THE THEORY OF UNBIASED ESTIMATION

By Paul R. Halmos

Syracuse University

1. Summary. Let $F(P)$ be a real valued function defined on a subset $\mathfrak{D}$ of the set $\mathfrak{D}^*$ of all probability distributions on the real line. A function f of n real variables is an unbiased estimate of F if for every system, $X_1, \cdots, X_n$, of independent random variables with the common distribution P, the expectation of $f(X_1 \cdots, X_n)$ exists and equals $F(P)$, for all P in $\mathfrak{D}$. A necessary and sufficient condition for the existence of an unbiased estimate is given (Theorem 1), and the way in which this condition applies to the moments of a distribution is described (Theorem 2). Under the assumptions that this condition is satisfied and that $\mathfrak{D}$ contains all purely discontinuous distributions it is shown that there is a unique symmetric unbiased estimate (Theorem 3); the most general (non symmetric) unbiased estimates are described (Theorem 4); and it is proved that among them the symmetric one is best in the sense of having the least variance (Theorem 5). Thus the classical estimates of the mean and the variance are justified from a new point of view, and also, from the theory, computable estimates of all higher moments are easily derived. It is interesting to note that for n greater than 3 neither the sample nth moment about the sample mean nor any constant multiple thereof is an unbiased estimate of the nth moment about the mean. Attention is called to a paradoxical situation arising in estimating such non linear functions as the square of the first moment.

2. Introduction. Consider the set $\mathfrak{D}^*$ of all probability distributions on the real line. The elements P of $\mathfrak{D}^*$ may be regarded as either set functions $P(E)$, defined for all Borel subsets E of the real line, (probability measures) or monotone non decreasing functions $P(x)$ of a real variable x, (cumulative distribution functions). Suppose that $F = F(P)$ is a real numerically valued function of distributions. For example $F(P)$ may be the expectation or the standard deviation of the distribution P, or it may be the amount of probability P assigns to some fixed set E_0. The problem of unbiased estimation is to find a function (statistic) of a sample of n from a population with distribution P, in such a way that the expected value of this function is equal to the value of $F(P)$ identically in P. More precisely, if $F(P)$ is defined on a subset $\mathfrak{D}$ of $\mathfrak{D}^*$, then an unbiased estimate of order n over $\mathfrak{D}$ is a real valued function $f = f(x_1 \cdots x_n)$ of n real variables, which is such that for every system $X_1, \cdots, X_n$ of independent random variables with the common distribution P (belonging to $\mathfrak{D}$), the expected value $E\{f(X_1, \cdots, X_n)\}$ exists and is equal to $F(P)$.

The problems posed in this paper are the following. (I) Which functions $F(P)$ admit an unbiased estimate? (II) What are all possible unbiased estimates of a given function $F(P)$? (III) Is there a reasonable definition of "best unbiased estimate" which enables one to select from all unbiased estimates of a fixed function $F(P)$ a unique best one?[1]

I shall present below a complete solution of these problems, under the assumption that the domain of estimation, $\mathfrak{D}$, is sufficiently large. The results also shed light on some classical concepts. It is possible, for instance, to exhibit computable unbiased estimates for all moments of a distribution about its expected value, and to prove that the known estimates of the expectation and the variance are essentially unique.

The vague concept of sufficiently large estimation domain $\mathfrak{D}$ is easily made precise. For any Borel set E on the real line let $\mathfrak{D}*(E)$ be the set of all those distributions which assign the probability 1 to some finite subset of E. Thus, for example, if E consists of exactly two points then $\mathfrak{D}*(E)$ is the set of all possible probability distributions in a dichotomy. A subset $\mathfrak{D}$ of $\mathfrak{D}^*$ will be said to be finitely closed over E if $\mathfrak{D}*(E) \subseteq \mathfrak{D}$. Finitely closed domains are "sufficiently large."

It is clear that some restriction (from below) on the size of $\mathfrak{D}$ is essential for a discussion of the characterization problem (II) and the uniqueness problem (III). For if, for example, the domain $\mathfrak{D}$ is artificially restricted to contain only one distribution, then there will always be a plethora of completely unrelated and uninteresting solutions of the problem of unbiased estimation, none of which can be said to be preferable to any other one. It is true, however, that the assumption of finite closure is too restrictive. The general problems of unbiased estimation are still unsolved over such interesting and useful domains as the set of all continuous distributions, and the set of all absolutely continuous distributions. There are also more special problems connected with special classes of distributions (e.g. the normal and the rectangular distributions), as well as the general problem of characterizing the domains which are sufficiently large to make a uniqueness theorem possible. I hope to return to these problems in the near future.

