The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1434–1438

ON UNBIASED ESTIMATION OF DENSITY FUNCTIONS

By A. H. Seheult[1] and C. P. Quesenberry

Durham University and North Carolina State University

1. Introduction and Summary. Let $X^{(n)} = (X_1, \cdots, X_n)$ be a random sample of size n from the distribution of a real-valued random variable X with an absolutely continuous distribution function F and a density function f. Rosenblatt (1956) showed that in this setting there exists no unbiased estimator of f based on the order statistics. His result follows from the fact that the empirical distribution function is not absolutely continuous. He also assumed that f is continuous, but this condition is unnecessary. Rosenblatt's result also arises as a consequence of general results by Bickel and Lehmann (1969) on unbiased estimation in convex families, such as the family of all such F (above).

A number of writers (Kolmogorov (1950), Schmetterer (1960), Ghurye and Olkin (1969)) have obtained unbiased estimators of particular normal-related families as well as for other estimable functions. Washio, Morimoto and Ikeda (1956) considered related questions for the Koopman–Pitman family of densities, and Tate (1959) confined his attention to functions of scale and location parameters. A question arises as to exactly when unbiased—uniform minimum variance unbiased (UMVU)—estimators of density functions exist and when they do not. In a recent publication, Lumel'skii and Sapozhnikov (1969) considered such a question in relation to estimating the density function at a point, whereas, in this paper our definition of unbiasedness requires the estimator to be unbiased at *every* point. The so-called "Bayesian" methods they employ yield estimators for most of the well-known families of distributions as well as for several types of p-dimensional discrete distributions.

In Section 2 we formulate the problem in a fairly general setting and obtain results in terms of unbiased estimators of probability measures (or distribution functions) which always exist. In Section 3 we consider examples to illustrate the theory of the preceding section and in Section 4 give a theorem which generalizes a lemma stated by Ghurye and Olkin (1969) which formalizes the approach used by Schmetterer (1960) for obtaining unbiased estimators of certain types of parametric functions.

2. The Existence theorem. Suppose $(\mathscr{X}, \mathscr{A}, \mu)$ is a σ-finite Euclidean measure space and that $\mathscr{P}$ is a family of probability measures P on $\mathscr{A}$ dominated by μ. Denote the Radon–Nikodym derivative (density) of P with respect to μ by p and the family of such densities by p. Let X be a random variable with distribution $P(\in \mathscr{P})$ and density $p(\in p)$; $X^{(n)} = (X_1, \cdots, X_n)$ be n independent random variables

Received February 9, 1970.

[1] This research was partially supported by the National Research Council of Canada when the author was a Research Associate at Carleton University, Ottawa.

identically distributed as X, and $x^{(n)} = (x_1, \cdots, x_n)$ an observation on $X^{(n)}$. Denote by $\mathscr{X}^{(n)}$ the sample space of observations $x^{(n)}$, by $\mathscr{A}^{(n)}$ the product σ-algebra of subsets of $\mathscr{X}^{(n)}$, and by $Q^{(n)}$ the product measure on $\mathscr{A}^{(n)}$ corresponding to any measure Q on $\mathscr{A}$. Let T be a statistic that maps $(\mathscr{X}^{(n)}, \mathscr{A}^{(n)})$ to a Euclidean space $(\mathscr{T}, \mathscr{B})$ and denote by $Q_*{}^T$ a version of the conditional probability measure (given T) induced on $\mathscr{A}^{(n)}$ from Q on $\mathscr{A}$. Then the class of cylinder sets of the form $\mathscr{X}^{(n-1)} \times A$ for $A \in \mathscr{A}$ is a sub-σ-algebra of $\mathscr{A}^{(n)}$ and we define a probability measure Q^T on $\mathscr{A}$ by $Q^T(A) = Q_*{}^T(\mathscr{X}^{(n-1)} \times A)$ for every $A \in \mathscr{A}$. (Actually, the restriction to Euclidean spaces is unnecessarily restrictive; as long as there exist regular conditional probability measures the results of this section remain valid.)

