The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1665–1671

ON CHOOSING A DELTA-SEQUENCE

BY MICHAEL WOODROOFE

University of Michigan

1. Introduction. We will be concerned with estimates of the density f from which a random sample $X_1, \cdots, X_n$ has been drawn. In particular, we will consider some modifications of the following type of estimate:

$$f_n(x; t_n) = t_n \int K(t_n(x-y))\, dF_n(y) \tag{1.1}$$
$$= (1/n) \sum_{i=1}^{n} t_n K(t_n(x-X_i)).$$

Here K is a real-valued, bounded, symmetric, absolutely integrable function on $\boldsymbol{R}^1$ for which

$$\int K(y)\, dy = 1; \tag{1.2}$$

t_n is an increasing sequence of positive real numbers for which $t_n \to \infty$ with $t_n = o(n)$ as $n \to \infty$; and F_n denotes the sample distribution function of $X_1, \cdots, X_n$. Such estimates were originally proposed by Rosenblatt [3] and were studied in some detail by Parzen [2].

It is known ([1] and [2]) that the asymptotic behavior of (1.1) depends on the smoothness of f near x and on the sequence t_n. Moreover, the optimal choice of t_n in the sense of minimizing the asymptotic expression for mean square error also depends on the smoothness of f near x and is therefore unknown to the statistician. Here we will consider some modifications of (1.1) which may be described as follows: first estimate f and its derivatives using (1.1) with a t_n sequence as described above; next, use these initial estimates to estimate the optimal t_n sequence, $t_n = \tau_n = \tau_n(f, x, K)$ say, by $\hat{\tau}_n = \hat{\tau}_n(x, K, X_1, \cdots, X_n)$ say; and finally, estimate f by (1.1) with $\hat{\tau}_n$ replacing t_n. Two such modifications are considered; and in both cases we are able to show that under the appropriate regularity conditions

$$E[f_n(x; \hat{\tau}_n) - f(x)]^2 \sim E[f_n(x; \tau_n) - f(x)]^2 \tag{1.3}$$

as $n \to \infty$ where $\sim$ means that the ratio of the two sides tends to one. As may be expected, the proofs of (1.3) constitute rather involved exercises in large sample theory. In order to shorten them, we have developed some special methods and notation which we hope will be of methodological interest in its own right. Briefly, we have developed an algebra of o_E and O_E for handling mean convergence of random variables. This algebra is analogous to the algebra of o_p and O_p.

The paper consists of five sections. In Section two we collect some facts about sample densities of the form (1.1) and state precisely the effect of the smoothness of f on their asymptotic behavior; in Section three we present the algebra of o_E and O_E; and in Sections four and five we present the main theorems.

Received October 3, 1969.