MOMENT ESTIMATORS FOR THE PARAMETERS OF A MIXTURE OF TWO BINOMIAL DISTRIBUTIONS[1]

By W. R. Blischke[2]

Cornell University

1. Introduction and summary. Mixtures of distributions present two types of problems. The first is the problem of identifiability; that is, given that a distribution function F is a probability mixture of distribution functions belonging to some family $\mathfrak{F}$, is the mixture unique? This topic has been dealt with quite extensively in recent papers by Robbins [9], Teicher [10] and [11], and others.

The second problem is that of estimating the parameters of the individual distribution functions comprising the mixture and the mixing measure. This is clearly possible only if the given mixture is identifiable. K. P. Pearson [5] and C. R. Rao [6] consider the problem of estimation for a mixture of two normal distributions and P. Rider [7 and 8] has recently constructed estimators for mixtures of two of either the exponential, Poisson, binomial, negative binomial or Weibull distributions. In each of these cases the method of maximum likelihood yields highly intractable equations. All of the above estimators have been constructed by the method of moments. In this paper moment estimators will be constructed for a mixture of two binomial distributions, (n, p_1) and (n, p_2). The construction presented here parallels that of Rider [8]. The limiting distributions of the estimators and their asymptotic relative efficiency will be computed.

It will be shown that as the binomial parameter $n \to \infty$ the asymptotic efficiency of the moment estimators tends to unity.

Finally, moment estimators will be constructed for the binomial parameters when the mixing parameter α is known. In this case an apparently anomalous result is obtained in that the asymptotic efficiencies of the analogous moment estimators when α is known tend to 0 rather than 1 as $n \to \infty$.

It is pointed out that it is possible, however, to construct moment estimators whose efficiency tends to 1 in this case as well. The estimators so constructed do not depend on the known proportion α when $n \geqq 3$. This suggests a possible explanation of the above anomaly: This fact that the asymptotic efficiencies tend to 0 rather than 1 as $n \to \infty$ may be due to the failure of these estimators to take into account sample deviations from the true proportions in which the respective populations are present in the mixture.

2. Construction of moment estimators for a mixture of two binomial distributions. Let $Y_1, \cdots, Y_m$ be independent and identically distributed chance vari-

Received May 7, 1961; revised November 16, 1961.

[1] Research supported by the office of Naval Research, Project No. NR 042-212, Contract No. Nonr-401(39). Reproduction in whole or in part is permitted for any purpose of the United States Government. This research forms part of a dissertation submitted to Cornell University in partial fulfillment of the requirements for the Ph.D. degree.

[2] Present address: North Carolina State College.

ables, each having distribution

$$P(Y_i = y) = \binom{n}{y} [\alpha p_1^y(1 - p_1)^{n-y} + (1 - \alpha)p_2^y(1 - p_2)^{n-y}] \quad \text{if } y = 0, 1, \cdots, \text{or } n \tag{1}$$

$$= 0 \quad \text{otherwise,}$$

where $0 < \alpha < 1$, $0 < p_1 < p_2 < 1$, and $n \geqq 3$ is integral. Equation (1) is a mixture of two binomial distributions and is easily seen to be identifiable only if $n \geqq 3$. Under this assumption, estimators for the parameters p_1, p_2, and α may be constructed as functions of certain sample factorial moments. It should be noted that in the following construction, as in the example considered by Rider [7], the assumption $p_1 < p_2$ is essential. The estimators do not have the stated properties if $p_1 = p_2$. (Note that it is actually required only that $p_1 \neq p_2$; the particular assumption that $p_1 < p_2$ is for notational convenience.) The construction of the moment estimators is as follows:

Define the kth sample factorial moment to be

$$F_k = \frac{1}{m} \sum_{i=1}^{m} \frac{Y_i(Y_i - 1)\cdots(Y_i - k + 1)}{n(n - 1)\cdots(n - k + 1)} \qquad k = 1, \cdots, n, \tag{2}$$

and denote its expectation by f_k. (Equation (2) differs by a constant multiplier from the usual definition of a factorial moment. Cf. Kendall and Stuart [3], Sections 3.7–3.10.) To compute f_k, define the chance variables

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

for $j = 0, \cdots, n$. Then $U_0, \cdots, U_n$ are jointly multinomially distributed with parameters $\pi_j = P\{Y_i = j\}, j = 0, \cdots, n$. Furthermore, F_k can now be written

$$F_k = \frac{1}{m} \sum_{j=0}^{n-k} \binom{n-k}{j} \binom{n}{k+j}^{-1} U_{j+k}.$$