We shall say that an estimator $\tilde{P} = \tilde{P}(\cdot\,; T)$ is an unbiased—uniform minimum variance unbiased (UMVU)—estimator of P if $\tilde{P}(A; T)$ is an unbiased—UMVU—estimator of $P(A)$ for every $A \in \mathscr{A}$. Hereafter we shall assume all statistics are sufficient for $\mathscr{P}$. The following is an immediate consequence of properties of conditional expectations, sufficiency, and the Rao–Blackwell, Lehmann–Scheffé theory. Particular cases have been used or alluded to by a number of writers including Tate (1959), Barton (1961), Laurent (1963), Pugh (1963), Basu (1964), Folks, *et al* (1965), and Sathe and Varde (1969).

LEMMA 1. *There always exists an unbiased estimator of P given by $\tilde{P} = P^T$, where P^T is any determination of the conditional probability measure given T. Moreover, if T is a complete and sufficient statistic then P^T is the* UMVU *estimator of P.*

By putting $\mathscr{X} = R^k$ and $A = \{(a_1, \cdots, a_k)\colon a_1 \leqq x_1, \cdots, a_k \leqq x_k\}$ we have the particular case of estimating distribution functions. In view of the foregoing we shall consider only estimators $\tilde{P}$ which are themselves probability measures on $\mathscr{A}$ for each $t \in \mathscr{T}$.

We say that an estimator $\tilde{p} = \tilde{p}(\cdot\,; T)$ is an unbiased—UMVU—estimator of p if $\tilde{p}(x; t)$ is $\mathscr{A} \times \mathscr{B}$-measurable and $E\tilde{p}(x; T) = p(x)$ a.e. μ. Note that the usual subscript P (or P_T—the corresponding induced probability measure on $\mathscr{B}$) on E and the qualifying phrase, "for all $P \in \mathscr{P}$", have been suppressed. This practice will be adhered to in the sequel and should present no difficulties.

THEOREM 1. *An unbiased estimator $\tilde{p}$ of p exists if and only if there is an estimator $\tilde{P} = P^T$ of P such that* (a) *for each $t \in \mathscr{T}$ the estimator is absolutely continuous with respect to μ, and* (b) *the Radon–Nikodym derivative is $\mathscr{A} \times \mathscr{B}$-measurable. Moreover, when such a $\tilde{P}$ exists an unbiased estimator of p is given by the Radon–Nikodym derivative, $d\tilde{P}/d\mu$.*

PROOF. If $E\tilde{p} = p$, put $\tilde{P}(A) = \int_A \tilde{p}\,d\mu$; then, $E\tilde{P}(A) = E\int_A \tilde{p}\,d\mu = \int_A (E\tilde{p})\,d\mu = P(A)$, by applying Fubini's Theorem. Hence $\tilde{P}$ is unbiased, and μ-continuous by construction.

Next, if $E\tilde{P} = P$ and $\tilde{P} \ll \mu$, put $\tilde{p} = d\tilde{P}/d\mu$. Then, $\int_A p\,d\mu = E\int_A \tilde{p}\,d\mu = \int_A (E\tilde{p})\,d\mu$. Hence, by the Radon–Nikodym Theorem, $E\tilde{p} = p$ a.e. μ.

THEOREM 2. *If T is a complete and sufficient statistic for $\mathscr{P}$ then a* UMVU *estimator $\tilde{p}$ of p exists if and only if P^T satisfies conditions* (a) *and* (b) *of Theorem 1. Moreover, when such a $\tilde{p}$ exists, it is given by $dP^T/d\mu$.*

PROOF. If $\tilde{p}$ is UMVU it must be a function of T, and, hence, $\int_A \tilde{p}\, d\mu$ is an UMVU estimator of P. Hence, by Lemma 1, $P^T \ll \mu$.

Next, if $P^T \ll \mu$ then $dP^T/d\mu$ is an unbiased estimator of p by Theorem 1. Since it is a function of T it is UMVU.

