THE BIVARIATE BETA DISTRIBUTION: COMPARISON OF MONTE CARLO GENERATORS AND EVALUATION OF PARAMETER ESTIMATES

James H. Macomber
Buddy L. Myers

ABSTRACT

The bivariate and multivariate beta distributions may provide appropriate stochastic models for a number of processes, particularly those involving random proportions. Researchers may therefore find it necessary to estimate the parameters of such distributions or generate Monte Carlo samples with known parameter values. Two possible generating techniques for beta bivariates are presented and compared in this paper. Estimating equations for the three parameters of the bivariate beta distribution are presented. These use the method of moments, the only tractable estimating technique, and an analysis of their properties is also presented.

This paper focuses on the bivariate beta distribution, but a user of a higher-dimensioned beta model will be able to make use of the discussion herein to provide assistance in determining many of the properties of such a model.

INTRODUCTION

The continued interest in simulation and Monte Carlo techniques as research tools implies a need to continue to develop their capabilities. We seek to do this in this project by extending knowledge about Monte Carlo generation of a particular statistical model: the bivariate beta distribution. In addition, we seek to evaluate estimates of the distribution's parameters derived by the method of moments. When a researcher might apply a bivariate beta model, he would sample from some bivariate beta-distributed population estimate the parameters of the distribution, and possibly use these parameter estimates to provide the basis for generating Monte Carlo samples. The results of our project will help provide guidance in such activities involving bivariate beta distributed populations, and from this work a researcher may be able to infer the properties of higher-dimensioned multivariate beta distributions.

The researcher wants to know, of course, when the bivariate, or multivariate, beta distribution may provide an appropriate statistical model for some real-world process. The beta distribution describes the random division of a continuous interval. The univariate beta distribution describes one such segmentation; the bivariate beta describes two such segmentations, and so forth. A principal use of the general multivariate beta distribution would therefore lie in modeling random proportions, when it is reasonable to assume that they are continuous.

The density function of the bivariate beta distribution is given by:

$$f_\beta^{(3)}(p_1,p_2|m_j) = \begin{cases} \dfrac{\Gamma(m_1+m_2+m_3)}{\Gamma(m_1)\Gamma(m_2)\Gamma(m_3)}\, p_1^{m_1-1}p_2^{m_2-1}(1-p_1-p_2)^{m_3-1} \text{ for} \\ \qquad p_1,p_2>0,\ p_1+p_2 \le 1 \qquad (1) \\ 0, \text{ elsewhere.} \end{cases}$$

The $m_j > 0$ for $j = 1, 2, 3$. These are the parameters of the distribution and may or may not be integral values. The letter gamma represents the gamma function. For example,

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx = a!/a .$$

In equation 1, p_1 and p_2 represent a pair of observations; that is, a beta bivariate.

Mauldon (9) has examined some of the mathematical properties of the general multivariate beta distribution. These include uniqueness, continuity, boundedness, and invariance under linear transformation of observed values. Note that, because of this last property, p_1 and p_2 can actually take on any values in the $(0,z)$ interval as long as the following condition is met:

$$p_1 + p_2 \le z .$$

In this paper, we are limiting ourselves to studying what Mauldon refers to as the basic beta distribution where the constraints noted in equation 1 hold. In addition to this, Mauldon shows that any marginal distribution of a k-dimensional beta distribution is itself a (k-1)-dimensional beta distribution. Mosimann (10) has also made contributions regarding properties of the multivariate beta distribution. He derives expressions for its first and second moments as well as expectations for the elements of its variance-covariance matrix. In addition, he demonstrates a direct relationship between the k-dimensional beta distribution and a set of (k+1) independent gamma distributions. Finally, Fielitz and Myers (2) have derived estimating equations for the three parameters of the bivariate beta distribution by using the method of moments. It appears, incidentally, that there is no other feasible estimation technique for the parameters of the bivariate or multivariate beta distribution. (13)

Several other applications of the bivariate beta distribution as a statistical model have also been suggested in the literature. One of these involves interpretation of spurious correlations, or correlations among proportions. (10) Pearson's product-moment correlation coefficient is inappropriate as a measure of association when observed values are constrained as they are in the case of random proportions. The correlation among proportions can be more appropriately measured through the covariance of a particular set of beta multivariates. If there are three classes to be considered, the bivariate beta distribution could provide a vehicle for interpreting correlations among them. The multivariate beta distribution may also be useful in modeling variable transition probabilities in a Markov chain where a Bayesian scheme is used to estimate steady-state probabilities or where a Markovian model is based on macro data and individual transitions from one state to another are not available. (8,6) A final suggested application comes from Quandt. (12) Stochastic models involving voter preferences may make use of a bivariate beta distribution when, say, there are three possible attitudes toward an issue (for, against, and undecided), and the proportion of voters in each attitude class is of interest.

Since there are a number of areas where a bivariate, or multivariate, beta distribution may provide an appropriate statistical model, we have undertaken this project in order to increase its potential usefulness. Specifically, we seek to develop a useful Monte Carlo generator for beta bivariates when parameter values are known. In addition, the Monte Carlo samples will be used to estimate the parameters of the distribution from which the samples come. This will allow comparison of known, true parameter values with estimated values in turn enabling us to evaluate certain estimator properties.