Thus

$$\begin{aligned} f_k &= \sum_{j=0}^{n-k} \binom{n-k}{j} \binom{n}{k+j}^{-1} \pi_{j+k} \\ &= \alpha p_1^k + (1 - \alpha)p_2^k. \end{aligned} \tag{3}$$

In constructing estimators, the moments given in (3) are considered as equations in the three unknowns α, p_1, and p_2. Any three such equations may be solved for the three parameters. We shall consider f_1, f_2, and f_3. Note that

$$f_2 - f_1^2 = \alpha(1 - \alpha)(p_1 - p_2)^2$$

and

$$f_3 - f_1 f_2 = \alpha(1 - \alpha)(p_1 + p_2)(p_1 - p_2)^2,$$

so that $p_1 + p_2 = (f_3 - f_1f_2)/(f_2 - f_1^2) = a$, say. Now from f_1,

$$\alpha = \frac{f_1 - p_2}{p_1 - p_2}, \tag{4}$$

which, when substituted into f_2, yields

$$f_2 = \frac{f_1 - p_2}{p_1 - p_2}(p_1^2 - p_2^2) + p_2^2 = (f_1 - p_2)a + p_2^2.$$

Thus

$$p_2^2 - ap_2 + f_1a - f_2 = 0. \tag{5}$$

Solving for p_1 instead of p_2 yields equation (5) with p_1 replacing p_2. Thus the restriction that $p_1 < p_2$ results in the unique solution

$$p_2, p_1 = \tfrac{1}{2}a \pm \tfrac{1}{2}(a^2 - 4af_1 + 4f_2)^{\frac{1}{2}},$$

which together with equation (4) expresses p_1, p_2 and α as functions of f_1, f_2, and f_3.

Moment estimators are now specified by substituting F_j for f_j, $j = 1, 2, 3$, in equations (4) and (6). This yields the estimators

$$\begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}A - \frac{1}{2}(A^2 - 4AF_1 + 4F_2)^{\frac{1}{2}} \\ \frac{1}{2}A + \frac{1}{2}(A^2 - 4AF_1 + 4F_2)^{\frac{1}{2}} \\ (F_1 - \hat{p}_2)/(\hat{p}_1 - \hat{p}_2) \end{bmatrix}, \tag{7}$$

where

$$A = \frac{F_3 - F_1F_2}{F_2 - F_1^2}.$$

The estimators of equation (7) have the unpleasant property of assuming complex as well as indeterminate values with positive probability, though this probability tends to 0 as $m \to \infty$. The event may be avoided altogether by using instead, e.g.,

$$\begin{aligned} &\begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \hat{\alpha} \end{bmatrix} \quad \begin{array}{l} \text{if } A^2 - 4AF_1 + 4F_2 > 0 \\ \text{and } (A^2 - 4AF_1 + 4F_2)^{\frac{1}{2}} \leqq \min(A, 2 - A), \end{array} \\ &\begin{bmatrix} F_1 \\ F_1 \\ 0 \end{bmatrix} \quad \text{otherwise.} \end{aligned} \tag{8}$$

Because of the discreteness of the problem, exact distributions of the estimators of equations (7) and (8) are not difficult to obtain. In fact, one can, formally at least, write expectations and variances for the above expressions. Because of the complexity of the functions involved, however, these moments cannot be written in a simple form and consequently shed little light on the behavior of the estimators.

The following two sections of this paper are devoted to asymptotic properties of the estimators of equation (7) (equivalently, of (8)), and a comparison of these estimators with the maximum likelihood estimators p_1^* , p_2^* , α^*.

3. Asymptotic behavior of the moment estimators. The limiting joint distribution of $\hat{p}_1$, $\hat{p}_2$, $\hat{\alpha}$ as $m \to \infty$ is readily obtained by application of an obvious extension of a theorem of Hoeffding and Robbins ([2], Theorem 4). The theorem requires the existence of a differential of the transformation specified by $H_1(F_1, F_2, F_3) = \hat{p}_1$, $H_2(F_1, F_2, F_3) = \hat{p}_2$, $H_3(F_1, F_2, F_3) = \hat{\alpha}$ evaluated at $f = (f_1, f_2, f_3)$ and that $E|F_i|^3 < \infty$, $i = 1, 2, 3$. These conditions are satisfied for the example under consideration. We conclude that $\hat{p}_1$, $\hat{p}_2$, $\hat{\alpha}$ are asymptotically normally distributed with respective means p_1 , p_2 , α and covariance matrix $\Sigma_{\hat{p}_1,\hat{p}_2,\hat{\alpha}}$ having entries

