Estimation of the shape parameter k of the negative binomial distribution

Mohammad F. Al-Saleh [*,1], Fatima K. Al-Batainah

Department of Statistics, Yarmouk University, Jordan

Abstract

Using approximate Bayesian methods, an estimator of the shape parameter of the negative binomial distribution is introduced. Most of the available estimators of this parameter, such as the method of moment and the maximum likelihood estimators (MME and MLE), give inadmissible values when the mean of the observed sample is larger than its variance. In this case, the MME gives values while the MLE is infinite. Such observed sample in not unlikely when the mean of the population is close to its variance. The suggested estimator always has valid values, and turned out to be more efficient for most values of the parameter space. Family size data is used for illustration of the procedure.
© 2002 Elsevier Science Inc. All rights reserved.

Keywords: Negative binomial distribution; Method of moments estimator; Maximum likelihood estimator; Bayes-like estimator

1. Introduction

Negative binomial distributions have been found to provide useful representations in many fields. Its applicability in birth-and-death processes had been shown by Furry [9] and Kendal [12]. Medical and military applications have been described by Chew [7] and Bennett and Birch [4]. It was used to model family size by Rao et al. [17]. Red mites on apple leaves [6] is a very

---

[*] Corresponding author.
E-mail address: malsaleh@squ.edu.om (M.F. Al-Saleh).
[1] Recently on leave at Department of Mathematics, Sultan Qaboos University, P.O. Box 36, 123 Al-Khodh, Oman.

diverse application which was examined by Suzanne et al. [18]. Elliott's [8] highlights its usefulness for analyzing samples of fresh water fauna. Wilson and Room [23], Binns [5] have used the negative binomial for modeling entomological data. For more details on these and other applications, see Johnson and Kotz [10].

The negative binomial random variable X is a non-negative discrete random variable with probability function, $NB(p, k)$, given by

$$f(x; p; k) = cp^k(1-p)^x; \quad x = 0, 1, 2, \ldots$$

$$0 < p < 1; \ k > 0 \quad \text{and} \quad c = \frac{\Gamma(x+k)}{\Gamma(x)\Gamma(x+1)}.$$

Note that, the mean of X is $\mu = kq/p$, the variance is $\sigma^2 = kq/p^2$; Clearly, $\mu < \sigma^2$.

This distribution has been used extensively on biological data and consequently, there has been some attempt to give an ecological meaning to the mean μ and the shape parameter k. μ has been thought of as the density of organisms in the area of interest, because an increase in μ, results when the population increases in size or become more dense, see Wilson and Room [23]. Anscombe [3], noted that there is a theoretical evidence that k depends on the intrinsic power of a species to reproduce it self and therefore, k can be thought of as a measure of aggregation.

In this paper we will concentrate on the estimation of k.

Let $X_1, X_2, \ldots, X_n$ be an independent and an identically distributed random variables (iid) from $NB(p, k)$, then the method of moment estimator (MME) of k is

$$\hat{k}_m = \frac{\overline{X}^2}{S^2 - \overline{X}} \quad \text{where} \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

The maximum likelihood estimator (MLE) is the solution of the equation:

$$\sum_{i=1}^{n}\sum_{r=0}^{x_i-1}\frac{1}{k+r} + n\ln\left(\frac{k}{k+\bar{x}}\right) = 0,$$

which has a unique solution provided that $\bar{x} < s^2$. These estimators have been considered by Pieters et al. [16], Suzanne et al. [18], Walter [21], Anraku and Yanagimoto [2], and others. The zero-class estimator was considered by Pieters et al. [16]. Maximum quasi-likelihood estimator was considered by McCullagh and Nelder [14]. It is noted by these authors and others, that both the MME and the MLE are superior to other estimators considered. However, these estimators are of no use when $s^2 \leqslant \bar{x}$. If μ is much smaller than k, then $\sigma^2 = \mu + \mu^2/k \cong \mu$; so, it is likely in this case to obtain samples with $s^2 < \bar{x}$, which yield negative MME and infinite MLE. Thus the estimation of k is most

difficult when μ is small and k is large. Suzanne et al. [18] indirectly estimated k, through its reciprocal ($\alpha = 1/k$), to avoid problem caused by infinite value of the MLE when $s^2 \cong \bar{x}$.

