The Annals of Statistics
1987, Vol. 15, No. 1, 1–19

SPECIAL INVITED PAPER

PREDICTION AND DESIGN[1,2]

BY DONALD YLVISAKER

University of California, Los Angeles

In various settings, the observation of a stochastic process at a finite number of locations leads to natural prediction and design questions. General problems of this type are introduced and then related to specific areas of application. A class of processes called G-MAPs is studied with reference to their predictive and other behavior. These processes include many familiar ones and, through being tied to Markov processes, allow a fresh view of prediction. Among other things, G-MAPs stand as reasonably workable possibilities for Bayesian priors in some complex contexts.

0. Introduction. The paper is concerned with problems of prediction and design based on finite observation of a stochastic process. The central questions are posed in Section 1 in an abstract way and include two formulations of observations with error. Adoption of the general context is then supported by a list of particular settings which are subsumed, together with some information on the literature and applications which attend them. In this way we make contact with problems which arise in the computer simulation of product performance, in time series sampling and in geostatistics, for example.

In Section 2 some processes called G-MAPs are introduced as objects of interest in such a study. Material given about them is borrowed in part from a program of Dynkin (1980, 1984) intended to bring the theory of Markov processes to bear on problems of mathematical physics. This is closely related to the work on harness processes in Hammersley (1967) and Williams (1973) and the work of Künsch (1979). The main theme of the presentation here, however, is dictated by the prediction interests expressed in Section 1.

In Section 3, we consider prediction and design directly for G-MAPs, observed with or without error. What emerges is a fresh perspective that stems from a coupling of the observed process with a Markov process. Some technical assistance derives from this connection, and there is the added prospect of determining designs empirically through simulation of the associated process. The main thrust of the paper though is to bring forward some models of stochastic behavior over quite arbitrary sets. Section 4 touches briefly on the case of an unknown mean, while proofs are deferred until Section 5.

1. Prediction and design. This section first fixes the notation and assumptions leading to general prediction and design questions at (1.1)–(1.4). There

Received September 1985; revised May 1986.
[1]This paper was presented as a Special Invited IMS Lecture at the Annual Meeting at Lake Tahoe, August 22, 1984.
[2]This work was supported by NSF Grant MCS 83-01587.
AMS 1980 *subject classifications.* Primary 62M20; secondary 62K05, 60G15.
Key words and phrases. Prediction, design, Markov processes, Gaussian fields, Markov property, Bayesian models.

follows a discussion of some concrete settings in which these problems arise in a natural way.

Begin with a set T of "sites" and let X be a Gaussian process indexed by T, with $EX = 0$ and $\operatorname{Cov} X = G$. (Normality can be replaced by a second-order assumption if the conditional expectations below are given their wide-sense interpretation.) Let $\Gamma(X)$ denote a (normal) random variable determined linearly on X—such a variable will be given specific form, typically $\Gamma(X) = \int X_t \, d\Gamma_t$ for some appropriate integration.

The first problem is the following: given the values X_s, $s \in S$ for a finite set $S \subset T$, form:

$$\text{(1.1)} \qquad \begin{array}{l} \text{(i) the conditioned process } E^S X = E(X | X_s,\ s \in S); \\ \text{(ii) the conditioned variable } E^S\Gamma(X) = E(\Gamma(X) | X_s,\ s \in S). \end{array}$$

Because of normality one can write, for each $t \in T$, $(E^S X)_t = \sum_{s \in S} c_t^S(s) X_s$, say. In conjunction with (1.1)(ii) notice that $E^S\Gamma(X)$ can be obtained from $E^S X$ by regularization: $E^S\Gamma(X) = \Gamma(E^S X)$. Thus if $\Gamma(X) = \int X_t \, d\Gamma_t$, one finds $E^S\Gamma(X) = \sum_{s \in S} X_s \int c_t^S(s) \, d\Gamma_t$. The process $X - E^S X$ will be referred to as the error process, $\Gamma(X - E^S X)$ as an error variable.

EXAMPLE 1.1. Let $T = \{1, 2, \ldots, N\}$ and $S = \{k+1, \ldots, N\}$. If X and G are partitioned in the usual way, $E^S X = (G_{22}^{-1} G_{21}, I)' X_2$. Moreover, if $\Gamma(X) = \gamma' X$, then $E^S\Gamma(X) = \gamma' E^S X$.

EXAMPLE 1.2. For another familiar setting, let X be a Wiener process on $T = [0, \infty)$ with $S = \{s_1, \ldots, s_n\}$, $0 < s_1 < \cdots < s_n \leq 1$. Take $\Gamma(X) = \int_0^1 X_t \, dt$. Now $E^S X$ is a linear interpolation of the observed values of the process on S, with $(E^S X)_t = X_{s_n}$ for $t \geq s_n$. By regularization one predicts $\Gamma(X)$ via the trapezoidal rule $E^S\Gamma(X) = \int_0^1 (E^S X)_t \, dt$.

Two variants of (1.1) will be considered as well. In the first of these the process X is observed with error. Formally, let ε be a white-noise process on T which is independent of X. Now given $X_s + \varepsilon_s$ for $s \in S$, S a finite subset of T, form:

$$\text{(1.2)} \qquad \begin{array}{l} \text{(i) } \mathscr{E}^S X = E(X | X_s + \varepsilon_s,\ s \in S); \\ \text{(ii) } \mathscr{E}^S\Gamma(X) = \Gamma(\mathscr{E}^S X). \end{array}$$

Going further, one may allow that the process X is observed with error and with replication allowed. Suppose for this that $\{\varepsilon^{(j)}\}$ is a sequence of independent white-noise processes on T which are independent of X. Given $X_s + \varepsilon_s^{(1)}, \ldots, X_s + \varepsilon_s^{(n_s)}$ for s in a finite subset S of T form:

$$\text{(1.3)} \qquad \begin{array}{l} \text{(i) } \bar{\mathscr{E}}^S X = E[X | X_s + \varepsilon_s^{(1)}, \ldots, X_s + \varepsilon_s^{(n_s)},\ s \in S]; \\ \text{(ii) } \bar{\mathscr{E}}^S\Gamma(X) = \Gamma(\bar{\mathscr{E}}^S X). \end{array}$$

In (1.2) predictions are linear in the observations while in (1.3) they are linear in $X_s + \bar{\varepsilon}_s^{(n_s)}$, the average observation at s.

For a sample of limited size, it is natural to consider the design questions which accompany prediction. Thus, corresponding to (1.1), determine $S \subset T$

with $|S| = n$ to minimize

(1.4) (i) the error process $X - E^S X$;
(ii) the error variable $\Gamma(X - E^S X)$.

Evidently (1.4)(i) requires some notion of a small process, while an adequate measure of size in (1.4)(ii) is $E\Gamma^2(X - E^S X)$, since the error variable is normal with mean zero. Analogous design problems go with (1.2) and (1.3) in the manner that (1.4) follows (1.1).

EXAMPLE 1.3. Continue with Example 1.2, $S \subset [0, 1]$ and $|S| = n$. For (1.4)(ii) take $s_j = 2j/(2n + 1)$, $j = 1, \ldots, n$. An apparent answer to (1.4)(i) is $s_j = j/n$, $j = 1, \ldots, n$, since the error process is then composed of n independent and identically distributed Brownian bridges.

A direct interpretation of X is that of a Bayesian model of a response function on T. If X is itself observable, then (1.1) represents an up-dated mean, and the error process $X - E^S X$ measures the resulting indeterminacy. Similar remarks apply to (1.2) and (1.3). Passage to (Bayesian) design at (1.4) suggests proper planning of observation sites. Other ways to look at X surface occasionally below.

Here then are some rather more specific settings for these problems. The different contexts brought forward represent a personal view and literature citations should not be thought of as in any way complete.

