# UNBIASED ESTIMATION IN CONVEX FAMILIES[1]

### By P. J. Bickel and E. L. Lehmann

### *University of California, Berkeley*

**1. Introduction and summary.** Suppose that we observe random variables $X_1, \cdots, X_n$ which are identically and independently distributed according to some distribution $F$ where $F$ ranges over a family $\mathfrak{F}$.

The following question was posed by J. Steffensen in this abstract context in [7].

(1) If a functional $q$ is given on $\mathfrak{F}$, when does there exist a statistic $\delta(X_1, \cdots, X_n)$ such that,

$$(1.1) \qquad E_F(\delta(X_1, \cdots, X_n)) = q(F),$$

for all $F \, \varepsilon \, \mathfrak{F}$? (As usual $E_F$ denotes the expectation under the assumption that $F$ is the common distribution of the $X_i$.) In other words, when does there exist an unbiased estimate of $q(F)$ based on $n$ observations?

A question which naturally follows from question 1 was raised and considered by Halmos in [2].

(2) Let the *degree* of $q$ be the smallest $n \leq \infty$ such that (1.1) holds for all $F \, \varepsilon \, \mathfrak{F}$ and some $\delta$. Characterize the degree of estimable functionals $q$ for various families $\mathfrak{F}$. (By estimable we mean merely that (1.1) should hold for some finite $n$ and $\delta$.)

In his paper [2] Halmos dealt with question 2 as well as the more important issue of existence of best unbiased estimates in the context of families $\mathfrak{F}$ which are very large, and asked whether anything can be said about more interesting families $\mathfrak{F}$.

The next two sections of our paper deal exclusively with question 2 for families $\mathfrak{F}$ which are closed under finite mixtures (convex combinations). In Section 2, some examples lead to our main result, Theorem 2.1, which gives an elementary characterization of degree for functionals defined on families $\mathfrak{F}$ as above. The theorem is followed by further examples which illustrate its utility. (However, the required convexity property excludes the usual parametric families.) Theorem 2.2, an extension of a theorem of Halmos [2], characterizes the degree of a product of estimable functionals. It is applied to give the degree of the $r$th cumulant of a distribution when $\mathfrak{F}$ is the family of all distributions with moments of all orders.

The third section deals with the problem of degree when we observe samples from two or more populations with different distributions. Theorem 3.1 shows that the notion of degree can be successfully extended in this situation. The theorem is followed by several examples in which the degree of various functionals such as two–sample distance criteria is computed.

Finally, in Section 4 we address ourselves to question 1. Theorem 4.1 gives a simple characterization of estimable functionals when the family $\mathfrak{F}$ is $\mathfrak{F}(\mu)$ where,

$$(1.2) \qquad \mathfrak{F}(\mu) = \{F : F \text{ is absolutely continuous with respect to } \mu\}$$

for a fixed $\sigma$ finite measure $\mu$. The characterization is suggested by Theorem 2.1. After some examples we state without proof a further theorem (4.2) along these lines and illustrate its applicability.

## 2. Degree.
We begin with some definitions. Throughout, $\mathfrak{F}$ denotes a family of distribution functions (usually on the real line).

DEFINITION 2.1. Suppose that $q$ is a functional on $\mathfrak{F}$. We say $q$ is *estimable with n observations* if there exists $\delta$ such that (1.1) is satisfied for all $F \varepsilon \mathfrak{F}$ and that $q$ is *estimable* if it is estimable with $n$ observations for some $n$. The *degree* of an estimable functional $q$ is the smallest $n$ for which $q$ is estimable.

Do estimability and degree depend on the domain $\mathfrak{F}$ over which $q$ is defined? Consider, for example, the class $\mathfrak{F}$ of all distributions with finite variance $\sigma^2(F)$ and let $q$ be the standard deviation $\sigma(F)$. As we shall show shortly $q$ is not estimable. It becomes estimable, however, if we restrict the domain to $\mathfrak{F}_0$ consisting of a single distribution $F_0 \varepsilon \mathfrak{F}$. Similarly $\sigma^2(F)$ itself will be seen to be of degree 2 in $\mathfrak{F}$; it is clearly of degree 0 in $\mathfrak{F}_0$ while it is of degree 1 in the class $\mathfrak{F}_1$ of all distributions $F \varepsilon \mathfrak{F}$ with $\int x \, dF(x) = a$ where $a$ is any given constant.

DEFINITION 2.2. $\mathfrak{F}$ is *convex* if for every $F$, $G \varepsilon \mathfrak{F}$ and $0 \leq \alpha \leq 1$, $\alpha F + (1 - \alpha)G \varepsilon \mathfrak{F}$.

Given $\mathfrak{F}$ convex and a functional $q$ on $\mathfrak{F}$ we define for fixed $F$ and $G$ in $\mathfrak{F}$ a function $Q_{F,G}$ on $[0, 1]$ by

$$(2.1) \qquad Q_{F,G}(\alpha) = q[\alpha F + (1 - \alpha)G].$$

If $q$ is estimable then $Q_{F,G}$ is a polynomial for every $F$ and $G$. For if (1.1) holds, then

$$
\begin{aligned}
& q[\alpha F + (1 - \alpha)G] \\
(2.2) \quad & = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta(x_1, \cdots, x_n)[\alpha \, dF(x_1) + (1 - \alpha) \, dG(x_1)] \\
& \qquad \cdots [\alpha \, dF(x_n) + (1 - \alpha) \, dG(x_n)].
\end{aligned}
$$

It should now be clear that if $\mathfrak{F}$ is convex and the distributions $F \varepsilon \mathfrak{F}$ have finite variance $\sigma^2(F)$ then the standard deviation $\sigma(F)$ is not estimable for any $n$. The above remark and the further considerations of this paper do not require the distributions $F$ to be univariate. Suppose, for example, that $\mathfrak{F}$ is the class of all bivariate distributions $F$, whose components have finite variances. Then the correlation coefficient of the components considered as a functional of $F$ is not estimable. Similarly, if $A$ and $B$ denote fixed sets on the real line and $(U, V)$ is a pair of random variables with joint distribution $F$ and $\mathfrak{F}$ is the class of all $F$ for which $P(V \varepsilon B) > 0$, then the conditional probability $P(U \varepsilon A \mid V \varepsilon B)$ is not estimable.

