ON THE ESTIMATION OF MIXING DISTRIBUTIONS

By D. C. Boes

Colorado State University

1. Introduction and summary. Let $\mathfrak{F} = \{F(x; y), y \varepsilon E\}$ be a family of cumulative distribution functions (cdf's) in the variable x indexed by $y \varepsilon E$, where E is a measurable subset of the real line. Assume that $F(x; y)$ is measurable in y for all x. Then, for any nondegenerate cdf G, whose induced probability measure assigns measure one to E, the cdf $H_G(x) = \int_E F(x; y)\, dG(y)$ is called a "mixture" of $\mathfrak{F}$, and G is called the "mixing distribution." The family $\mathfrak{F}$ will be called the kernel of the mixture.

Among the many interesting problems engendered by mixtures of distributions is one of great practical importance, namely, the estimation of the mixing distribution. This problem is simply stated—on the basis of observations from the mixture H_G, estimate the mixing distribution G. However, before actual estimation can be meaningfully investigated, identifiability of the family of mixtures, defined presently, must be verified. Let $\mathcal{G}$ denote the class of mixing distributions G, and $\mathfrak{H}$ the induced class of mixtures H for some specified family $\mathfrak{F}$. A mixture of $\mathfrak{F}$ will be called "identifiable" if $H(x) = \int F(x; y)\, dG^*(y) = \int F(x;y)\, dG(y)$ implies $G^* = G$. If every member H of $\mathfrak{H}$ is identifiable, then $\mathfrak{H}$ will be said to be identifiable. For identifiable families $\mathfrak{H}$, the problem of estimation of the mixing cdf when the elements of $\mathfrak{F}$ occurring in the mixture are known is dealt with here. In particular, the problem of unbiased estimation for finite mixtures is considered. In a finite mixture the kernel is any finite set of known but arbitrary cdf's and the mixing distributions are discrete assigning positive weight to each of the cdf's in the kernel.

The estimation of the mixing ratio when two arbitrary known cdf's are mixed is discussed at length in Section 2. Identifiability is then evident. Necessary and sufficient conditions on two distribution mixtures for uniform attainment of the Cramér-Rao lower bound are derived. The class of θ-efficient estimators is found; also, as it turns out, the minimax unbiased estimator is a member of this θ-efficient class and it is characterized. In Section 3, the results of Section 2 are extended to any finite mixture. For identifiable finite mixtures, necessary and sufficient conditions for the existence of an estimator which uniformly attains the minimal ellipsoid of concentration are given. The θ-efficient family of estimators is derived; also, estimators within the θ-efficient family which are consistent asymptotically normal efficient are characterized.

2. On the estimation of the mixing proportion in a mixture of two distributions.

2.1. *Introduction.* Let

$$\mathfrak{H} = \{H_\theta(x) : -\tfrac{1}{2} < \theta < \tfrac{1}{2}, H_\theta(x) = (\tfrac{1}{2} + \theta)F(x) + (\tfrac{1}{2} - \theta)G(x)\}$$

Received 6 January 1965; revised 1 July 1965.

denote the family of mixtures of any two fixed (distinct) one-dimensional cdf's $F(x)$ and $G(x)$. Here, the mixing distribution is discrete, assigning mass $(\frac{1}{2} + \theta)$ to the cdf $F(x)$ and mass $(\frac{1}{2} - \theta)$ to the cdf $G(x)$. Let μ be any σ-finite measure on the Borel sets of the real line which dominates both F and G, and hence also $H_\theta(x)$. Then, by the Radon-Nikodym theorem, there exist densities $h_\theta(x)$ such that $P_\theta\{A\} = \int_A h_\theta(x)\, d\mu(x)$ for any Borel set A of the real line, where P_θ is the probability measure corresponding to the cdf $H_\theta(x)$. Let $h_\theta(\vec{x})$ and $H_\theta(\vec{x})$ denote respectively the joint density and joint cumulative distribution functions of $X_1, \cdots, X_n$, where $X_1, \cdots, X_n$ are independent identically distributed random variables with common cdf H_θ.

For a given random sample $x_1, \cdots, x_n$ from a population with distribution H_θ an estimate of the parameter θ is desired. Here identifiability, of the family $\mathcal{H}$ is evident, for if $(\frac{1}{2} + \theta_1)F(x) + (\frac{1}{2} - \theta_1)G(x) \equiv (\frac{1}{2} + \theta_2)F(x) + (\frac{1}{2} - \theta_2)G(x)$, then $(\theta_1 - \theta_2)[F(x) - G(x)] \equiv 0$ and since $F(x) \not\equiv G(x)$, we have $\theta_1 = \theta_2$.

The estimation problem confronted here is unusual by virtue of the nature of the *a priori* information about the unknown distribution $H_\theta(x)$ and the fact that the parameter enters linearly. This latter fact prompts one to solve for the parameter θ in the equation $H_\theta(B) = (\frac{1}{2} + \theta)F(B) + (\frac{1}{2} - \theta)G(B)$ and form a host of estimators $\hat{\theta}_n(B) = \{\hat{H}_\theta(B) - \frac{1}{2}[F(B) + G(B)]\}/[F(B) - G(B)]$ simply by replacing $H_\theta(B)$ by an estimator $\hat{H}_\theta(B)$. [The notation $F(A) = P_F\{A\}$ will be adopted and adhered to hereafter, where F is the cdf corresponding to the probability measure P_F and A is any Borel set. All cdf's are taken to be left continuous.] $H_n(B)$, defined to be the proportion of observations $X_1 \cdots X_n$ that fall in B, is an unbiased estimator of $H_\theta(B)$ and provides us with the family of estimators

$$\hat{\theta}_n(B) = \{H_n(B) - \tfrac{1}{2}[F(B) + G(B)]\}/[F(B) - G(B)] \tag{1}$$

for any Borel set B such that $F(B) \neq G(B)$. In particular, if B is restricted to the infinite closed rays $(-\infty, x]$, (1) becomes

$$\hat{\theta}_n(x) = \{H_n(x) - \tfrac{1}{2}[F(x) + G(x)]\}/[F(x) - G(x)] \tag{2}$$

for any x such that $F(x) \neq G(x)$; here, $H_n(x)$ is the sample cdf (taken as left continuous). The family (1) provides a myriad of estimators, each of which is unbiased, converges with probability one to the true parameter, and has variance of order $O(1/n)$, the exact value being given by

$$\operatorname{Var} \hat{\theta}_n(B) = \{(\tfrac{1}{2} + \theta)[F(B) - F^2(B)] + (\tfrac{1}{2} - \theta)[G(B) - G^2(B)]\}/ \\ n[F(B) - G(B)]^2 + [(\tfrac{1}{2} + \theta)(\tfrac{1}{2} - \theta)/n].$$

The estimators defined in (1) can take values outside the interval $(-\frac{1}{2}, \frac{1}{2})$. This drawback could be rectified by truncating the estimator, but then unbiasedness would be lost. Admittedly, the desirability of the property of unbiasedness is questionable, however, for the most part, unbiased estimation is considered herein.