T is a finite set. Suppose first that X is observable but full observation is deemed to be not feasible. A suggestive possibility of this kind is spelled out.

Let $T = \{-1, 1\}^k$ and regard X as a measure of the performance of a product, where performance depends on k factors each of which has two levels. The objective is to design for and predict unobserved performance. This set-up is pertinent if one measures performance via computer simulation, for example. Measuring actual performance in various environments promises error-prone observations with replication possible and even desirable.

For the given T, functions of particular interest are the grand mean $\Gamma_0(X) = 2^{-k}\sum X_t$ and the rth order interaction terms

$$\Gamma_{j_1,\ldots,j_r}(X) = 2^{-k}\sum_s X_s s_{j_1} \cdots s_{j_r}.$$

With this notation X can be decomposed as

$$\begin{aligned} X_t &= \sum_s X_s \left[2^{-k}\prod_{j=1}^{k}(1 + s_j t_j)\right] \\ &= 2^{-k}\left[\sum_s X_s + \sum_{j=1}^{k}\left(\sum_s X_s s_j\right)t_j + \sum_{j_1 \neq j_2}\left(\sum_s X_s s_{j_1} s_{j_2}\right)t_{j_1}t_{j_2}\right. \\ &\qquad \left. + \cdots + \left(\sum_s X_s s_1 \cdots s_k\right)t_1 \cdots t_k\right] \\ &= \Gamma_0(X) + \sum_{j=1}^{k}\Gamma_j(X)t_j + \sum_{j_1 \neq j_2}\Gamma_{j_1, j_2}(X)t_{j_1}t_{j_2} + \cdots + \Gamma_{1,\ldots,k}(X)t_1 \cdots t_k. \end{aligned}$$

Then specific design problems select S to minimize one of, for example,

$$(1.5)\qquad \begin{array}{ll} \text{(i)} & \max_{r;\, j_1,\ldots,j_r} \alpha_r E\Gamma^2_{j_1,\ldots,j_r}(X - E^S X), \quad \text{for some } \alpha_r \geq 0, \\ \text{(ii)} & \max_t E(X_t - E^S X_t)^2. \end{array}$$

If instead X is observed with error, then $E^S X$ should be replaced by $\mathscr{E}^S X$ or $\bar{\mathscr{E}}^S X$. In the last instance this becomes a more standard analysis of variance set-up. Though nothing has been done here with modelling the error variance, we are close to Bayesian specifications as in Lindley and Smith (1972).

T is an interval of R^1. Papers by Kimeldorf and Wahba (1970a, 1970b) emphasize the form of Bayesian response prediction when X is observed with or without error.

Problems connected with direct observation of X might be viewed as time series sampling with the goal of regression estimation, signal detection, etc. The design aspect of predicting one or more linear functionals was treated in an early series of papers by Sacks and Ylvisaker (1966, 1968, 1970a, 1970b), Wahba (1971) and Hájek and Kimeldorf (1974). Some improvements are found in Eubank, Smith and Smith (1981, 1982). A more general approach to sampling questions is taken in the work of Schoenfelder and Cambanis (1982), Cambanis and Masry (1983) and Bucklew and Cambanis (1984). There is, as well, an excellent survey article by Cambanis (1984).

By allowing X to be a finite-dimensional process observed with error, one meets the general area of Bayesian design for linear models, see for instance Pilz (1981) or the recent work of Chaloner (1984) and the references therein. At the same time infinite-dimensional processes X have been taken as models for regression prediction and design, with an eye towards model robustness issues. After Wahba (1978a), some examples are Steinberg (1985), Wecker and Ansley (1983) and Sacks and Ylvisaker (1985).

The challenging question of design to make the error process small, (1.4)(i), is relatively unexplored. Some results are in Speckman (1976) for the case of an autoregressive X. Some recent work in complexity has dealt with related problems, viewing the distribution of X as a measure on a suitable Banach space. A survey of this is in Wózniakowski (1985).

Other T's. A natural progression carries one to questions of sampling over intervals of higher dimensional space, but technical issues abound in this.

For direct observation of a process X with the goal of designing for optimal prediction of linear functionals, there is some information on convergence rates in Ylvisaker (1975) and Wahba (1978b). Rates of convergence for stratified or random sampling designs can be found in Tubilla (1975) and Schoenfelder (1982).

In a noteworthy paper, Micchelli and Wahba (1979) consider a general T, and provide lower bounds on the behavior of designs at (1.4)(i) and (1.4)(ii) in terms of certain eigenvalues. This is then related to observation of specific processes on R^d.

In the direction of applications, there is an entire literature associated with the technique of Kriging in earth and hydrosciences. Then X may be viewed as a Bayesian model of, say, ore content of a field indexed by T. Interest focusses on the prediction of an average content or of the field itself following a set of observed measurements. Access to this literature is available in the survey of Journal (1983) or in the recent paper by Yakowitz and Szidarovsky (1985).

2. A class of models. Throughout the section X is to be a Gaussian process indexed by the set T, it has mean zero, covariance G, and the finite-dimensional distributions of X are assumed to be nonsingular. Certain processes are singled out and various facts will be established about them. Especially, Dynkin's program which aims to couple Gaussian fields with Markov processes is studied in the context of the problems of Section 1.

Here are the processes of immediate interest.

DEFINITION 2.1. X is called a G-MAP on T if for any $t \in T$ and finite subset S of T,

$$E^S X_t = \sum p_t^S(s) X_s, \tag{2.1}$$

where $p_t^S(s) \geq 0$ and $\sum p_t^S(s) \leq 1$.

Note in this connection that $E^S X_t$ must already be a linear combination of the given values X_s, $s \in S$, as at (1.1), that the nonsingularity of X makes the representation (2.1) unique, and that the notation now emphasizes the fact that these coefficients form a (sub-) distribution on S. The name given such processes is an acronym for Gaussian, Markov-associated process, and this link will come out below.

From (2.1) one sees readily that the correlation function of a G-MAP is nonnegative by conditioning on one-point sets S. The main working relation is the one which follows repeated conditioning:

$$p_t^S(s) = \sum_{u \in U} p_t^U(u) p_u^S(s), \qquad S \subset U. \tag{2.2}$$

Further consequences of (2.1) are extracted in the first theorem.

THEOREM 2.1. *If X is a G-MAP on T, then*:

(i) (Heredity) *X is a G-MAP on T', $T' \subset T$.*

(ii) (Translation) *$-X_s$ and $X_t - X_s$, $t \neq s$, form a G-MAP on T for any $s \in T$.*

(iii) (Conditioning) *The error process $X - E^S X$ is a G-MAP on $T - S$ for any finite $S \subset T$.*

(v) (Decomposition) *T is a disjoint union of sets T_α so that X on T_α and X on T_β are independent G-MAPs if $\alpha \neq \beta$, and if $s, t \in T_\alpha$ for some α, then $G(s, t) > 0$.*

A family of independent normal random variables provides a trivial example of a G-MAP. General examples are produced in the following way. Start with a continuous time Markov process Z on the state space T. Assume Z has a symmetric transition function P_ξ and a finite-valued Green's function:

$$G(s,t) = \int_0^\infty P_\xi(s,t)\,d\xi. \tag{2.3}$$

It turns out that G is nonnegative definite on $T \times T$ so there is a Gaussian process X indexed by T, with mean zero and covariance G.

THEOREM 2.2 [Dynkin (1980)]. *The Gaussian process X with covariance G given by* (2.3) *is a G-MAP on T, and p_t^S is the first-hit distribution of Z on S starting from state t.*

It should be noted that Dynkin refers to X as *the Gaussian field associated with Z*. There are related results on *Harness processes* in Williams (1973), tracing back to Hammersley (1967), and on *Gaussian Markov random fields* in Künsch (1979). In these papers the focus is rather on the construction of processes given by conditional specifications and on their Riesz decompositions. In particular the L_2 or normal aspect present here does not correspond to the deepest part of that theory.