We shall return to the question of estimability in Section 4. In the present sec-

tion we assume that $q$ is estimable and try to determine its degree. In this connection (2.2) is again helpful. It shows that if there exists a pair $F$, $G$ for which $Q_{F,G}$ is a polynomial of degree $n$, then the degree of $q$ cannot be less than $n$. If then an estimate of $q$ can be found, which is based on $n$ observations, this establishes $n$ as the degree of $q$.

EXAMPLE 2.1. As an example, consider the variance $\sigma^2(F)$ of a univariate distribution, with $\mathfrak{F}$ the class of all distributions with finite variance. Then $Q$ is of degree 2 and hence $q$ cannot be estimated with one observation. Since $\frac{1}{2}(X_2 - X_1)^2$ is an unbiased estimate based on two observations, this shows $\sigma^2(F)$ to be of degree 2.

EXAMPLE 2.2. For a second example let $q(F) = \int F \, dF$. The associated $Q$ is,

$$(2.3) \quad Q_{F,G}(\alpha) = \alpha^2 \int (F - G) \, d(F - G) + \alpha[\int (F - G) \, dG + \int G \, d(F - G)]$$
$$+ \int G \, dG.$$

If $\mathfrak{F}$ is the class of all univariate continuous distributions $q(F)$ is identically equal to $\frac{1}{2}$ and hence of degree 0. On the other hand, if $\mathfrak{F}$ is the class of all distributions, $Q$ is easily seen to be of degree 2 for suitable $F$, $G$. Since $\delta(X_1, X_2) = 1$ or 0 as $X_1 \leqq X_2$ or $X_1 > X_2$ is an estimate of $Q(F)$ based on two observations this shows that $q$ is of degree 2. One can similarly show that if $\mathfrak{F}$ is the family of all continuous *bivariate* distributions then the degree of $q$ is again 2.

The above examples suggest that, roughly speaking, a functional $q$ is of degree $n$ provided $Q$ is a polynomial of degree $n$. This is made precise by the following theorem.

THEOREM 2.1. *Suppose $\mathfrak{F}$ is convex and $q$ is an estimable functional on $\mathfrak{F}$. Then $q$ is of degree $n$ if and only if*

(i) *for every $F$ and $G \, \varepsilon \, \mathfrak{F}$, $Q_{F,G}$ is a polynomial in $\alpha$ of degree $\leqq n$;*
*and*

(ii) *for every $F \, \varepsilon \, \mathfrak{F}$ there exists some $G$ for which $Q_{G,F}(\alpha)$ is a polynomial of degree $n$ in $\alpha$.*

The essence of the proof is contained in the following lemma.

LEMMA 2.1. (i) *If $q$ is of degree $\leqq n$, then $Q_{F,G}$ is a polynomial of degree $\leqq n$ for all $F$ and $G$.*

(ii) *If there exists an $F$ such that $Q_{F,G}$ is a polynomial of degree $< n$ for all $G$ and if $q$ is estimable, then $q$ is of degree $< n$.*

PROOF. (i) By (2.2) if $q$ is estimable with $n$ observations, and $\delta(x_1, \cdots, x_n)$ is an estimate of $q$, then,

$$(2.4) \qquad\qquad Q_{F,G}(\alpha) = \sum_{r=0}^{n} \alpha^r (1 - \alpha)^{n-r} C_r(F, G),$$

is a polynomial of degree $\leqq n$ and (i) is immediate.

(ii) We identify the coefficients $C_r$ by,

$$(2.5) \quad C_r(F, G) = \sum_{r=0}^{n} \sum_{\{i_1, \cdots, i_r\}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta(x_1, \cdots, x_n)$$
$$\cdot \prod_{j=1}^{r} dF(x_{i_r}) \prod_{i \varepsilon B_n(i_1, \cdots, i_r)} dG(x_i),$$

where $B_n(i_1, \cdots, i_r) = \{1, \cdots, n\} - \{i_1, \cdots, i_r\}$ and $\{i_1, \cdots, i_r\}$ ranges over all distinct subsets of size $r$ of $\{1, \cdots, n\}$.

The hypothesis of (ii) implies that for a fixed $F$

$$(2.6) \qquad \sum_{r=0}^n (-1)^{n-r} C_r(F, G) = 0.$$

Now let us define,

$$(2.7) \quad \delta_{(i_1, \cdots, i_r)}(x_{j_1}, \cdots, x_{j_s}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta(x_1, \cdots, x_n) \, dF(x_{i_1}) \cdots dF(x_{i_r}),$$

where $\{j_1, \cdots, j_s\} = B_n\{i_1, \cdots, i_r\}$. $\delta_{(i_1, \cdots, i_r)}$ is defined a.s. $(F)$ by (2.7) and may be set equal to 0 when the right hand side of (2.7) is undefined. In terms of the $\delta_{(i_1, \cdots, i_r)}$, (2.6) may be rewritten as

$$
\begin{aligned}
(2.8) \qquad &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta(x_1, \cdots, x_n) \, dG(x_1) \cdots dG(x_n) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[ \sum_{r=1}^n (-1)^{n-r+1} \sum_{(i_1, \cdots, i_r)} \delta_{(i_1, \cdots, i_r)}(x_{j_1}, \cdots, x_{j_s}) \right] \\
&\quad \cdot dG(x_1) \cdots dG(x_n)
\end{aligned}
$$

which in turn may be read,

$$(2.9) \quad E_G(\delta(X_1, \cdots, X_n)) = E_G\left[ \sum_{r=1}^n (-1)^{n-r+1} \delta_{(i_1, \cdots, i_r)}(X_1, \cdots, X_s) \right]$$

for all $G \, \varepsilon \, \mathcal{F}$. Since $r \geq 1$, $s$ is less than $n$ and we see that

$$\sum_{r=1}^n (-1)^{n-r+1} \delta_{(i_1, \cdots, i_r)}(X_1, \cdots, X_s)$$

is an estimate of $q$ based on at most $n - 1$ observations. Assertion (ii) of the lemma follows.