3. Examples. In this section we consider applications of Theorem 2 to particular families of distributions. Where convenient we will characterize examples by the vector $\{p_\theta, \mathscr{X}, \Theta; T\}$, where p_θ is the density to be estimated, Θ is the parameter space, and T is the complete and sufficient statistic.

When μ is counting measure on a countable set of points of $\mathscr{X}$, $\mathscr{P}$ is a family of discrete distributions and, hence, for *any* statistic T, P^T (for $P \in \mathscr{P}$) is a discrete distribution on a subset of the original set of points of $\mathscr{X}$. Therefore, if T is a complete and sufficient statistic a UMVU estimator of p always exists. For such discrete distributions, the simplest method for obtaining the UMVU estimators is by direct application of the Rao–Blackwell, Lehmann–Scheffé theory. Patil (1963) obtained such estimators of arbitrarily truncated *Noak distributions* of which the binomial, Poisson, negative binomial, and logarithmic distributions are special cases. The reader is referred to Patil's paper for these results. For the remainder of this section μ will be taken to be Lebesgue measure of the appropriate dimension.

EXAMPLE 3.1. *The family of all k-dimensional absolutely continuous distributions.* In this setting the k-dimensional empirical distribution function is the UMVU estimator of the unknown distribution function. Hence, by Theorem 2, no unbiased estimator of the density exists. Rosenblatt (1956) obtained the result for $k = 1$ under the unnecessary additional assumption that the unknown density function should be continuous.

EXAMPLE 3.2. *Truncation distributions.* This family of distributions includes two types given by

$$\text{(3.1)} \qquad \text{Type I: } \{k_1(\theta)h_1(x)I_B(x),\ R,\ (a, b);\ Y_1\}$$
$$\text{Type II: } \{k_2(\theta)h_2(x)I_A(x),\ R,\ (a, b);\ Y_n\},$$

where the interval (a, b) is either finite, semi-infinite or infinite, and k_1 and h_1, and k_2 and h_2 satisfy

$$\text{(3.2)} \qquad 1/k_1(\theta) = \int_\theta^b h_1(x)\,dx, \qquad \text{for all} \quad \theta \in (a, b);$$
$$1/k_2(\theta) = \int_a^\theta h_2(x)\,dx, \qquad \text{for all} \quad \theta \in (a, b).$$

$I_A(x)$ and $I_B(x)$ are the indicator functions of the open intervals (a, θ) and (θ, b), and the statistics Y_1 and Y_n are the smallest and largest order statistics, respectively. Tate (1959) has obtained the following UMVU distribution function estimators:

$$\text{(3.3)} \qquad \text{Type I: } \tilde{F}(x;\ Y_1) = 0, \qquad a < x < Y_1 < b,$$
$$= \frac{1}{n} + \left(1 - \frac{1}{n}\right)\frac{\int_{Y_1}^x h_1(u)\,du}{\int_{Y_1}^b h_1(u)\,du}, \quad a < Y_1 \leqq x < b.$$

$$\text{(3.4)} \qquad \text{Type II: } \tilde{F}(x;\, Y_n) = \left(1-\frac{1}{n}\right)\frac{\int_a^x h_2(u)\,du}{\int_a^{Y_n} h_2(u)\,du}, \qquad a < x < Y_n < b,$$
$$= 1, \qquad a < Y_n \leqq x < b.$$

Both of the above estimators are mixed distributions with a jump of $1/n$ at $x = Y_1$ in the first case and at $x = Y_n$ in the second case. Hence, by Theorem 2, there exists no MVU estimator of the density function. As a particular case consider the uniform distribution on $(0, \theta)$. The Type II estimator specializes to:

$$\text{(3.5)} \qquad \tilde{F}(x;\, Y_n) = (n-1)x/nY_n, \qquad 0 < x < Y_n < \infty,$$
$$= 1, \qquad 0 < Y_n \leqq x < \infty.$$

(It is worthwhile to note that, for $n > 1$, $(n-1)/nY_n$ is the UMVU estimator of $1/\theta$).