Bayesian methods have been used for the more popular problem of estimating the parameter p of the distribution. In this direction, see Papageorgiou and Wesolowski [15], Walter and Hamedani [20] and Wede and Pieters [22]. In this paper we will not use Bayesian methods directly; we will use approximate Bayesian technique to estimate the parameter k. Since for large positive integer k, X_i is approximately normal, we may use a suitable prior on the set $\{(\mu, \sigma^2) : 0 < \mu < \sigma^2\}$ and use Bayesian methods to derive an estimator of k. This estimator turned out to have valid values for all samples. Furthermore, the suggested estimator is superior to both MME and MLE for the difficult case (small μ and large k) and not bad otherwise. This approach of normal approximation before estimation, was used by Blumenthal and Dahiya (1981) and Al-Saleh and Shakhtreh [1], for estimating the binomial parameter, n.

The rest of the paper is organized as follows. In Section 2, the estimator is derived. Numerical comparison is reported in Section 3. Real data application is introduced in Section 4. Concluding remarks are given in Section 5.

2. Derivation of the estimator

Assume $X_1, \ldots, X_n$ are iid NB(p, k). If k is a positive integer then X_i can be represented as $X_i = \sum_{j=1}^{k} Y_j$ where Y_j are iid NB$(p, 1)$. For large k, X_i is approximately $N(\mu, \sigma^2)$; $0 < \mu < \sigma^2$ (by Central Limit Theorem). Assume $\pi(\mu|\sigma^2) = 1/\sigma^2$, for $0 \leqslant \mu < \sigma^2$ and zero otherwise (a non-informative uniform prior of μ given σ^2); and $\pi(\sigma^2) = 1/\sigma^2$ for $\sigma^2 > 0$ (Jeffery prior of σ^2), then the joint prior of μ and σ^2 is $\pi(\mu, \sigma^2) = 1/\sigma^4$ for $0 \leqslant \mu < \sigma^2 < \infty$. Hence, the posterior distribution of $(\mu, \sigma^2)|\underline{x}$ is approximated by

$$\pi(\mu, \sigma^2|\underline{x}) \cong \frac{(1/\sigma^2)^2 (1/\sigma^2)^{n/2} e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2}}{\int_0^\infty \int_0^{\sigma^2} \left(\frac{1}{\sigma}\right)^{n/2+2} e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2} \, d\mu \, d\sigma^2}$$

which can be written as

$$\pi(\mu, \sigma^2|\underline{x}) \cong \frac{h_1(\sigma^2) \frac{\sqrt{n}}{\sigma} \phi\left(\frac{\sqrt{n}(\mu - \bar{x})}{\sigma}\right)}{\int_0^\infty \int_0^{\sigma^2} h_1(\sigma^2) \frac{\sqrt{n}}{\sigma} \phi\left(\frac{\sqrt{n}(\mu - \bar{x})}{\sigma}\right) d\mu \, d\sigma^2}$$

$$\cong \frac{h_1(\sigma^2) \frac{\sqrt{n}}{\sigma} \phi\left(\frac{\sqrt{n}(\mu - \bar{x})}{\sigma}\right)}{E\left(\Phi\left(\frac{(\sigma^2 - \bar{x})\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\bar{x}\sqrt{n}}{\sigma}\right)\right)},$$

where $h_1(\sigma^2)$ is the density of the inverse gamma random variable, $\mathrm{IG}(\alpha, \beta)$, with $\alpha = (n+1)/2$, $\beta = 2/\sum(x_i - \bar{x})^2$; $\phi(z)$ is the density of the standard normal random variable, Φ is its cumulative distribution function and the expectation is, of σ^2 given $\underline{x}$, with respect to $\mathrm{IG}(\alpha, \beta)$ density.