3. Existence. A function $F(P)$, defined on a domain $\mathfrak{D} \subseteq \mathfrak{D}^*$, will be called homogeneous over $\mathfrak{D}$, of degree $k = 1, 2, \cdots$, if there exists a real valued function $\varphi = \varphi(x_1, \cdots, x_k)$ of k real variables which is such that for every P in $\mathfrak{D}$ the Lebesgue-Stieltjes integral[2]

$$\int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_k)$$

[1] My interest in these problems stems from conversations and correspondence with Reinhold Baer, who first called my attention to the problem of finding unbiased estimates for the moments about the expected value. The general questions of existence and uniqueness of unbiased estimates were raised explicitly by J. F. Steffensen in a footnote on p. 18 of his book, *Some Recent Researches in the Theory of Statistics and Actuarial Science*, Cambridge Univ. Press, 1930.

[2] All integrals in this paper are to be extended over the entire Euclidean space of indicated dimension.

exists and is equal to $F(P)$, and if the integer k is minimal with respect to the property of the existence of such a representation.

THEOREM 1. *A necessary and sufficient condition that F have an unbiased estimate of order n over $\mathfrak{D}$ is that it be homogeneous over $\mathfrak{D}$ of degree $k \leq n$.*

PROOF. To prove sufficiency, suppose that

$$F(P) = \int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_k)$$

for all P in $\mathfrak{D}$, with $k \leq n$. Define f by

$$f(x_1, \cdots, x_k, x_{k+1}, \cdots, x_n) = \varphi(x_1, \cdots, x_k).$$

Then if $X_1, \cdots, X_n$ are independent random variables with the same distribution P (belonging to $\mathfrak{D}$)

$$\begin{aligned} E\{f(X_1, \cdots, X_n)\} &= \int \cdots \int f(x_1, \cdots, x_n)\, dP(x_1) \cdots dP(x_n) \\ &= \int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_n) \\ &= \int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_k) = F(P). \end{aligned}$$

The necessity of the condition is even more trivial: the definition of an unbiased estimate of order n is such that the existence of one is equivalent to homogeneity of degree $\leq n$.

As a special case, and an important illustration of how the degree is evaluated, consider the moments $F_m = F_m(P)$ of a distribution P about the origin,

$$F_m(P) = \int x^m\, dP(x),$$

and the moments $\bar{F}_m(P)$ about the expected value $F_1(P)$,

$$\bar{F}_m(P) = \int (x - F_1(P))^m\, dP(x).$$

THEOREM 2. *If $\mathfrak{D}$ is any subset of $\mathfrak{D}^*$ contained in the domain of definition of each of the functions $F_1, \cdots, F_r$, and finitely closed over $\{0, 1\}$ (where $\{0, 1\}$ denotes the set containing the two numbers 0 and 1 only), and if $k_1, \cdots, k_r$ are arbitrary non negative integers, then the function*

$$F(P) = F_1^{k_1}(P) \cdots F_r^{k_r}(P)$$

is homogeneous over $\mathfrak{D}$ of degree exactly $k = k_1 + \cdots + k_r$.

PROOF. The representation of F by a k-fold integral,

$$F(P) = \int \cdots \int x_1 \cdots x_{k_1} x_{k_1+1}^2 \cdots x_{k_1+k_2}^2 \cdots x_{k_1+\cdots+k_r}^r\, dP(x_1) \cdots dP(x_k),$$

shows that F is homogeneous of degree $\leq k$. That the degree of F is indeed equal to k is proved as follows. Suppose that

$$F(P) = \int \cdots \int \varphi(x_1, \cdots, x_h) \, dP(x_1) \cdots dP(x_h)$$

for all P in $\mathfrak{D}$. Observe that if P is the singular distribution which assigns probability 1 to the point 1 on the real line then the identity of the two representations of F reduces to $\varphi(1, \cdots, 1) = 1$; similarly assigning the total probability to 0 implies that $\varphi(0, \cdots, 0) = 0$. More generally, choose P so that it assigns the probability p, $(0 \leq p \leq 1)$, to the point 1, and the probability $q = 1 - p$ to 0. It follows that

$$p^k = p^h + p^{h-1}q\,\varphi_1 + \cdots + pq^{h-1}\,\varphi_{h-1},$$

where φ_i is the sum of all $\varphi(x_1, \cdots, x_h)$, over those $h$tuples $(x_1, \cdots, x_h)$ which contain exactly i 0's and $(h - i)$ 1's. If q is replaced by $1 - p$ in the right side of the last equation, the resulting equation is supposed to be satisfied by all p, $0 \leq p \leq 1$. If, however, $h < k$, then the two sides of the equation are polynomials of different degrees; hence $h \geq k$.