$$\sigma_{ii'} = \sum_{j=1}^{3} \mu_2(F_i)\zeta_{ij}\zeta_{i'j} + \sum_{\substack{j,j'=1\\ j\neq j'}}^{3} \mu_{11}(F_j, \; F_{j'})\zeta_{ij}\zeta_{i'j'}, \tag{9}$$

where $\zeta_{ij} = (\partial H_i/\partial F_j) \mid_f$.

We proceed to the computation of $\Sigma_{\hat{p}_1,\hat{p}_2,\hat{\alpha}}$. To compute this covariance matrix, the covariance matrix of (F_1, F_2, F_3) and the first order partial derivatives $\partial H_i/\partial F_j$ evaluated at $f = (f_1, f_2, f_3)$ are required for $i, j = 1, 2, 3$.

The moment generating function of the distribution of equation (1) is

$$\phi(t) = \alpha(1 - p_1 + p_1e^t)^n + (1 - \alpha)(1 - p_2 + p_2e^t)^n. \tag{10}$$

From equation (10) the necessary moments of the F_i are obtained. By equation (3) these can be written as

$$\begin{aligned}
\mu_2(F_1) &= \frac{1}{mn}[f_1 + (n-1)f_2] - \frac{1}{m}(f_1)^2 \\
\mu_2(F_2) &= \frac{1}{mn(n-1)}[2f_2 + 4(n-2)f_3 + (n-2)(n-3)f_4] - \frac{1}{m}(f_2)^2 \\
\mu_2(F_3) &= \frac{1}{mn(n-1)(n-2)}[6f_3 + 18(n-3)f_4 + 9(n-3)(n-4)f_5 \\
&\qquad + (n-3)(n-4)(n-5)f_6] - \frac{1}{m}(f_3)^2 \\
\mu_{11}(F_1, F_2) &= \frac{1}{mn}[2f_2 + (n-2)f_3] - \frac{1}{m}f_1 f_2 \\
\mu_{11}(F_1, F_3) &= \frac{1}{mn}[3f_3 + (n-2)f_4] - \frac{1}{m}f_1 f_3 \\
\mu_{11}(F_2, F_3) &= \frac{1}{mn(n-1)}[6f_3 + 6(n-3)f_4 + (n-3)(n-4)f_5] - \frac{1}{m}f_2 f_3 .
\end{aligned} \tag{11}$$

Finally,

$$(12)\qquad \left[\frac{\partial H_i}{\partial F_j}\bigg|_f\right] = \frac{1}{(p_1 - p_2)^2}\begin{bmatrix} \dfrac{p_2(p_2 + 2p_1)}{\alpha} & -\dfrac{p_1 + 2p_2}{\alpha} & \dfrac{1}{\alpha} \\ \dfrac{p_1(p_1 + 2p_2)}{1 - \alpha} & -\dfrac{2p_1 + p_2}{1 - \alpha} & \dfrac{1}{1 - \alpha} \\ \dfrac{6p_1 p_2}{p_2 - p_1} & -\dfrac{3(p_1 + p_2)}{p_2 - p_1} & \dfrac{2}{p_2 - p_1} \end{bmatrix}$$

After some computation the elements of the asymptotic covariance matrix of $\hat{p}_1$, $\hat{p}_2$, $\hat{\alpha}$ are now found to be

