CONF-800644--1

ESTIMATORS FOR THE TRUNCATED BETA-BINOMIAL DISTRIBUTION

Corwin L. Atwood

ABSTRACT

Let X have a beta-binomial(m,p,θ) distribution, truncated such that $X > t$ for $t = 0$ or 1. Suppose that independent observations of X are available. A consistent estimator of (p,θ) is given, based on the first three sample moments. This may be used as a start for maximum likelihood estimation or jackknifing. The standard assumptions for a $C(\alpha)$ test that X is truncated binomial do not hold. However a test is proposed based on jackknifing the sample variance of X. Some Monte Carlo comparisons are given. For moderately small data sets, these comparisons show that (1) the moment estimator is often superior to the MLE, and (2) the $C(\alpha)$ test is superior to other proposed tests, in spite of its lack of theoretical justification.

1. SUMMARY

Suppose that X has a beta-binomial(m,p,θ) distribution, truncated such that $X > t$ for $t = 0$ or 1. Suppose also that independent observations of X are available. The maximum likelihood equations are easy to write down, but solving them for p and θ requires a numerical iteration procedure. A consistent estimator for (p,θ) is presented here, based on the first three sample moments of X. For small samples, this moment estimate may give impossible values, but if it gives possible values it can be used as it is, or as a basis for jackknifing, or as the starting point for the maximum likelihood iterations.

Diagnostic checks on the model assumptions are suggested. Simple standarized residuals are proposed for investigating whether the data deviate in some unspecified way from the assumptions. A departure from the assumptions which is of special interest is $\theta=0$, i.e., X is truncated binomial. This can be tested based on a jackknife confidence interval for θ. Or the $C(\alpha)$ test statistic can be used, even though the assumption for asymptotic optimality of the test does not hold. Or a simple test based on the sample variance of X can be used. For small or moderate samples, the significance levels are all based on jackknifing.

Monte Carlo simulations are used to compare the various estimators and the various tests of $\theta=0$. For the range of parameters and sample sizes considered, these comparisons show the following. Jackknifing the moment estimator is generally not helpful, and often very harmful. Neither the maximum likelihood estimator nor the moment estimator dominates the other. The jackknife confidence intervals for p and θ are not reliable for sample sizes as small as 30. The $C(\alpha)$ test is clearly superior to the other proposed tests of $\theta=0$.

2. BASIC CALCULATIONS

A random variable Y has a beta-binomial distribution with parameters m, p, and θ if

$$P(Y = k) = \binom{m}{k} \frac{\prod_{r=0}^{k-1} (p+r\theta) \prod_{r=0}^{m-k-1} (q+r\theta)}{\prod_{r=0}^{m-1} (1 + r\theta)} \qquad (1)$$

where m is a positive integer, k is an integer from 0 to m, and the other parameters satisfy $0 \leq p \leq 1$, $q = 1 - p$, $\theta \geq 0$. The distribution arises if π has a beta(α,β) distribution and if Y given π has a binomial(m,π) distribution. The parametrization (1) follows Griffiths (1973) and Tarone (1979), and is obtained by setting

$$p = E(\pi) = \alpha/(\alpha + \beta)$$

$$\theta = 1/(\alpha + \beta).$$

If $\theta = 0$, then Y has a binomial(m,p) distribution.

In this paper X will have a truncated beta-binomial distribution, with $X > t$, where $t = 0$ or 1. That is, for $k > t$

$$P(X = k) = P(Y = k) \;/\; P(Y > t) \qquad (2)$$

with the distribution of Y given by (1). One way in which this truncated distribution could arise is as follows (Georgin and Roy 1978.) A system of m components is occasionally hit by a shock. If a shock hits, then each component fails with some probability π, and the components behave independently of each other. On any one shock, π is the same for all the components, but π varies from shock to shock according to a beta distribution. The variability of π means that the shocks differ in severity. Under these assumptions, the number of failed components on a random shock has a beta-binomial distribution. If the shocks themselves are not observable, but if instead only instances of failed components can be observed, then the observed failures follow a truncated beta-binomial distribution with $t = 0$. Now suppose, in addition, that components can fail individually, without any shock. If these failures cannot be distinguished from instances when a shock causes just one component to fail, then the relevant data set for inference about the shocks has a truncated beta-binomial distribution with $t = 1$.

We will assume that $m \geq t + 3$. For if $m = t + 2$, then X can take only two values, so the distribution of X is determined by a single parameter.

Suppose now that $x_1, \ldots x_n$ are independent observations of a truncated beta-binomial random variable, with known m and t and unknown p and θ. Abbreviate Expression (1) by $R(k)$. Then $\log L$, the logarithm of the likelihood is

$$\sum_{i=1}^{n} \log P(X_i = x_i)$$

$$= K + \sum_{i=1}^{n} \log R(x_i) - n \log \left[1 - \sum_{k=0}^{t} R(k) \right]$$

$$= K + \sum_{i=1}^{n} \left[\sum_{r=0}^{x_i - 1} \log(p + r\theta) + \sum_{r=0}^{m - x_i - 1} \log(q + r\theta) - \sum_{r=0}^{m-1} \log(1 + r\theta) \right]$$

$$- n \log T$$

defining T. Here K does not depend on p or θ.

There is no low-dimensional sufficient statistic for p and θ. The proof of this assertion is similar to that of Example 2.16 in Cox and Hinkley (1974).