For any class of sets $\mathcal{A}$, define A^* such that

$$|F(A^*) - G(A^*)| = \sup_{A \varepsilon \mathcal{A}} |F(A) - G(A)|, \qquad \text{then}$$

$$\sup_{A \varepsilon \mathcal{A}} |H_{\hat{\theta}_n(A^*)}(A) - H_\theta(A)| \leqq \sup_{A \varepsilon \mathcal{A}} |H_n(A) - H_\theta(A)| ;$$

i.e., $H_{\hat{\theta}_n(A^*)}(A)$ is a better estimator of H_θ than is $H_n(A)$ from the standpoint of the implicit definition of distance. It should definitely be possible to improve on the sample cdf in estimating the distribution $H_\theta(x)$ since the former in no way takes cognizance of the fact that $H_\theta(x)$ is a two distribution mixture. The family (1) will play an important role in the subsequent investigation.

Bayes estimators encountered here possess a somewhat unusual characteristic if the risk function is defined as the square error; that is, a Bayes estimator of sample size n depends on the first $n + 1$ moments of the *a priori* distribution and not on the *a priori* distribution itself. This singularity follows immediately from the fact that the parameter enters the distribution function linearly.

2.2. *Necessary and sufficient conditions for the attainment of the Cramér-Rao lower bound.* Let $\{h_\theta(\vec{x}), \theta \varepsilon \Omega\}$ be a one-parameter family of probability density functions on R^n with $P_\theta\{\vec{X} \varepsilon A\} = \int_A h_\theta(\vec{x})\, d\mu(\vec{x})$ for any Borel set A of R^n, where μ is a σ-finite measure independent of θ, and Ω is any open set of real numbers. Let $d(\vec{x})$ be an estimator of θ, that is, a measurable mapping of R^n into R^1. Conditions under which the Cramér-Rao bound is a valid lower bound for the variance of an estimator $d(\vec{x})$ of a parameter θ are well known. Consider the case where:

(a) Ω is the entire real line or an open interval of the real line.

(b) $\log h_\theta(\vec{x})$ is an absolutely continuous function of θ for almost all x.

(c) Both $\int h_\theta(\vec{x})\, d\mu(\vec{x})$ and $\int d(\vec{x}) h_\theta(\vec{x})\, d\mu(\vec{x})$ are differentiable with respect to θ under integral sign.

(d) $0 < E_\theta[(\partial/\partial\theta) \log h_\theta(\vec{X})]^2 < \infty$.

If E_θ denotes expectation with respect to P_θ, the preceding conditions are ample for the validity of

$$(3) \qquad \begin{aligned} \sigma_d^2(\theta) &\geqq [E_\theta\, d(\vec{X})(\partial/\partial\theta) \log h_\theta(\vec{X})]^2 / E_\theta[(\partial/\partial\theta) \log h_\theta(\vec{X})]^2 \\ &= [(\partial/\partial\theta) E_\theta\, d(\vec{X})]^2 / E_\theta[(\partial/\partial\theta) \log h_\theta(\vec{X})]^2. \end{aligned}$$

PROPOSITION 1. *If Conditions* (a) *through* (d) *are satisfied for* $h_\theta(\vec{x})$ *and* $d(\vec{x})$ *and if* $\sigma_d^2 > 0$ *for all* $\theta \varepsilon \Omega$, *then a necessary and sufficient condition that* $\sigma_d^2(\theta)$ *coincides with the Cramér-Rao lower bound is that* $h_\theta(\vec{x}) = A(\vec{x}) \exp \cdot [d(\vec{x}) b(\theta) + c(\theta)]$ *where* $b'(\theta) \not\equiv 0$ *and* $A(\vec{x}) = 0$ *if and only if* $h_\theta(\vec{x}) = 0$.

PROOF. Fend [4].

REMARK. Consider the case where $h_\theta(\vec{x}) = \prod_{i=1}^n h_\theta(x_i)$ and $\mu(\vec{x})$ is a product measure on R^n each component of which is μ (a σ-finite measure on R^1). Then the joint density $h_\theta(\vec{x})$ is of Darmois-Koopman form if and only if the one-dimensional density $h_\theta(x)$ is of Darmois-Koopman form, i.e.,

$$(4) \qquad h_\theta(x) = V(x) \exp [t(x) s(\theta) + u(\theta)].$$

In the problem to be investigated (see Section 2.1) $\mu(\vec{x})$ is as stipulated in the above remark and

$$(5) \qquad h_\theta(\vec{x}) = \textstyle\prod_1^n [(\frac{1}{2} + \theta)f(x_i) + (\frac{1}{2} - \theta)g(x_i)].$$

It suffices to confine attention to $S_H^+ = \{x: h_\theta(x) > 0\} = \{x: f(x) > 0$ or $g(x) > 0\}$. Let $d(\vec{x})$ be a Borel measurable mapping from R^n into R^1, and

$$(6) \qquad \mathfrak{D} = \{d(\vec{x}): \sigma_d^2(\theta) > 0 \text{ for all } \theta \in (-\tfrac{1}{2}, \tfrac{1}{2})\}.$$

It can be shown that the Regularity Conditions (a) through (d) are satisfied for $\Omega = (-\frac{1}{2}, \frac{1}{2})$, h_θ as in (5) and $d \in \mathfrak{D}$.

