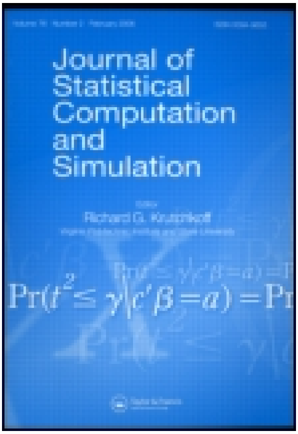


Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

# A note on bayesian estimation and prediction for the beta-binomial model

Jack Lee [a] & Y. L. Lio [a]
[a] Institute of Statistics National, Chiao Tung University, Hsinchu, Taiwan, R.O.C.
Published online: 10 Jan 2011.

*J. Statist. Comput. Simul.*, 1999, Vol. 63, pp. 73–91
Reprints available directly from the publisher
Photocopying permitted by license only





# A NOTE ON BAYESIAN ESTIMATION AND PREDICTION FOR THE BETA-BINOMIAL MODEL

JACK C. LEE[a] and Y. L. LIO[b,*]

[a] *Institute of Statistics National, Chiao Tung University, Hsinchu 30050, Taiwan, R.O.C.;*
[b] *Department of Mathematics, University of South Dakota, Vermillion, SD 57069 USA*

*(Received 16 February 1998; In final form 4 November 1998)*

The beta-binomial model which is generated by a simple mixture model has been widely applied in the social, physical, and health sciences. Lee and Sabavala (1987) proposed a Bayesian approach with a conjugate-type beta family of priors for suitably transformed parameters in the beta-binomial, and demonstrated the simulations for a special case of two trials. The main purpose of this paper is to extend the study of Lee and Sabavala (1987) by a numerical double integration. This method can be used for the case of general trials. When the number of trials is two, the results are similar to those from Lee and Sabavala (1987). However, the predictions for the real problems are much better than the results in Lee and Sabavala (1987).

*Keywords:* Conjugate-type priors; simulation; television viewing

## 1. INTRODUCTION

The beta-binomial model is generated in the following manner. Consider a population in which for each member some event occurs as the outcome of a Bernoulli trial with probability $p$. Thus given $0 < p < 1$, the number of occurrences for $k$ in $r$ trials has the binomial distribution,

$$\Pr(k|r,p) = \binom{r}{k} p^k (1-p)^{r-k}, \quad k = 0, 1, \ldots, r. \tag{1}$$

*Corresponding author.

Suppose that $p$ varies across the population according to a beta distribution,

$$f(p|\alpha,\beta) = p^{\alpha-1}(1-p)^{\beta-1}/B(\alpha,\beta), \ \ \alpha > 0, \ \ \beta > 0, \ \ 0 < p < 1, \quad (2)$$

where $B(\alpha,\beta)$ is the complete beta function. Since $p$ is not observable, the probability distribution of $k$ in $r$ trials, given $\alpha$ and $\beta$, for a randomly chosen member is the following simple mixture model

$$\Pr_{BB}(k|r,\alpha,\beta) = \int \Pr(k|r,p)f(p|\alpha,\beta)dp.$$

Therefore, the beta-binomial model for the occurrences $k$ can be rewritten as

$$\Pr_{BB}(k|r,\alpha,\beta) = \binom{r}{k} B(\alpha+k,\beta+r-k)/B(\alpha,\beta), \ \ k = 0,1,\ldots,r. \quad (3)$$

This model was initiated by Pearson (1925) in an experimental investigation of Bayes' theorem, and formally proposed by Skellam (1948). Since then, the beta-binomial model has been applied in mental testing (Huynh, 1979; Lord, 1965; Wilcox, 1981), toxicological experimentation (Williams, 1975), epidemiology (Griffiths, 1973), media exposure (Green, 1970), and buying behavior (Massy, Montgomery and Morrison, 1970), among others. Up until Lee and Sabavala (1987), the most commonly used estimation procedures were the method of moments, minimum Chi-square, and maximum likelihood (ML) (Griffiths, 1973; Kalwani, 1980; Kleinman, 1973; Morrison, 1966; Wilcox, 1979). Lee and Sabavala (1987) proposed a Bayesian procedure for inference on the parameters and prediction of the event frequencies and pointed out the advantage of the Bayesian procedures over the ML procedures in computation. Because the ML estimators were not available in a closed form for beta-binomial models, the ML solutions could not be well defined or could fall outside the valid range of parameters through an iterative computing process. A key initial step in the Bayesian framework of Lee and Sabavala's (1987) was to reparameterize the model in term of $\mu = \alpha/(\alpha+\beta)$ and $\rho = 1/(\alpha+\beta+1)$. These parameters $\mu$ and $\rho$ have meaningful interpretations in the beta binomial model, but $\alpha$ and $\beta$ do not (see Sabavala and Morrison, 1977). Moreover, $\mu$ and $\rho$ have values restricted to the interval 0 and 1. Therefore, beta priors for $\mu$ and $\rho$ can be used for the Bayesian framework in the model. Kahn and Raftery (1996) proposed Bayesian Logistic regression and Bayesian estimation for the parameter $p$. The approach is different from the beta-binomial model in which $p$ is not the parameter. Lee and Sabavala (1987) demonstrated the simulations for special case of two trials and suggested a Pearson type I distribution to approximate the posterior distributions, and then using the approach for the case of two trials to predict the new event frequencies in a real problem which had input data from the case of more than two trials. Our objective here is to use a numerical integration method directly to solve the general case (*i.e.*, $r$ can be any positive integer) of Bayesian estimation and prediction in the beta-binomial model with typical input data from Lee and Sabavala (1987). In general, the estimations for the special case of two trials are similar to the Bayesian estimations from Lee and Sabavala (1987). However, the predictions of event frequencies in the real application problems with more than two trials are much better than the predictions of Lee and Sabavala's (1987).

