




# Confidence intervals and prediction intervals for two-parameter negative binomial distributions

Md Mahadi Hasan and K. Krishnamoorthy 

Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA, USA

## ABSTRACT

Problems of finding confidence intervals (CIs) and prediction intervals (PIs) for two-parameter negative binomial distributions are considered. Simple CIs for the mean of a two-parameter negative binomial distribution based on some large sample methods are proposed and compared with the likelihood CIs. Proposed CIs are not only simple to compute, but also better than the likelihood CIs for moderate sample sizes. Prediction intervals for the mean of a future sample from a two-parameter negative binomial distribution are also proposed and evaluated for their accuracy. The methods are illustrated using two examples with real life data sets.

## ARTICLE HISTORY

Received 19 December 2022  
Accepted 22 November 2023


## KEYWORDS

Joint sampling approach; maximum likelihood estimates; over-dispersion; poisson distribution; prediction intervals; score method

## 1. Introduction

Poisson models are commonly postulated for count data where the mean and variance are approximately equal. However, there are many situations where the mean of count data is smaller than the variance and over-dispersion is observed in the data, and Poisson model is not appropriate for analyzing such count data. The two-parameter negative binomial (NB) distribution has become increasingly popular as a more flexible alternative to the Poisson distribution especially when the data exhibit over-dispersion. Application of the NB model to over-dispersed data is noted in Anscombe [1]. This distribution, also known as the Poisson-gamma distribution, is often appropriate for data for aggregated organisms. The ‘over-dispersion’ arises when the organisms are ‘clumped’, ‘clustered’ or ‘aggregated’ in space or time, whereas the ‘under-dispersion’ arises from a more regular positioning than that produced by a Poisson mechanism (Ross and Preece [12]). The NB model has also been extensively used in crash data analysis, because crash data are usually characterized by over-dispersion (Park and Lord [11]). Sanchez and He [15] have noted that internet data on the number of packets per unit of time do not fit a Poisson distribution. They provided an example where the mean number of packets arriving per second is 384.262 and the variance is 15,306.35 which is much larger than the mean. Using a quantile-quantile plot, these authors demonstrated that a two-parameter negative binomial distribution fits the internet data well.

**CONTACT** K. Krishnamoorthy  [krishna@louisiana.edu](mailto:krishna@louisiana.edu)  Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA.

 Supplemental data for this article can be accessed online at <http://dx.doi.org/10.1080/02664763.2023.2297157>.

A reviewer has noted that there are other new discrete distributions available to model an over-dispersed count data. In particular, Mazucheli et al. [9] have introduced two discrete analogs for the Shanker distribution as alternatives to model over-dispersed data sets. Although, these new models maybe interesting in their own way, in this article we pay attention only to the available well-known two-parameter NB model for over-dispersed count data. The NB model that we consider involves two parameters, namely, the mean  $\mu$  and dispersion parameter  $\theta$  to capture the extra variation observed in the data. To describe a two-parameter negative binomial model, consider a family of mixed Poisson distributions with the following probability mass function (PMF)

$$P(X = x) = \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} f(\lambda) d\lambda, \tag{1}$$

where  $\lambda$  is the mean of a Poisson model and  $f(\lambda)$  is a known density function of the mixture distribution for  $\lambda$ . If  $f(\lambda)$  is the gamma density with the shape parameter  $\theta$  and the scale parameter  $\mu/\theta$ , then it can be readily verified that the marginal distribution of  $X$  has the PMF

$$P(X = x | \mu, \theta) = \frac{\Gamma(x + \theta)}{\Gamma(\theta)\Gamma(x + 1)} \left(\frac{\mu}{\mu + \theta}\right)^x \left(\frac{\theta}{\mu + \theta}\right)^\theta, \\ x = 0, 1, 2, \dots, \mu > 0, \theta > 0. \tag{2}$$

The distribution with the above PMF is called two-parameter negative binomial or Poisson-gamma distribution. It can also be developed from the usual negative binomial distribution with the PMF

$$P(Y = y | r, p) = \binom{r + y - 1}{y} p^r (1 - p)^y, \quad y = 0, 1, 2, \dots \tag{3}$$

where the random variable  $Y$  represents the number of failures until the occurrence of the  $r$ th success in a sequence of independent Bernoulli trials each with success probability  $p$ . By parameterizing  $\theta = r$  and  $\mu = E(X) = r(1 - p)/p$ , the above PMF can be written as

$$P(Y = y | \mu, \theta) = \frac{\Gamma(\theta + y)}{\Gamma(y + 1)\Gamma(\theta)} \left(\frac{\mu}{\mu + \theta}\right)^y \left(\frac{\theta}{\mu + \theta}\right)^\theta, \quad y = 0, 1, 2, \dots$$

where  $\theta$  is real positive, and both  $\theta$  and  $\mu$  are unknown. The mean and variance of the above distribution are  $E(Y) = \mu$  and  $\text{Var}(Y) = \mu + \mu^2/\theta$ .

For the usual negative binomial distribution that arises in inverse sampling, confidence intervals, prediction intervals and tolerance intervals are available in the literature; see Khurshid et al. [8] and Dang and Krishnamoorthy [6] and the references therein. However, as noted by Shilane et al. [13], determining confidence intervals for the mean of a NB distribution is not so straightforward, particularly when sample sizes are small. These authors have proposed a few methods and noted that some commonly used methods exhibit poor coverage in the case of high dispersion. They have proposed some CIs based on the asymptotic result that the sample mean of a NB distribution has a gamma distribution for large samples. In this article, we shall explore other accurate methods, such as the score and likelihood methods, to find CIs for the mean counts.

Another problem of interest that we shall address in this article is the construction of prediction intervals (PIs). In crash data analysis and other areas of applications, one may be interested in predicting the number of outcomes given the data at present (Wood [18]). The prediction interval for the mean counts of a future sample from the two-parameter NB distribution under consideration is also of practical importance. For a given sample  $X$  of size  $n$  from a  $NB(\mu, \theta)$  distribution, the problem is to find a PI for the mean  $\bar{Y}$  of a future sample of size  $m$  from the same  $NB(\mu, \sigma)$  distribution. Specifically, for a given confidence level  $1 - \alpha$ , the problem is to find two real valued functions  $L(X; \alpha)$  and  $U(X; \alpha)$  so that

$$P_{X,Y} (L(X, n; \alpha) \leq \bar{Y} \leq U(X, n; \alpha)) = 1 - \alpha,$$

for all  $\mu$  and  $\theta$ . Sheaffer and Leavenworth [14] have provided an approximate PI based on the Wald approximation.

In this article, we first describe all available methods to find confidence intervals and prediction intervals for two-parameter negative binomial distributions. We propose simple yet accurate score confidence intervals for the mean. We also propose PIs for the mean of a future sample using the joint sampling approach which produced accurate PIs for binomial and Poisson distributions (Krishnamoorthy and Peng [7]). In the following section, we provide CIs for the mean based on the Wald approach, likelihood method and the score method. We also propose a modification to the Wald CI to improve the coverage probability. In Section 3, we address the problem of prediction of the sample mean of a future sample from the negative binomial distribution based on a current sample available from the same distribution. We provide the Wald PI, the likelihood PI and a PI based on the joint sampling approach. In Section 4, we carried out simulation studies to judge the accuracy of the proposed CIs and PIs and to compare them in terms of precision. Two examples with real life data are used to illustrate the construction of CIs and PIs in Section 5 and some concluding remarks are given in Section 6.

## 2. Confidence intervals

Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a sample from a  $NB(\mu, \theta)$  distribution. Define the sample mean and variance as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

respectively.

### 2.1. Wald confidence intervals

Using the Wald [16] theorem, we see that

$$\frac{n(\bar{X} - \mu)^2}{S^2} \sim Z^2, \quad \text{asymptotically,}$$

where  $Z$  is the standard normal random variable. Let  $c_\alpha^2 = Z_{1-\alpha/2}^2$ , where  $Z_q$  denote the 100q percentile of the standard normal distribution. For a given  $(\bar{X}, S^2, \alpha)$ , the Wald CI for

the mean is formed by the two roots of the equation  $n(\bar{X} - \mu)^2/S^2 = c_\alpha^2$ , which are

$$\bar{X} \pm c_\alpha S/\sqrt{n}. \tag{4}$$

Shilane et al. [13] have considered the above Wald CI to estimate the mean  $\mu$ .