Corollary. *If $\mathfrak{D}$ is any subset of $\mathfrak{D}^*$ contained in the domain of definition of the function $\bar{F}_m$ and finitely closed over $\{0, 1\}$ then $\bar{F}_m$ is homogeneous over $\mathfrak{D}$ of degree exactly m and, consequently, it has unbiased estimates over $\mathfrak{D}$ of order n if and only if $m \leq n$.*

Proof. Since

$$\begin{aligned}\bar{F}_m(P) &= \int (x - F_1(P))^m \, dP(x) \\ &= \sum_{j=0}^m (-1)^j \binom{m}{j} F_1^j(P) \int x^{m-j} \, dP(x) \\ &= \sum_{j=0}^m (-1)^j \binom{m}{j} F_1^j(P) F_{m-j}(P),\end{aligned}$$

the conclusions of the corollary are implied by Theorems 1 and 2.

4. Symmetry. Theorem 1 may be regarded as a solution of the existence problem (I). An examination of its proof shows, however, that the estimates there constructed are very unsatisfactory indeed. In the special case $F = F_1$, for instance, the estimate becomes $f(x_1, \cdots, x_n) = x_1$. The first element of a sample of n is, to be sure, an unbiased estimate of the expectation of the distribution, but it is intuitively clear that, since it ignores most of the information at hand, it is not a good one. In order to exhibit the best estimates it becomes necessary to study the symmetric ones. Recall that a function $f = f(x_1, \cdots, x_n)$ is symmetric if it is invariant under all permutations of its arguments. The proof of the main theorem of this section, the theorem of uniqueness for symmetric unbiased estimates, is based on two lemmas.

LEMMA 1. *If $Q = Q(p_1, \cdots, p_n)$ is a homogeneous polynomial of degree $>$ 0 in n real variables, such that whenever $0 \leq p_i \leq 1$, $i = 1, \cdots, n$, and $p_1 + \cdots + p_n = 1$ then $Q(p_1, \cdots, p_n) = 0$, then Q must be identically zero.*

PROOF. (Induction on n.) For $n = 1$ the lemma is trivial. Assume therefore that $n > 1$ and that the lemma is true for $n - 1$. Observe that the hypothesis is equivalent to the vanishing of Q for all systems of non negative arguments (without the restriction $p_1 + \cdots + p_n = 1$), since any such system $\{p_i\}$ can be replaced by $\{p_i/(p_1 + \cdots + p_n)\}$. If in Q the variables $p_1, \cdots, p_{n-1}$ are given any non negative values, then the hypothesis implies that the resulting polynomial in p_n vanishes for all non negative values of p_n, and therefore identically. Consequently the coefficients of the powers of p_n in Q, which are themselves homogeneous polynomials in $p_1, \cdots, p_{n-1}$, vanish for non negative arguments and therefore (by the induction hypothesis) identically.[3]

LEMMA 2. *If $\mathfrak{D}$ is a set of distributions finitely closed over a Borel set E of the real line and if the symmetric function $f(x_1, \cdots, x_n)$ is such that for every distribution P in $\mathfrak{D}$ the Lebesgue-Stieltjes integral*

$$\int \cdots \int f(x_1, \cdots, x_n)\, dP(x_1) \cdots dP(x_n)$$

exists and has the value zero, then $f(x_1, \cdots, x_n) = 0$ whenever $x_i \in E$, $i = 1, \cdots, n$.

PROOF. Consider any point $(x_1^0, \cdots, x_n^0)$ with $x_i^0 \in E$, $i = 1, \cdots, n$, and any distribution P (in $\mathfrak{D}^*(E)$) which assigns the probability 1 to the subset $\{x_1^0, \cdots, x_n^0\}$ of E. If the probability of x_i^0 is p_i, $i = 1, \cdots, n$, then the integral

$$\int \cdots \int f(x_1, \cdots, x_n)\, dP(x_1) \cdots dP(x_n)$$

is a homogeneous polynomial (of degree n) in the n variables $p_1, \cdots, p_n$. The hypotheses of Lemma 1 are satisfied—it follows that this polynomial vanishes identically. The symmetry of f implies that the coefficient of the term $p_1 \cdots p_n$ is exactly $n!f(x_1^0, \cdots, x_n^0)$, thereby establishing the conclusion of the lemma.

