

MIXTURES OF TWO BINOMIAL DISTRIBUTIONS

BU-121-M

W. R. Blischke*

July, 1960

1. Introduction. In the theory of mixtures of distributions two types of problems are confronted. The first is the problem of "identifiability," that is, given that a population, f , is composed of a mixture of two or more distributions, say $f(x) = \sum_{i=1}^k \alpha_i f_i(x)$, where $\sum_{i=1}^k \alpha_i = 1$ and $0 < \alpha_i < 1$ for $i=1, \dots, k$, is the mixture unique? To show that a mixture of distributions is identifiable it must be shown that if also $f(x) = \sum_{i=1}^{k'} \alpha'_i f'_i(x)$, then $k=k'$, $\alpha_i = \alpha'_i$, and $f_i = f'_i$. Most of the work in the literature is concerned with this problem. (See [3], [4], and [5]).

The second problem is that of estimation, that is, given that a mixture is identifiable, how shall the parameters of the individual distributions comprising the mixture and the mixing parameters (the α 's) be estimated? (If the mixture is not identifiable, of course, one cannot hope to be able to estimate the individual parameters and thus "identify" the mixture.) The present paper will deal with mixtures of two binomial distributions.

2. The problem of identifiability of mixtures of binomials. Ordinarily a "mixture" of two binomial distributions is defined as a mixture of two binomials with parameters, say, $(1, p_{(1)})$ and $(1, p_{(2)})$. Thus, if α ($0 < \alpha < 1$) is the mixing parameter, in the mixed distribution the probability of a "success" is $\alpha p_{(1)} + (1-\alpha)p_{(2)}$. This makes the mixture again a binomial distribution, that is, if n independent observations are obtained then the chance variable Y ="number of 'successes' out of n " has the distribution

$$P(Y=y) = \binom{n}{y} [\alpha p_{(1)} + (1-\alpha)p_{(2)}]^y [\alpha(1-p_{(1)}) + (1-\alpha)(1-p_{(2)})]^{n-y} .$$

According to [5] this is a sufficient condition that the distribution be unidentifiable.

*

Biometrics Unit, Department of Plant Breeding, Cornell University, Ithaca, N.Y.

The unidentifiability becomes immediately apparent when one attempts to apply an estimation procedure such as the method of maximum likelihood. The maximum likelihood equations all result in the equation

$$\frac{Y}{n} = \alpha p(1) + (1-\alpha)p(2)$$

so only the parameter of the mixture can be estimated and not the parameters of the individual distributions comprising the mixture.

Now, can the problem be formulated so that the result is a mixture of binomials which is identifiable? Suppose, for example, that the two distributions in the mixture differ not only in the parameters $p(i)$, but also in the type of "failures" which can result, that is, $p(1)$ and $p(2)$ are the respective probabilities of a "success" but "failures" can be distinguished with regard to the population involved. Denote these, respectively, "failures of type 1" and "failures of type 2." Thus, if α is the mixing parameter, then for a single observation,

$$P(\text{"success"}) = \alpha p(1) + (1-\alpha)p(2) ,$$

$$P(\text{"failure of type 1"}) = \alpha(1-p(1)) ,$$

and

$$P(\text{"failure of type 2"}) = (1-\alpha)(1-p(2)) .$$

It is easily seen that the distribution of n independent such observations is trinomial:

$$P(n_1 \text{ "successes", } n_2 \text{ "failures of type 1", } n_3 \text{ "failures of type 2"}) \\ = \frac{n!}{n_1!n_2!n_3!} [\alpha p(1) + (1-\alpha)p(2)]^{n_1} [\alpha(1-p(1))]^{n_2} [(1-\alpha)(1-p(2))]^{n_3} ,$$

where $n_1+n_2+n_3=n$.

This is again unidentifiable (for example, both $\alpha = \frac{1}{2}$, $p(1) = \frac{1}{4}$, $p(2) = \frac{1}{2}$, and $\alpha = \frac{5}{8}$, $p(1) = \frac{2}{5}$, $p(2) = \frac{1}{3}$ give parameters $\frac{3}{8}$, $\frac{3}{8}$, and $\frac{1}{4}$, respectively, in the resulting trinomial), and the method of maximum likelihood yields

$$\frac{n_1}{n} = \widehat{\alpha p(1) + (1-\alpha)p(2)}$$

$$\frac{n_2}{n} = \widehat{\alpha(1-p(1))}$$

$$\frac{n_3}{n} = \widehat{(1-\alpha)(1-p(2))}$$

These (non-linear) equations are not sufficient to obtain a solution for the three parameters.

