The Annals of Statistics
1982, Vol. 10, No. 2, 526–538

RANDOM DESIGNS FOR ESTIMATING INTEGRALS OF STOCHASTIC PROCESSES[1]

BY CAROL SCHOENFELDER AND STAMATIS CAMBANIS

Duke University and University of North Carolina

The integral of a second-order stochastic process Z over a d-dimensional domain is estimated by a weighted linear combination of observations of Z in a random design. The design sample points are possibly dependent random variables and are independent of the process Z, which may be nonstationary. Necessary and sufficient conditions are obtained for the mean squared error of a random design estimator to converge to zero as the sample size increases towards infinity. Simple random, stratified and systematic sampling designs are considered; an optimal simple random design is obtained for fixed sample size; and the mean squared errors of the estimators from these designs are compared. It is shown, for example, that for any simple random design there is always a better stratified design.

1. Introduction. We consider the problem of estimating the weighted average of a second-order stochastic process $\{Z(t): t \in A\}$, $A \subset \mathbb{R}^d$. The process is assumed to be continuous in quadratic mean with covariance $R_0(s, t)$, mean $m(t)$, and second moment function $R(s, t) = R_0(s, t) + m(s)m(t)$ such that $0 < \int_A R(t, t)\, dt < \infty$. We wish to estimate

$$I = \int_A Z(t)\phi(t)\, dt, \tag{1.1}$$

where it is assumed that $E(I^2)$ is non-zero and finite. Then a linear estimate of I, using weights and observations at random sample points, is given by

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^{n} c_n(X_{in})Z(X_{in}), \tag{1.2}$$

where the sample points $\{X_{in}\}_i$ are (possibly dependent) random variables taking values in A and independent of Z. The accuracy of the approximation is measured by the mean squared error (mse) $e_n^2 = E(\bar{Z}_n - I)^2$. Other measures such as $E|\bar{Z}_n - I|$ are of interest but will not be investigated here.

The sequence of weights c_n and designs $\{X_{in}\}_i$ should be chosen so that as the sample size n tends to infinity the mse e_n^2 should tend to zero. Necessary and sufficient conditions for this are derived in Section 2 under fairly general assumptions. Next we find the weights and designs that minimize the mse for fixed sample size n under the simple random sampling scheme (Section 3) and we discuss stratified and systematic sampling schemes (Sections 4 and 5). In Section 6 these sampling schemes are compared for fixed sample size n. The problem of finding the weights and designs that minimize the mse for stratified and systematic sampling asymptotically, rather than for fixed sample size, is treated in Schoenfelder (1978).

In the available literature it is assumed that each sample point X_{in} is uniformly distributed over its range, A is bounded, $\phi = 1/|A|$, $c_n = 1$, and Z is wide-sense stationary. Zubrzycki (1958) obtains expressions for the mse under simple random, stratified and

Received November 1979; revised August 1981.

[1] This research was supported by the Air Force Office of Scientific Research under Grant AFOSR-75-2796.

AMS 1970 *subject classification.* Primary 60G00; secondary 62K05.

Key words and phrases. Random designs, estimation of integrals of stochastic processes, simple random sampling, stratified sampling, systematic sampling.

systematic sampling schemes and compares their relative sizes under certain assumptions on the correlation function of Z; Tubilla (1975) obtains asymptotic expressions for the mse under assumptions on the quadratic mean differentiability of Z. Random designs have been studied by many others, including Cochran (1946) and Quenouille (1949), who take A to consist of a finite number of points, and have found application in areas such as krieging (Zubrzycki, 1958; David, 1978), forestry land surveys (Williams, 1956), and quality control (Jowett, 1952).

Here we do not take Z to be stationary, the sample points to be uniformly distributed and the weights to be constant. Instead, we allow the sample points to be nonuniformly distributed and $\bar{Z}_n$ to assume the slightly more general form (1.2). Then we use the variability of Z in the estimation of I and obtain better estimators, when Z is not stationary, than those generally used. Under certain conditions, we show that when Z is stationary, the choice of weight and design generally used is best.

It should be noted that for any random design there always exists a better nonrandom design (in the mse sense). Optimal nonrandom designs are often difficult to obtain (Sacks and Ylvisaker, 1970). Optimal random designs are seen to be easy to obtain in certain cases (Propositions 3.1, 4.1, Examples 4.2, 4.3, 5.1). Further, for certain nonrandom designs that are reasonable but not optimal, it is possible to find better random designs (Examples 3.3, 6.3).

2. The general case. In this section, under certain mild assumptions, we find necessary and sufficient conditions for the mean square error e_n^2 to converge to zero as $n \to \infty$.

Let G_{in} be the distribution of X_{in} and J_{ijn} the joint distribution of X_{in} and X_{jn}, $i, j = 1, \cdots, n$. Then the mse can be written

$$e_n^2 = E[E\{(\bar{Z}_n - I)^2 | X_{1n}, \cdots X_{nn}\}] = \int_A \int_A R(s, t) c_n(s) c_n(t) J_n(ds, dt)$$

(2.1)

$$- 2 \int_A \int_A R(s, t)\phi(s) c_n(t)\, ds\, dG_n(t) + \int_A \int_A R(s, t)\phi(s)\phi(t)\, ds\, dt$$

where

$$G_n = n^{-1} \textstyle\sum_{i=1}^n G_{in} \quad \text{and} \quad J_n = n^{-2} \sum_{i=1}^n \sum_{j=1}^n J_{ijn}$$

are distributions over A and $A \times A$, respectively, and J_n is symmetric in its arguments with marginal G_n. The averaged marginal distribution G_n measures the intensity of sampling, and the averaged bivariate distribution J_n measures, in part, the dependence of the design on itself. When the sample points are pairwise independent,

$$J_n(s, t) = G_n(s) G_n(t) + \frac{1}{n} \{G_n(\min(s, t)) - \frac{1}{n} \textstyle\sum_{i=1}^n G_{in}(s) G_{in}(t)\}, \tag{2.2}$$

where $\min(s, t) = (\min(s_1, t_1), \cdots, \min(s_d, t_d))$ for $s, t \in \mathbb{R}^d$. When the design is nonrandom (and hence independent of itself as well as having independent sample points), $nG_n(B)$ counts the number of sample points in the set B and

$$J_n(s, t) = G_n(s) G_n(t). \tag{2.3}$$

In fact, as proved in Section 7, the converse is also true.

PROPOSITION 2.1. *$J_n(s, t) = G_n(s)G_n(t)$ on $A \times A$ if and only if the design is nonrandom.*

For simplicity, we concentrate on the important case where the second moment function R of Z is strictly positive definite in the sense that if μ is a signed measure on A such that $\int_A R^{1/2}(t, t)\, d|\mu|(t) < \infty$ and $\int_A \int_A R(s, t)\, d\mu(s)\, d\mu(t) = 0$, then $\mu = 0$.