$$(13)\qquad \begin{aligned} \sigma^2_{\hat{p}_1} &= \frac{1}{\alpha^2\delta^4}[p_2^2(p_2 + 2p_1)^2\mu_2(F_1) + (p_1 + 2p_2)^2\mu_2(F_2) + \mu_2(F_3) \\ &\quad - 2p_2(p_2 + 2p_1)(p_1 + 2p_2)\mu_{11}(F_1, F_2) + 2p_2(p_2 + 2p_1)\mu_{11}(F_1, F_3) \\ &\quad - 2(p_1 + 2p_2)\mu_{11}(F_2, F_3)] \\ &= \frac{p_1 q_1}{\alpha m n} + \binom{n}{2}^{-1}\frac{4\alpha p_1^2 q_1^2 + \alpha' p_2^2 q_2^2}{m\alpha^2\delta^2} + \binom{n}{3}^{-1}\frac{\alpha p_1^3 q_1^3 + \alpha' p_2^3 q_2^3}{m\alpha^2\delta^4} \\ \sigma^2_{\hat{p}_2} &= \frac{p_2 q_2}{\alpha' m n} + \binom{n}{2}^{-1}\frac{\alpha p_1^2 q_1^2 + 4\alpha' p_2^2 q_2^2}{m\alpha'^2\delta^2} + \binom{n}{3}^{-1}\frac{\alpha p_1^3 q_1^3 + \alpha' p_2^3 q_2^3}{m\alpha'^2\delta^4} \\ \sigma^2_{\hat{\alpha}} &= \frac{\alpha\alpha'}{m} + 9\binom{n}{2}^{-1}\frac{\alpha p_1^2 q_1^2 + \alpha' p_2^2 q_2^2}{m\delta^4} + 4\binom{n}{3}^{-1}\frac{\alpha p_1^3 q_1^3 + \alpha' p_2^3 q_2^3}{m\delta^6} \\ \sigma_{\hat{p}_1,\hat{p}_2} &= 2\binom{n}{2}^{-1}\frac{\alpha p_1^2 q_1^2 + \alpha' p_2^2 q_2^2}{m\alpha\alpha'\delta^2} + \binom{n}{3}^{-1}\frac{\alpha p_1^3 q_1^3 + \alpha' p_2^3 q_2^3}{m\alpha\alpha'\delta^4} \\ \sigma_{\hat{p}_1,\hat{\alpha}} &= 3\binom{n}{2}^{-1}\frac{2\alpha p_1^2 q_1^2 + \alpha' p_2^2 q_2^2}{m\alpha\delta^3} + 2\binom{n}{3}^{-1}\frac{\alpha p_1^3 q_1^3 + \alpha' p_2^3 q_2^3}{m\alpha\delta^5} \\ \sigma_{\hat{p}_2,\hat{\alpha}} &= 3\binom{n}{2}^{-1}\frac{\alpha p_1^2 q_1^2 + 2\alpha' p_2^2 q_2^2}{m\alpha'\delta^3} + 2\binom{n}{3}^{-1}\frac{\alpha p_1^3 q_1^3 + \alpha' p_2^3 q_2^3}{m\alpha'\delta^5} \end{aligned}$$

where $q_1 = 1 - p_1$, $q_2 = 1 - p_2$, $\alpha' = 1 - \alpha$, and $\delta = p_2 - p_1$.

4. Asymptotic relative efficiency of the moment estimators. The asymptotic efficiency (ARE) of a consistent asymptotically normally distributed estimator $\hat{\theta}$ of a parameter θ relative to the maximum likelihood estimator θ^* is computed as ARE $(\hat{\theta}) = \sigma^2_{\theta^*}/\sigma^2_{\hat{\theta}}$, where $\sigma^2_{\theta^*}$ is the Cramér-Rao lower bound and $\sigma^2_{\hat{\theta}}/m$ is the variance in the limiting distribution of $\hat{\theta}$. If $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_r)$ is an estimator of a vector parameter $\theta = (\theta_1, \cdots, \theta_r)$, with the components of $\hat{\theta}$ asymptotically jointly normally distributed with mean θ and covariance matrix $(1/m)\Sigma_{\hat{\theta}}$, asymptotic relative efficiency may be computed in this way for each component of θ, or the components may be considered jointly. In the latter case, the joint asymptotic efficiency (JARE) of $\hat{\theta}$ relative to the maximum likelihood estimator θ^* is computed as the square of the ratio of the areas of the ellipses of concentration of the respective asymptotic normal distributions. (Cf. [1], chapter 32).

