2002

Bayesian Inference for the Negative Binomial Distribution via Polynomial Expansions

Eric T. Bradlow
University of Pennsylvania

Bruce G. S Hardie

Peter S. Fader
University of Pennsylvania

Bayesian Inference for the Negative Binomial Distribution via Polynomial Expansions

Abstract
To date, Bayesian inferences for the negative binomial distribution (NBD) have relied on computationally intensive numerical methods (e.g., Markov chain Monte Carlo) as it is thought that the posterior densities of interest are not amenable to closed-form integration. In this article, we present a "closed-form" solution to the Bayesian inference problem for the NBD that can be written as a sum of polynomial terms. The key insight is to approximate the ratio of two gamma functions using a polynomial expansion, which then allows for the use of a conjugate prior. Given this approximation, we arrive at closed-form expressions for the moments of both the marginal posterior densities and the predictive distribution by integrating the terms of the polynomial expansion in turn (now feasible due to conjugacy). We demonstrate via a large-scale simulation that this approach is very accurate and that the corresponding gains in computing time are quite substantial. Furthermore, even in cases where the computing gains are more modest our approach provides a method for obtaining starting values for other algorithms, and a method for data exploration.

Keywords
beta-prime distribution, empirical bayes methods, Pearson type VI distribution

Disciplines
Business | Marketing | Statistics and Probability

Closed-Form Bayesian Inference for the Negative Binomial Distribution

Eric T. Bradlow
Bruce G.S. Hardie
Peter S. Fader[1]

June 2000

[1]Eric T. Bradlow (ebradlow@wharton.upenn.edu) is Assistant Professor of Marketing and Statistics, and Peter S. Fader is Associate Professor of Marketing, The Wharton School of the University of Pennsylvania, Philadelphia, PA 19104-6371, USA. Bruce G.S. Hardie is Assistant Professor of Marketing, London Business School, London, NW1 4SA, UK.

Abstract

To date, Bayesian inferences for the negative binomial distribution (NBD) have relied on computationally intensive numerical methods (e.g., Markov chain Monte Carlo) as it is thought that the posterior densities of interest are not amenable to closed-form integration. In this paper, we present closed-form solutions to the Bayesian inference problem for the NBD. The key insight is to approximate the ratio of two gamma functions using a polynomial expansion, which then allows for the use of a conjugate prior. Given this approximation, we arrive at closed-form expressions for the moments of both the marginal posterior densities and the predictive distribution. We demonstrate that this approach is very accurate and that the corresponding gains in computing time are quite substantial.

1 Introduction

For the past 80 years, numerous researchers have modeled count data assuming the individual-level count, X_i, is distributed Poisson(λ_i) and the rate parameter λ_i is distributed gamma(r, α). In other words, the marginal distribution of X_i follows a negative binomial distribution (NBD):

$$\begin{aligned}
[x_i|r,\alpha] &= \int_0^\infty [x_i|\lambda_i][\lambda_i|r,\alpha]\, d\lambda_i \\
&= \int_0^\infty \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!} \frac{\alpha^r \lambda_i^{r-1} e^{-\alpha\lambda_i}}{\Gamma(r)}\, d\lambda_i \\
&= \frac{\Gamma(r+x_i)}{\Gamma(r)x_i!} \left(\frac{\alpha}{\alpha+1}\right)^r \left(\frac{1}{\alpha+1}\right)^{x_i} \qquad (1)
\end{aligned}$$

The first application of this characterization of the NBD was presented by Greenwood and Yule (1920) to model accident statistics. It has subsequently been used to model phenomena as diverse as the purchasing of consumer packaged goods (Ehrenberg 1959), salesperson productivity (Carroll, Lee, and Rao 1986), and library circulation (Burrell 1990).

The standard inference approach has been to estimate r and α using maximum likelihood from the marginal distribution of X_i given in (1). This so-called empirical Bayes method then bases inferences on the estimated marginal distribution $[x_i|\hat{r},\hat{\alpha}]$. For example, a common application is to make predictions of future counts, Y_i, conditional on past behavior, e.g., $E(Y_i|x_i,\hat{r},\hat{\alpha}) = (\hat{r}+x_i)/(\hat{\alpha}+1)$. These traditional analyses treat r and α as fixed and known at $\hat{r}$ and $\hat{\alpha}$. A key shortcoming of this approach is that it ignores the variability (estimation uncertainty) of $\hat{r}$ and $\hat{\alpha}$ as estimates of r and α. Furthermore, it does not allow the incorporation of any prior information in the analysis, such as insights that may have been derived from previous studies or from subjective information.

A natural solution to these problems is to adopt a Bayesian approach in which prior distributions are specified for the NBD model parameters r and α. In recent years, several researchers have utilized Bayesian inference for the NBD (e.g., Deely and Smith 1998;

Klugman 1992; Schlüter, Deely, and Nicholson 1997). However, none of these have obtained *closed-form inferences*, as we do here. To demonstrate these past shortcomings, and our approach, we set up the general problem formulation as follows.