Finally, suppose that two binomial distributions with parameters $(n, p(1))$ and $(n, p(2))$, where $p(1) < p(2)$ and $n > 1$, are mixed, that is, one of the populations is selected, with selection probabilities α and $(1-\alpha)$, respectively, and the number of "successes" in n trials observed. In this case the chance variable $Y =$ "number of successes" has as its distribution

$$P(Y=y) = \alpha \binom{n}{y} p(1)^y (1-p(1))^{n-y} + (1-\alpha) \binom{n}{y} p(2)^y (1-p(2))^{n-y},$$

and maximum likelihood yields

$$\frac{Y}{n} = \widehat{\alpha p(1) + (1-\alpha)p(2)}$$

Suppose, however, that this experiment is repeated m times ($m > 1$). Denote by Y_i the chance variable "number of 'successes' at the i^{th} trial", $i=1, \dots, m$. Let $\alpha' = 1-\alpha$. Then

$$\begin{aligned} P_{Y_i}(y_i) &= P(Y_i = y_i) \\ (1) \quad &= \alpha \binom{n}{y_i} p(1)^{y_i} (1-p(1))^{n-y_i} + \alpha' \binom{n}{y_i} p(2)^{y_i} (1-p(2))^{n-y_i}, \end{aligned}$$

and the joint distribution of Y_1, \dots, Y_m is

$$(2) \quad p_{Y_1, \dots, Y_m}(y_1, \dots, y_m) = \prod_{i=1}^m \binom{n}{y_i} [\alpha p(1)^{y_i} (1-p(1))^{n-y_i} + \alpha' p(2)^{y_i} (1-p(2))^{n-y_i}]$$

= L, say.

In this case the mixture is identifiable but the problem of estimation is a difficult one.

The method of maximum likelihood yields the equations

$$\frac{\partial \log L}{\partial p(1)} = \sum_{i=1}^m \frac{\alpha [Y_i p(1)^{Y_i-1} (1-p(1))^{n-Y_i} - (n-Y_i) p(1)^{Y_i} (1-p(1))^{n-Y_i-1}]}{\alpha p(1)^{Y_i} (1-p(1))^{n-Y_i} + \alpha' p(2)^{Y_i} (1-p(2))^{n-Y_i}} = 0,$$

or

$$0 = \sum_{i=1}^m \frac{Y_i - np(1)}{\alpha p(1)^{Y_i} (1-p(1))^{n-Y_i} + \alpha' \left(\frac{p(2)}{p(1)}\right)^{Y_i-1} \left(\frac{1-p(2)}{1-p(1)}\right)^{n-Y_i-1} p(2)^{Y_i-1} (1-p(2))^{n-Y_i-1}}$$

$$\frac{\partial \log L}{\partial p(2)} = \sum_{i=1}^m \frac{Y_i - np(2)}{\alpha' p(2)^{Y_i} (1-p(2))^{n-Y_i} + \alpha \left(\frac{p(1)}{p(2)}\right)^{Y_i-1} \left(\frac{1-p(1)}{1-p(2)}\right)^{n-Y_i-1} p(1)^{Y_i-1} (1-p(1))^{n-Y_i-1}} = 0,$$

and

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^m \frac{Y_i p(1)^{Y_i-1} (1-p(1))^{n-Y_i} - Y_i p(2)^{Y_i-1} (1-p(2))^{n-Y_i}}{\alpha p(1)^{Y_i} (1-p(1))^{n-Y_i} + \alpha' p(2)^{Y_i} (1-p(2))^{n-Y_i}} = 0.$$

These equations are insoluble even for $m=2$.

It is possible, however, to construct estimators of the three parameters in this case. The remainder of this paper will be devoted to a discussion of two such estimators, one of which is constructed by considering the tails of the distribution given by formula (2). The second procedure to be considered makes use of all the information in a sample.

3. The method of the tails. The m independent chance variables Y_i each can take on the values $0, 1, \dots$, or n . Suppose the m observations are presented in the form of a sample frequency distribution. Roughly speaking, one might hope to be able somehow to "partition" this frequency distribution and to then construct estimators of the individual $p_{(i)}$'s as a result of such an operation. This approach leads one to consider only the extreme classes in the sample frequency distribution in constructing estimators for the $p_{(i)}$'s, the idea being that such classes will contain "mostly" individuals from only one of the two component populations.

More precisely, this argument leads to the following construction: Let U_x , for $x=0, \dots, n$, be the number of experiments which result in exactly x "successes." Then

$$P(U_x=r) = P(r \text{ } Y_i \text{'s are } x, (m-r) \text{ not-}x)$$

$$= \binom{m}{r} [P(Y_1=x)]^r [1-P(Y_1=x)]^{m-r}$$

since the Y_i are independent and identically distributed. Thus by (1),

$$P(U_x=r) = \binom{m}{r} \left[\binom{n}{x} \left\{ \alpha p_{(1)}^x (1-p_{(1)})^{n-x} + \alpha' p_{(2)}^x (1-p_{(2)})^{n-x} \right\} \right]^r$$

$$(3) \quad \cdot \left[1 - \binom{n}{x} \left\{ \alpha p_{(1)}^x (1-p_{(1)})^{n-x} + \alpha' p_{(2)}^x (1-p_{(2)})^{n-x} \right\} \right]^{m-r}$$

To simplify the succeeding computations, denote

$$(4) \quad p_x = \binom{n}{x} \left[\alpha p_{(1)}^x (1-p_{(1)})^{n-x} + \alpha' p_{(2)}^x (1-p_{(2)})^{n-x} \right]$$

Thus equation (3) becomes

$$P(U_x=r) = \binom{m}{r} p_x^r (1-p_x)^{m-r} .$$

By a similar computation the joint distribution of U_0, \dots, U_n is seen to be multinomial:

$$(5) \quad P_{U_0, \dots, U_n}(u_0, \dots, u_n) = \frac{m!}{u_0! \dots u_n!} p_0^{u_0} p_1^{u_1} \dots p_n^{u_n},$$

for u_0, \dots, u_n non-negative integers whose sum is m .