For considering specific versions for examples of the general set-up above, observe that finiteness of the Green's function means transience of Z. Transience can be effected by killing an otherwise recurrent process and this will be a common feature below. Then view Z as a (sub-) Markov process on T or, equivalently, as a Markov process on $T \cup \{\Delta\}$, where Δ is the absorbing death state. For the process X, the death point Δ corresponds to a variable $X_\Delta \equiv 0$. By insisting that X_Δ be observed, the measures p_t^S can be made into probability measures: for $\Delta \in S$, set

$$\bar{p}_t^S(s) = p_t^S(s), \quad \text{for } s \neq \Delta, \qquad \bar{p}_t^S(\Delta) = 1 - \sum_{s \neq \Delta} p_t^S(s). \tag{2.4}$$

For $\Delta \in S \subset U$ one has

$$\bar{p}_t^S(s) = \sum_{u \in U} \bar{p}_t^U(u)\bar{p}_u^S(s), \qquad s \in S, \tag{2.5}$$

by repeated conditioning as at (2.2). In passing we note that Δ is a nontrivial state if T is finite, but is not necessarily one if T is countable.

The geometry of the situation deserves a brief mention here. If one views X_s, $s \in S$, and $X_\Delta \equiv 0$ as the vertices of a convex polygon in L_2, then, according to (2.1), the projection of a vertex X_s on an opposite face is not exterior to that face. From Theorem 2.1(ii) the origin can be translated to any other vertex and thus plays no special role. Edges of these polygons form acute angles at vertices where they intersect, though this is not enough to characterize them. A search for a geometric term to cover the full set-up has not been successful.

EXAMPLE 2.1. Let $T = \{1, \ldots, N\}$ and consider a Markov process Z which jumps uniformly to a new state at exponential rate 1, but which is terminated independently of this motion at exponential rate λ. It is easy to see that G is the covariance matrix of equi-correlated random variables with positive correlation. A corresponding normal vector X is a G-MAP on T.

EXAMPLE 2.2. Let Z be a symmetric stable process on the line with index $\alpha > 1$ and which is independently killed at exponential rate 1. The Green's function is given by

$$G(s, t) = G(|s - t|) = \frac{1}{\pi} \int_0^\infty \frac{\cos|s - t|x}{1 + x^\alpha}\, dx.$$

For $\alpha = 2$ one has the (closed) covariance form $e^{-|s-t|}$ of the Ornstein–Uhlenbeck process on R^1.

EXAMPLE 2.3. If Z is a Brownian motion on the positive half-line which is killed on hitting zero, one has $G(s, t) = A \cdot s \wedge t$ and the Wiener process is seen to be a G-MAP. Here the first-hit distributions of finite sets are easily obtained and provide an "explanation" of the predictions in Example 1.2.

EXAMPLE 2.4. Consider a symmetric random walk on the integers making steps of size ± 1 and ± 2 at rate 1, and which is killed independently at rate λ. An appeal to the equations satisfied by the Green's function [see (5.2) below, for example] leads to a recurrence relation for $G(k) = G(s, s + k)$, $k = 0, 1, \ldots$. One then obtains correlation sequences of the form

$$G(k)/G(0) = \theta\beta_0^k + (1 - \theta)(-\beta_1)^k,$$

where $\theta = \beta_0(1 - \beta_1^2)/[\beta_0(1 - \beta_1^2) + \beta_1(1 - \beta_0^2)]$ and $0 < \beta_1 < \beta_0 < 1$. In particular, each such Green's function is the covariance of a stationary second-order autoregressive sequence. A comparison with the autoregressive correlation sequences as depicted in Figure 3.3(b) of Box and Jenkins (1976) places the Green's function examples in the first quadrant there, though they do not exhaust it [one cannot have $G(2)/G(0) < (G(1)/G(0))^2$, for instance]. Other correlation sequences corresponding to stationary G-MAPs can be explicitly obtained by pushing this program further.

EXAMPLE 2.5. The Brownian-sheet process X on the first quadrant of R is not a G-MAP. The covariance is given by $G((s, \sigma), (t, \tau)) = (s \wedge t)(\sigma \wedge \tau)$ and a computation gives, for example,

$$E\left[X_{(2,2)} | X_{(2,1)}, X_{(1,1)}, X_{(1,2)}\right] = X_{(2,1)} - X_{(1,1)} + X_{(1,2)}.$$

On the other hand, there is much more to be said about certain prediction problems in this case [cf. Dynkin (1980)].

Now consider the converse direction to the result in Theorem 2.2. In fact it is not true that all G-MAPs can be directly obtained by the method explored in Examples 2.1–2.4. In particular, the p_t^S given by (2.1) serve to determine an appropriate Z if T is finite, but need not do so otherwise.

EXAMPLE 2.6. Let X be the Wiener process of Example 2.3 restricted to the set $\{2\} \cup \{1 - 1/n,\ n \geq 1\}$. Then X is a G-MAP [Theorem 2.1(i)], but the "first-hit" distribution $p_2^{\{1-1/n,\ n \leq k\}}$ assigns unit mass to the point $1 - 1/k$ and yet, in conjunction with (2.2), one finds that $p_2^{\{1-1/n,\ n \geq 1\}}$ must assign zero probability to $\{1 - 1/n,\ n \leq k\}$ for every k.

We next state a converse result when T is countable. Thus assuming the existence of a G-MAP satisfying some conditions one has an associated Markov process on T. The present result hardly exhausts the subject, in particular, there is the theory of Markov process construction from first-hit distributions, as in Shih (1971). However, by playing down continuous parameter problems, we can avoid larger technicalities and still have what will be argued to be the right context for the design problems discussed in Section 3.

THEOREM 2.3. *Let X be a G-MAP on a countable set T so that for some $\theta > 0$ and all $t \in T$,*

(i) $\theta \leq E[X_t - E[X_t|X_s,\ s \neq t]]^2 \leq EX_t^2 \leq \theta^{-1}$,

(ii) *the probability measures $\{\bar{p}_t^S,\ \Delta \in S \subset T \cup \{\Delta\},\ |S| < \infty\}$ at (2.4) are tight.*

There is a Markov process Z on T having Green's function $G = \operatorname{Cov} X$.

The upper bound in (i) of Theorem 2.3 seems largely a convenience in the proof. On the other hand, the simple Example 2.6 already points up the need for something akin to the remaining hypotheses it carries. Theorem 2.5 of Künsch (1979) covers similar territory but begins with what should be the transition mechanism of Z, and it is restricted to finitely many possible transitions from any state t. In the course of the proof of Theorem 2.3, the one-step transitions of Z are recovered from the measures $\bar{p}_t^S$ and (ii) is the essential and implicit restriction on them.

From this point on, we suppose X is a G-MAP associated with a Markov process Z as at (2.3). It turns out that X has some simply described Markov properties itself, and remarkably so since T is quite arbitrary. This proceeds as follows. Let S and U be (not necessarily finite) subsets of T.

DEFINITION 2.2. A process X on T is called Markov (S, U) provided X_s and X_u are independent given $\{X_t,\ t \in S \cap U\}$, for $s \in S$ and $u \in U$.

If X is a G-MAP, another way to express conditional independence given $\{X_t,\ t \in S \cap U\}$ is to say that the decomposition of T for $X - E^{S \cap U}X$ afforded by Theorem 2.1(iv) has $T_\alpha \subset S - S \cap U$ or $T_\alpha \subset U - S \cap U$ for all α. The main result is

THEOREM 2.4 [Dynkin (1980)]. *A G-MAP X associated with the Markov process Z is Markov (S, U) if Z cannot move from S to U without passing through $S \cap U$.*

For a simple illustration note that the G-MAP of Example 2.4 is Markov with respect to the sets $\{k|k \leq 1\}$ and $\{k|k \geq 0\}$. At the same time a little reflection makes it apparent that conditional independence after finite observation is much less a feature of processes on, say, $\mathbb{Z}^2$.