Let the vector $\mathbf{x} = (x_1, \ldots, x_n)$ represent the observed (count) data for n subjects measured across a time period of unit length. Assuming $x_1, \ldots, x_n$ are generated i.i.d. NBD(r, α), we have

$$\begin{aligned}
[\mathbf{x}|r,\alpha] &= \prod_{i=1}^{n} \left[\frac{\Gamma(r+x_i)}{\Gamma(r)x_i!} \left(\frac{\alpha}{\alpha+1}\right)^r \left(\frac{1}{\alpha+1}\right)^{x_i} \right] \\
&= \left[\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{\Gamma(r)x_i!} \right] \left(\frac{\alpha}{\alpha+1}\right)^{nr} \left(\frac{1}{\alpha+1}\right)^{\sum_{i=1}^n x_i}
\end{aligned} \tag{2}$$

Specifying a prior $[r, \alpha]$, it follows that the k-th marginal posterior moment of r given data $\mathbf{x}$ is given by

$$\begin{aligned}
E(r^k|\mathbf{x}) &= \int r^k [r|\mathbf{x}]\, dr \\
&= \iint r^k [r,\alpha|\mathbf{x}]\, d\alpha\, dr, \qquad \text{and then from Bayes rule} \\
&= \frac{\iint r^k [\mathbf{x}|r,\alpha][r,\alpha]\, d\alpha\, dr}{\iint [\mathbf{x}|r,\alpha][r,\alpha]\, d\alpha\, dr}
\end{aligned} \tag{3}$$

Similarly, the k-th marginal posterior moment of α is given by

$$\begin{aligned}
E(\alpha^k|\mathbf{x}) &= \int \alpha^k [\alpha|\mathbf{x}]\, d\alpha \\
&= \iint \alpha^k [r,\alpha|\mathbf{x}]\, dr\, d\alpha \\
&= \frac{\iint \alpha^k [\mathbf{x}|r,\alpha][r,\alpha]\, dr\, d\alpha}{\iint [\mathbf{x}|r,\alpha][r,\alpha]\, d\alpha\, dr}
\end{aligned} \tag{4}$$

(An equivalent expression can be written for $E(Y_i^k|\mathbf{x})$, the k-th moment of the predictive distribution.) Performing the requisite integrations allows the analyst to make the inferences of interest; i.e., to obtain parameter estimates, predictions, etc., from the appropriate marginal posterior distributions.

To highlight the challenges associated with a Bayesian analysis of the NBD, we substitute (2) into (3), which yields

$$E(r^k|\mathbf{x}) = \frac{\iint r^k \left[\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{\Gamma(r)x_i!}\right] \left(\frac{\alpha}{\alpha+1}\right)^{nr} \left(\frac{1}{\alpha+1}\right)^{\sum_{i=1}^{n} x_i} [r,\alpha]\, d\alpha\, dr}{\iint \left[\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{\Gamma(r)x_i!}\right] \left(\frac{\alpha}{\alpha+1}\right)^{nr} \left(\frac{1}{\alpha+1}\right)^{\sum_{i=1}^{n} x_i} [r,\alpha]\, d\alpha\, dr} \tag{5}$$

It is easy to specify a prior for α such that we can integrate α out of the above expression (such a distribution is well understood and appears in many places including some of the earlier mentioned Bayesian NBD references). To date, however, no researcher has been able to specify a prior for r such that there exists a closed-form solution to the above expression, and likewise for (4). (A closed-form solution trivially exists for the degenerate case of a point-mass prior for r, e.g., Klugman (1992).) Therefore, researchers have resorted to computationally intensive numerical methods (e.g., Markov chain Monte Carlo) in order to make the inferences of interest.

In this paper, we present a closed-form solution to the Bayesian inference problem for the NBD. We do this by specifying a prior for the model parameters r and α which, after utilizing a polynomial expansion to approximate the ratio of two gamma functions (details provided in Sections 2.1 and 2.2), leads to a closed-form solution for the relevant marginal posterior moments. The remainder of this paper is as follows. Section 2 develops closed-form expressions for the moments of the marginal posterior densities for both r and α. In Section 3, we present a simulation study that demonstrates the accuracy of our approach and the substantial savings in computing time associated with its use. Section 4 shows the corresponding moments (using the same method as in Section 2) for the evaluation of the moments of the predictive distribution. In particular, we derive closed-form expressions for the conditional mean and variance. We conclude (Section 5) with a few brief summary comments.

2 Moments of the Marginal Posterior Densities

Our goal is to come up with closed-form expressions for $E(r^k|\mathbf{x})$ and $E(\alpha^k|\mathbf{x})$. We start by assuming independent priors for r and α with the following marginals:

i. *Marginal prior for* α: Consider a prior of the form

$$[\alpha|\delta_1,\delta_2] \propto \left(\frac{\alpha}{\alpha+1}\right)^{\delta_1-1}\left(\frac{1}{\alpha+1}\right)^{\delta_2+1}$$

When $\delta_1 = 1$ and $\delta_2 = -1$, we have an improper uniform prior on α. Specifying the proportionality constant as $1/B(\delta_1,\delta_2)$, we have a beta-prime prior on α (Johnson, Kotz, and Balakrishnan (JKB) 1995, p. 248):

$$[\alpha|\delta_1,\delta_2] = \frac{1}{B(\delta_1,\delta_2)}\left(\frac{\alpha}{\alpha+1}\right)^{\delta_1-1}\left(\frac{1}{\alpha+1}\right)^{\delta_2+1}, \quad \delta_1,\delta_2 > 0 \tag{6}$$

which is equivalent to assuming $\alpha/(\alpha+1) \sim \text{beta}(\delta_1,\delta_2)$. The mean and variance of the beta-prime distribution are

$$E(\alpha) = \frac{\delta_1}{\delta_2-1}$$
$$\text{var}(\alpha) = \frac{\delta_1(\delta_1+\delta_2-1)}{(\delta_2-1)^2(\delta_2-2)}$$