3. ESTIMATORS

3.1 Maximum Likelihood Estimators

Differentiation of log L is straightforward. We obtain

$$\frac{\partial}{\partial p} \log L$$

$$= \sum_{i=1}^{n} \left[\sum_{r=0}^{x_i-1} \frac{1}{p + r\theta} - \sum_{r=0}^{m-x_i-1} \frac{1}{q + r\theta} \right]$$

$$+ n \sum_{k=0}^{t} R(k) \left[\sum_{r=0}^{k-1} \frac{1}{p + r\theta} - \sum_{r=0}^{m-k-1} \frac{1}{q + r\theta} \right] / T$$

$$\frac{\partial}{\partial \theta} \log L$$

$$= \sum_{i=1}^{n} \left[\sum_{r=0}^{x_i-1} \frac{r}{p + r\theta} + \sum_{r=0}^{m-x_i-1} \frac{r}{q + r\theta} - \sum_{r=0}^{m-1} \frac{r}{1 + r\theta} \right]$$

$$+ n \sum_{k=0}^{t} R(k) \left[\sum_{r=0}^{k-1} \frac{r}{p + r\theta} + \sum_{r=0}^{m-k-1} \frac{r}{q + r\theta} - \sum_{r=0}^{m-1} \frac{r}{1 + r\theta} \right] / T$$

Formally, the maximum likelihood estimates p and θ are the solutions of

$$T \frac{\partial}{\partial p} \log L = 0$$

$$T \frac{\partial}{\partial \theta} \log L = 0.$$

We have not verified that these equations have a unique solution. Indeed, Kleinman (1973) seems unable to verify that the solution is unique in the simpler untruncated case. Therefore a numerical procedure for solving the maximum likelihood equations should also search at scattered values of p and θ to see if log L seems to be unimodal.

3.2 Moment Estimators

An an alternative to maximum likelihood estimation, one can base estimates on the sample moments of X. Let X and Y be related by Equations (1) and (2). We first obtain the population moments of X.

Define the notation

$$b^{(j)} = b(b-1)\ldots(b-j+1)$$

for any real b and positive integer j. Define $b^{(0)} = 1$. Then a little algebra shows that

$$P(Y > t)\ E\left[X^{(j)}\right]$$

$$= \sum_{k=t+1}^{m} k^{(j)}\ P(Y = k)$$

$$= m^{(j)} \frac{\prod_{r=0}^{j-1} (p + r\theta)}{\prod_{r=0}^{j-1} (1 + r\theta)} \sum_{h=t+1}^{m} \binom{m-j}{k-j} \prod_{r=0}^{k-j-1} (p + j\theta + r\theta) \prod_{r=0}^{m-k-1} (q + r\theta)/$$

$$\prod_{r=0}^{m-j-1} (1 + j\theta + r\theta)$$

$$= m^{(j)} \frac{\prod_{r=0}^{j-1} (p + r\theta)}{\prod_{r=0}^{j-1} (1 + r\theta)} \sum_{h=t-j+1}^{m'} \binom{m'}{h} \prod_{r=0}^{h-1} (p' + r\theta') \prod_{r=0}^{m'-h+1} (q' + r\theta')/$$

$$\prod_{r=0}^{m'-1} (1 + r\theta')$$

where $m' = m-j$, $p' = (p + j\theta)/(1 + j\theta)$, $q' = 1 - p'$, and $\theta' = \theta/(1 + j\theta)$. So finally

$$EX^{(j)} = m^{(j)} \frac{\prod_{r=0}^{j-1} (p + r\theta)}{\prod_{r=0}^{j-1} (1 + r\theta)} \frac{P(Y' > t-j)}{P(Y > t)}$$

where

$Y \sim$ beta-binomial(m,p,θ)

$Y' \sim$ beta-binomial$(m-j, [p + j\theta]/[1 + j\theta], \theta/[1 + j\theta])$.

In particular, if $t=0$ then

$$EX = \frac{mp}{P(Y > 0)} .$$

If t=1, then

$$EX = \frac{mp\left[1-\prod_{r=0}^{m-2}(q + r\theta) \;/\; \prod_{r=1}^{m-1}(1 + r\theta)\right]}{P(Y > 1)} .$$

If t=0 or 1, then

$$EX(X-1) = \frac{m(m-1)p(p + \theta)}{(1 + \theta)\, P(Y > t)}$$

and

$$EX(X-1)(X-2) = \frac{m(m-1)(m-2)\; p(p + \theta)(p + 2\theta)}{(1 + \theta)(1 + 2\theta)\; P(Y > t)}$$

The moment estimators presented below are based on the three moments EX, $EX^{(2)}$ and $EX^{(3)}$, and the corresponding sample moments. Denote the sample moments by

$$S_i = \frac{1}{n} \sum_{j=1}^{n} X^{(i)}$$

for i=1 to 3. Define

$$d_0(S_1, S_2, S_3) = (m-2)S_2^2 + (m-1)(m-2)S_1 S_2 - 2(m-1)S_1 S_3$$

$$d_1(S_1, S_2, S_3) = d_0(S_1, S_2, S_3) + 2mS_3 - 2m(m-2)S_2.$$

THEOREM. *Let* $X > t$ *have truncated beta-binomial*(m,p,θ) *distribution, with* $0 < p < 1$, *and* $\theta \geq 0$.

(a) *Suppose* $t=0$. *Then*

$$p^* = [2(m-2)S_2^{\,2} - S_2S_3 - (m-1)S_1S_3] / d_0(S_1, S_2, S_3)$$

and

$$\theta^* = [(m-1)S_1S_3 - (m-2)S_2^{\,2}] / d_0(S_1, S_2, S_3)$$

converge in probability to p *and* θ.