Now if $p(1) < p(2)$, then for large n , $p(1)^n \ll p(2)^n$ and $(1-p(1))^n \gg (1-p(2))^n$,

so

$$\begin{aligned} p_1 &= \binom{n}{1} [\alpha p(1) (1-p(1))^{n-1} + \alpha' p(2) (1-p(2))^{n-1}] \\ &\approx n \alpha p(1) (1-p(1))^{n-1}, \end{aligned}$$

and

$$p_0 \approx \alpha (1-p(1))^n.$$

Thus

$$\frac{p_1}{p_0} \approx \frac{np(1)}{1-p(1)},$$

so

$$(6) \quad p(1) \approx \frac{p_1}{p_1 + np_0}.$$

Similarly,

$$(7) \quad p(2) \approx \frac{np_n}{p_{n-1} + np_n}.$$

Now the maximum likelihood estimators of the parameters in the multinomial distribution given by (5) are $\hat{p}_x = \frac{U_x}{m}$. This, in view of (6) and (7), suggests the "estimators"

$$(8) \quad \hat{p}(1) = \begin{cases} \frac{U_1}{U_1+nU_0} & \text{unless } U_0=U_1=0 \\ 1 & \text{if } U_0=U_1=0 \end{cases}$$

$$\hat{p}(2) = \begin{cases} \frac{nU_n}{nU_n+U_{n-1}} & \text{unless } U_n=U_{n-1}=0 \\ 0 & \text{if } U_n=U_{n-1}=0 \end{cases}$$

We now proceed to compute the distribution, expected value, and bias of $\hat{p}(1)$. Simple substitutions in these quantities for $\hat{p}(1)$ will yield the analogous quantities for $\hat{p}(2)$.

3(a). Distribution of $\hat{p}(1)$. The joint distribution of U_0, U_1 is

$$(9) \quad p_{U_0, U_1}(u_0, u_1) = \sum_{\substack{u_2, \dots, u_n \\ u_2 + \dots + u_n = m - u_0 - u_1}} \frac{m!}{u_0! \dots u_n!} p_0^{u_0} \dots p_n^{u_n}$$

$$= \frac{m!}{u_0! u_1! (m - u_0 - u_1)!} p_0^{u_0} p_1^{u_1} (1 - p_0 - p_1)^{m - u_0 - u_1}$$

where $0 \leq u_0 \leq m$, $0 \leq u_1 \leq m$, and $u_0 + u_1 \leq m$. Furthermore, for $0 < x < 1$,

$$(10) \quad P(\hat{p}(1) = x) = P\left(\frac{U_1}{U_1+nU_0} = x\right)$$

$$= P\left(U_1 = \frac{nx}{1-x} U_0\right)$$

$$= P\left(U_1=1, U_0 = \frac{1-x}{nx}\right) + P\left(U_1=2, U_0 = \frac{2-2x}{nx}\right) + \dots$$

Note that since U_0 and U_1 take on only integral values, (10) is positive only if x is of the form $\frac{k}{cn+k}$, where c and k are positive integers with $c+k \leq m$.

Thus (10) becomes

$$P(\hat{p}(1) = \frac{k}{cn+k}) = \sum_{j=1}^{m-1} P\left(U_1=j, U_0 = \frac{j(1-\frac{k}{cn+k})}{n(\frac{k}{cn+k})}\right)$$

$$\begin{aligned}
 &= \sum_{j=1}^{m-1} P(U_1=j, U_0 = \frac{jc}{k}) \\
 (11) \quad &= \sum_{j=1}^{m-1} p_{U_0, U_1}(\frac{cj}{k}, j)
 \end{aligned}$$

Since $p_{U_0, U_1}(\frac{cj}{k}, j) = 0$ unless cj/k is integral and $cj/k + j \leq m$, many of the terms in the sum (11) will vanish for general c, k . If, however, c and k are relatively prime, then (11) can be written as

$$(12) \quad P(\hat{p}_{(1)} = \frac{k}{cn+k}) = \sum_{j=1}^{\lfloor \frac{m}{c+k} \rfloor} \frac{m!}{(cj)!(kj)!(m-cj-kj)!} p_0^{cj} p_1^{nj} (1-p_0-p_1)^{m-jc-jk}$$

for $c=1, \dots, m-1$, $k=1, \dots, m-c$, and c, k relatively prime, where $[z]$ symbolizes the greatest integer less than or equal to z .