If prior information is available for α (from past data or otherwise), then values of δ_1 and δ_2 can be determined by moment matching to arrive at an informative prior.

ii. *Marginal prior for* r: Consider a prior of the form

$$[r|a,b,z_1,z_2] \propto \frac{(r-z_1)^a}{(r-z_2)^b}$$

When $a = b = 0$, we have an improper uniform prior on r. Specifying the proportionality constant as $[\Gamma(b)(z_1-z_2)^{b-a-1}]/[\Gamma(b-a-1)\Gamma(a+1)]$, we have a Pearson Type

VI prior on r (JKB, p. 344):

$$[r|a,b,z_1,z_2] = L_1 \frac{(r-z_1)^a}{(r-z_2)^b} \text{ where } L_1 = \frac{\Gamma(b)(z_1-z_2)^{b-a-1}}{\Gamma(b-a-1)\Gamma(a+1)} \tag{7}$$

where $b > a > -1$ and $r \geq z_1 > z_2$. Note that the beta-prime density (6) is a Pearson Type VI density with $a = \delta_1 - 1$, $b = \delta_1 + \delta_2$, $z_1 = 0$ and $z_2 = -1$. The mean and variance of this distribution are

$$E(r) = z_1 + \frac{(z_1 - z_2)(a+1)}{b-a-2}$$
$$\text{var}(r) = \frac{(z_1-z_2)^2(a+1)(b-1)}{(b-a-2)^2(b-a-3)}$$

(Note: there are errors in JKB for these equations which are corrected here). The four parameters z_1, z_2, a, and b can be expressed in terms of the first four moments of r to arrive at an informative prior. The rationale for this choice of marginal prior will be discussed in Section 2.1.

2.1 Moments of the Marginal Posterior of r

To derive an expression for the k-th moment of the posterior density of r, we first substitute (6) into (5) and perform the integration over α. As can be seen by the forms in (5) and (6), the term $(\frac{\alpha}{\alpha+1})^{nr}$ from the NBD combines readily with $(\frac{\alpha}{\alpha+1})^{\delta_1-1}$ from the beta-prime prior for α (likewise for the $\frac{1}{\alpha+1}$ terms). This then yields

$$E(r^k|\mathbf{x}) = \frac{\int r^k \left[\prod_{i=1}^n \frac{\Gamma(r+x_i)}{\Gamma(r)x_i!}\right] \frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1-1)}[r]\,dr}{\int \left[\prod_{i=1}^n \frac{\Gamma(r+x_i)}{\Gamma(r)x_i!}\right] \frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1-1)}[r]\,dr} \tag{8}$$

where $C_1 = \sum_{i=1}^n x_i + \delta_2 + 1$.

The task now facing us is to substitute (7) into (8) and perform the integrations over r. The problem here is that r appears in two ratios of gamma functions in each integral — a form that is not analytically tractable. To get past this hurdle, we will first express each

ratio of gamma functions as a polynomial in r, thus making it easy to integrate over the prior on r by integrating the product of the polynomials term by term. The first ratio of gamma functions involving $\Gamma(r+x_i)/\Gamma(r)$ can be expressed exactly as a polynomial in r. The second ratio $\Gamma(nr+\delta_1)/\Gamma(nr+\delta_1+C_1-1)$ is approximated by a polynomial in r. As we describe, by including a sufficiently large number of terms, this approximation can be made as close as necessary.

Consider the first ratio of gamma functions. Letting $x^* = \max(x_1, \ldots, x_n)$ and n_j the number of observations (e.g., people) with count $x_i = j$, we have

$$\begin{aligned}\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{\Gamma(r)} &= \left(\frac{\Gamma(r+1)}{\Gamma(r)}\right)^{n_1} \left(\frac{\Gamma(r+2)}{\Gamma(r)}\right)^{n_2} \cdots \left(\frac{\Gamma(r+x^*)}{\Gamma(r)}\right)^{n_{x^*}} \\ &= r^{n_1} \left[r(r+1)\right]^{n_2} \cdots \left[r(r+1)\cdots(r+x^*-1)\right]^{n_{x^*}} \\ &= r^{s_1}(r+1)^{s_2} \cdots (r+x^*-1)^{s_{x^*}} \qquad (9)\end{aligned}$$

where $s_i = n_i + n_{i+1} + \cdots + n_{x^*}$.