(b) *Suppose* $t=1$. *If* $d_1(EX, EX^{(2)}, EX^{(3)}) \neq 0$, *then*

$$p^* = [2(m-2)S_2^{\,2} - S_2S_3 - (m-3)S_1S_3 - 2(m-2)S_1S_2] / d_1(S_1, S_2, S_3)$$

and

$$\theta^* = [(m-1)S_1S_3 - (m-2)S_2^{\,2} + (m-2)S_1S_2 - mS_3] / d_1(S_1, S_2, S_3)$$

converge in probability of p *and* θ.

The proof is given in the appendix.

In part (b), the assumption $d_1(EX,EX^{(2)},EX^{(3)}) \neq 0$ seems to be true always. A computer calculaion has shown that it is true if p takes any of the values 0.0001, 0.9999, or an integer multiple of 0.01 up to 0.99, if $\theta/(1+\theta)$ takes any of these values or the value 0, and if m is any integer from 4 through 100.

For small samples, p^* and θ^* may be impossible, i.e., the denominator in the expressions may be zero, or if it is nonzero the estimates may not satisfy $0 \leq p \leq 1$, $\theta \geq 0$. In any of these cases another estimation procedure must be used.

Johnson and Kotz (1969) mention the asymptotic efficiencies of moment estimators for various other truncated distributions. They are "remarkably high" (over 90 percent) for the truncated binomial distribution, (p. 75), over 70 percent for the truncated Poisson (p. 106), and "very inefficient" in one case and over 55 percent in another case for the truncated negative binomial (p. 137). The efficiency of moment estimation seems to depend greatly on the underlying distribution and on the details of how the moments are used. The moment estimator presented here performs quite well in the Monte Carlo examples of Section 5.

In the Monte Carlo studies described below, moment estimates are used, when they exist, as the starting points in an iteration sequence to obtain the maximum likelihood estimates.

3.3 Jackknife Estimators

A jackknife estimator may be constructed, based on either the maximum likelihood estimator or the moment estimator. Far less computation is required to use the moment estimator. In the Monte Carlo trials described below, jackknife estimates and confidence intervals are based on the moment estimates.

4. DIAGNOSTIC CHECKS ON THE MODEL

4.1 Alternate Model: Arbitrary

Let $z_k = P(X = k)$ for $t < k \leq m$, and let $\hat{z}_k$ be the MLE, obtained by substituting the maximum likelihood estimators for p and θ into the expression for $P(X=k)$. Let N_k be the number of observations with $X=k$. Conditional on the total number of observations n, N_k is binomial(n,z_k). So the kth "standardized residual",

$$U_k = \frac{N_k - n\,\hat{z}_k}{[n\,\hat{z}_k\,(1 - \hat{z}_k)]^{1/2}} ,$$

has mean and variance approximately 0 and 1. Large or small values of the U_k's, or strong patterns, indicate that X is not truncated beta-binomial.

If the sample size is large, a likelihood ratio test or X^2 goodness of fit test could be performed.

4.2 Alternate Model: $\theta = 0$

Three tests of

H_0: $\theta = 0$

will be given, all based on jackknife confidence intervals for various quantities.

<u>Test based on estimate of θ</u>

A jackknife confidence interval for θ can be constructed, based on θ^* or θ. If the lower end point is positive, H_0 would be rejected.

C(α) test

In the untruncated case, Tarone (1979) gives the C(α) test for H_0: $\theta = 0$. Tarone's test is asymptotically optimal (Neyman 1959). The test is based on the statistic

$$C = \frac{\partial}{\partial\theta} \log L(p, \theta)\big|_{\theta=0} .$$

A necessary and sufficient condition for asymptotic optimality (Neyman 1959) is that

$$E_{p,\theta=0}\left[\frac{\partial}{\partial p}\frac{\partial}{\partial\theta} \log L(p,\theta)\big|_{\theta=0} \right] = 0 . \qquad (3)$$

In the truncated case, Equation (3) does not hold. (In particular, in the simplest case, when t=0 and m=2, the left side of (3) equals $-2np/(1-q^2)$.) Therefore, this theoretical reason for using the C(α) test is not present.

However, C can be used anyway. Direct algebraic computation shows that

$$E_{p,\theta=0}\left[\frac{\partial}{\partial\theta} \log L(p,\theta)\Big|_{\theta=0} \right] = 0$$

when t=0 or 1. Therefore, C should be compared to 0. With a small or moderate data set, a jackknife confidence interval can be constructed for EC. If the lower end of the interval is positive then H_0 would be rejected.

Test based on sample variance of X

It is well-known that the (untruncated) beta-binomial(m,p,θ) distribution has a larger variance than does the binomial(m,p) distribution. If the variances of the two truncated distributions also have this relation, then the sample variance can be used to distinguish between the distributions. In fact, the situation is as follows.

Let $VB(p)$ and $VBB(p,\theta)$ denote the variance of a truncated binomial(m,p) and truncated beta-binomial(m,p,θ) distribution, respectively, both truncated at the same $t \geq 0$. From (1), it is clear that as $\theta \to \infty$, $P(Y=0) \to 1-p$ and $P(Y=m) \to p$. Therefore, $EX \to m$ and $EX^2 \to m^2$, so $VBB(p,\theta) \to 0$. Therefore, $VBB(p,\theta) < VB(p)$ for some values of θ. This inequality is the reverse of what might naively have been expected.