Now,

$$E(\mu|\underline{x}) = \int_0^\infty \int_0^{\sigma^2} \mu\pi(\mu, \sigma^2|x)\,\mathrm{d}\mu\,\mathrm{d}\sigma^2$$

which can be simplified to

$$E(\mu|\underline{x}) = \hat{\mu} \cong \bar{x} - \frac{E\left[\frac{\sigma}{\sqrt{n}}\left(\phi\left(\frac{(\sigma^2-\bar{x})\sqrt{n}}{\sigma}\right) - \phi\left(-\frac{\bar{x}\sqrt{n}}{\sigma}\right)\right)\right]}{E\left[\Phi\left(\frac{(\sigma^2-\bar{x})\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\bar{x}\sqrt{n}}{\sigma}\right)\right]}.$$

Similarly,

$$E(\sigma^2|\underline{x}) = \int_0^\infty \int_0^{\sigma^2} \sigma^2\pi(\mu, \sigma^2|\bar{x})\,\mathrm{d}\mu\,\mathrm{d}\sigma^2$$

which can be simplified to

$$\hat{\sigma}^2 \cong \frac{E\left(\sigma^2\left[\Phi\left(\frac{(\sigma^2-\bar{x})\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\bar{x}\sqrt{n}}{\sigma}\right)\right]\right)}{E\left(\Phi\left(\frac{(\sigma^2-\bar{x})\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\bar{x}\sqrt{n}}{\sigma}\right)\right)}.$$

$\hat{\mu}$ and $\hat{\sigma}^2$ are called the approximate generalized Bayes estimates of μ and σ^2, respectively. Using $\hat{\mu}$ and $\hat{\sigma}^2$, k can be estimated by

$$\hat{k}_\mathrm{b} = \frac{\hat{\mu}^2}{\hat{\sigma}^2 - \hat{\mu}}.$$

Note that $\hat{\mu} \leqslant \hat{\sigma}^2$ and hence $\hat{k}_\mathrm{b} > 0$. Thus $\hat{k}_\mathrm{b}$, unlike the MME and MLE, always give admissible answers.

3. Numerical comparisons

In this section the suggested estimator of k, $\hat{k}_\mathrm{b}$, and the MME and the MLE, are compared. The method of simulation is described in the following steps:

1. For a given k, p and n, generate a random sample of size n from $\mathrm{NB}(k, p)$, x_1, $x_2, \ldots, x_n$. Let $\alpha = (n+1)/2$ and $\beta = 2/\sum(x_i - \bar{x})^2$.
2. Simulate L values of σ^2 from $\mathrm{IG}(\alpha, \beta)$.
3. Calculate the corresponding L values of $h^*(\sigma^2)$, $h^{**}(\sigma^2)$, and $h^{***}(\sigma^2)$ to obtain $h^*(\sigma_1^2), \ldots, h^*(\sigma_L^2)$, $h^{**}(\sigma_1^2), \ldots, h^{**}(\sigma_L^2)$, and $h^{***}(\sigma_1^2), \ldots, h^{***}(\sigma_L^2)$, where

$$h^*(\sigma^2) = \frac{\sigma}{\sqrt{n}} \left(\phi \left(\frac{(\sigma^2 - \bar{x})\sqrt{n}}{\sigma} \right) - \phi \left(-\frac{\bar{x}\sqrt{n}}{\sigma} \right) \right),$$

$$h^{**}(\sigma^2) = \Phi \left(\frac{(\sigma^2 - \bar{x})\sqrt{n}}{\sigma} \right) - \Phi \left(-\frac{\bar{x}\sqrt{n}}{\sigma} \right) \text{ and } h^{***}(\sigma^2) = \sigma^2 h^{**}(\sigma^2).$$