Finally,

$$\begin{aligned}
 (13) \quad P(\hat{p}_{(1)} = 0) &= P(U_1 = 0, U_0 > 0) \\
 &= \sum_{j=1}^m p_{U_0, U_1}(j, 0) \\
 &= \sum_{j=1}^m \frac{m!}{j!(m-j)!} p_0^j (1-p_0-p_1)^{m-j} \\
 &= (1-p_1)^m - (1-p_0-p_1)^m,
 \end{aligned}$$

and

$$\begin{aligned}
 (14) \quad P(\hat{p}_{(1)} = 1) &= P(U_0 = 0) \\
 &= (1-p_0)^m.
 \end{aligned}$$

3(b). Expected value of $\hat{p}_{(1)}$. From (12), (13), and (14), $E(\hat{p}_{(1)})$ can be computed directly as

$$(15) \quad E(\hat{p}_{(1)}) = \sum x P(\hat{p}_{(1)} = x)$$

$$= \sum_{k=1}^{m-1} \sum_{r=1}^{m-k} \left[\frac{r}{kn+r} \sum_{j=1}^{\lfloor \frac{m}{r+k} \rfloor} \frac{m!}{(kj)!(rj)!(m-rj-kj)!} p_0^{kj} p_1^{rj} (1-p_0-p_1)^{m-j(r+k)} \right]$$

k, r relatively prime

$$+ (1-p_0)^m$$

Equation (15) can be rearranged in various ways but cannot be put into any easily recognizable form and certainly does not reduce to a simple function of $p_{(1)}$ and $p_{(2)}$.

An approximation to $E(\hat{p}_{(1)})$ can be obtained by the following trick. Write $U_1 + nU_0 = W$, and let $\epsilon_W = (W - EW)$ and $\epsilon_{U_1} = (U_1 - EU_1)$. Then

$$\begin{aligned} \hat{p}_{(1)} &= \frac{U_1}{W} \\ &= \frac{EU_1 + \epsilon_{U_1}}{EW + \epsilon_W} \\ &= (EU_1 + \epsilon_{U_1}) \left[\frac{1}{EW} - \frac{\epsilon_W}{(EW)^2} + \frac{\epsilon_W^2}{(EW)^3} - \dots \right] \\ &= \frac{EU_1 + \epsilon_{U_1}}{EW} \cdot \sum_{j=0}^{\infty} (-1)^j \left(\frac{\epsilon_W}{EW} \right)^j \end{aligned}$$

for all ϵ_W for which this expansion is valid, i.e., for ϵ_W such that

$$\begin{aligned} |W - EW| &= |\epsilon_W| \\ &< |EW| \\ &= EW ; \end{aligned}$$

thus for W such that

$$0 < W < 2EW .$$

If this condition held with probability 1, then

$$(16) \quad E(\hat{p}_1) = \frac{1}{EW} E \left[(EU_1 + \epsilon_{U_1}) \sum_{j=0}^{\infty} (-1)^j \left(\frac{\epsilon_W}{EW} \right)^j \right],$$

in which case initial terms of the series could be evaluated to give a reasonable approximation to $E(\hat{p}_1)$. The condition in this case does not hold for all possible values of W . It is true, however, that ϵ_W/EW converges in probability to zero. According to Hansen, et. al. ([2], p. 107 ff) this is sufficient for the initial terms in the series in (16) to yield a useful approximation to $E(\hat{p}_1)$.

Now equation (16) expresses $E(\hat{p}_1)$ in terms of the moments of U_1 and W , and hence in terms of the moments of U_0 and U_1 . The joint moment generating function of U_0, U_1 is ([1], p. 418):

$$(17) \quad [p_0 e^{t_0} + p_1 e^{t_1} + (1-p_0-p_1) e^{t_2}]^m.$$

From (17) we find

$$\begin{aligned} EU_1 &= mp_1, \\ EW &= EU_1 + nEU_0 \\ &= mp_1 + nmp_0, \\ E\epsilon_W^2 &= E[(U_1 + nU_0) - (mp_1 + nmp_0)]^2 \\ &= mp_1(1-p_1) + n^2mp_0(1-p_0) - 2nmp_0p_1, \end{aligned}$$

and

$$E\epsilon_{U_1}\epsilon_W = mp_1(1-p_1-np_0).$$

Thus, since $E\epsilon_{U_1} = E\epsilon_W = 0$, ignoring terms of higher order than 2,

$$\begin{aligned}
 E(\hat{p}(1)) &= \frac{mp_1}{mp_1+nmp_0} + \frac{mp_1}{(mp_1+nmp_0)^3} [mp_1(1-p_1)+n^2mp_0(1-p_0)-2nmp_0p_1] \\
 &\quad - \frac{mp_1(1-p_1-np_0)}{(mp_1+nmp_0)^2} \\
 &= \frac{p_1}{p_1+np_0} + \frac{n(n-1)p_0p_1}{m(p_1+np_0)^3} \\
 &= \frac{\alpha(1-p(1))^{n-1}p(1)+\alpha'p(2)(1-p(2))^{n-1}}{\alpha(1-p(1))^{n-1}+\alpha'(1-p(2))^{n-1}} \\
 &\quad + \frac{n(n-1)}{m} \frac{[\alpha(1-p(1))^n+\alpha'(1-p(2))^n][n\alpha p(1)(1-p(1))^{n-1}+n\alpha'p(2)(1-p(2))^{n-1}]}{[n\alpha(1-p(1))^{n-1}+n\alpha'(1-p(2))^{n-1}]^3} \\
 &= \frac{p(1)+[\frac{\alpha'}{\alpha}p(2)(\frac{1-p(2)}{1-p(1)})^{n-1}]}{1+\frac{\alpha'}{\alpha}(\frac{1-p(2)}{1-p(1)})^{n-1}} \\
 &\quad + (\frac{n-1}{nm}) \frac{[\alpha(1-p(1))^n+\alpha'(1-p(2))^n][\alpha p(1)(1-p(1))^{n-1}+\alpha'p(2)(1-p(2))^{n-1}]}{[\alpha(1-p(1))^{n-1}+\alpha'(1-p(2))^{n-1}]^3}
 \end{aligned}$$