However, the relevant comparison is not between $VB(p)$ and $VBB(p,\theta)$ with the same p. Rather, to investigate whether the distribution is binomial or beta-binomial, in one case we would estimate p with θ assumed to be 0, and in the other case we would estimate both p and θ. The two estimates of p would be different! Let $\hat{p}_0$ be the maximum likelihood estimator of p, assuming $\theta=0$. It is the solution of

$$S_1 = mp \left[1 - \sum_{k=0}^{t-1} \binom{m-1}{k} p^k q^{m-1-k} \right] \Big/ \left[1 - \sum_{k=0}^{t} \binom{m}{k} p^k q^{m-k} \right] \tag{4}$$

where S_1 is the sample mean of X, as before. Suppose now that X is really truncated beta-binomial(m,p,θ), and let p_0 be the limit of $\hat{p}_0$ as $n \to \infty$, i.e., the solution of (4) with S_1 replaced by EX. To decide if $\theta=0$, we compare the truncated binomial(m,p_0) and the truncated beta-binomial(m,p,θ) distributions. So the relevant variances to compare are $VB(p_0)$ and $VBB(p,\theta)$.

Suppose that $0 < p < 1$, $\theta > 0$. If $m \geq t + 3$, $t=0$ or 1, and $m \leq 100$, then computer calculations indicate that $VBB(p,\theta) > VB(p_0)$. The calculations were performed on a CDC 176, letting p and $\theta/(1 + \theta)$ each take values 0.0001, 0.9999 and all integer multiples of 0.01 up to 0.99. A few typical plots of VB and VBB are shown in Figures A-C.

This justifies use of the sample variance to test whether $\theta=0$. Let $S^2 = S_2 - S_1(S_1 - 1)$ be the sample variance of X. Let $VB(\hat{p}_0)$ be the estimated variance of X obtained by substituting the MLE $\hat{p}_0$ into the expression for the variance of a truncated binomial random variable. For a small or moderate sized sample, contruct a (one-sided) jackknife confidence interval based on

$$\frac{S^2}{VB(\hat{p}_0)} - 1.$$

If the lower end point is positive, reject the hypothesis $\theta=0$.

5. MONTE CARLO RESULTS

5.1 Description of the Procedure

Monte Carlo experiments were performed as follows. For the values of t, m, p, and θ of interest, psendorandom samples of the desired size were generated. Ninety-six such experiments were performed corresponding to the following values:

$t=0$ or 1
$m=5$ or 20
$p=0.1, 0.5$, or 0.9
$\theta=0$, 1/3, 1, or 3
sample size = 10 or 30.

The number of samples generated in an experiment varied from 1300 to 5000. This is because the samples and corresponding estimates were stored, so that the calculations for a particular sample would not have to be repeated whenever the sample reoccurred. Sampling was continued until 5000 samples or 1300 distinct samples were obtained. A few replications indicate that this many samples gives numerical results which are accurate to one or two significant figures. This is adequate for the comparisons made. Each sample was generated as a multinomial sample, using the IMSL (1979) subroutine GGMLT.

For each sample, the moment, jackknifed moment, and maximum likelihood estimates of p and θ were calculated. Jackknife confidence intervals were also calculated for p and θ, based on the pseudovalues of the moment estimators, as described more fully below. The three tests of $\theta=0$ mentioned in Section 4 were performed. Another set of quantities was also estimated, based on the three estimates of (p,θ), and confidence intervals were calculated based on the pseudovalues of the moment estimates. These quantities were

$$A_k = \prod_{r=0}^{k-1} (p + r\theta) \Big/ \left[\prod_{r=0}^{k-1} (1 + r\theta)\, P(Y > t) \right]$$

for $t < k \leq 4$. The quantities A_k are of natural interest in the shock example of Section 1. In that example, if the shocks occur with rate μ, then observed failure instances occur with rate $\lambda = \mu\, P(Y > t)$, and a particular set of k components fails with rate λA_k.

Jackknife confidence intervals were not calculated by assuming a normal distribution for the pseudovalues, since the distributions were often highly skewed. Rather, from a sample of size n, the ith and $(n + 1 - i)$th ordered pseudovalues were used as a confidence interval with nominal level $1 - (2i - 1)/n$. This nominal level is based on a continuity correction. In this way, 90 percent intervals were found, using n=10 and i=1, or n=30 and i=2. One-sided tests of $H_0{:}\theta=0$ were based on such intervals, with a nominal significance level of 0.05.

For some samples, the estimates of Section 3 were impossible or undefined. In such cases, the following estimates were used. (a) If every element of the sample had value t+1, then the moment estimate and MLE of p and θ were set to 0. (b) If every element of the sample had value m, then the moment estimate and MLE of θ were set to 0, and the moment estimate and MLE of p were set to 1. (c) If the moment estimates were impossible (i.e., violating $0 \leq p \leq 1$, $\theta \geq 0$), or if they were undefined (because $d_0(S_1,S_2,S_3)$, resp. $d_1(S_1,S_2,S_3)$, was zero), then the following procedure was followed. In Equation (A1), resp (A5), θ was set to 0, the population moments were replaced by the sample moments, and the results was solved for p^*. Then p was set to p^* in (A2), the population moments were replaced by the sample moments, and the result was solved for θ^*. If this θ^* was negative, it was set to 0. (d) If the jackknifed moment estimates were impossible, i.e., violating $0 \leq p \leq 1$ or $\theta \geq 0$, then the moment estimates were used.

The maximum likelihood estimates were found using the IMSL (1979) conjugate gradient subroutine ZXCGR, starting at the moment estimates.

5.2 Conclusions

Of course, all the conclusions given below are valid only for the range of m and sample size considered.

Estimators of p and θ

The jackknifed moment estimator should not be used. In terms of mean squared error, it is never much better than the simple moment estimator, it is usually worse, and sometimes much worse.

Neither the moment estimator nor the maximum likelihood estimator dominates the other, in terms of mean squared error. If $\theta=0$, then the moment estimator seems slightly preferable. If $\theta=1$ or 3, then the MLE is better when $p=.1$, and the moment estimator is much better when $p=.5$ or .9. If $\theta=1/3$ and if $m=5$ or the sample size is 10, then the pattern is usually like that for larger θ. If $\theta=1/3$ and $m=20$ and the sample size is 30, then the MLE is preferable. Tables 1 through 3 give more details.