We expand the second ratio of gamma functions as follows:

$$\begin{aligned}\frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1-1)} &= \prod_{j=2}^{C_1} \frac{1}{nr+\delta_1+C_1-j} \\ &= \prod_{j=2}^{C_1} \frac{1}{(nr+\delta_1+C_1)(1-j/(nr+\delta_1+C_1))} \\ &= y^{1-C_1} \prod_{j=2}^{C_1} \frac{1}{1-j/y} \\ &= y^{1-C_1} \prod_{j=2}^{C_1} \left(1 + \frac{j}{y} + \frac{j^2}{y^2} + \cdots\right)\end{aligned}$$

where $y = nr + \delta_1 + C_1$. Taking the first m terms of the infinite series $(1 + \frac{j}{y} + \frac{j^2}{y^2} + \ldots)$, and then the first m terms of the corresponding product of these polynomials, we obtain the key approximation!:

$$\frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1-1)} \approx y^{1-C_1}\left(1+\frac{U_1}{y}+\frac{U_2}{y^2}+\cdots+\frac{U_m}{y^m}\right)$$

where the polynomial coefficients $U_1, \ldots U_m$ can be calculated by recursive polynomial multiplication. (Note: we have developed a fast and simple recursive algorithm to compute the U's. In addition, standard software packages such as Mathematica will perform this calculation directly.) Letting $K_1 = (\delta_1 + C_1)/n$, we have

$$\frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1-1)} \approx [n(r+K_1)]^{1-C_1}\sum_{l=0}^{m}\frac{U_l}{[n(r+K_1)]^l} \tag{10}$$

where $U_0 = 1$. (This generalizes to m terms the one-term approximation for the ratio of two gamma functions given in Johnson, Kotz, and Kemp (1992, equation 1.33) as well as the two-term expansion given in Abramowitz and Stegun (1972, equation 6.1.47).)

Now, if we let $z_1, z_2 \in \{-x^*, \ldots, -1, 0\}$ and assume a, b integer, then for

$$v_t = \begin{cases} 1 & \text{if } |z_1| = t-1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad w_t = \begin{cases} -1 & \text{if } |z_2| = t-1 \\ 0 & \text{otherwise} \end{cases}$$

(7) and (9) combine to give us

$$\begin{aligned}\prod_{i=1}^{n}\frac{\Gamma(r+x_{1i})}{\Gamma(r)}[r] &= L_1\left[r^{s_1+v_1a+w_1b}\cdots(r+x^*-1)^{s_{x^*}+v_{x^*}a+w_{x^*}b}(r+x^*)^{s_{x^*+1}+v_{x^*+1}a+w_{x^*+1}b}\right] \\ &= L_1\left[r^{h_1}(r+1)^{h_2}\cdots(r+x^*)^{h_{x^*+1}}\right]\end{aligned}$$

where $h_t = s_t + v_t a + w_t b$ and $s_{x^*+1} = 0$. (While the restriction of a, b to integers is somewhat restrictive, it still leaves a very flexible class of priors for r.) Letting $h = h_1 + h_2 + \cdots + h_{x^*+1}$, we can write this product as a polynomial in r:

$$\prod_{i=1}^{n} \frac{\Gamma(r+x_{1i})}{\Gamma(r)}[r] = L_1\left[a_0 + a_1 r + \cdots + a_h r^h\right] = L_1 \sum_{j=0}^{h} a_j r^j \tag{11}$$

where the polynomial coefficients can be computed for given $h_1, \ldots, h_{x^*+1}$ using the same algorithm used to compute the U's in (10). (Note that $a_0 = a_1 = \cdots = a_{h_1-1} = 0$, as there are no terms less than h_1.)

Substituting (10) and (11) into (8), we get

$$\begin{aligned} E(r^k|\mathbf{x}) &\approx \frac{\int r^k \left(\sum_{j=0}^{h} a_j r^j\right)(r+K_1)^{1-C_1} \sum_{l=0}^{m} \frac{U_l}{[n(r+K_1)]^l}\, dr}{\int \left(\sum_{j=0}^{h} a_j r^j\right)(r+K_1)^{1-C_1} \sum_{l=0}^{m} \frac{U_l}{[n(r+K_1)]^l}\, dr} \\ &= \frac{\sum_{j=0}^{h} a_j Q_{j+k}}{\sum_{j=0}^{h} a_j Q_j} \end{aligned} \tag{12}$$

where

$$\begin{aligned} Q_j &= \int r^j (r+K_1)^{1-C_1} \sum_{l=0}^{m} \frac{U_l}{[n(r+K_1)]^l}\, dr \\ &= \sum_{l=0}^{m} \frac{\Gamma(C_1-j+l-2)}{\Gamma(C_1+l-1)} \frac{\Gamma(j+1)}{K_1^{C_1-j+l-2}} \frac{U_l}{n^l} \end{aligned} \tag{13}$$

which follows naturally (from the normalizing constant) from the definition of the Pearson Type VI density in (7). That is, the Pearson Type VI density is a conjugate family for the NBD after employing the approximation to $\Gamma(nr+\delta_1)/\Gamma(nr+\delta_1+C_1-1)$ given in (10). Thus, closed-form posterior inferences are now feasible for r.

In order for the integral in (13) to be the normalizing constant of the Pearson Type VI, and hence defined, we must have $C_1 - j + l - 2 > 0 \; \forall\, j, l$. The maximum value that j can take on is $h+k$ ((11) is of order h, and we are computing $E(r^k|\mathbf{x})$; thus $h+k$ terms at most) while the minimum value that l can take on is 0. Therefore, Q_j is defined $\forall\, j$ when $C_1 - h - k > 2$. Now $h = \sum_{i=1}^{x^*+1} h_i = \sum_{i=1}^{x^*+1} s_i + a - b = \sum_{i=1}^{x^*+1} \sum_{j=1}^{x^*} n_j + a - b = \sum_{i=1}^{n} x_i + a - b$. Recall $C_1 = \sum_{i=1}^{n} x_i + \delta_2 + 1$. Therefore $C_1 - h - k > 2 \Leftrightarrow \delta_2 > a - b + k + 1$. If we are simply

interested in the mean of the marginal posterior of r ($k = 1$), the approximation presented in (12) holds, provided the parameters of the marginal prior distributions for r and α satisfy the condition $\delta_2 > a - b + 2$ (which is not very restrictive).