It transpires that the collection of covariances G satisfying (2.3) is not closed with respect to addition, i.e., the sum of independent G-MAPs need not be another one. For example, a direct computation shows that the covariance matrices,

$$G_1 = \begin{pmatrix} 1 & \frac{1}{8} & \frac{1}{2} \\ \frac{1}{8} & 1 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & 1 \end{pmatrix}, \qquad G_2 = \begin{pmatrix} 1 & \frac{1}{8} & \frac{1}{4} \\ \frac{1}{8} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & 1 \end{pmatrix},$$

correspond to G-MAPs, but under $G_1 + G_2$ the prediction of X_1 on X_2 and X_3 assigns a small negative weight to X_2. In passing we note that nonclosure is not alleviated by weakening Definition 2.1 to insist only that the coefficient $p_t^S(s)$ satisfy $\Sigma|p_t^S(s)| \leq 1$ (as would arise, for example, by changing the sign of a G-MAP on some fixed subset T' of T). Two suitable covariances are now

$$G_3 = \begin{pmatrix} 1 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & 1 & 0 \\ \frac{3}{4} & 0 & 1 \end{pmatrix}, \qquad G_4 = \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 1 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} \end{pmatrix},$$

but under $G_1 + G_2$, the prediction of X_1 on X_2 and X_3 yields positive coefficients with sum exceeding 1.

While the closure property does not hold for independent summation generally, the special case of observing a G-MAP with independent errors does produce another G-MAP. For this let δ_{s_0} be the indicator function of $s_0 \in T$.

THEOREM 2.5. *If X is a G-MAP on T and $\varepsilon_0, \varepsilon_1$ are i.i.d. centered normal random variables independent of X,*

(i) *$X + \varepsilon_0 \delta_{s_0}$ is a G-MAP on T,*

(ii) *if s_0' is adjoined to T, $X_{s_0}^* = X_{s_0} + \varepsilon_0$, $X_{s_0'}^* = X_{s_0} + \varepsilon_1$, $X_s^* = X_s$ for $s \neq s_0$, then X^* is a G-MAP on $T \cup \{s_0'\}$.*

This result will suffice to cover the models discussed at (1.2) and (1.3), see Propositions 3.1 and 3.2 about this.

To conclude the section it is noted that G-MAPs are nonsmooth processes. There is considerable documentation of this available for Examples 2.2 and 2.3, but another characterization of roughness is given here since T is quite arbitrary. To set the stage let the G-MAP X be observed only with error and with replication possible, (1.3). Given observations $X_s + \varepsilon_s^{(1)}, \ldots, X_s + \varepsilon_s^{(n_s)}$ for $s \in S$, $\Sigma n_s = n$, $\operatorname{var} \varepsilon_s^{(i)} \equiv \theta$, consider the prediction of $\Gamma(X) = \Sigma_{u \in U} \gamma_u X_u$ for a finite

$U \subset T$. Predict X_u by $\Sigma q_u^S(s)(X_s + \bar{\varepsilon}_s)$, and compute

$$E\Big(\sum \gamma_u\big(X_u - \sum q_u^S(s)(X_s + \bar{\varepsilon}_s)\big)\Big)^2$$

$$= E\Big(\sum \gamma_u\big(X_u - \sum q_t^S(s)X_s\big)\Big)^2 + \sum_s \frac{\theta}{n_s}\Big(\sum_u \gamma_u q_u^S(s)\Big)^2. \tag{2.6}$$

If the integer nature of the n_s is ignored in apportioning observations among the sites and the right-hand side of (2.6) is minimized subject only to $\Sigma n_s = n$, we reach the *approximate* design problem:

$$\text{minimize } E\Big(\sum \gamma_u\big(X_u - \sum q_u^S(s)X_s\big)\Big)^2 + \frac{\theta}{n}\Big(\sum_s \Big|\sum_u \gamma_u q_u^S(s)\Big|\Big)^2, \tag{2.7}$$

by choice of S and q_u^S, $u \in U$. With no integer restriction on n_s we may consider that (2.7) applies to any finite $S \subset T$.

THEOREM 2.6. *If X is a G-MAP on T and if S^* is optimum for* (2.7), *then* $S^* \subset U$.

The sense of Theorem 2.6 is that when the design problem is posed in this approximate way one cannot rely on the smoothness of X for extrapolation purposes, contrary to what can transpire for other processes. Here is a simple example to illustrate the last remark. Let the covariance matrix be

$$G = \begin{pmatrix} 1 & \frac{2}{3} & 0 \\ \frac{2}{3} & 1 & \frac{2}{3} \\ 0 & \frac{2}{3} & 1 \end{pmatrix},$$

so X is not a G-MAP, let $\theta/n = \frac{1}{9}$ and consider the prediction of $X_1 + X_3$. The best predictor based on $X_1 + \bar{\varepsilon}_1$ and $X_3 + \bar{\varepsilon}_3$ alone takes $n/2$ observations at each site with the accompanying prediction of $\frac{9}{11}(X_1 + \bar{\varepsilon}_1) + \frac{9}{11}(X_3 + \bar{\varepsilon}_3)$. Then (2.7) reduces to $\frac{4}{11}$. Meanwhile, for a scheme which takes $\frac{1}{3}n$ observations at each of the three sites and predicts by $\frac{1}{2}(X_1 + \bar{\varepsilon}_1) + \frac{1}{2}(X_2 + \bar{\varepsilon}_2) + \frac{1}{2}(X_3 + \bar{\varepsilon}_3)$, one finds that (2.7) has value $\frac{1}{3}$.

3. Prediction and design again. Throughout the section X is a G-MAP on a countable set T, and X satisfies the conditions of Theorem 2.3. Thus the covariance G of X is the Green's function of an associated Markov process Z and (2.3) can be read as

$$\begin{aligned} G(s,t) &= E(\text{time } Z \text{ spends at } s, \text{ starting at } t) \\ &= E_t(\text{time } Z \text{ spends at } s). \end{aligned} \tag{3.1}$$

The perspective here is that a countable T provides a sensible level on which to think about designs. The same point is made in Li (1984), where an interval of possible observation is replaced by a discrete set of sites, lending tractability and leading to some interesting conclusions about optimum designs. After some

general discussion of prediction and prediction error, attention is paid to the error in observation variants mentioned in Section 1.

Let Γ be a probability measure on T and take $\Gamma(X) = \int X_t \, d\Gamma_t = \sum \gamma_t X_t$. If one predicts $\Gamma(X)$ based on observing X on S, then

$$E^S\Gamma(X) = \sum \gamma_t E^S X_t = \sum \gamma_t \sum p_t^S(s) X_s = \sum p_\Gamma^S(s) X_s, \tag{3.2}$$

where p_Γ^S is the first-hit distribution for Z on S when the initial distribution is Γ. Theorem 2.2 appears here as the special case of a degenerate Γ, and one can consider signed measures $\Gamma = \Gamma^+ - \Gamma^-$ in the same way.