Section 2 describes the models in a Bayesian framework with typical input data from Lee and Sabavala (1987). Section 3 presents results of estimations for a general case of trial.

A real data application is presented in Section 4, and Section 5 contains concluding remarks.



## 2. FORMULA FOR ESTIMATIONS AND PREDICTIONS

For a fixed number of trials $r$, assume that a sample of size $n$ is selected from the model given in (3). Then the data will consist of the number of occurrences, $k$, for each of the $n$ units. This data can be summarized as $\{n_k;\ k = 0, 1, \ldots, r\}$, where $n_k$ is the number of units with $k$ occurrences and $n = \sum_{k=0}^{r} n_k$. Hence, the likelihood function is given by

$$L(\alpha, \beta\,|r, \{n_k\}) = \prod_{k=0}^{r} \{\Pr_{BB}(k|r, \alpha, \beta)\}^{n_k}. \tag{4}$$

Let $\mu = \alpha/(\alpha + \beta)$ and $\rho = 1/(\alpha + \beta + 1)$, then the reparameterized likelihood function of $\mu$ and $\rho$ can be presented as

$$L(\mu, \rho | r, \{n_k\}) \propto \left(\prod_{i=1}^{r-1}\left(\frac{A_i}{C_i}\right)^{n_0}\right) \left(\prod_{k=1}^{r-1}\left(\left(\left(\prod_{i=0}^{k-1} B_i\right)\left(\prod_{i=0}^{r-k-1} A_i\right)\right) \Big/ \left(\prod_{i=0}^{r-1} C_i\right)\right)^{n_k}\right)\left(\prod_{i=0}^{r-1}\left(\frac{B_i}{C_i}\right)^{n_r}\right), \tag{5}$$



where $A_i = [(1-\mu)(1-\rho) + i\rho]$, $B_i = [\mu(1-\rho) + i\rho]$, and $C_i = [(1-\rho) + i\rho]$ for any given nonnegative integer $i$. It should be mentioned that the proportionality for (5) is dependent of $r$ and $\{n_k; k = 0, 1, \ldots, r\}$. Since $\mu$ and $\rho$ have values between 0 and 1, the conjugate-type beta prior distributions for $\mu$ and $\rho$ will be applied here. Assuming that $\mu$ and $\rho$ are independent *a priori*, then the joint prior distribution can be expressed as

$$g(\mu, \rho) \propto \mu^{\gamma_1 - 1}(1-\rho)^{\gamma_2 - 1}\rho^{\delta_1 - 1}(1-\rho)^{\delta_2 - 1}, \quad 0 < \mu < 1, \ 0 < \rho < 1, \tag{6}$$

where $\gamma_1$, $\gamma_2$, $\delta_1$ and $\delta_2$ are positive real numbers given. Combining the likelihood function with the prior $g(\mu, \rho)$, the joint posterior distribution $\Pr(\mu, \rho \,|\, r, \{n_k\})$ of $\mu$ and $\rho$ given data $\{n_k; k = 0, 1, \ldots, r\}$ can be shown to satisfy that

$$\Pr(\mu, \rho \,|\, r, \{n_k\}) \propto \left(\prod_{i=1}^{r-1}\left(\frac{A_i}{C_i}\right)^{n_0}\right) \left(\prod_{k=1}^{r-1}\left(\left(\left(\prod_{i=0}^{k-1} B_i\right)\left(\prod_{i=0}^{r-k-1} A_i\right)\right) \Big/ \left(\prod_{i=0}^{r-1} C_i\right)\right)^{n_k}\right)\left(\prod_{i=0}^{r-1}\left(\frac{B_i}{C_i}\right)^{n_r}\right)$$

When $r = 2$, (7) can be simplified to be

$$\Pr(\mu, \rho | r = 2, \{n_k\}) \propto [(1-\mu)A_1]^{n_0}[\mu B_1]^{n_2} \cdot \rho^{\delta_1 - 1}(1-\rho)^{n_1 + \delta_2 - 1} \cdot \mu^{n_1 + \gamma_1 - 1}(1-\mu)^{n_1 + \gamma_2 - 1} \tag{8}$$

We keep the factor $[(1-\mu)A_1]^{n_0}[\mu B_1]^{n_2}$ in (8) to reduce the possibility of underflow problems in large sample (*i.e.*, $n$ is large). Expanding the factor $[(1-\mu)A_1]^{n_0}[\mu B_1]^{n_2}$, the formula (8) can be expressed as a weighted sum of finitely many conjugate-type beta priors that is an equivalent result of Lee and Sabavala (1987). However, formula (8) is much better for programming a numerical integration when $r=2$.