The confidence intervals, based on the pseudovalves from the jackknifed moment estimators, are not reliable. The frequency that the interval covers the true value is often much less than the nominal level of 90 percent, although it does improve when the sample size increases from 10 to 30. Table 4 gives examples.

Estimators of A_k

Jackknifing the moment estimator of A_k is slightly worse than using the moment estimator itself, in terms of mean squared error. The moment estimator and maximum likelihood estimator seem about equally good, with the MLE slightly better in most cases.

The jeckknife intervals for A_k are generally conservative, except in the two cases when $(p,\theta) = (.1,0)$ or $(.9,3)$. In these two cases the observed coverage frequencies were low.

Tests of H_0: $\theta=0$

The $C(\alpha)$ test is clearly superior to the other two. All three tests have size which is less than the nominal value of .05. However, the $C(\alpha)$ test has far better power. If the sample size is 10, no test has very good power. Table 5 shows one case with sample size 30.

APPENDIX

The moment estimator theorem is proved here.

PROOF. (a) If t=0 then

$$EX^{(2)}/EX = (m-1)(p + \theta) \;/\; (1 + \theta) \tag{A1}$$

$$EX^{(3)}/EX^{(2)} = (m - 2)(p + 2\theta)/(1 + 2\theta) \tag{A2}$$

Solving for p and θ yields equations which are identical to the equations in part (a) of the theorem, except the population moments $EX^{(i)}$ appear in place of the corresponding sample moments S_i. It is tedious but direct to show that

$$d_0(EX,\ EX^{(2)},\ EX^{(3)}) = \frac{m^2 (m - 1)^2 \ (m - 2)\ p^2 \ (p + \theta)(1 - p)}{(1 + \theta)^2 \ (1 + 2\theta) \quad P(Y > t)^{\ 2}}$$

which is nonzero if $0 < p < 1$. Therefore p and θ are continuous functions of the moments. Therefore, by well-known results in probability theory, p^* and θ^* converge to the corresponding functions of the population moments, i.e., to p and θ.

(b) If t=1, then the analogue of Equation (A1) becomes very cumbersome. An easier method is to imitate Rider (1955) and introduce a third variable, and to solve three equations, as follows. Let $X > 1$ be the truncated version of Y. Then

$$EY = mp \tag{A3}$$

$$EY^{(2)} = m\ (m - 1)\ p\ (p + \theta)\ /\ (1 + \theta) \tag{A4}$$

To obtain equations in terms of the moments of X, observe that

$$EY = \sum_{0}^{1} k\ P(Y = k) + P(Y > 1) \sum_{2}^{\infty} k\ P(x = k)$$

$$= P(Y = 1) + P(Y > 1)\ EX$$

and

$$EY^{(2)} = P(Y > 1)\ EX^{(2)}$$

Substitute these expressions into Equations (A3) and (A4). From Equation (1) calculate that

$$P(Y = 1) = \frac{mp}{q + mp + (m - 1)\theta} [1 - P(Y > 1)].$$

Substitute this, solve the two resulting equations to eliminate $P(Y > 1)$, and obtain finally

$$EX - mp = EX^{(2)} (1 + \theta) - (m - 1)(p + \theta)\ EX. \qquad (A5)$$

This is one equation which p and θ satisfy. The other equation is Equation (A2). Solving Equations (A2) and (A5) yields equations which are identical to those in part (b) of the theorem, except that they are in terms of $EX^{(i)}$ instead of S_i. Here the assumption has been used that $d_1(EX, EX^{(2)}, EX^{(3)})$ is nonzero. So, as in part (a), p^* and θ^* converge to p and θ.

REFERENCES

Cox, D. R., and Hinkley, D. V. (1974), Theoretical Statistics, London: Chapman and Hall Ltd.

Georgin, Jean Pierre, and Roy, Christophe (1978), "Une modelisation des effects des causes communes de défaillnce: le modèle semi catastrophique," DSN Report No. 195, Paris: OECD Nuclear Energy Agency. Translated as SINDOC (78) 85, NRC Translation 484, U.S. Nuclear Regulatory Agency.

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.

IMSL Library Reference Manual, Edition 7 (1979), Houston: International Mathematical and Statistical Libraries, Inc.

Johnson, Norman L., and Kotz, Samuel (1969), Discrete Distributions, New York: John Wiley and Sons.

Kleinman, Joel C. (1973) "Proportions with Extraneous Variance: Single and Independent Samples," Journal of the American Statistical Association, 68, 46-54.

Neyman, Jerzy (1959) "Optimal Asymptotic Tests of Composite Statistical Hypothesis," in Probability and Statistics, ed. U. Grenander, New York: John Wiley and Sons, 213-234.

Rider, Paul R. (1955), "Truncated Binomial and Negative Binomial Distributions," Journal of the American Statistical Association, 50, 877-883.

Tarone, Robert E. (1979), "Testing the Goodness of Fit of the Binomial Distribution," Biometrika, 66, 585-590.