Since the areas of the ellipses of concentration are proportional to the determinants of the respective covariance matrices,

$$\text{JARE}\ (\hat{\theta}) = \frac{\det\ (\Sigma_{\theta^*})}{\det\ (\Sigma_{\hat{\theta}})}. \tag{14}$$

The elements of the inverse of $\Sigma_{\theta^*} = \Sigma_{p_1^*, p_2^*, \alpha^*}$ for the example under consideration are easily computed, e.g., the upper left hand element of $\Sigma^{-1}_{p_1^*, p_2^*, \alpha^*}$ is

$$E\left(\frac{\partial \log P(Y)}{\partial p_1}\right)^2 = \sum_{y=0}^{n} \left[\frac{\partial P(y)}{\partial p_1} \frac{1}{P(y)}\right]^2 P(y)$$

$$= \alpha^2 \sum_{y=0}^{n} \frac{\binom{n}{y}^2 p_1^{2y-2}(1 - p_1)^{2(n-y-1)}(y - np_1)^2}{P(y)}$$

$$= \frac{\alpha^2}{p_1^2(1 - p_1)^2} \sum_{y=0}^{n} \frac{[P_1(y)(y - np_1)]^2}{P(y)},$$

where $P(y) = P(Y = y)$ is given by equation (1), and $P_1(y) = \binom{n}{y} p_1^y(1 - p_1)^{n-y}$. With the further notation $P_2(y) = \binom{n}{y} p_2^y(1 - p_2)^{n-y}$, the desired matrix may be written

$$\Sigma^{-1}_{p_1^*, p_2^*, \alpha^*} = \begin{pmatrix} \alpha^2\Sigma \frac{[P_1(y)(y - np_1)]^2}{p_1^2 q_1^2 P(y)} & \alpha\alpha'\Sigma \frac{P_1(y)P_2(y)(y - np_1)(y - np_2)}{p_1 p_2 q_1 q_2 P(y)} & \alpha\Sigma \frac{P_1(y)(y - np_1)[P_1(y) - P_2(y)]}{p_1 q_1 P(y)} \\ \cdots & \alpha'^2\Sigma \frac{[P_2(y)(y - np_2)]^2}{p_2^2 q_2^2 P(y)} & \alpha'\Sigma \frac{P_2(y)(y - np_2)[P_1(y) - P_2(y)]}{p_2 q_2 P(y)} \\ \cdots & \cdots & \Sigma \frac{[P_1(y) - P_2(y)]^2}{P(y)} \end{pmatrix} \tag{15}$$

The elements of $(1/m)\Sigma_{\hat{p}_1, \hat{p}_2, \hat{\alpha}}$ are given by equation (13). Asymptotic efficiencies can now be computed as indicated above.

It is interesting to note the behavior of the asymptotic relative efficiencies as functions of the parameter n. It is easily seen that the moment estimators are a solution to the maximum likelihood equations if $n = 3$, so that for $n = 3$ the asymptotic efficiencies are unity. For $n > 3$ this is not true, however, since in fact $\det\ (\Sigma_{p_1^*, p_2^*, \alpha^*}) < \det\ (\Sigma_{\hat{p}_1, \hat{p}_2, \hat{\alpha}})$ in this case.

It will now be shown that as $n \to \infty$, however, JARE $(\hat{p}_1, \hat{p}_2, \hat{\alpha}) \to 1$. By equation (13), as $n \to \infty$

$$\Sigma_{\hat{p}_1, \hat{p}_2, \hat{\alpha}} = \begin{pmatrix} \frac{p_1(1 - p_1)}{\alpha n} & 0 & 0 \\ 0 & \frac{p_2(1 - p_2)}{(1 - \alpha)n} & 0 \\ 0 & 0 & \alpha(1 - \alpha) \end{pmatrix} + O(n^{-2}). \tag{16}$$

Thus to prove the assertion it suffices to show that as $n \to \infty$,

$$(17) \qquad \Sigma^{-1}_{p_1^*, p_2^*, \alpha^*} = \begin{pmatrix} \frac{n\alpha}{p_1(1-p_1)} & 0 & 0 \\ 0 & \frac{n(1-\alpha)}{p_2(1-p_2)} & 0 \\ 0 & 0 & \frac{1}{\alpha(1-\alpha)} \end{pmatrix} + O(n^{-2}).$$

We shall prove the convergence for the upper left-hand elements of the matrices of equation (17). The convergence of the other entries follows similarly.

We wish to show that as $n \to \infty$,

$$(18) \qquad \sum_{y=0}^{n} \frac{P_1^2(y)(y - np_1)^2}{P(y)} = \frac{np_1(1-p_1)}{\alpha} + O(n^{-2}).$$

The proof follows from the following result of Okamoto ([4], Theorem 1): If X is a binomially distributed chance variable with parameters n and p, and c a non-negative constant, depending possibly on n or p, then

(i) $P((X/n) - p \geqq c) < e^{-2nc^2}$,

(ii) $P((X/n) - p \leqq -c) < e^{-2nc^2}$.

