The Asymptotic Behavior of Likelihood Estimators for the Dispersion Parameter of the Negative Binomial Distribution

Neville Weber*
University of Sydney, Sydney, Australia
neville.weber@sydney.edu.au

Katherine Courtney
University of Sydney, Sydney, Australia

Abstract

The random variable X has a negative binomial distribution, NBD(μ, α), if $P(X = x) = \Gamma(x+\alpha)/[\Gamma(\alpha)\ x!] \times p^\alpha(1-p)^x$, for $x = 0, 1, \ldots$, where $\mu = EX = \alpha(1-p)/p$, $\alpha > 0$ and $0 < p < 1$. There is general agreement that the sample mean is the best estimator for the parameter μ. Several alternative estimators have been proposed for the dispersion parameter, α, the simplest being the method of moments estimator, which is defined provided the sample variance is greater than the sample mean. In simulation studies the method of moments estimator, the conditional likelihood estimator (CLE) and the maximum likelihood estimator (MLE) have been found to be the most reliable estimators, but all three exhibit erratic behavior for particular parameter combinations. We develop asymptotic results to confirm that the CLE and MLE exhibit the same asymptotic behavior as the simple moment estimator. These results help to explain the observed heavy tailed behavior and provide an alternative perspective on when the various estimators are defined.

Keywords: Negative binomial distribution; maximum likelihood estimator; conditional likelihood estimator; weak limit theorem.

1. Introduction

The random variable X has a negative binomial distribution NBD(μ, α) if

$$P(X = x) = \frac{\Gamma(x + \alpha)}{\Gamma(\alpha)\ x!} p^\alpha (1 - p)^x, \text{ for } x = 0, 1, \ldots,$$

where $\mu = EX = \alpha(1-p)/p$, $\alpha > 0$ and $0 < p < 1$. We can write $p = \alpha/(\mu + \alpha)$.
It is easy to show that $\text{Var}(X) = \mu(1+\mu/\alpha)$. The negative binomial has found wide applicability for modelling count data which are overdispersed relative to the Poisson model (see Tripathi (1985) for examples).
The sample mean is an efficient estimator for the parameter μ. However there have been many estimators proposed for the dispersion parameter α (see van de Ven (1993) and Al-Khasawneh (2010) and references therein). Simulation studies have shown that the most widely used estimators; the moment estimator, maximum likelihood estimator (MLE) and conditional likelihood estimator (CLE), are not defined in certain cases and they all demonstrate heavy tailed behavior when the α parameter is large relative to μ.
The classical approximations for the binomial distribution illustrate that different insights can be obtained by considering arrays where the parameter varies with the sample size. For example, by allowing the probability parameter in the binomial to vary with the sample size we obtain the classical Poisson approximation as an alternative to the normal approximation. This paper provides a series of asymptotic results based on arrays of negative binomial distributions that help to explain the observed erratic behavior in the dispersion parameter estimators.

2. Results

Let $\{X_{ni}\}_{i=1}^n$ be an array of row-wise independent random variables such that $X_{ni} \sim NBD(\mu_n, \alpha_n)$. Let $\bar{X}_n = n^{-1}\sum_{i=1}^n X_{ni}$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_{ni} - \bar{X}_n)^2$. If $S_n^2 > \bar{X}_n$ the moment estimator for α_n is

$$\hat{\alpha}_{M,n} = \bar{X}_n^2/(S_n^2 - \bar{X}_n). \tag{1}$$

The likelihood function is

$$L_n = \prod_{i=1}^{n} \frac{\Gamma(X_{ni}+\alpha_n)}{\Gamma(\alpha_n)\ X_{ni}!} \left(\frac{\alpha_n}{\mu_n+\alpha_n}\right)^{\alpha_n} \left(\frac{\mu_n}{\mu_n+\alpha_n}\right)^{X_{ni}}.$$

Let $\mathcal{L}_n$ denote the log likelihood. The MLE for μ_n is the sample mean and the MLE for α_n, $\hat{\alpha}_{L,n}$, is the solution of

$$\frac{\partial \mathcal{L}_n}{\partial \alpha_n} = n^{-1} \sum_{i=1}^{n} I(X_{ni} \geq 1) \sum_{r=1}^{X_{ni}} (\alpha_n + r - 1)^{-1} - \log\left(1 + \frac{\bar{X}_n}{\alpha_n}\right) = 0. \tag{2}$$