TABLE 1. Mean Squared Errors of Estimates of p and θ, when t=1, m=5, sample size = 10[a]

True θ	True p: p = .1		True p: p = .5		True p: p = .9	
θ=0	.0062	.0001	.016	.019	.004	0.69
	.0075	.0038	.026	.729	.008	5.66
	.0063	.0015	.042	.072	.013	2.62
	MOM		MOM		MOM	
θ=1/3	.096	.099	.031	0.13	.021	3.5
	.142	.472	.053	4.66	.043	36.0
	.054	.094	.084	0.55	.151	28.9
	MLE		MOM		MOM	
θ=1	.25	0.81	.073	0.9	.018	6.7
	.30	4.22	.090	16.0	.054	57.4
	.11	0.62	.129	3.3	.212	63.7
	MLE		MOM		MOM	
θ=5	.45	7.6	.15	8.4	.018	11.5
	.47	23.2	.15	34.6	.056	60.4
	.21	7.9	.18	31.1	.151	67.0
	MLE		MOM		MOM	

a Each column of three numbers contains the mean squared errors of the moment estimator, jackknifed moment estimator, and maximum likelihood estimator, respectively. In each cluster of six numbers, the left column corresponds to estimators of p, the right column to estimators of θ. The best estimator of (p,θ) is indicated in each case by capital letters MOM, JK, or MLE. If the estimators are all about equally good, ANY is written.

TABLE 2. Mean Squared Errors of Estimates of p and θ, when t=1, m=20, sample size = 30[a]

True θ	True p: p = .1		p = .5		p = .9	
θ=0	.00060	.00021	.00044	.00012	.00015	.00010
	.00072	.00487	.00044	.00013	.00015	.00011
	.00094	.00043	.00039	.00008	.00015	.00010
	MOM		MLE		ANY	
θ=1/3	.028	0.046	.0063	.028	.0011	.041
	.050	1.267	.0051	.103	.0010	.022
	.008	0.024	.0040	.019	.0009	.003
	MLE		MLE		MLE	
θ=1	.12	0.49	.029	0.44	.013	4.1
	.14	14.33	.032	54.06	.018	264.8
	.02	0.16	.037	0.61	.038	23.5
	MLE		MOM		MOM	
θ=3	.29	5.2	.08	4.8	.012	15.0
	.36	141.3	.11	314.8	.026	455.1
	.07	1.7	.11	9.1	.198	731.6
	MLE		MOM		MOM	

a See footnote for Table 1.

TABLE 3. Best Estimators of (p, θ)[a]

m	sample size = 10					
	t = 0			t = 1		
m=5	MOM	MOM	MOM	MOM	MOM	MOM
	(MLE)	MOM	MOM	MLE	MOM	MOM
	MLE	MOM	MOM	MLE	MOM	MOM
	MLE	MOM	MOM	MLE	MOM	MOM
m=20	MOM	ANY	ANY	MOM	ANY	ANY
	MLE	MLE	(MOM)	MLE	MOM	MOM
	MLE	MOM	MOM	MLE	MOM	MOM
	MLE	MOM	MOM	MLE	MOM	MOM
	sample size = 30					
	t = 0			t = 1		
m=5	MOM	MLE	MLE	MOM	MOM	JK
	MLE	MOM	MOM	MLE	MOM	MOM
	MLE	MOM	MOM	MLE	MOM	MOM
	MLE	MOM	MOM	MLE	MOM	MOM
m=20	ANY	ANY	ANY	MOM	MLE	ANY
	MLE	MLE	(MLE)	MLE	MLE	MLE
	MLE	MOM or MLE	MOM	MLE	MOM	MOM
	MLE	MOM	MOM	MLE	MOM	MOM

a Each cluster of twelve abbreviations is a condensation of a table like Table 1 or Table 2. The numbers are omitted, and the best estimators are given, in the same arrangement as in Tables 1 and 2. Parentheses indicate that the best estimators for p and θ differed, but that the indicated choice seemed somewhat better.

TABLE 4. Observed Frequency that Confidence Intervals for p and θ Cover True Values[a]

True θ	True p		
	p = .1	p = .5	p = .9
	a. t=1, m=5, sample size = 10		
θ=0	.67	.90	.98
	.99	.94	.98
θ=1/3	.34	.80	.85
	.11	.38	.38
θ=1	.28	.54	.71
	.26	.31	.19
θ=3	.22	.32	.44
	.18	.19	.05
	b. t = 1, m = 20, sample size = 30		
θ=1	.99	1.00	1.00
	1.00	1.00	1.00
θ=1/3	.76	1.00	1.00
	.63	1.00	.95
θ=1	.61	.94	.95
	.60	.89	.89
θ=3	.53	.61	.68
	.51	.62	.24

a In each pair of numbers, the upper number corresponds to p, the lower number to θ. The nominal confidence coefficient was .90 in each case.

TABLE 5. Observed Frequency that H_0: $\theta=0$ is Rejected, for t=1, m=20, sample size = 30[a]

	Basis of test		
True p	Estimate of θ	$C(\alpha)$ test	Variance of X
p=.1	.00	.00	.00
	.03	.93	.09
	.03	1.00	.01
	.04	1.00	.05
p=.5	.00	.00	.00
	.00	1.00	.00
	.02	1.00	.01
	.04	1.00	.03
p=.9	.00	.00	.00
	.00	.77	.00
	.04	.54	.02
	.24	.49	.23

a Each column of four numbers gives the frequency that H_0 is rejected, at four values of θ: 0, 1/3, 1 and 3. A good test would have th first number $\leq$ .05 and the other three numbers large.