GENERATION OF BETA BIVARIATES

The qualities of an acceptable bivariate beta generator are mathematical consistency with the density function and an ability to function for any feasible set of parameter values. These are of course in addition to the usual attributes of speed, low cost, reproducability, ease of implementation, and reliability as evidenced by an ability for samples to pass statistical tests. One suggested generator, based on order statistics, was rejected at the outset because it would work only for integral parameter values. Two others which could potentially possess all of the stated qualities were developed and these generators will now be discussed.

AN INVERSION GENERATOR

The first generator is based on what is commonly called the inversion method (14, pp. 315-316.) and requires derivation of equations for both marginal and conditional mass functions. Since the derivation is lengthy, it is not presented here. We can make use of incomplete beta functions for this generator because, as noted earlier, the marginal distribution of a bivariate beta is itself a univariate beta distribution (9), and it can then be shown that the conditional distribution is a transformed univariate beta. The equations for the generator are

$$R_{j-1} = K_1 \int_0^{p_1} p_1^{m_1-1}(1-p_1)^{m_2+m_3-1}dp_1 \,, \tag{2}$$

$$R_j = K_2 \int_0^{x^*} x^{*m_2-1}(1-x^*)^{m_3-1}dx^* \,, \text{ and} \tag{3}$$

$$X^* = p_2/(1-p_1) \,, \text{ where} \tag{4}$$

R_{j-1} and R_j are random numbers uniformly distributed over the (0,1) interval (for j= 1,2...,2n),

$$K_1 = \frac{\Gamma(m_1+m_2+m_3)}{\Gamma(m_1)\Gamma(m_2+m_3)} \,, \text{ and}$$

$$K_2 = \frac{\Gamma(m_1+m_2+m_3)}{\Gamma(m_2)\Gamma(m_3)(1-p_1)^{m_2+m_3-1}} \,.$$

By drawing a pair of uniform random numbers we can sequentially solve equations 2, 3, and 4, obtaining a beta bivariate represented by p_1 and p_2, and having parameters m_1, m_2, and m_3. From an input-output standpoint, this generator is desirable since one pair of uniform random numbers yields one bivariate beta pair.

Unfortunately, equations 2, 3, and 4 are not so easily solved. Equations 2 and 3 contain integrals which can not make use of direct quadrature formulas; a numeric technique must be used. We found that, for given p_1 and x^* values, Simpson's rule works reasonably well in evaluating the integral terms, particularly with large parameter values since convergence is fairly rapid. Finding p_1 and x^* which satisfy the equations represents another problem. We settled on a binary search technique, setting initial values at 0.5. The equation is evaluated and a second value, either 0.25 or 0.75, is chosen. Subsequent steps are halved and we are able to converge on p_1 and x^* values which provide approximate solutions to equations 2 and 3. In outline form, our algorithm works like this:

1. Draw uniform random number, R.

2. Set $p_1 = 0.5, \Delta p_1 = 0.25$.

3. Evaluate equation 2 (Simpson's rule sensitivity set at 10^{-8}).

4. If difference between left and right side

is less than 10^{-7} go to step 7.

5. If R is greater than the right side, $p_1 = p_1 + \Delta p_1$, else $p_1 = p_1 - \Delta p_1$.

6. $\Delta p_1 = \Delta p_1 / 2$, go to 3.

7. Repeat steps 1 through 6 for x*.

8. $p_2 = x^*/(1-p_1)$

9. Repeat steps 1 through 8 n times for sample of n beta bivariates, p_1 and p_2.

This algorithm seems to have several of the advantages mentioned. As we shall see, however, it is seriously deficient in terms of speed. We could speed it up by reducing the precision noted in steps 3 and 4, but this would obviously have an adverse effect on sample quality.

A TRANSFORMATION GENERATOR

The second generator is based on a transformation of three independent gamma univariates all having identical scale parameters and having m_1, m_2, and m_3 as respective shape parameters. The transformation equations follow directly from Mosimann's proof (10, pp. 74-75.). They are

$$p_1 = \frac{Y_1}{Y_1+Y_2+Y_3} \quad \text{and} \quad p_2 = \frac{Y_2}{Y_1+Y_2+Y_3}\,, \qquad (5)$$

where $Y_1 \sim \Gamma(m_1,a)$, $Y_2 \sim \Gamma(m_2,a)$, and $Y_3 \sim \Gamma(m_3,a)$. The p_1 and p_2 pair is a beta bivariate. This generator thus requires three gamma univariates which, if we let a =1 , is a simple task. Fishman (3, pp. 203-211.) provides a useful set of algorithms for generating the required gamma univariates. He also presents an algorithm for beta univariates using a parallel transformation with two gamma univariates instead of three. This generator can thus be regarded as a multivariate extension of a well-known univariate generator.

One minor difficulty with this generating scheme is its large appetite for uniform random numbers. It requires up to $(9 + k_1 + k_2 + k_3)$ uniform random numbers where k_i is the largest integer in m_i. For example, a sample of 100 bets bivariates having $m_1 = 9\cdot5$, $m_2 = 9\cdot5$, and $m_3 = 9\cdot5$ would require 100(9 + 9 + 9 + 9) = 3600 uniform random numbers. From the input-output standpoint, this generator is much less efficient than the inversion generator; the latter requires only 200 uniform random numbers for the same hypothetical example. The speed of the transformation generator is so much greater though that this comparison may be insignificant. The algorithm for the transformation generator is outlined as follows:

