2 to be exceeded, are slightly higher in case of the suggested method that in case of method of moments, these differences do not seem to be much significant and can be considered due to sample fluctuation. Moreover, it is relatively very quick to be obtained and so it may be preferred to others where very quick results are required.

Keywords: estimation, Binomial distribution, GNBD

INTRODUCTION

The negative binomial distribution was perhaps the first probability distribution, considered in statistics, whose variance is larger than its mean. On account of wide variety of available discrete distributions, the research workers in applied fields have begun to wonder which distribution would be most suitable one in a particular case and how to choose it. With the aim of reducing this problem, Jain and Consul (1971) gave a Generalized Negative Binomial Distribution (GNBD) by compounding the negative binomial distribution with another parameter which takes into account the variations in the mean and variance. This GNBD reduces to the binomial or the negative binomial distribution as particular cases and converges to a Poisson-type distribution in which the variance may be more than, equal to or less than the mean, depending upon the value of the parameter.

For 1 > α > 0 and $|\alpha\beta|$< 1 the Generalized Negative Binomial Distribution (GNBD) has been defined by

$$b_\beta(x,n,\alpha)=\frac{n\Gamma[n+\beta x]}{x!\Gamma[n+\beta x-x+1]}\alpha^x(1-\alpha)^{n+\beta x-x},\ n>0,\ x=0,1,2,\ldots, \quad \text{[Eq.1]}$$

such that

$$b_\beta(x, n, \alpha) = 0 \text{ for } x \le m \text{ if } n + \alpha m < 0.$$

Kemp (1986) showed that the ML method of estimation can be regarded as a scoring method using weighted sums of discrepancies between observed and expected frequencies. Famoye and Lee (1992) adopting the Kemp's approach applied the weighted discrepancies method for the estimation of Generalized Poisson Distribution

[1] Department of Statistics, Patna University, Patna, Bihar (India)

Corresponding authors Email: akumar1904@rediffmail.com

(GPD). In addition to it they also applied the minimum chi- square method of estimation to estimate GPD. An attempt has been made here to study the parameters of the GNBD using the weighted discrepancies method and also the minimum chi-square method. Both the method produces such equations which are not simple to be solved directly and hence some iteration method has to be applied to find the solutions. However the same difficulty has been found also in the case of GPD by Famoye and Lee (1992).

Weighted Discrepancies Method

Let f_x denote the observed frequencies; x=0,1,…,k. Obviously, k is the largest of the observations.

Let $N = \sum_{x=0}^{k} f_x$

The corresponding relative frequencies are given by

$$n_x = \frac{f_x}{n}; x = 0,1,2,\ldots,k \qquad \text{[Eq.2]}$$

The log likelihood function for the GNBD can be written as

$$\log L = \sum_x N n_x \log P(x;\alpha,\beta,n) \qquad \text{[Eq.3]}$$

And the likelihood equations as

$$\sum_{x=0}^{k} n_x \frac{\partial}{\partial n} \log P_x = 0$$
$$\sum_{x=0}^{k} n_x \frac{\partial}{\partial \alpha} \log P_x = 0 \qquad \text{[Eq.4]}$$
$$\sum_{x=0}^{k} n_x \frac{\partial}{\partial \beta} \log P_x = 0$$

Where P_x is taken for P(x; α, β, n) for the sake of simplicity.

From the fact that $\sum_x P_x = 1$ we have

$$\sum_{x=0}^{k} P_x \frac{\partial}{\partial n} \log P_x = 0$$
$$\sum_{x=0}^{k} P_x \frac{\partial}{\partial \alpha} \log P_x = 0 \qquad \text{[Eq.5]}$$
$$\sum_{x=0}^{k} P_x \frac{\partial}{\partial \beta} \log P_x = 0$$

From equation (3) and (4), we get

$$\sum_{x=0}^{k} (n_x - P_x) \frac{\partial}{\partial n} \log P_x = 0$$
$$\sum_{x=0}^{k} (n_x - P_x) \frac{\partial}{\partial \alpha} \log P_x = 0 \qquad \text{[Eq.6]}$$
$$\sum_{x=0}^{k} (n_x - P_x) \frac{\partial}{\partial \beta} \log P_x = 0$$

Substituting the corresponding expression for the derivatives in (Eq.6) from (Eq.3), (Eq.4) and (Eq.5), we have

$$\sum_{x=0}^{k} (n_x - P_x) \left[\frac{(N - f_0)}{n} + N \log(1-\alpha) + \sum_{x=2}^{k} \sum_{j=1}^{x-1} \frac{f_x}{(n + \beta x - j)} \right] = 0 \qquad \text{[Eq.7]}$$

$$\sum_{x=0}^{k} (n_x - P_x) \left[\frac{N\bar{x}}{\alpha} - \frac{N(n + (\beta - 1)\bar{x})}{(1-\alpha)} \right] = 0 \qquad \text{[Eq.8]}$$

and

$$\sum_{x=0}^{k} (n_x - P_x) \left[N\bar{x} \log(1-\alpha) + \sum_{x=2}^{k} \sum_{j=1}^{x-1} \frac{x f_x}{(n + \beta x - j)} \right] = 0 \qquad \text{[Eq. 9]}$$

To obtain the weighted discrepancies estimates, these equations can be solved iteratively by the Newton-Raphson method. The initial values of n, a and β can be taken as the moment estimate of these parameters.