It should be pointed out that the explicit proportionality for (7) is not available when $r>2$. Therefore, for a numerical calculation of $\Pr(\mu, \rho|r, \{n_k\})$, the proportionality for (7) will be estimated through the following steps during the simulations.

Given a positive integer $z$, let $\{(\mu_i, \rho_j): i=0,1,\ldots,z; j=0,1,\ldots,z\}$ be a set of grid points over $[0, 1]\times[0, 1]$ such that $\{\mu_i = 0,1,\ldots,z\}$ and $\{\rho_j: j=0,1,\ldots,z\}$ equally partition over [0, 1], respectively.

(I) Calculate a Riemann sum for the function in the right side of (7) with the grid points $\{(\mu_i, \rho_j): i=0,1,\ldots,z; j=0,1,\ldots,z\}$.
(II) Increase the grid points by doubling $z$ and repeat step (I).

This procedures will be continued until two successive Riemann sums, $\mathrm{RSS}_z$ and $\mathrm{RSS}_{2z}$, satisfy $(\mathrm{RSS}_{2z}-\mathrm{RSS}_z)/\mathrm{RSS}_z < 10^{-5}$. Then the last Riemann sum will be used as the proportionality for (7) and the final grid points will be used for all the following integrations. The marginal posterior density of $\mu$ can also be approximated at the set of final grid points $\{\mu_i;\ i=1,\ldots,z\}$ through a Riemann sum of $\Pr(\mu_i, \rho|r, \{n_k\})$ over the set of final grid points $\{\rho_j: j=1,\ldots,z\}$ and so can the marginal posterior density of $\rho$ at the set of final grid points $\{\rho_j: j=1,\ldots,z\}$ through a Riemann sum of $\Pr(\mu, \rho_j|r, \{n_k\})$ over the set of final grid points $\{\mu_i: i=1,\ldots,z\}$. Furthermore, using a computing procedure, GCIN, in the International Mathematical and Statistical Library (IMSL, 1991) for evaluating the inverse of a general continuous cumulative distribution function given ordinates of the density, any $p$th quantile estimates for the posterior distributions of $\mu$ and $\rho$ can be constructed, respectively. Therefore, the posterior intervals of $\mu$ and $\rho$ can be easily obtained, respectively.

Note that given the data $\{n_k: k=0,1,\ldots,r\}$, we are concerned with the inference on $\mu$ and $\rho$ (or $\alpha$ and $\beta$), in the beta-binomial model, rather than on $p$, in the binomial model. It is further noted that the Bayesian estimation of Kahn and Raftery (1996) is for $p$. If the marginal posteriors for $\mu$ and $\rho$, respectively, are available, then the Bayesian estimators of $\mu$ and $\rho$ may be chosen as means, medians, or modes from these marginal posterior densities. Interval estimates may be found using the symmetric interval, or the asymmetric interval (*i.e.*, with 0 as the lower bound and finding the upper bound, or with 1 as the upper bound and finding the lower bound), or the highest marginal posterior density interval. Except the case of two trials, the closed forms of marginal posteriors for $\mu$ and $\rho$, respectively, are apparently not available. Therefore, the following double integrations (9) and (10) will be used to find the Bayesian estimates for $\mu$ and $\rho$, respectively. (*i.e.*, Bayes' estimates under quadratic loss).

$$E(\mu) = \int\int \mu \Pr(\mu, \rho | r, \{n_k\}) d\mu d\rho \tag{9}$$

and

$$E(\rho) = \int\int \rho \Pr(\mu, \rho | r, \{n_k\}) d\mu d\rho \tag{10}$$

where $\Pr(\mu, \rho\,|\,r, \{n_k\})$ is given in (7) for $r > 2$, and is given in (8) for $r = 2$. Therefore, the weighted sums of $\mu$ and $\rho$ with the weight of $\Pr(\mu, \rho\,|\,r, \{n_k\})$ on a set of the final grid points can be used as the estimates of $E(\mu)$ and $E(\rho)$, respectively. Furthermore, using the evaluations of $\Pr(\mu, \rho\,|\,r, \{n_k\})$ on a set of the final grid points as the ordinates of the density, any quantile can be obtained through the GCIN procedure of IMSL. Therefore, given $0 < \alpha < 1$, any $1-\alpha$ confidence interval can be obtained. For example, finding the $q$ $(=1-\alpha)$ quantile, $\mu_q$, of $\mu$, then $(0, \mu_q)$ is an $1-\alpha$ asymmetric confidence interval for $\mu$.

For Bayesian Prediction in the beta binomial model, we predict the relative frequency of $k'$, the number of event occurrences in some future $r'$ opportunities for a specific member of the population, conditional on having observed $k$ out of $r$ for the same member and on data $\{n_k;\ k = 0, 1, \ldots, r\}$. In making these predictions, it is assumed that $p$ for each member of the population remains constant. Hence, the Bayes prediction formula turns out to be

$$\Pr(k'|r', k, r, \{n_k\}) = \int\int \Pr(k'|r', k, r, \alpha, \beta) \Pr(\mu, \rho|r, \{n_k\}) d\mu d\rho, \tag{11}$$

where $\Pr(k'|r', k, r, \alpha, \beta) = \Pr_{BB}(k'|r', \alpha + k, \beta + r - k)$. (In view of Lee and Sabavala, 1987). Again, the prediction of the relative frequency of $k'$ can be approximated by the weighted sum of $\Pr(k'|r', k, r, \alpha, \beta)$ with the weight of $\Pr(\mu, \rho|r, \{n_k\})$, on a set of the final grid points. When $r=2$ and $r'=1$ and $k'=1$, Lee and Sabavala (1987) presented the prediction formula (11) through a linear combination of complicated terms that are related to the beta functions.