1. Set i = 1.

2. Compute k, the largest integer in m_i.

3. Compute $b = m_i - k$.

4. If k = 0, set V=0, h=0, and go to step 9.

5. Generate uniform random numbers R_h for h=1,...,k.

6. Compute $V = -\ln(\prod_{h=1}^{k} R_h)$.

7. If b > 0, go to step 9.

8. Set W = 0, Z = 0, and go to step 13.

9. Generate R_{h+1}, R_{h+2}, R_{h+3}

10. Compute $Z = -\ln(R_{h+1})$

11. Compute $w_1 = (R_{h+2})^{1/b}$ and $w_2 = R_{h+3})^{1/(1-b)}$

12. Compute $W = w_1/(w_1+w_2)$

13. Compute $Y_i = V + WZ$

14. Compute i = i+1.

15. If $i \leq 3$, go to step 2.

16. Compute $p_1 = Y_1/(Y_1+Y_2+Y_3)$ and $p_2 = Y_2/(Y_1+Y_2+Y_3)$; p_1 and p_2 are a beta bivariate pair.

17. Repeat steps 1 through 16 n times for a sample of n beta bivariates having parameters m_1, m_2, and m_3.

Note that the branch routine including steps 9 through 12 is used for nonintegral parameter values. This algorithm, although perhaps less straightforward than that of the inversion generator, involves a much less complex set of computations and is therefore much faster.

SIMULATION PROCEDURE AND SAMPLE TESTING

This section describes how we have utilized the generators described and how we have evaluated the parameter estimates obtained from the Monte Carlo samples.

CHOOSING A UNIFORM GENERATOR

One potentially serious problem with uniform random generators used in multivariate applications is that many of them have been shown to produce clearly nonrandom sequences in n-space. This is due to the inability of a cyclical sequence of K numbers to locate all of the k^2 points in two-dimensional space, and so forth for higher-dimensional space. Kennedy (4) and Lewis (7) have done considerable work with what is termed the generalized feedback shift register (GFSR) generator. The GFSR generator has several advantages over more commonly encountered congruential generators. One of these is its ability to repeat values within a full period of its cycle, and thus it has potential n-space uniformity. Another advantage is its transportability; it can produce identical sequences on different computers. For these reasons, the GFSR generator was chosen for this project. FORTRAN function subroutines which duplicated sequences obtained by Kennedy were in existence for the Honeywell system at Kent State University, the system used for this study. Since the GFSR generator has been thoroughly tested for both fit and randomness (4, pp. 10-11), our ability to duplicate a tested sequence eliminates the problem of having to test it again.

Several parameters are specified by the user of the GFSR generator routines. As long as these are carefully chosen in accordance with tested results, there should be no difficulties with the generated sequences. The generator makes use of a primitive polynomial, $x^p + x^q + 1$, and a specified delay, d, which indicates the number of columns skipped in randomly taking a set of binary digits from a random matrix of binary digits. The parameters p, q, and d are user-selected in a particular fashion. Values of p = 98, q = 27, and d $\geq$ 100p were used in our sampling since they have been shown to give satisfactory results. The chosen q value should not be close to zero, p/2, or p (4, p. 10). For the above p and q values, d should be relatively prime to 2^p-1 (4, p. 7). The period of the resulting sequence will be 2^p-1, or $2^{98}-1$ in our case. This is certainly a sufficient period for our purposes since we used sequences no longer than 2^{18} for any set of samples. Also note that by reinitializing the generator with a new d value we can create a new sequence with period $2^{98}-1$. Once the GFSR generator is initialized, it is fairly rapid and, because of its other advantages including potential n-space randomness, we elected to use it in this study.

TESTING SAMPLES

In order to empirically establish the validity of our generators in production of acceptable Monte Carlo samples, several tests are performed on each sample to compare the samples with theoretical bivariate beta distributions. Tests were used to detect departures from both fit and randomness.

First, a Kolmogorov test is performed on each marginal sample distribution. That is, each set of n p_1 values and each set of n p_2 values are tested against the theoretical beta distribution from which they should have come. Since we use a computerized routine to perform these tests, it is necessary to specify discrete intervals in the domains of p_1 and p_2 respectively. These intervals are chosen arbitrarily to be 0.001 over each domain's interval (0,1). By using Simpson's rule to evaluate the incomplete beta functions, $f(p_1)$ and $f(p_2)$, for each increment of p_1 and p_2, we are able to closely approximate the theoretical values of these functions. The routine developed then finds the maximum vertical distance between the theoretical steps and the ordered sample's mass function. This distance is then compared with internally stored Kolmogorov critical values (α= 0.05) and, if this distance is excessive, the sample is rejected and another is drawn. Note that rejection of either sample set, p_1 or p_2 values, causes a complete new sample to be drawn.

The advantage of this test is that is exact for small sample sizes. Since we are interested in small to moderately large sample sizes, it is appropriate for our purposes. A disadvantage is the necessity to rely on discrete intervals in the population distribution function. For large parameter values, the bivariate beta distribution is extremely concentrated, as will be seen shortly. The proportion of samples, for given sample size, which fail the Kolmogorov test of fit increases somewhat as parameters become large. We feel that this is at least partly due to the construction of the test itself. In this case many sample observations may lie between two successive discrete interval values. This probably causes misleading measurements of the maximum vertical distance in some cases and therefore a number of otherwise acceptable samples may be rejected. It can also be argued, however, that the test becomes more conservative with fixed discrete intervals on the domains as parameter values increase. Discrete intervals smaller than 0.001 may be in order but the cost in additional computing time for the theoretical distribution functions is rather high.