2.2 Moments of the Marginal Posterior of α

Turning our attention to α, the moments of its marginal posterior can be computed in a similar manner. Substituting (2) and (6) into (4) and simplifying, we get

$$E(\alpha^k|\mathbf{x}) = \frac{1}{\prod_{l=1}^{k}(C_1 - l - 1)} \frac{\int \left[\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{\Gamma(r)}\right] \frac{\Gamma(nr+\delta_1+k)}{\Gamma(nr+\delta_1+C_1-1)}[r]\, dr}{\int \left[\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{\Gamma(r)}\right] \frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1-1)}[r]\, dr} \tag{14}$$

Applying the same logic used to derive (10), we have

$$\begin{aligned} \frac{\Gamma(nr+\delta_1+k)}{\Gamma(nr+\delta_1+C_1-1)} &= \prod_{j=2}^{C_1-k} \frac{1}{nr+\delta_1+C_1-j} \\ &\approx [n(r+K_1)]^{1+k-C_1} \sum_{l=0}^{m} \frac{V_l^k}{[n(r+K_1)]^l} \end{aligned} \tag{15}$$

where $V_0^k = 1$, and we have a separate set of polynomial coefficients, $V_1^k, \ldots V_m^k$ for each of the k moments of the posterior of α. These can again be computed using the previously mentioned recursive algorithm.

Substituting (11) and (15) into (14), we get

$$\begin{aligned} E(\alpha^k|\mathbf{x}) &\approx \frac{n^k}{\prod_{l=1}^{k}(C_1 - l - 1)} \frac{\int \left(\sum_{j=0}^{h} a_j r^j\right) (r+K_1)^{1+k-C_1} \sum_{l=0}^{m} \frac{V_l^k}{[n(r+K_1)]^l}\, dr}{\int \left(\sum_{j=0}^{h} a_j r^j\right) (r+K_1)^{1-C_1} \sum_{l=0}^{m} \frac{U_l}{[n(r+K_1)]^l}\, dr} \\ &= \frac{n^k}{\prod_{l=1}^{k}(C_1 - l - 1)} \frac{\sum_{j=0}^{h} a_j R_j^k}{\sum_{j=0}^{h} a_j Q_j} \end{aligned} \tag{16}$$

where

$$R_j^k = \int r^j (r+K_1)^{1+k-C_1} \sum_{l=0}^{m} \frac{V_l^k}{[n(r+K_1)]^l}\, dr$$
$$= \sum_{l=0}^{m} \frac{\Gamma(C_1+l-j-k-2)}{\Gamma(C_1+l-k-1)} \frac{\Gamma(j+1)}{K_1^{C_1+l-j-k-2}} \frac{V_l^k}{n^l} \tag{17}$$

which again follows from the normalizing constant of the Pearson Type VI density in (7). Thus, closed-form posterior inferences are obtained for α. In order for (17) to be defined, we must have $C_1 + l - j - k - 2 > 0 \ \forall\, j, l$, which leads to the same condition as before ($\delta_2 > a - b + k + 1$).

3 A Simulation Study

To assess the accuracy and computational feasibility of our approach, we performed a large-scale simulation study using three main factors: r (3 levels), α (3 levels), and m, the number of polynomial expansion terms (5 levels) for the ratio of two gamma functions given in (10) and (15). We selected these as simulation factors as we wanted to know: (a) whether the quality of the estimates varied with the true (latent) values of r and α, and (b) whether the number of terms utilized in the approximating expansion played a significant role in determining accuracy. A fully crossed design was implemented yielding 45 simulation conditions.

For each condition, a given simulation consisted of $N = 500$ observations from a negative binomial distribution with the given r and α. Two hundred replicates were generated for each of the 45 conditions; that is 200 data sets of size 500. The results reported for each condition is the average over these replicates.

To select simulation values for r and α, we utilized a re-parameterization of the NBD values of r and α in terms of $\rho = \text{corr}(X_i, Y_i) = 1/(\alpha + 1)$, the correlation between two negative binomial draws in non-overlapping periods of equal length for a given individual,

and $P_0 = P(X_i = 0) = [\alpha/(\alpha+1)]^r = (1-\rho)^r$, the fraction of 0 values. (ρ also equals the square of the correlation between the observed count (X_i) and the individual's latent rate (λ_i).) This re-parameterization is convenient as both ρ and P_0 are bounded between 0 and 1, allowing us to select a grid along the unit square. In addition, ρ and P_0 are common and intuitive quantities which can be selected by a researcher based on historical figures or expert judgment.