4. Approximate

$$E(h^*(\sigma^2)|\underline{x}) \text{ by } \hat{E}(h^*(\sigma^2)|\underline{x}) = \frac{1}{L} \sum_{i=1}^{L} h^*(\sigma_i^2),$$

$$E(h^{**}(\sigma^2)|\underline{x}) \text{ by } \hat{E}(h^{**}(\sigma^2)|\underline{x}) = \frac{1}{L} \sum_{i=1}^{L} h^{**}(\sigma_i^2)$$

and

$$E(h^{***}(\sigma^2)|\underline{x}) \text{ by } \hat{E}(h^{***}(\sigma^2)|\underline{x}) = \frac{1}{L} \sum_{i=1}^{L} h^{***}(\sigma_i^2).$$

5. Approximate

$$\hat{\mu} \text{ by } \hat{\mu}^* = \bar{x} - \frac{\hat{E}(h^*(\sigma^2)|\underline{x})}{\hat{E}(h^{**}(\sigma^2)|\underline{x})},$$

$$\hat{\sigma}^2 \text{ by } \hat{\sigma}^{*2} = \frac{\hat{E}(h^{***}(\sigma^2)|\underline{x})}{\hat{E}(h^{**}(\sigma^2)|\underline{x})} \quad \text{and } \hat{k}_b \text{ by } \frac{\hat{\mu}^{*2}}{\hat{\sigma}^{*2} - \hat{\mu}^{*2}}.$$

6. Repeat 1–5, m times to obtain $\hat{k}_1, \hat{k}_2, \ldots, \hat{k}_m$.
7. Calculate $\text{MS}\hat{\text{E}}(\hat{k}_m) \cong \sum_{i=1}^{m} (\hat{k} - k)^2 / m$.
8. The efficiency of $\hat{k}_b$ with respect to $\hat{k}_m$ is approximated by $\text{eff}(\hat{k}_b; \hat{k}_m) = \text{MS}\hat{\text{E}}(\hat{k}_m)/\text{MS}\hat{\text{E}}(\hat{k}_b)$.

The efficiency with respect to $\hat{k}_l$, is defined similarly.

The MSE of $\hat{k}_b$ and its efficiency with respect to the $\text{MME}(\hat{k}_m)$ and the $\text{MLE}(\hat{k}_l)$, for the cases of $n = 50$ and 100 are reported in Tables 1 and 2. L is taken to be 2000, and m is taken to be 10000. The MSE of the MME and the MLE are taken from Willson et al. (1984). The range of μ and k were chosen by the authors to represent the range, which is most commonly found in entomology. In their paper, samples with s^2 smaller than $\bar{x}$ were discarded. Thus, the performance of the MME and MLE reported in their tables is actually an over estimation of the actual performance of the two estimators.

Based on these tables, we may conclude the following:

1. The MSE of $\hat{k}_b$ and $\hat{k}_m$ are increasing with k for each fixed μ and decreasing with μ for each fixed k, for almost all the cases. Hence, it is harder to estimate k when μ is small and k is large. This is actually the case when $\mu \approx \sigma^2$.

Table 1
Comparisons of MLE, MME and $\hat{k}_b$ based on a sample of size 50

μ	k	Count	MSE($\hat{k}_l$)	MSE($\hat{k}_m$)	Count	MSE($\hat{k}_b$)	eff($\hat{k}_b; \hat{k}_l$)	eff($\hat{k}_b; \hat{k}_l$)
1	1	66	18	3.6	76	0.58	31.0	6.6
	3	1397	150	23	1529	1.5	100	15.3
	5	3278	240	31	2747	4.6	52.2	6.7
3	1	0	0.15	0.22	0	0.26	0.58	0.85
	3	19	56	16	37	5.2	10.8	3.08
	5	254	740	97	267	13.2	56.8	7.5
5	1	0	0.10	0.15	0	0.16	0.56	0.83
	3	0	3.3	3.2	0	0.67	1.0	0.97
	5	15	330	52	23	1.41	23.1	3.6