## 3. SIMULATIONS

The main purpose of this section is to estimate the parameters, $\mu$ and $\rho$, for each of the following four models by using the computing simulation process described in Section 2. These four models, given by Lee and Sabavala (1987), represent quite distinct beta-binomial models. For more detail, see Lee and Sabavala (1987).

$$\begin{aligned} &\text{Model 1}: \mu = 0.40, \quad \rho = 0.50. \\ &\text{Model 2}: \mu = 0.40, \quad \rho = 0.25. \\ &\text{Model 3}: \mu = 0.50, \quad \rho = 0.33. \\ &\text{Model 4}: \mu = 0.80, \quad \rho = 0.29. \end{aligned}$$

To examine the impact of different priors, the prior distributions considered here are type A with $\gamma_1 = 1.6$, $\gamma_2 = 2.4$, $\delta_1 = 2.0$, and $\delta_2 = 2.0$; type B with $\gamma_1 = 2.0$, $\gamma_2 = 2.0$, $\delta_1 = 3.6$, and $\delta_2 = 1.2$; and type C with $\gamma_1 = 1.0$, $\gamma_2 = 1.0$, $\delta_1 = 1.0$, $\delta_2 = 1.0$. It is noted that the type C prior is the uniform prior, and type A and type B are non-uniform priors. Table I summarizes basic statistics for these priors. It is expected that type A could be the "best" for Model 1, and type B could be the "best" for $\mu$ in Model 3. But type B can be construed to be "misinformed". (See Lee and Sabavala, 1987). The other parameters in the simulation are the number of trials ($r$), the sample size ($n$), and the number of replications (100). Then, the procedures for given the number of trial $r$ and the number of sample size $n$ are as follow. First, generate $n$ values of $k$, the number of successes in $r$ opportunities from (3) by using a FORTRAN program, for a given model defined above. Second, The resulting frequency distribution, $\{n_k;\ k=0, 1, \ldots, r\}$, is used as an input in utilizing (7), (8), (9), and (10) to compute Bayes'

TABLE I Summary of basic statistics for three types of priors

| *Priors* | $\mu$ | | $\rho$ | |
|---|---|---|---|---|
| | *Mean* | *Variance* | *Mean* | *Variance* |
| A | 0.40 | 0.048 | 0.50 | 0.050 |
| B | 0.50 | 0.050 | 0.75 | 0.032 |
| C | 0.50 | 0.083 | 0.50 | 0.083 |

estimates of $\mu$ and $\rho$ with each type of priors. These steps are repeated for 100 times, and summary measures of the discrepancy between the estimates and the true values are computed. For point estimates of $\mu$ and $\rho$, the summary measures are the mean absolute deviations (MAD) and the squared root of mean square errors (RMSE). For interval estimate (at a given probability level), the summary measures are the average width of the interval (AW) and the proportion of intervals including the true value (INCL) among 100 replications. It takes about twelve minutes to run the entire process for each model and each priors with $r = 3$ in an Inter 486 with 16 MB RAMPC.

Tables II–VII summarize the simulation results for the four models with the three different priors. Table VIII shows the comparisons between symmetric and asymmetric intervals with uniform priors. The AW results in Tables II–VII are for symmetric intervals. The AW for asymmetric intervals are obtained as following: Measuring the length of the shortest confident interval among the symmetric interval and two asymmetric ones, mentioned in Section 2, for each replication, then the AW for asymmetric interval is the average of the lengths of the 100 shortest confident intervals. Therefore, if the AW for asymmetric intervals is smaller than the AW for symmetric intervals, then some asymmetric intervals could be better than symmetric ones. It is noted that the results for the case of two trials are similar to those from Lee and Sabavala (1987). Tables II–VII show the following common conclusions: The performances of Bayes' estimates by using Prior A and Prior B are uniformly better than the performances of Bayes' estimates by using Prior C for Model 1.

For Model 2, it seems that the Bayes' estimates for $\mu$ by using Prior B are better than the Bayes' estimates for $\mu$ by using the others; but the Bayes' estimates for $\rho$ by using Prior B are not better than the Bayes' estimates for $\rho$ by using the others. For Model 3, by using Prior B the

TABLE II Comparison of estimation with different priors using simulated data with $n = 20$ and $r = 2$

| *Estimator* | *Prior A* | | | | *Prior B* | | | | *Prior C* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator of $\mu$ | | | | | | | | | | | | |
| MAD | .066 | .067 | .066 | .073 | .064 | .057 | .062 | .071 | .071 | .062 | .070 | .061 |
| RMSE | .083 | .074 | .084 | .090 | .081 | .072 | .079 | .087 | .088 | .079 | .089 | .079 |
| AW | .280 | .267 | .277 | .241 | .287 | .274 | .283 | .246 | .288 | .272 | .283 | .232 |
| INCL | .89 | .96 | .90 | .82 | .92 | .95 | .90 | .83 | .88 | .93 | .88 | .87 |
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator for $\rho$ | | | | | | | | | | | | |
| MAD | .100 | .118 | .104 | .114 | .121 | .219 | .180 | .239 | .129 | .105 | .117 | .105 |
| RMSE | .121 | .146 | .132 | .146 | .146 | .242 | .209 | .258 | .154 | .138 | .145 | .138 |
| AW | .490 | .462 | .472 | .522 | .434 | .438 | .436 | .483 | .545 | .501 | .518 | .567 |
| INCL | .97 | .92 | .92 | .95 | .92 | .56 | .68 | .61 | .90 | .98 | .93 | .98 |

Note: The probability for interval estimates is .90.