The conditional likelihood estimator, introduced by Kalbfleisch and Sprott (1973), maximizes the likelihood obtained after conditioning on the sufficient statistic for μ_n. The sufficient statistic for μ_n is $\sum_{i=1}^{n} X_{ni}$ and $\sum_{i=1}^{n} X_{ni} \sim NBD(n\mu_n, n\alpha_n)$, so the conditional distribution is

$$P(X_{n1} = x_{n1}, .., X_{nn} = x_{nn} | \sum_{i=1}^{n} X_{ni} = a_n) = \frac{\Gamma(n\alpha_n)\ a_n!}{\Gamma(n\alpha_n + a_n)} \prod_{i=1}^{n} \frac{\Gamma(\alpha_n + x_{ni})}{\Gamma(\alpha_n)\ x_{ni}!}.$$

Let $\mathcal{L}_{n1}$ denote the logarithm of the conditional likelihood. The CLE, $\hat{\alpha}_{C,n}$, satisfies

$$\frac{\partial \mathcal{L}_{n1}}{\partial \alpha_n} = n^{-1} \sum_{i=1}^{n} I(X_{ni} \geq 1) \sum_{r=1}^{X_{ni}} (\alpha_n + r - 1)^{-1} - \sum_{r=1}^{n\bar{X}_n} (n\alpha_n + r - 1)^{-1} = 0. \tag{3}$$

Lemma 1 *For $\alpha \geq 0$*

$$\sum_{r=1}^{k} (\alpha + r - 1)^{-1} = \frac{k}{\alpha} - \frac{k(k-1)}{2\alpha(\alpha+1)} + \ldots + \frac{(-1)^{k-1} k(k-1)\ldots 1}{k\alpha(\alpha+1)\ldots(\alpha+k-1)}.$$

By applying the identity in Lemma 1 to (2) and (3) we can relate the CLE and MLE to the moment estimator via an appropriate expansion. We find that the three estimators have similar asymptotic behavior.
For large n the estimator $\bar{X}_n$ approximates μ_n and S_n^2 is close to $\mu_n(1+\mu_n/\alpha_n)$, and so the moment estimator for α_n, with $S_n^2 - \bar{X}_n$ in the denominator, is unstable when μ_n/α_n is small. We will show that the asymptotic behavior of $\hat{\alpha}_{M,n}$ depends on the rate at which the ratio approaches 0.
By analogy with Hall (1994), which dealt with estimators based on the binomial distribution, we will consider the distributions of the estimators assuming (μ_n, α_n) satisfy the following conditions as $n \to \infty$:

$$\alpha_n \to \infty \tag{4}$$
$$\mu_n \to \mu, \quad \text{where } 0 < \mu \leq \infty \tag{5}$$
$$\frac{\mu_n}{\alpha_n} \to 0 \tag{6}$$
$$\lambda_n = n(\mu_n/\alpha_n)^2 \to \lambda, \quad \text{where } 0 \leq \lambda \leq \infty. \tag{7}$$

Theorem 1 *Let Z be a standard normal random variable and let $\hat{\alpha}_n$ denote the moment estimator for α. Assume that (4) to (7) hold.*

(i) If $\lambda = 0$,

$$\sqrt{\frac{2}{\lambda_n}} \frac{\hat{\alpha}_n}{\alpha_n} \xrightarrow{\mathcal{D}} Z^{-1}.$$

(ii) If $0 < \lambda < \infty$,

$$\frac{\hat{\alpha}_n}{\alpha_n} \xrightarrow{\mathcal{D}} \frac{1}{1 + \sqrt{\frac{2}{\lambda}} Z}.$$

(iii) If $\lambda = \infty$,

$$\sqrt{\frac{\lambda_n}{2}}\left(\frac{\hat{\alpha}_n - \alpha_n}{\alpha_n}\right) \xrightarrow{\mathcal{D}} Z.$$

Thus if λ is finite then the estimator displays heavy tailed behavior as Z^{-1} has Cauchy-like tails. This corresponds to α_n growing much faster than μ_n and is consistent with the patterns observed in the simulations of van de Ven (1993). The limit result for the moment estimator was reported in van de Ven and Weber (1999).