To describe the prediction error attached to (3.2), notice that the error process $X - E^S X$ is itself associated with the process Z killed on entrance to the set S. This follows by writing

$$\begin{aligned} E(X - E^S X)_t (X - E^S X)_u &= G(t, u) - \sum \sum p_t^S(s) p_u^S(v) G(s, v) \\ &= G(t, u) - \sum p_t^S(s) G(s, u) \\ &= E_t(\text{time } Z \text{ spends at } u) \\ &\quad - E_t(\text{time } Z \text{ spends at } u \text{ after hitting } S). \end{aligned}$$

Then if τ is a random site with distribution Γ,

$$\begin{aligned} E\Gamma^2(X - E^S X) &= \sum \sum \gamma_t \gamma_u E_t(\text{time } Z \text{ spends at } u, \text{ before } S) \\ &= E_\Gamma(\text{time } Z \text{ spends at } \tau, \text{ before } S). \end{aligned} \tag{3.3}$$

The effect of adding or deleting sites at which to observe is implicit in, for $S \subset U$,

$$\begin{aligned} E\Gamma^2(X - E^S X) &= E\Gamma^2(X - E^U X) + E\Gamma^2(E^U X - E^S X) \\ &= E_\Gamma(\text{time } Z \text{ spends at } \tau, \text{ before } U) \\ &\quad + E_\Gamma(\text{time } Z \text{ spends at } \tau, \text{ before } U \text{ and after } S). \end{aligned} \tag{3.4}$$

Beyond the intuitive appeal for the question of choosing a design set S, (3.3) and (3.4) raise the possibility of determining designs empirically by simulating Z. Recall the set-up with $T = \{-1, 1\}^k$ from Section 1, for example. Let $|S| = n$. Particular processes, random walks say, are easily simulated on T and, for any fixed S, one can measure the worth of S according to criteria suggested at (1.5). To run a full-fledged competition in such a manner, algorithms for testing many subsets are required. This problem is currently being studied.

For consideration of observation with error, suppose first that replicated observations are not possible. Write $X_t^\varepsilon = X_t + \varepsilon_t$, where ε denotes a white-noise process with bounded variance function θ on T.

PROPOSITION 3.1. *If X satisfies the conditions of Theorem 2.3, then so does X^ε. Hence there is a Markov Z^ε with state space T and Green's function G^ε, $G^\varepsilon(s, t) = G(s, t) + \theta_t \delta_{s,t}$.*

Not much is readily said about the resulting process Z^ε, but it is clear that pleasant properties of the original process Z are easily destroyed in this sequence. Take the Wiener process of Example 2.3 confined to integers $t \le n+1$ for instance, and let $\theta_t = t^2$. Then an elementary calculation shows that $(p^\varepsilon)_{n+1}^{\{1,n\}}(1) \to \frac{1}{2}$ as $n \to \infty$ while $(p^\varepsilon)_{n+1}^{\{1,n\}}(n) \to 0$.

When the possibility of replicated observations is added, as at (1.3), modifications are necessary to produce an associated Markov process. Start by taking $T^N = \{(t, j) | t \in T,\ j = 1, \ldots, N\}$ with $X^N_{(t,j)} = X_t + \varepsilon_t^{(j)}$, where the $\varepsilon^{(j)}$ are independent white-noise processes with a common bounded variance function θ on T.

PROPOSITION 3.2. *If X satisfies the conditions of Theorem 2.3, then so does X^N. Hence there is a Markov process Z^N with state space T^N and Green's function G^N, $G^N((s, i), (t, j)) = G(s, t) + \theta_t \delta_{s,t} \delta_{i,j}$.*

Again, substantial information about Z^N has not been accumulated.

4. The unknown mean. Taking the viewpoint that the G-MAP X represents a Bayesian model of a response function, the assumption has so far been that the mean of X is known, hence it is zero with no loss of generality. In these brief remarks we consider observation of $Y = m + X$, where X is subject to the conditions of Theorem 2.3 and m is to be modelled.

Assume X is independent of m, random or not, and assess the performance of a given predictor $\sum c_t^S(s) Y_s$ of Y_t through

$$(4.1)\quad E\Big(Y_t - \sum c_t^S(s) Y_s\Big)^2 = E\Big(X_t - \sum c_t^S(s) X_s\Big)^2 + E\Big(m_t - \sum c_t^S(s) m_s\Big)^2.$$

Supposing first that m is random with known mean (zero) and $\operatorname{Cov} m = G'$, notice that minimizing (4.1) returns one to the initial problem (1.1) with G replaced by $G + G'$. This Bayes set-up has an exact minimax analogue as follows. Let m be an unknown function in the reproducing kernel Hilbert space $H(G')$ associated with a covariance kernel G' and suppose $\|m\|^2_{G'} \le 1$. Then

$$\begin{aligned}\max_{\|m\| \le 1} \Big(m_t - \sum c_t^S(s) m_s\Big)^2 &= \max\Big(m, G'(\cdot, t) - \sum c_t^S(s) G'(\cdot, s)\Big)^2_{G'} \\ &= \Big\|G'(\cdot, t) - \sum c_t^S(s) G'(\cdot, s)\Big\|^2_{G'} \\ &= E\Big(X'_t - \sum c_t^S(s) X'_s\Big)^2,\end{aligned}$$

where X' is a process with mean zero and $\operatorname{Cov} X' = G'$. In particular, under the assumption,

$$(4.2)\qquad \|m\|^2_G \le A,$$

it follows readily that $Y_t^S = \sum p_t^S(s) Y_s$ is a minimax predictor of Y_t. Moreover

regularization obtains and

$$\min_{c_t^S} \max_{\|m\|^2 \le A} E\Big(\Gamma(Y) - \sum c_t^S(s) Y_s\Big)^2 = \max_{\|m\|^2 \le A} E\Gamma^2(Y - Y^S) \tag{4.3}$$
$$= (1 + A) E\Gamma^2(X - E^S X).$$

Hence the mean zero prediction and design problems are exactly duplicated in this case, while maximum prediction errors are inflated.

A parametric modelling of m is common in Kriging applications. Taking $m = \sum_{j=0}^{k} \beta_j f_j$ one finds the maximum mean squared error at (4.1) is finite only if c_t^S satisfies the unbiasedness conditions

$$f_j(t) = \sum c_t^S(s) f_j(s), \qquad j = 0, 1, \ldots, k. \tag{4.4}$$

These conditions have an immediate influence on the resulting minimization of $E(X_t - \sum c_t^S(s) X_s)^2$ at (4.1), and generally one finds some optimal weights to be negative. There is, however, one case of some interest where this is not so, and we record those details.

Let m be an arbitrary unknown constant. The unbiasedness condition (4.4) now asks that $\sum c_t^S(s) = 1$ in order that the maximum value of (4.1) be finite. The subsequent constrained minimization of

$$E\Big(X_t - \sum c_t^S(s) X_s\Big)^2 = E\Big(X_t - \sum p_t^S(s) X_s\Big)^2 + E\Big(\sum \big(p_t^S(s) - c_t^S(s)\big) X_s\Big)^2$$

yields a minimizing vector $\mathbf{c}_t^S = \mathbf{p}_t^S + [(1 - \sum p_t^S(s))/1' G_S^{-1} \mathbf{1}] G_S^{-1} \mathbf{1}$, in which G_S denotes the covariance matrix of X restricted to the set S. In fact some simple manipulations show that the sth element of $G_S^{-1}\mathbf{1}$ is given by

$$\frac{\mathrm{Det}\big(G_{S-\{s\}}\big)}{\mathrm{Det}(G_S)}\Big(1 - \sum p_s^{S-\{s\}}(u)\Big) \ge 0. \tag{4.5}$$

Hence the unbiased weights are at least as large as those from mean zero prediction, while the maximum mean square error is increased by an amount $(\mathbf{1}' G_S^{-1} \mathbf{1})^{-1} (1 - \sum p_t^S(s))^2$.