We show in Theorem 2.3 that when $G_n \Rightarrow G$, $J_n \Rightarrow J$, and $c_n \to c$ in a prescribed manner, then the necessary and sufficient conditions for $e_n^2 \to 0$ as $n \to \infty$ are

$$c(t)\, dG(t) = \phi(t)\, dt \quad \text{on } A, \tag{2.4}$$

$$J(s, t) = G(s)G(t) \quad \text{on } A \times A. \tag{2.5}$$

The natural consistency condition (2.4) is a result of the approximation of the integral (1.1) by the sum (1.2). Condition (2.5) describes an asymptotic average pairwise independence property for the design. It holds for designs whose sample points are pairwise independent; included here are simple random and stratified sampling schemes but not systematic sampling. For nonrandom designs, such an average pairwise independence property holds in addition for each sample size n (Proposition 2.1).

The following lemma, whose proof is given in Section 7, is used in proving Theorem 2.3.

LEMMA 2.2. *Let J be a distribution function on $A \times A$ with marginal G, and c a Borel function on A, satisfying*

$$\int_A R(t, t)c^2(t)\, dG(t) < \infty, \tag{2.6}$$

$$\int_A \int_A f(s)f(t)J(ds, dt) \geq 0 \tag{2.7}$$

for all $f \in \mathscr{L}_2[dG]$. Then

$$\begin{aligned} &\int_A \int_A R(s, t)c(s)c(t)J(ds, dt) - 2\int_A \int_A R(s, t)\phi(s)c(t)\, ds\, dG(t) \\ &\quad + \int_A \int_A R(s, t)\phi(s)\phi(t)\, ds\, dt \\ &= \int_{(0,1]} r dQ(r) + \int_A \int_A R(s, t)\{c(s)\, dG(s) - \phi(s)\, ds\}\{c(t)\, dG(t) - \phi(t)\, dt\}, \end{aligned} \tag{2.8}$$

where Q is nonnegative, of bounded variation, and nondecreasing in r. If, moreover, R is strictly positive definite and c is nonzero a.e. $[dG]$, *then $\int r dQ(r) = 0$ if and only if $J(s, t) = G(s)G(t)$ on $A \times A$.*

When applied to J_n, G_n, and c_n, Lemma 2.2 provides a decomposition of e_n^2 into two nonnegative terms, of which the first, $\int r dQ_n(r)$, vanishes iff the design is nonrandom (cf. Proposition 2.1), and the second vanishes only if the design is random. For simple random, stratified, and systematic sampling designs, Zubrzycki (1958) and Tubilla (1975), in effect, assume

$$c_n(t)\, dG_n(t) = \phi(t)\, dt \quad \text{on } A, \tag{2.9}$$

and thus the mse is given solely by the first term $\int r dQ_n(r)$. Observe that (2.9) implies that $\bar{Z}_n$ is an unbiased estimator of I for any realization of the process Z; i.e., $E\{\bar{Z}_n \mid Z(\cdot)\} = I$.

THEOREM 2.3. *The* mse $e_n^2 \to 0$ *if and only if* (2.4) *and* (2.5) *hold, under the following assumptions*: *R is strictly positive definite; ϕ is nonzero* a.e. [Leb] *on A; J_n converges weakly to some distribution J on $A \times A$ with marginal G;*

$$G(\{t \in A : \exists t_n \to t \text{ s.t. } c_n(t_n) \not\to c(t)\}) = 0$$

for some Borel function c; and (2.6) *holds.*

PROOF. It follows from the assumptions (cf. Billingsley, 1968, page 34) that

$$e^2 = \lim_{n\to\infty} e_n^2 = \int_A\int_A R(s, t)c(s)c(t)J(ds, dt) - 2\int_A\int_A R(s, t)\phi(s)c(t)\, ds\, dG(t)$$
$$+ \int_A\int_A R(s, t)\phi(s)\phi(t)\, ds\, dt.$$

By first considering continuous and bounded f, it can be easily seen that (2.7) holds. Thus by Lemma 2.2, e^2 is given by the RHS of (2.8), the sum of two nonnegative terms. Since R is strictly positive definite, the second of these terms is zero iff $c(s)\, dG(s) = \phi(s)\, ds$, which implies c is nonzero a.e. $[dG]$ since ϕ is nonzero a.e. [Leb]. Then, again by Lemma 2.2, the first of these two terms is zero iff (2.5) holds. □

In Theorem 2.3, when R is not strictly positive definite, then (2.4) and (2.5) are sufficient for $e_n^2 \to 0$ but not necessary.

As was noted, nonrandom designs and designs whose sample points are pairwise independent satisfy condition (2.5). Using (2.3) and (2.2), respectively, we can write e_n^2 as the sum of two nonnegative terms, of which the first is zero and $O(1/n)$, respectively, and the second equals the second term on the RHS of (2.8) with c_n, G_n replacing c, G. Thus, without use of Lemma 2.2, we have that for nonrandom designs and for designs with pairwise independent sample points (assuming $G_n \Rightarrow G$ and c, G and R are as in Theorem 2.3) $e_n^2 \to 0$ iff (2.4) holds.

3. Simple random sampling. In Zubrzycki (1958) and Tubilla (1975), simple random sampling (srs) is defined such that the sample points $\{X_{in}\}$ are independent and identically distributed (iid) with uniform distribution on A, i.e. $dG_n(t) = dt/|A|$; also ϕ is constant, $c_n = |A|\phi$, and thus (2.9) holds. Here we generalize the definition of srs by allowing $\{X_{in}\}$ to be iid G where G does not depend on n, and we choose $c_n \equiv c$ satisfying (2.9) for general ϕ. Under these assumptions the mse for srs is given by

$$e_{r,n}^2 = \frac{1}{n}\left\{\int_A R(t, t)c^2(t)\, dG(t) - \int_A\int_A R(s, t)\phi(s)\phi(t)\, ds\, dt\right\}. \tag{3.1}$$

Using a standard variational argument, we have the following.

PROPOSITION 3.1. *The* mse $e_{r,n}^2$ *is minimized if and only if the sampling distribution* G *has density proportional to*

$$R^{1/2}(t, t)|\phi(t)| \text{ on } A. \tag{3.2}$$

Under this optimal srs *design*

$$e_{r,n}^2 = \frac{1}{n}\left[\left\{\int_A R^{1/2}(t, t)|\phi(t)|\, dt\right\}^2 - \int_A\int_A R(s, t)\phi(s)\phi(t)ds\, dt\right]$$
$$\bar{Z}_n = \frac{a}{n}\sum_{i=1}^n \frac{\operatorname{sgn}\{\phi(X_{in})\}}{R^{1/2}(X_{in}, X_{in})} Z(X_{in}),$$

where $a = \int_A R^{1/2}(t, t)|\phi(t)|\, dt$, *and* $\operatorname{sgn}(t) = 1$ *for* $t \geq 0$, $\operatorname{sgn}(t) = -1$ *for* $t < 0$.