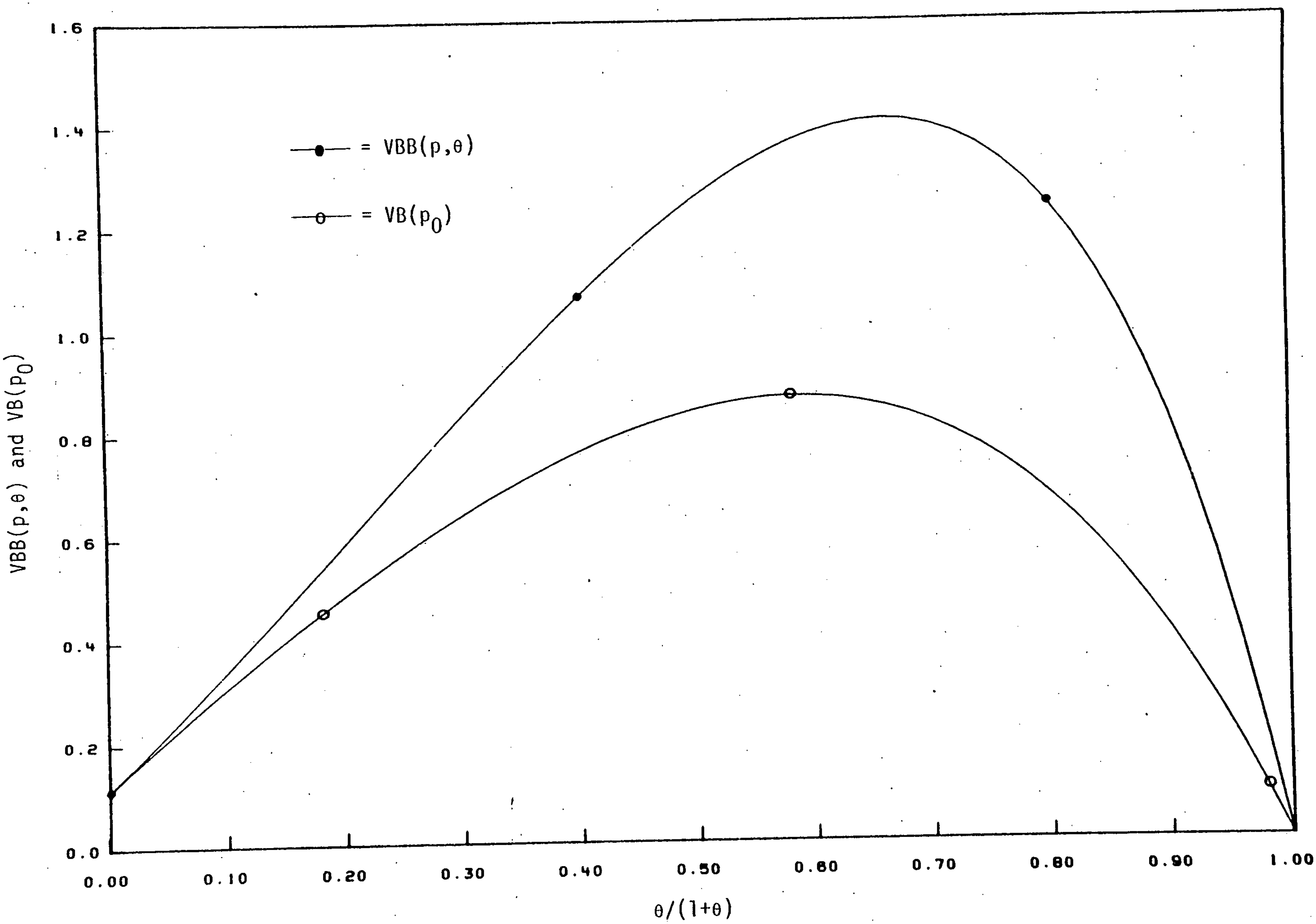

Figure A. Variance of beta-binomial(m,p,θ) and binomial(m,p_0) when $m=5$, $p=0.1$

