Ann. Inst. Statist. Math.
Vol. 46, No. 4, 707–722 (1994)

SAMPLING DESIGNS FOR REGRESSION COEFFICIENT ESTIMATION WITH CORRELATED ERRORS*

YINGCAI SU[1] AND STAMATIS CAMBANIS[2]

[1]*Department of Statistics, University of Arizona, Tucson, AZ 85721, U.S.A.*
[2]*Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, U.S.A.*

(Received March 1, 1993; revised December 9, 1993)

Abstract. The problem of estimating regression coefficients from observations at a finite number of properly designed sampling points is considered when the error process has correlated values and no quadratic mean derivative. Sacks and Ylvisaker (1966, *Ann. Math. Statist.*, **39**, 66–89) found an asymptotically optimal design for the best linear unbiased estimator (BLUE). Here, the goal is to find an asymptotically optimal design for a simpler estimator. This is achieved by properly adjusting the median sampling design and the simpler estimator introduced by Schoenfelder (1978, Institute of Statistics Mimeo Series No. 1201, University of North Carolina, Chapel Hill). Examples with stationary (Gauss-Markov) and nonstationary (Wiener) error processes and with linear and nonlinear regression functions are considered both analytically and numerically.

Key words and phrases: Regression coefficient estimation, sampling designs, correlated errors.

1. Introduction, results and examples

The problem of interest is to estimate regression coefficients from observations at a finite number of appropriately designed sampling points when the error process has correlated values. The performance of an estimator is measured by its mean square error (MSE). The sampling points designed so as to minimize the MSE form an optimal sampling design which is generally difficult to find. To circumvent this difficulty Sacks and Ylvisaker (1966) proposed an asymptotic approach and showed that the regular sampling designs, determined by a properly chosen sampling density, are asymptotically optimal when the best linear unbiased estimator (BLUE) is used and the error process has no quadratic mean derivative. Schoenfelder (1978) considered the median sampling design along with a simpler estimator, but the procedure is not asymptotically optimal for this problem (Cambanis (1985)). In this paper we adjust appropriately the median sampling

* Research supported by the Air Force Office of Scientific Research Contract No. 91-0030.

design and the simpler estimator and show that the adjusted scheme has an asymptotically optimal performance.

One parameter regression setup. We first consider the simple regression model

$$X(t) = \beta f(t) + N(t), \qquad t \in [0,1], \tag{1.1}$$

where f is a known regression function, β is an unknown parameter and the errors $N(t)$ form a correlated random process with mean zero, $EN(t) \equiv 0$, and known covariance function $EN(t)N(s) = R(t,s)$. Here f is assumed to have comparable smoothness with the noise covariance and more specifically, to be of the form

$$f(t) = \int_0^1 R(t,s)\phi(s)ds, \tag{1.2}$$

where ϕ is a known continuous function on $[0,1]$. The error process $N(t)$ is assumed to have no quadratic mean derivative but its covariance function $R(t,s)$ is assumed to have continuous mixed partial derivatives of order up to two off the diagonal of the unit square, and continuous limits for $R^{0,1}$ at the diagonal from above and below, denoted by $R^{0,1}(t,t\pm) = \lim_{s\to t\pm 0} \partial R(t,s)/\partial s$. The function of jumps of $R^{0,1}$ along the diagonal,

$$\alpha(t) = R^{0,1}(t,t-) - R^{0,1}(t,t+),$$

is assumed to be continuous, ≥ 0 and not identically 0, and plays a crucial role in the asymptotics.

If $X(t)$ can be observed over the entire interval $[0,1]$, then the BLUE of β is

$$\hat\beta = s^{-2}\int_0^1 X(t)\phi(t)dt,$$

where

$$s^2 = \int_0^1\int_0^1 \phi(t)R(t,s)\phi(s)dtds \left(\triangleq \iint \phi R\phi\right) = \int_0^1 \phi(t)f(t)dt$$

and its MSE is

$$\mathrm{MSE}(\hat\beta) = \mathrm{Var}(\hat\beta) = s^{-2}$$

(see Parzen (1961), Sacks and Ylvisaker (1966)). Here, we want to estimate β from observations of $X(t)$ at n sample points $T_n = \{t_{n,k}\}_1^n$ in the interval $[0,1]$ using a linear estimator

$$\beta_{T_n} = \sum_{k=1}^n C_{n,k}X(t_{n,k}) = C'_{T_n}X_{T_n},$$

where $C'_{T_n} = (C_{n,1},\ldots,C_{n,n})$ are coefficients and $X'_{T_n} = (X(t_{n,1}),\ldots,X(t_{n,n}))$. Then $\mathrm{MSE}(\beta_{T_n}) = \mathrm{Bias}^2(\beta_{T_n}) + \mathrm{Var}(\beta_{T_n})$, where

$$\mathrm{Bias}(\beta_{T_n}) = \beta(C'_{T_n}f_{T_n} - 1), \qquad \mathrm{Var}(\beta_{T_n}) = C'_{T_n}R_{T_n}C_{T_n},$$

$f'_{T_n} = (f(t_{n,1}), \ldots, f(t_{n,n}))$ and $R_{T_n} = (R(t_{n,k}, t_{n,j}))_{n\times n}$ is assumed to be invertible for every n. We wish to choose the coefficients C_{T_n} and the sampling designs T_n in such a way that $\mathrm{Bias}(\beta_{T_n})$ and $\mathrm{Var}(\beta_{T_n})$ are as close as possible to 0 and s^{-2}, respectively.