We selected the 3×3 grid corresponding to 0.25, 0.50, 0.75 for each of ρ and P_0 leading to the values of r and α given in columns 3 and 4 of Table 1. For m, the number of polynomial expansion terms used, we selected five values: 1, 2, 10, 50, and 300. Experimentation with our approach indicated that these values would demonstrate the increased accuracy of the estimates as m increases. Computing time using an HP-UX 9000 server, and code written in Fortran, for generating 200 data sets of size 500 and estimating the corresponding moments using our approach was 2, 3, 7, 45, and 800 seconds respectively for $m = 1, 2, 10, 50, 300$. This clearly suggests very practical computing time for a given single data set, especially when compared to the numerical methods (e.g., MCMC) traditionally used in the Bayesian setting.

To complete the simulation specification, we needed to select values of the hyperparameters $(a, b, z1, z2)$, the Pearson Type VI parameters for the prior on r, and (δ_1, δ_2), the beta-prime parameters for the prior on α. These values were then inserted into equations (12) and (16), along with the 500 negative binomial draws, to obtain estimates $E(r|\mathbf{x})$ and $E(\alpha|\mathbf{x})$ for each simulation.

We ran each of the 45 simulation conditions under three choices of these hyperparameters, all having $z_1 = 0$ and $z_2 = -1$ indicating a beta-prime prior on r:

	prior on r	prior on α
i)	improper uniform	weakly informative with mean 1
	$(a = 0, b = 0)$	$(\delta_1 = 2, \delta_2 = 3)$
ii)	weakly informative with mean 1	weakly informative with mean 1
	$(a = 1, b = 5)$	$(\delta_1 = 2, \delta_2 = 3)$
iii)	weakly informative with mean 1	improper uniform
	$(a = 1, b = 5)$	$(\delta_1 = 1, \delta_2 = -1)$

We note that the condition with two uniform priors does not satisfy the constraint $\delta_2 > a - b + 2$, described earlier, and thus could not be explored. As our results indicated no significant differences across the three choices of the hyperpriors (determined by a repeated measures ANOVA), the results reported below are the average across the three sets. As with any Bayesian analysis, though, this is due, in part, to the large sample size selected ($N = 500$) in comparison to the small amount of information in the priors.

Reported in Table 1 are the Mean Absolute Estimation Errors (MAEE) for the first moments, $\text{avg}(|\widehat{E(r|\mathbf{x})} - r|)$ and $\text{avg}(|\widehat{E(\alpha|\mathbf{x})} - \alpha|)$, for each of the 45 simulation conditions and aggregated across various cells. Columns 1–5 convey the 3×3 grid for ρ and P_0, and the corresponding values of r, α and $E(X_i) = r/\alpha$, respectively. The results indicate a number of interesting features. First, as expected, the estimation error decreases with an increasing number of expansion terms m. In fact, we observe a very poor fit for a small number of terms and a good fit for the $m = 300$ condition. Secondly, estimation error increases with $E(X)$. This suggests that when the sample average is large, many more terms in the expansion m are necessary to provide adequate estimates. But even in the worst case (row 3), the 300-term approximation is excellent and, if necessary, additional terms could be added with very little difficulty or computational burden.

4 Moments of the Predictive Distribution

The approach developed above for evaluating the moments of the marginal posterior densities can be applied to the evaluation of the moments of the predictive distribution ($E(Y_i^k|\mathbf{x})$). As prediction is among the most common applications of the NBD model in practice, closed-

Table 1: Mean Absolute Estimation Error for $(\widehat{E(r|\mathbf{x})}, \widehat{E(\alpha|\mathbf{x})})$ with $X_i \sim \text{NBD}(r, \alpha)$. The simulation conditions are derived from $\rho = \text{corr}(X_i, Y_i) = 1/(\alpha + 1)$, $P_0 = P(X_i = 0) = (1 - \rho)^r$, and m = number of terms used in the approximation of the ratio of two gamma functions.

					m					
P_0	ρ	r	α	$E(X_i)$	1	2	10	50	300	Mean
0.25	0.25	4.82	3.00	1.61	(27.62, 17.54)	(22.24, 14.05)	(7.05, 4.51)	(0.98, 0.62)	(0.05, 0.07)	(11.59, 7.36)
0.25	0.50	2.00	1.00	2.00	(33.73, 17.31)	(27.57, 14.18)	(11.05, 5.78)	(1.68, 0.93)	(0.10, 0.06)	(14.81, 7.66)
0.25	0.75	1.00	0.33	3.00	(33.04, 11.91)	(27.39, 9.94)	(13.20, 5.03)	(3.77, 1.57)	(0.12, 0.08)	(15.49, 5.71)
0.50	0.25	2.41	3.00	0.83	(6.04, 7.78)	(4.80, 6.24)	(1.08, 1.33)	(0.71, 0.87)	(0.00, 0.04)	(2.53, 3.26)
0.50	0.50	1.00	1.00	1.00	(6.41, 7.11)	(5.33, 5.90)	(2.15, 2.49)	(0.31, 0.31)	(0.05, 0.06)	(2.85, 3.18)
0.50	0.75	0.50	0.33	1.50	(4.51, 3.97)	(4.16, 3.56)	(2.44, 2.25)	(0.70, 0.70)	(0.01, 0.04)	(2.36, 2.10)
0.75	0.25	1.00	3.00	0.33	(0.88, 2.98)	(0.72, 2.26)	(0.27, 0.57)	(0.32, 0.99)	(0.03, 0.00)	(0.44, 1.37)
0.75	0.50	0.42	1.00	0.42	(1.05, 3.22)	(0.87, 2.73)	(0.37, 1.05)	(0.15, 0.35)	(0.01, 0.02)	(0.49, 1.48)
0.75	0.75	0.21	0.33	0.60	(0.75, 2.20)	(0.71, 2.05)	(0.42, 1.23)	(0.11, 0.27)	(0.00, 0.01)	(0.40, 1.15)
Mean					(12.67, 8.22)	(10.42, 6.77)	(4.22, 2.69)	(0.97, 0.73)	(0.05, 0.05)	