We have

$$\left| \sum_{y=0}^{n} \frac{P_1^2(y)(y-np_1)^2}{P(y)} - \frac{np_1(1-p_1)}{\alpha} \right| = \left| \sum_{y=0}^{n} P_1(y)(y-np_1)^2 \left[\frac{P_1(y)}{P(y)} - \frac{1}{\alpha} \right] \right|$$

$$= \frac{1-\alpha}{\alpha} \sum_{y=0}^{n} \frac{P_1(y)P_2(y)(y-np_1)^2}{\alpha P_1(y) + (1-\alpha)P_2(y)} \leqq \frac{1-\alpha}{\alpha^2} n^2 \sum_{y \leqq \frac{1}{2}n(p_1+p_2)} P_2(y)$$

$$+ \frac{1}{\alpha} n^2 \sum_{y \geqq \frac{1}{2}n(p_1+p_2)} P_1(y) < \left(\frac{n}{\alpha}\right)^2 e^{-\frac{1}{2}n(p_1-p_2)^2}$$

by Okamoto's theorem with $c = \frac{1}{2}(p_2 - p_1)$. Since $p_1 \neq p_2$ by assumption, the last expression can be made arbitrarily small by choosing n sufficiently large. This verifies equation (18).

As an illustration of the above phenomena, the ARE for the individual estimators and the JARE are given as functions of n for $\alpha = .4$ and $(p_1, p_2) = (.3, .6)$ and $(.1, .8)$ in Table 1. It is evident from the table that the ARE depends upon how close p_1 and p_2 are. It would be of interest to determine as functions of α, p_1, p_2 (or $p_2 - p_1$) the minimum asymptotic efficiencies as well as the values of n at which the minima occur. This problem has not been investigated.

5. Estimation of the binomial parameters when α is known. If in the distribution of equation (1) the mixing parameter α is known, then $n \geqq 2$ is required for identifiability. In this case the estimators which are maximum likelihood when $n = 2$ are obtained by solving the first two factorial moments for the two

TABLE 1

Asymptotic efficiencies of the moment estimators as functions of n for $(p_1, p_2, \alpha) = (.3, .6, .4)$ and $(.1, .8, .4)$.

n	$p_1 = .3,\ p_2 = .6$				$p_1 = .1,\ p_2 = .8$			
	ARE			JARE	ARE			JARE
	$\hat{p}_1$	$\hat{p}_2$	$\hat{\alpha}$		$\hat{p}_1$	$\hat{p}_2$	$\hat{\alpha}$	
3	1	1	1	1	1	1	1	1
4	.975	.962	.966	.960	.876	.816	.885	.771
5	.953	.932	.937	.921	.857	.819	.908	.739
6	.929	.903	.907	.882	.800	.790	.907	.666
8	.889	.859	.858	.816	.788	.793	.933	.646
10	.856	.827	.821	.763	.792	.805	.950	.651
20	.768	.752	.766	.623	.855	.874	.985	.749
49	.736	.744	.872	.562	.937	.946	.998	.886
∞	1	1	1	1	1	1	1	1

unknown parameters. The solution again requires that $p_1 < p_2$. It should be noted that here this is a much more restrictive assumption than in the previous case. Unless $\alpha = \frac{1}{2}$, it is not sufficient when α is known to know simply that $p_1 \neq p_2$; it must be known specifically that α is the proportion in which the population having smaller mean is present in the mixture (Cf. Rider's discussion on this point in [7], p. 145).

Under this assumption a development similar to that of Section 2 yields as functions of the first two factorial moments the moment estimators

$$\tag{19} \begin{aligned} \bar{p}_1 &= F_1 - \left[\frac{1-\alpha}{\alpha}(F_2 - F_1^2)\right]^{\frac{1}{2}} \\ \bar{p}_2 &= F_1 + \left[\frac{\alpha}{1-\alpha}(F_2 - F_1^2)\right]^{\frac{1}{2}}. \end{aligned}$$

These estimators are efficient for $n = 2$. We shall now show, however, that the asymptotic relative efficiencies of the estimators of equation (19) tend to 0 rather than to 1 as $n \to \infty$.