Notice that the sampling density of the optimal srs design does not depend on the sample size. Also, the best design as well as the estimator $\bar{Z}_n$ depend only on the values of R on the diagonal. The complete structure of R enters only in the value of the mse.

EXAMPLE 3.2. *Stationary process with a trend.* Let A be bounded and Z have mean $m(t)$ and covariance $R_0(s, t) = C(t - s)$ where C is continuous and $\sigma^2 = C(0)$. Then the best srs design is given by

$$g(t) = \{m^2(t) + \sigma^2\}^{1/2} |\phi(t)| / Q,$$

where $Q = \int_A \{m^2(t) + \sigma^2\}^{1/2} |\phi(t)|\, dt$, yielding

$$\bar{Z}_n = \frac{Q}{n} \sum_{i=1}^n \frac{\operatorname{sgn}(\phi(X_{in}))}{\{m^2(X_{in}) + \sigma^2\}^{1/2}} Z(X_{in})$$

$$ne_{r,n}^2 = Q^2 - \left(\int_A m\sigma\right)^2 - \int_A \int_A C(t-s)\phi(s)\phi(t)\, ds\, dt.$$

If $\phi = |A|^{-1}$ and $m(t)$ is constant (no trend), the sampling density of the best srs design is the uniform density over A, yielding the usual estimator ($c_n = 1$) with mse as obtained by Zubrzycki (1958).

EXAMPLE 3.3. *d-dimensional Brownian motion.* Assume that $A = (0, 1]^d$, $\phi = 1$, and that Z has mean zero and covariance $R_0(\mathbf{s}, \mathbf{t}) = \Pi \min(s_i, t_i)$. Define $\Pi(\mathbf{t}) = \Pi\, t_i$. The best srs design has sampling density

$$g(\mathbf{t}) = (3/2)^d\, \Pi^{1/2}(\mathbf{t}),$$

yielding

$$\bar{Z}_n = (2/3)^d \cdot \frac{1}{n} \sum_{i=1}^n \frac{Z(\mathbf{X}_{in})}{\Pi^{1/2}(\mathbf{X}_{in})}, \quad ne_{r,n}^2 = \{(4/9)^d - (1/3)^d\}\sigma^2,$$

a significant departure from the usual design and estimator. We have $ne_{r,n}^2\, \sigma^{-2} \approx .111$ for $d = 1$, $\approx .086$ for $d = 2$, $\approx .051$ for $d = 3$. It may also be seen that $ne_{r,n}^2 \sigma^{-2}$ has a maximum at $d = 1$ and is strictly decreasing in d to zero. The rate of convergence of $e_{r,n}^2$ to zero as $n \to \infty$ is $O(n^{-1})$, while the rate of convergence of the mse to zero for a *nonrandom* product design for stochastic processes of this type is $O(n^{-2/d})$ (Ylvisaker, 1975). This points out the nonoptimality of product designs. The results of Ylvisaker (1975) and especially Wahba (1978) indicate, however, that a superposition of nonrandom product designs may give a better rate.

In concluding this section, we note that the performance of the "optimal" srs design is robust with respect to insufficient knowledge of the second moment function R. Let $e_{r,n}^2(S \mid R)$ be the mse when the design is chosen via (3.2) as if S were the second moment function when R is the true second moment function of Z. Then

$$e_{r,n}^2(S \mid R) = \frac{1}{n} \left\{ \int_A S^{1/2}(t, t) |\phi(t)|\, dt \int_A \frac{R(s, s)}{S^{1/2}(s, s)} |\phi(s)|\, ds - \int_A \int_A R(s, t)\phi(s)\phi(t)\, ds\, dt \right\}$$

and it can be shown that when S is appropriately close to R,

$$d(S, R \mid R) = n |e_{r,n}^2(S \mid R) - e_{r,n}^2(R \mid R)|$$

is close to zero. For details see Schoenfelder (1978).

4. Stratified sampling. Stratified sampling (sts) consists of independently choosing a simple random sample of size $n_i (\geq 1)$ from each A_{im}, $i = 1, \cdots, m$ ($m \leq n$, $\Sigma\, n_i = n$) where $\{A_{im}\}_{i=1}^m$ is a partition of A. Thus in sts the sample points X_{jim} ($j = 1, \cdots, n_i$, $i = 1, \cdots, m$, $\Sigma\, n_i = n$) are chosen independently of each other so that X_{jim} has distribution G_{im} concentrated on A_{im} and so that c_n and $G_n = \Sigma\, (n_i/n) G_{im}$ satisfy (2.9); Zubrzycki (1958) and Tubilla (1975) take $n_i \equiv 1$ with G_{im} the uniform distribution over A_{im} and c_n and ϕ constant. The mse for sts is given by

$$e_{st,n}^2 = \sum_{i=1}^m \left\{ \frac{1}{n_i} \int_{A_{im}} R(t, t) \frac{\phi^2(t)}{g_{im}(t)}\, dt - \int_{A_{im}} \int_{A_{im}} R(s, t)\phi(s)\phi(t)\, ds\, dt \right\}, \tag{4.1}$$

which reduces to (3.1) when $m = 1$. We have the following optimality result which is an immediate consequence of Proposition 3.1.

PROPOSITION 4.1. *For fixed partition* $\{A_{im}\}_i$ *of* A, $e^2_{st,n}$ *is minimized when the sampling distributions* G_{im} *have densities* g_{im} *proportional to* $R^{1/2}(t, t)|\phi(t)|$ *on* A_{im}. *Under this design, with* $Q_{im} = \int_{A_{im}} R^{1/2}(t, t)|\phi(t)|\,dt$, *we have*

$$e^2_{st,n} = \sum_{i=1}^m \left\{\frac{Q^2_{im}}{n_i} - \int_{A_{im}}\int_{A_{im}} R(s, t)\phi(s)\phi(t)\,ds\,dt\right\}$$

$$\bar{Z}_n = \sum_{i=1}^m \frac{Q_{im}}{n_i} \sum_{j=1}^{n_i} Z(X_{jin}) \frac{\operatorname{sgn}\{\phi(X_{jin})\}}{R^{1/2}(X_{jin}, X_{jin})}.$$

EXAMPLE 4.2. *Stationary process with a trend.* Under the assumptions of Example 3.2, the mse under sts is minimized for fixed partition when the sampling density is

$$g_{in}(t) = \frac{1}{Q_{im}} \{m^2(t) + \sigma^2\}^{1/2}|\phi(t)| \quad \text{on } A_{im}, \quad i = 1, \cdots, m$$

where $Q_{im} = \int_{A_{im}} \{m^2(t) + \sigma^2\}^{1/2}|\phi(t)|\,dt$, yielding

$$\bar{Z}_n = \sum_{i=1}^m \frac{Q_{im}}{n_i} \sum_{j=1}^{n_i} \frac{\operatorname{sgn}\{\phi(X_{jin})\}}{\{m^2(X_{jin}) + \sigma^2\}^{1/2}} Z(X_{jin})$$

$$e^2_{st,n} = \sum_{i=1}^m \left[\frac{1}{n_i} Q^2_{im} - \int_{A_{im}}\int_{A_{im}} \{C(t - s) + m(s)m(t)\}\phi(s)\phi(t)\,ds\,dt\right].$$

When the mean $m(t)$ is constant (no trend), $\phi = 1$, $A = [0, 1]$, each A_{in} is an interval, and $n_i = 1$ for all i, the mse can be rewritten

$$e^2_{st,n} = \sum_{i=1}^n \int_{-\infty}^{\infty} \left\{|A_{in}|^2 - \frac{4}{\lambda^2}\sin^2\left(\frac{\lambda|A_{in}|}{2}\right)\right\} dF(\lambda)$$

where F is the spectral distribution corresponding to R. This is minimized when $|A_{in}| = 1/n$ for all i, the choice in the literature.