For a closer approximation, the third order terms are

$$\begin{aligned}
 &= \frac{EU_1 E\epsilon_W^3}{(EW)^4} + \frac{E\epsilon_{U_1} \epsilon_W^2}{(EW)^3} \\
 &= \frac{-mp_1 [mp_1(1-p_1)(1-2p_1)-3nmp_0p_1(1-2p_1)-3n^2mp_0p_1(1-2p_0)+n^3mp_0(1-p_0)(1-2p_0)]}{(mp_1+nmp_0)^4}
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{mp_1(1-p_1)(1-2p_1) - 2nmp_0p_1(1-2p_1) - n^2mp_0p_1(1-2p_0)}{(mp_1+nmp_0)^3} \\
 & = \frac{-np_0p_1}{m^2(p_1+nmp_0)^4} [(1-2p_1)(2np_0-1) + n(1-2p_0)(n-2p_1)] .
 \end{aligned}$$

Approximations ignoring terms of higher order than 3 will be tabulated in section 5 for several values of the parameters.

3(c). Estimation of α . In the case of mixtures of two binomials, always

$$(18) \quad p_n = \alpha p_{(1)}^n + (1-\alpha)p_{(2)}^n .$$

Thus, given any estimators of $p_{(1)}$ and $p_{(2)}$, an estimator of α can be constructed by solving (18) for α and replacing p_n , $p_{(1)}$ and $p_{(2)}$ by estimators of these parameters. Using the estimators of equation (8), this gives

$$\begin{aligned}
 (19) \quad \hat{\alpha} &= \frac{\frac{U_n}{m} - \hat{p}_{(2)}^n}{\hat{p}_{(1)}^n - \hat{p}_{(2)}^n} \\
 &= \frac{\left(\frac{nU_n}{U_{n-1} + nU_n}\right)^n - \frac{U_n}{m}}{\left(\frac{nU_n}{U_{n-1} + nU_n}\right)^n - \left(\frac{U_1}{U_1 + nU_0}\right)^n}
 \end{aligned}$$

Many other estimators of α can be constructed using $\hat{p}_{(1)}$ and $\hat{p}_{(2)}$. For example, since $p_0 = \alpha(1-p_{(1)})^n + (1-\alpha)(1-p_{(2)})^n$, by the same argument as above we get the estimator

$$(20) \quad \hat{\alpha}_1 = \frac{\frac{U_0}{m} - \left(\frac{U_{n-1}}{nU_n + U_{n-1}}\right)^n}{\left(\frac{nU_n}{U_1 + nU_0}\right)^n - \left(\frac{U_{n-1}}{nU_n + U_{n-1}}\right)^n}$$

Another possibility is

$$(21) \quad \hat{\alpha}_2 = \frac{\hat{\alpha} + \hat{\alpha}_1}{2} .$$

No attempt has been made to investigate the properties of these estimators.

4. Estimation of $p_{(1)}$ and $p_{(2)}$ using the entire sample frequency distribution. Another estimator of $(p_{(1)}, p_{(2)})$ can be constructed by noting the relationship between these parameters and the p_j . From equation (4) write out p_j and p_{j+1} :

$$(22) \quad \binom{n}{j}^{-1} p_j = \alpha p_{(1)}^j (1-p_{(1)})^{n-j} + \alpha' p_{(2)}^j (1-p_{(2)})^{n-j}$$

$$(23) \quad \binom{n}{j+1}^{-1} p_{j+1} = \alpha p_{(1)}^{j+1} (1-p_{(1)})^{n-j-1} + \alpha' p_{(2)}^{j+1} (1-p_{(2)})^{n-j-1} .$$

Add (22) and (23). The right hand side of the resulting sum is

$$(24) \quad \alpha p_{(1)}^j (1-p_{(1)})^{n-j-1} + \alpha' p_{(2)}^j (1-p_{(2)})^{n-j-1} ,$$

a polynomial in the $p_{(i)}$ of degree $(n-1)$. Now add to (24) the corresponding expression with j replaced by $(j+1)$. This gives

$$(25) \quad \binom{n}{j}^{-1} p_{j+2} + \binom{n}{j+1}^{-1} p_{j+1} + \binom{n}{j+2}^{-1} p_{j+2} = \alpha p_{(1)}^j (1-p_{(1)})^{n-j-2} + \alpha' p_{(2)}^j (1-p_{(2)})^{n-j-2} .$$