LEMMA 1. *If h_θ is as in (5) and the Cramér-Rao bound is uniformly attained by the variance of some $d(\vec{x}) \in \mathfrak{D}$, then there exists a decomposition of S_H^+ into sets Λ_1, Λ_2 of positive μ-measure such that $t(x) = \begin{cases} t_1 \text{ on } \Lambda_1 \\ t_2 \text{ on } \Lambda_2 \end{cases}$. Here, $t(x)$ is as in (4).*

PROOF. By Proposition 1 and the remark following it, $h_\theta(x)$ has the Darmois-Koopman form, i.e.,

$$\begin{aligned} h_\theta(x) &= \tfrac{1}{2}[f(x) + g(x)] + \theta[f(x) - g(x)] \\ &\equiv V(x) \exp[t(x)s(\theta) + u(\theta)]. \end{aligned}$$

Differentiating the above twice with respect to θ gives: $-[u'(\theta)]^2 - u''(\theta) \equiv [t(x)s'(\theta)]^2 + 2t(x)s'(\theta)u'(\theta) + t(x)s''(\theta)$. Since $\sigma_d^2 > 0$, t takes at least two unequal values, say t_1 and t_2. Let t_3 also be a value of $t(x)$; then from the preceding $-[u'(\theta)]^2 - u''(\theta) \equiv [t_is'(\theta)]^2 + 2t_is'(\theta)u'(\theta) + t_is''(\theta)$ for $i = 1, 2, 3$; hence $[t_1s'(\theta)]^2 + 2t_1s'(\theta)u'(\theta) + t_1s''(\theta) \equiv [t_2s'(\theta)]^2 + 2t_2s'(\theta)u'(\theta) + t_2s''(\theta)$; or, $(t_1^2 - t_2^2)[s'(\theta)]^2 \equiv -(t_1 - t_2)2s'(\theta)\, u'(\theta) - (t_1 - t_2)s''(\theta)$, i.e, $(t_1 + t_2)(s')^2 \equiv -2s'u' - s''$. Now either $t_3 = t_1$ or in similar fashion, $(t_1 + t_3)(s')^2 \equiv -2s'u' - s''$, and since $s'(\theta) \not\equiv 0$, $t_1 + t_2 = t_1 + t_3$ or $t_3 = t_2$. Thus, $\Lambda_i = \{x: t(x) = t_i\}$, has positive μ-measure, $i = 1, 2$, and $\Lambda_1 \cup \Lambda_2 = S_H^+$.

THEOREM 1. *If the variance of some $d \in \mathfrak{D}$ coincides with the Cramér-Rao bound, there exists a decomposition of S_H^+ into sets Λ_1, Λ_2 of positive μ-measure such that*

$$(7) \qquad h_\theta(x) = V(x) \cdot K_i(\theta) \qquad \textit{for} \quad x \in \Lambda_i,\ i = 1, 2.$$

An equivalent condition is

$$(8) \qquad f(x)/g(x) \textit{ is constant (possibly } \infty) \textit{ on } \Lambda_i, \qquad i = 1, 2.$$

Conversely, if (7) or (8) holds, the Cramér-Rao bound is uniformly attained by the variance of the estimator $d(\vec{x}) = nt_1H_n(\Lambda_1) + nt_2H_n(\Lambda_2)$ for any two real unequal numbers t_1 and t_2. Recall that $nH_n(\Lambda_i)$ = the number of $x_1, \cdots, x_n$ which fall in Λ_i.

PROOF. If the variance of some $d \in \mathfrak{D}$ coincides with the Cramér-Rao bound, $h_\theta(x)$ has the Darmois-Koopman form, i.e., $h_\theta(x) = V(x) \exp[t(x)s(\theta) + u(\theta)]$. Moreover, by Lemma 1, $t(x) = \begin{cases} t_1 \text{ on } \Lambda_1 \\ t_2 \text{ on } \Lambda_2 \end{cases}$; therefore,

$$h_\theta(x) = V(x) \exp [t(x)s(\theta) + u(\theta)] = V(x) \exp [t_1 s(\theta) + u(\theta)] \text{ on } \Lambda_1$$
$$= V(x) \exp [t_2 s(\theta) + u(\theta)] \text{ on } \Lambda_2 .$$

Conversely, since $x \varepsilon S_H{}^+$, necessarily $K_j(\theta) > 0, j = 1, 2$. Thus, for any two real unequal numbers t_1, t_2, we may define

$$u(\theta) = [t_2 \log K_1(\theta) - t_1 \log K_2(\theta)]/(t_2 - t_1);$$
$$s(\theta) = [\log K_1(\theta) - \log K_2(\theta)]/(t_1 - t_2).$$

Then $h_\theta(x)$ has the Darmois-Koopman form, $V(x) \exp [t(x)s(\theta) + u(\theta)]$, with $t(x) = \begin{cases} t_1 \text{ on } \Lambda_1 \\ t_2 \text{ on } \Lambda_2 \end{cases}$. Hence by Proposition 1, $\sigma_d^2(\vec{x})(\theta)$ coincides with the Cramér Rao lower bound where $d(\vec{x}) = \sum_{i=1}^n t(x_i) = nt_1 H_n(\Lambda_1) + nt_2 H_n(\Lambda_2)$. The asserted equivalence is easily verified.

Corollary. *If $h_\theta(x)$ satisfies the factorization criterion of Theorem* 1 *and* Λ_1 *is as in Theorem* 1, *then*

$$\hat{\theta}_n(\Lambda_1) = \{H_n(\Lambda_1) - \tfrac{1}{2}[F(\Lambda_1) + G(\Lambda_1)]\}/[F(\Lambda_1) - G(\Lambda_1)]$$

is the uniformly minimum variance unbiased estimator of θ. Note that $\hat{\theta}_n(\Lambda_1)$ *belongs to family* (1). *Also, by the strong law of large numbers* (S.L.L.N.), $\hat{\theta}_n(\Lambda_1)$ *converges almost certainly to the true parameter.*

2.3 *θ-efficient class.* In the preceding section, necessary and sufficient conditions on the distributions F and G for the uniform attainment of the Cramér-Rao lower bound were derived. Achievement of this requirement led to a narrow circumscription of the distributions (F and G) being mixed. Consequently, in the remaining but preponderant cases, a less stringent property will be adopted, namely, that of θ-efficiency.

Let $\mathfrak{J}$ constitute a class of estimators of θ.

Definition. An estimator, $t' \varepsilon \mathfrak{J}$, is called θ^o-efficient in $\mathfrak{J}$ if $\sigma_{t'}^2(\theta^o) \leqq \sigma_t^2(\theta^0)$ for every $t \varepsilon \mathfrak{J}$.