PROOF OF THEOREM 2.1. (a) Let $q$ be of degree $n$. Then for every $F$ there exists $G$ such that $Q$ is of degree $n$. If there did not, it would follow from part (i) of the lemma that there exists an $F$ for which $Q$ is of degree $<n$ for all $G$ and hence by part (ii) of the lemma, $q$ would be of degree $<n$.

(b) Suppose that conditions (i) and (ii) of the theorem hold and that $q$ is estimable. Then it follows from part (i) of the lemma that $q$ cannot be of degree $<n$ and from part (ii) that $q$ cannot be of degree $>n$.

We have considered several examples in which condition (i) of Theorem 2.1 proved useful. Here are some further illustrations in which the full force of the theorem is used.

EXAMPLE 2.3. Let $\mathcal{F}$ be the family of all distributions $F$ which are absolutely continuous with continuous densities $f$ and suppose that we wish to estimate $q(F) = f(a)$ for some fixed known $a$. In this situation $Q$ is evidently of degree $\leq 1$ for all $F$, $G$. Yet $q$ is not estimable. To see this, note that in view of Theorem 2.1, if $q$ is estimable it must be estimable with 1 observation. Therefore, it suffices to exhibit a subfamily $\mathcal{F}_0 \subset \mathcal{F}$ for which $q$ is not estimable with 1 observation. Let $\mathcal{F}_0 = \{f : f(t) = \frac{1}{2}\lambda \exp(-\lambda |t|) \text{ for some } \lambda > 0 \text{ and all } t\}$. If $q$ were estimable with 1 observation in $\mathcal{F}_0$ we would have,

$$(2.10) \qquad \int_{-\infty}^{\infty} \delta(t)[\exp\{-\lambda(|t| - |a|)\}] \, dt = 1$$

for some estimate $\delta$ and all $\lambda > 0$. Differentiating both sides we get,

$$(2.11) \qquad \int_{-\infty}^{\infty} \delta(t)[|t| - |a|][\exp\{-\lambda(|t| - |a|)\}\, dt] = 0.$$

The uniqueness of the Laplace transform implies,

$$(2.12) \qquad \delta(t)[|t| - |a|] = 0 \text{ a.e.}$$

and this is inconsistent with (2.10). (This result was proved in a different manner by Rosenblatt (1956) in [6]. The present proof is no simpler but has the slight advantage of not being restricted to nonnegative estimates.)

EXAMPLE 2.4. Consider the following generalization of Tukey's [9] gross error model.

Let $\mathfrak{F} = \{F : F = (1 - \epsilon)\Phi + \epsilon H, H \,\varepsilon\, \mathfrak{IC}, 0 \leq \epsilon \leq 1\}$ where $\Phi$ is the standard normal distribution, $\mathfrak{IC}$ is a convex, complete (in the sense of [4]) family of distributions equivalent to $\Phi$. Suppose, moreover, that $\mathfrak{F}$ is identifiable in the sense that,

$$(2.13) \qquad (1 - \epsilon_1)\Phi + \epsilon_1 H_1 = (1 - \epsilon_2)\Phi + \epsilon_2 H_2$$

implies $\epsilon_1 = \epsilon_2$ and $H_1 = H_2$. The identifiability and completeness conditions are satisfied by taking for $\mathfrak{IC}$ all mixtures of rectangular distributions, all distributions with support bounded from below etc.

The functional to be estimated is $q((1 - \epsilon)\phi + \epsilon H) = \epsilon$. Clearly, $\mathfrak{F}$ is convex and $Q$ is linear. Again, $q$ is not estimable. For were it estimable by Theorem 2.1 it would be of degree 1. But,

$$(2.14) \qquad (1 - \epsilon) \int_{-\infty}^{\infty} \delta(x)\, d\Phi(x) + \epsilon \int_{-\infty}^{\infty} \delta(x)\, dH(x) = \epsilon$$

implies

$$(2.15) \qquad \int_{-\infty}^{\infty} \delta(x)\, d\Phi(x) = 0$$

and

$$(2.16) \qquad \int_{-\infty}^{\infty} \delta(x)\, dH(x) = 1.$$

It follows from the completeness of $\mathfrak{IC}$ that these equations cannot be satisfied if $\mathfrak{IC}$ is equivalent to $\Phi$.

EXAMPLE 2.5. For our final example let $\mathfrak{F}_0$ be the location parameter family of distributions generated by the standard Cauchy distribution $F_0$ whose density is given by,

$$(2.17) \qquad f_0(t) = \pi^{-1}(1 + t^2)^{-1}.$$

Let $\mathfrak{F} = \mathfrak{C}(\mathfrak{F}_0)$ be the smallest convex family of distributions containing $\mathfrak{F}_0$. Suppose that for some $F \,\varepsilon\, \mathfrak{F}$

$$(2.18) \qquad F(t) = \sum_{i=1}^{r} \alpha_i F_0(t - \theta_i) = \sum_{j=1}^{s} \beta_j F_0(t - \eta_j)$$

for all $t$, where the $\alpha$'s and $\beta$'s are $\geq 0$ and $\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{s} \beta_j = 1$. Then it follows easily from a consideration of the associated characteristic functions that

$r = s$ and the set of $\alpha$'s must coincide with the set of $\beta$'s as must the set of $\theta$'s with the set of $\eta$'s. In view of this we can define the "center of symmetry" $q(F)$ of a distribution $F$ given by (2.18) by,

$$(2.19) \qquad q(F) = \sum_{i=1}^{r} \alpha_i \theta_i .$$

The $Q$ function corresponding to $q$ is linear, but again $q$ is not estimable since $q$ is of degree $>1$ on $\mathfrak{F}_0$. This was shown in [1] by Ghosh and Singh using some results by Pollard.

