The Annals of Mathematical Statistics
1971, Vol. 42, No. 6, 1815–1842

CURVE ESTIMATES[1]

BY MURRAY ROSENBLATT

University of California, San Diego

1. Introduction. There is a large class of problems in which the estimation of curves arises naturally (see [15], [34]). It is curious that one of the earliest extensive investigations of this type involves the estimation of the spectral density function when sampling from a stationary sequence ([1], [17], [27], [33]). Even though the simple histogram has been used for years, it was only later that the simpler question of estimating a probability density function was dealt with at some length ([26], [25], [9]). Because the final character of the usual results obtained in both problem areas is quite similar, and the arguments are much more transparent in the case of the probability density function, we shall develop the results for the probability density function first. Later some corresponding results for spectra will be given. The similarities and differences in the two areas will be noted. Since the literature is rather extensive by now, any presentation of theory as given can only be a selection of topics and cannot claim to be exhaustive or perhaps even representative. There are a number of attractive open problems that one can suggest solutions to on heuristic grounds. A few of these problems will be examined. In most cases it is clear that one will not use the techniques to be proposed in estimating a density function unless there is a good deal of data (many observations), little a priori information about the density function available, but a great need to get additional information about the density function, even if it is fairly crude.

2. Estimating the probability density function by independent observations. Consider a population with absolutely continuous distribution function $F(x)$ and probability density function $f(x) = F'(x)$. A simple sequence of estimates is determined by the choice of an *integrable bounded weight function* $w(u)$ with

$$\int w(u)\,du = 1 \tag{1}$$

and a *sequence of bandwidths* $b(n) \downarrow 0$ as $n \to \infty$. Notice that this implies that $\int w^2(u)\,du < \infty$. An estimate $f_n(x)$ of $f(x)$

$$f_n(x) = \frac{1}{nb(n)} \sum_{j=1}^{n} w\left(\frac{x - X_j}{b(n)}\right) \tag{2}$$

is determined by a sample $X_1, \cdots, X_n$ of independent observations from the population. If w is chosen to be nonnegative, the estimate $f_n(x)$ itself will be a probability

Received February 11, 1971.

[1] Invited Wald Lectures given at the annual meeting of the Institute of Mathematical Statistics held at Laramie, Wyoming, August 25–27, 1970; research supported by Office of Naval Research Contract N00014–69–A–0200–6016, Task NR–042–217. Reproduction in whole or in part is permitted for any purpose of the United States Government.