θ^o-efficient unbiased estimators have been discussed in the literature under general background conditions; for example, see Barankin [2], Stein [6], and Bahadur [1]. In particular, the existence and uniqueness of θ-efficient unbiased estimators has been proved.

Let $\mathfrak{U}$ denote the class of unbiased estimators of θ, and let

$$I(\theta) = E_\theta[(\partial/\partial\theta) \log h_\theta(X)]^2.$$

The following proposition follows immediately from standard conditions for efficiency.

Proposition 2. *Let $X_1, \cdots, X_n$ be independent identically distributed random variables with common* cdf $H_\theta(x)$. *For any $\theta^o \varepsilon (-\frac{1}{2}, \frac{1}{2})$ the estimator of θ,*

$$\hat{\theta}_n(\theta^o) = \theta^o + [1/nI(\theta^o)] \sum_{i=1}^n [f(X_i) - g(X_i)]/h_{\theta^o}(X_i) \tag{9}$$

is θ^o-efficient in $\mathfrak{U}$.

REMARK. Each estimator $\hat{\theta}_n(\theta^o)$, $-\frac{1}{2} < \theta^o < \frac{1}{2}$, is admissible (where the risk function is taken to be the variance) within the class of unbiased estimators.

REMARK. $\hat{\theta}_n(\theta^o)$ converges almost certainly to θ.

PROOF. A necessary and sufficient condition for independent identically distributed random variables to obey the S.L.L.N. is that $E|X| < \infty$. But $E_\theta|[f(X) - g(X)]/h_{\theta^o}(X)| < \infty$ since

$$|f(x) - g(x)|/h_\theta(x) \leqq \{1/\min [\tfrac{1}{2} + \theta, \tfrac{1}{2} - \theta]\}$$

hence the S.L.L.N. is applicable.

LEMMA 2. *The Cramér-Rao lower bound, $[nI(\theta)]^{-1}$, is a concave infinitely differentiable function of θ.*

PROOF. Infinite differentiability follows from the dominated convergence theorem. Now, n times the Cramér-Rao lower bound $= 1/I(\theta) = 1/\int [(f - g)^2/h_\theta]$ and

$$(10) \qquad (d/d\theta)(1/I(\theta)) = -I'(\theta)/I^2(\theta) = \int [(f - g)^3/h_\theta^2]/(\int (f - g)^2/h_\theta)^2.$$

$$(11) \qquad (d^2/d\theta^2)(1/I(\theta)) = \{-2 \int [(f - g)^2/h_\theta] \int [(f - g)^4/h_\theta^3] + 2(\int (f - g)^3/h_\theta^2)^2\}/(\int (f - g)^2/h_\theta)^3.$$

Now

$$(\int ((f - g)/h)^3\, dH)^2 = (\int ((f - g)/h)((f - g)/h)^2\, dH)^2 \leqq \int ((f - g)/h)^2\, dH \cdot \int ((f - g)/h)^4\, dH$$

by Schwarz's inequality, i.e., the numerator of (11) is negative, but the denominator is positive, so $(d^2/d\theta^2)(1/I(\theta)] < 0$, i.e., the Cramér-Rao lower bound is a concave function.

LEMMA 3. *$\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta)$ is concave quadratic function of θ.*

PROOF. $\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta) = n^{-1}[-(\theta - \theta^o)^2 + [1/I^2(\theta^o)] \int ((f - g)/h_{\theta^o})^2 h_\theta]$.

The point at which the maximum of $\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta)$ occurs is readily calculated to be

$$(12) \qquad \theta_{\max} = \theta^o + [1/2I^2(\theta^o)] \int [(f - g)^3/h^2_{\theta_o}].$$

THEOREM 2. *If $\bar{\theta}^o$ is the unique value for which the maximum of the Cramér-Rao bound in $[-\frac{1}{2}, \frac{1}{2}]$ is attained, the estimator $\hat{\theta}_n(\bar{\theta}^o)$ is minimax in $\mathfrak{U}$.*

PROOF. Suppose first that the Cramér-Rao lower bound has a maximum within $(-\frac{1}{2}, \frac{1}{2})$. Then $\bar{\theta}^o$ is such that (10) is equal to zero. Taking θ^o in (12) to be $\bar{\theta}^o$, we see that $\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta)$ also attains its maximum at $\bar{\theta}^o$, so that $\hat{\theta}_n(\bar{\theta}^o)$ is minimax. On the other hand, if the maximum of Cramér-Rao lower bound in $[-\frac{1}{2}, \frac{1}{2}]$ occurs at one of the end points, say $\bar{\theta}^o = \frac{1}{2}$, then the numerator of the right hand side of (10) is greater than zero for every θ in $(-\frac{1}{2}, \frac{1}{2})$, and hence by (12) $\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta)$ attains its maximum for some value to the right of θ^o for all θ^o; in particular, this is true for $\theta^o = \frac{1}{2}$. Hence $\max_{|\theta| \leqq \frac{1}{2}} (\sigma^2_{\hat{\theta}_n(\frac{1}{2})}(\theta))$ is attained at $\theta = \frac{1}{2}$. Similarly for $\bar{\theta}^o = -\frac{1}{2}$.

The following theorems show that the estimators (9), though seemingly unrelated to the family (1) of estimators, are actually, in some instances, averages of these. Before stating the theorems a useful lemma is stated.

LEMMA 4. *For any two measurable sets A and B,* Cov $[H_n(A), H_n(B)] = [H(A \cdot B) - H(A)H(B)]/n$, *where H_n is the sample* cdf *from a population with distribution H. In particular,* Cov $[H_n(A), H_n(B)] = [-H(A)H(B)]/n$ *if A and B are disjoint and* Var $[H_n(A)] = \{H(A)[1 - H(A)]\}/n$.