5. Proofs. We will verify Theorems 2.1, 2.3, 2.5 and 2.6, together with Propositions 3.1 and 3.2.

PROOF OF THEOREM 2.1. (i) follows from Definition 2.1. For (ii) let S be a finite set, $s \in S$, and observe that

$$\begin{aligned} &E\big[X_t - X_s | - X_s, X_u - X_s, u \in S - \{s, t\}\big] \\ &\quad = E\big[X_t - X_s | X_s, X_u, u \in S - \{s, t\}\big] \\ &\quad = \sum_{t \ne s} p_t^{S-\{t\}}(u) X_u - X_s\big(1 - p_t^{S-\{t\}}(s)\big) \\ &\quad = \sum_{t \ne s} p_t^{S-\{t\}}(u)(X_u - X_s) - X_s\Big(1 - \sum_u p_t^{S-\{t\}}(u)\Big). \end{aligned} \tag{5.1}$$

If one conditions (5.1) on $X_u - X_s$, $u \in S - \{s, t\}$, it is evident that (2.1) will

follow for $E[X_t - X_s | X_u - X_s,\ u \in S - \{s, t\}]$ provided (2.1) holds for $E[-X_s | X_u - X_s,\ u \in S - \{s\}]$. Now $E[-X_s | X_u - X_s,\ u \in S - \{s\}] = \Sigma\alpha_u(X_u - X_s)$, where the α's satisfy

$$(-X_s, X_v - X_s) = \sum \alpha_u (X_u - X_s, X_v - X_s), \qquad v \in S - \{s\}.$$

But if G is the covariance matrix of X_u, $u \in S$, then $\alpha_u = (G^{-1}\mathbf{1})_u / \mathbf{1}'G^{-1}\mathbf{1}$ since

$$\begin{aligned}
&\sum_{u \neq s} \frac{(G^{-1}\mathbf{1})_u}{\mathbf{1}'G^{-1}\mathbf{1}} [G(u, v) - G(u, s) - G(v, s) + G(s, s)] \\
&\quad = \sum_{u \in S} \frac{(G^{-1}\mathbf{1})_u}{\mathbf{1}'G^{-1}\mathbf{1}} [G(u, v) - G(u, s) - G(v, s) + G(s, s)] \\
&\quad = -G(v, s) + G(s, s).
\end{aligned}$$

It has been noted at (4.5) that each α_u is nonnegative and so $\Sigma_{u \neq s}\alpha_u = 1 - \alpha_s \leq 1$.

If $t \notin S$ and $U \cap S = \varnothing$, then (iii) follows from

$$\begin{aligned}
&E\left[(X - E^S X)_t \middle| (X - E^S X)_u,\ u \in U\right] \\
&\quad = E\left\{E\left[(X - E^S X)_t \middle| X_v,\ v \in U \cup S\right] \middle| (X - E^S X)_u,\ u \in U\right\} \\
&\quad = E\left[\sum p_t^{U \cup S}(v) X_v - \sum p_t^S(s) X_s \middle| (X - E^S X)_u,\ u \in U\right] \\
&\quad = E\left[\sum p_t^{U \cup S}(v) X_v - \sum_s \sum_v p_t^{U \cup S}(v) p_v^S(s) X_s \middle| (X - E^S X)_u,\ u \in U\right] \\
&\quad = E\left[\sum p_t^{U \cup S}(v)(X - E^S X)_v \middle| (X - E^S X)_u,\ u \in U\right] \\
&\quad = \sum_{u \in U} p_t^{U \cup S}(u)(X - E^S X)_u,
\end{aligned}$$

in which (2.2) is invoked at the third step, and one uses $X - E^S X = 0$ on S.

To see (iv) take $s \sim t$ to mean $p_t^{\{s\}}(s) = G(t, s)/G(s, s) > 0$. The relation $\sim$ is reflexive and symmetric, so suppose $s \sim t$, $t \sim u$ and $s \nsim u$. Then a calculation of $E[X_s | X_t, X_u]$ yields

$$p_s^{\{u, t\}}(u) = \frac{G(s, u)G(t, t) - G(s, t)G(u, t)}{G(u, u)G(t, t) - G^2(u, t)} = 0,$$

contradicting $G(s, t)G(u, t) \neq 0$ with $G(s, u) = 0$. □

PROOF OF THEOREM 2.3. Let t be fixed and take $S_N = \{s_0, s_1, \ldots, s_N\} \nearrow S$, $s_0 \equiv \Delta$. From (2.5)

$$\bar{p}_t^{S_N}(s_j) = \bar{p}_t^{S_{N+1}}(s_{N+1}) \bar{p}_{s_{N+1}}^{S_N}(s_j) + \bar{p}_t^{S_{N+1}}(s_j), \qquad j = 0, 1, \ldots, N,$$

and therefore $\bar{p}_t^{S_N}(s)$ decreases to a limit, $\bar{p}_t^S(s)$, as $N \to \infty$ for all $s \in S$. It is easy to check that the limit is independent of the particular approach of S_N to S and, from (ii), that $\bar{p}_t^S$ is a full probability measure on S.

Set $V_N = (E^{S_N}X)_t = \Sigma_{S_N}\bar{p}_t^{S_N}(s)X_s$ and $W_N = \Sigma_{S_N}\bar{p}_t^{S}(s)X_s$. Observe that $V_n \to_{i.m.} (E^SX)_t$ and that

$$\begin{aligned} E(V_N - W_N)^2 &= E\Big(\sum_{S_N}\big(\bar{p}_t^{S_N}(s) - \bar{p}_t^{S}(s)\big)X_s\Big)^2 \\ &\le \theta^{-1}\Big(\sum\big(\bar{p}_t^{S_N}(s) - \bar{p}_t^{S}(s)\big)\Big)^2 \\ &= \theta^{-1}\Big(1 - \sum_{S_N}\bar{p}_t^{S}(s)\Big)^2 \to 0 \end{aligned}$$

as $N \to \infty$. Thus $W_N \to_{i.m.} (E^SX)_t$ and one may write without ambiguity that $(E^SX)_t = \Sigma_S\bar{p}_t^S(s)X_s$. Nonsingularity at (i) together with repeated conditioning shows that (2.5) holds for countable subsets S and U of $T \cup \{\Delta\}$.

Next consider $\bar{p}_t^{T-\{t\}}$ as a transition mechanism, say $\bar{p}_t^{T-\{t\}}(s) = \pi(t, s)$, for $s \neq t$ with Δ as a death state. Let Z be a regular step process on T determined by π and the speed function $\gamma_t = (E(X_t - E[X_t|X_s,\ s \neq t])^2)^{-1}$ killed on hitting Δ. It is to be shown that the Green's function G' of Z is the covariance G of X.

First note that since $E[X_t - E[X_t|X_s,\ s \neq t]]^2 > 0$ and

$$E\big[E[X_t|X_s,\ s \neq t]|X_t\big] = E\big[\sum\pi(t, s)X_s|X_t\big] = \big[\sum\pi(t, s)\bar{p}_s^{\{t\}}(t)\big]X_t,$$

then $\Sigma\pi(t, s)\bar{p}_s^{\{t\}}(t) = \mu < 1$. Bringing in (2.5),

$$\begin{aligned} \mu &= \sum\pi(t, s_1)\sum\pi(s_1, s_2)\bar{p}_{s_2}^{\{t\}}(t) = \cdots \\ &= \sum_{s_1 \neq t}\pi(t, s_1)\pi(s_1, t) + \cdots \\ &\quad + \sum_{s_i \neq t}\pi(t, s_1)\pi(s_1, s_2)\cdots\pi(s_{n-1}, t) \\ &\quad + \sum_{s_i \neq t}\pi(t, s_1)\pi(s_1, s_2)\cdots\pi(s_{n-1}, s_n)\bar{p}_{s_n}^{\{t\}}(t), \end{aligned}$$

and therefore eventual return of Z to t is uncertain for any $t \in T$. It follows that the Green's function G' of Z is finite everywhere and satisfies the equations

$$\text{(5.2)} \qquad \begin{cases} G'(t, t) = \gamma_t^{-1} + \sum\pi(t, u)G'(u, t), \\ G'(t, s) = \sum\pi(t, u)G'(u, s), \qquad s \neq t. \end{cases}$$

In fact G satisfies these same equations (repeated conditioning) so it remains to be seen that G is a minimal solution to (5.2).