EXAMPLE 4.3. *Wiener process.* Let $R(s, t) = \min(s, t)$, $A = [0, 1]$, and $\phi = 1$, and assume each A_{in} is an interval and $n_i = 1$ for all i. Then $e^2_{st,n}$ is minimized when g_n is proportional to $t^{1/2}$ on A_{in} for each i, and $\bar{A}_{in} = [x_{i-1,n}, x_{in}\}$ where $x_{0n} = 0$, $x_{nn} = 1$, and for $i = 1, \cdots, n - 1$,

$$\frac{x_{i+1,n}}{x_{in}} = \left(\frac{1}{2} + \cos\theta_i\right)^2, \quad \theta_i = \frac{1}{3}\cos^{-1}\left\{3\left(\frac{x_{i-1,n}}{x_{in}}\right)^2 - 4\left(\frac{x_{i-1,n}}{x_{in}}\right)^{3/2}\right\}.$$

When $n = 10$, the endpoints x_{in} are approximately .000, .116, .217, .316, .414, .512, .610, .708, .805, .903, 1.000.

5. Systematic sampling. Systematic sampling (sys) consists of choosing a simple random sample from the first stratum and then choosing the samples on each remaining stratum by applying some specified transformation on the sample from the first stratum. For ease of notation we consider only the case where a sample of size one is taken from each stratum specifically, consider a partition $\{A_{in}\}_i$ of A and a set of transformations $T_{ijn}: A_{jn} \to A_{in}$, $1 - 1$ and onto, such that $T_{ijn} = T_{ikn}T_{kjn}$ for all i, j, k. The sample point X_{1n} has distribution G_{1n} concentrated on A_{1n} and the remaining sample points are defined by $X_{in} = T_{i1n}X_{1n}$ with X_{in} having distribution G_{in} on A_{in}, $i = 2, \cdots, n$, and with c_n and $G_n = \sum G_{in}/n$ satisfying (2.9). (In the literature G_{in} is the uniform distribution on A_{in}, the transformations are translations, and c_n and ϕ are constant.) The mse for sys is given by

$$e^2_{sy,n} = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n \int_{A_{in}} R(t, T_{jin}t)\phi(t)\frac{\phi(T_{jin}t)}{g_n(T_{jin}t)}\,dt - \int_A\int_A R(s, t)\phi(s)\phi(t)\,ds\,dt. \tag{5.1}$$

It does not appear that in general this expression can be minimized. However, minimization may be feasible in certain very special cases, as the following example illustrates.

EXAMPLE 5.1. Assume that R is stationary and nonnegative on $A = [0, 1)$, $\phi = 1$, $c_n g_n = 1$, $A_{in} = [t_{i-1}, t_i)$, and that $T_{ijn}t = (t - t_{j-1})(\Delta_i/\Delta_j) + t_{i-1}$ is a translation from A_{jn} to A_{in} where $\Delta_i = t_i - t_{i-1}$. Then

$$e^2_{sy,n} = \sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_i}{\Delta_j} \int_{t_{j-1}}^{t_j} \frac{R(T_{ijn}u, u)}{g_{jn}(u)} du - \int_0^1 \int_0^1 R(s, t)\, ds\, dt$$

$$\geq \sum_{j=1}^n \frac{1}{\Delta_j} \left[\int_{t_{j-1}}^{t_j} \left\{ \sum_{i=1}^n \Delta_i R(T_{ijn}u, u) \right\}^{1/2} du \right]^2 - \int_0^1 \int_0^1 R(s, t)\, ds\, dt,$$

with equality if and only if $g_{in}(u)$ is proportional to $\{\sum_{i=1}^n \Delta_i R(T_{ijn}u, u)\}^{1/2}$ on A_{jn}. In particular, if we let $\Delta_i = 1/n$ for all i (the choice in the literature), the mse is minimized over all consistent g_n when $g_n = 1$, the choice in the literature.

6. Comparisons. In this section we establish the intuitively clear result that, in general, sts is better than srs in the mse sense; i.e., $e^2_{st,n} \leq e^2_{r,n}$. We then give conditions under which $e^2_{st,n}$ is larger (smaller) than $e^2_{sy,n}$ and note that when Z is stationary, these conditions reduce to those given in Zubrzycki (1958). Finally we consider the relationship between sys and srs when $e^2_{st,n} < e^2_{sy,n}$.

PROPOSITION 6.1. *For every n-point* srs *design which satisfies* (2.4) *and* (2.6) *there exists a better n-point* sts *design in the sense that* $e^2_{st,n} \leq e^2_{r,n}$.

PROOF. We show this for a sts design with $n = m (n_i \equiv 1)$. Let G be the sampling distribution for srs and g its density on $A \cap (\phi \neq 0)$. For simplicity, assume $\phi \neq 0$ a.e. Consider a sts design such that for each $i = 1, \cdots, n$, $G(A_{in}) = 1/n$ and X_{in} has distribution concentrated on A_{in} with density proportional to g and such that $c_n(t)\, dG_n(t) = \phi(t)\, dt$ on A. Then

$$e^2_{r,n} - e^2_{st,n} = \frac{1}{2n} \sum_{i,j=1}^n \int_A \int_A R(s, t)\phi(s)\phi(t)(1_{A_{in}} - 1_{A_{jn}})(s)(1_{A_{in}} - 1_{A_{jn}})(t)\, ds\, dt \geq 0. \quad \square$$

Similarly it can be shown *that for every n-point* sts *design with some* $n_i > 1$ *there exists a better* sts *design with all* $n_i = 1$, obtained by subdividing any elements of the partition for which $n_i > 1$. Thus, with no loss of generality, we assume $n_i \equiv 1$ in Proposition 5.1 and will continue to make this assumption.