TABLE III Comparison of estimation with different priors using simulated data with $n = 20$ and $r = 3$

| *Estimator* | *Prior A* | | | | *Prior B* | | | | *Prior C* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator of $\mu$ | | | | | | | | | | | | |
| MAD | .060 | .055 | .060 | .063 | .058 | .054 | .056 | .062 | .064 | .059 | .063 | .056 |
| RMSE | .077 | .070 | .075 | .081 | .075 | .068 | .071 | .080 | .082 | .074 | .079 | .072 |
| AW | .265 | .245 | .254 | .222 | .272 | .254 | .261 | .230 | .272 | .247 | .257 | .212 |
| INCL | .92 | .91 | .89 | .86 | .93 | .95 | .92 | .86 | .90 | .90 | .89 | .85 |
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator for $\rho$ | | | | | | | | | | | | |
| MAD | .090 | .096 | .077 | .101 | .101 | .166 | .115 | .193 | .107 | .095 | .091 | .100 |
| RMSE | .114 | .121 | .095 | .130 | .123 | .188 | .140 | .213 | .135 | .118 | .109 | .124 |
| AW | .409 | .390 | .395 | .457 | .380 | .380 | .379 | .437 | .441 | .417 | .427 | .488 |
| INCL | .94 | .92 | .99 | .92 | .89 | .64 | .78 | .69 | .93 | .96 | .97 | .95 |

Note: The probability for interval estimates is .90.

TABLE IV Comparison of estimation with different priors using simulated data with $n = 20$ and $r = 4$

| *Estimator* | *Prior A* | | | | *Prior B* | | | | *Prior C* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator of $\mu$ | | | | | | | | | | | | |
| MAD | .061 | .052 | .058 | .059 | .060 | .052 | .054 | .059 | .065 | .056 | .060 | .053 |
| RMSE | .078 | .066 | .074 | .076 | .076 | .065 | .069 | .076 | .083 | .070 | .077 | .067 |
| AW | .258 | .230 | .242 | .212 | .265 | .239 | .249 | .220 | .265 | .230 | .244 | .201 |
| INCL | .90 | .88 | .90 | .84 | .93 | .94 | .91 | .84 | .89 | .87 | .88 | .85 |
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator for $\rho$ | | | | | | | | | | | | |
| MAD | .076 | .077 | .075 | .099 | .087 | .127 | .100 | .168 | .086 | .079 | .086 | .097 |
| RMSE | .094 | .096 | .092 | .125 | .105 | .147 | .123 | .191 | .107 | .096 | .103 | .121 |
| AW | .367 | .345 | .352 | .416 | .350 | .345 | .345 | .408 | .388 | .363 | .373 | .435 |
| INCL | .96 | .93 | .94 | .90 | .92 | .75 | .82 | .71 | .93 | .95 | .95 | .91 |

Note: The probability for interval estimates is .90.

TABLE V Comparison of estimation with different priors using simulated data with $n = 30$ and $r = 2$

| *Estimator* | *Prior A* | | | | *Prior B* | | | | *Prior C* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator of $\mu$ | | | | | | | | | | | | |
| MAD | .061 | .054 | .061 | .053 | .060 | .052 | .058 | .052 | .064 | .056 | .063 | .046 |
| RMSE | .075 | .065 | .075 | .066 | .074 | .064 | .073 | .065 | .079 | .069 | .079 | .060 |
| AW | .236 | .222 | .231 | .196 | .241 | .228 | .235 | .201 | .240 | .224 | .233 | .190 |
| INCL | .93 | .93 | .89 | .89 | .92 | .94 | .90 | .90 | .91 | .92 | .87 | .90 |
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator for $\rho$ | | | | | | | | | | | | |
| MAD | .094 | .093 | .087 | .114 | .103 | .164 | .136 | .208 | .115 | .096 | .106 | .110 |
| RMSE | .119 | .127 | .116 | .145 | .126 | .194 | .167 | .233 | .145 | .129 | .131 | .143 |
| AW | .436 | .405 | .421 | .476 | .393 | .389 | .393 | .444 | .478 | .430 | .459 | .513 |
| INCL | .94 | .93 | .95 | .93 | .90 | .71 | .77 | .65 | .91 | .95 | .95 | .95 |

Note: The probability for interval estimates is .90.