Another disadvantage is that we are independantly testing sample elements which are in fact paired. Clearly, it would be preferable to test the two dimensions jointly rather than to perform independent tests along each dimension. Unfortunately, two-sample variations of Kolmogorov-type tests, such as the Cramér-von Mises two-sample test, generally assume independence of pairs and hypothesize that two unknown distributions are identical or not identical. Neither of these actions is consistent with testing the fit of beta bivariates. There is apparently no workable method of testing the complete sample jointly along both dimensions through the use of a Kolmogorov-type test.

For larger sample sizes, however, we are able to overcome this problem by using a chi-square test of fit. We use a technique similar to the algorithm described for the inversion generator to establish an isoprobability grid of the sample space. The main difference is that we use known mass function values rather than random numbers. For each parameter combination, the computer routine calculates a set of p_1 and p_2 values which segment the sample space into 25 isoprobability regions. This is equivalent to establishing a 5 x 5 contingency table having equal expected cell frequencies. This design allows us to relax the usual requirement of 5 observations per cell (1, pp. 142-143.) and we test all samples where $n \geq 50$. A routine classifies all observed pairs into one of the 25 isoprobability cells, and differences between actual and expected cell frequencies provides a means of calculating a test statistic which is approximately chi-square distributed with 16 degrees of freedom. If our test statistic exceeds the chi-square critical value (α = 0.05), the Monte Carlo sample is rejected.

A test using fixed p_1 and p_2 values to set up a grid would be of little value since all of the parameters of the bivariate beta distribution can be regarded as shape parameters. We would be faced with situations where expected cell frequencies would vary greatly and the assumptions underlying this sort of chi-square test would be violated.

Note that for each parameter combination, a different set of lines defining the mesh is determined. After lines defining 20 percent slices are found along the domain of p_1, conditional p_2 values are found for each slice. The p_2 slices will not be straight lines through the sample space but rather will be staggered as each four percent isoprobability cell is constructed.

It might be preferable to use different mesh sizes in the grid for different sample sizes. That is, we could establish smaller expected frequencies for large sample sizes with a finer mesh having, say, 36 cells rather than 25. We felt that the possible increase in information about the Monte Carlo samples would probably not be worth the cost in additional computing time which is fairly substantial for just 25 cells. The present design of the test permits it to be used for all feasible parameter combinations and sample sizes. Furthermore, it allows us to test simultaneously an entire bivariate beta sample for theoretical fit and thus makes use of more of the information in a sample than the Kolmogorov tests performed on each marginal distribution separately.

One sample quality which is often ignored in random variate generation is that of randomness. Researchers generally assume that, if the uniform random number generator which they are using produces sequences which pass tests of randomness, then this quality will carry over into transformed variates. It is possible for multidimensional transformations to cause a loss of apparent randomness in a sequence (7, p. 4). Because of this, we are also interested in testing randomness in the sample sequences, or perhaps it is better to say detecting departures from randomness.

In order to approach this problem, a survey of many nonparametric tests of randomness in paired observations was conducted (1, among others). Several tests appeared to hold promise but had to be rejected upon close examination. For example, the Olmstead-Tukey test of association was considered. The assumptions of the test are met by the samples to be considered (1, p. 336). However, the hypotheses used in applying this as a test of randomness are

H_o: There is no serial correlation between observed pairs (the p_1 and p_2 values in this case) , and

H_a: There is significant serial correlation between the pairs.

Given the restrictions on p_1 and p_2 noted earlier in equation 1, we would expect that large p_1 values would be accompanied by small p_2 values and vice versa. We would expect a significant negative correlation between p_1 and p_2 values. This version of the Olmstead-Tukey test therefore would not be useful in examining randomness in a set of beta bivariates.

We also considered tests based on the ranks of differences between observed pairs. However, rank tests generally require assumptions of symmetry in the distribution of differences. Since beta distributions are not as a rule symmetric, there is no reason to expect symmetry in differences between pairs. Thus this sort of test also had to be discarded.

Most other classical tests of randomness also had to be rejected. The gap test, for example, is based on the length of strings of digits between recurring pairs of digits. It is reasonably straightforward where any successive digit is equally likely, or independent of its predecessor, as with uniform univariates. This is not the case with the bivariate beta distribution, and determining theoretical gaps in this case would be a monumental, if not impossible, task.

The only test of randomness which we are able to readily apply to bivariate beta sequences is the Wald-Wolfowitz runs test for lengths of runs above and below the median. This test is applied three times to each sample set, once each on the set of n p_1 values, on the set of n p_2 values, and on the set of n (p_1+p_2) values. The first two tests can detect departures from randomness only along one dimension or the other. The last test on (p_1+p_2) values represents an attempt to use as much of the information contained in the complete sample as possible; some of course will be lost in degrading the data from interval to nominal, binary coding. The hypotheses of the test are

H_o: The sequence is generated by a random process, and

H_a: The variables in the sequence are either dependent on other variables in the sequence or distributed differently from one another.

This is a situation where rejection of the null in too many cases (More than 1 in 20 samples for α = 0.05.) would lead one to suspect a lack of randomness in the generated sequences. Consistent failure to reject the null does not establish randomness but it does increase one's confidence in the quality of the generated sequences.

This discussion points out a need for further development of techniques for detecting various forms of nonrandomness in Monte Carlo generated sequences. Since transformations may cause a degradation of randomness, it seems appropriate that users of such sequences attempt to empirically detect such a problem. Unfortunately, tests currently available do not seem to be fully capable of doing so, particularly in multivariate cases such as the asymmetric bivariate beta distribution. Development of new tests or modification of existing ones would seem to be in order.