Similar examples in which $Q$ is of degree $>1$ for each pair $F$, $G$ and $q$ is not estimable, can easily be constructed.

Before proving our next result, we need an elementary lemma.

LEMMA 2.2. *Let $\mathfrak{F}$ be convex and $q_1$, $q_2$ defined on $\mathfrak{F}$ be estimable with $n$ observations. Then either there exists $F \varepsilon \mathfrak{F}$ such that $q_1(F)q_2(F) \neq 0$ or one of $q_1$, $q_2$ is identically $0$.*

PROOF. Either the second alternative of the lemma holds or we may assume there exist $F^*$, $G^*$ such that $q_1(F^*) \neq 0$, $q_2(G^*) \neq 0$. Then,

$$q_1(\alpha F^* + (1 - \alpha)G^*)q_2(\alpha F^* + (1 - \alpha)G^*)$$

is a product of two polynomials of degree at most $n$ neither of which vanishes identically. Hence there exists $\alpha_0$ such that

$$q_1(\alpha_0 F^* + (1 - \alpha_0)G^*)q_2(\alpha_0 F^* + (1 - \alpha_0)G^*) \neq 0$$

and the lemma is proved.

We now establish a generalization of Theorem 2 of [2].

THEOREM 2.2. *Suppose $q_1$, $\cdots$, $q_r$ are estimable functions of degree $k_1$, $\cdots$, $k_r$ respectively, defined on a convex domain $\mathfrak{F}$. Then, $q = q_1 \cdots q_r$ is estimable of degree $k_1 + \cdots + k_r$.*

PROOF. It suffices to check the result for $r = 2$ since the general case then follows by induction.

Clearly, $q$ has degree at most $k_1 + k_2$. To see this, note that if $\delta^{(1)}(X_1, \cdots, X_{k_1})$ and $\delta^{(2)}(X_1, \cdots, X_{k_2})$ are unbiased estmates of $q_1$ and $q_2$ based on $k_1$ and $k_2$ observations respectively, then $\delta(X_1, \cdots, X_{k_1+k_2}) = \delta^{(1)}(X_1, \cdots, X_{k_1}) \cdot \delta^{(2)}(X_{k_1+1}, \cdots, X_{k_1+k_2})$ is an unbiased estimate $q$ based on $k_1 + k_2$ observations.

To show that the degree of $q$ cannot be less than $k_1 + k_2$ let $K(F, G)$, $K_1(F, G)$, $K_2(F, G)$ denote the coefficients of $\alpha^{k_1+k_2}$, $\alpha^{k_1}$ and $\alpha^{k_2}$ in the $Q_{(F,G)}$ polynomials corresponding to $q$, $q_1$ and $q_2$ respectively. From (2.5) we see that for fixed $F$, $K_1(F, G)$ and $K_2(F, G)$ treated as functions of $G$ are estimable with $k_1$ and $k_2$ observations respectively. Since $q_1$ and $q_2$ are of degree $k_1$ and $k_2$ respectively Theorem 2.1 implies that whatever be $F$, neither $K_1(F, G)$ nor $K_2(F, G)$ vanish identically. Hence, by Lemma 2.2 neither does $K(F, G) = K_1(F, G)K_2(F, G)$. Therefore, whatever be $F$, $q(\alpha F + (1 - \alpha)G)$ is of degree $(k_1 + k_2)$ for some $G$ and the theorem follows from Theorem 2.1.

Along the same lines it is easy to see that if $\mathfrak{F}$ is convex, $q_1$ is of degree $m$ on $\mathfrak{F}$ and $q_2$ is of degree $n$ of $\mathfrak{F}$ with $m < n$ then $q_1 + q_2$ is of degree $n$ on $\mathfrak{F}$.

EXAMPLE 2.6. As an application of this remark and Theorem 2.1, consider $\mathfrak{F} = \{F : \int_{-\infty}^{\infty} |x|^k \, dF(x) < \infty$ for all $k\}$. The $r$th *cumulant* of $F$, $K_r(F)$ is defined as the coefficient of $(it)^r/r!$ in the formal expansion of $\log \int_{-\infty}^{\infty} e^{itx} \, dF(x)$. Clearly,

$$(2.20) \qquad K_r(F) = \sum_{\{(j_1, \cdots, j_r) : \Sigma i j_i = r\}} c(j_1, \cdots, j_r) \alpha_1^{j_1}(F) \cdots \alpha_r^{j_r}(F)$$

where $\alpha_k(F) = \int_{-\infty}^{\infty} x^k \, dF(x)$. Since $c(r, 0, \cdots, 0) \neq 0$ and since $\alpha_k(F)$ is of degree 1, Theorem 2.2 and our remark show that $K_r(F)$ is of degree $r$.

## 3. Degree: Two-sample problem.

Let us now extend the definition and characterization of degree to functionals of two distributions. The further extension to an arbitrary number of distributions is trivial. Suppose that we observe $X_1, \cdots, X_m$ independently and identically distributed according to $F_1 \varepsilon \mathfrak{F}_1$ and $Y_1, \cdots, Y_n$ independently and identically distributed according to $F_2 \varepsilon \mathfrak{F}_2$ where $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are convex.