To obtain the corresponding results for the MLE and CLE we need a slightly stricter constraint. For the case where $\lambda < \infty$ introduce the new condition

$$\mu_n \left(\frac{\mu_n}{\alpha_n}\right)^{j_1} \to 0 \;\text{ for some positive integer } j_1. \tag{8}$$

For the case when $\lambda = \infty$ we will impose the condition

$$\sqrt{n}\mu_n \left(\frac{\mu_n}{\alpha_n}\right)^{j_2} \leq 1, \;\text{ for some finite } \; j_2 > 0. \tag{9}$$

Theorem 2 *Results (i) - (ii) of Theorem 1 hold under conditions (4)-(8) where $\hat{\alpha}_n$ refers to either the MLE or the CLE for α_n. Further, if λ is infinite and (4)-(7) and (9) are satisfied, then result (iii) holds for the CLE and the MLE.*

3. Proof Outlines

Proof Outline for Theorem 1

Let $Z_{ni} = X_{ni} - \mu_n$. Write

$$\begin{aligned} S_n^2 - \bar{X}_n &= (n-1)^{-1}[\sum_{i=1}^{n}(X_{ni} - \bar{X}_n)^2 - (n-1)\bar{X}_n] \\ &= (n-1)^{-1}[\sum_{i=1}^{n}(Z_{ni}^2 - EZ_{ni}^2 - Z_{ni}) - n\bar{Z}_n^2 + \bar{X}_n + nEZ_{n1}^2 - n\mu_n] \\ &= (n-1)^{-1}[V_n - n\bar{Z}_n^2 + \bar{X}_n + n\mu_n^2/\alpha_n], \end{aligned} \tag{10}$$

where V_n is a sum of independent, zero mean random variables. If conditions (4) - (6) hold then we can apply the classical Central Limit Theorem to get

$$\Psi_n = V_n/(\sqrt{2n}\mu_n) \xrightarrow{\mathcal{D}} Z,$$

as it is easy to check that Liapounov's condition holds using fourth moments of the summands.

Thus the moment estimator can be written as

$$\frac{\hat{\alpha}_{M,n}}{\alpha_n} = \frac{[(n-1)/n]\;(\bar{X}_n/\mu_n)^2}{1 + \sqrt{\frac{2}{\lambda_n}}\Psi_n - \bar{Z}_n^2/\mu_n^2 + \bar{X}_n/(n\mu_n^2)}, \tag{11}$$

where $\lambda_n = n(\mu_n/\alpha_n)^2$. Under conditions (4) and (5), for any $\epsilon > 0$,

$$\begin{aligned} P(|\bar{X}_n/\mu_n - 1| > \epsilon) &\leq Var\bar{X}_n/(\epsilon^2\mu_n^2) \\ &= \frac{1}{n\epsilon^2\mu_n} + \frac{1}{n\epsilon^2\alpha_n} \to 0, \end{aligned}$$

and so $\bar{X}_n/\mu_n \xrightarrow{p} 1$. Similarly $\bar{Z}_n^2/\mu_n^2 \xrightarrow{p} 0$. Hence if $\lambda_n \to \lambda$ and $0 < \lambda < \infty$ then the result is immediate.

Note

$$\alpha_n(S_n^2-\bar{X}_n)/\mu_n^2 = 1+\sqrt{\frac{2}{\lambda_n}}\Psi_n + o_p(1). \tag{12}$$

By rearranging (10) we can obtain the results for the other two cases. If $\lambda = 0$ then

$$\sqrt{\frac{2}{\lambda_n}}\frac{\hat{\alpha}_{M,n}}{\alpha_n} \simeq \frac{1}{\sqrt{\frac{\lambda_n}{2}}+\Psi_n}.$$

If $\lambda = \infty$ then it is easy to show that

$$\sqrt{\frac{\lambda_n}{2}}\left(\frac{\hat{\alpha}_{M,n}-\alpha_n}{\alpha_n}\right) = \frac{\Psi_n + o_p(1)}{1+\sqrt{\frac{2}{\lambda_n}}\Psi_n + o_p(1)}.$$