Begin by iterating with G in (5.2). Thus for the first equation

$$\begin{aligned} G(t, t) = \gamma_t^{-1} &+ \sum\pi(t, s_1)\pi(s_1, t)G(t, t) \\ &+ \sum\pi(t, s_1)\pi(s_1, s_2)\pi(s_2, t)G(t, t) + \cdots \\ &+ \sum_{s_i \neq t}\pi(t, s_1)\cdots\pi(s_{n-1}, t)G(t, t) \\ &+ \sum_{s_i \neq t}\pi(t, s_1)\cdots\pi(s_{n-1}, s_n)G(s_n, t), \end{aligned} \qquad \text{(5.3)}$$

and for the second,

$$G(t,s) = \pi(t,s)G(s,s) + \sum_{s_1 \neq s} \pi(t,s_1)\pi(s_1,s)G(s,s) + \cdots$$

$$+ \sum_{s_i \neq s} \pi(t,s_1)\cdots\pi(s_{n-1},s)G(s,s) \tag{5.4}$$

$$+ \sum_{s_i \neq s} \pi(t,s_1)\cdots\pi(s_{n-1},s_n)G(s_n,s).$$

Let R_n^1 denote the last term in (5.3), R_n^2 the last term in (5.4). It will be seen that R_n^1 and $R_n^2 \to 0$ as $n \to \infty$ for all t and s.

Consider R_n^1 and let $\varepsilon > 0$ be given. Determine a finite set of points $\eta = \eta(\varepsilon, t)$ so that $\bar{p}_t^S(\eta^C) < \varepsilon$ for all finite S, by assumption (ii). If $s \notin \eta$ one has $\bar{p}_t^{\{s\}}(s) = G(s,t)/G(s,s) < \varepsilon$ and thus $G(s,t) < \varepsilon\theta^{-1}$. If $s \in \eta$ then $G(s,t) < \theta^{-1}$. Observe that

$$R_n^1 \leq \varepsilon\theta^{-1} + \theta^{-1} \sum_{\substack{s_i \neq t \\ s_n \in \eta}} \pi(t,s_1)\cdots\pi(s_{n-1},s_n)$$

$$\leq \varepsilon\theta^{-1} + \theta^{-1}P_t[Z \text{ is in } \eta \text{ after } n \text{ transitions}]$$

and $\sum P_t[Z$ is in η after n transitions$] = E_t[$number of transitions to $\eta] < \infty$, since uncertain return to any given state ensures that the number of transitions to η has exponential tails. Hence $R_n^1 \leq 2\varepsilon\theta^{-1}$ for sufficiently large n. $R_n^2 \to 0$ as $n \to \infty$ by the same argument. One may now use (5.3) to conclude that $G(t,t) = \gamma_t^{-1}(P_t[Z \text{ does not return to } t])^{-1} = G'(t,t)$ and then, from (5.4), that $G(t,s) = G'(t,s)$ for $s \neq t$. □

PROOF OF THEOREM 2.5. The notation from (2.1) is used for the process X and $\theta = \operatorname{var}\varepsilon_0$. For (i) it suffices to consider

$$E\left[X_t | X_{s_0} + \varepsilon_0, X_s, s \in S - \{s_0\}\right]$$

$$= E\left[\sum p_t^S(s)X_s | X_{s_0} + \varepsilon_0, X_s, s \in S - \{s_0\}\right] \tag{5.5}$$

$$= \sum_{s \neq s_0} p_t^S(s)X_s + p_t^S(s_0)E\left[X_{s_0} | X_{s_0} + \varepsilon_0, X_s, s \in S - \{s_0\}\right]$$

and, in turn,

$$E\left[X_{s_0} | X_{s_0} + \varepsilon_0, X_s, s \in S - \{s_0\}\right] = a\left(X_{s_0} + \varepsilon_0\right) + \sum_{s \neq s_0} c_s X_s, \tag{5.6}$$

say. Condition (5.6) on X_s, $s \in S - \{s_0\}$ to obtain

$$\sum p_{s_0}^{S-\{s_0\}}(s)X_s = a\sum p_{s_0}^{S-\{s_0\}}(s)X_s + \sum_{s \neq s_0} c_s X_s.$$

Therefore the left-hand side of (5.6) is

$$a\left(X_{s_0} + \varepsilon_0\right) + (1-a)\sum p_{s_0}^{S-\{s_0\}}(s)X_s. \tag{5.7}$$

Now multiply (5.6) by $X_{s_0} + \varepsilon_0$ and compute the expectation, using (5.7), to find

$$G(s_0, s_0) = a(G(s_0, s_0) + \theta) + (1 - a)\sum p_{s_0}^{S-\{s_0\}}(s)G(s_0, s).$$

Then it follows that

$$\begin{aligned}(1 - a)&\left[G(s_0, s_0) - \sum p_{s_0}^{S-\{s_0\}}(s)G(s_0, s)\right] \\ &= (1 - a)E\left[X_{s_0} - E\left[X_{s_0}|X_s,\ s \in S - \{s_0\}\right]\right]^2 \\ &= a\theta.\end{aligned}$$

This shows that $0 < a < 1$, hence $E[X_{s_0}|X_{s_0} + \varepsilon_0, X_s,\ s \in S - \{s_0\}]$ satisfies (2.1) as does $E[X_t|X_{s_0} + \varepsilon_0, X_s,\ s \in S - \{s_0\}]$, from (5.5).

To establish (ii) note beyond (5.5)–(5.7) only that

$$\begin{aligned}E&\left[X_t|X_{s_0} + \varepsilon_0, X_{s_0} + \varepsilon_1, X_s,\ s \in S - \{s_0\}\right] \\ &= E\left[X_t|X_{s_0} + \bar{\varepsilon}, X_s,\ s \in S - \{s_0\}\right].\end{aligned}$$

□

PROOF OF THEOREM 2.6. Begin with a general $S \not\subset U$ and observe that (2.7) is larger than

$$E\Big(\sum \gamma_u X_u - E\left[\sum \gamma_u \sum q_u^S(s)X_s|X_u,\ u \in U\right]\Big)^2 + \frac{\theta}{n}\left(\sum_s \Big|\sum_u \gamma_u q_u^S(s)\Big|\right)^2$$

$$= E\Big(\sum \gamma_u X_u - \sum_u \gamma_u \sum_s q_u^S(s)\sum p_s^U(v)X_v\Big)^2 + \frac{\theta}{n}\left(\sum_s \Big|\sum_u q_u^S(s)\Big|\right)^2.$$

Now

$$\sum_v \Big|\sum_s \sum_u \gamma_u q_u^S(s)p_s^U(v)\Big| \le \sum_v \sum_s p_s^U(v)\Big|\sum_u \gamma_u q_u^S(s)\Big| \le \sum_s \Big|\sum_u \gamma_u q_u^S(s)\Big|.$$

