SOME REMARKS ON THE USE OF MOMENT ESTIMATORS FOR THE PARAMETERS OF A MIXTURE OF TWO BINOMIAL DISTRIBUTIONS

BU-126-M Wallace R. Blischke December, 1960

Abstract

Moment estimators are constructed for a mixture of two binomial distributions having parameters $(n,p_{(1)})$ and $(n,p_{(2)})$ with $p_{(1)}<p_{(2)}$ and $n \geq 3$ and mixing parameter $\alpha(0<\alpha<1)$. The solution given is

$$\begin{pmatrix} \hat{p}_{(1)} \\ \hat{p}_{(2)} \\ \hat{\alpha} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}a - \frac{1}{2}(a^2-4af_1+4f_2)^{\frac{1}{2}} \\ \frac{1}{2}a + \frac{1}{2}(a^2-4af_1+4f_2)^{\frac{1}{2}} \\ (f_1-\hat{p}_{(2)})/(\hat{p}_{(1)}-\hat{p}_{(2)}) \end{pmatrix} \quad \text{if } a^2-4af_1+4f_2 \geq 0 \text{ and } (a^2-4af_1+4f_2)^{\frac{1}{2}} \leq \min(a,2-a)$$

$$= \frac{1}{n}\begin{pmatrix} f_1 \\ f_1 \\ 0 \end{pmatrix} \quad \text{otherwise}$$

where

$$f_j = \sum_{i=0}^{n-j} \binom{n-j}{i}\binom{n}{i+j}^{-1} \frac{U_{i+j}}{m}$$

$$a = \frac{f_3-f_1f_2}{f_2-f_1^2}$$

with

m = total number of observations

U_t = number of observations taking on the value t $(0 \leq t \leq n)$.

Because of their complexity, it has not been possible to determine expectations or variances of the above estimators.

It is noted that it is possible to construct many such moment estimators. A "better" moment estimator would be one that is a function of the minimal sufficient statistic

$$U^* = \left(\sum_{j=1}^{n} jU_j, \cdots, \sum_{j=1}^{n} j^n U_j\right).$$

SOME REMARKS ON THE USE OF MOMENT ESTIMATORS FOR THE PARAMETERS OF A MIXTURE OF TWO BINOMIAL DISTRIBUTIONS

Wallace R. Blischke*

BU-126-M

December, 1960

Let $Y_1, \ldots, Y_m$ be independent and identically distributed chance variables having distribution

(1) $$P(Y_i=y_i)=\binom{n}{y_i}\ [\alpha p_{(1)}^{y_i}(1-p_{(1)})^{n-y_i}+(1-\alpha)p_{(2)}^{y_i}(1-p_{(2)})^{n-y_i}]$$

where $0 < \alpha < 1$, $0 < p_{(1)} < p_{(2)} < 1$, and $n \geqq 3$ is integral. Equation (1) is a mixture of the two binomial distributions with mixing parameter α. The problem is to estimate $p_{(1)}$, $p_{(2)}$, and α. As is the case in most mixture problems (cf [2], and [3], p. 300) the only standard estimation procedure yielding equations that are at all tractable is the method of moments. Moment estimators for these parameters were given in [1]. The solution there was based on the following construction:

Define the chance variable

$$U_j = \text{"number of } Y_i\text{'s taking on the value } j\text{"}$$

for $0 \leqq j \leqq n$. Let

$$p_i = \binom{n}{j}[\alpha p_{(1)}^{j}(1-p_{(1)})^{n-j}+(1-\alpha)p_{(2)}^{j}(1-p_{(2)})^{n-j}]$$

Then the joint distribution of $U_o, \ldots, U_n$ is

$$p_{U_o,\ldots,U_n}(u_0,\ldots,u_n) = \frac{m!}{u_0!\ldots u_n!}\, p_0^{u_o}\cdots p_n^{u_n} \quad \text{if } 0 \leqq u_i \leqq m \text{ for } i=1,\ldots,n$$

$$\text{and } \sum_{i=1}^{n} u_i = m$$

$$= 0 \text{ otherwise}$$

* Biometrics Unit, Plant Breeding Department, Cornell University

As in [1], define

(2) $$F_{k,j} = \sum_{g=0}^{k} \binom{k}{g}\binom{n}{j+g}^{-1} p_{j+g} \quad (\text{for } 0 \leq k \leq n,\ 0 \leq j \leq n-k).$$

Then

$$F_{k,j} = \alpha p_{(1)}^{j}(1-p_{(1)})^{n-j-k} + (1-\alpha)p_{(2)}^{j}(1-p_{(2)})^{n-j-k}$$

The corresponding sample quantities, $f_{k,j}$, are defined by substituting in (2) the maximum likelihood estimates $\hat{p}_{j+g} = \frac{1}{m} U_{j+g}$.

Moment estimators for the three parameters α, $p_{(1)}$, and $p_{(2)}$, can now be constructed by solving a chosen set of at least three functions $F_{k,j}$ for these three parameters and substituting $f_{k,j}$'s for $F_{k,j}$'s in the solutions. In [1] the simultaneous equations used were $F_{n-1,1}$, $F_{n-2,2}$, $F_{n-4,4}$, and $F_{n-8,8}$. A more judicious choice of $F_{k,j}$'s will yield much more satisfactory estimators.