THEOREM 3. *Suppose the positive spectrum of H_θ, S_H^+, admits a countable (possibly finite) decomposition into sets Λ_i, $i = 1, 2, \cdots$, such that*[1] $f(x)/g(x) = \lambda_i \geqq 0$, on Λ_i, *where $\lambda_i \neq \lambda_j$ for $i \neq j$. Let $f_i = F(\Lambda_i)$, $g_i = G(\Lambda_i)$, $h_i = H_{\theta^o}(\Lambda_i)$, and $F_i = F(\bigcup_{j=1}^i \Lambda_j) = \sum_{j=1}^i f_j$, $G_i = \sum_{j=1}^i g_j$, $H_i = \sum_{j=1}^i h_j$, $i = 1, 2, \cdots$. Further, let $d_i = (f_i - g_i)/h_i - (f_{i+1} - g_{i+1})/h_{i+1}$, $\beta_i = [1/I(\theta^o)](F_i - G_i)\, d_i$, (so that $\sum \beta_j = 1$) and put $D_i = [H_n(\bigcup_{j=1}^i \Lambda_j) - \frac{1}{2}(F_i + G_i)]/(F_i - G_i)$ if $F_i - G_i \neq 0$, $D_i = 0$ if $F_i = G_i$, $i = 1, 2, \cdots$. Then $\hat{\theta}_n(\theta^o) = \sum_{j=1}^\infty \beta_j D_j$* is *$\theta^o$-efficient in $\mathfrak{U}$.*

PROOF. $\sum_1^\infty \beta_j = [1/I(\theta^o)]\sum_1^\infty (F_i - G_i)\, d_i = [1/I(\theta^o)]\sum_1^\infty [(f_i - g_i)^2/h_i]$; however $I(\theta^o) = \int [(f - g)^2/h_{\theta^o}] = \sum_{i=1}^\infty \int_{\Lambda_i} (f - g)^2/h_{\theta^o} = \sum_{i=1}^\infty (f_i - g_i)^2 h_i$, hence $\sum_{j=1}^\infty \beta_j = 1$. It remains to prove that $\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta^o) = [1/nI(\theta^o)]$. $\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta^o) = \sum_{j=1}^\infty \beta_j \sum_{j=1}^\infty \beta_i$ Cov $[D_i, D_j]$. By employing Lemma 4 and noting $\bigcup_{\alpha=1}^j \Lambda_\alpha \subset \bigcup_{\alpha=1}^i \Lambda_\alpha$ if $j < i$ and $\bigcup_{\alpha=1}^j \Lambda_\alpha \supset \bigcup_{\alpha=1}^i \Lambda_\alpha$ if $i \leqq j$, we obtain

$$\begin{aligned}\sum_{i=1}^\infty \beta_i \operatorname{Cov}[D_i, D_j] &= \sum_{i=1}^j \beta_i[H_i(1 - H_j)/n(F_i - G_i)(F_j - G_j)] \\ &\quad + \sum_{i=j+1}^\infty \beta_i[H_j(1 - H_i)/n(F_i - G_i)(F_j - G_j)] \\ &= [1/nI(\theta^o)(F_j - G_j)] \\ &\quad \cdot [F_j - G_j - H_j((f_1 - g_1)/h_1) + H_j(\textstyle\sum_1^\infty (1 - H_i)\, d_i)].\end{aligned}$$

However, $\sum_1^m (1 - H_i)\, d_i = (f_1 - g_1)/h_1 - [(f_{m+1} - g_{m+1})/h_{m+1}](1 - \sum_{i=1}^m h_i) - \sum_{i=1}^m (f_i - g_i)$, hence $(f_1 - g_1)/h_1 = \sum_1^\infty (1 - H_i)\, d_i$; whence

$$\sum_{i=1}^\infty \beta_i \operatorname{Cov}[D_i, D_j] = 1/nI(\theta^o)$$

and so $\sum_{j=1}^\infty \beta_j \sum_{i=1}^\infty \beta_i$ Cov $[D_i, D_j] = [1/nI(\theta^o)]\sum_1^\infty \beta_j = 1/nI(\theta^o)$.

COROLLARY 1. *If F and G are both discrete distributions, then the θ^o-efficient estimator of θ is a linear combination of estimators belonging to Family* (1) .

COROLLARY 2. *Under the assumptions of Theorem* 3, *if the Λ_j's can be indexed so that* $0 \leqq \lambda_1 < \lambda_2 < \cdots < \lambda_j < \cdots$, *(in particular, whenever the decomposition of S_H^+ is finite) then the β_j's form a discrete probability distribution.*

PROOF. It suffices to prove $\beta_j \geqq 0$, $j = 1, 2, \cdots$. Now $\beta_j \geqq 0$ if and only if $(F_j - G_j)\, d_j \geqq 0$. Also, $d_j < 0$ is readily proved to be equivalent to $\lambda_j < \lambda_{j+1}$. Suppose $F_j - G_j = C > 0$ for some j, which requires that $\lambda_{j+1} > 1$. Then $F_{j+1} - G_{j+1} = f_{j+1} + F_j - g_{j+1} - G_j = g_{j+1}(\lambda_{j+1} - 1) + F_j - G_j \geqq F_j - G_j = C$. Proceeding inductively $F_j - G_j > 0$ implies $F_i - G_i \geqq C$ for every $i > j$ which contradicts $\lim_i (F_i - G_i) = 0$. Thus $F_j - G_j \leqq 0$ every j, and combining this with $d_j < 0$ for every j, we have $\beta_j \geqq 0$, $j = 1, 2, \cdots$.

[1] If $S_F^+ - S_G^+ \neq \emptyset$, then $S_F^+ - S_G^+ = \Lambda_j$ for some j, and $\lambda_j = \infty$ for that j; also, if $S_G^+ - S_F^+ \neq \emptyset$, then $S_G^+ - S_F^+ = \Lambda_i$ for some i, and $\lambda_i = 0$ for that i.

THEOREM 4. *If F and G are absolutely continuous with respect to Lebesgue measure, their densities f and g are almost everywhere differentiable, and g/f is monotone nondecreasing, then* $\hat{\theta}_n(\theta^o) = \int \hat{\theta}_n(x)w(x)\,dx$ *is the θ^o-efficient unbiased estimator of θ, where* $\hat{\theta}_n(x) = \{H_n(x) - \frac{1}{2}[F(x) + G(x)]\}/[F(x) - G(x)]$ *as before, (see Equation 2) and*

$$w(x) = \{[F(x) - G(x)]/I(\theta^o)\} \cdot \{[f(x) - g(x)]h'_{\theta^o}(x) - [f'(x) - g'(x)]h_{\theta^o}(x)\}/h^2_{\theta^o}(x)$$

and $\int w(x)\,dx = 1$ *and* $w(x) \geqq 0$, *i.e.,* $w(x)$ *is a probability density function.*