SELECTION OF TRUE PARAMETER VALUES, SAMPLE SIZES, AND REPLICATIONS.

Our first task was to determine which of the two proposed generators performs better at producing beta bivariates. We quickly found that, based on the statistical tests of fit and randomness that we were able to use, each generator produces generally acceptable bivariate beta sequences. There is, however, a great difference in the computing time required. The inversion generater performs very poorly, even with large parameter values where convergence is very rapid in our application of Simpson's rule to evaluate the integrals in equations 2 and 3. In spite of its consumption of uniform random numbers, the transformation generator is clearly preferable as will be seen in the results section.

Since we are interested in examining the properties of method-of-moments parameter estimates for the bivariate beta distribution, we must generate a large number of bivariate beta parameter values and sample sizes. The transformation generator is used for this purpose.

The bivariate beta parameters m_1, m_2, and m_3 can take on any positive, real values. Clearly, not all of the infinite number of possible parameter combinations can be examined in terms of a general Monte Carlo experiment involving the bivariate beta distribution. We should, however, involve as many parameter combinations as are manageable and at the same time represent a broad spectrum of possibilities.

In order to select the experimental combinations of m_1, m_2, and m_3, each parameter was systematically varied from 0.5 through 10.0, in increments of 0.5. For each of the resulting 8000 combinations, the mean values, the variances, and the covariance were computed using equations given by Mosimann for the general multivariate beta distribution and adapted for the bivariate case (10, p. 68). These formulae are:

$$E(p_1) = \frac{m_1}{m_1+m_2+m_3} \quad \text{and} \quad E(p_2) = \frac{m_2}{m_1+m_2+m_3} \quad ;$$

$$E(\mathrm{Var}\ p_1) = \frac{m_1(m_2+m_3)}{(m_1+m_2+m_3+1)(m_1+m_2+m_3)^2} \quad \text{and}$$

$$E(\mathrm{Var}\ p_2) = \frac{m_2(m_2+m_3)}{(m_1+m_2+m_3+1)(m_1+m_2+m_3)^2} \quad ; \quad \text{and}$$

$$E(\mathrm{Cov}\ p_1\ p_2) = -\sqrt{\frac{m_1\ m_2}{(m_2+m_3)(m_1+m_3)}} \quad .$$

The experimental true parameter values are selected from the 8000 examined combinations with an eye toward fairly even representation of the possible theoretical values of the preceding expected statistics. Based on this admittedly arbitrary criterion, we believe that our evaluation of parameter estimates can be adequately carried out using the following true parameter values:

$$m_1 = \{1.5,\ 3.0,\ 5.0,\ 7.0,\ 8.5,\ 10.0\}$$

$$m_2 = \{1.0,\ 2.5,\ 5.0,\ 7.5,\ 10.0\}, \text{ and}$$

$$m_3 = \{1.5,\ 5.0,\ 7.5,\ 10.0\}.$$

This results in a sample of 120 parameter combinations from a theoretically infinite number. Parameter planes are reasonably evenly filled out, and a good representation of distribution statistic values is brought into the experiment. At the same time, a manageable number of combinations is selected.

Sample sizes are chosen with the intent of examining small to medium-large sample properties of both Monte Carlo samples and parameter estimates. With this in mind, sample sizes of 10, 15, 20, 30, 40, 50, 100, 150, and 200 are drawn. This selection will enable examination of small-sample properties as well as provide information on asymptotic properties of parameter estimators. Larger sample sizes may be regarded as advisable in examining asymptotic properties, but early computer runs suggested that little would be gained in terms of additional experimental information at the expense of significantly increased computing time.

The number of replications of each sample-size, parameter-set combination to be performed was chosen by considering both theoretical and empirical factors. A substantial number are required so that we can obtain distributions of parameter-estimate bias. The central limit theorem indicates that sampling distributions of parameter estimates will become normal as the number of samples becomes infinite. The number of samples recommended for the use of a normal approximation is generally described as "large." Practically speaking, the number 30 is usually regarded as sufficient for invoking the central limit theorem where it is applicable, which rests largely on the sampling distribution having finite variance. This is the case with method-of-moments parameter estimates for the bivariate beta distribution since all parameter estimates will be real, positive numbers. To verify the choice of 30 replications, several sensitivity analyses were performed. Using several sample-size, parameter-set combinations, runs ranging from 20 to 40 replications were performed in order to assess the sampling distributions of parameter estimates. The variances of these distributions tended to stabilize between 25 and 35 replications depending on parameter values and sample size. We believe therefore that our choice of 30 replications of each combination is sufficient to provide reliable information on the behavior of method-of moments parameter estimates.

In summary, 120 parameter sets, nine sample sizes, and 30 replications will be used to generate the data for evaluating parameter estimates. This means that a total of 32,400 bivariate beta distributed samples must be drawn. This total does not include the runs performed to evaluate the alternative generators. A discussion of the methods used to analyze this data is presented in the next section.