PROPOSITION 6.2. *Consider a* sts *design in which* $n = m$, $n_i = 1$ *and the* ith *sample point has distribution* G_{in} *on* A_{in}, $i = 1, \cdots, n$, *and a* sys *design in which the* ith *sample point again has distribution* G_{in} *on* A_{in}, $i = 1, \cdots, n$, *and the transformations* $\{T_{ijn}\}$ *are consistent with this choice. Let* c_n *be the same for both designs. For* $s, t \in A_{in}$, *let*

$$k_{jin}(s, t) = R(s, T_{jin}t) c_n(T_{jin}t),$$

$$M_{jin} = \int_{A_{in}} k_{jin}(t, t)\, dG_{in}(t) - \int_{A_{in}} \int_{A_{in}} k_{jin}(s, t)\, dG_{in}(s)\, dG_{in}(t).$$

(i) *If* $M_{jin} \geq 0$ *for all* $i \neq j$, *then* $e^2_{st,n} \leq e^2_{sy,n}$.
(ii) *If* $M_{jin} \leq 0$ *for all* $i \neq j$, *then* $e^2_{sy,n} \leq e^2_{st,n}$.

PROOF. As a result of these assumptions

$$n^2(e^2_{sy,n} - e^2_{st,n}) = E\{\sum_{i,j} R(X_{in}, T_{jin}X_{in})c_n(X_{in})c_n(T_{jin}X_{in}) - R(X_{in}, X_{jn})c_n(X_{in})c_n(X_{jn})\},$$

where each X_{in} has distribution G_{in}. For each i, let Y_{in} be a random variable independent of and identically distributed with X_{in}. Then since $X_{jn} = T_{jin}X_{in}$ a.s., $T_{jin}Y_{in}$ has the same distribution as X_{jn}, yielding

$$n^2(e^2_{sy,n} - e^2_{st,n}) = \sum_{i\neq j} E\{k_{jin}(X_{in}, X_{in}) - k_{jin}(X_{in}, Y_{in})\} = \sum_{i\neq j} M_{jin},$$

which completes the proof. □

When R is stationary, $\phi = 1/|A|$ for A bounded, $g_n = 1$ on A and the T_{jin} are translations by τ_{jin}, then

$$M_{jin} = R(\tau_{jin}) - |A_{in}|^{-2} \int_{A_{jn}} \int_{A_{jn}} R(t - s)\, ds\, dt,$$

and Proposition 6.2 reduces to a result due to Zubrzycki (1958).

We would like to obtain sufficient conditions that are easier to verify than $M_{jin} \geq 0$ (≤ 0). Observe that if for each $i \neq j$, k_{jin} satisfies the inequality

$$k(s, s) + k(t, t) - k(s, t) - k(t, s) \geq 0 \quad (\text{or} \leq 0) \tag{6.1}$$

a.e. $dG_{in} \times dG_{in}$ on $A_{in} \times A_{in}$, we obtain, upon integration with respect to $dG_{in} \times dG_{in}$, $M_{jin} \geq 0$ (or ≤ 0). Thus (6.1) is a sufficient condition. It has the intuitive appeal that it can be written as

$$E[\{Y(s) - Y(t)\}\{Y(T_{jin}s) - Y(T_{jin}t)\}] \geq 0 \quad (\text{or} \leq 0),$$

where $Y(s) = Z(s)c_n(s)$, and so requires that the increments of the processes Y and $Y \circ T_{jin}$, over the same interval of A_{in}, should be positively (or negatively) correlated a.e. $[dG_{in}]$ for all $i \neq j$. When $d = 1$, $c_n = 1/|A|$ for bounded A, and the T_{jin} are order preserving, the increments of the two processes are positively correlated if $R(s, t) = \min(s, t)$ or $\exp(-\rho|s - t|)$ and negatively correlated if $R(s, t) = \exp\{-\rho(s - t)^2\}$ for $\rho < 1/(8|A|^2)$. The error comparisons when $d = 2$, $c_n = 1/|A|$, and $R(s, t) = \exp(-\rho|s - t|)$ have been established by Zubrzycki (1958). Comparisons based on the asymptotic size of the mse's are explored in Schoenfelder (1978).

If $e^2_{sy,n} \leq e^2_{st,n}$, then $e^2_{sy,n} \leq e^2_{st,n} \leq e^2_{r,n}$. If $e^2_{sy,n} \geq e^2_{st,n}$, then, in general, the order between the mse's in srs and sys is not fixed. It is shown in Schoenfelder (1978) that for large n, under certain mild assumptions, $e^2_{sy,n} \leq e^2_{r,n}$. On the other hand, as is seen in the following example, we may have $e^2_{r,n} \leq e^2_{sy,n}$ for certain fixed n.

EXAMPLE 6.3. Let $\phi = 1$, $A = (0, 1]$, and $Z(t) = Y(t) + a\sin(2\nu\pi t + \theta)$ where $0 \leq \theta < 2\pi$, $a \geq 0$, and Y is a stochastic process having mean 0 and covariance $E\{Y(s)Y(t)\} = \exp(-\rho|s - t|)$ where $0 < \rho < \infty$. For srs, let $g = c = 1$. For sts, let $g_n = c_n = 1$ and $A_{in} = ((i - 1)/n, i/n]$. For sys, let $g_n = c_n = 1$, $A_{in} = ((i - 1)/n, i/n]$ and $T_{ijn}t = t + (j - i)/n$ when $t \in A_{jn}$. Consider a midpoint sampling scheme (mps) (cf. Tubilla, 1974) in which, in the notation of Section 2, $c_n = 1$ and $X_{in} = (2i - 1)/(2n)$ for all i. Then if ν is a multiple of n, the sample size,

$$e^2_{st,n} = \frac{1}{n}\left(1 + \frac{a^2}{2}\right) - C(\rho, n), \quad e^2_{r,n} = \frac{1}{n}\left\{1 + \frac{a^2}{2} - C(\rho, 1)\right\},$$

$$e^2_{sy,n} = B(\rho, n) + \frac{a^2}{2} - C(\rho, 1), \quad e^2_{mp,n} = B(\rho, n) + a^2\sin^2\theta - 2D(\rho, n) + C(\rho, 1),$$

where

$$B(\rho, n) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \exp\left(-\frac{\rho}{n} |i-j| \right),$$

$$C(\rho, n) = \frac{2}{\rho} \left\{ \frac{1}{n} - \frac{1}{\rho} (1 - e^{-\rho/n}) \right\}, \quad D(\rho, n) = \frac{2}{\rho} - \frac{1}{n\rho} (1 - e^{-\rho}) \operatorname{sech}\left(\frac{\rho}{2n} \right).$$

