ON ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND MODE[1]

By Emanuel Parzen

Stanford University

0. Introduction. Given a sequence of independent identically distributed random variables $X_1, X_2, \cdots, X_n, \cdots$ with common probability density function $f(x)$, how can one estimate $f(x)$?

The problem of estimation of a probability density function $f(x)$ is interesting for many reasons. As one possible application, we mention the problem of estimating the hazard, or conditional rate of failure, function $f(x)/\{1 - F(x)\}$, where $F(x)$ is the distribution function corresponding to $f(x)$. In this paper we discuss the problem of estimation of a probability density function and the problem of determining the mode of a probability density function. Despite the obvious importance of these problems, we are aware of only two previous papers on the subject of estimation of the probability density function (Rosenblatt [5] and Whittle [6]).

In this paper we show how one may construct a family of estimates of $f(x)$, *and of the mode*, which are consistent and asymptotically normal. We shall see that there are a multitude of possible estimates. We do not examine here the question of which estimate to use.

The problem of estimating a probability density function is in some respects similar to the problem of estimating the spectral density function of a stationary time series; the methods employed here are inspired by the methods used in the treatment of the latter problem (see Parzen [4] for references). The problem of estimating the mode of a probability density function is somewhat similar to the problem of maximum likelihood estimation of a parameter; the methods employed here are inspired by the methods used in the treatment of the latter problem (see Le Cam [2] for references).

1. A class of estimates of the probability density function. Let $X_1, X_2, \cdots, X_n$ be independent random variables identically distributed as a random variable X whose distribution function $F(x) = P[X \leqq x]$ is absolutely continuous,

$$F(x) = \int_{-\infty}^{x} f(x')\, dx', \tag{1.1}$$

with probability density function $f(x)$.

As an estimate of the value $F(x)$ of the distribution function at a given point x, it is natural to take the sample distribution function

---

Received August 29, 1961.

[1] This work was supported in part by Office of Naval Research Contract Nonr-225(21) at Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.