*Modified Wald Confidence Intervals*

Instead of using  $S^2$  in the above pivotal quantity, we use the expression

$$n^{-1} \sum_{i=1}^n (X_i - \mu)^2 \simeq S^2 + (\bar{X} - \mu)^2$$

and develop CI based on the asymptotic result that

$$\frac{n(\bar{X} - \mu)^2}{S^2 + (\bar{X} - \mu)^2} \sim Z^2.$$

Solving the equation  $n(\bar{X} - \mu)^2 = c_\alpha^2(S^2 + (\bar{X} - \mu)^2)$  for  $\mu$ , we find an approximate  $100(1 - \alpha)\%$  confidence interval for  $\mu$  as

$$\bar{X} \pm c_\alpha \frac{S}{\sqrt{n - c_\alpha^2}}. \tag{5}$$

Note that the above CI is defined only when  $n > c_\alpha^2 = z_{1-\alpha/2}^2$ , the squared  $100(1 - \alpha/2)$  percentile of the standard normal distribution. For a 95% CI  $z_{.975}^2 = 3.8415$  and for a 99% CI it is  $z_{.995}^2 = 6.6349$ . So in order to construct a modified Wald CI with any practical level of confidence, the sample size  $n$  should be at least 7. If the sample size is very small, (say, 3), then one can construct only a 90% (or less than 90%) CI based on the modified Wald method. Furthermore, the two CIs (4) and (5) are approximately the same for large  $n$ ; however, we will see that the latter one has better coverage probabilities for moderate sample sizes. We refer to the CI (5) as the modified Wald (M-Wald) confidence interval.

**2.2. Likelihood confidence intervals for  $\mu$**

Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a sample from a NB( $\mu, \theta$ ) distribution. The log-likelihood function can be expressed as

$$\begin{aligned} \ln L(\mu, \theta | \mathbf{X}) &= n\theta \ln \frac{\theta}{\mu + \theta} - n \ln \Gamma(\theta) + n\bar{X} \ln \frac{\mu}{\mu + \theta} \\ &+ \sum_{i=1}^n \ln \Gamma(X_i + \theta) - \sum_{i=1}^n \ln \Gamma(X_i + 1), \end{aligned} \tag{6}$$

where  $\bar{X}$  is the sample mean. It is easy to check that the partial derivative  $\partial \ln L(\mu, \theta | \mathbf{X})/\partial \mu = 0$  gives  $\hat{\mu} = \bar{X}$ . Taking partial derivative with respect to  $\theta$  and replacing  $\mu$  with  $\bar{X}$ , we see that the maximum likelihood estimator (MLE) of  $\theta$  is the solution of

the equation

$$h(\theta) = n \ln \frac{\theta}{\bar{X} + \theta} + \sum_{i=1}^n \psi(\theta + X_i) - n\psi(\theta) = 0, \tag{7}$$

where  $\psi(x) = \frac{d \ln \Gamma(x)}{dx}$  is the digamma function. Wilson et al. [17] have shown that there is at least one positive root for  $\theta$  in the above equation provided  $S^2 > \bar{X}$ . Aragon et al. [2] have shown that this condition is necessary and sufficient for the existence and uniqueness of the MLE. Dai et al. [5] have shown that if the sample mean is less than 3/2, then the condition  $S^2 > \bar{X}$  will be the sufficient and necessary condition for the above equation has a unique root. Bandara et al. [3] have provided a faster method of computing the MLE if one exists.

The root of Equation (7) can be found using the Newton-Raphson iterative scheme

$$\theta_{\text{new}} = \theta_{\text{old}} - h(\theta_{\text{old}})/h'(\theta_{\text{old}}), \tag{8}$$

where

$$h'(\theta) = n\bar{X}/[\theta(\bar{X} + \theta)] + \sum_{i=1}^n \psi'(\theta + X_i) - n\psi'(\theta),$$

and  $\psi'(x)$  is the trigamma function. The moment estimate  $\hat{\theta}_M = \bar{X}/(S^2/\bar{X} - 1)$  can be used as an initial value for  $\theta_{\text{old}}$ . The R function `glm.nb(x ~ 1, link = identity)`, available in the package ‘MASS,’ can also be used to find the MLEs.

*Confidence Interval for  $\mu$*

From the Fisher information matrix, the variance of  $\hat{\mu} = \bar{X}$  is obtained as  $(\mu + \mu^2/\theta)/n$ . Using an estimate of the variance and the asymptotic normality of the MLE, we find an approximate CI for  $\mu$  as

$$\bar{X} \pm \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\bar{X} + \bar{X}^2/\hat{\theta}}, \tag{9}$$

where  $z_\alpha$  is the  $100\alpha$  percentile of the standard normal distribution. We refer to this CI as the likelihood CI.

**2.3. Score confidence intervals**

Another CI for the mean  $\mu$  can be obtained by mimicking the idea of the score CI for a binomial proportion. In particular, we use the model-based variance expression  $\text{Var}(\bar{X}) = n^{-1}(\mu + \mu^2/\theta)$ , and use the result that

$$\frac{n(\bar{X} - \mu)^2}{\mu + \mu^2/\theta} \sim Z^2, \quad \text{asymptotically.}$$

Solving the equation for  $\frac{n(\bar{X} - \mu)^2}{\mu + \mu^2/\theta} = z_{1-\alpha/2}^2 = c_\alpha^2$  for  $\mu$ , and replacing  $\theta$  by an estimate  $\hat{\theta}$ , we find an approximate CI as

$$\frac{\bar{X} + c_\alpha^2/(2n)}{1 - c_\alpha^2/(n\hat{\theta})} \pm \frac{c_\alpha/\sqrt{n}}{1 - c_\alpha^2/(n\hat{\theta})} \sqrt{\frac{c_\alpha^2}{4n} + \bar{X} + \frac{\bar{X}^2}{\hat{\theta}}} \tag{10}$$

The above CI (10) is referred to as the score confidence interval. We use the MLE  $\hat{\theta}$  in the above CI.

The Wald and modified Wald CIs are straightforward to compute as they require only the sample mean and variance. The likelihood CIs and the score CIs require the computation of the MLE of  $\theta$ . The MLE of  $\theta$  is easy to obtain using the Newton-Raphson iterative scheme in (8) or the R package `glm.nb()`. We evaluate these methods for accuracy and precision in Section 4.1.

### 3. Prediction intervals

Let  $X_1, \dots, X_n$  be a sample from a  $NB(\mu, \theta)$  distribution. Let  $\bar{Y}$  denote the mean of a future sample from the same negative binomial distribution. In the following, we shall describe some prediction intervals for  $\bar{Y}$  based on the available sample  $X_1, \dots, X_n$ .

#### 3.1. Wald prediction intervals

The Wald PI is similar to the one for binomial distributions proposed in page 207 of Nelson [10]. This PI is based on the result that

$$\frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1), \quad \text{asymptotically,} \tag{11}$$

where  $S^2$  is the sample variance based on  $X_1, \dots, X_n$ . The PI for  $\bar{Y}$  is given by

$$\bar{X} \pm c_\alpha S\sqrt{\frac{1}{n} + \frac{1}{m}}, \tag{12}$$

where  $c_\alpha = z_{1-\alpha/2}$  is the upper  $\alpha/2$  quantile of the standard normal distribution. Instead of using the usual sample variance, one could use

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left( X_i - \frac{n\bar{X} + m\bar{Y}}{m+n} \right)^2 \simeq S^2 + \left( \frac{m}{m+n} \right)^2 (\bar{X} - \bar{Y})^2$$

in (11) and solving the equation

$$\frac{(\bar{X} - \bar{Y})^2}{S^2 \left( \frac{1}{n} + \frac{1}{m} \right) + \frac{m}{n(m+n)} (\bar{X} - \bar{Y})^2} = c_\alpha^2$$

for  $\bar{Y}$ , we find the PI for  $\bar{Y}$  as

$$\bar{X} \pm \frac{c_\alpha S}{\sqrt{1 - \frac{mc_\alpha^2}{n(m+n)}}} \sqrt{\frac{1}{n} + \frac{1}{m}}, \tag{13}$$

where  $c_\alpha^2 = z_{1-\alpha/2}^2$ . We refer to the above PI as the modified Wald PI.