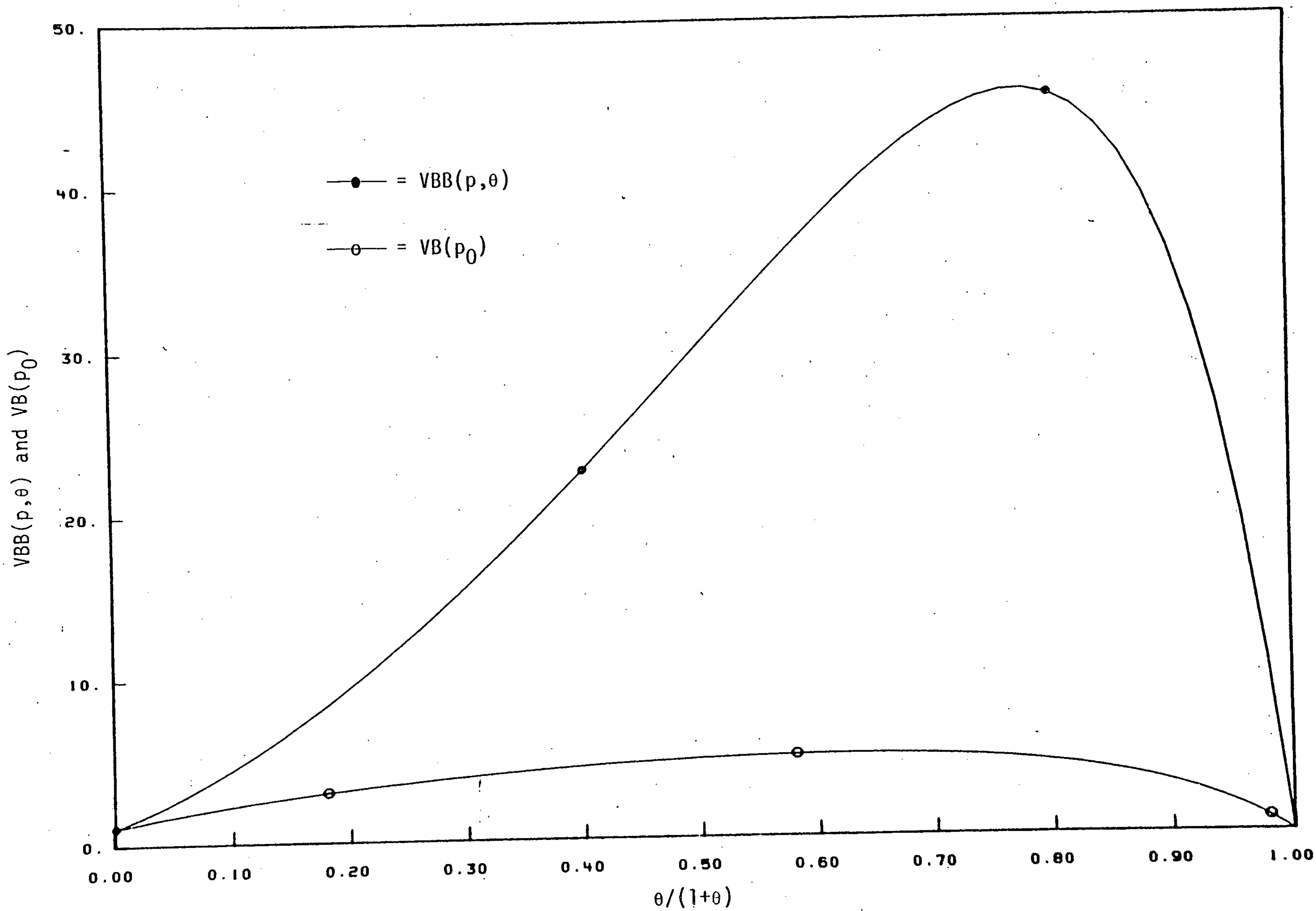

Figure B. Variance of beta-binomial(m,p,θ) and binomial(m,p_0) when m=20, p=0.1

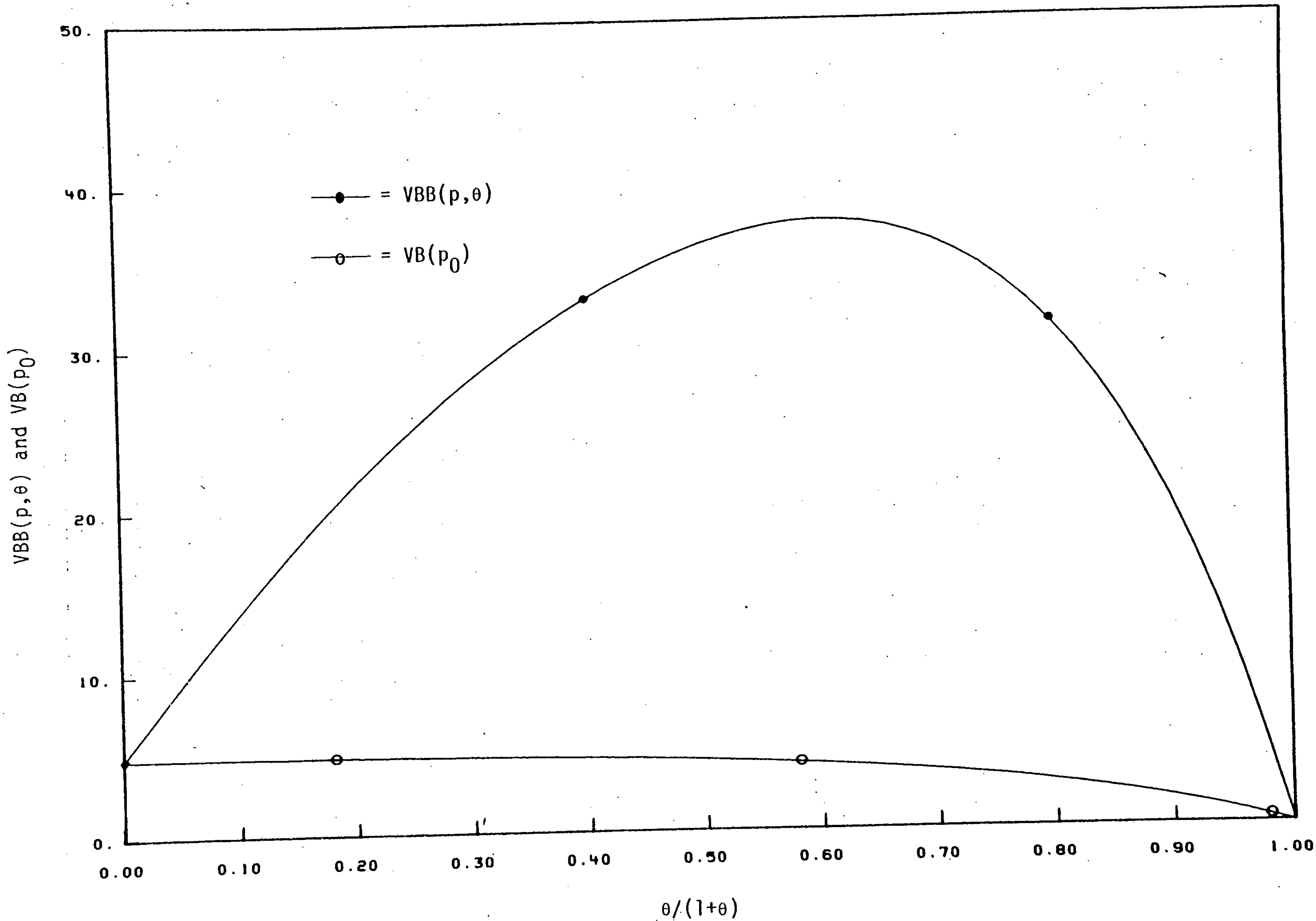

Figure C. Variance of beta-binomial(m,p,θ) and binomial(m,p_0) when $m=20$, $p=0.6$