It is easily seen that $e_{st,n}^2 < e_{r,n}^2$ if $n > 1$. Since $B(\rho, n) > C(\rho, 1)$ it follows that $e_{r,n}^2 < e_{sy,n}^2$ if $a^2 > 2\{1 - C(\rho, 1)\}/(n-1)$, which is satisfied if $a^2 > 2/(n-1)$. Further, $e_{sy,n}^2 < e_{mp,n}^2$ if and only if

$$a^2(.5 - \sin^2\theta) \le \frac{2}{n\rho} (1 - e^{-\rho}) \left(\operatorname{sech}\left(\frac{\rho}{2n} \right) - \frac{2n}{\rho} \right),$$

which is satisfied if $a^2(.5 - \sin^2\theta) < -1/(6n^2)$. Thus if the process Y is such that ν is a multiple of n, and a and θ satisfy $a^2 > 2(n-1)$, $a^2(.5 - \sin^2\theta) < -1/(6n^2)$, then we have for this fixed n that

$$e_{st,n}^2 < e_{r,n}^2 < e_{sy,n}^2 < e_{mp,n}^2.$$

7. Proofs of results.

Proof of Proposition 2.1. The "only if" part is clear. For the "if" part, assume $J_n(s, t) = G_n(s)G_n(t)$. Then for any Borel set $B \subset A$,

$$\sum_{i=1}^{n} P(X_{in} \in B) \le n^2 J_n(B, B) = n^2 G_n^2(B) = \{\sum_{i=1}^{n} P(X_{in} \in B)\}^2,$$

and thus $\sum P(X_{in} \in B) = 0$, or else ≥ 1. Let D be the set of atoms of G_n. Partitioning $A - D$ into $n+1$ sets A_j such that $G_n(A_j) \le 1/(n+1)$ for all j, we have

$$\sum_{i=1}^{n} P(X_{in} \in A_j) = nG_n(A_j) \le \frac{n}{n+1} < 1$$

for all j. Thus $G_n(A_j) = 0$ for all j and G_n is purely discrete.

Finally we show that particular points cannot appear in some realizations of the design and not in others, or with differing multiplicities, giving a random design. Indeed $n^2 J_n(\{a\}, \{a\}) = [n\ G_n\{a\}]^2$ implies $\Sigma_{ij} P(X_{in} = a, X_{jn} = a) = \{\Sigma_i P(X_{in} = a)\}^2$, i.e., $E\{\Sigma_i 1_{\{a\}}(X_{in})\}^2 = [E\{\Sigma_i 1_{\{a\}}(X_{in})\}]^2$. It follows that $\Sigma_i 1_{\{a\}}(X_{in}) =$ constant a.s., and thus, with probability one, the number of sample points which coincide with a is a fixed integer (in $[1, n]$ when $a \in D$, equal to 0 otherwise). Thus the design is nonrandom. □

Proof of Lemma 2.2. To prove Lemma 2.2 we need the following lemmas which use a decomposition for bivariate distributions due to Chesson (1976). Chesson considers bivariate distributions for random elements, but here we consider only vector-valued random variables. Let X and Y be random variables taking values in $\mathbb{R}^d$ and defined on some probability space $(\Omega, \mathscr{A}, P)$. Denote the joint distribution function of X and Y by J, their marginal distribution functions by F and G, and by $\mathscr{F}$ and $\mathscr{G}$ the σ-fields generated by X and Y, respectively. Define

$$\mathscr{H} = \{(f(X), g(Y)) : f \in \mathscr{L}_2(\mathbb{R}^d, \mathscr{B}^d, dF), g \in \mathscr{L}_2(\mathbb{R}^d, \mathscr{B}^d, dG)\}.$$

Chesson notes that $\mathscr{H}$ is a real Hilbert space under the inner product

$$< (f_1(X), g_1(Y)), (f_2(X), g_2(Y)) > = \frac{1}{2} E\{f_1(X) f_2(X) + g_1(Y) g_2(Y)\}.$$

Then using the bounded self-adjoint operator B in $\mathscr{H}$ defined by

$$B(f(X), g(Y)) = (E\{g(Y) \mid X\}, E\{f(X) \mid Y\})$$

and its spectral decomposition, he proves the following result, which we restate as follows for vector-valued random variables.

THEOREM 7.1. (Chesson, 1976). *There exists a unique family of subspaces* $\mathscr{M}_r$ *of* $\mathscr{H}$, $0 \le r \le 1$, *such that*

(i) $\bigcap_{r>r'} \mathscr{M}_r = \mathscr{M}_{r'}$, $0 \le r' < 1$, *and* $\mathscr{M}_0 = \{0\}$.

(ii) *If* $\{(\xi_u(X), \eta_u(Y))\}_{u\in T(r)}$ *is an orthonormal basis for* $\mathscr{M}_r$, *then* $\{\xi_u\}_{u\in T(r)}$ *and* $\{\eta_u\}_{u\in T(r)}$ *are orthonormal families of functions on the spaces* $\mathscr{L}_2(\mathbb{R}^d, \mathscr{B}^d, dF)$ and $L_2(\mathbb{R}^d, \mathscr{B}^d, dG)$ *respectively.*

(iii) *For* $(f(X), g(Y)) \in \mathscr{H}$ *and* $\{(\xi_u(X), \eta_u(Y))\}_{u\in T(r)}$ *as in* (ii),

$$E\{f(X)g(Y)\} = \int_{(0,1]} r\,dQ(r),$$

where the right-continuous function Q *of bounded variation is given by*

$$Q(r) = \textstyle\sum_{u\in T(r)} \left\{\displaystyle\int f(x)\xi_u(x)\, dF(x)\right\}\left\{\int g(y)\eta_u(y)\, dG(y)\right\}.$$

(iv) *If* $(f(X), g(Y)) \in \mathscr{M}_\beta \ominus \mathscr{M}_\alpha$, $0 \le \alpha < \beta \le 1$, *and* $Ef^2(X) = 1 = Eg^2(Y)$, *then* $\alpha < E\{f(X)g(Y)\} \le \beta$.

This theorem suggests the following result which is used in obtaining a representation for $\int\int R(s,t)c(s)c(t)J(ds, dt)$ in the proof of Lemma 2.2.

LEMMA 7.2. *For any function* $h \in \mathscr{L}_2(dF \times dG)$ *such that*

$$h(x, y) = \textstyle\sum_{i,j=1}^{\infty} a_{ij} f_i(x) g_i(y) \text{ in } \mathscr{L}_2(dF \times dG) \tag{7.1}$$

where

$$\textstyle\sum_{i,j=1}^{\infty} |a_{ij}| < \infty, \tag{7.2}$$

$\{f_i\}$ *is an orthonormal set in* $\mathscr{L}_2(dF)$, $\{g_i\}$ *is an orthonormal set in* $\mathscr{L}_2(dG)$, *and* $\{(\xi_u, \eta_u)\}_{u\in T(r)}$, F, *and* G *as in Theorem* 7.1,

$$Eh(X, Y) = \int_{(0,1]} rQ(dr; h) \tag{7.3}$$

where for fixed h, Q *is right continuous and of bounded variation and is given by*

$$Q(r; h) = \textstyle\sum_{u\in T(r)} \displaystyle\int\int h(x, y)\xi_u(x)\eta_u(y)\, dF(x)\, dG(y). \tag{7.4}$$