**3.2. Likelihood prediction intervals**

Let  $\hat{\theta}$  be the MLE of  $\theta$  based on the sample  $X_1, \dots, X_n$ . Instead of using the moment estimate  $S^2$  for the variance in (11), we can use the MLE  $\hat{\sigma}_{mle}^2 = \bar{X} + \bar{X}^2/\hat{\theta}$ , and obtain the PI

$$\bar{X} \pm c_\alpha \hat{\sigma}_{mle} \sqrt{\frac{1}{n} + \frac{1}{m}}. \tag{14}$$

We refer to the above PI as the likelihood PI.

**3.3. Prediction intervals based on the joint sampling approach**

To find a prediction interval for the mean  $\bar{Y}$  of a sample of size  $m$  from a  $NB(\mu, \theta)$  distribution based on an available sample of size  $n$  from the same distribution, we consider the asymptotic result that

$$\frac{(\bar{X} - \bar{Y})^2}{\hat{\sigma}_{xy}^2 \left(\frac{1}{n} + \frac{1}{m}\right)} \sim Z^2, \tag{15}$$

where  $Z \sim N(0, 1)$ ,

$$\hat{\sigma}_{xy}^2 = \hat{\mu}_{xy} + \frac{\hat{\mu}_{xy}^2}{\hat{\theta}}, \quad \mu_{xy} = \frac{n\bar{X} + m\bar{Y}}{m + n}$$

and  $\hat{\theta}$  is the MLE of  $\theta$  based on the sample  $X_1, \dots, X_n$ . Noting that  $\hat{\mu}_{xy}(1/m + 1/n) = \bar{X}/m + \bar{Y}/n$ , we can write

$$\hat{\sigma}_{xy}^2 \left(\frac{1}{m} + \frac{1}{n}\right) = \left(\frac{\bar{X}}{m} + \frac{\bar{Y}}{n}\right) + \frac{mn}{(m+n)\hat{\theta}} \left(\frac{\bar{X}}{m} + \frac{\bar{Y}}{n}\right)^2.$$

Substituting the above expression in (15) and solving the equation

$$\frac{(\bar{X} - \bar{Y})^2}{\left(\frac{\bar{X}}{m} + \frac{\bar{Y}}{n}\right) + \frac{mn}{(m+n)\hat{\theta}} \left(\frac{\bar{X}}{m} + \frac{\bar{Y}}{n}\right)^2} = c_\alpha^2$$

for  $\bar{Y}$ , we can obtain a PI for  $\bar{Y}$ . Specifically, the two roots (with respect to  $\bar{Y}$ ) of the above equation form a  $100(1 - \alpha)\%$  PI for  $\bar{Y}$ . Letting

$$a = 1 - \frac{mc_\alpha^2}{n(m+n)\hat{\theta}}, \quad b = \left(1 + \frac{c_\alpha^2}{(m+n)\hat{\theta}}\right)\bar{X} + \frac{c_\alpha^2}{2n}, \quad \text{and}$$

$$c = \left(1 - \frac{nc_\alpha^2}{m(m+n)\hat{\theta}}\right)\bar{X}^2 - \frac{c_\alpha^2}{m}\bar{X},$$

the PI for  $\bar{Y}$  can be expressed as

$$\frac{b \mp \sqrt{b^2 - ac}}{a}. \tag{16}$$

The above PI is referred to as the joint sampling prediction interval (JS-PI). This joint sampling approach has been used to find PIs for binomial and Poisson distributions (Krishnamoorthy and Peng [7]) and CIs in a calibration problem (Brown [4]).

**Remark 3.1:** Letting  $m \rightarrow \infty$  in the above PIs in (12), (13), (14) and (16), it can be readily verified that the PIs simplify to their corresponding CIs for  $\mu$ .

## 4. Simulation studies

### 4.1. Coverage probabilities and precisions of the confidence intervals for the mean

To judge the accuracy and precision of the proposed CIs, we carried out extensive simulation studies along the lines of simulation study by Shilane et al. [13]. For our coverage and precision studies, we consider  $\mu = 3, 5, 10$  and  $20$ ,  $\theta = 0.5, 1, 4$  and  $10$ , and  $n = 10(10)90$ . Thus, we consider  $4 \times 4 \times 9 = 144$  sample size and parameter configurations and estimated coverage probabilities and precisions of the Wald, modified Wald, score and the likelihood confidence intervals at each of these configurations. The estimated values, based on 10,000 simulation runs, are reported in Table 1. Recall that a positive MLE for  $\theta$  is guaranteed provided  $s^2 > \bar{x}$ , and so in our simulation studies we discarded samples for which the mean  $\bar{x}$  is greater than the variance  $s^2$ .

We observe from the reported estimates in Table 1 that the score CI is satisfactory for all parameter configurations provided sample size is 30 or large. Even for small samples of size 10, the coverage probability is around 0.920 for most cases and they are accurate enough for practical purposes for samples of sizes 20 or more. The next satisfactory and simple CI is the modified Wald CI in (5). For  $\theta \geq 2$ , this CI controls the coverage probability close to the nominal level for all sample sizes and values of  $\mu$  considered for the study. This modified Wald CIs are as good as the score CIs except for small values of  $\theta$ . See the results for  $\theta = 0.5$ . We may prefer this simple modified Wald CI to others for  $n \geq 30$  and  $\theta \geq 2$ . The likelihood CIs and the Wald CIs are inferior to other CIs in terms of the coverage probability. In some cases, the likelihood CIs are shorter than others because of the smaller coverage probabilities than the nominal level.

We also estimated the coverage probabilities for large samples of sizes 100, 125, 150, 175 and 200,  $\theta = 0.5, 1, 4$  and  $10$  and  $\mu = 3$  and  $10$ . These estimated coverage probabilities along with the expected widths are given in Table 2. Examination of the estimated values in the table clearly indicates that the score CI controls the coverage probability very close to the nominal level 0.95 for all values of parameter and sample size configurations considered. Score CIs are slightly better than the likelihood CIs in terms of coverage probability. We once again see that the modified Wald CI performs satisfactorily except for  $\theta = 0.5$ . Even though the score CI and the M-Wald CI exhibit similar properties, the latter one is preferable for simplicity. Note that computation of the M-Wald CI requires only the sample mean, variance and standard normal percentiles. For large sample sizes and for  $\theta \geq 1$ , the M-Wald CIs are preferable to other CIs.

We have carried out extensive simulation studies including many values of  $\mu$ , but we here report only a few selected values, namely, 3, 5, 10 and 20. In general, we observed that the coverage probabilities of a CI is not much affected by the values of  $\mu$ . So our comparisons are valid for all  $\mu$ . However, our comparisons in the preceding paragraphs are valid only for  $\theta \geq 0.5$ . Other parameter configurations ( $\theta$  in the set  $\{0.025, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ ) considered in Shilane et al. [13] have been omitted because of some numerical complexities. In particular, for samples generated from  $NB(\mu, \theta)$  with small  $\theta < 0.5$ , we observed that the R function `g1m.nb(x = 1,`