The entries of the covariance matrix $(1/m)\Sigma_{\bar{p}_1,\bar{p}_2}$ are found by the methods of Section 3 to be

$$\tag{20} \begin{aligned} \sigma_{\bar{p}_1}^2 &= \frac{p_1 q_1}{\alpha m n} + \frac{\alpha' \delta^2}{4\alpha m} + \binom{n}{2}^{-1} \frac{\alpha p_1^2 q_1^2 + \alpha' p_2^2 q_2^2}{4\alpha^2 m \delta^2} \\ \sigma_{\bar{p}_2}^2 &= \frac{p_2 q_2}{\alpha' m n} + \frac{\alpha \delta^2}{4\alpha' m} + \binom{n}{2}^{-1} \frac{\alpha p_1^2 q_1^2 + \alpha' p_2^2 q_2^2}{4\alpha'^2 m \delta^2} \\ \sigma_{\bar{p}_1,\bar{p}_2} &= \frac{\delta^2}{4m} - \binom{n}{2}^{-1} \frac{\alpha p_1^2 q_1^2 + \alpha' p_2^2 q_2^2}{4\alpha\alpha' m \delta^2}, \end{aligned}$$

where the notation is as in Section 3. The relevant comparison in this case is between $\Sigma_{\bar{p}_1,\bar{p}_2}$ and the inverse of

$$\Sigma^{-1}_{p_1^{**},p_2^{**}} = \begin{pmatrix} \dfrac{\alpha^2}{p_1^2 q_1^2} \Sigma \dfrac{[P_1(y)(y - np_1)]^2}{P(y)} & \dfrac{\alpha\alpha'}{p_1 p_2 q_1 q_2} \Sigma \dfrac{P_1(y)P_2(y)(y - np_1)(y - np_2)}{P(y)} \\ \cdots & \dfrac{\alpha'^2}{p_2^2 q_2^2} \Sigma \dfrac{[P_2(y)(y - np_2)]^2}{P(y)} \end{pmatrix}. \tag{21}$$

By the results of the previous section, the entries of $\Sigma_{p_1^{**},p_2^{**}}$ are $O(n^{-1})$ as $n \to \infty$. By equation (20), however, the entries of $\Sigma_{\bar{p}_1,\bar{p}_2}$ are $O(1)$, so that asymptotic efficiencies are $O(n^{-1})$ as $n \to \infty$.

Thus the estimators of equation (19) do not have the desirable properties for large n of the moment estimators of the previous sections. It has been suggested that this apparent anomaly may be due to the choice of moments for construction of the estimators, and that if some other functions had been chosen, efficient estimators could be obtained by a straightforward construction when α is known as well. It is not at all evident, however, that this should be an inherent property of this class of estimators. An alternate possible explanation is that the estimators of equation (19) do not take into account sample deviations from the true proportions in which the respective populations are present in the mixture. This explanation may be intuitively more appealing, particularly in view of the following result:

Note that if $n \geqq 3$, the estimators $\hat{p}_1$ and $\hat{p}_2$ (of equation (7)) are meaningful and have the same properties as before. This follows since $\sigma^2_{\hat{p}_1}$, $\sigma^2_{\hat{p}_2}$ and $\sigma_{\hat{p}_1,\hat{p}_2}$ remain the same so that by equation (13),

$$\Sigma_{\hat{p}_1,\hat{p}_2} = \begin{pmatrix} \dfrac{p_1(1 - p_1)}{\alpha n} & 0 \\ 0 & \dfrac{p_2(1 - p_2)}{(1 - \alpha)n} \end{pmatrix} + O(n^{-2}).$$

Furthermore, applying the results of the previous section to equation (21), we get

$$\Sigma^{-1}_{p_1^{**},p_2^{**}} = \begin{pmatrix} \dfrac{\alpha n}{p_1(1 - p_1)} & 0 \\ 0 & \dfrac{(1 - \alpha)n}{p_2(1 - p_2)} \end{pmatrix} + O(n^{-2})$$

so that ARE of $\hat{p}_1$ and $\hat{p}_2$ tends to 1 as $n \to \infty$, as claimed.

We can now combine the estimators $\hat{p}_1$, $\hat{p}_2$ and $\bar{p}_1$, $\bar{p}_2$ to form estimators, p_1' and p_2', say, which are maximum likelihood (and hence efficient) for $=2$

and have asymptotic efficiency tending to unity for large n, namely

$$(22) \qquad \begin{aligned} (p_1', p_2') &= (\bar{p}_1, \bar{p}_2) \quad \text{if} \quad n = 2 \\ &= (\hat{p}_1, \hat{p}_2) \quad \text{if} \quad n \geqq 3. \end{aligned}$$

Thus it is possible to construct efficient estimators for large n when α is known. Notice, however, that the estimators so constructed do not make use of the knowledge of α when $n \geqq 3$. This may indicate that the latter explanation of the abnormal behavior of $\bar{p}_1$, $\bar{p}_2$ given above has some merit.

