TECHNICAL REPORT No. 134

November 1967

SOME PARAMETRIC CASES IN WHICH
A PROBABILITY DENSITY IS NOT ESTIMABLE

by

Joseph Putter
SOME PARAMETRIC CASES
IN WHICH A PROBABILITY DENSITY IS NOT ESTIMABLE

Joseph Putter*
University of Wisconsin

1. Introduction

Unbiased estimators of probability densities are known for several parametric families [2, 3, 4]. However, all these estimators are based on samples of size \( n \geq 2 \). This seems to call for an explanation, especially in cases where the estimators are functions of the sample mean only. A density (at a given point of the sample space) is essentially just a given function of the unknown parameter, and regarded from this point of view, is there any reason why more than one observation should be required for its unbiased estimation?

The investigation of the problem was initially motivated by the selection problem described in Section 3 below. The purpose of this note is to answer the question posed above, and also to point out some cases in which the non-estimability of a probability density is not directly connected with sample size. Section 2 gives a general theorem on density estimators based on one observation for a complete family, and a stronger theorem for shift and scale parameter families. A related theorem of a different type is proved for the normal case. In Section 3, an application is given, involving the non-estimability of a selection outcome regardless of sample size.

* On leave from the Volcani Institute of Agricultural Research, Bet-Dagan, Israel.
The intuitive meaning of the requirement \( n \geq 2 \) can perhaps be seen most clearly by considering a complete family of discrete distributions. In this case, the unique unbiased estimator of \( \Pr \{ X = x_k \mid \theta \} \) based on one observation of \( X \) is the indicator of \( x_k \). Theorems 1 and 2 of this note are based on the non-existence of analogous estimators in the continuous case. Theorem 3, however, seems to suggest a different interpretation, possibly in terms of information-content concepts; it also indicates that some integrability conditions used to prove the existence or non-existence of unbiased estimators [1, 3] may be dispensed with in certain cases.

2. Non-existence of unbiased density estimators based on one observation.

**Theorem 1.** Let \( \{ p(x; \theta) \} \) be a complete family of densities w.r.t. a measure \( \mu \); let \( A \) be a set (in the sample space \( S \)) of positive \( \mu \)-measure, containing no points of positive \( \mu \)-measure, such that \( p(x; \theta) > 0 \) for all \( x \in A \) and all \( \theta \) (in the parameter space); and let \( Y \) be a random variable whose probability density is \( p(y; \theta) \) for an unknown \( \theta \). There exists no function \( g(x, y) \) such that \( \mathbb{E} g(x, Y) = p(x; \theta) \) for all \( x \in A \) and all \( \theta \).

**Proof.** The required property of \( g \) can be written in the form

\[
\int_S g(x, y) p(y; \theta) d\mu(y) = p(x; \theta).
\]

Integrating this over any non-null set \( B \subseteq A \), and using the Fubini theorem, we obtain

\[
\mathbb{E} \int_B g(x, Y) d\mu(x) = \int_B p(x; \theta) d\mu(x).
\]

But, by the completeness assumption, this implies that, except for a null set of \( y \)'s,
\[ \int_B g(x,y) \, d\mu(x) = \begin{cases} 1 & \text{for } y \in B \\ 0 & \text{for } y \notin B \end{cases} \]

Hence,

\[ \int \int g(x,y) \, d\mu(x) \, d\mu(y) = 0 \]

for any set \( C \subseteq A \times S - D \), where \( D \) is the diagonal \( \{x=y\} \), and therefore \( g(x,y) = 0 \) for almost all \((x,y)\) with \( x \neq y \) and \( x \in A \), which contradicts the assumptions.

Theorem 1 can be used, in special cases, to prove the non-existence of an unbiased density estimator for any point of the sample space. This will hold whenever an unbiased estimator of the density at a point \( x \) generates unbiased estimators of the density at all points in a neighborhood of \( x \). The two most important cases of this seem to be those of univariate density families with a shift or scale parameter.

**Theorem 2.** Let \( \{p(x;\theta)\} \) be a complete family of probability densities (w.r.t Lebesgue measure) on the real line, satisfying (for all \( x \) and \( \theta \)) either

\begin{enumerate}
  \item[(i)] \( p(x;\theta) = f(x-\theta) \) (shift-parameter family), or
  \item[(ii)] \( p(x;\theta) = \theta f(\theta x) \) (scale-parameter family).
\end{enumerate}

Let \( x_0 \) be any fixed real number, and let \( Y \) be a random variable with the probability density \( p(y;\theta) \) for an unknown \( \theta \). There exists no function \( h(y) \) such that \( Eh(Y) = p(x_0;\theta) \) for all \( \theta \).