Table 2
Comparisons of MLE, MME and $\hat{k}_b$ based on a sample of size 100

μ	k	Count	MSE($\hat{k}_l$)	MSE($\hat{k}_m$)	Count	MSE($\hat{k}_b$)	eff($\hat{k}_b; \hat{k}_l$)	eff($\hat{k}_b; \hat{k}_l$)
1	1	3	0.57	0.55	2	0.334	1.71	1.65
	3	489	150	38	515	2.26	66.4	16.8
	5	1561	240	68	1547	3.5	68.3	19.36
3	1	0	0.059	0.084	0	0.09	0.644	0.92
	3	0	2.2	2.3	2	2.16	1.02	1.07
	5	16	170	40	32	10.6	16.91	3.98
5	1	0	0.04	0.064	0	0.07	0.563	0.9
	3	0	0.85	0.94	0	0.96	0.885	0.98
	5	0	6	5.7	0	5.3	1.13	1.08

2. The efficiency of $\hat{k}_b$ is much larger than 1 for small μ and large k. This means, in this hard case of estimation, $\hat{k}_b$ performs much better than $\hat{k}_m$. For most of other cases the efficiency is fairly close to 1, which means that the two estimators have similar performance.
3. For small μ and large k the number of bad samples (counts) is very large. It increases with k for fixed μ. This means, there are situations when $\hat{k}_m$ cannot be used while $\hat{k}_b$ can be obtained and used. Furthermore, the actual MSE of $\hat{k}_m$ can be higher than the reported values.
4. The MSE of $\hat{k}_b$ and $\hat{k}_m$ are decreasing when we move from $n = 50$ to 100, except when the counts of bad samples is large, in this case, the MSE increases with n for $\hat{k}_m$. We believe that this is due to the number of counts being reduced when n gets large. Actually, due to discarding bad samples, we do not expect the calculated MSE for $\hat{k}_m$ to behave as it is expected, because these calculated values are not the actual values.

All of the above conclusions are also valid for the case of the MLE instead of the MME.

4. Application: family size distribution

The distribution of completed family size (or sipship size) has been a subject of interest for human biologists, geneticists, demographers, and social scientists. There is good empirical evidence that the distribution is nearly that of a negative binomial [13]. Waller et al. [19] noticed that, the number of childless families is much greater than the expected number of childless families, when they fit the negative binomial distribution to the observed frequencies of completed family size from various sources, although the fit is good for the rest of the distribution. This led them to suggest that the childless families are a mixture of two types of families: (1) one is biologically fertile and could have children but did not. So this type is childless due to chance, and hence should be a part of the general negative binomial distribution of family size; (2) the other type is either biologically or electively not fertile (sterile) and thus have no children due to known reasons. This type should not be a part of the general negative binomial distribution of family size. Three methods of fitting the negative binomial distribution to the observed data have been employed. The first method consists of approximating the parameters p and k of the negative binomial distribution directly, from the observed data. In the second method, the zero class is considered as missing and the parameters k and p are estimated on the basis of the incomplete (truncated) distribution. The third method consists of iteration from the sets of initial trial values p, k obtained from methods I or II. We will use these methods to fit a negative binomial distribution to the family size data obtained from the census (1994) of Jordan.

Based on the general census of Jordan Department of Statistics [11], the observed family size frequency distribution along with the fitted frequencies for the three methods is given in Table 3. Fig. 1 represents the observed and fitted curves for the three methods. The following table contains a summary of the approximated values of μ, σ^2, p, and k for the three methods.

Parameters	Method I	Method II	Method III
μ	4.32151	4.87267	4.60081
σ^2	9.187	7.6346	8.49145
$\hat{p} = \mu/\sigma^2$	0.47034	0.638235	0.541818
$\hat{k} = \mu^2/(\sigma^2 - \mu)$	3.837	8.596491	5.44063

It can be seen from Table 3 and the curves, that the last method is the most appropriate method of fitting; so we adopt this method. Based on this method, the values of p and k for the fitted negative binomial distribution are 0.542 and 5.441, respectively.