PROOF. By Chesson's Theorem, since $f_i \in \mathscr{L}_2(dF)$, $g_j \in \mathscr{L}_2(dG)$,

$$E\{f_i(X)g_j(Y)\} = \int_{(0,1]} rQ(dr; f_i g_j)$$

where

$$Q(r; f_i g_j) = \textstyle\sum_{u\in T(r)} E\{(f_i\xi_u)(X)\}E\{(g_j\eta_u)(Y)\}.$$

Let

$$Q_n(r; h) = \textstyle\sum_{i,j=1}^{n} a_{ij} Q(r; f_i g_j). \tag{7.5}$$

An application of Fubini's Theorem easily justified by (7.2) gives

$$E\{h(X, Y)\} = \textstyle\sum_{i,j=1}^{\infty} a_{ij} E\{f_i(X)g_j(Y)\} = \lim_{n\to\infty} \displaystyle\int_{(0,1]} rQ_n(dr; h).$$

We shall show that (i) for each r, $Q_n(r; h) \to Q(r; h)$ as $n \to \infty$ where $Q(r; h)$ is given by (7.4); (ii) the total variation $V(Q_n) \le M < \infty$ for all n; and (iii) $Q(r; h)$ is right continuous. Observe that (i) and (ii) imply that $Q(\cdot; h)$ is of bounded variation and (7.3) holds.

(i) Let Z have the same distribution, G, as Y but be independent of X. By orthonormality and the Schwarz Inequality

$$\sum_{u\in T(r)} |E\{(f_i\xi_u)(X)(g_j\eta_u)(Z)\}| \le [E\{f_i^2(X)\}\cdot E\{g_j^2(Y)\}]^{1/2} = 1 \tag{7.6}$$

and

$$E\{|(f_i\xi_u)(X)(g_j\eta_u)(Z)|\} \le 1.$$

Thus as desired

$$\lim_{n\to\infty} Q_n(r; h) = \sum_{i,j=1}^{\infty} \sum_{u\in T(r)} a_{ij} E\{(f_i\xi_u)(X)(g_j\eta_u)(Z)\}$$
$$= \sum_{u\in T(r)} E\{h(X, Z)\xi_u(X)\eta_u(Z)\}.$$

(ii) Denoting by V the total variation, we have from (7.5)

$$V(Q_n) \le \sum_{i,j=1}^{n} |a_{ij}|\, V(Q(\cdot; f_i g_j)).$$

From Chesson (1976), $Q(r; fg)$ may be written as the difference of two increasing functions of r, as follows (where $P(0, r]$ is the orthogonal projection from $\mathscr{H}$ onto $\mathscr{M}_r$):

$$Q(r; fg) = \langle P(0, r](f(X), g(Y)), (f(X), g(Y))\rangle$$
$$- \langle P(0, r](f(X), -g(Y)), (f(X), -g(Y))\rangle$$

giving for the total variation

$$V(Q(\cdot; fg)) = \langle P(0, 1](f(X), g(Y)), (f(X), g(Y))\rangle$$
$$+ \langle P(0, 1](f(X), -g(Y)), (f(X), -g(Y)), (f(X), -g(Y))\rangle \le Ef^2(X) + Eg^2(Y).$$

Thus

$$V(Q_n) \le 2 \sum_{i,j=1}^{n} |a_{ij}| \le 2 \sum_{i,j=1}^{\infty} |a_{ij}| < \infty.$$

(iii) From (7.5) and (7.6)

$$|Q_n(r; h) - Q_m(r; h)| \le \sum_{i,j=n+1}^{m} |a_{ij}|.$$

Thus Q_n is uniformly Cauchy, and since it converges pointwise to Q, it converges uniformly to Q. We have also by (7.5) that $Q_n(r; h)$ is right continuous since, by Chesson's Theorem, $Q(r; fg)$ is right continuous. The uniform convergence of Q_n to Q then implies that Q is right continuous. □

Observe that (1, 1) is an eigenvector of B with corresponding eigenvalue 1 and hence $(1, 1) \in \mathscr{M}_1$. Thus the orthonormal basis $\{(\xi_u, \eta_u)\}_{u\in T(1)}$ for $\mathscr{M}_1$ may be chosen as $\{(1, 1)\} \cup \{(\xi_u, \eta_u)\}_{u\in S(1)}$ for some set $S(1)$. Also by part (i) of Chesson's Theorem, $\cap_{r>0} \mathscr{M}_r = \{0\}$ and thus $Q(0+) = 0$. In the proof of Lemma 2.2 we let $S(r) = T(r)$, $0 < r < 1$.

LEMMA 7.3. *Assume that the bivariate distribution of X and Y is symmetric and such that $E\{f(X)f(Y)\} \ge 0$ for all f, with $Ef^2(X) < \infty$. Then every element $(f(X), g(Y))$ of $\mathscr{M}_r$, $0 < r \le 1$, is such that $f = g$ a.s.* $[dF]$.

PROOF. Define the operator $C: \mathscr{H} \to \mathscr{H}$ by

$$C(f(X), g(Y)) = (g(X), f(Y)).$$

Under the assumption of symmetry, C is well-defined on $\mathscr{H}$ and is an isomorphism. It may easily be seen that C commutes with B and thus with $P(0, r]$, $0 < r \le 1$. Assume $(f(X),$

$g(Y)) \in \mathscr{M}_r$. Then

$$(g(X), f(Y)) = CP(0, r](f(X), g(Y)) = P(0, r](g(X), f(Y))$$

implies $(g(X), f(Y)) \in \mathscr{M}_r$. Since $\mathscr{M}_r$ is a subspace $((f-g)(X), (g-f)(Y)) \in \mathscr{M}_r$. Part (iv) of Chesson's Theorem then implies $E\{(f-g)(X)(g-f)(Y)\} \geq 0$, with equality iff $f = g$ a.s. $[dF]$. By hypothesis $E\{(f-g)(X)(f-g)(Y)\} \geq 0$, and thus $f = g$ a.s. $[dF]$. □

LEMMA 7.4. *If the bivariate distribution of X and Y satisfies the hypothesis of Lemma* 7.3 *and if h is such that*

$$h(x, y) = \textstyle\sum_{i=1}^{\infty} a_i f_i(x) f_i(y) \text{ in } \mathscr{L}_2(dF \times dF)$$

where $a_i \geq 0$, $i = 1, 2, \cdots, \sum_{i=1}^{\infty} a_i < \infty$, *and* $\{f_i\}$ *is an orthonormal set in* $\mathscr{L}_2(dF)$, *then* $Q(r; h)$ *of Lemma* 7.2 *is nondecreasing in r.*