After k such steps beginning with p_j , we get (for $0 \leq k \leq n - j$)

$$(26) \quad \sum_{g=0}^k \binom{k}{g} \binom{n}{j+g}^{-1} p_{j+g} = \alpha p(1)^j (1-p(1))^{n-j-k} + \alpha' p(2)^j (1-p(2))^{n-j-k}$$

$$= F_{k,j}, \text{ say.}$$

Thus, for example,

$$F_{n-1,1} = \alpha p(1) + \alpha' p(2)$$

$$F_{n-2,1} = \alpha p(1)(1-p(1)) + \alpha' p(2)(1-p(2))$$

$$F_{n-k,k} = \alpha p(1)^k + \alpha' p(2)^k \quad (k \leq n) .$$

Thus

$$(27) \quad F_{n-1,1}^2 = \alpha^2 p(1)^2 + 2\alpha\alpha' p(1)p(2) + \alpha'^2 p(2)^2,$$

so, since $\alpha' = 1 - \alpha$,

$$(28) \quad F_{n-2,2} - F_{n-1,1}^2 = (\alpha - \alpha^2) p(1)^2 + (\alpha' - \alpha'^2) p(2)^2 - 2\alpha\alpha' p(1)p(2)$$

$$= \alpha(1-\alpha)(p(1) - p(2))^2$$

Thus, since $p(1) < p(2)$,

$$(29) \quad \sqrt{F_{n-2,2} - F_{n-1,1}^2} = \sqrt{\alpha(1-\alpha)} (p(2) - p(1)) .$$

Similarly,

$$(30) \quad \sqrt{F_{n-4,4} - F_{n-2,2}^2} = \sqrt{\alpha(1-\alpha)} (p(2)^2 - p(1)^2) .$$

Hence

$$(31) \quad + \sqrt{\frac{F_{n-4,4}^{-F_{n-2,2}^2}}{F_{n-2,2}^{-F_{n-1,1}^2}}} = \frac{p(2)^2 - p(1)^2}{p(2) - p(1)}$$

$$= p(1) + p(2) \cdot$$

By exactly the same argument, also

$$(32) \quad + \sqrt{\frac{F_{n-8,8}^{-F_{n-4,4}^2}}{F_{n-4,4}^{-F_{n-2,2}^2}}} = p(1)^2 + p(2)^2 \cdot$$

Now (31) and (32) must be solved for $p(1)$ and $p(2)$. This is made possible by the added restriction that $p(1) < p(2)$. For then

$$(33) \quad p(2) - p(1) = \frac{+ \sqrt{2(p(1)^2 + p(2)^2) - (p(1) + p(2))^2}}{2 + \sqrt{\frac{F_{n-8,8}^{-F_{n-4,4}^2}}{F_{n-4,4}^{-F_{n-2,2}^2}} - \frac{F_{n-4,4}^{-F_{n-2,2}^2}}{F_{n-2,2}^{-F_{n-1,1}^2}}}}$$

Thus

$$(34) \quad \left\{ \begin{aligned} p(1) &= \frac{1}{2} \left\{ + \sqrt{\frac{F_{n-4,4}^{-F_{n-2,2}^2}}{F_{n-2,2}^{-F_{n-1,1}^2}}} - + \sqrt{2 + \sqrt{\frac{F_{n-8,8}^{-F_{n-4,4}^2}}{F_{n-4,4}^{-F_{n-2,2}^2}} - \frac{F_{n-4,4}^{-F_{n-2,2}^2}}{F_{n-2,2}^{-F_{n-1,1}^2}}} \right\} \\ p(2) &= + \sqrt{\frac{F_{n-4,4}^{-F_{n-2,2}^2}}{F_{n-2,2}^{-F_{n-1,1}^2}}} - p(1) \end{aligned} \right.$$

Finally, since $F_{n-1,1} = \alpha p(1) + \alpha' p(2)$,

$$(35) \quad \alpha = \frac{p(2)^{-F_{n-1,1}}}{p(2)^{-p(1)}} \cdot$$

Thus, knowing the multinomial parameters in a population which is a mixture of two binomials, it is possible to compute exactly the two binomial parameters and the mixing parameter. Hence estimators can again be constructed by substituting for the $F_{k,j}$ in (34) and (35) the corresponding sample quantities.

Specifically, the maximum likelihood estimators of the parameters p_j in equation (5) are $\hat{p}_j = U_j/m$. Thus it is possible to estimate the $F_{k,j}$ by, say,

$$(36) \quad f_{k,j} = \frac{1}{m} \sum_{g=0}^k \binom{k}{g} \binom{n}{j+g}^{-1} U_{j+g} .$$

Note, however, that the $f_{k,j}$ are not one-to-one functions of the \hat{p}_j , so the $f_{k,j}$ are not maximum likelihood estimators of the $F_{k,j}$. (They are unbiased estimators, though, as we shall see.)