TABLE VI Comparison of estimation with different priors using simulated data with $n = 30$ and $r = 3$

| *Estimator* | *Prior A* | | | | *Prior B* | | | | *Prior C* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator of $\mu$ | | | | | | | | | | | | |
| MAD | .057 | .050 | .053 | .048 | .055 | .048 | .051 | .048 | .059 | .052 | .055 | .044 |
| RMSE | .071 | .063 | .065 | .062 | .069 | .062 | .063 | .062 | .074 | .066 | .068 | .057 |
| AW | .221 | .200 | .211 | .181 | .226 | .207 | .216 | .186 | .226 | .200 | .213 | .174 |
| INCL | .91 | .92 | .92 | .88 | .91 | .95 | .94 | .89 | .89 | .88 | .89 | .89 |
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator for $\rho$ | | | | | | | | | | | | |
| MAD | .086 | .074 | .072 | .093 | .080 | .110 | .091 | .160 | .098 | .081 | .084 | .094 |
| RMSE | .108 | .098 | .090 | .119 | .104 | .138 | .114 | .181 | .123 | .105 | .104 | .114 |
| AW | .354 | .332 | .343 | .401 | .334 | .327 | .330 | .391 | .375 | .351 | .366 | .430 |
| INCL | .91 | .92 | .95 | .95 | .91 | .80 | .86 | .67 | .89 | .93 | .93 | .97 |

Note: The probability for interval estimates is .90.

TABLE VII Comparison of estimation with different priors using simulated data with $n = 30$ and $r = 4$

| *Estimator* | *Prior A* | | | | *Prior B* | | | | *Prior C* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator of $\mu$ | | | | | | | | | | | | |
| MAD | .056 | .048 | .052 | .046 | .054 | .047 | .049 | .046 | .058 | .050 | .053 | .043 |
| RMSE | .070 | .059 | .064 | .059 | .068 | .058 | .062 | .059 | .073 | .061 | .067 | .054 |
| AW | .215 | .188 | .200 | .171 | .219 | .195 | .205 | .177 | .219 | .188 | .201 | .164 |
| INCL | .91 | .92 | .88 | .88 | .91 | .93 | .91 | .90 | .88 | .92 | .87 | .86 |
| Model | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Estimator for $\rho$ | | | | | | | | | | | | |
| MAD | .074 | .065 | .069 | .085 | .071 | .088 | .080 | .132 | .081 | .072 | .077 | .086 |
| RMSE | .091 | .081 | .084 | .108 | .090 | .109 | .098 | .155 | .099 | .087 | .095 | .104 |
| AW | .312 | .291 | .300 | .361 | .302 | .291 | .295 | .357 | .325 | .304 | .314 | .375 |
| INCL | .92 | .94 | .94 | .89 | .92 | .84 | .86 | .72 | .89 | .93 | .90 | .93 |

Note: The probability for interval estimates is .90.

TABLE VIII Comparison of symmetric and asymmetric interval of $\rho$ using simulated data and uniform priors and $r = 2$

| | *AW* | | *INCL* | | *Proportion of* |
|---|---|---|---|---|---|
| | *Sym.* | *Asym.* | *Sym.* | *Asym.* | *asym. better* |
| $n = 20$ | | | | | |
| Model 1 | .545 | .538 | .90 | .83 | .27 |
| Model 2 | .501 | .484 | .98 | .95 | .60 |
| Model 3 | .518 | .506 | .93 | .91 | .45 |
| Model 4 | .567 | .542 | .98 | .96 | .67 |
| $n = 30$ | | | | | |
| Model 1 | .478 | .476 | .91 | .85 | .11 |
| Model 2 | .430 | .415 | .95 | .91 | .58 |
| Model 3 | .459 | .451 | .95 | .91 | .38 |
| Model 4 | .513 | .498 | .95 | .92 | .53 |

Note: The probability for interval estimates is .90.

Bayes' estimates for $\mu$ can be better than the Bayes' estimates by using Prior A; and Prior C can be the best Prior for the Bayes' estimates of $\mu$. But the Prior B is not better than either Prior A or Prior C for the Bayes' estimates of $\rho$. For Model 4, the Prior C is the best prior for the Bayes' estimates of $\mu$. Prior C can be better than Prior A for the Bayes' estimates of $\rho$. Prior B is not better than either Prior A or Prior C for the Bayes' estimates of $\rho$. In general, for a given sample size $n$, the estimates of $\mu$ and $\rho$ are more accurate for larger $r$; meanwhile, for a given $r$, the estimates for $\mu$ and $\rho$ are better for larger $n$. The symmetric and asymmetric intervals have also been compared. Table VIII summarizes the comparisons between symmetric and asymmetric intervals for $\rho$ with the uniform priors. Generally, the asymmetric interval for $\rho$ could be better than symmetric one. But the advantage of asymmetric interval for $\rho$ is not significant when $r > 2$. The asymmetric intervals for $\mu$ are never narrower than the symmetric intervals for $\mu$. Therefore, the results of the comparisons for $r > 2$ or for $\mu$ are not included in the table.