**Table 1.** Coverage probabilities and (expected widths) of 95% CIs for the mean.

	Wald	M-Wald	Likelihood	Score	Wald	M-Wald	Likelihood	Score
$\mu = 3$								
$n$	$\theta = .5$				$\theta = 1$			
10	-	-	-	-	.872(3.92)	.918(5.00)	.871(3.99)	.873(5.04)
20	.882(3.75)	.901(4.17)	.923(7.87)	.890(3.92)	.896(2.90)	.920(3.22)	.902(2.95)	.933(3.77)
30	.899(3.12)	.915(3.34)	.937(4.54)	.909(3.24)	.920(2.40)	.934(2.57)	.923(2.43)	.938(2.81)
40	.910(2.74)	.920(2.88)	.941(3.53)	.918(2.81)	.926(2.09)	.938(2.20)	.931(2.11)	.940(2.35)
50	.915(2.46)	.922(2.56)	.942(3.00)	.921(2.51)	.928(1.88)	.936(1.96)	.933(1.90)	.945(2.06)
60	.919(2.25)	.926(2.33)	.943(2.65)	.926(2.30)	.930(1.72)	.937(1.78)	.933(1.73)	.945(1.86)
70	.925(2.09)	.931(2.15)	.949(2.40)	.929(2.13)	.935(1.60)	.940(1.64)	.938(1.61)	.947(1.70)
80	.928(1.96)	.933(2.01)	.946(2.21)	.933(1.99)	.936(1.50)	.940(1.53)	.938(1.50)	.948(1.58)
90	.929(1.85)	.933(1.89)	.948(2.05)	.934(1.88)	.937(1.41)	.943(1.44)	.940(1.42)	.949(1.48)
$n$	$\theta = 4$				$\theta = 10$			
10	.943(2.95)	.975(3.76)	.940(3.44)	.932(2.88)	.961(2.71)	.986(3.45)	.952(2.62)	.952(2.90)
20	.938(2.02)	.958(2.24)	.938(2.12)	.936(1.99)	.950(1.82)	.972(2.03)	.946(1.79)	.955(1.86)
30	.935(1.63)	.950(1.75)	.938(1.68)	.941(1.62)	.956(1.46)	.969(1.56)	.953(1.44)	.953(1.47)
40	.941(1.41)	.951(1.48)	.941(1.44)	.938(1.40)	.949(1.25)	.960(1.32)	.947(1.24)	.947(1.26)
50	.940(1.26)	.950(1.31)	.942(1.28)	.938(1.25)	.951(1.11)	.958(1.16)	.948(1.10)	.949(1.11)
60	.946(1.15)	.952(1.19)	.947(1.17)	.944(1.15)	.950(1.01)	.957(1.04)	.947(1.00)	.950(1.01)
70	.943(1.06)	.950(1.09)	.945(1.07)	.943(1.06)	.948(0.93)	.955(0.96)	.947(0.93)	.950(0.93)
80	.946(0.99)	.952(1.02)	.948(1.01)	.946(0.99)	.945(0.87)	.951(0.89)	.945(0.86)	.947(0.87)
90	.947(0.94)	.952(0.96)	.947(0.95)	.946(0.94)	.949(0.82)	.954(0.83)	.948(0.81)	.950(0.82)
$\mu = 5$								
$n$	$\theta = .5$				$\theta = 1$			
10	-	-	-	-	.878(6.30)	.922(8.03)	.877(6.37)	.891(14.4)
20	.883(6.09)	.904(6.77)	.923(10.6)	.892(6.33)	.905(4.55)	.925(5.07)	.910(4.63)	.930(5.89)
30	.898(5.04)	.915(5.40)	.940(7.25)	.910(5.21)	.917(3.82)	.934(4.09)	.920(3.86)	.938(4.45)
40	.907(4.42)	.916(4.65)	.938(5.68)	.915(4.54)	.925(3.31)	.935(3.48)	.929(3.34)	.941(3.71)
50	.915(3.98)	.924(4.14)	.942(4.85)	.922(4.07)	.932(2.98)	.942(3.10)	.935(3.01)	.941(3.25)
60	.918(3.66)	.926(3.78)	.950(4.30)	.928(3.73)	.931(2.73)	.940(2.82)	.937(2.75)	.944(2.93)
70	.925(3.41)	.930(3.51)	.945(3.88)	.929(3.46)	.940(2.53)	.946(2.61)	.942(2.55)	.945(2.70)
80	.929(3.18)	.933(3.26)	.941(3.58)	.933(3.23)	.940(2.37)	.945(2.43)	.941(2.39)	.944(2.50)
90	.932(3.01)	.936(3.07)	.943(3.34)	.935(3.04)	.940(2.23)	.945(2.28)	.942(2.24)	.948(2.35)
$n$	$\theta = 4$				$\theta = 10$			
10	.927(4.16)	.969(5.31)	.923(4.59)	.920(4.05)	.946(3.63)	.983(4.62)	.938(3.49)	.943(3.76)
20	.932(2.89)	.954(3.22)	.936(3.02)	.929(2.85)	.953(2.47)	.972(2.75)	.949(2.42)	.950(2.48)
30	.934(2.37)	.951(2.54)	.939(2.43)	.932(2.35)	.943(1.98)	.956(2.12)	.940(1.96)	.945(1.99)
40	.936(2.05)	.947(2.16)	.939(2.10)	.934(2.04)	.950(1.70)	.961(1.79)	.949(1.69)	.943(1.71)
50	.939(1.84)	.947(1.92)	.945(1.86)	.938(1.83)	.949(1.52)	.958(1.58)	.947(1.51)	.948(1.52)
60	.942(1.68)	.950(1.74)	.941(1.70)	.941(1.67)	.947(1.38)	.954(1.43)	.945(1.37)	.946(1.39)
70	.943(1.56)	.949(1.60)	.943(1.58)	.941(1.55)	.946(1.28)	.952(1.31)	.943(1.27)	.944(1.28)
80	.942(1.46)	.947(1.50)	.945(1.47)	.942(1.45)	.942(1.19)	.949(1.22)	.941(1.19)	.944(1.19)
90	.941(1.38)	.947(1.41)	.949(1.39)	.940(1.37)	.946(1.13)	.952(1.15)	.945(1.12)	.947(1.13)
	Wald	M-Wald	Likelihood	Score	Wald	M-Wald	Likelihood	Score
$\mu = 10$								
$n$	$\theta = .5$				$\theta = 1$			
10	-	-	-	-	.875(12.0)	.922(15.4)	.875(12.1)	.896(20.7)
20	.880(11.8)	.901(13.1)	.884(12.3)	.928(19.1)	.909(8.76)	.928(9.75)	.915(8.91)	.941(11.2)
30	.903(9.92)	.917(10.6)	.914(10.2)	.942(14.1)	.920(7.31)	.935(7.83)	.926(7.38)	.943(8.51)
40	.909(8.66)	.920(9.11)	.918(8.88)	.946(11.0)	.926(6.36)	.936(6.69)	.931(6.41)	.942(7.11)
50	.920(7.80)	.929(8.12)	.930(7.98)	.945(9.47)	.931(5.70)	.940(5.93)	.936(5.75)	.943(6.24)
60	.920(7.16)	.927(7.40)	.928(7.28)	.946(8.37)	.935(5.23)	.942(5.40)	.938(5.27)	.946(5.63)
70	.924(6.64)	.931(6.83)	.931(6.75)	.947(7.59)	.933(4.84)	.939(4.98)	.938(4.87)	.946(5.16)
80	.930(6.23)	.936(6.39)	.941(6.32)	.946(7.00)	.938(4.53)	.943(4.64)	.942(4.55)	.948(4.78)
90	.926(5.86)	.931(5.99)	.932(5.94)	.947(6.50)	.940(4.28)	.945(4.38)	.940(4.30)	.948(4.49)
$n$	$\theta = 4$				$\theta = 10$			
10	.912(7.14)	.961(9.09)	.905(6.92)	.915(7.75)	.939(5.65)	.978(7.20)	.929(5.42)	.932(5.69)
20	.923(5.08)	.947(5.66)	.921(5.01)	.928(5.27)	.935(3.90)	.957(4.34)	.929(3.82)	.931(3.91)
30	.932(4.18)	.949(4.48)	.930(4.15)	.936(4.29)	.939(3.17)	.953(3.39)	.937(3.13)	.937(3.17)
40	.939(3.62)	.950(3.81)	.939(3.59)	.942(3.68)	.943(2.75)	.954(2.89)	.941(2.72)	.942(2.75)
50	.943(3.24)	.951(3.37)	.943(3.22)	.945(3.29)	.939(2.46)	.949(2.56)	.938(2.44)	.940(2.46)

(continued)