PROOF. $\int w(x)\,dx = [1/I(\theta^o)] \int (F - G)\, d((g - f)/h_{\theta_o}) = [1/I(\theta^o)] \cdot [F(x) - G(x)][(f(x) - g(x))/h_{\theta^o}(x)] \,|_{-\infty}^{\infty} + [1/I(\theta^o)] \int (f - g)^2/h_{\theta^o} = 1$. (The 2nd equality is obtained by an integration by parts.) Clearly $w(x) \geqq 0$ if and only if $(F - G)[(f - g)h' - (f' - g')h] \geqq 0$. But g/f nondecreasing implies that $(f - g)h' - (f' - g')h \geqq 0$, and that $F(x) - G(x) \geqq 0$ for every x. Hence $w(x)$ is a probability density function. Next, using Fubini's theorem,

$$E_\theta \int \hat{\theta}_n(x)w(x)\,dx = \int \{[H_\theta(x) - \tfrac{1}{2}[F(x) + G(x)]]/[F(x) - G(x)]\}w(x)\,dx$$
$$= \int \theta w(x)\,dx = \theta,$$

i.e., $\hat{\theta}_n(\theta^o)$ is unbiased. It remains to prove that $\hat{\theta}_n(\theta^o)$ is θ^o-efficient. $\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta^o) = = E_{\theta^o}(\hat{\theta}_n(\theta^o) - \theta^o)^2 = \int\int \{\mathrm{Cov}\,[H_n(x), H_n(y)]/\{F(x) - G(x)][F(y) - G(y)]\} w(x)w(y)\,dx\,dy$ by Fubini's theorem. Also, utilizing the preceding lemma and integrating by parts, we obtain

$$\int \{\mathrm{Cov}\,[H_n(x), H_n(y)]/[F(x) - G(x)]\}w(x)\,dx = [F(y) - G(y)/nI(\theta^o)],$$

hence $\sigma^2_{\hat{\theta}_n(\theta^o)}(\theta^o) = 1/nI(\theta^o)$.

3. On the estimation of the mixing distribution in a mixture of a finite number of distributions.

3.1 *Introduction.* Let

$$(13) \quad \mathcal{H} = \{H_\theta(x) : H_\theta(x) = \textstyle\sum_{i=1}^{k+1} \theta_i F_i(x);$$
$$\theta_i > 0,\ i = 1, \cdots k + 1;\ k \geqq 2;\ \textstyle\sum_{i=1}^{k+1} \theta_i = 1\}$$

denote the family of mixtures of any fixed set of $k + 1$ (distinct) cdf's, $F_1, \cdots, F_{k+1}$. Here, the mixing distribution is discrete assigning mass θ_i to distribution F_i, $i = 1, \cdots, k + 1$. Let μ by any σ-finite measure which dominates F_i, $i = 1, \cdots, k + 1$. By the Radon-Nikodym theorem, there exist densities $h_\theta(x) = \sum_{i=1}^{k+1} \theta_i f_i(x)$. As before, let $h_\theta(\vec{x})$ and $H_\theta(\vec{x})$ denote the joint density and joint cumulative distribution functions respectively of $X_1, \cdots, X_n$, where $X_1, \cdots, X_n$ are independent identically distributed random variables with common cdf H_θ.

In contradistinction to a mixture of only two distributions, identifiability of the family $\mathcal{H}$ no longer holds in general. Teicher [8] proved that a necessary and sufficient condition that $\mathcal{H}$ in (13) be identifiable is that there exist $k + 1$

real values $x_1, \cdots, x_{k+1}$ such that the determinant $|F_i(x_j)| \neq 0$, $i, j = 1, \cdots, k + 1$.

In the following, it will always be assumed that the family of mixtures at hand is identifiable, otherwise the problem of estimation is not meaningful.

Since $\theta_1, \theta_2, \cdots, \theta_{k+1}$ are linearly dependent, the parameter space can be taken as the open set $\Theta = \{\theta: \theta_i > 0,\ i = 1, \cdots, k,\ \sum_1^k \theta_i < 1\}$ of k-dimensional Euclidean space.

3.2. *Necessary and sufficient conditions for the uniform attainment of the minimum ellipsoid of concentration.* Let $\{h_\theta(\vec{x}),\ \theta \in \Omega\}$ be a k-parameter family of probability density functions defined on R^n with $P_\theta\{\vec{X} \in B\} = \int_B h_\theta(\vec{x})\, d(\vec{x})$ for any Borel set B of R^n, where μ is a σ-finite measure independent of θ, and Ω is any k-dimensional interval of Euclidean k-space. Denote by $\mathfrak{U}$ the class of all estimators, $\hat{\theta}(\vec{x}) = (\hat{\theta}_1(\vec{x}), \cdots, \hat{\theta}_k(\vec{x}))$, of $\theta = (\theta_1, \cdots, \theta_k)$ for which the $\hat{\theta}_i(\vec{x})$'s are linearly independent with positive probability for all $\theta \in \Omega$ and for which $E_\theta(\hat{\theta}_i(\vec{X})) = \theta_i$, $i = 1, \cdots, k$. Equivalently, $\mathfrak{U}$ is the class of all unbiased estimators of θ whose variance-covariance matrices are positive definite for all θ in Ω. The ellipsoid of concentration of any estimator $\hat{\theta}(\vec{x})$ of $\mathfrak{U}$ is defined as the interior and boundary of $\sum_{i,j=1}^k \sigma^{ij}(\theta)(u_j - \theta_j)(u_i - \theta_i) = k + 2$, where $(\sigma^{ij}(\theta))^{-1} = (\sigma_{ij}(\theta)) = (E_\theta(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j))$. Conditions under which there exists a minimal ellipsoid of concentration, i.e., a (largest) ellipsoid centered at θ and lying wholly within the ellipsoid of concentration of any unbiased estimator, have been stated in the literature. A recent short proof of a result of this kind appears in [6]. These conditions are k-dimensional analogues of the regularity conditions for the validity of the Cramér-Rao lower bound of Section 2.2. For example, consider the case where:

(a) Ω is an open interval of Euclidean k-space.

(b) $\hat{\theta}(\vec{x})$ belongs to $\mathfrak{U}$.

(c) $\log h_\theta(\vec{x})$ is an absolutely continuous function of θ_i for almost all x, $i = 1, \cdots, k$.

(d) $\int h_\theta(\vec{x})\, d\mu(\vec{x})$ and $\int \hat{\theta}_i(\vec{x}) h_\theta(\vec{x})\, d\mu(\vec{x})$, $i = 1, \cdots, k$, are differentiable with respect to θ_j under the integral sign, $j = 1, \cdots, k$.