In particular, estimators can be constructed on the basis of only three of the $F_{k,j}$'s. For example, consider $F_{n-1,1}$, $F_{n-2,2}$, and $F_{n-3,3}$. For brevity, denote the functions $F_{n-k,k}$ simply by F_k. The construction is as follows: First, the equations

(3) $$F_1 = \alpha p_{(1)} + (1-\alpha)p_{(2)}$$

(4) $$F_2 = \alpha p_{(1)}^2 + (1-\alpha)p_{(2)}^2$$

(5) $$F_3 = \alpha p_{(1)}^3 + (1-\alpha)p_{(2)}^3$$

are solved for the three parameters as follows: Note that

$$F_2 - F_1^2 = \alpha(1-\alpha)(p_{(1)} - p_{(2)})^2$$

Now

$$F_1F_2 = \alpha^2 p_{(1)}^3 + \alpha(1-\alpha)p_{(1)}^2 p_{(2)} + \alpha(1-\alpha)p_{(1)}p_{(2)}^2 + (1-\alpha)^2 p_{(2)}^3,$$

so

$$\begin{aligned} F_3 - F_1F_2 &= \alpha(1-\alpha)(p_{(1)}^3 - p_{(1)}^2 p_{(2)} - p_{(1)}p_{(2)}^2 + p_{(2)}^3) \\ &= \alpha(1-\alpha)(p_{(1)} + p_{(2)})(p_{(1)} - p_{(2)})^2 \\ &= (p_{(1)} + p_{(2)})(F_2 - F_1^2) \end{aligned}$$

Thus

$$p_{(1)}+p_{(2)} = \frac{F_3 - F_1F_2}{F_2 - F_1^2} = A, \text{ say.}$$

Now solve (3) and (4) for α to obtain

$$\alpha = \frac{F_1 - p_{(2)}}{p_{(1)} - p_{(2)}} ,$$

which, when substituted in (4) yields

$$F_2 = \frac{F_1 - p_{(2)}}{p_{(1)} - p_{(2)}} (p_{(1)}^2 - p_{(2)}^2) + p_{(2)}^2$$
$$= (F_1 - p_{(2)})A + p_{(2)}^2$$

so that the result is the quadratic equation

(6) $$p_{(2)}^2 - Ap_{(2)} + F_1A - F_2 = 0$$

Solution of equations (3), (4), and (5) for $p_{(1)}$ instead of $p_{(2)}$ yields equation (6) with $p_{(1)}$ replacing $p_{(2)}$. The restriction that $p_{(1)} < p_{(2)}$ thus results in the unique solution

(7) $$p_{(1)}, p_{(2)} = \frac{1}{2}A \mp \frac{1}{2}(A^2 - 4AF_1 + 4F_2)^{1/2}.$$

Estimators can now be defined by substituting $f_{k,j}$'s for $F_{k,j}$'s. Write f_1, f_2, f_3, and a for the sample quantities (corresponding to $F_1, \ldots, A$, respectively) obtained by this substitution. This yields the moment estimators

(8) $$\begin{pmatrix} \hat{p}_{(1)} \\ \hat{p}_{(2)} \\ \hat{\alpha} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}a - \frac{1}{2}(a^2 - 4af_1 + 4f_2)^{1/2} \\ \frac{1}{2}a + \frac{1}{2}(a^2 - 4af_1 + 4f_2)^{1/2} \\ \dfrac{f_1 - \hat{p}_{(2)}}{\hat{p}_{(1)} - \hat{p}_{(2)}} \end{pmatrix} \quad \text{if } a^2 - 4af_1 + 4f_2 \geq 0 \text{ and } (a^2 - 4af_1 + 4f_2)^{1/2} \leq \min(a, 2-a)$$

$$= \frac{1}{n} \begin{pmatrix} f_1 \\ f_1 \\ 0 \end{pmatrix} \quad \text{otherwise}$$

(Note that the definition of the estimators when $a^2-4af_1+4f_2 < 0$ or $(a^2-4af_1+4f_2) > \min(a, 2-a)$ is arbitrary. In some cases one may wish to use instead, e.g., $(\hat{p}_{(1)}, \hat{p}_{(2)}, \hat{\alpha}) = (1, 0, 0)$.)

It has not been possible to compute expectations or variances of these estimators. There is empirical evidence, however, that the estimators of equation (8) are considerably "better" than those of reference [1].

Now some refinements can be made in these estimators. The only possibly justification (at this point, at least) for using moment estimators here is that

$$Ef_{k,j} = F_{k,j}, \text{ all } k, j.$$

It is not true, however, that

$$Ef_{k,j}f_{k',j'} = F_{k,j}F_{k',j'} .$$

Unbiased estimators of such products, though, can easily be computed. In particular, since

$$E\frac{U_j}{m} = p_j, \; j = 0,\dots,n,$$

it is easily seen that

$$\frac{m}{m-1}\left[f_1f_2 - \frac{1}{m^2n^2(n-1)}\sum_{j=2}^{n} j^2(j-1)U_j\right] = g_{1,2}, \text{ say,}$$

and

$$\frac{m}{m-1}\left[f_1^2 - \frac{1}{m^2n^2}\sum_{j=1}^{n} j^2U_j\right] = g_{1,1}, \text{ say}$$

are unbiased estimators of F_1F_2 and F_1^2, respectively. One can now look at the "refined" estimators obtained by replacing f_1f_2 by $g_{1,2}$, etc., in (8). Other similar "refinements" are possible.