EVALUATION OF PARAMETER ESTIMATORS

The method of moments appears to be the only tractable method of estimating the parameters of a bivariate beta distribution through sample observations. The method of maximum likelihood, for example, usually gives consistent, asymptotically efficient, and asymptotically normally-distributed parameter estimates. We attempted to derive maximum likelihood estimators but found that the likelihood function, when differentiated with respect to each parameter, does not have an analytical solution. In fact we wound up with an infinite number of terms to be summed with each term itself an infinite sum. Given this difficulty, there is little reason to expect that even some approximation to the maximum likelihood estimators would be worth the trouble. Similar analytic difficulties arise with other estimating techniques such as minimum chi-square and so forth. The basis of the difficulty lies in being faced with an undifferentiable function.

Method-of-moments estimators are generally mean-square-error consistent and asymptotically normal. They are often useful when other techniques get bogged down in mathematical manipulations, certainly the case with the bivariate beta distribution. Exact statistical properties must, however, be established in each case individually (5, p. 172). We will first outline the derivation of the estimators and then discuss the properties which are to be analyzed.

First, let $M = m_1+m_2+m_3$, and then let

$$E(p_1) = \frac{m_1}{M} = x_{11} , \qquad (6)$$

$$E(p_2) = \frac{m_2}{M} = x_{21} , \text{ and} \qquad (7)$$

$$E(p_1^2) = \frac{m_1(m_2+m_3)+m_1^2(M+1)}{(M^2)(M+1)} = x_{12}, \qquad (8)$$

where x_{11}, x_{21}, and x_{12} are sample moments obtained from

$$x_{11} = \frac{1}{n} \sum_{j=1}^{n} p_{1j} \qquad (9)$$

$$x_{21} = \frac{1}{n} \sum_{j=1}^{n} p_{2j} \qquad (10)$$

$$x_{12} = \frac{1}{n} \sum_{j=1}^{n} p_{1j}^2 \qquad (11)$$

where the j subscript is added to the observed values to index sample elements. The two first-order moment equations and the first second-order moment equation have been selected, giving three equations with three unknowns (m_1, m_2, and m_3). We could select other moment combinations, but symmetry of approach suggests this one (2, p. 8). Note that the third first-order moment is a linear combination of the other two. Hence one second-order moment is necessary for three independent equations.

Sample data are used to solve equation 9, 10, and 11 for x_{11}, x_{21}, and x_{12} respectively. These values then go into the equations below to solve for m_1^*, m_2^*, and m_3^*, estimates of m_1, m_2, and m_3. From equations 6, 7, and 8, the computational formulas are

$$m_1^* = \frac{(x_{11}-x_{12})x_{11}}{x_{12}-(x_{11})^2} , \qquad (12)$$

$$m_2^* = \frac{(x_{11}-x_{12})x_{21}}{x_{12}-(x_{11})^2} , \text{ and} \qquad (13)$$

$$m_3^* = \frac{(x_{11}-x_{12})(1-x_{11}-x_{21})}{x_{12}-(x_{11})^2} . \qquad (14)$$

For a more complete discussion of these equations and their derivations, see Fielitz and Myers (2, pp. 9-12).

Desirable small-sample estimator properties are unbiasedness, minimum variance, and linearity. "BLU" estimators possess these properties. The estimators we have shown here may or may not be best in the minimum variance sense. We have no way of evaluating this property because we have no other estimators with which to compare the method-of-moments estimators and because Cramer-Rao lower bounds are intractable. They are also not linear since a second-order moment must be used in calculating them.

We can, however, empirically evaluate their bias. For each sample-size, parameter-set combination, the difference between the predetermined parameter values and estimated parameters is stored. A distribution of these differences is obtained by noting differences for each of the 30 replications of that particular combination. We assume that the mean of these differences represents an unbiased estimate of estimator bias for that particular combination. We also assume that the distribution is approximately normal and use the variance of the bias distribution as an estimate of the variance of the estimator. This enables testing of the following hypotheses:

H_0: $m_i^* = m_i$ vs. H_a: $m_i^* \neq m_i$, using a simple z-score test. This is the only small-sample property which can be objectively analyzed.

Desirable large-sample properties are asymptotic unbiasedness, consistency, asymptotic efficiency and sufficiency. First, sufficiency is assured since all available observations are used in estimating the parameters. Asymptotic efficiency depends first on having consistent estimators. We may be able to show that the estimators are consistent, but we can not establish the second condition required for asymptotic efficiency: that no other consistent estimator has smaller asymptotic variance. Again we have, as a basis of comparison, neither other estimators nor theoretical minimums provided by Cramer-Rao lower bounds.

We are left with empirically evaluating asymptotic bias and consistency; each is approached in a similar manner. Asymptotic unbiasedness implies that

$\lim_{n\to\infty}$ {Bias $(\hat{\theta})$} = 0 where $\hat{\theta}$ is some estimator of θ, the true parameter. Similarly, consistency implies that

$\lim_{n\to\infty}$ {Var$(\hat{\theta})$ + (Bias $\hat{\theta})^2$} = 0. We are able to evaluate these properites through the following model:

Bias $m_i^* = An^B$, where we hypothesize that A>0, and B<0. Taking natural logarithms, the following regression model is obtained:

ln(Bias m_i^*) = lnA+B·ln(n), where sample size n is the independent variable and bias measurements provide the dependent variable. A similar model is developed for evaluating consistency, except that the dependent variable is the sum of estimator variance and the square of the bias. The hypotheses are

H_o: m_i^* is not asymptotically unbiased,

H_a: m_i^* is asymptotically unbiased.