DEFINITION 3.1. A real functional $q$ on $\mathfrak{F}_1 \times \mathfrak{F}_2$ is said to be estimable with $(m, n)$ observations if there exists a statistic $\delta(X_1, \cdots, X_m ; Y_1, \cdots, Y_n)$ such that

$$(3.1) \qquad E_{(F_1, F_2)}[\delta(X_1, \cdots, X_m ; Y_1, \cdots, Y_n)] = q(F_1, F_2)$$

for all $F_1 \varepsilon \mathfrak{F}_1$, $F_2 \varepsilon \mathfrak{F}_2$.

In generalization of (2.1) it is useful to introduce a function $Q$ through

$$(3.2) \qquad Q(\alpha_1, \alpha_2) = q[\alpha_1 F_1 + (1 - \alpha_1)G_1, \alpha_2 F_2 + (1 - \alpha_2)G_2]$$

where $0 \leqq \alpha_1, \alpha_2 \leqq 1$ and where we suppress the dependence of $Q$ on $F_i$, $G_i$, $i = 1, 2$.

If $q$ is estimable, it is seen as in Section 2 that $Q(\alpha_1, \alpha_2)$ is a polynomial for every $F_1$ and $F_2$. This remark is again enough to establish the non-estimability of many functions such as, for example, $q(F_1, F_2) = |F_1(x) - F_2(x)|$ for fixed $x$, or $q(F_1, F_2) = \int |F_1(x) - F_2(x)| \, d(F_1(x) + F_2(x))/2$.

Let us now consider how to define the degree of an estimable functional $q$. We might for example determine the smallest value $n_0$ of $n$ for which there exists an $m$ such that $q$ is estimable with $(m, n_0)$ observations and then the smallest value $m_0$ of $m$ for which $q$ is estimable with $(m, n_0)$ observations, and define the degree of $q$ on $(m_0, n_0)$. If we interchange the role of $m$ and $n$ in this definition, it is not clear a priori that we shall arrive at the same answer. To show that this is so, and at the same time arrive at a more symmetric definition of degree, consider the partial ordering among pairs of integers $(m, n)$ according to which $(m, n) <$ $(m', n')$ provided $m \leqq m'$ and $n \leqq n'$ with at least one of the two inequalities being strict. If $q$ is an estimable function, denote by $S_q$ the set of all pairs $(m, n)$ for which $q$ is estimable. Then we shall show below that the set $S_q$ has a least element with respect to the ordering $\leqq$; that is, there exists a pair of integers $(m_0, n_0)$ in $S_q$ such that $m \geqq m_0$, $n \geqq n_0$ for all $(m, n) \varepsilon S_q$ and that in fact $S_q$ consists exactly of all pairs $(m, n)$ with $m \geqq m_0$ and $n \geqq n_0$. This pair $(m_0, n_0)$ will be called the *degree* of the functional $q$.

The existence of a least element easily follows from the following properties of $S_q$ :

(i) $(m, n) \, \varepsilon \, S_q \Rightarrow (m + a, n + b) \, \varepsilon \, S_q$ for all $a, b \geq 0$;

(ii) $(m, n + 1) \, \varepsilon \, S_q$ and $(m + 1, n) \, \varepsilon \, S_q \Rightarrow (m, n) \, \varepsilon \, S_q$ .

PROOF OF (ii). If $q$ is estimable with $(a, b)$ observations, then $Q$ defined by (3.2) is a polynomial in $\alpha_1$ , $\alpha_2$ of degree at most $(a, b)$. (We say that the degree of a polynomial in $\alpha_1$ , $\alpha_2$ is $(a, b)$ if the maximum power of $\alpha_1$ that is present is $a$ and the maximum power of $\alpha_2$ that is present is $b$.) Thus for $q$ to be estimable both with $(m, n + 1)$ and $(m + 1, n)$ observations, $Q$ must be a polynomial of degree at most $(m, n)$ whatever be $F_i$ , $G_i$ . Writing $Q$ out explicitly, taking the $\alpha_1^{m+1}$ terms into account for the moment, we have

$$(3.3) \quad Q(\alpha_1 , \alpha_2)$$

$$= \sum_{k_1=0}^{m+1} \sum_{k_2=0}^{n} C_{k_1,k_2}(F_1 , F_2 , G_1 , G_2)\alpha_1^{k_1} (1 - \alpha_1)^{m+1-k_1}\alpha_2^{k_2} (1 - \alpha_2)^{n-k_2}$$

where the $C_{k_1,k_2}$ are defined by formulae analogous to (2.5) in terms of the estimate $\delta(X_1 , \cdots , X_{m+1} , Y_1 , \cdots , Y_n)$ based on $(m + 1, n)$ observations. Since $Q$ is of degree at most $(m, n)$ the coefficient of $\alpha_1^{m+1}$ must vanish so that,

$$(3.4) \quad \sum_{k_2=0}^{n} \alpha_2^{k_2} (1 - \alpha_2)^{n-k_2} \left( \sum_{k_1=0}^{m} (-1)^{m-k_1+1} C_{k_1,k_2}(F_1 , F_2 , G_1 , G_2) \right) \equiv 0.$$

By the completeness of the binomial family of distributions this implies that,

$$(3.5) \quad \sum_{k_1=0}^{m+1} (-1)^{m+1-k_1} C_{k_1,k_2}(F_1 , F_2 , G_1 , G_2) = 0$$

for all $k_2$ and in particular for $k_2 = 0$. By using the resulting equation, the result follows as did Lemma 2.1 from equation (2.6).

As in the univariate case, to determine the degree of a functional $q$, it is often enough to note that if $Q$ is a polynomial of degree $(m, n)$, the degree of $q$ cannot be less than $(m, n)$. If an estimate of $q$ can then be found based on $(m, n)$ observations, this establishes $(m, n)$ as the degree of $q$. As an illustration, let $q = \int F_1 \, dF_2$ with $\mathfrak{F}_1 = \mathfrak{F}_2$ , the class of all continuous distributions. Here

$$Q(\alpha_1 , \alpha_2) = \alpha_1\alpha_2 \int (F_1 - G_1) \, d(F_2 - G_2)$$

$$+ \alpha_1 \int (F_1 - G_1) \, dG_2 + \alpha_2 \int G_1 d(F_2 - G_2) + \int G_1 \, dG_2$$

is clearly of degree $(1, 1)$. Since

$$\delta(X_1 , Y_1) = 1 \quad \text{or} \quad 0 \quad \text{as} \quad X_1 \leqq Y_1 \quad \text{as} \quad X_1 > Y_1$$

is an unbiased estimate of $q$, the degree of $q$ is $(1, 1)$.