**Proof.** If \( h(Y) \) is an unbiased estimator of \( p(x_0;\theta) \), then (for any \( x \)) an unbiased estimator of \( p(x;\theta) \) is given by \( g(x,Y) = h(Y+x_0-x) \) in case (i), and by \( g(x,Y) = \frac{x_0}{x} h\left(\frac{x_0}{x} Y\right) \) in case (ii), contradicting Theorem 1.

Theorem 2 can be easily extended to cover multivariate cases in which the density is of the form \( \Pi_1 p(x_1;\theta_1) \), or \( \Pi_1 p(x_1;\theta) \), etc. This explains
why, for the shift and scale families considered by Tate [3], a sample of size \( n > k \) is necessary for the unbiased estimation of \( \prod_{i=1}^{k} p(x_i; \theta) \).

In the normal case, Theorem 2 can be used to obtain the following result, which connects the estimability of a normal density with the variances of available statistics rather than with sample size.

**Theorem 3.** Denote by \( \phi \) the normal density

\[
\phi(x; \theta, \sigma^2) = (2\pi\sigma^2)^{-1} \exp[-(x-\theta)^2/2\sigma^2].
\]

Let \( x, \sigma^2 \) and \( \tau^2 \) be fixed and known, and let \( Y \) be \( N(\theta, \tau^2) \) with unknown \( \theta \). An unbiased estimator \( f(Y) \) of \( \phi(x; \theta, \sigma^2) \) exists if and only if \( \tau^2 < \sigma^2 \).

**Proof.** If \( \tau^2 < \sigma^2 \), a unique unbiased estimator of \( \phi(x; \theta, \sigma^2) \) is given by \( \phi(x; Y, \sigma^2 - \tau^2) \). (A particular case, with \( \tau^2 = \sigma^2/n \), is Kolmogorov's [2] estimator based on a sample of size \( n \geq 2 \).) The case \( \tau^2 = \sigma^2 \) is covered by Theorem 2. If \( \tau^2 > \sigma^2 \), let \( U \) be \( N(0, \sigma^2 \tau^2/(\tau^2 - \sigma^2)) \) and independent of \( Y \). In the joint distribution of \( Y \) and \( U \), the weighted average \( Z = (\sigma^2 Y + (\tau^2 - \sigma^2)U)/\tau^2 \), which is \( N(\theta, \sigma^2) \), is a sufficient statistic for \( \theta \). Hence, \( g(Z) = E(f(Y) | Z) \) does not depend on \( \theta \). But if \( f(Y) \) is an unbiased estimator of \( \phi(x; \theta, \sigma^2) \), then so is \( g(Z) \), contradicting Theorem 2.

**Remark.** The fact that a square integrable unbiased estimator of \( \phi(x; \theta, \sigma^2) \) based on \( Y \) exists if and only if \( \tau^2 < \sigma^2 \) is a special case of a theorem proved by Ghosh and Lingh ([1], Theorem 2.5). However, the above proof shows that square integrability is not relevant in this case.
3. Application to a selection problem

As indicated by Theorem 3, a normal density may turn out to be non-estimable for reasons unconnected with sample size. This is actually the case in the following application of Theorem 2.

Let \( Y_j \) be the mean of \( n \) independent observations, from a population which is \( N(\mu_j, \sigma^2) \) with known \( \sigma^2 \) \( (j=1, 2) \), and let that population which has yielded the larger value of \( Y_j \) be selected for permanent cultivation, say. The expected mean of the selected population is

\[
M = \mu_1 \Pr \{ Y_1 > Y_2 \} + \mu_2 \Pr \{ Y_2 > Y_1 \},
\]

and we consider the problem of estimating \( M \). Denoting by \( F_j \) the c.d.f. of \( Y_j \), we have

\[
M = \mu_1 \int_{-\infty}^{\infty} F_2(x) dF_1(x) + \mu_2 \int_{-\infty}^{\infty} F_1(x) dF_2(x)
\]

\[
= \int_{-\infty}^{\infty} x[F_1(x)dF_2(x) + F_2(x)dF_1(x)]
\]

\[
- \int_{-\infty}^{\infty} [(x-\mu_1)F_2(x)dF_1(x) + (x-\mu_2)F_1(x)dF_2(x)]
\]

\[= J_1 - J_2, \text{ say.}\]

Now, \( J_1 \) is the expected value of \( \max(Y_1, Y_2) \), and straightforward integration by parts yields \( J_2 = 2n^{-\frac{1}{2}} \phi(0; \mu_1-\mu_2, 2\sigma^2/n) \). By Theorem 2, since \( Y_1 - Y_2 \) is a sufficient statistic for \( \mu_1 - \mu_2 \), there exists no unbiased estimator of \( J_2 \), and therefore also of \( M \).
4. Acknowledgments

The author wishes to express his gratitude to Professors R. Johnson and J. H. Klotz for helpful discussions of the problem.

REFERENCES