Proof Outline for Theorem 2
To obtain the results for the CLE we need further notation. For simplicity we will suppress the n subscript where it does not cause any ambiguity. Let

$$S_{nj} = \frac{n\bar{X}_n(n\bar{X}_n-1)\dots(n\bar{X}_n-j+1)}{n\alpha(n\alpha+1)\dots(n\alpha+j-1)} \tag{13}$$

and

$$T_{nj} = \frac{\sum_{i=1}^{n} X_{ni}(X_{ni}-1)\dots(X_{ni}-j+1)}{n\alpha(\alpha+1)\dots(\alpha+j-1)}. \tag{14}$$

Applying Lemma 1 to equation (3) yields the following representation. The expression contains an infinite sum but since the X_{ni} are integer valued the summands will be 0 after some point as $S_{nj}=0$ if $j \geq n\bar{X}_n+1$ and $T_{nj}=0$ if $j \geq \max_{i\leq n}\{X_{ni}\}+1$.

Lemma 2

$$\begin{aligned}&\frac{2\alpha(\alpha+1)(n\alpha+1)}{n(n-1)\alpha_n(S_n^2-\bar{X}_n)}\frac{\partial\mathcal{L}_1}{\partial\alpha}\\ &= \left(\frac{\hat{\alpha}_{M,n}}{\alpha_n}-\frac{\alpha}{\alpha_n}\right)+\frac{2\alpha(\alpha+1)(n\alpha+1)}{(n-1)\alpha_n(S_n^2-\bar{X}_n)}\sum_{j=3}^{\infty}\frac{(-1)^j}{j}(S_{nj}-T_{nj})-\frac{S_n^2}{n\alpha_n(S_n^2-\bar{X}_n)}.\end{aligned}$$

Following Hall (1994), let $C_1(n)\geq 1$ be a slowly varying function satisfying $C_1(n)\to\infty; C_1(n)\mu_n/\alpha_n \to 0$ and $\mu_n(C_1(n)\mu_n/\alpha_n)^{j_1}\to 0$ as $n\to\infty$. We restrict attention to estimators $\hat{\alpha}$ satisfying $\hat{\alpha}/\alpha_n \in [C_1(n)^{-1}, C_1(n)]$. Let j_0 be an arbitrary fixed integer such that $j_1+3\leq j_0<\infty$ where j_1 is the constant from condition (8). The Markov inequality and standard arguments can be used to establish the following.

Lemma 3 *Under conditions (4) and (6)*

$$\sum_{j=j_0+1}^{\infty}\frac{(-1)^j}{j}T_{nj} = O_P\left(\frac{\mu_n^3}{\alpha^2(\alpha+1)}\left(\frac{C_1\mu_n}{\alpha_n}\right)^{j_0-2}\right),$$

and

$$\sum_{j=j_0+1}^{\infty}\frac{(-1)^j}{j}S_{nj} = O_P\left(\frac{\mu_n^3}{\alpha^2(\alpha+1)}\left(\frac{C_1\mu_n}{\alpha_n}\right)^{j_0-2}\right).$$

To determine the order of $\sum_{j=3}^{j_0}(-1)^j(S_{nj}-T_{nj})/j$ we need to mean adjust the terms and consider $\sum_{j=3}^{j_0}(-1)^j(S_{nj}-ES_{nj})/j$, $\sum_{j=3}^{j_0}(-1)^j(T_{nj}-ET_{nj})/j$ and $\sum_{j=3}^{j_0}(-1)^j(ES_{nj}-ET_{nj})/j$ separately.

Lemma 4 *Under conditions (4) - (6) and (8)*

$$\begin{aligned}\sum_{j=3}^{\infty}\frac{(-1)^j}{j}(S_{nj}-T_{nj}) &= O_P\left(\frac{\mu_n^2}{\sqrt{n}\alpha^2(\alpha+1)}+\frac{\mu_n^2}{\alpha^2(\alpha+1)}\left(\frac{C_1\mu_n}{\alpha_n}\right)\right),\ \textit{for } \mu_n\geq 1,\\ &= O_P\left(\frac{\sqrt{\mu_n}}{\sqrt{n}\alpha^2(\alpha+1)}+\frac{\mu_n^2}{\alpha^2(\alpha+1)}\left(\frac{C_1\mu_n}{\alpha_n}\right)\right),\ \textit{for } \mu_n< 1.\end{aligned}$$