As another example, consider $q = F_1(c) - F_2(c)$ where $c$ is any constant. Then

$$Q(\alpha_1 , \alpha_2) = \alpha_1[F_1(c) - G_1(c)] - \alpha_2[F_2(c) - G_2(c)] + G_1(c) - G_2(c)$$

is of degree $(1, 1)$. Also $I_{[X_1 \leqq c]} = I_{[Y_1 \leqq c]}$ where $I_{[E]}$ denotes the indicator function of the event $E$, is an unbiased estimate of $q$, and hence shows $q$ to be of degree $(1, 1)$. In the present case, it is noteworthy that the $\alpha_1\alpha_2$-term is missing in $Q$;

this makes it possible to estimate $q$ effectively with one observation rather than two. For let $X_1$ or $Y_1$ be observed with probability $\frac{1}{2}$ each, and let $\delta' = I_{[X_1 \leq c]}$ when the observed random variable is $X_1$ and $\delta' = I_{[Y_1 > c]}$ when the observed variable is $Y_1$. Then if $\delta = 2\delta' - 1$, we find

$$E(\delta') = \tfrac{1}{2}[F_1(c) + 1 - F_2(c)] \quad \text{and hence} \quad E(\delta) = q.$$

Of course, the variance of $\delta$ is larger, (in fact exactly twice as large), than that of the earlier estimate. (Such a reduction of the effective number of observations required to obtain an unbiased estimate of $q$ is possible whenever $q$ is the sum of a number of estimable functions $q_i$ each of which has a total degree $m_i + n_i$ less than the total degree $m + n$ of $q$.)

As another example, suppose that $X_1$, $X_2$ and $Y_1$, $Y_2$ are independently distributed according to continuous distributions $F_1$ and $F_2$ and let $q$ denote the probability that the two $X$'s are either both smaller or both larger than the two $Y$'s

$$q(F_1, F_2) = P(X_1, X_2 < Y_1, Y_2 \quad \text{or} \quad Y_1, Y_2 < X_1, X_2)$$
$$= \int (1 - F_2)^2 \, dF_1^2 + \int (1 - F_1)^2 \, dF_2.$$

Here it is easily seen that the coefficients of $\alpha_1^2 \alpha_2$ and of $\alpha_1 \alpha_2^2$ are in general different from zero while the coefficient of $\alpha_1^2 \alpha_2^2$ is zero. The latter fact follows immediately from the well-known representation (Lehmann (1951))

$$q(F_1, F_2) = \tfrac{1}{3} + \tfrac{1}{2} \int (F_1 - F_2)^2 \, d(F_1 + F_2).$$

It is clear in the present case that $q$ can be estimated by observing with probability $\frac{1}{2}$ each, either two $X$'s and one $Y$ or two $Y$'s and one $X$. This fact is related to an unbiased estimate of $q$ proposed by Renyi (1953).

The above remarks extend in the obvious way to more than two samples, and the natural generalizations of Theorem 2.1 and 2.2 hold trivially also in the case of two or more samples.

**4. Existence of unbiased estimates.** In Theorem 2.1 we have shown that if $q$ is estimable on a convex $\mathfrak{F}$ the $Q$ functions must be polynomials. On the other hand, Examples 2.3–2.5 show that this condition is not sufficient for estimability. In this section we shall show that this requirement is essentially sufficient if $\mathfrak{F}$ is large enough. Our main result deals with the case $\mathfrak{F} = \mathfrak{F}(\mu)$ defined in the introduction.

THEOREM 4.1. *If $q$ is a real functional on $\mathfrak{F}(\mu)$ then $q$ is estimable with $n$ observations if and only if,*
   (i) *$q$ is uniformly bounded on $\mathfrak{F}(\mu)$;*
   (ii) *$Q_{(F,G)}$ is a polynomial of degree at most $n$ in $\alpha$ for each $F, G \varepsilon \mathfrak{F}(\mu)$.*

We note that since $\mathfrak{F}(\mu)$ is convex it follows from Theorem 2.1 that $q$ is of degree $n$ if and only if (i) and (ii) hold and $Q_{F,G}(\alpha)$ is of degree $n$ for every $F$ and some $G$.

We need some preliminary definitions and results. Let $L_1(\mu) = \{f : \int |f| \, d\mu < \infty\}$

and $L_1^+(\mu) = \{f \, \varepsilon \, L_1(\mu) : f \geqq 0\}$. Evidently $\mathfrak{F}(\mu) = \{f \, \varepsilon \, L_1^+(\mu) : \int f \, d\mu = 1\}$ in the sense of the natural 1–1 correspondence between a distribution and its density.

DEFINITION 4.1. A functional $B$ on $L_1^+(\mu)$ to $R$ is said to be *linear* if $B(0) = 0$ and $B(a_1 f + a_2 g) = a_1 B(f) + a_2 B(g)$ for $f, g \, \varepsilon \, L_1^+(\mu)$, $a_1, a_2 \geqq 0$.

Similarly, a functional $B$ on $L_1^+(\mu) \times \cdots \times L_1^+(\mu)$ is said to be *n-linear* if $B$ is linear in each argument when the others are held fixed.