Substituting the word "consistent" for "unbiased" in the above gives the hypotheses for consistency. Our acceptance or rejection is based on other test results, namely the set of tests performed on the regression models to determine the effect of sample size on these attributes. Significant regression results will lead us to reject the nulls above. Grouped data are used in the analyses for both properties. This is accounted for in determining critical values for the regression model hypotheses. In the interest of manageability, grouping of variance data has to be simplified. Rather than attempt to compute 280,840 covariance terms for each of the 9 sample sizes, we assume that the positive covaraince terms will tend to be offset by negative covariance terms. This allows us to use mean variance values for each sample size where this variance is that of the bias distribution for a particular sample size. In turn, this greatly simplifies the data to be used in the consistency regression model, but we can not be certain of the cost of this assumption in terms of accuracy.

In summary, we are limited to empirically examining the following estimator properties: bias, asymptotic bias, and consistency. The latter two must be examined indirectly through the use of regression models, the second of which may not be conclusive because of the method chosen to group variance data from the bias distributions.

RESULTS

The two bivariate beta generators have no significant discernible differences in the quality of the variates which they produce. There was no significant difference between them in the proportion of tests failed, where the tests included runs tests for randomness, Kolmogorov tests for fit, and chi-square tests for fit for n >50. We conclude that the principal difference is the computer processing time required for each. Table 1 shows the times required for several parameter-set, sample size combinations.

TABLE 1

Comparison of Generator Speed

Predetermined Combinations			Processing Time in Seconds	
Parameters m_1,m_2,m_3	Sample Size n	Replications	Inver.	Transfrm.
3 ,3 ,3	10	2	23.1	2.6
5 ,5 ,5	20	2	21.7	1.5
5 ,5 ,5	50	2	59.3	8.7
8.5,9.5,3	10	2	15.7	1.4
8.5,9.5,3	50	1	49.4	14.3

Note that the times include the time required to perform the tests mentioned above, so generating time would be somewhat less. This also explains some of the differences which might not be expected. For example, we might expect that the tranformation generator would take more time for the second combination than the first, since more uniform numbers are required as are twice as many transformations. Instead, we find a smaller time of 1.5 seconds for the second combination as opposed to 2.6 seconds for the first. This is due to the fact that the increased speed of the Kolmogorov test, due to more rapid convergence in using Simpson's rule to evaluate incomplete beta functions, more than offsets the increased generating time. The combinations shown in Table 1 are typical. They reveal that the transformation generator is on the order of ten times faster than the inversion generator; the advantage would be even more pronounced if tests of fit and randomness were not performed on each sample.

Table 2 shows several typical parameter combinations and the test results for each combination; the transformation generator is used in each case. Test results are shown as frequencies of test failure. This gives a potential user an indication of the reliability of the generator. Note that rejection of a sample on either the Kolmogorov or chi-square tests of fit caused a new sample to be drawn for the purpose of evaluating parameter estimates; we wanted to have 30 valid replications and wanted to avoid using questionable samples to estimate parameters.

TABLE 2

Frequency of Tests Failures
(30 valid replications)

Parameter Set m_1,m_2,m_3	Sample Size n	Test Kolm.	Failures Chi-Sq.	(α=0.05) Runs
1.5 ,1 ,5	30	1	-	1
1.5 ,1 ,5	100	3	2	Ø
5 ,5 ,5	50	4	1	2
5 ,5 ,5	200	2	3	1
8.5 ,10 ,10	20	2	-	1
8.5 ,10 ,10	150	4	2	2

Note that the failure frequencies for the Kolmogorov and runs tests reflect two tests and three tests on each sample respectively. Given this, and with $\alpha = 0.05$, we would expect to reject some valid samples. Also note that, as discussed earlier, large parameter values tended to give more frequently rejected samples.

Our evaluation of parameter estimators produced some interesting results. First, our evaluation of estimator bias proved to be inconclusive. For the 3240 (120 parameter combinations, 9 sample sizes, 3 parameter estimates) distributions of estimator bias which were generated, all but eight had means which were within one standard deviation from zero. All of the remaining eight distributions were within two standard deviations of zero. Using a parametric test, in other words, we are unable to reject the null hypothesis of no estimator bias in every combination examined. We had expected to find statistical evidence of some positive bias, particularly for small sample sizes (2, p. 6).

If we look at the distributions of bias nonparametrically, different results are obtained. If estimators are unbiased, we would expect an approximately equal frequency of positive and negative bias distribution means. However, over 95 percent of these means were positive. This would cause us to reject the null of no bias and might lead to a conclusion that there is some degree of positive bias in method-of-moments estimators. We realize that such a conclusion is weak on the grounds of scientific objectivity; most readers would agree that constructing a second hypothesis test like this, using the same data, is not desirable or even unacceptable. We discuss it here simply to show that, although bias appears to be statistically insignificant, there may nevertheless be a small amount of positive bias in method-of-moments parameter estimators. This is an area where more work is needed to make such a conclusion credible.

Conclusions regarding asymptotic bias and consistency are less tenuous. Stated very simply, we found that sample size explains almost all of the variation in both bias and consistency terms noted in the previous section. That is, we are led to conclude that, for method-of-moments estimators,

$$\lim_{n\to\infty} (\text{Bias } m_i^*) = 0 \qquad \text{and}$$

$$\lim_{n\to\infty} (\text{Var}(m_i^*) + (\text{Bias } m_i^*)^2) = 0$$

Results for the regression of bias on sample size are shown in Table 3.