**Table 1.** Continued.

	Wald	M-Wald	Likelihood	Score	Wald	M-Wald	Likelihood	Score
60	.939(2.97)	.946(3.07)	.938(2.95)	.945(3.00)	.945(2.25)	.952(2.33)	.943(2.23)	.946(2.25)
70	.948(2.75)	.954(2.83)	.948(2.74)	.948(2.78)	.945(2.08)	.950(2.14)	.943(2.07)	.945(2.08)
80	.944(2.57)	.949(2.64)	.943(2.57)	.944(2.60)	.946(1.95)	.950(2.00)	.945(1.94)	.946(1.95)
90	.945(2.42)	.949(2.48)	.945(2.41)	.946(2.44)	.946(1.84)	.952(1.88)	.945(1.83)	.945(1.84)
$\mu = 20$								
$n$			$\theta = .5$				$\theta = 1$	
10	-	-	-	-	.875(23.5)	.923(29.9)	.898(42.2)	.870(23.5)
20	.883(23.3)	.905(25.9)	.892(24.3)	.933(42.7)	.906(17.2)	.928(19.1)	.909(17.4)	.930(21.8)
30	.901(19.4)	.916(20.8)	.910(20.1)	.941(27.4)	.917(14.2)	.931(15.2)	.921(14.4)	.943(16.5)
40	.915(17.1)	.927(18.0)	.924(17.6)	.946(21.9)	.926(12.3)	.936(13.0)	.933(12.5)	.940(13.8)
50	.918(15.3)	.926(16.0)	.925(15.7)	.947(18.6)	.931(11.1)	.939(11.5)	.934(11.2)	.948(12.1)
60	.921(14.1)	.927(14.6)	.929(14.4)	.945(16.5)	.932(10.2)	.939(10.5)	.936(10.2)	.946(10.9)
70	.927(13.0)	.932(13.4)	.935(13.4)	.952(15.0)	.940(9.47)	.944(9.74)	.940(9.52)	.947(10.0)
80	.926(12.2)	.932(12.5)	.932(12.4)	.949(13.7)	.937(8.86)	.943(9.08)	.939(8.90)	.948(9.35)
90	.930(11.6)	.934(11.9)	.936(11.7)	.948(12.8)	.939(8.37)	.943(8.56)	.942(8.41)	.948(8.79)
$n$			$\theta = 4$				$\theta = 10$	
10	.905(12.9)	.956(16.5)	.896(12.5)	.906(13.8)	.923(9.46)	.968(12.0)	.910(9.05)	.914(9.44)
20	.921(9.38)	.944(10.4)	.919(9.24)	.929(9.71)	.931(6.71)	.953(7.46)	.925(6.57)	.926(6.71)
30	.930(7.71)	.944(8.26)	.929(7.65)	.938(7.90)	.941(5.48)	.955(5.87)	.938(5.41)	.940(5.48)
40	.937(6.70)	.949(7.04)	.936(6.65)	.940(6.82)	.938(4.76)	.949(5.01)	.937(4.71)	.938(4.76)
50	.938(6.02)	.948(6.26)	.939(5.98)	.943(6.10)	.945(4.26)	.955(4.43)	.944(4.23)	.945(4.26)
60	.941(5.51)	.947(5.70)	.939(5.48)	.941(5.57)	.941(3.90)	.949(4.03)	.940(3.87)	.940(3.90)
70	.942(5.10)	.948(5.24)	.944(5.08)	.945(5.15)	.941(3.60)	.948(3.71)	.939(3.58)	.943(3.60)
80	.945(4.77)	.950(4.89)	.945(4.75)	.945(4.81)	.946(3.38)	.951(3.46)	.945(3.36)	.945(3.37)
90	.949(4.50)	.954(4.60)	.948(4.49)	.948(4.54)	.947(3.19)	.952(3.26)	.947(3.17)	.945(3.18)

link = identity) \$theta does not converge within the maximum number of iterations and so the returned values may not be accurate or the true MLEs. The R function and our own R code based on the Newton-Raphson method all produce similar errors for small values of  $\theta$ . Notice that Shilane et al. [13] were able to compute the coverage probabilities of the Wald, gamma and other CIs, because they are functions of sample mean and variance, not functions of the MLEs. Furthermore, the reported coverage probabilities of these CIs are very poor. The coverage probabilities are ranging from 0.20 to 1.00 when the nominal level is 0.95. These coverage studies indicate that the CIs considered in their paper are very unsatisfactory when  $\theta$  is small.

We also noted that simulation studies by Shilane et al. [13] include samples with all zeros. As in real application, a two-parameter NB distribution is postulated only when the sample variance  $s^2$  is larger than the mean  $\bar{x}$ , in our simulation studies we have included only samples for which  $s^2 > \bar{x}$ . To check the performance of the Wald CIs for small values of  $\theta$ , we estimated the coverage probabilities of the Wald and M-Wald CIs and reported them in Table 3. Coverage probabilities of both CIs are much less than the nominal level for small values of  $\theta$ . This is true even for large samples. These two CIs are not useful in situations where we expect  $\theta$  to be small.

On an overall basis, we recommend the score CIs for all practical situations where  $\theta \geq 0.5$  and the sample size is 20 or more. For other cases, the score CIs may be less satisfactory. For samples, say  $n < 20$ , if a researcher has evidence to believe that  $\theta > 2$ , then the Wald or M-Wald CIs can also be used.

**Table 2.** Coverage probabilities and (expected widths) of 95% CIs for the mean (large samples).

	Wald	M-Wald	Likelihood	Score	Wald	M-Wald-1	Likelihood	Score
$\mu = 3$								
$n$	$\theta = .5$				$\theta = 1$			
100	.927(1.76)	.932(1.80)	.932(1.79)	.946(1.94)	.937(1.34)	.942(1.37)	.938(1.34)	.947(1.40)
125	.936(1.58)	.940(1.61)	.939(1.60)	.948(1.71)	.938(1.20)	.942(1.22)	.941(1.20)	.948(1.24)
150	.937(1.45)	.940(1.47)	.941(1.46)	.947(1.54)	.943(1.10)	.945(1.11)	.944(1.10)	.948(1.13)
175	.940(1.34)	.943(1.36)	.944(1.35)	.947(1.41)	.949(1.02)	.951(1.03)	.948(1.02)	.948(1.04)
200	.940(1.25)	.941(1.26)	.942(1.26)	.951(1.31)	.942(0.95)	.950(0.96)	.942(0.95)	.951(0.97)
$n$	$\theta = 4$				$\theta = 10$			
100	.943(0.89)	.948(.911)	.942(.899)	.946(.901)	.948(.778)	.952(.793)	.947(.775)	.945(.779)
125	.947(.801)	.950(.813)	.946(.799)	.946(.806)	.947(.693)	.950(.704)	.947(.691)	.948(.694)
150	.942(.731)	.945(.741)	.942(.730)	.944(.735)	.949(.632)	.952(.641)	.949(.631)	.948(.633)
175	.945(.677)	.948(.685)	.945(.676)	.948(.680)	.951(.585)	.953(.591)	.951(.584)	.951(.585)
200	.948(.633)	.949(.639)	.947(.633)	.948(.636)	.945(.546)	.949(.552)	.945(.545)	.945(.547)
$\mu = 10$								
$n$	$\theta = .5$				$\theta = 1$			
100	.936(5.59)	.941(5.70)	.942(5.66)	.950(6.14)	.939(4.07)	.943(4.16)	.942(4.09)	.951(4.25)
125	.935(5.02)	.938(5.10)	.940(5.07)	.946(5.40)	.943(3.64)	.945(3.70)	.948(3.77)	.945(3.66)
150	.934(4.58)	.937(4.64)	.938(4.63)	.947(4.88)	.942(3.33)	.945(3.37)	.950(3.43)	.948(3.34)
175	.937(4.24)	.939(4.29)	.942(4.27)	.951(4.47)	.946(3.08)	.948(3.12)	.948(3.16)	.946(3.09)
200	.940(3.98)	.942(4.01)	.943(4.00)	.949(4.16)	.944(2.89)	.946(2.92)	.947(2.95)	.945(2.90)
$n$	$\theta = 4$				$\theta = 10$			
100	.945(2.30)	.950(2.35)	.946(2.30)	.948(2.32)	.948(1.74)	.952(1.78)	.947(1.74)	.947(1.74)
125	.947(2.06)	.951(2.10)	.947(2.06)	.950(2.08)	.947(1.56)	.951(1.59)	.947(1.56)	.946(1.56)
150	.945(1.88)	.948(1.91)	.945(1.88)	.947(1.89)	.946(1.42)	.948(1.44)	.945(1.42)	.946(1.42)
175	.952(1.74)	.953(1.76)	.952(1.74)	.950(1.75)	.948(1.32)	.951(1.33)	.948(1.32)	.947(1.32)
200	.947(1.63)	.950(1.65)	.947(1.63)	.948(1.64)	.949(1.23)	.951(1.24)	.948(1.23)	.949(1.23)