Under conditions (4)-(6) $S_n^2/(n\mu_n^2)\overset{p}{\to}0$, and $\sqrt{\lambda_n}S_n^2/(n\mu_n^2)\overset{p}{\to}0$. Using the representation (10) and adapting the arguments outlined for Theorem 1 we obtain the results (i) and (ii) for the CLE.
To deal with the case where $\lambda_n\to\infty$ more delicate arguments are required to bound the remainder terms in the following representation, particularly the term involving the difference of expectations $ET_{nj}-ES_{nj}$.

$$\sqrt{\frac{\lambda_n}{2}}\frac{2\alpha(\alpha+1)(n\alpha+1)}{n(n-1)\alpha_n(S_n^2-\bar{X}_n)}\frac{\partial\mathcal{L}_1}{\partial\alpha}=\sqrt{\frac{\lambda_n}{2}}\left(\frac{\hat{\alpha}_{M,n}}{\alpha_n}-1\right)-\sqrt{\frac{\lambda_n}{2}}\left(\frac{\alpha}{\alpha_n}-1\right)+\sqrt{\frac{\lambda_n}{2}}\frac{S_n^2}{n\alpha_n(S_n^2-\bar{X}_n)}+$$

$$+\sqrt{\frac{\lambda_n}{2}}\frac{2\alpha(\alpha+1)(n\alpha+1)}{(n-1)\alpha_n(S_n^2-\bar{X}_n)}\left(\sum_{j=3}^{\infty}\frac{(-1)^j}{j}(S_{nj}-ES_{nj})-\sum_{j=3}^{\infty}\frac{(-1)^j}{j}(T_{nj}-ET_{nj})-\sum_{j=3}^{\infty}\frac{(-1)^j}{j}(ET_{nj}-ES_{nj})\right).$$

To obtain the result for the MLE we apply Lemma 1 and a Taylor series expansion of the log function in (2). Let $Q_{nj}=(\bar{X}_n/\alpha)^j$ and $R_{nj_0}=(-\bar{X}_n/\alpha)^{j_0+1}/((j_0+1)(1+\xi)^{j_0+1})$ for some ξ satisfying $0\leq\xi\leq\bar{X}_n/\alpha$.

$$\frac{\partial\mathcal{L}}{\partial\alpha}=\frac{n[\bar{X}_n^2-\alpha((n-1)S_n^2/n-\bar{X}_n)]}{2\alpha^2(\alpha+1)}+n\sum_{j=3}^{j_0}\frac{(-1)^j}{j}(Q_{nj}-T_{nj})+nR_{nj_0}+n\sum_{j=j_0+1}^{\infty}\frac{(-1)^{j-1}}{j}T_{nj}. \qquad (15)$$

Reorganizing the expression we can explicitly relate the MLE to the moment estimator plus a remainder term as follows.

Lemma 5

$$\begin{aligned}\frac{2\alpha^2(\alpha+1)}{n\alpha_n((n-1)S_n^2/n-\bar{X}_n)}\frac{\partial\mathcal{L}}{\partial\alpha} &= \left(\frac{\hat{\alpha}_{M,n}}{\alpha_n}-\frac{\alpha}{\alpha_n}\right)+\frac{\hat{\alpha}_{M,n}}{\alpha_n}\left(\frac{S_n^2}{n((n-1)S_n^2/n-\bar{X}_n)}\right)\\ &+\frac{2\alpha^2(\alpha+1)}{\alpha_n((n-1)S_n^2/n-\bar{X}_n)}\left(\sum_{j=3}^{j_0}\frac{(-1)^j}{j}(Q_{nj}-T_{nj})+R_{nj_0}+\sum_{j=j_0+1}^{\infty}\frac{(-1)^{j-1}}{j}T_{nj}\right).\end{aligned}$$

Standard arguments lead to the following observations.