To compare the values of the three estimates using an actual sample, we use a sample that was taken from the population of Jordan, by the Public Statistical

Table 3
Observed and fitted number of family size

Family size (x)	Observed number of families (f_x)	Fitted number for method		
		I	II	III
0	59979	35240.3	12410.9	21320.30
1	64047	71620.5	38596.8	53147.95
2	78838	91743.2	66997.6	78419.45
3	82384	94545.1	85610.1	89114.78
4	77575	85593.7	89787.7	86159.26
5	68431	71060.1	81832.0	74536.82
6	57795	55433.1	67084.8	59427.09
7	46127	41258.0	50605.8	44501.63
8	35983	29604.7	35691.3	31707.65
9	25766	20623.3	23810.1	21696.11
10	18410	14022.3	15157.3	14355.17
11	8303	9341.8	9270.3	9232.49
12	5315	6119.6	5795.39	5795.39
13+	7843	10590.3	8722.63	8722.63
Total	636796	636796	598137	598137

Department in 1996. Table 4 contains the frequency distribution of this sample. Assuming that the population in 1996 is not very different from that in 1994, we may regard this sample as a random sample from the original population. The values of the three estimates are summarized in the table below:

Method	$\hat{p}$	$\hat{k}$
MME($\hat{k}_m$)	0.5971	6.4196
MLE($\hat{k}_l$)	0.5787	5.9500
Bays-like($\hat{k}_b$)	0.5970	6.4169

It can be seen that the values of the three estimates are relatively close to the approximate actual values of p and k. Note that the values of $\bar{x}$ and s^2 for this sample are 4.33 and 7.26 respectively, so the variance is about twice as large as the mean. Therefore, any of the three estimators is likely to give a reasonable value.

To see how the three estimators deal with bad samples, random samples were continuously selected from the population, until three bad samples of size 50 each were obtained. The frequency distribution of each of these samples is given in Table 4. In these three samples, large values of family size are not observed. If we did not know the sampling distribution, then we may conclude that either the population is not negative binomial or the value of k is very large. Each of the MME and MLE fails to give reasonable values for the three
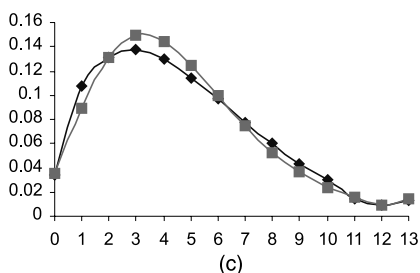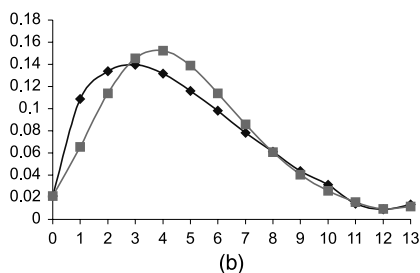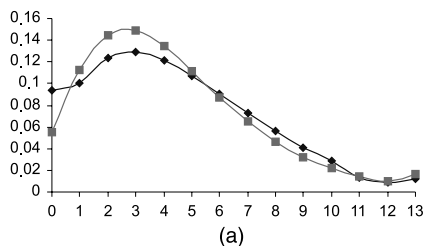
Fig. 1. Observed of fitted curve of family size using (a) method I, (b) method II and (c) method III.

samples, while this new estimator has valid values, though far from the actual one. The following are the values of the estimates:

Sample I: $\hat{k}_m = -37.7;$ $\hat{k}_l = \infty; \hat{k}_b = 34.59$

Sample II: $\hat{k}_m = -41.9;$ $\hat{k}_l = \infty; \hat{k}_b = 29.01$ (1)

Sample III: $\hat{k}_m = -134.0;$ $\hat{k}_l = \infty; \hat{k}_b = 20.67$

5. Concluding remarks

In this paper, a new estimator of the parameter k of the negative binomial random variable is introduced. This estimator turned out to be substantially more efficient that the MLE and the MME, for the cases for which the two