form moments in this instance have great practical value. Let Y_i be the count variable for individual i in a non-overlapping time period, where the two periods corresponding to X_i, Y_i are of equal length. In this section, we focus on evaluating the mean and variance of the predictive distribution: $E(Y_i|\mathbf{x})$ and $\text{var}(Y_i|\mathbf{x})$. The method easily generalizes to higher moments.

Given r and α, the mean and variance of Y_i, conditional on x_i, are:

$$E(Y_i|x_i, r, \alpha) = \frac{r + x_i}{\alpha + 1} \text{ and } \text{var}(Y_i|x_i, r, \alpha) = \frac{r + x_i}{\alpha + 1} + \frac{r + x_i}{(\alpha + 1)^2}$$

It follows that

$$E(Y_i|\mathbf{x}) = \iint \frac{r + x_i}{\alpha + 1} [r, \alpha|\mathbf{x}] \, d\alpha \, dr \tag{18}$$

and

$$\begin{aligned} \text{var}(Y_i|\mathbf{x}) &= \iint \left(\frac{r + x_i}{\alpha + 1} + \frac{r + x_i}{(\alpha + 1)^2} \right) [r, \alpha|\mathbf{x}] \, d\alpha \, dr \\ &= E(Y_i|\mathbf{x}) + \iint \frac{r + x_i}{(\alpha + 1)^2} [r, \alpha|\mathbf{x}] \, d\alpha \, dr \end{aligned} \tag{19}$$

Central to evaluating (18) and (19) is the evaluation of the following integral:

$$\iint \frac{r + x_i}{(\alpha + 1)^k} [r, \alpha|\mathbf{x}] \, d\alpha \, dr = \frac{\iint \frac{r+x_i}{(\alpha+1)^k} [\mathbf{x}|r, \alpha][r][\alpha] \, d\alpha \, dr}{\iint [\mathbf{x}|r, \alpha][r][\alpha] \, d\alpha \, dr} \tag{20}$$

Substituting (2) and (6) into (20) and simplifying, we get

$$\iint \frac{r + x_i}{(\alpha + 1)^k} [r, \alpha|\mathbf{x}] \, d\alpha \, dr = \prod_{l=1}^{k} (C_1 + l - 2) \frac{\int (r + x_i) \left[\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{\Gamma(r)} \right] \frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1+k-1)} [r] \, dr}{\int \left[\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{\Gamma(r)} \right] \frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1-1)} [r] \, dr} \tag{21}$$

Using the same logic as in (10), we have

$$\frac{\Gamma(nr+\delta_1)}{\Gamma(nr+\delta_1+C_1+k-1)} = \prod_{j=2-k}^{C_1} \frac{1}{nr+\delta_1+C_1-j}$$
$$\approx [n(r+K_1)]^{1-C_1-k} \sum_{l=0}^{m} \frac{W_l^k}{[n(r+K_1)]^l} \tag{22}$$

where $W_0^k = 1$, and we have a separate set of polynomial coefficients, $W_1^k, \ldots W_m^k$ for $k = 1, 2$, computed using the previously mentioned recursive algorithm.

Substituting (11) and (22) into (21), we get

$$\begin{aligned}
&\iint \frac{r+x_i}{(\alpha+1)^k}[r,\alpha|\mathbf{x}]\, d\alpha\, dr \\
&\quad\approx \frac{\prod_{l=1}^{k}(C_1+l-2)}{n^k} \frac{\int (r+x_i)\left(\sum_{j=0}^{h} a_j r^j\right)(r+K_1)^{1-C_1-k} \sum_{l=0}^{m} \frac{W_l^k}{[n(r+K_1)]^l}\, dr}{\int \left(\sum_{j=0}^{h} a_j r^j\right)(r+K_1)^{1-C_1} \sum_{l=0}^{m} \frac{U_l}{[n(r+K_1)]^l}\, dr} \\
&\quad= \frac{\prod_{l=1}^{k}(C_1+l-2)}{n^k} \frac{\sum_{j=0}^{h} a_j \left[S_{j+1}^k + x_i S_j^k\right]}{\sum_{j=0}^{h} a_j Q_j}
\end{aligned} \tag{23}$$

where

$$\begin{aligned}
S_j^k &= \int r^j (r+K_1)^{1-C_1-k} \sum_{l=0}^{m} \frac{W_l^k}{[n(r+K_1)]^l}\, dr \\
&= \sum_{l=0}^{m} \frac{\Gamma(C_1+k+l-j-2)}{\Gamma(C_1+k+l-1)} \frac{\Gamma(j+1)}{K_1^{C_1+k+l-j-2}} \frac{W_l^k}{n^l}
\end{aligned} \tag{24}$$

which again follows from the normalizing constant of the Pearson Type VI density in (7).