TABLE 2

Asymptotic relative efficiences of the moment estimators $\bar{p}_1$, $\bar{p}_2$ and p_1', p_2' as functions of n for $(p_1, p_2, \alpha) = (.3, .6, .4)$ and $(.1, .8, .4)$.

n	ARE($\bar{p}_1$)	ARE($\bar{p}_2$)	JARE	ARE(p_1')	ARE(p_2')	JARE
			$p_1 = .3, \quad p_2 = .6$			
2	1	1	1	1	1	1
3	.903	.737	.636	.039	.044	.011
4	.968	.775	.738	.084	.093	.032
5	.954	.760	.722	.124	.137	.056
6	.941	.751	.707	.167	.180	.083
10	.880	.731	.641	.321	.335	.205
20	.661	.621	.437	.587	.598	.448
49	.281	.334	.151	.729	.739	.584
∞	0	0	0	1	1	1
			$p_1 = .1, \quad p_2 = .8$			
2	1	1	1	1	1	1
3	.590	.880	.589	.616	.649	.564
4	.408	.648	.370	.758	.729	.650
5	.301	.527	.261	.801	.778	.690
6	.226	.428	.186	.782	.776	.664
10	.119	.259	.089	.791	.804	.667
20	.058	.141	.042	.855	.874	.758
49	.024	.063	.018	.935	.946	.888
∞	0	0	0	1	1	1

For the examples of Section 4 the asymptotic efficiencies of the estimators of equations (19) and (22) are given in Table 2. Note from the table that if $p_2 - p_1$ is small the estimators $\bar{p}_1$, $\bar{p}_2$ are considerably more efficient unless n is quite large, while for larger difference between the parameters the primed estimators are always more efficient. This indicates that a "better" procedure would be to use $\bar{p}_1$, $\bar{p}_2$ if $2 \leqq n < k$, say, and $\hat{p}_1$, $\hat{p}_2$ if $n > k$, where k is some function of p_1, p_2, α (or $p_2 - p_1$ and α) determined in such a way that the JARE of the combined estimator is as high as possible.

Another possibility would be to construct estimators for p_1 and p_2 independently by maximizing individual asymptotic efficiencies. This problem has not been investigated but is evidently related to the problem of determining minimum efficiency discussed in the preceding section.

6. Acknowledgments. The author is especially indebted to Professor D. S. Robson for his guidance and encouragement during the course of this investigation, and to Professors William Kruskal and Wassily Hoeffding and the reviewer for their many helpful comments. In particular, the author is grateful to Professor Hoeffding for the reference to the work of Okamoto and the subsequent proof of efficiency given in Section 4. The author's original proof was considerably longer. Thanks are also due Professor R. L. Anderson for his comments on the lack of efficiency when α is known.

REFERENCES

[1] CRAMÉR, HAROLD (1946). *Mathematical Methods of Statistics*, Princeton Univ. Press.

[2] HOEFFDING, WASSILY and ROBBINS, HERBERT (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15** 773–780.

[3] KENDALL, MAURICE G. and STUART, ALAN (1958). *The Advanced Theory of Statistics.* **1** Hafner Pub., New York.

[4] OKAMOTO, MASASHI (1958). Some inequalities relating to the partial sum of binomial probabilities. *Ann. Inst. Statist. Math.* **10** 29–35.

[5] PEARSON, KARL (1894). Contributions to the mathematical theory of evolution. *Phil. Trans. Roy. Soc.* **185A** 71–110.

[6] RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* Wiley, New York.

[7] RIDER, PAUL R. (1961). The method of moments applied to a mixture of two exponential distributions. *Ann. Math. Statist.* **32** 143–147.

[8] RIDER, PAUL R. (1961). Estimating the parameters of mixed Poisson, binomial, and Weibull distributions by the method of moments. *Bull. Inst. Internat. Stat.* **39** (Part 2).

[9] ROBBINS, HERBERT (1948). Mixtures of distributions. *Ann. Math. Statist.* **19** 360–369.

[10] TEICHER, HENRY (1960). On the mixture of distributions. *Ann. Math. Statist.* **31** 55–72.

[11] TEICHER, HENRY (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.