## 4. EXAMPLE

Lee and Sabavala (1987) illustrated the special meaning of the parameters $\mu$ and $\rho$ in the application of the beta-binomial model to the commercially available media models in detail. In this section, we will apply the estimation formulae (7), (9), and (10) and the prediction formula (11) to the television-viewing data sets taken from Lee and Sabavala (1987). These data were obtained from a sample of 60 homes located in a particular television market and collected for a study by a television network on the strength of loyalty for news programs. With complicated presentations for the estimation and the prediction formulae in Lee and Sabavala (1987); Lee and Sabavala (1987) first analyzed the data from day 1 and day 2 with the uniform priors, and suggested the Pearson type I distribution approximation to the marginal posteriors. Then, used the Pearson type approximated distributions as the priors to analyze the data from days 3–4. Finally, predicted the day 5 viewing by using the posteriors from the Bayes' analysis on day 3 and day 4 as input for the prediction formula with $r' = 1, k' = 1$, and $r = 2$. This approach was used under the beta-binomial model for the case of $r = 2$. In this section, for each program/time slot, we would like to use the observations of days 1–4 as input for the Bayesian estimates of $\mu$ and $\rho$ and the Bayesian predictions of day 5 viewing. Therefore, $r' = 1$, $k' = 1$, $r = 4$, and $\{n_k;\ k = 0, \ldots, 4\}$ as listed in Table IX, will be the data for estimation and prediction. In view of the empirical results from Sabavala and Morrison (1977), we will assume that $\mu$ and $\rho$ have independent prior distributions. The Prior A, Prior B, and Prior C will be used here as the prior distributions of $\mu$ and $\rho$. This approach is apparently different from the one used by Lee and Sabavala (1987).

Table IX lists the day 5 predictions and Tables X–XII list the means, standard deviations, and the percentiles of the estimated posterior distribution (for each of the parameters). The Bayesian intervals can easily be obtained from Tables X–XII. In view of these results, it can be seen that the point estimates (Tabs. X–XII) of each parameter under each program/time slot are slightly different with different priors, the posterior means and medians for each parameter in each program are fairly close, and the estimated posterior quantiles of each parameter in each program/time slot are almost independent of Priors A, B, and C. In general, the day 5 predictions for each program seem to be quite independent of the priors used and are better than the Bayesian predictions of Lee and Sabavala (1987). However, these prediction results are similar to those obtained by the "plug-in" method when the ML approach is used in the scheme proposed by Lee and Sabavala (1987).

TABLE IX Analysis of television-viewing data:Bayesian Prediction of Day 5

| *Program (slot)* | *No.* 0 | *Times* 1 | *Viewed* 2 | *on Days* 3 | 1−4 4 |
|---|---|---|---|---|---|
| 1. Number of Homes | 30 | 12 | 3 | 3 | 12 |
| Actual number viewing on Day 5 | 1 | 1 | 1 | 3 | 11 |
| Bayesian Day 5 Prediction 1 | 1.06 | 2.96 | 1.37 | 2.01 | 10.57 |
| Bayesian Day 5 Prediction 2 | 1.00 | 2.98 | 1.39 | 2.03 | 10.71 |
| Bayesian Day 5 Prediction 3 | 1.04 | 2.96 | 1.38 | 2.01 | 10.59 |
| 2. Number of Homes | 47 | 6 | 1 | 1 | 5 |
| Actual number viewing on Day 5 | 1 | 0 | 0 | 1 | 4 |
| Bayesian Day 5 Prediction 1 | .62 | 1.37 | .44 | .66 | 4.36 |
| Bayesian Day 5 Prediction 2 | .57 | 1.40 | .45 | .67 | 4.48 |
| Bayesian Day 5 Prediction 3 | .59 | 1.37 | .44 | .66 | 4.37 |
| 3. Number of Homes | 44 | 8 | 1 | 3 | 4 |
| Actual number viewing on Day 5 | 0 | 0 | 0 | 2 | 3 |
| Bayesian Day 5 Prediction 1 | 1.00 | 1.81 | .43 | 1.90 | 3.35 |
| Bayesian Day 5 Prediction 2 | .91 | 1.85 | .44 | 1.96 | 3.45 |
| Bayesian Day 5 Prediction 3 | .97 | 1.81 | .43 | 1.90 | 3.35 |
| 4. Number of Homes | 32 | 7 | 5 | 9 | 7 |
| Actual number viewing on Day 5 | 2 | 2 | 1 | 6 | 6 |
| Bayesian Day 5 Prediction 1 | 1.42 | 1.74 | 2.26 | 5.91 | 6.02 |
| Bayesian Day 5 Prediction 2 | 1.33 | 1.75 | 2.29 | 5.99 | 6.11 |
| Bayesian Day 5 Prediction 3 | 1.41 | 1.74 | 2.26 | 5.91 | 6.03 |

Prediction 1 with Prior A; Prediction 2 with Prior B; Prediction 3 with Prior C.

TABLE X Bayesian estimates with Prior A

| *Program (slot)* | | *Mean* | *SD* | 2.5% | 5.0% | 25% | 50% | 75% | 95% | 97.5% |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $\mu$ | .231 | .046 | .147 | .159 | .198 | .228 | .260 | .310 | .327 |
| | $\rho$ | .584 | .072 | .439 | .463 | .535 | .585 | .634 | .700 | .721 |
| 2 | $\mu$ | .097 | .031 | .046 | .052 | .075 | .094 | .116 | .154 | .168 |
| | $\rho$ | .614 | .098 | .416 | .448 | .549 | .618 | .684 | .769 | .794 |
| 3 | $\mu$ | .125 | .034 | .068 | .076 | .101 | .123 | .146 | .185 | .199 |
| | $\rho$ | .535 | .094 | .349 | .378 | .471 | .536 | .601 | .688 | .714 |
| 4 | $\mu$ | .243 | .044 | .162 | .173 | .212 | .241 | .272 | .319 | .335 |
| | $\rho$ | .532 | .073 | .388 | .411 | .483 | .533 | .583 | .651 | .672 |