(e) The matrix $(I_{ij}(\theta))$ is positive definite for all $\theta \in \Omega$, where

$$I_{ij}(\theta) = \int [(\partial/\partial\theta_j) \log h_\theta(\vec{x}) \cdot (\partial/\partial\theta_i) \log h_\theta(\vec{x})] h_\theta(\vec{x})\, d\mu(\vec{x}).$$

If the preceding conditions are satisfied, then the minimum ellipsoid of concentration is given by

$$\sum_{i,j=1}^k I_{ij}(\theta)(u_i - \theta_i)(u_j - \theta_j) = k + 2. \tag{14}$$

DEFINITION. An estimator $\hat{\theta}(\vec{x}) = (\hat{\theta}_1(\vec{x}), \cdots, \hat{\theta}_k(\vec{x}))$ will be called *jointly efficient* for $\theta = (\theta_1, \cdots, \theta_j)$ within $\mathfrak{U}$ if the ellipsoid of concentration of $\hat{\theta}(\vec{x})$, centered at θ, coincides with the minimum ellipsoid of concentration.

PROPOSITION 3. *If the preceding Conditions* (a) *through* (e) *are satisfied for* $h_\theta(\vec{x})$, *and* $\hat{\theta}(\vec{x}) \in \mathfrak{U}$, *then a necessary and sufficient condition that* $(\hat{\theta}_1(\vec{x}), \cdots, \hat{\theta}(\vec{x}))$ *be jointly efficient is that* $h_\theta(\vec{x}) = A(\vec{x}) \exp [\sum_i^k \hat{\theta}_i(\vec{x}) b_i(\theta) + c(\theta)]$ *for some*

$A(\bar{x})$, $b_i(\theta)$, and $c(\theta)$, where $A(\bar{x}) = 0$ if and only if $h_\theta(\bar{x}) = 0$, and $b_1(\theta)$, $b_2(\theta)$, $\cdots$, $b_k(\theta)$, $c(\theta)$ are linearly independent and each is differentiable with respect to θ_α, $\alpha = 1, \cdots, k$.

PROOF. This proposition is a generalization of Proposition 1. A detailed proof for it, as well as for the following remark, is given in [3].

REMARK. Consider the case where $h_\theta(\bar{x}) = \prod_{i=1}^n h_\theta(x_i)$ and $\mu(\bar{x})$ is a product measure on R^n each component of which is μ (a σ-finite measure on R^1). Then the one-dimensional density $h_\theta(x)$ is of Darmois-Koopman form if and only if the oint density $h_\theta(\bar{x})$ is of Darmois-Koopman form.

Consider now the family of densities corresponding to $\mathcal{H}$ of (13), viz.,

$$\begin{aligned} \mathcal{H}' &= \{h_\theta(x) = \textstyle\sum_1^{k+1} \theta_i f_i(x); \\ \theta_{k+1} &= 1 - \textstyle\sum_1^k \theta_i, \theta_i > 0, i = 1, \cdots, k+1\} \\ &= \{\textstyle\sum_1^k \theta_i[f_i(x) - g(x)] + g(x), \theta = (\theta_1, \cdots, \theta_k) \in \Theta\} \end{aligned} \tag{15}$$

where $g(x) \equiv f_{k+1}(x)$ and $\Theta = \{\theta: \sum_{i=1}^k \theta_i < 1, \theta_i > 0, i = 1, \cdots, k\}$. It can be shown that the Regularity Conditions (a) through (e) are satisfied for $\Omega = \Theta$, $h_\theta(x)$ defined by (15) and $\hat{\theta} \in \mathfrak{U}$, where $\mathfrak{U}$ is, as earlier, the class of unbiased estimators of θ whose covariance matrix is positive definite for all $\theta \in \Theta$. The following theorem, a generalization of Theorem 1 follows from Proposition 3 and the preceding remark. A proof appears in [3].

THEOREM 5. *Suppose $\mathcal{H}'$ of (15) is identifiable. A necessary and sufficient condition that there exists a jointly efficient estimator within $\mathfrak{U}$ is that there exists a decomposition of S_H^+ into sets $\Lambda_1, \cdots, \Lambda_{k+1}$ of positive μ-measure such that*

$$h_\theta(x) = V(x)\cdot K_j(\theta) \quad \text{for} \quad x \in \Lambda_j, j = 1, \cdots, k+1. \tag{16}$$

An equivalent condition is that there exists a decomposition of S_H^+ into sets $\Lambda_1, \cdots, \Lambda_{k+1}$ of positive μ-measure such that $f_\alpha(x)$ is proportional to $f_\beta(x)$ on each Λ_j, $j = 1, \cdots, k+1$ for $\alpha, \beta = 1, \cdots, k+1$.

Theorem 5 gives necessary and sufficient conditions for the existence of a jointly efficient estimator given that $\mathcal{H}'$ (see Equation 15) is identifiable. Unfortunately, not all families $\mathcal{H}'$ which satisfy the factorization criterion (16) of Theorem 5 are identifiable. The following proposition serves two purposes; first, it provides a method for checking the identifiability of a family of finite mixtures for which a jointly efficient estimator exists; and second, it prefaces the derivation of the jointly efficient estimator given in the succeeding corollary.

PROPOSITION 4. *For the family $\mathcal{H}'$, suppose that $h_\theta(x) = V(x)\cdot K_j(\theta)$ for $x \in \Lambda_j$, $j = 1, \cdots, k+1$ where the Λ_j's are disjoint sets of positive μ-measure such that $\bigcup_{j=1}^{k+1} \Lambda_j = S_H^+$. Then $\mathcal{H}'$ is identifiable if and only if there exists a subset $\{\Lambda_{i_1}, \cdots, \Lambda_{i_k}\}$ of $\{\Lambda_1, \cdots, \Lambda_{k+1}\}$ such that $|F_\alpha(\Lambda_{i_\beta}) - G(\Lambda_{i_\beta})| \neq 0, \alpha, \beta = 1, \cdots, k$.*

PROOF. Similar to the proof of Theorem 1 in Teicher [8].