If $\varphi = \varphi(x_1, \cdots, x_k)$ is any function of k real variables and if n is a positive integer, $n \geq k$, it is convenient to write

$$\varphi^{[n]} = \varphi^{[n]}(x_1, \cdots, x_n)$$

for the average of the values of φ over all points obtained from $(x_1, \cdots, x_n)$ by extracting ordered subsets of k x's. Thus, for instance,

$$(x_1 x_2)^{[3]} = \tfrac{1}{3}\,(x_1x_2 + x_1x_3 + x_2x_3)$$

and

$$(x_1)^{[n]} = \tfrac{1}{n}\,(x_1 + \cdots + x_n).$$

[3] I am indebted to J. B. Rosser and R. J. Walker for this proof; my original proof of Lemma 1 was more complicated.

THEOREM 3. *Let $\mathfrak{D}$ be a set of distributions finitely closed over a Borel set E of the real line and let F be a homogeneous function of degree k,*

$$F(P) = \int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_k)$$

over $\mathfrak{D}$. If $f(x_1, \cdots, x_n)$ is a symmetric unbiased estimate of F over $\mathfrak{D}$, of order $n \geq k$, then for every point $(x_1, \cdots, x_n)$ with $x_i \in E$, $i = 1, \cdots, n$, $f(x_1, \cdots, x_n)$ is equal to the symmetrized function $\varphi^{[n]}(x_1, \cdots, x_n)$.

PROOF. Observe first that

$$\int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_k)$$

remains invariant if $(x_1, \cdots, x_k)$ is replaced by $(x_{i_1}, \cdots, x_{i_k})$, where $\{i_1, \cdots, i_k\}$ is any subset of $\{1, \cdots, n\}$, since the change is merely a matter of notation. It follows that

$$F(P) = \int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_k)$$

$$= \int \cdots \int \varphi^{[n]}(x_1, \cdots, x_n)\, dP(x_1) \cdots dP(x_n),$$

so that $\varphi^{[n]}$ is indeed an unbiased estimate of F. Since $\varphi^{[n]}$ is also symmetric, $f - \varphi^{[n]}$ satisfies the hypotheses of Lemma 2, and the desired conclusion follows from an application of that lemma.

5. Characterization. For any Borel set E on the real line let $\mathfrak{D}^*(E)$ be the set of all those distributions which assign the probability 0 to the complement of E. Thus, clearly, $\mathfrak{D}_*(E) \subsetneq \mathfrak{D}^*(E)$; if E is the entire real line then $\mathfrak{D}^*(E) = \mathfrak{D}^*$; if E consists of a finite number of points then $\mathfrak{D}_*(E) = \mathfrak{D}^*(E)$.

THEOREM 4. *Let $\mathfrak{D}$ be a set of distributions finitely closed over a Borel set E of the real line and contained in $\mathfrak{D}^*(E)$, and let F be a homogeneous function of degree k,*

$$F(P) = \int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_k)$$

over $\mathfrak{D}$. A necessary and sufficient condition that the function $f = f(x_1, \cdots, x_n)$ be an unbiased estimate of F over $\mathfrak{D}$, of order $n \geq k$, is that the Lebesgue-Stieltjes integral

$$\int \cdots \int f(x_1, \cdots, x_n)\, dP(x_1) \cdots dP(x_n)$$

exist for every P in $\mathfrak{D}$ and that for every point $(x_1, \cdots, x_n)$ with $x_i \in E$, $i = 1, \cdots, n$, the symmetrized function $f^{[n]}(x_1, \cdots, x_n)$ be equal to $\varphi^{[n]}(x_1, \cdots, x_n)$.

PROOF. If f is an unbiased estimate then $f^{[n]}$ is a symmetric unbiased estimate and therefore, by Theorem 3, equal to $\varphi^{[n]}$; the converse follows from the facts that

$$\int \cdots \int f(x_1, \cdots, x_n)\, dP(x_1) \cdots dP(x_n)$$
$$= \int \cdots \int f^{[n]}(x_1, \cdots, x_n)\, dP(x_1) \cdots dP(x_n)$$

and that (as a consequence of the hypothesis $\mathfrak{D} \subseteq \mathfrak{D}^*(E)$) the equality of $f^{[n]}$ and $\varphi^{[n]}$ for points whose coordinates are in E implies the equality of their integrals.