Method of Minimum Chi-Square

We know that

$$\chi^2 = \sum_{x=0}^{k} \frac{(n_x - P_x)^2}{P_x}$$

has approximately chi-square distribution. For minimizing it, we differentiate *w.r.t.* n, α and β and so the minimum chi-square equations are obtained as

$$\sum_{x=0}^{k}(n_x - P_x)\left(1+\frac{n_x}{P_x}\right)\frac{\partial}{\partial n}\log P_x = 0$$

$$\sum_{x=0}^{k}(n_x - P_x)\left(1+\frac{n_x}{P_x}\right)\frac{\partial}{\partial \alpha}\log P_x = 0 \qquad \text{[Eq.10]}$$

$$\sum_{x=0}^{k}(n_x - P_x)\left(1+\frac{n_x}{P_x}\right)\frac{\partial}{\partial \beta}\log P_x = 0$$

Again substituting for the derivatives their corresponding expressions we get a system of equations as

$$\sum_{x=0}^{k}(n_x - P_x)\left(1+\frac{n_x}{P_x}\right)\left[\frac{(N-f_0)}{n} + N\log(1-\alpha) + \sum_{x=2}^{k}\sum_{j=1}^{x-1}\frac{f_x}{(n+\beta x - j)}\right] = 0 \qquad \text{[Eq.11]}$$

$$\sum_{x=0}^{k}(n_x - P_x)\left(1+\frac{n_x}{P_x}\right)\left[\frac{N\bar{x}}{\alpha} - \frac{N(n+(\beta-1)\bar{x})}{(1-\alpha)}\right] = 0 \qquad \text{[Eq.12]}$$

and

$$\sum_{x=0}^{k}(n_x - P_x)\left(1+\frac{n_x}{P_x}\right)\left[N\bar{x}\log(1-\alpha) + \sum_{x=2}^{k}\sum_{j=1}^{x-1}\frac{x f_x}{(n+\beta x - j)}\right] = 0 \qquad \text{[Eq.13]}$$

These equations like weighted discrepancies method are not straight to be solved directly and hence some method of iteration of finding solution such as Newton-Raphson method may be applied.

It may be seen that the weights in the likelihood equations depend only upon the observed frequencies while the weighted discrepancies method and the minimum chi-square method, both have weights depending upon the parameters as well as the observed frequencies.

Interval Estimation

Let $(X_1, X_2, \ldots, X_N)$ be a random sample from the GNBD with parameter β. As the GNBD is a member of the MPSD, $Y=\sum_{i=1}^{N}X_i$ is sufficient and complete statistic for α. The distribution of Y is also GNBD and

$$P(Y=y) = \frac{nN}{nN+\beta y}\binom{nN+\beta y}{y}\frac{\left[\alpha(1-\alpha)^{\beta-1}\right]^{y}}{(1-\alpha)^{-nN}};\ y=0,1\ldots$$

The 100(1-α)% lower and upper confidence limits for α in the GNBD can be obtained by solving

$$\sum_{y=y_2}^{\infty}\frac{nN}{nN+\beta y}\binom{nN+\beta y}{y}\frac{\left[\alpha(1-\alpha)^{\beta-1}\right]^{y}}{(1-\alpha)^{-nN}} = \frac{1}{2}\alpha_1$$

and

$$\sum_{y=0}^{y_1}\frac{nN}{nN+\beta y}\binom{nN+\beta y}{y}\frac{\left[\alpha(1-\alpha)^{\beta-1}\right]^{y}}{(1-\alpha)^{-nN}} = \frac{1}{2}\alpha_1$$

It may not be possible to solve the above equations algebraically for y_1 an y_2 when n, α and β are given. These equations may be solved by applying some iteration method like the Newton-Raphson method. In case of large samples the 100(1-α_1) % confidence limits for the parameter α can be obtained as

$$\frac{\bar{X} - Z_{\frac{\alpha_1}{2}}\left(\frac{s}{\sqrt{N}}\right)}{n+\beta\left(\bar{X} - z_{\frac{\alpha_1}{2}}\left(\frac{s}{\sqrt{n}}\right)\right)}$$

and

$$\frac{\bar{X} + Z_{\frac{\alpha_1}{2}}\left(\frac{s}{\sqrt{N}}\right)}{n+\beta\left(\bar{X} - z_{\frac{\alpha_1}{2}}\left(\frac{s}{\sqrt{n}}\right)\right)}$$

where $Z_{\frac{\alpha_1}{2}}$ is the value of standard normal variate towards right of which the area under standard normal curve is and

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$$

An Alternative Estimator

Most often in practice the observed distributions are obtained only up to a finite value of a variable i.e. up to finite number of classes having non-zero frequencies. Far such cases an estimating technique can be applied conveniently as follows.