COROLLARY. *Let $\mathcal{H}'$ be identifiable and satisfy the factorization criterion* (16) *with the sets* Λ_j *indexed so that* $|F_i(\Lambda_j) - G(\Lambda_j)| \neq 0$, $i, j = 1, \cdots, k$. *Let* $\hat{\theta}^* = (\hat{\theta}_1^*, \cdots, \hat{\theta}_k^*)$ *be the solution to the equations*

$$H_n(\Lambda_j) - G(\Lambda j) = \sum_{i=1}^k \hat{\theta}_i[F_i(\Lambda_j) - G(\Lambda_j)], \quad j = 1, \cdots, k,$$

then $\hat{\theta}^*$ *is the jointly efficient estimator of* θ.

3.3. θ^o-*efficient estimators*. An estimator, $\hat{\theta}(\theta^o) = (\hat{\theta}_1(\theta^o), \cdots, \hat{\theta}_k(\theta^o))$ of $\theta = (\theta_1, \cdots, \theta_k)$, is called θ^o-efficient if its ellipsoid of concentration coincides with the minimum ellipsoid of concentration at the point θ^o. Let

$$(17) \quad T_i = T_i(\vec{x}; \theta^o) = n^{-1} \sum_{j=1}^n [f_i(x_j) - g(x_j)]/h_{\theta^o}(x_j)\}, \quad i = 1, \cdots, k,$$

and $J_{ij}(\theta) = E((\partial/\partial\theta_i) \log h_\theta(X)(\partial/\partial\theta_j) \log h_\theta(X))$. Note that $I_{ij}(\theta) = nJ_{ij}(\theta)$ for the case of independent and identically distributed random variables, where $I_{ij}(\theta)$ is as in (14).

PROPOSITION 5. *The estimator* $\hat{\theta}(\theta^o) = (\hat{\theta}_1(\theta^o), \cdots, \hat{\theta}_k(\theta^o))$ *of* θ, *defined by*

$$(18) \quad \hat{\theta}_i(\theta^o) = \theta_i^o + \sum_{j=1}^k J^{ij}(\theta^o)T_j, \quad i = 1, \cdots, k.$$

is θ^o-*efficient, where* T_j *is as in* (17) *and* $(J^{ij}(\theta)) = (J_{ij}(\theta))^{-1}$.

PROOF. $E_\theta T_i = \sum_{j=1}^k (\theta_j - \theta_j^o)J_{ij}(\theta^o)$; hence, from (18), $E_\theta(\hat{\theta}_i - \theta_i^o) = \sum_{j=1}^k J^{ij}(\theta^o)E_\theta T_j$, i.e., $\hat{\theta}_i$ is unbiased. Also, $\sigma_{ij}(\theta) = \text{Cov}[\hat{\theta}_i(\theta^o), \hat{\theta}_j(\theta^o)] = \sum_\alpha \sum_\beta J^{i\alpha}(\theta^o)J^{\beta j}(\theta^o) \text{Cov}[T_\alpha, T_\beta]$, $i, j = 1, \cdots, k$; and, $\text{Cov}[T_\alpha, T_\beta] = n^{-1}[\int [(f_\alpha - g)/h_{\theta_0}]\cdot[(f_\beta - g)/h_{\theta_0}]\cdot h_\theta \, d\mu(x) - E_\theta[(f_\alpha - g)/h_{\theta^o}]E_\theta(f_\beta - g)/h_{\theta^o}]$, whence $\text{Cov}[T_\alpha, T_\beta]|_{\theta=\theta^o} = n^{-1}J_{\alpha\beta}(\theta^o)$, and so $\sigma_{ij}(\theta^o) = n^{-1} \sum_{\alpha=1}^k \sum_{\beta=1}^k J^{i\alpha}(\theta^o)J^{\beta j}\cdot(\theta^o)J_{\alpha\beta}(\theta^o) = n^{-1}J^{ij}(\theta^o)$. Equivalently, $nJ_{ij}(\theta^o) = \sigma^{ij}(\theta^o)$, hence the minimum ellipsoid of concentration (14) coincides with the ellipsoid of concentration of $\hat{\theta}(\theta^o)$ at the point θ^o.

REMARK. If $\tilde{\theta}_n^* = (\theta_1^*, \cdots, \theta_k^*)$ is any estimator of θ for which $\tilde{\theta}_n^* = \theta + O_P(n^{-\frac{1}{2}})$, then $\hat{\theta}_i(\tilde{\theta}_n^*) = \theta_i^* + \sum_{j=1}^k J^{ij}(\tilde{\theta}_n^*)T_j^*$, $i = 1, \cdots, k$, in addition to being $\tilde{\theta}_n^*$-efficient, is a consistent asymptotically normal efficient estimator of θ, where $T_j^* = n^{-1} \sum_{i=1}^n [f_j(x_i) - g(x_i)]/h_{\tilde{\theta}_n^*}(x_i)$, $j = 1, \cdots, k$. Here O_P is the probability order symbol introduced in [5]. A sequence of random variables is $O_P[f(n)]$ if for every $\epsilon > 0$ there exist constants $A = A(\epsilon) > 0$ and $N = N(\epsilon) > 0$ such that $P\{|Y_n/f(n)| > A\} < \epsilon$ for every $n > N$.

4. Acknowledgment. This manuscript contains part of the author's dissertation written under the direction of Professor Henry Teicher; his guidance and encouragement during the preparation were invaluable.

REFERENCES

[1] BAHADUR, R. R. (1958). On unbiased estimates of uniformly minimum variance. *Sankhyā*. **18** 211–224.

[2] BARANKIN, E. W. (1949). Locally best unbiased estimates. *Ann. Math. Statist.* **20** 477–501.

[3] BOES, D. C. (1963). On the estimation of mixing distributions. Ph.D. thesis. Purdue University, Lafayette, Indiana.

[4] FEND, A. V. (1959). On the attainment of Cramér-Rao and Bhattacharyya bounds for the variance of an estimate. *Ann. Math. Statist.* **30** 381–388.

[5] MANN, H. B. and WALD, A. (1943). On stochastic limit and order relationships. *Ann. Math. Statist.* **14** 217–226.

[6] SHALAEVSKII, O. V. (1961). A short proof of the Cramér-Rao inequality. *Theor. Prob. Appl.* **6** 323–324.

[7] STEIN, CHARLES (1950). Unbiased estimates with minimum variance. *Ann. Math. Statist.* **21** 406–415.

[8] TEICHER, HENRY (1963). Identifiability of finite mixtures. *Ann. Math. Statist.* **34** 1265–1269.