Estimators of the $p_{(i)}$ can now be constructed on the basis of the $f_{k,j}$, namely

$$(37) \quad \left\{ \begin{aligned} \hat{p}_{(1)}^* &= \frac{1}{2} \left\{ + \sqrt{\frac{f_{n-4,4} - f_{n-2,2}^2}{f_{n-2,2} - f_{n-1,1}^2}} - + \sqrt{2 + \sqrt{\frac{f_{n-8,8} - f_{n-4,4}^2}{f_{n-4,4} - f_{n-2,2}^2}} - \frac{f_{n-4,4} - f_{n-2,2}^2}{f_{n-2,2} - f_{n-1,1}^2}} \right\} \\ \hat{p}_{(2)}^* &= + \sqrt{\frac{f_{n-4,4} - f_{n-2,2}^2}{f_{n-2,2} - f_{n-1,1}^2}} - \hat{p}_{(1)}^* \\ \hat{\alpha}^* &= \frac{\hat{p}_{(2)}^* - f_{n-1,1}}{\hat{p}_{(2)}^* - \hat{p}_{(1)}^*} \end{aligned} \right.$$

5. Remarks on the estimators $\hat{p}_{(i)}$ and $\hat{p}_{(i)}^*$. The following table gives approximations for $E(\hat{p}_{(1)})$ for $n=10$ and several values of $(p_{(1)}, p_{(2)}, \alpha)$ ignoring terms of higher order than 3 in the series expansion of equation (16):

P(1)	P(2)	α	Approximation to $E\hat{p}_{(1)}$			
.20	.70	.20	.20029	+	$\frac{5.37}{m}$	- $\frac{209.07}{m^2}$
.30	.70	.20	.30078	+	$\frac{23.40}{m}$	- $\frac{3,133.17}{m^2}$
.35	.70	.20	.35133	+	$\frac{49.33}{m}$	- $\frac{12,950}{m^2}$
.40	.70	.20	.40233	+	$\frac{1065.4}{m}$	- $\frac{57,482}{m^2}$
.40	.80	.20	.40008	+	$\frac{107.15}{m}$	- $\frac{58,255}{m^2}$
.50	.90	.20	.49999	+	$\frac{575.94}{m}$	- $\frac{1,620,666}{m^2}$
.55	.95	.20	.55000	+	$\frac{1492.9}{m}$	- $\frac{10,841,610}{m^2}$
.40	.70	.40	.40088	+	$\frac{53.47}{m}$	- $\frac{14,440}{m^2}$
.40	.70	.50	.40059	+	$\frac{42.80}{m}$	- $\frac{9,240.51}{m^2}$
.20	.40	.20	.24619	+	$\frac{4.79}{m}$	- $\frac{141.22}{m^2}$
.30	.80	.20	.30003	+	$\frac{23.42}{m}$	- $\frac{3141.63}{m^2}$
.30	.70	.30	.30046	+	$\frac{15.61}{m}$	- $\frac{1,385.37}{m^2}$
.50	.75	.25	.50141	+	$\frac{458.07}{m}$	- $\frac{1,024,928}{m^2}$
.10	.60	.25	.10101	+	$\frac{.8421}{m}$	- $\frac{7.86}{m^2}$
.50	.90	.50	.49999	+	$\frac{230.40}{m}$	- $\frac{259,057}{m^2}$
.10	.90	.05	.10000	+	$\frac{4.18}{m}$	- $\frac{229.06}{m^2}$
.10	.40	.50	.10761	+	$\frac{.4348}{m}$	- $\frac{1.54}{m^2}$
.10	.30	.50	.11887	+	$\frac{.4407}{m}$	- $\frac{1.39}{m^2}$
.10	.30	.30	.13910	+	$\frac{.7460}{m}$	- $\frac{4.19}{m^2}$
.10	.20	.50	.12570	+	$\frac{.3792}{m}$	- $\frac{.8409}{m^2}$

Although admittedly a small sample, the calculations tabulated above serve to point out some interesting properties of the estimator $\hat{p}_{(1)}$ and of the approximation to $E(\hat{p}_{(1)})$.

In the first place, note that the first term in the approximation does not depend on m . Thus even for very large m the estimator $\hat{p}_{(1)}$ may be considerably biased. For example, for $p_{(1)}=.20$, $p_{(2)}=.40$ and $\alpha=.20$, if m is so large that the terms depending on m are negligible, we still get $E(\hat{p}_{(1)}) = .2462$. This large bias (.0462) is caused by the fact that with this set of parameters the approximation

$$\alpha p_{(1)}^n + \alpha' p_{(2)}^n = \alpha p_{(1)}^n$$

is poor. This will generally be true for $p_{(i)}$ which are quite close together, although the $p_{(i)}$ can be much nearer one another and still have $E(\hat{p}_{(i)})$ reasonably close to $p_{(i)}$ if α is near .5 than if α is near 0 or 1.

Note also from the above table that m must be large in order that the series expansion of equation (16) be valid. (Although m does not have to be large for this to hold for the last four cases tabulated, it will have to be very large in order that the analogous approximation to $E(\hat{p}_{(2)})$ be valid in these cases.) It is apparent that for small m the series (the first two terms of which are tabulated above) will, in most cases, diverge. In each case, however, if m is large enough to make the first term depending on m small (e.g., $< .01$), it will be large enough to make all succeeding terms small.

Finally, note that the definition of the estimators given in equation (8) for the case $\frac{0}{0}$ is arbitrary. Furthermore, the definition of $\hat{p}_{(i)}$ in these cases does not affect the above results since the approximation for $E(\hat{p}_{(1)})$ is given under the assumption that m is large enough that the indeterminate form has very small probability of occurring.