Table 4
Family size sample (1996) and three bad samples 1, 2 and 3

Size (x)	Frequency (96)	Frequency (1)	Frequency (2)	Frequency (3)
0	204	1	2	1
1	577	0	3	5
2	747	2	2	6
3	767	6	7	12
4	759	13	13	6
5	651	10	8	6
6	519	7	6	6
7	436	5	6	4
8	267	1	3	0
9	173	2	0	1
10	111	3	0	0
11	68	0	0	0
12	34	0	0	0
13+	29	0	0	0
Total	5342	50	50	50

estimators are not proper to use. When all of the three estimators are valid, the new estimator is still not bad. This estimator is recommended when prior information suggest that the mean of the population is much smaller than its variance; in this case both of the other two estimators might give improper values.

References

[1] Al-Saleh, Shakhtreh, Estimating the parameter n of the binomial distribution using moment generating function approach, Pakistan Journal of Statistics 13 (1997) 227–243.
[2] Anraku, Yanagimoto, Estimation for the negative binomial distribution based on the conditional likelihood, Common Statistics 19 (1990) 771–806.
[3] Anscombe, The transformation of Poisson, binomial and negative binomial data, Biometrika 35 (1949) 246–254.
[4] Bennett, Birch, Sampling inspection tables for comparison of two groups using the negative binomial model, Trabajos 15 (1964) 1–12.
[5] Binns, Behavioral dynamics and the negative binomial distribution, Oikos 47 (1986) 315–318.
[6] Bliss, Fisher, Fitting the negative binomial distribution to biological data, Biometrics 9 (1953) 176–200.
[7] Chew, Applications of the negative binomial distribution with probability of misclassification, Virginia Journal of Science (New Series) 15 (1964) 34–40.
[8] Elliott, Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates, Freshwater Biological Association Scientific Publication 25, Freshwater Biological Association, Ambleside, England, 1979.
[9] Furry, On fluctuation phenomena in the passage of high energy electrons through lead, Physical Review 52 (1937) 569–581.
[10] Johnson, Kotz, Univariate Discrete Distributions, John Wiley and Sons, Inc., 1992.

[11] Jordan Department of Statistics, Reports of General Cencus, Amman, Jordan, 1994.

[12] A. Kendal, Stochastic processes and population growth, Journal of the Royal Statistical Society, Series B 11 (1949) 230–282.

[13] Kojima, Kelleher, Survival of mutant genes, American Nature 96 (1962) 329–346.

[14] McCullagh, Nelder, Generalized Linear Models, Chapman and Hall, London, 1983.

[15] Papageorgiou, Wesolowski, Posterior mean identifies the prior distribution in NB and related models, Statistics and Probability letters 36 (1997) 127–134.

[16] Pieters, Gates, Matis, Sterling, Small sample comparison of different estimators of negative binomial parameters, Biometrics 33 (1977) 718–723.

[17] Rao, Mazumdar, Waller, Li, Correlation between the numbers of two types of children in a family, Biometrics 29 (1973) 271–279.

[18] Suzanne, Clark, Perry, Estimation of the negative binomial parameter k by maximum quasi-likelihood, Biometrics 45 (1989) 309–316.

[19] Waller, Rao, Li, Heterogenity of childless families, Social Biology 20 (2) (1973) 133–138.

[20] Walter, Hamedani, Bayes empirical Bayes estimation of natural exponential families with quadratic variance functions, Annals of Statistics 19 (1991) 1191–1224.

[21] Walter, Maximum likelihood estimation for the negative binomial dispersion parameter, Biometrics 46 (1990) 863–867.

[22] Wede, Pieters, Eye fixations on advertisements and memory for brands. A model and findings, Marketing Science 19 (2000) 297–312.

[23] Wilson, Room, Clumping patterns of fruit and arthropods in cotton, with implications for binomial sampling, Environmental Entomology 12 (1983) 50–54.