DEFINITION 4.2. An $n$ linear functional on $L_1^+(\mu) \times \cdots \times L_1^+(\mu)$ is said to be *bounded* if,

$$(4.1) \quad \sup \{|B(f_1, \cdots, f_n)| : \|f_1\|$$
$$= \cdots = \|f_n\| = 1, f_i \, \varepsilon \, L^+(\mu), i = 1, \cdots, n\} < \infty.$$

DEFINITION 4.3. A functional $q$ on $L_1^+(\mu)$ is said to be a *polynomial* if for every, $f, g \, \varepsilon \, L_1^+(\mu)$, $q(f + ag)$ is a polynomial in $a$ for $a \geqq 0$.

DEFINITION 4.4. A functional $q$ on $L_1^+(\mu)$ is said to be *homogeneous of degree n* if $q(cf) = c^n q(f)$ for every $f \, \varepsilon \, L_1^+(\mu)$, $c \geqq 0$.

These definitions are all slight modifications of the standard definitions given for these terms when $L_1(\mu)$ rather than just $L_1^+(\mu)$ is the domain of interest (cf. Hille and Phillips [3], pages 40, 760). The following results established in [3], pages 760–764 for homogeneous polynomials on $L_1(\mu)$ and $n$ linear functionals on $L_1(\mu) \times \cdots \times L_1(\mu)$ can easily be seen to continue to hold for the objects we have defined.

LEMMA 4.1. *If $q$ is a polynomial, homogeneous of degree $n$ on $L_1^+(\mu)$, there exists an $n$ linear functional $B$ on $L_1^+(\mu) \times \cdots \times L_1^+(\mu)$ such that $q(f) = B(f, f, \cdots, f)$.*

This is Theorem 26.2.2 of [3] specialized and modified.

LEMMA 4.2. *Suppose $B$ is an $n$ linear functional on $L_1^+(\mu) \times \cdots \times L_1^+(\mu)$ such that,*

$$(4.2) \qquad \sup \{B(f_1, \cdots, f_n) : \|f_i\| = 1\} < \infty,$$

*whenever all co-ordinates except the $i$th are fixed, for $i = 1, \cdots, n$. Then, $B$ is bounded.*

REMARK. The proof requires an appropriate modification of the principle of uniform boundedness which is easily seen to be valid. This is essentially Theorem 26.2.5 of [3].

Finally, we need a deeper result whose statement and proof for the situation when the domain of interest is $L_1(\mu)$ may be found, for example, in Trèves [8] (Corollary 2, page 473).

LEMMA 4.3. *If $B$ is a bounded $n$ linear functional on $L_1^+(\mu) \times \cdots \times L_1^+(\mu)$ then there exists a measurable function $\delta$ on $R^n$ such that,*

$$(4.3) \quad B(f_1, \cdots, f_n)$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta(x_1, \cdots, x_n) f_1(x_1) \cdots f_n(x_n) \mu(dx_1) \cdots \mu(dx_n).$$

Furthermore,

$$(4.4) \qquad (\mu) \operatorname{ess\,sup} |\delta| < \infty$$

where

(4.5)     $(\mu)\text{ess sup } |\delta|$

$$= \inf_B \sup \{|\delta(x_1, \cdots, x_n)| : (x_1, \cdots, x_n) \, \varepsilon \, B, \, \mu^{(n)}(B^c) = 0\}$$

and $\mu^{(n)}$ is the product measure on $R^n$ each of whose co-ordinate measures is $\mu$ and $B^c$ is the complement of $B$.

Finally we define,

DEFINITION 4.5. The *n extension of q*, written $q^{(n)}$ is a functional on $L_1^+(\mu)$ given by,

(4.6)                         $q^{(n)}(f) = \|f\|^n q(f/\|f\|)$.

It is obvious that the *n* extension of any functional $q$ on $\mathfrak{F}(\mu)$ is homogeneous of degree $n$.

We now proceed to the proof of Theorem 4.1.

PROOF.

*Sufficiency.* Construct $q^{(n)}$.

(4.7)              $q^{(n)}(f + ag) = \|f + ag\|^n q(f + ag/\|f + ag\|)$.

Since $f, g$ and $a$ are $\geqq 0$, $\|f + ag\| = \|f\| + a\|g\|$.
Now if (ii) holds,

(4.8)     $q\left(\dfrac{f + ag}{\|f + ag\|}\right) = q\left(\dfrac{f}{\|f\|}\left(\dfrac{\|f\|}{\|f + ag\|}\right) + \dfrac{g}{\|g\|}\left(\dfrac{a\|g\|}{\|f + ag\|}\right)\right)$

is a polynomial of degree at most $n$ in $a\|g\|/\|f + ag\|$. Expanding both factors in (4.7) suitably we conclude that $q^{(n)}$ is a polynomial. Since $q^{(n)}$ is homogeneous of degree $n$ Lemma 4.1 yields an $n$ linear functional $B$ on $L_1^+(\mu) \times \cdots \times L_1^+(\mu)$ such that $q^{(n)}(f) = B(f, \cdots, f)$. From (i) we see that

$$\sup \{|B(f, \cdots, f)| : \|f\| = 1\} < \infty$$

and hence, by linearity, that $B$ is bounded. Lemma 4.3 completes the proof of the sufficiency of our conditions.

*Necessity.* The necessity of (ii) is obvious. Given an estimate $\delta$ of $q$ define an $n$ linear functional $B$ on $L_1^+(\mu) \times \cdots \times L_1^+(\mu)$ by,

(4.9)   $B(f_1, \cdots, f_n)$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta(x_1, \cdots, x_n) f_1(x_1) \cdots f_n(x_n) \mu(dx_1) \cdots \mu(dx_n).$$

$B$ is well defined at least for $f_1 = \cdots = f_n$. Showing that $B$ is finite on all $L_1^+(\mu) \times \cdots \times L_1^+(\mu)$ merely involves breaking $\delta$ up into its positive and negative parts and then noting that $B(c_1 f_1 + \cdots + c_n f_n, \cdots, c_1 f_1 + \cdots + c_n f_n)$ is finite for all $c_1, \cdots, c_n \geqq 0$. By Lemma 4.3 we see that our necessity assertion is equivalent to showing that $B$ is bounded. By Lemma 4.2 and the Fubini Theorem we see that we need only prove the result for $n = 1$. But, in this case, define $B_m(f) = \int_{-\infty}^{\infty} \delta_m(x) f(x) \mu(dx)$

where                         $\delta_m = \delta$     if $|\delta| \leqq m$,

                                $= 0$     otherwise.