PROOF. If $\{(\xi_u(X), \xi_u(Y))\}_{u \in T(r)}$ is an orthonormal basis for $\mathscr{M}_r$ (cf. Lemma 7.3), we have that Q of Lemma 7.2 is given by

$$Q(r; h) = \textstyle\sum_{i=1}^{\infty} a_i \sum_{u \in T(r)} \left\{ \displaystyle\int f_i(x)\, \xi_u(x)\, dF(x) \right\}^2.$$

Define $\mathscr{N}_r = \{f : (f(X), f(Y)) \in \mathscr{M}_r\}$, $0 \leq r \leq 1$. It may be easily seen that if $\{\xi_u(X), \xi_u(Y)\}_{u \in T(r)}$ is a complete orthonormal set (CONS) in $\mathscr{M}_r$, then $\{\xi_u\}_{u \in T(r)}$ is a CONS in $\mathscr{N}_r$. Denoting by f_{ir} the projection of f_i onto $\mathscr{N}_r$, we have

$$Q(r; h) = \textstyle\sum_{i=1}^{\infty} a_i \| f_{ir} \|^2$$

where $\|f\|$ is the norm of f in $\mathscr{L}_2(dF)$. From part (i) of Chesson's Theorem, $\bigcap_{r > r'} \mathscr{N}_r = \mathscr{N}_{r'}$ which implies $\|f_{ir}\| > \|f_{ir'}\|$, $0 \leq r' < r \leq 1$, and, since the a_i are nonnegative, that $Q(r; h)$ is nondecreasing in r. □

Now we are in a position to prove Lemma 2.2.

PROOF (OF LEMMA 2.2). Denote the LHS of (2.8) by e^2. That e^2 has the desired form is shown as follows, using the above results.

Step 1. We first show that $R(s, t)c(s)c(t) = K(s, t)$, say, is of the form (7.1) with (7.2) holding.

Since R continuous implies $\mathscr{R}(R)$, the reproducing kernel Hilbert space of R, is separable, it can be easily shown that $\mathscr{R}(K)$ is separable. Since K is Borel measurable and $\mathscr{R}(K)$ is separable, the second order process with mean 0 and covariance K has a measurable modification (Cambanis, 1975). Also, (2.6) implies that K is the kernel of an integral operator K in $\mathscr{L}_2[dG]$. It then follows from Theorems 1 and 3 and the proof of Theorem 1 in Cambanis (1973), which may easily be extended to real-valued measurable stochastic processes defined on a Borel set in $\mathbb{R}^d$, that for all $s, t \in A$,

$$K(s, t) = \textstyle\sum_{k=1}^{\infty} \lambda_k a_k(s) a_k(t) + r(s, t), \tag{7.7}$$

where $\{\lambda_k\}$ and $\{a_k\}$ are the (positive) eigenvalues and the corresponding eigenfunctions of K; that $\sum_{k=1}^{\infty} \lambda_k < \infty$; and that the covariance function $r(s, t)$ satisfies $r(t, t) = 0$ a.e. $[dG]$. Thus $r(s, t) = 0$ a.e. $[dG \times dG]$, and (7.7) holds with the series converging in $\mathscr{L}_2[dG \times dG]$, i.e., K is of the form (7.1) with (7.2) holding.

Step 2. Using Lemmas 7.2 and 7.3 we can then write

$$\int_A \int_A R(s, t)c(s)c(t) J(ds, dt) = \int_{(0,1]} r\,dQ(r) + \int_A \int_A R(s, t)c(s)c(t)\, dG(s)\, dG(t),$$

where

$$Q(r) = \sum_{u \in S(r)} \int_A \int_A R(s, t)(c\xi_u)(s)(c\xi_u)(t)\, dG(s)\, dG(t);$$

$\{\xi_u\}_{u \in S(r)}$ forms an orthonormal set in $\mathscr{L}_2[dG]$ for each r; and the sets $\mathscr{M}_r$ with basis $\{(\xi_u, \xi_u)\}_{u \in S(r)}$, $0 < r \le 1$, and $\mathscr{M}_0 = \{0\}$ satisfy $\cap_{r > r'} \mathscr{M}_r = \mathscr{M}_{r'}$, $0 \le r' < 1$. Also Q is of bounded variation and nondecreasing, and $Q(0+) = 0$. Thus (2.8) holds.

Assume R is strictly positive definite and c is nonzero a.e. $[dG]$. Then $\int r\,dQ(r) = 0$ iff $Q \equiv 0$; iff $(c\,\xi_u)(s)\, dG(s) = 0$ on A for all $u \in S(r)$, $0 < r \le 1$; iff $\xi_u = 0$ a.e. $[dG]$, for all $u \in S(r)$, $0 < r \le 1$; and hence iff $J(s, t) = G(s)G(t)$ on $A \times A$. □

Acknowledgment. The authors are grateful to the referees and the editor whose comments helped improve the presentation of the paper.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

CAMBANIS, S. (1973). Representation of stochastic processes of second order and linear operations. *J. Math. Anal. Appl.* **41** 603–620.

CAMBANIS, S. (1975). The measurability of a stochastic process of second order and its linear space. *Proc. Amer. Math. Soc.* **47** 467–475.

CHESSON, P. (1976). The canonical decomposition of bivariate distributions. *J. Multivariate Anal.* **6** 526–537.

COCHRAN, W. (1946). Relative accuracy of systematic and stratified random samples for a certain class of population. *Ann. Math. Statist.* **17** 164–177.

DAVID, M. (1978). Sampling and estimation problems for three dimensional spatial stationary and nonstationary stochastic processes as encountered in the mineral industry. *J. Statist. Plann. Inference* **2** 211–244.

JOWETT, G. (1952). The accuracy of systematic sampling from conveyor belts. *Appl. Statist.* **1** 50–59.

QUENOUILLE, M. (1949). Problems in plane sampling. *Ann. Math. Statist.* **20** 355–375.

SACKS, J. and YLVISAKER, D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.* **37** 66–89.

SACKS, J. and YLVISAKER, D. (1970). Statistical designs and integral approximation. In *Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress,* 115–136, Canadian Mathematical Congress, Montreal.

SCHOENFELDER, C. (1978). Random designs for estimating integrals of stochastic processes. *Institute of Statistics Mimeo Series No. 1201,* Univ. of North Carolina at Chapel Hill.

TUBILLA, A. (1975). Error convergence rates for estimates of multidimensional integrals of random functions. *Technical Report No. 72,* Dept. of Statist., Stanford Univ.

WAHBA, G. (1978). Interpolating surfaces: High order convergence rates and their associated designs, with application to X-ray image reconstruction. *Technical Report No. 523,* Dept. of Statist., Univ. of Wisconsin at Madison.

WILLIAMS, R. (1956). The variance of the mean of systematic samples. *Biometrika* **43** 137–148.

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

ZUBRZYCKI, S. (1958). Remarks on random, stratified, and systematic sampling in a plane. *Colloq. Math.* **6** 251–264.

DEPARTMENT OF MATHEMATICS
DUKE UNIVERSITY
DURHAM, NC 27706

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NC 27514