TABLE 3

Regression Results: $\ln(\text{Bias } m_i^*) = \ln A + B \cdot \ln(n)$

Mean Bias (Dep. Var.)

	m_1	m_2	m_3
lnA	8.277	8.047	8.249
Standard error	0.135	0.129	0.179
t-Statistic	61.43	62.19	46.09
B	-1.199	-1.139	-1.150
Standard error	0.0346	0.0332	0.0459
t-Statistic	-34.64	-34.27	-25.02
Multiple R (adj.)	0.9967	0.9966	0.9937
R^2 (adj.)	0.9934	0.9932	0.9874
F-Statistic (ANOVA)	1200.1	1174.8	626.2

We see from this that for each parameter, our hypothesis of B<0 is borne out; all three t-statistic values show significance well beyond the 99 percent level for 7 degrees of freedom. Similarly, the F-statistic values are all significant beyond the 99 percent level for 1, 7 degrees of freedom.

Another observation is the significance of the intercept value, A. The t-statistic values for these would lend support to our earlier speculation that samll-sample parameter estimates are likely to be positively biased.

The results for evaluation of consistency are given in Table 4. The significance of all the regression results leads us to conclude that method-of-moments parameter estimators are consistent. We must temper this conclusion with the fact that we have used somewhat simplified methods of calculating the terms in the consistency limit equation. It would be better perhaps to state that the estimators are probably consistent, but we cannot say so with total certainty.

TABLE 4

Regression Results
$\ln(\text{Var}(m_i^*) + (\text{Bias } m_i^*)^2) = \ln A + B \cdot \ln(n)$

Mean Consistency Terms for

	m_1	m_2	m_3
ln A	6.423	6.585	6.629
Standard error	0.3364	0.3759	0.3182
t-Statistic	19.10	17.52	20.83
B	-1.415	-1.447	-1.141
Standard error	0.0864	0.0965	0.0817
t-Statistic	-16.38	-14.99	-17.27
Multiple R (adj.)	0.9854	0.9826	0.9868
R^2 (adj.)	0.9789	0.9655	0.9738
F-Statistic (ANOVA)	268.4	224.8	298.3

CONCLUSIONS

First, we have found that both proposed generators produce samples which pass certain tests of validity; however, the transformation generator is clearly preferable for producing beta bivariates because of its speed. We would recommend that users of this generator test their samples for validity since using, for example, some other uniform generator could adversely affect sample quality in terms of fit and randomness. In addition, there is a general need for developing new ways of detecting departures from randomness when generating asymmetric multivariate samples, whatever the distribution of the population involved.

A total of 32,400 valid bivariate beta samples were produced using the transformation generator which was in turn based on the GFSR uniform generator. These were used to empirically evaluate certain properties of method-of-moments parameter estimators for the bivariate beta distribution. Our results show that estimator bias is not statistically significant, but other consideration suggest that there may be a small amount of bias in the estimators for small samples. More work on this property may be in order. Further, our results show that the estimators are asymptotically unbiased and seem to be consistent. No other important estimator properties can be evaluated conclusively because we can not be certain whether or not method-of-moments estimators are minimum variance estimators. Finally, since our findings are consistent with findings involving the univariate beta distribution (2), we infer that similar conclusions probably hold for higher-dimensioned beta distributions.

BIBLIOGRAPHY

1. Conover, W. J., Practical Nonparametric Statistics. John Wiley & Sons, Inc., New York, 1971.
2. Fielitz, Bruce D. and Meyers, Buddy L., "Estimation of Parameters in the Beta Distribution," Decision Sciences, Vol. 6, January, 1975, pp. 1-13.
3. Fishman, George S., Concepts and Methods in Discrete Event Digital Simulation. John Wiley & Sons, Inc., New York, 1973.
4. Kennedy, William J., "Recent Advances in Techniques for Generating Pseudo-Random Numbers," paper presented at Joint National Meeting of ORSA-TIMS, Las Vegas, Nevada, 1975.
5. Kmenta, Jan, Elements of Econometrics. The Macmillan Company, New York, 1970.
6. Lee, T.C., Judge, G. G., and Zellner, A., "Maximum Likelihood and Bayesian Estimation of Transition Probabilities," Journal of the American Statistical Association, Vol. 63, December, 1968, pp. 1162-1179.
7. Lewis, T. G., Distribution Sampling for Computer Simulation. D. C. Heath and Company, Lexington, Mass., 1975.
8. Martin, J. J., Bayesian Decision Problems and Markov Chains. John Wiley & Sons, Inc., New York, 1967.
9. Mauldon, J. G., "A Generalization of the Beta-Distribution," Annals of Mathematical Statistics, Vol. 30, 1959, pp. 509-520.
10. Mosimann, James. E., "On the Compound Multinomial Distribution, the Multivariate β-distribution, and Correlations among Proportions," Biometrika, Vol. 49, January, 1962, pp. 65-82.
11. Myers, Buddy L. and Enrick, Norbert L., Statistical Functions. Kent State University Press, Kent, Ohio, 1970.
12. Quandt, Richard E., "A Stochastic Model of Elections in Two-Party Systems," Journal of the American Statistical Association, Vol. 69, June, 1974, pp. 315-324.
13. Romesburg, H. Charles, "Estimation of Parameters in the Beta Distribution: Comment," Decision Sciences, Vol. 7, January, 1976, p. 162.
14. Shreider, Yu. A., The Monte Carlo Method, trans. by G. J. Tee. Pergamon Press, London, 1967.
15. Wilkes, M. V., A Short Introduction to Numerical Analysis. Cambridge University Press, Cambridge, England, 1966.