Then, $B_m(f)$ is a bounded linear functional and $B_m(f) \to B(f)$ for each $f$. The Banach-Steinhaus Theorem yields that $B$ is bounded and the theorem follows.

Extension of the theorem to the $k$–sample case is easy.

We turn to some examples to which the theorem or its $k$–sample extension may be applied.

EXAMPLE 4.1. Let $F_0$ be a fixed distribution. Take $\mu$ to be Lebesgue measure and let $q(F)$ be the "goodness of fit" statistic.

$$(4.10) \qquad q(F) = \int_{-\infty}^{\infty} (F - F_0)^2 h(F_0)\, dF_0$$

where $h$ is positive and bounded.

The requirements of the theorem are satisfied for $n = 2$ and $q$ is estimable on $\mathfrak{F}(\mu)$ with two observations. Of course here, as is typically the case in a"natural' situation, direct construction of the estimate of $q$ is easy. Theorem 2.1 guarantees that the degree is indeed 2 since the coefficient of $\alpha^2$ in $Q_{F,G}(\alpha)$ is

$$\int_{-\infty}^{\infty} (F - G)^2 h(F_0)\, dF_0 .$$

EXAMPLE 4.2. On the other hand, if $\mu$ is Lebesgue measure,

$$(4.11) \qquad q(F) = \int_{-\infty}^{\infty} (F - F_0)^2\, dF$$

$$= \int_{-\infty}^{\infty} F^2\, dF - 2 \int_{-\infty}^{\infty} FF_0\, dF + \int_{-\infty}^{\infty} F_0^2\, dF$$

is also of degree 2 since $\int_{-\infty}^{\infty} F^2\, dF = \frac{1}{3}$ and the coefficient of $\alpha^2$ in $Q_{F,G}(\alpha)$ is $\int_{-\infty}^{\infty} (F - G)^2\, dF_0$.

The same considerations can be applied to two–sample distance criteria such as,

$$(4.12) \qquad q(F_1, F_2) = \int_{-\infty}^{\infty} (F_1 - F_2)^2\, d(F_1 + F_2),$$

a functional we have already encountered in Section 3. Theorem 4.1 (or rather its extension) shows that $q$ is estimable with $(3,3)$ observations. Closer examination of $Q$ in this case shows that the degree is in fact $(2,2)$, and as was noted in Section 3, if an extraneous randomization is permitted only one observation from one population and two from the other are needed.

Finally, we give a theorem extending Theorem 4.1 to "unbounded" functionals defined on subdomains of $\mathfrak{F}(\mu)$.

Let $\{A_m\}$ be a nested sequence of measurable subsets of $R^n$, with $A_1 \subset A_2 \subset \cdots$ such that $\mu(R^n - \bigcup_m A_m) = 0$.

Let $\mathfrak{F}_m(\mu)$ be the set of all members of $\mathfrak{F}(\mu)$ which vanish off $A_m$.

THEOREM 4.2. *Let $q$ be a functional defined on a domain $\mathfrak{F}$ such that,*

$$(4.13) \qquad \mathfrak{F} \supset \bigcup_m \mathfrak{F}_m(\mu)$$

*and*

$$(4.14) \qquad q \geqq 0.$$

*Then $q$ is estimable with $n$ observations if and only if,*

(i) $q$ *is bounded on each* $\mathfrak{F}_n(\mu)$.

(ii) $Q_{F,G}(\alpha)$ *is a polynomial of degree at most $n$ for each $F$, $G$ $\varepsilon$ $\mathfrak{F}$.*

(iii) *For any $f$ $\varepsilon$ $\mathfrak{F}$,*

(4.15) $$q\left(f\psi_m/\|f\psi_m\|\right) \to q(f)$$

*where $\psi_m$ is the indicator of $A_m$.*

The proof is straightforward. That (4.14) is needed in conjunction with the other conditions may be shown by example.

Using this result we can, for instance, drop the boundedness restriction on $h(F_0)$ in Example 4.1 and need only require that the functional $q$ defined by (4.10) be defined on the set of all densities $f$ such that the right hand side of (4.10) is finite.

**Acknowledgment.** We are indebted to L. LeCam for the simple proof of necessity in Theorem 4.1 we have given, and to G. Bredon and C. Moore for an earlier (unused) proof of the same part.

## REFERENCES

[1] GHOSH, J. K. and SINGH, R. (1966). Unbiased estimation of location and scale parameters. *Ann. Math. Statist.* **37** 1671–1675.

[2] HALMOS, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17** 34–43.

[3] HILLE, E. and PHILLIPS, R. S. (1957). Functional analysis and semi-groups. *Amer. Math. Soc. Coll. Publ.* **31**.

[4] LEHMANN, E. L. and SCHEFFÉ, HENRY (1950). Completeness, similar regions and unbiased estimation. *Sankhyā* **10** 305–340.

[5] RENYI, A. (1953). Neue Kriterien zum Vergleich Zweier Stichproben. *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **2** 243–265.

[6] ROSENBLATT, MURRAY (1956). Remark on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

[7] STEFFENSEN, J. (1930). *Some Recent Researches in the Theory of Statistics and Actuarial Science.* Cambridge Univ. Press.

[8] TRÈVES, J. F. (1967). *Topological Vector Spaces, Distributions, and Kernels.* Academic Press, New York.

[9] TUKEY, JOHN W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics.* Stanford Univ. Press.