DESIGNS FOR REGRESSION PROBLEMS WITH CORRELATED ERRORS

By Jerome Sacks[1] and Donald Ylvisaker[2]

Northwestern University and University of Washington

1. Introduction. The regression model which will underlie the discussion in this paper is as follows: An observation taken at a point t in the closed, bounded interval $[a, b]$ has the form

$$Y(t) = \sum_{i=1}^{k} \beta_i f_i(t) + X(t) \tag{1.1}$$

where $f_1, \cdots, f_k$ are given regression functions; $\beta_1, \cdots, \beta_k$ are unknown parameters; and the error, $X(t)$, is a random variable with $EX(t) = 0$ and $EX^2(t) < \infty$. The model is not quite specified yet since, if more than one observation is taken, say at t_1 and t_2 with both t_1 and t_2 in $[a, b]$, we must say something about the joint behavior of the random variables $X(t_1)$ and $X(t_2)$. Additionally, something should be said about the possibility of "repeating" an observation, i.e., whether two or more observations may be taken at the same t.

The most thoroughly discussed model has been the one which assumes uncorrelated errors with constant variance and repeatable observations. In this case, if n observations are taken at m distinct points $t_1, \cdots, t_m$ with μ_i observations taken at t_i, $\mu_1 + \cdots + \mu_m = n$, the questions of best linear unbiased estimation of $\beta_1, \cdots, \beta_k$, or of some set of linear combinations of the β_i's, have answers which have been known for some time. In recent years, the corresponding design problems, involving optimum choice of the t_i's and μ_i's, have received extensive attention, notably in the papers of Kiefer and Wolfowitz [14] and [15], and Kiefer [10], [11], [12] and [13]. In this paper, we are concerned with such an n observation design problem when the error process has a smooth correlation structure and observations are not repeatable. Our perspective is most easily understood by viewing the error process X as a time series and the experimenter as sampling in time. In this stochastic process context, where an infinite observation set is considered feasible, linear estimation of the β_i's is well understood, due especially to the efforts of Grenander [3], Hájek [4], [5] and [6], and Parzen [17], [18] and [19] (this theory subsumes, of course, the standard estimates for finite observation sets). Within this framework, the design problem can be looked at also from a regret point of view, comparing n-observation designs with observation over all of [0, 1]; this is of interest since the estimators based on interval observation are frequently unknown or, if known, may be difficult to compute; the estimators based on n observations require the inversion of an $n \times n$ matrix and this may be more feasible. The best es-

Received 10 February 1965; revised 18 August 1965.

[1] Research sponsored by NSF grant G-24500.

[2] Research sponsored by NSF grant G-25211.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org