Hence (2.7) is larger than

$$E\Big(\sum \gamma_u X_u - \sum \gamma_u \sum q_u^S(s)\sum p_s^U(v)X_v\Big)^2 + \frac{\theta}{n}\left(\sum_v \Big|\sum_s \sum_u \gamma_u q_u^S(s)p_s^U(v)\Big|\right)^2,$$

and S is not optimum. □

PROOF OF PROPOSITION 3.1. According to Theorem 2.5(i), $X + \varepsilon_{s_0}\delta_{s_0}$ is a G-MAP on T if X is, so by induction $X + \sum_{j=0}^n \varepsilon_{s_j}\delta_{s_j}$ is a G-MAP for any finite subset $\{s_0, \ldots, s_N\}$ of T. This is sufficient to establish (2.1) for X^ε:

$$E(X_t^\varepsilon|X_s^\varepsilon,\ s \in S) = \sum p_t^{\varepsilon, S}(s)X_s^\varepsilon,$$

for a (sub-) distribution $p_t^{\varepsilon, S}$ on S. In fact X^ε satisfies the conditions of Theorem 2.3 guaranteeing the existence of Z^ε as asserted. On the one hand, condition (i) is clearly satisfied since this was assumed for X itself. Second, tightness of the measures $\{\bar{p}_t^{\varepsilon, S}\}$ follows by induction. Especially, in the step X goes to $X + \varepsilon_{s_0}\delta_{s_0}$, the resulting $p_t^{0, S}$ can be identified from (5.5)–(5.7) as $ap_t^S + (1 - a)p_t^{S-\{s_0\}}$. Thus no added weight goes to the "tails" and tightness is assured. □

PROOF OF PROPOSITION 3.2. Theorem 2.5 and a simple induction argument show that X^N is a G-MAP on T^N, so it remains to be seen that Theorem 2.3 applies to X^N. Again condition (i) is easily verified so the matter comes down to the tightness condition (ii). Now if for the original process X, $\bar{p}_t^S(B_\eta) < \eta$, for some finite set B_η uniformly in S, then $B_\eta \times \{1, \ldots, N\}$ serves the same role for the process X^N. To see this note from the preceding proof that $E[X_t | X_s + \bar{\varepsilon}_s, s \in S] = \Sigma_s p_t^{*S}(s)(X_s + \bar{\varepsilon}_s)$, where p_t^{*S} is a probability measure satisfying $p_t^{*S}(B_\eta) < \eta$ as well. But $(p^N)_t^S$ is determined from p_t^{*S} by dividing the weights $p_t^{*S}(s)$ equally among those replicates at s which are in the conditioning set. □

Acknowledgments. I have benefitted from discussions of Markov processes with Rick Durrett and Sid Port. Persi Diaconis supplied harness references and, along with Jerry Sacks, has been very supportive in this whole endeavor. Toby Mitchell clarified aspects of my thinking through his talks at two Design Workshops, though he may deny this. Finally, several improvements have resulted from the thoughtful work of the referees.

REFERENCES

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis: forecasting and control.* Holden-Day, San Francisco.

BUCKLEW, J. A. and CAMBANIS, S. (1984). Estimating random integrals from noisy observations: sampling designs and their performance. Technical Report 86, Center for Stochastic Processes, Univ. North Carolina.

CAMBANIS, S. (1984). Sampling designs for time series. *Handbook for Statistics* **5** 337–362. North-Holland, Amsterdam.

CAMBANIS, S. and MASRY, E. (1983). Sampling designs for the detection of signals in noise. *IEEE Trans. Inform. Theory* **IT-29** 83–104.

CHALONER, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* **12** 283–300.

DYNKIN, E. B. (1980). Markov processes and random fields. *Bull. Amer. Math. Soc.* **3** 975–999.

DYNKIN, E. B. (1984). Gaussian and non-Gaussian random fields associated with Markov processes. *J. Funct. Anal.* **55** 344–376.

EUBANK, R. L., SMITH, P. L. and SMITH, P. W. (1981). Uniqueness and eventual uniqueness of optimal designs in some time series models. *Ann. Statist.* **9** 486–493.

EUBANK, R. L., SMITH, P. L. and SMITH, P. W. (1982). A note on optimal and asymptotically optimal designs for certain time series models. *Ann. Statist.* **10** 1295–1301.

HÁJEK, J. and KIMELDORF, G. (1974). Regression designs in autoregressive stochastic processes. *Ann. Statist.* **2** 520–527.

HAMMERSLEY, J. M. (1967). Harnesses. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **3** 89–117. Univ. California Press.

JOURNEL, A. G. (1983). Geostatistics. *Encyclopedia of Statistical Sciences* **3** 424–431. Wiley, New York.

KIMELDORF, G. and WAHBA, G. (1970a). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41** 495–502.

KIMELDORF, G. and WAHBA, G. (1970b). Spline functions and stochastic processes. *Sankhyā Ser. A* **32** 173–180.

KÜNSCH, H. (1979). Gaussian Markov random fields. *J. Fac. Sci. Tokyo Univ.* **26** 53–73.

LI, K.-C. (1984). Robust regression designs when the design space consists of finitely many points. *Ann. Statist.* **12** 269–282.

LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 1–41.

MICCHELLI, C. A. and WAHBA, G. (1979). Design problems for optimal surface interpolation. Technical Report 565, Dept. Statistics, Univ. Wisconsin.

PILZ, J. (1981). Robust Bayes and minimax-Bayes estimation and design in linear regression. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 163–177.

SACKS, J. and YLVISAKER, D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.* **37** 68–89.

SACKS, J. and YLVISAKER, D. (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Statist.* **39** 49–69.

SACKS, J. and YLVISAKER, D. (1970a). Designs for regression problems with correlated errors III. *Ann. Math. Statist.* **41** 2057–2074.

SACKS, J. and YLVISAKER, D. (1970b). Statistical designs and integral approximation. *Proc. Twelfth Bien. Sem. Canad. Math. Cong.* 115–136. Canad. Math. Cong., Montreal.

SACKS, J. and YLVISAKER, D. (1985). Model robust design in regression: Bayes theory. In *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 667–679. Wadsworth, Monterey, Calif.

SCHOENFELDER, C. (1982). Random designs for estimating integrals of stochastic processes: asymptotics. Technical Report 6, Center for Stochastic Processes, Univ. North Carolina.

SCHOENFELDER, C. and CAMBANIS, S. (1982). Random designs for estimating integrals of stochastic processes. *Ann. Statist.* **10** 526–538.

SHIH, C. T. (1971). Construction of Markov processes from hitting distributions II. *Ann. Math. Statist.* **42** 97–114.

SPECKMAN, P. (1976). Approximation of autoregressive Gaussian processes. Ph.D. dissertation, Dept. Mathematics, Univ. California, Los Angeles.

STEINBERG, D. M. (1985). Model robust response surface designs: Scaling two-level factorials. *Biometrika* **72** 513–526.

TUBILLA, A. (1975). Error convergence rates for estimates of multidimensional integrals of random functions. Ph.D. dissertation, Dept. Statistics, Stanford Univ.

WAHBA, G. (1971). On the regression design problem of Sacks and Ylvisaker. *Ann. Math. Statist.* **42** 1035–1043.

WAHBA, G. (1978a). Interpolating surfaces: high order convergence rates and their associated designs, with application to X-ray image reconstruction. Technical Report 523, Dept. Statistics, Univ. Wisconsin.

WAHBA, G. (1978b). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

WECKER, W. E. and ANSLEY, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *J. Amer. Statist. Assoc.* **78** 81–89.

WILLIAMS, D. A. (1973). Some basic theorems on harnesses. In *Stochastic Analysis* (*A Tribute to the Memory of Rollo Davidson*) (D. G. Kendall and E. F. Harding, eds.) 349–363. Wiley, London.

WÓZNIAKOWSKI, H. (1985). A survey of information-based complexity. *J. Complexity* **1** 11–44.

YAKOWITZ, S. J. and SZIDAROVSKY, F. (1985). A comparison of Kriging and nonparametric regression methods. *J. Multivariate Anal.* **16** 21–53.

YLVISAKER, D. (1975). Designs on random fields. *A Survey on Statistical Design and Linear Models* 593–607. North-Holland, Amsterdam.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, LOS ANGELES
LOS ANGELES, CALIFORNIA 90024