**Table 3.** Coverage probabilities of 95% CIs for the mean for small values of  $\theta$ .

$\mu = 5$								
$n$	$\theta = .025$		$\theta = .075$		$\theta = .20$		$\theta = .40$	
	Wald	M-Wald	Wald	M-Wald	Wald	M-Wald	Wald	M-Wald
10	.625	.650	.665	.699	.758	.801	.812	.858
20	.600	.613	.724	.741	.815	.837	.868	.890
30	.634	.642	.763	.775	.854	.866	.894	.909
40	.654	.662	.793	.802	.865	.877	.904	.914
50	.690	.697	.814	.822	.887	.895	.911	.919
60	.713	.718	.830	.836	.892	.900	.918	.925
70	.741	.746	.850	.856	.897	.904	.920	.926
80	.748	.752	.852	.857	.909	.914	.922	.928
90	.756	.760	.866	.871	.908	.913	.924	.927

**4.2. Coverage probabilities and precisions of the prediction intervals**

All PIs are simple to compute, and as noted earlier, they all simplify to corresponding CIs for large  $m$ , and so performances of PIs for large  $m$  should be similar to those of the CIs in Section 4.1. To understand the performance of the proposed PIs for small to moderate values of  $m$ , we estimated their coverage probabilities and expected widths and presented them in Table 4. We estimated the coverage probabilities and precisions of all PIs for sample sizes ranging from 20 to 90, future sample sizes ranging from 5 to 50,  $\mu = 3, 7$  and 12, and  $\theta = 0.5, 1, 4$  and 10. All estimates are based on 10,000 simulation runs. We observe from Table 4 that the Wald, M-Wald and likelihood PIs are very similar in most cases, and the M-Wald PIs are slightly better than the other two. We once again note that the M-Wald

**Table 4.** Coverage probabilities and (expected widths) of 95% PIs for the mean of a future sample size  $m$ .

		Wald	M-Wald	Likelihood	JS-PI	Wald	M-Wald	Likelihood	JS-PI
$\mu = 3$									
		$\theta = .5$				$\theta = 1$			
$n$	$m$								
20	5	.913(8.39)	.916(8.56)	.920(8.77)	.948(9.58)	.926(6.52)	.928(6.65)	.927(6.61)	.942(6.89)
30	5	.925(8.31)	.926(8.39)	.929(8.58)	.950(8.92)	.931(6.37)	.933(6.43)	.933(6.44)	.944(6.57)
40	10	.920(6.08)	.922(6.14)	.929(6.25)	.950(6.51)	.937(4.68)	.940(4.72)	.942(4.73)	.951(4.82)
50	30	.922(4.03)	.925(4.09)	.929(4.11)	.943(4.38)	.929(3.07)	.932(3.11)	.936(3.09)	.941(3.19)
70	20	.935(4.44)	.937(4.47)	.943(4.52)	.948(4.64)	.938(3.38)	.940(3.40)	.941(3.41)	.947(3.45)
90	50	.934(3.11)	.936(3.14)	.941(3.15)	.949(3.26)	.940(2.37)	.942(2.39)	.942(2.38)	.946(2.42)
$\theta = 4$									
$n$	$m$								
20	5	.929(4.38)	.934(4.47)	.927(4.35)	.935(4.40)	.938(3.80)	.942(3.87)	.945(3.83)	.948(3.86)
30	5	.939(4.28)	.941(4.32)	.937(4.25)	.942(4.27)	.942(3.69)	.946(3.73)	.946(3.71)	.949(3.72)
40	10	.938(3.13)	.941(3.16)	.937(3.11)	.941(3.13)	.940(2.71)	.943(2.73)	.943(2.71)	.944(2.72)
50	30	.938(2.05)	.941(2.08)	.938(2.04)	.941(2.06)	.947(1.77)	.950(1.80)	.947(1.77)	.949(1.78)
70	20	.944(2.26)	.946(2.27)	.944(2.25)	.946(2.26)	.945(1.95)	.946(1.96)	.945(1.94)	.946(1.95)
90	50	.942(1.57)	.944(1.59)	.942(1.57)	.940(1.58)	.945(1.36)	.946(1.37)	.945(1.35)	.944(1.36)
$\mu = 7$									
		$\theta = .5$				$\theta = 1$			
$n$	$m$								
20	5	.909(18.78)	.911(19.1)	.914(19.5)	.948(21.3)	.929(14.0)	.931(14.3)	.931(14.2)	.943(14.8)
30	5	.923(18.46)	.924(18.6)	.927(19.0)	.952(19.8)	.928(13.7)	.929(13.8)	.934(13.9)	.944(14.1)
40	10	.925(13.69)	.927(13.8)	.933(14.0)	.952(14.6)	.937(10.1)	.939(10.2)	.940(10.2)	.951(10.4)
50	30	.925(8.96)	.929(9.1)	.934(9.17)	.952(9.74)	.936(6.65)	.940(6.75)	.940(6.69)	.947(6.89)
70	20	.936(9.96)	.938(10.0)	.942(10.1)	.952(10.3)	.941(7.35)	.942(7.39)	.944(7.39)	.949(7.48)
90	50	.935(6.95)	.937(7.0)	.942(7.04)	.949(7.27)	.939(5.11)	.941(5.15)	.942(5.13)	.946(5.21)
$\theta = 4$									
$n$	$m$								
20	5	.935(8.42)	.939(8.59)	.931(8.30)	.935(8.39)	.936(6.67)	.940(6.80)	.934(6.58)	.934(6.61)
30	5	.937(8.19)	.939(8.27)	.935(8.12)	.939(8.16)	.942(6.45)	.944(6.51)	.939(6.38)	.939(6.39)
40	10	.939(6.01)	.941(6.07)	.938(5.97)	.941(6.00)	.942(4.73)	.943(4.78)	.940(4.69)	.940(4.70)
50	30	.944(3.93)	.947(3.98)	.943(3.91)	.944(3.94)	.943(3.09)	.946(3.14)	.941(3.07)	.941(3.08)
70	20	.942(4.32)	.944(4.35)	.943(4.31)	.945(4.32)	.942(3.41)	.945(3.43)	.942(3.39)	.944(3.39)
90	50	.944(3.01)	.946(3.04)	.944(3.01)	.943(3.02)	.943(2.37)	.945(2.39)	.942(2.36)	.943(2.37)
$\mu = 12$									
		$\theta = .5$				$\theta = 1$			
$n$	$m$								
20	5	.909(31.6)	.912(32.3)	.915(33.1)	.947(36.0)	.924(23.3)	.926(23.7)	.927(23.7)	.943(24.6)
30	5	.919(31.3)	.920(31.5)	.925(32.5)	.948(33.6)	.938(22.9)	.940(23.1)	.942(23.1)	.952(23.6)
40	10	.930(22.9)	.931(23.1)	.935(23.6)	.954(24.6)	.936(16.9)	.938(17.0)	.941(17.0)	.948(17.4)
50	30	.927(15.1)	.930(15.3)	.934(15.5)	.947(16.4)	.928(11.0)	.931(11.2)	.935(11.1)	.942(11.4)
70	20	.932(16.8)	.934(16.9)	.940(17.1)	.952(17.5)	.935(12.2)	.936(12.3)	.939(12.3)	.945(12.4)
90	50	.938(11.7)	.940(11.8)	.943(12.0)	.951(12.3)	.939(8.55)	.941(8.62)	.944(8.59)	.948(8.72)
$\theta = 4$									
$n$	$m$								
20	5	.928(13.2)	.933(13.5)	.926(13.0)	.929(13.2)	.934(9.90)	.938(10.1)	.930(9.72)	.931(9.76)
30	5	.941(12.9)	.942(13.0)	.940(12.8)	.942(12.8)	.942(9.63)	.944(9.72)	.939(9.50)	.940(9.52)
40	10	.944(9.54)	.946(9.63)	.943(9.47)	.942(9.51)	.942(7.06)	.944(7.13)	.940(6.99)	.939(7.00)
50	30	.944(6.23)	.948(6.32)	.945(6.19)	.945(6.23)	.945(4.61)	.949(4.67)	.943(4.57)	.943(4.58)
70	20	.944(6.83)	.945(6.87)	.944(6.80)	.946(6.82)	.941(5.08)	.942(5.11)	.940(5.05)	.941(5.05)
90	50	.947(4.77)	.949(4.80)	.947(4.75)	.947(4.77)	.943(3.54)	.944(3.57)	.941(3.53)	.943(3.53)

PIs are easy to compute as it requires only the mean and variance of the sample. Among all four PIs, the PI based on the joint sampling approach (JS-PI) performs very satisfactorily by controlling the coverage probability close to the nominal level for all sample size and parameter configurations.