Let us consider that t is the highest observed value having non-zero frequency i.e. (t-1) is the lowest value having zero frequency. From one of the conditions of pmf of the GNBD, we have

$$n + \beta t - t \geq 0 \text{ for } x \leq 0$$

considering the equality we get,

$$n + \beta t - t = 0$$

This gives

$$n = t(1-\beta)$$

We have the expression of mean of the GNBD as

$$\mu_1' = \frac{n\alpha}{1-\alpha\beta} = \frac{n\alpha}{\theta} \text{ ; where } \theta = (1-\alpha\beta)$$

Substituting the value of n from above and after a little simplification, we get

$$\mu_1' = t - \frac{t(1-\alpha)}{\theta}$$

This gives $\theta = \dfrac{t(1-\alpha)}{t-\mu_1'}$

Again the variance of GNBD is given by

$$\mu_2 = \frac{n\alpha(1-\alpha)}{(1-\alpha\beta)^3}$$

Therefore, $\dfrac{\mu_2}{\mu_1'} = \dfrac{1-\alpha}{\theta^2}$

Using the value of θ, we get

$$\frac{\mu_2}{\mu_1'} = \frac{(1-\alpha)(t-\mu_1')^2}{t^2(1-\alpha)^2}$$

This is some simplification and replacing the population moments by the respective sample moments, we get fully the estimate of α as

$$\hat{\alpha} = 1 - \frac{(t-\overline{X})^2\,\overline{X}}{t^2 s^2}$$

and that of θ as

$$\hat{\theta} = \frac{(t-\overline{X})\,\overline{X}}{t\, s^2}$$

The estimate of β and n can be obtained from

$$\hat{\beta} = \frac{(1-\hat{\theta})}{\hat{\alpha}}$$

and $\hat{n} = t(1-\hat{\beta})$

Goodness of Fit

An attempt has been made to fit the GNBD to some observed distributions estimating the parameters α, β and η by the suggested alternative method. To know how much good or bad the fits are due to this method in comparison to those due to the method of moments, the fits of the GNBD by the methods of moments have also been shown. We have used a couple of sports data sets used by Sinha (1984) for fitting the NBD. The expected frequencies according to both the methods along with the estimates of the parameters and the values of chi-square are given in the following tables.

Table 1: Wickets taken by Sobers in 158 completed innings in test cricket

Wickets taken	Observed frequency	Expected frequency	
		Method of Moments	Suggested Method
0	39	47.96	45.91
1	41	43.51	46.79
2	31	31.33	32.70
3	23	19.16	18.49
4	8	10.04	8.95
5	5	}6.00	}5.75
6	1		
Total	158	158.00	158.00
Mean	1.4873418		
χ^2		1.355	2.3163
d.f.		1	2
P (χ^2)		0.251	0.318
$\hat{\alpha}$		0.7462185	0.5759178
$\hat{\beta}$		0.7555587	0.7573010
$\hat{n}$		0.8693978	1.4561940

Table 2: Runs scored by Vishwanath in 142 completed innings

Runs (unit of 30)	Observed frequency	Expected frequency	
		Method of Moments	Suggested Method
0	73	73.75	71.56
1	35	32.96	36.04
2	16	17.75	18.15
3	9	9.41	8.84
4	6	}8.13	}7.41
5	2		
6	1		
Total	142	142.00	142.00
Mean	0.9436619		
χ^2		0.4174	0.6577
d.f.		1	2
P (χ^2)		0.524	0.721
$\hat{\alpha}$		0.7509945	0.5988346
$\hat{\beta}$		0.8321722	0.8749750
$\hat{n}$		0.4712604	1.7501499

CONCLUSION

It is encouraging to observe from these tables that the suggested estimator is giving more or less as good results as given by the moment estimators. It may be seen from tables 1 and 2 that the values of P (χ^2), the probability of the observed value of χ^2 to be exceeded, are slightly higher in case of the suggested method than in case of method of moments. However, these differences do not seem to be much significant and can be considered due to sample fluctuation. Further, the suggested method has one definite advantage over the other methods in certain situations. It can be applied when the method of moments fails to give estimates of the parameters. Moreover, it is relatively very quick to be obtained and so it may be preferred to others where very quick results are required.

REFERENCES

Famoye F and Lee CMS. 1992. Estimation of generalized poisson distribution. *Commun. Statist. Simul. And Comput.* **21**(1) 173 – 188.

Jain GC and Consul PC. 1971. A generalized negative binomial distribution, *SIAM J. Appl. Math.* **21**(4) 501 – 513.

Kemp AW. 1986. Weighted discrepancies and maximum likelihood estimation for discrete distributions. *Commun. Statist. Theor. Meth.* **15**(3) 783 - 803.

Sinha AK. 1984. The negative binomial distribution and its role in operational research and population studies, *Ph. D. Thesis*, Patna University, Patna, India.

Correct Citation: Kumar A, Bharati RC, Singh SK, Mishra A and Singh KM.2014. Methods of estimation of generalized negative binomial distribution. *Journal of AgriSearch* **1**(3): 183-187