In order for (24) to be defined, we must have $C_1 + k + l - j - 2 > 0 \;\forall\, j, l$. Thus S_j^k is defined $\forall\, j$ when $C_1 - h - k > 2 \Leftrightarrow \delta_2 > a - b + 1$. This is automatically satisfied when we constrain the parameters of the prior distributions such that the approximation of the moments of the marginal posterior distributions of r and α, (12) and (16), hold (i.e., $\delta_2 > a - b + 2$). Thus no new constraints need to be applied.

Substituting (23) into (18) and (19), we get the following expressions for the conditional

mean and variance:

$$E(Y_i|\mathbf{x}) \approx \frac{C_1 - 1}{n} \frac{\sum_{j=0}^{h} a_j \left[S_{j+1}^1 + x_i S_j^1 \right]}{\sum_{j=0}^{h} a_j Q_j} \tag{25}$$

and

$$\text{var}(Y_i|\mathbf{x}) \approx E(Y_i|\mathbf{x}) + \frac{C_1(C_1 - 1)}{n^2} \frac{\sum_{j=0}^{h} a_j \left[S_{j+1}^2 + x_i S_j^2 \right]}{\sum_{j=0}^{h} a_j Q_j} \tag{26}$$

5 Conclusions

The negative binomial distribution is widely used to model count data. However, the standard inference approach estimates the model parameters using maximum likelihood, treating them fixed and known at $\hat{r}$ and $\hat{\alpha}$. In contrast, the Bayesian approach allows for uncertainty in parameter estimates and the incorporation of prior information.

It had previously been thought that Bayesian inference for the NBD must make use of numerical integration since expressions for the posterior densities of interest contain a model parameter within a ratio of gamma functions. However, in this paper we have presented closed-form solutions to the Bayesian inference problem for the NBD. The key insight is to approximate the ratio of two gamma functions using a polynomial expansion, which then allows for the use of a conjugate prior. Given this approximation, we arrive at closed-form expressions for the moments of both the marginal posterior densities and the predictive distribution. In fact, the posterior densities themselves are a special case of the results given. We demonstrate that this approach is very accurate and that the corresponding gains in computing time are quite substantial.

The accuracy of our approach depends on the number of terms used in the approximation to the ratio of the two gamma functions. Even for a large number (i.e., 300) of terms, however, the computational time required to make the relevant inferences is minimal. When compared to standard numerical methods (e.g., MCMC), the gains in computing time are

quite substantial. This is of great practical importance as it provides analysts with the flexibility of the Bayesian approach to inference for the NBD without the excessive computational burden traditionally associated with such an approach.

The approach developed in this paper need not be limited to the NBD. We hope that this work spurs on other researchers to search for closed-form solutions in other situations where the application of Bayesian techniques has forced the analyst to resort to computationally intensive methods. A natural starting point would be to examine other common distributions such as the beta-binomial and the Pareto. Some researchers have attempted to implement Bayesian approaches for these models, e.g. Lee and Sabavala (1987) and Arnold and Press (1989), but in both cases they had to resort to numerical methods of some sort.

References

Abramowitz, Milton and Irene A. Stegun, eds. (1972), *Handbook of Mathematical Functions*, New York: Dover Publications.

Arnold, Barry C. and S. James Press (1989), "Bayesian Estimation and Prediction for Pareto Data," *Journal of the American Statistical Association*, **84** (December) 1079–84.

Burrell, Quentin L. (1990), "Using the Gamma-Poisson Model to Predict Library Circulations," *Journal of the American Society for Information Science*, **41** (April), 164–70.

Carroll, Vincent P., Hau L. Lee, and Ambar G. Rao (1986), "Implications of Salesforce Productivity Heterogeneity and Demotivation: A Navy Recruiter Case Study," *Management Science*, **32** (November), 1371–88.

Deely, John J. and Adrian F. M. Smith (1998), "Quantitative Refinements for Comparisons of Institutional Performance," *Journal of the Royal Statistical Society, Series A*, **161**, 5–12.

Ehrenberg, A. S. C. (1959), "The Pattern of Consumer Purchases," *Applied Statistics*, **8** (March), 26–41.

Greenwood, Major and G. Udny Yule (1920), "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeat Accidents," *Journal of the Royal Statistical Society*, **83** (March), 255–79.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan (1995), *Continuous Univariate Distributions, Volume 2*, 2nd edn., New York: John Wiley & Sons.

Johnson, Norman L., Samuel Kotz, and Adrienne Kemp (1992), *Univariate Discrete Distributions*, 2nd edn., New York: John Wiley & Sons.

Kilpatrick, S. James, Jr. (1977), "An Empirical Study of the Distribution of Episodes of Illness Recorded in the 1970–71 National Morbidity Survey," *Applied Statistics*, **26**, 26–33.

Klugman, Stuart A. (1992), *Bayesian Statistics in Actuarial Science*, Boston: Kluwer Academic Publishers.

Lee, Jack C. and Darius J. Sabavala (1987), "Bayesian Estmation and Prediction for the Beta-Binomial Model," *Journal of Business and Economic Statistics*, **5** (July) 357–67.

Schlüter, P. J., J. J. Deely, and A. J. Nicholson (1997), "Ranking and Selecting Motor Vehicle Accident Sites by Using a Hierarchical Bayesian Model," *The Statistician*, **46**, 293–316.