Theorem 4 exhibits all possibilities for unbiased estimates (over domains satisfying the hypotheses). Given a point $(x_1, \cdots, x_n)$, suppose that the number of different points obtained from it by permutations of the coordinates is N. (If the x_i are all different then $N = n!$). An unbiased estimate is obtained if f is defined arbitrarily over $N - 1$ of these points and if its value on the Nth point is chosen so that the identity $f^{[n]} = \varphi^{[n]}$ is satisfied. As long as the arbitrary choices at the (possibly) uncountably infinite point groups are not too wild and not too large (i.e. are such that the resulting function f is measurable and integrable), f will indeed be an unbiased estimate. Typical nonpathological examples of unsymmetric unbiased estimates are weighted averages of the permuted values of $\varphi(x_1, \cdots, x_k)$, similar to the unweighted average $\varphi^{[n]}(x_1, \cdots, x_n)$.

6. Uniqueness. The assumption of symmetry is a rather natural one to require of an estimate: it amounts to requiring that the estimated value should be independent of the order in which the observations are made. Theorems 3 and 4 establish that the concept of symmetry is inherently associated with unbiased estimation and that, under this assumption, there is a unique unbiased estimate (whenever there is one at all). These theorems, therefore, constitute a partial answer to the uniqueness problem (III): symmetry, after all, is a possible interpretation of "good" estimate. From another point of view the answer to the problem of "best" estimate is contained in the following theorem.

THEOREM 5. *Under the hypotheses of Theorem 4, among all unbiased estimates of*

$$F(P) = \int \cdots \int \varphi(x_1, \cdots, x_k)\, dP(x_1) \cdots dP(x_k)$$

the symmetric one, $\varphi^{[n]}(x_1, \cdots, x_n)$ *is the one with least variance or, equivalently, the least second moment*

$$\int \cdots \int \{\varphi^{[n]}(x_1, \cdots, x_n)\}^2\, dP(x_1) \cdots dP(x_n).$$

PROOF. Observe first that if $X_1, \cdots, X_n$ are independent random variables with the same distribution P then, if f is an unbiased estimate of $F(P)$, the variance of $f(X_1, \cdots, X_n)$ is given by

$$E\{f(X_1, \cdots, X_n)\}^2 - E^2\{f(X_1, \cdots, X_n)\}.$$

Since the second term is the same for all f, namely $F^2(P)$, minimizing the variance is indeed equivalent to minimizing

$$E\{f(X_1, \cdots, X_n)\}^2 = \int \cdots \int \{f(x_1, \cdots, x_n)\}^2 \, dP(x_1) \cdots dP(x_n).$$

This quantity need not be finite even for f's and P's for which $E\{f(X_1, \cdots, X_n)\}$ exists. It will be shown, however, to be minimized by $\varphi^{[n]}$ in the sense that

$$E\{\varphi^{[n]}(X_1, \cdots, X_n)\}^2 \leq E\{f(X_1, \cdots, X_n)\}^2$$

for all unbiased estimates f and all P, and that the inequality actually holds for some P.

For the proof consider any unbiased estimate f of F. For any given point $(x_1, \cdots, x_n)$ suppose that N is the number of different points obtained from it by permutations of the arguments, and denote by f_i, $i = 1, \cdots, N$, the values of f at these points. Since, according to Theorem 4, $f^{[n]} = \varphi^{[n]}$, it follows that

$$(\varphi^{[n]})^2 = \left(\frac{1}{N} \sum_{i=1}^{N} f_i\right)^2 \leq \frac{1}{N} \sum_{i=1}^{n} f_i^2 = (f^2)^{[n]},$$

Hence

$$\begin{aligned} \int \cdots \int \{\varphi^{[n]}(x_1, \cdots, x_n)\}^2 \, dP(x_1) \cdots dP(x_n) & \\ \leq \int \cdots \int \{f^2(x_1, \cdots, x_n)\}^{[n]} \, dP(x_1) \cdots dP(x_n) & \\ = \int \cdots \int f^2(x_1, \cdots, x_n) \, dP(x_1) \cdots dP(x_n). & \end{aligned}$$

This already establishes the minimal property of $\varphi^{[n]}$ in the weak sense.