The estimators $\hat{p}_{(1)}^*$, $\hat{p}_{(2)}^*$, and α^* pose many problems. Note in particular the following:

i) Computation of their expectations will be a formidable task, if not, indeed, an impossible one. Note, however, that

$$\begin{aligned} Ef_{k,j} &= \frac{1}{m} E \sum_{g=0}^k \binom{k}{g} \binom{n}{j+g}^{-1} U_{j+g} \\ &= \frac{1}{m} \sum_{g=0}^k \binom{k}{g} \binom{n}{j+g}^{-1} (EU_{j+g}) \\ &= \sum_{g=0}^k \binom{k}{g} \binom{n}{j+g}^{-1} p_{j+g} \\ &= F_{k,j} \quad \text{by (24),} \end{aligned}$$

so $f_{k,j}$ is an unbiased estimate of $F_{k,j}$.

ii) A more immediate problem is that some of the terms under radicals in (37) are negative with positive probability. Furthermore, the indeterminate form $\frac{0}{0}$ can occur. How shall the estimators be defined in such cases?

It is possible to construct other functions of the $F_{k,j}$ yielding combinations of $p_{(1)}$ and $p_{(2)}$ such as those in (31) and (32). In no case, however, are the above difficulties overcome.

iii) Finally, for more than two binomials the result of (26) still holds, that is, for a mixture of r binomials with parameters $(n, p_{(1)}), \dots, (n, p_{(r)})$ and mixing parameters $\alpha_1, \dots, \alpha_r$, the computation leading to equation (26) yields

$$(38) \quad F_{k,j} = \sum_{i=1}^r \alpha_i p_{(i)}^j (1-p_{(i)})^{n-j-k} .$$

In this more general case, however, the computations following (26) will not eliminate the α 's. Thus the above results do not suggest a solution of the problem of mixtures of more than two binomials.

It should be pointed out, however, that the sample frequency distribution approach which led to equation (5) is non-parametric in the sense that for any mixture $f = \sum_{i=1}^k \alpha_i f_i$ (in fact, for any distribution f), a multinomial distribution will result from such a formulation. Specifically, let X_1, \dots, X_m be independent and identically distributed, each having distribution $f = \sum_{i=1}^k \alpha_i f_i$. Let $x_0 = -\infty < x_1 < x_2 < \dots < x_n < x_{n+1} = +\infty$, and define for $j=0, \dots, n$, $U_j =$ number of X_i 's such that $x_j < X_i \leq x_{j+1}$. Then the joint distribution of U_0, \dots, U_n is the multinomial distribution of equation (5) with parameters

$$p_j = \sum_{i=1}^k \alpha_i \int_{x_j}^{x_{j+1}} f_i \cdot$$

This may be of some help in the search for methods of constructing estimators in other parametric problems.

As a concluding remark, the results of a small sampling study are presented. From a mixture of binomials with $n=10$, $p_{(1)}=.35$, $p_{(2)}=.70$, and $\alpha=.20$ samples of sizes $m=1000$ and $m=1500$ were drawn. A sample of size $m=1000$ was also drawn from a population with $n=10$, $p_{(1)}=.20$, $p_{(2)}=.70$, and $\alpha=.20$. The estimators $\hat{p}_{(1)}$, $\hat{p}_{(2)}$, $\hat{p}_{(1)}^*$, and $p_{(2)}^*$ were computed where possible for each sample and, in the first case, for the composite of the two samples. The results were:

$p(1)$	$p(2)$	α	m	$\hat{p}(1)$	$\hat{p}(2)$	$\hat{p}^*(1)$	$\hat{p}^*(2)$
.35	.70	.20	1000	.1549	.7260	-	-
.35	.70	.20	1500	.2857	.6398	-	-
.35	.70	.20	2500	.2199	.6751	.4529	.5762
.20	.70	.20	.000	.2207	.7726	.1068	.7799

The blanks indicate that it was not possible to compute the star-estimators. In each case this was due to a negative number appearing under a radical at some point in the computation.

Of course, no conclusions can be drawn on the basis of so few samples. It is clear, though, that the p_j 's must be uniformly very accurately estimated in order that the $p_{(i)}^*$ yield at all reasonable estimates. The $\hat{p}_{(i)}$ require that $p_0, p_1, p_{n-1},$ and p_n be uniformly very accurately estimated. Thus both estimators require extremely large sample sizes.

References

- [1] Cramer, Harold, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1957.
- [2] Hansen, Morris H., William N. Hurwitz, and William G. Madow, Sample Survey Methods and Theory, Vol. II, John Wiley and Sons, New York, 1953.
- [3] Robbins, Herbert, "Mixture of distributions," Annals of Math. Stat. Vol. 19 (1948), p. 360.
- [4] Robbins, Herbert, and E. J. G. Pitman, "Application of the method of mixtures to quadratic forms in normal variates," Annals of Math. Stat., Vol. 20 (1949), p. 552.
- [5] Teicher, Henry, "On the mixture of distributions," Annals of Math. Stat., Vol. 31 (1960), p. 55.