In Table 5, we report the coverage probabilities and expected widths of PIs for large sample sizes 100(25)200. It is clear from the reported values that accuracies of all methods increase with increasing sample size. However, for small values of  $\theta$ , the coverage probabilities of all PIs, except JS-PI, could slightly lower than the nominal level 0.95. If coverage probabilities of all PIs are very close to the nominal level, then their expected

**Table 5.** Coverage probabilities and (expected widths) of 95% PIs for the mean of a future sample size  $m$  (large sample size  $n$ ).

		Wald	M-Wald	Likelihood	JS-PI	Wald	M-Wald	Likelihood	JS-PI
$\mu = 3$									
		$\theta = .5$				$\theta = 1$			
$n$	$m$								
100	5	.942(8.13)	.942(8.14)	.943(8.22)	.951(8.25)	.949(6.15)	.950(6.15)	.951(6.17)	.956(6.18)
125	10	.945(5.81)	.945(5.82)	.949(5.88)	.951(5.91)	.947(4.43)	.947(4.43)	.948(4.44)	.950(4.45)
150	10	.948(5.80)	.948(5.80)	.950(5.85)	.952(5.87)	.950(4.40)	.950(4.40)	.952(4.41)	.951(4.41)
175	20	.944(4.19)	.944(4.20)	.947(4.22)	.951(4.24)	.948(3.19)	.948(3.19)	.948(3.19)	.949(3.20)
200	20	.946(4.17)	.947(4.17)	.948(4.19)	.953(4.21)	.944(3.16)	.944(3.16)	.946(3.17)	.948(3.17)
		$\theta = 4$				$\theta = 10$			
$n$	$m$								
100	5	.947(4.09)	.947(4.09)	.946(4.08)	.947(4.08)	.947(3.54)	.947(3.54)	.947(3.53)	.948(3.53)
125	10	.949(2.93)	.950(2.94)	.949(2.93)	.950(2.93)	.950(2.53)	.951(2.53)	.950(2.53)	.951(2.53)
150	10	.951(2.92)	.951(2.93)	.951(2.92)	.951(2.92)	.950(2.52)	.950(2.52)	.949(2.51)	.948(2.51)
175	20	.950(2.11)	.950(2.12)	.950(2.11)	.950(2.11)	.948(1.82)	.948(1.82)	.947(1.82)	.947(1.82)
200	20	.952(2.10)	.952(2.10)	.952(2.09)	.952(2.09)	.951(1.81)	.951(1.81)	.952(1.81)	.951(1.81)
$\mu = 7$									
		$\theta = .5$				$\theta = 1$			
$n$	$m$								
100	5	.940(18.1)	.940(18.11)	.942(18.3)	.950(18.4)	.954(13.3)	.954(13.3)	.954(13.3)	.957(13.4)
125	10	.946(13.0)	.946(13.05)	.950(13.1)	.957(13.3)	.943(9.56)	.943(9.57)	.944(9.60)	.948(9.62)
150	10	.947(12.9)	.947(12.99)	.951(13.0)	.958(13.3)	.951(9.52)	.952(9.53)	.952(9.55)	.954(9.57)
175	20	.944(9.38)	.944(9.39)	.947(9.44)	.951(9.49)	.944(6.89)	.944(6.89)	.947(6.91)	.949(6.92)
200	20	.947(9.33)	.947(9.34)	.950(9.38)	.953(9.42)	.951(6.84)	.951(6.85)	.952(6.86)	.954(6.87)
		$\theta = 4$				$\theta = 10$			
$n$	$m$								
100	5	.946(7.85)	.947(7.85)	.947(7.83)	.948(7.83)	.940(6.16)	.941(6.17)	.940(6.14)	.941(6.14)
125	10	.951(5.62)	.951(5.63)	.951(5.61)	.952(5.62)	.951(4.43)	.951(4.44)	.950(4.42)	.950(4.42)
150	10	.948(5.60)	.948(5.60)	.949(5.59)	.949(5.59)	.950(4.40)	.950(4.40)	.949(4.39)	.950(4.39)
175	20	.950(4.04)	.950(4.05)	.949(4.04)	.950(4.04)	.949(3.18)	.949(3.18)	.949(3.17)	.949(3.17)
200	20	.945(4.02)	.945(4.03)	.946(4.02)	.947(4.02)	.951(3.16)	.951(3.17)	.950(3.16)	.950(3.16)
$\mu = 12$									
		$\theta = .5$				$\theta = 1$			
$n$	$m$								
100	5	.941(30.6)	.941(30.7)	.941(31.0)	.949(31.1)	.949(22.2)	.949(22.2)	.951(22.3)	.956(22.3)
125	10	.944(22.0)	.944(22.0)	.949(22.2)	.954(22.3)	.945(15.9)	.945(15.9)	.947(15.9)	.949(16.0)
150	10	.948(21.9)	.949(21.9)	.951(22.1)	.956(22.2)	.949(15.9)	.949(15.9)	.952(15.9)	.954(15.9)
175	20	.943(15.8)	.943(15.8)	.944(15.9)	.950(16.0)	.945(11.4)	.945(11.5)	.946(11.5)	.948(11.5)
200	20	.952(15.7)	.952(15.7)	.954(15.8)	.957(15.9)	.949(11.4)	.949(11.4)	.949(11.4)	.951(11.4)
		$\theta = 4$				$\theta = 10$			
$n$	$m$								
100	5	.946(12.3)	.947(12.4)	.947(12.3)	.949(12.3)	.947(9.19)	.947(9.20)	.945(9.16)	.945(9.16)
125	10	.947(8.89)	.947(8.90)	.946(8.87)	.947(8.87)	.945(6.60)	.946(6.61)	.944(6.58)	.944(6.58)
150	10	.944(8.84)	.945(8.84)	.944(8.82)	.945(8.83)	.948(6.56)	.948(6.56)	.947(6.54)	.948(6.54)
175	20	.950(6.39)	.951(6.40)	.950(6.38)	.950(6.39)	.947(4.74)	.948(4.75)	.946(4.73)	.948(4.73)
200	20	.949(6.35)	.949(6.35)	.949(6.34)	.950(6.34)	.948(4.71)	.948(4.71)	.948(4.70)	.948(4.70)

widths are very close to each other. On an overall basis, the JS-PI can be recommended for applications.

## 5. Examples

**Example 5.1:** This example is arising from the analysis of traffic flow data in an internet communications network. The data include packet counts at each of  $n = 102$  consecutive seconds. Sanchez and He [15] and Shilane et al. [13] used a two-parameter negative binomial model to estimate the mean packet counts per second. It has been noted that packets arrive according to a Poisson process with over dispersion and so a negative binomial model was postulated for the analysis. The sample mean and standard deviation of the data are  $\bar{x} = 310.31$  and  $s = 94.54$ , respectively. Shilane et al. [13] have used the `glm.nb` method in R to compute the MLE  $\hat{\theta} = 10.59$  with a standard error of 1.52.