If the inequality were an equality for all P for which the terms are defined then, by Lemma 2, it would follow that

$$\{\varphi^{[n]}(x_1, \cdots, x_n)\}^2 = \{f^2(x_1, \cdots, x_n)\}^{[n]}$$

for all $(x_1, \cdots, x_n)$. Hence the Schwarz inequality, as applied above to the sum $\frac{1}{N} \sum_{i=1}^{N} f_i$, reduces to an equality; this can happen if and only if $(f_1, \cdots, f_N)$ is proportional to $\left(\frac{1}{N}, \cdots, \frac{1}{N}\right)$, i.e. if and only if all f_i are equal to each other. The validity of this statement for every point is equivalent to the symmetry of f and hence, by Theorem 3, to the statement $f = \varphi^{[n]}$. This concludes the proof of Theorem 5.

7. Concluding remarks. (1) The most obvious estimates of the moments, $F_m(P)$, of a distribution about the origin are the sample moments

$$\frac{1}{n}\sum_{i=1}^{n} x_i^m .$$

Their use is justified by the uniqueness theorems (3, 4, and 5) of this paper. Similarly one might think that the natural estimates of the moments, $\bar{F}_m(P)$, about the expected value $F_1(P)$, are best estimated by the sample moments

$$g_m(x_1, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^m$$

about the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Denote by $f_m(x_1, \cdots, x_n)$ the estimate of $\bar{F}_m(P)$ obtained by expanding $\bar{F}_m(P)$ in terms of the $F_j(P)$, as in the proof of the corollary to Theorem 2, and then estimating each term by the symmetric estimate considered in Theorems 3 and 4. Then an easy calculation shows that

$$f_2(x_1, \cdots, x_n) = \frac{n}{n-1} g_2(x_1, \cdots, x_n)$$

and

$$f_3(x_1, \cdots, x_n) = \frac{n^2}{(n-1)(n-2)} g_3(x_1, \cdots, x_n).$$

(These functions are the classical estimates of $\bar{F}_2$ and $\bar{F}_3$.) For $m > 3$, f_m can still be expressed in terms of g's, but no longer as a constant multiple of g_m. It appears that in general f_m is a linear combination of $g_1, \cdots, g_m$ with coefficients which are rational numbers whose denominators are $(n-1)(n-2)\cdots(n-m+1)$. This fact is another aspect of the non existence of unbiased estimates of order n for $\bar{F}_m$ when $m > n$.

(2) For any Borel set E on the real line denote by $F_E(P)$ the probability, $P(E)$, assigned by P to E. If $\varphi_E(x)$ is the characteristic function of the set E, the representation

$$F_E(P) = \int \varphi_E(x)\, dP(x)$$

shows that $F_E(P)$ is homogeneous of degree 1, and therefore possesses unbiased estimates of all orders. The symmetric unbiased estimate of order n is given, in perfect accordance with intuitive demands, by the function $f_E(x_1, \cdots, x_n)$ whose value is $\frac{1}{n}$ times the number of those coordinates x_i which belong to E.

(3) The situation in estimating such "non linear" functions as $(F_1(P))^2$ is somewhat paradoxical. In the first place it appears strange that there should be

essentially different processes for estimating the expected value and the square of the expected value. (Recall that since

$$(F_1(P))^2 = \int\int x_1 x_2 \, dP(x_1) \, dP(x_2),$$

the symmetric unbiased estimate of $(F_1(P))^2$, of order n, is $(x_1x_2)^{[n]}$.) Consider, for instance, the distribution P which assigns probability $\frac{1}{2}$ to each of the points $+1$ and -1. The symmetric unbiased estimate of order 2 for $F_1(P)$ is $\frac{1}{2}(x_1 + x_2)$, and for $(F_1(P))^2$ it is x_1x_2. Hence in the four possible cases

$$(1, 1), (1, -1), (-1, 1), (-1, -1)$$

the biased, incorrect estimate $\{\frac{1}{2}(x_1 + x_2)\}^2$ for $(F_1(P))^2$ yields

$$1, \quad 0, \quad 0, \quad 1,$$

whereas the unbiased, correct estimate yields

$$1, \quad -1, \quad -1, \quad 1.$$

The actual value of $(F_1(P))^2$ is, of course, 0. Hence it is true in this case that whenever the biased estimate is in error, the unbiased one errs by the same amount. To add insult to injury, the unbiased procedure even yields negative estimates for the essentially non negative quantity $(F_1(P))^2$. These considerations seem to indicate the necessity for caution in using unbiased estimates of "non linear" quantities, such for instance as $\bar{F}_m(P)$.