TABLE XI Bayesian estimates with Prior B

| *Program (slot)* | | *Mean* | *SD* | 2.5% | 5.0% | 25% | 50% | 75% | 95% | 97.5% |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $\mu$ | .238 | .047 | .153 | .165 | .205 | .235 | .268 | .319 | .336 |
| | $\rho$ | .609 | .071 | .466 | .490 | .562 | .611 | .658 | .722 | .742 |
| 2 | $\mu$ | .106 | .033 | .051 | .058 | .082 | .103 | .126 | .167 | .181 |
| | $\rho$ | .663 | .093 | .470 | .502 | .601 | .667 | .729 | .807 | .829 |
| 3 | $\mu$ | .134 | .035 | .074 | .082 | .109 | .131 | .156 | .197 | .211 |
| | $\rho$ | .580 | .091 | .396 | .426 | .518 | .582 | .644 | .727 | .751 |
| 4 | $\mu$ | .249 | .045 | .166 | .178 | .217 | .247 | .278 | .326 | .342 |
| | $\rho$ | .558 | .071 | .416 | .439 | .510 | .559 | .608 | .674 | .695 |

TABLE XII Bayesian estimates with Prior C

| *Program* (*slot*) | | *Mean* | *SD* | 2.5% | 5.0% | 25% | 50% | 75% | 95% | 97.5% |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $\mu$ | .229 | .047 | .145 | .157 | .196 | .227 | .259 | .310 | .327 |
| | $\rho$ | .587 | .074 | .439 | .463 | .538 | .589 | .638 | .706 | .726 |
| 2 | $\mu$ | .094 | .031 | .042 | .049 | .071 | .090 | .113 | .150 | .165 |
| | $\rho$ | .620 | .101 | .413 | .446 | .552 | .624 | .693 | .781 | .806 |
| 3 | $\mu$ | .122 | .034 | .065 | .072 | .098 | .119 | .143 | .182 | .196 |
| | $\rho$ | .535 | .097 | .342 | .372 | .468 | .536 | .603 | .693 | .721 |
| 4 | $\mu$ | .242 | .045 | .160 | .172 | .210 | .240 | .271 | .318 | .335 |
| | $\rho$ | .534 | .074 | .386 | .410 | .483 | .535 | .585 | .654 | .676 |



## 5. CONCLUDING REMARK

Bayesian method for the beta-binomial model has been shown to be a viable alternative to the maximum likelihood approach. By using numerical integration, the Bayesian approach for the beta-binomial model can be fruitfully implemented in real life applications.

### *Acknowledgments*

The authors are grateful to the Associate Editor, the referee, and Professor Don Edwards, for additional references and suggestions which much improved this paper.

Jack C. Lee is grateful for the research support from the National Science Council of Taiwan (88-2118-M-009-005).

### *References*

Green, J. D. (1970) Personal Media Probabilities. *Journal of Advertising Research*, **10**, 12–18.

Griffiths, D. A. (1973) Maximum Likelihood Estimation for the Beta Binomial Distribution and an Application to the Household Distribution of the Total Number of Cases of a Disease. *Biometrics*, **29**, 637–648.

Huynth, H. (1979) Statistical Inference for Two Reliability Indices in Mastery Testing Based on the Beta Binomial Model. *Journal of Educational Statistics*, **4**, 231–246.

International Mathematical and Statistical Libraries, Inc. (1991) IMSL, Houston, Texas.

Kahn Michael, J. and Raftery Adrian, E. (1996) Discharge Rates of Medicare Stroke Patients to Skilled Nursing Facilities: Bayesian Logistic Regression with Unobserved Heterogeneity. *Journal of the American Statistical Association*, **91**, 29–41.

Kalwani, M. U. (1980) Maximum Likelihood Estimation of Zero Order Models Given Variable Numbers of Purchases per Household. *Journal of Marketing Research*, **17**, 547–551.

Lee, J. C. and Sabavala, D. J. (1987) Bayesian Estimation and Prediction for the Beta-Binomial Model. *Journal of Business & Economic Statistics*, **5**, 357–367.

Lord, F. M. (1965) A Strong True-Score Theory, with Applications. *Psychometrika*, **30**, 234–270.

Massy, W. F., Montgomery, D. B. and Morrison, D. G. (1970) Stochastic Models of Buying Behavior. Cambridge MA, MIT Press.

Morrison, D. G. (1966) Testing Brand-Switching Models. *Journal of Marketing Research*, **3**, 401–409.

Pearson, E. S. (1925) Bayes' Theorem in the Light of Experimental Sampling. *Biometrika*, **17**, 388–442.

Sabavala, D. J. and Morrison, D. G. (1977) A Model of TV Show Loyalty. *Journal of Advertising Research*, **17**, 35–43.

Skellam, J. G. (1948) A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable between the Sets of Trials. *Journal of the Royal Statistical Society Ser. B*, **10**, 257–261.

Wilcox, R. R. (1979) Estimating the Parameters of the Beta Binomial Distribution. *Educational and Psychological Measurement*, **39**, 527–535.

Williams, D. A. (1975) The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity. *Biometrics*, **31**, 949–952.