4. Unbiased estimation of certain estimable functions. Let Φ be the class of functions ϕ of the form $\phi(P) = E_P h(X)$ such that ϕ exists for each $P \in \mathscr{P}$. The following is a direct generalization of Lemma 1 in Ghurye and Olkin (1969). These writers require the existence of an unbiased estimator of the density; a condition that is unnecessarily restrictive.

THEOREM 3. *An unbiased estimator of* $\phi \in \Phi$ *is given by* $\phi(P^T)$. *Moreover, if* T *is complete and sufficient (for* $\mathscr{P}$*)* $\phi(P^T)$ *is the* UMVU *estimator of* ϕ.

Consider the following examples to illustrate Theorem 3 when no unbiased estimator of the density function exists.

EXAMPLE 4.1. Let $k = 1$ in Example 3.1. Then the UMVU estimator of $\phi(F) = E_F X$ is given by $\int_R x\,dF_n(x) = \bar{X}$, the sample mean; where, $F_n(x)$ is the empirical distribution function. This, is of course, a well-known result.

EXAMPLE 4.2. The mean $\theta/2$ of the uniform distribution on $(0, \theta)$, on using $\tilde{F}$ in (3.5), is estimated by

$$(n-1)(nY_n)^{-1}\int_0^{Y_n} x\,dx + Y_n/n = (n+1)Y_n/2n;$$

once again, a well-known result.

Acknowledgment. The authors would like to thank the referee and Peter J. Bickel, Princeton University for drawing their attention to references [8] and [3], respectively.

REFERENCES

[1] BARTON, D. E. (1961). Unbiased estimation of a set of probabilities. *Biometrika.* **48** 227–229.
[2] BASU, A. P. (1964). Estimates of reliability for some distributions useful in life testing. *Technometrics.* **6** 215–219.
[3] BICKEL, P. J. and LEHMANN, E. L. (1969). Unbiased estimation in convex families. *Ann. Math. Statist.* **40** 1523–1536.

[4] FOLKS, J. L., PIERCE, D. A. and STEWART, C. (1965). Estimating the fraction of acceptable product. *Technometrics.* **7** 43–50.

[5] GHURYE, S. G. and OLKIN, I. (1969). Unbiased estimation of some multivariate probability densities and related functions. *Ann. Math. Statist.* **40** 1261–1271.

[6] KOLMOGOROV, A. N. (1950). Unbiased estimates. *Izv. Akad. Nauk SSSR, Ser. Mat.* **14** 303–326. *Amer. Math. Soc. Trans. No.* 98.

[7] LAURENT, A. G. (1963). Conditional distribution of order statistics and distribution of the reduced ith order statistic of the exponential model. *Ann. Math. Statist.* **34** 652–657.

[8] LUMEL'SKII, YA. P. and SAPOZHNIKOV, P. N. (1969). Unbiased estimation of density functions. *Theor. Probability Appl.* **14** 357–364.

[9] PATIL, G. P. (1963). Minimum variance unbiased estimation and certain problems of additive number theory. *Ann. Math. Statist.* **34** 1050–1056.

[10] PUGH, E. L. (1963). The best estimate of reliability in the exponential case. *Operations Res.* **11** 57–61.

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

[12] SATHE, Y. S. and VARDE, S. D. (1969). On minimum variance unbiased estimation of reliability. *Ann. Math. Statist.* **40** 710–714.

[13] SCHMETTERER, L. (1960). On a problem of J. Neyman and E. Scott. *Ann. Math. Statist.* **31** 656–661.

[14] TATE, R. F. (1959). Unbiased estimation: functions of location and scale parameters. *Ann. Math. Statist.* **30** 341–366.

[15] WASHIO, Y., MORIMOTO, H. and IKEDA, N. (1956). Unbiased estimation based on sufficient statistics. *Bull. Math. Statist.* **6** 69–94.