Lemma 6 *Assume conditions (4)-(7) hold. Then*

$$\begin{aligned}\sqrt{\frac{2}{\lambda_n}}\frac{\hat{\alpha}_{M,n}}{\alpha_n}\left(\frac{S_n^2}{n((n-1)S_n^2/n-\bar{X}_n)}\right) &= o_p(1),\ \textit{if } \lambda=0,\\ \frac{\hat{\alpha}_{M,n}}{\alpha_n}\left(\frac{S_n^2}{n((n-1)S_n^2/n-\bar{X}_n)}\right) &= o_p(1),\ \textit{if } 0<\lambda<\infty,\\ \sqrt{\frac{\lambda_n}{2}}\frac{\hat{\alpha}_{M,n}}{\alpha_n}\left(\frac{S_n^2}{n((n-1)S_n^2/n-\bar{X}_n)}\right) &= o_p(1),\ \textit{if } \lambda=\infty.\end{aligned}$$

Under conditions (4)-(6) we can show that $R_{nj_0}=O_P\left(\frac{\mu_n^3}{\alpha^2(\alpha+1)}(C_1\mu_n/\alpha_n)^{j_0-2}\right)$. By a similar argument to that required to establish Lemma 4 we get the following.

Lemma 7 *Under conditions (4) - (6) and (8)*

$$\sum_{j=3}^{j_0} \frac{(-1)^j}{j}(Q_{nj} - T_{nj}) + R_{nj_0} + \sum_{j=j_0+1}^{\infty} \frac{(-1)^{j-1}}{j} T_{nj} = O_P\left(\frac{\mu_n^2}{\sqrt{n}\alpha^2(\alpha+1)} + \frac{\mu_n^2}{\alpha^2(\alpha+1)}\left(\frac{C_1\mu_n}{\alpha_n}\right)\right), \text{ for } \mu_n \geq 1,$$

$$= O_P\left(\frac{\sqrt{\mu_n}}{\sqrt{n}\alpha^2(\alpha+1)} + \frac{\mu_n^2}{\alpha^2(\alpha+1)}\left(\frac{C_1\mu_n}{\alpha_n}\right)\right), \text{ for } \mu_n < 1.$$

Combining these observations we can establish results (i) and (ii) for the MLE. As with the CLE the $\lambda = \infty$ case requires slightly different constraints to bound the remainder terms.

4. Conclusions

The above results provide a theoretical interpretation of the erratic behavior of dispersion parameter estimators for the negative binomial distribution. From the first terms in the expansion in Lemma 2 we see that the CLE can always be calculated whenever the moment estimator exists, that is when $S^2 > \bar{X}$, and from the leading term in the expansion in (15) there is a positive solution for the MLE if $\frac{n-1}{n}S^2 > \bar{X}$, which is consistent with results in Levin and Reeds (1977).

Theorem 1 can be used to approximate the proportion of samples where the moment estimator is not defined. Assuming $\lambda < \infty$, the proportion is approximately $P\left((1 + \sqrt{\frac{2}{\lambda}}Z)^{-1} < 0\right) = \Phi(-\sqrt{\lambda/2})$. This uses the asymptotic variance of $V_n/(\sqrt{2n}\mu_n)$. A better finite sample approximation is $\Phi\left(-\sqrt{n}\mu_n/(\alpha_n\sqrt{2b_n})\right)$, where $b_n = 1 + \frac{1}{\alpha_n} + \frac{2\mu_n}{\alpha_n} + \frac{4\mu_n}{\alpha_n^2} + \frac{\mu_n^2}{\alpha_n^2} + \frac{3\mu_n^2}{\alpha_n^3}$.

References

Al-Khasawneh, M. F. (2010) Estimating the negative binomial dispersion parameter, Asian Journal of Mathematics & Statistics, 3, 1-15.

Kalbfleisch, J.D. and Sprott, D.A. (1973). Marginal and conditional likelihoods. Sankhya A 35, 311- 328.

Hall, P. (1994). On the erratic behavior of estimators of N in the binomial (N, p) distribution. J. Amer. Statist. Assoc. 89, 344-352.

Levin, B. and Reeds, J. (1977). Compound multinomial likelihood functions are unimodal: Proof of a conjecture of I.J. Good. Ann. Statist. 5, 79-87.

Tripathi, R. (1985). Negative binomial distribution. Encyclopedia of Statistical Sciences (Wiley, New York) 6, 169-177.

van de Ven, R. (1993). Estimating the shape parameter for the negative binomial distribution. J. Statist. Comput. Simul. 46, 111-123.

van de Ven, R. and Weber, N.C. (1999) On the asymptotic behavior of estimators of the shape parameter of the negative binomial distribution. Bulletin of the International Statistical Institute, ISI 52nd Session, 419-420.