A more rewarding pursuit would be to compute at least approximate variances for the estimators of equation (8) and determine the efficiency of these estimators relative to, e.g., the maximum likelihood estimators. Because of the complexity of these estimators it has not been possible to make such comparisons. (Two remarks are in order. Firstly, the method of moments is known generally to be inefficient. There is no reason to believe that this case will be an exception. Note, however, that for n=3, the moment estimators of equation (8) satisfy the maximum likelihood equations (cf. [1] so that the moment estimators are efficient for $n = 3$.) As previously noted, however, we are almost always faced with a situation, when dealing with mixtures, wherein the only standard estimation procedure yielding equations that are at all tractable is the method of moments. ([3], p. 300).

In such a situation it may be desirable to at least derive a moment estimator which is in some sense "optimal" among all possible moment estimators. (Note that many sets of three or more equations such as (3), (4) and (5) could be solved for $p_{(1)}$, $p_{(2)}$ and α. For example a derivation very similar to that given above will yield a solution to F_1, F_2 and F_4.)

In this connection it will be desirable to at least have an estimator which is a function of the sufficient statistic. This should be a minimal requirement for optimality.

Now a sufficient statistic for this problem is the vector $(U_o,\ldots,U_n)$. For we have

$$\begin{aligned} & p_{Y_1,\ldots,Y_m}(y_1,\ldots,y_m) \\ & = \prod_{i=1}^{m} \binom{n}{y_i} [\alpha p_{(1)}^{y_i}(1-p_{(1)})^{n-y_i} + (1-\alpha)\, p_{(2)}^{y_i}(1-p_{(2)})^{n-y_i}] \\ (9) \qquad & = \prod_{j=0}^{n} \binom{n}{j}^{u_j} [\alpha p_{(1)}^{j}(1-p_{(1)})^{n-j} + (1-\alpha)\, p_{(2)}^{j}(1-p_{(2)})^{n-j}]^{u_j} \\ & = [\prod_{j=0}^{n} \binom{n}{j}^{u_j}] \prod_{j=0}^{n} [\sum_{k=0}^{u_j} \binom{u_j}{k} \alpha^k p_{(1)}^{kj}(1-p_{(1)})^{kn-kj}(1-\alpha)^{u_j-k}\, p_{(2)}^{ju_j-jk} \\ & \qquad (1-p_{(2)})^{(n-j)(u_j-k)}] \end{aligned}$$

where

$u_j =$ number of y_i's equaling j.

Thus $p_{Y_1,\ldots,Y_m}$ depends on $Y_1,\ldots,Y_m$ only through $U_o,\ldots,U_n$, so by the Neyman Factorization Theorem $(U_o,\ldots,U_n) = U$, say, is sufficient. In fact, it is clear from equation (9) that U is a minimal sufficient statistic. No further reduction is possible.

Now the requirement that the estimators be functions of U is equivalent to the requirement that they be functions of $U_1,\ldots,U_n$ since $U_o=m-(U_1+\ldots+U_n)$. In other words, $(U_1,\ldots,U_n)$ is sufficient. This in turn implies that

$U^* = (\sum_{j=1}^{n} jU_j, \sum_{i=1}^{n} j^2U_j, \cdots, \sum_{j=1}^{n} j^nU_j)$ is sufficient; for U^* is a set of n independent linear equations in $U_1, \cdots, U_n$ (so that knowing U^* is equivalent to knowing $U_1, \cdots, U_n$). Note that

$$U^* = (\sum_{j=1}^{n} jU_j, \ldots, \sum_{j=1}^{n} j^nU_j)$$

$$= (\sum_{i=1}^{m} Y_i, \ldots, \sum_{i=1}^{m} Y_i^n)$$

so that, in terms of the original chance variables, the vector $(\Sigma Y_i, \ldots, \Sigma Y_i^n)$ is a sufficient statistic.

Thus a "good" estimator should be a function of the first n sample moments about 0. This can be achieve, for example, by solving the equations $F_1, \ldots, F_n$ and substituting $f_1, \ldots, f_n$ in the solution. No attempt has been made to derive an estimator which is a function of the entire set $f_1, \ldots, f_n$.

References

[1] Blischke, W. R., "Mixtures of two binomial distributions", Biometrics Unit Mimeo 121-M. Cornell University (July, 1960).

[2] Pearson, K. P., "Contributions to the mathematical theory of evaluation. 1. Dissection of frequency curves." Phil. Trans. Roy. Soc. A,185:71 (1894)

[3] Rao, C. R., Advanced Statistical Methods in Biometric Research. John Wiley & Sons, Inc., New York. (1952).