**Table 6.** Confidence intervals and prediction intervals and their [widths] for traffic data.

Method	Confidence intervals for the mean packet counts per second		
	90% CI	95% CI	99% CI
Wald	(294.9, 325.7) [30.8]	(292.0, 328.7) [36.7]	(286.2, 334.4) [48.2]
M-Wald	(294.7, 325.9) [31.2]	(291.6, 329.0) [37.4]	(285.4, 335.2) [49.9]
Likelihood	(294.5, 326.1) [31.6]	(291.5, 329.1) [37.6]	(285.6, 335.0) [49.5]
Score	(295.3, 326.9) [31.7]	(292.6, 330.3) [37.8]	(287.4, 337.1) [49.8]
Method	Prediction intervals for the mean packet counts for a sample of 30 seconds		
	90% PI	95% PI	99% PI
Wald	(278.0, 342.6) [64.6]	(271.8, 348.8) [77.0]	(259.7, 360.9) [101]
M-Wald	(277.9, 342.7) [64.8]	(271.7, 349.0) [77.3]	(259.4, 361.3) [102]
Likelihood	(277.2, 343.4) [66.3]	(270.8, 349.8) [79.0]	(258.4, 362.2) [104]
JS-PI	(278.0, 344.2) [66.3]	(271.9, 350.9) [79.0]	(260.3, 364.2) [104]

**Table 7.** Confidence intervals and prediction intervals and their [widths] for tick data.

Method	Confidence intervals for the mean number of ticks per sheep		
	90% CI	95% CI	99% CI
Wald	(5.490, 7.632) [2.14]	(5.285, 7.837) [2.55]	(4.884, 8.238) [3.35]
M-Wald	(5.472, 7.650) [2.18]	(5.254, 7.868) [2.61]	(4.811, 8.311) [3.50]
Likelihood	(5.550, 7.569) [2.02]	(5.360, 7.762) [2.40]	(4.982, 8.139) [3.16]
Score	(5.675, 7.729) [2.05]	(5.529, 7.996) [2.47]	(5.262, 8.570) [3.31]
Method	Prediction intervals for the mean number of ticks for a sample of 10 sheep		
	90% PI	95% PI	99% PI
Wald	(3.312, 9.810) [6.50]	(2.690, 10.43) [7.74]	(1.474, 11.65) [10.2]
M-Wald	(3.306, 9.815) [6.51]	(2.680, 10.44) [7.76]	(1.451, 11.67) [10.2]
Likelihood	(3.504, 9.618) [6.11]	(2.919, 10.20) [7.28]	(1.774, 11.35) [9.57]
JS-PI	(3.637, 9.762) [6.13]	(3.105, 10.41) [7.31]	(2.091, 11.71) [9.62]

CIs for the mean packet count per second based on different methods are given in Table 6. All the methods produced CIs that are quite similar because of the large sample size.

We also computed PIs for the mean of a future sample of size 30 and reported them in Table 6. Note that the Wald and M-Wald PIs are practically the same. The likelihood PI and the JS-PI are quite similar and they are slightly wider than the Wald and M-Wald PIs.

**Example 5.2:** The number of ticks  $x$  was counted on each of a sample of 82 sheep and the following frequency table, taken from Ross and Preece [12], was obtained:

$x$	0	1	2	3	4	5	6	7	8	9	10	11	12
$f$	4	5	11	10	9	11	3	5	3	2	2	5	0
$x$	13	14	15	16	17	18	19	20	21	22	23	24	25
$f$	2	2	1	1	0	0	1	0	1	1	1	0	2

For these data, the mean  $\bar{x} = 6.5610$ , the variance  $s^2 = 34.7678$  and the MLE  $\hat{\theta} = 1.7775$ . Note that the variance is much larger than the mean which indicates that the data is over-dispersed. Furthermore,  $n = \sum_{i=0}^{25} f_i = 82$ . We computed 90, 95 and 99% CIs for the mean and PIs for the mean of a future sample of size  $m = 10$  and reported them in Table 7.

We see in Table 7 that the Wald and M-Wald CIs are similar to some extent, and the score CI and the likelihood CI are somewhat similar. The PIs also exhibit similar results. In particular, the JS-PIs and the likelihood PIs are practically the same and they are narrower than the Wald and M-Wald PIs.

## 6. Concluding remarks

We have provided CIs for the mean and PIs for the mean of a future sample based on the MLE of  $\theta$  which can be computed provided the sample variance is greater than the sample mean. Of course, if a sample data does not satisfy this condition, then a NB model may not be postulated to analyze such count data. The score CI for the mean and the PI based on the joint sampling approach are very satisfactory having good coverage probabilities. The score CI and the JS-PI that we considered are simple to compute using the MLE of  $\theta$ , which can be computed using the R function `glm.nb`. We also noted that a simple modified Wald CI, which is comparable with the score CI for  $\theta \geq 2$ , can be easily computed using a scientific calculator. Our study has also revealed that the Wald CIs and M-Wald CIs are not satisfactory if  $\theta < 0.5$ . In general, no CI is satisfactory when  $\theta < 0.5$  and further research is needed to find satisfactory CIs for the mean when  $\theta$  is small.

We provided PIs that are also simple to compute, yet very satisfactory for sample sizes not too small. The prediction interval based on the joint sampling approach is easy to compute and very satisfactory for applications when sample sizes are moderate to large.

## Acknowledgments

The authors are grateful to two reviewers for providing useful comments and suggestions that enhanced earlier versions of this paper.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

K. Krishnamoorthy  <http://orcid.org/0000-0002-3919-918X>

## References

- [1] F.J. Anscombe, *The statistical analysis of insect counts based on the negative binomial distribution*, *Biometrics* 5 (1949), pp. 165–173.
- [2] J. Aragon, D. Eberly, and S. Eberly, *Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution*, *Stat. Probab. Lett.* 15 (1992), pp. 375–379.
- [3] U. Bandara, G. Ryan, and M. Riten, *On computing maximum likelihood estimates for the negative binomial distribution*, *Stat. Probab. Lett.* 148 (2019), pp. 54–58.
- [4] P.J. Brown, *Multivariate calibration*, *J. R. Stat. Soc. Ser. B* 44 (1982), pp. 287–321.
- [5] H. Dai, Y. Bao, and M. Bao, *Maximum likelihood estimate for the dispersion parameter of the negative binomial distribution*, *Stat. Probab. Lett.* 83 (2013), pp. 21–27.
- [6] B-A. Dang and K. Krishnamoorthy, *Confidence intervals, prediction intervals and tolerance intervals for negative binomial distributions*, *Stat. Pap.* 63 (2022), pp. 795–820.
- [7] K. Krishnamoorthy and J. Peng, *Improved closed-form prediction intervals for binomial and Poisson distributions*, *J. Stat. Plan. Inference* 141 (2011), pp. 1709–1718.
- [8] A. Khurshid, M.I. Ageel, and R. Lodhi, *On confidence intervals for the negative binomial distribution*, *Rev. Invest. Oper.* 26 (2005), pp. 59–70.
- [9] J. Mazucheli, W. Bertoli, and R. Oliveira, *Two useful discrete distributions to model overdispersed count data*, *Rev. Colomb. Estad.* 43 (2020), pp. 21–48.
- [10] W. Nelson, *Applied Life Data Analysis*, Wiley, Hoboken, NJ, 1982.

- [11] B-J Park and D. Lord, *Adjustment for maximum likelihood estimate of negative binomial dispersion parameter*. Transportation Research Record: Journal of the Transportation Research Board Vol. 2061, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 9–19.
- [12] G.J.S. Ross and D.A. Preece, *The negative binomial distribution*, J. R. Stat. Soc. Ser. D Stat. 34 (1985), pp. 323–335.
- [13] D. Shilane, S.N. Evans, and A.E. Hubbard, *Confidence intervals for negative binomial random variables of high dispersion*, Int. J. Biostat. 6 (2010), Article 10, pp. 1–10.
- [14] R.L. Sheaffer and R.S. Leavenworth, *The negative binomial model for counts in units of varying size*, J. Qual. Technol. 8 (1976), pp. 158–163.
- [15] J. Sanchez and Y. He, *Internet data analysis for the undergraduate statistics curriculum*, J. Stat. Educ. 13 (2005), <https://doi.org/10.1080/10691898.2005.11910568>.
- [16] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Trans. Am. Math. Soc. 54 (1943), pp. 426–482.
- [17] L.J. Wilson, J.L. Folks, and J.H. Young, *Complete sufficiency and maximum likelihood estimation for the two-parameter negative binomial distribution*, Metrika 33 (1986), pp. 349–362.
- [18] G.R. Wood, *Confidence and prediction intervals for generalized linear accident models*, Accid. Anal. Prev. 37 (2005), pp. 267–273.