For a fixed sampling design T_n, the optimal coefficients $\hat{C}_{T_n}$ are those of the BLUE $\hat{\beta}_{T_n}$, which minimize $\mathrm{Var}(\beta_{T_n})$ subject to $\mathrm{Bias}(\beta_{T_n}) = 0$,

$$\hat{C}'_{T_n} = f'_{T_n} R^{-1}_{T_n} / f'_{T_n} R^{-1}_{T_n} f_{T_n}, \tag{1.3}$$

and the corresponding variance is

$$\mathrm{Var}(\hat{\beta}_{T_n}) = (f'_{T_n} R^{-1}_{T_n} f_{T_n})^{-1}. \tag{1.4}$$

Asymptotically optimal sampling designs for BLUE. An optimal sampling design $T_n^o = \{t^o_{n,k}\}_1^n$ of size n, which minimizes $\mathrm{Var}(\hat{\beta}_{T_n})$, is generally difficult to find. In order to avoid this difficulty, Sacks and Ylvisaker (1966) introduced sequences of sampling designs $\{T_n^*\}_1^\infty$ which are asymptotically optimal as the sample size n tends to infinity in the sense that

$$\lim_{n\to\infty} \{\mathrm{Var}(\hat{\beta}_{T_n^*}) - s^{-2}\} \Big/ \left\{\inf_{T_n} \mathrm{Var}(\hat{\beta}_{T_n}) - s^{-2}\right\} = 1, \tag{1.5}$$

where the infinimum is taken over all designs of sample size n, and showed that

$$\lim_{n\to\infty} n^2 \left\{\inf_{T_n} \mathrm{Var}(\hat{\beta}_{T_n}) - s^{-2}\right\} = \frac{s^{-4}}{12} \int_0^1 \alpha(t) \frac{\phi^2(t)}{h_o^2(t)} dt,$$

where

$$h_o(t) = \{\alpha(t)\phi^2(t)\}^{1/3} \Big/ \int_0^1 \{\alpha(u)\phi^2(u)\}^{1/3} du, \tag{1.6}$$

assuming that $\alpha(t)$ is strictly positive and $\phi(t)$ has no zeros. From this and (1.5) we can conclude that a sequence of sampling designs $\{T_n\}_1^\infty$ is asymptotically optimal if and only if it satisfies

$$\lim_{n\to\infty} n^2 \{\mathrm{Var}(\hat{\beta}_{T_n}) - s^{-2}\} = \frac{s^{-4}}{12} \int_0^1 \alpha(t) \frac{\phi^2(t)}{h_o^2(t)} dt. \tag{1.7}$$

Regular and median sampling designs. Let $h(t)$ be a positive density on $[0, 1]$, with strictly increasing distribution $H(t) = \int_0^t h(s) ds$, $0 \le t \le 1$. The regular sampling design

$$T_n = \left\{ r_{n,k} = H^{-1}\left(\frac{k-1}{n-1}\right),\ k = 1, \ldots, n \right\}$$

includes the endpoints $r_{n,1} = 0$ and $r_{n,n} = 1$. The median sampling design

$$\left\{ m_{n,k} = H^{-1}\left(\frac{2k-1}{2n}\right),\ k = 1, \ldots, n \right\}$$

consists of the h-medians of each $(r_{n+1,k}, r_{n+1,k+1})$. With $h(t) \equiv 1$, both regular and median sampling are periodic.

Sacks and Ylvisaker (1966) showed that *the regular sequence of sampling designs determined by h_o of (1.6) is asymptotically optimal when the BLUE $\hat{\beta}_{T_n}$ is used.* In view of (1.3), the coefficient $\hat{C}_{n,k}$ of each $X(r_{n,k})$ depends on all sampling points and its computation may be liable to numerical instabilities.

Schoenfelder (1978) and Cambanis and Masry (1983) have used simpler coefficients, which, for our problem, are

$$C_{n,k} = n^{-1}s^{-2}(\phi/h)(r_{n,k}), \qquad k = 1, \ldots, n,$$

along with regular and median sampling designs. The resulting estimators $\beta_n(r,h)$ and $\beta_n(m,h)$ are generally biased. The median sampling design using simpler coefficients $\beta_n(m,h_0)$ is asymptotically optimal for the integral estimation and the signal detection problems (Cambanis (1985)). However, for the regression problem, its asymptotic behavior is

$$\lim_{n\to\infty} n^2\{\mathrm{MSE}[\beta_n(m,h)] - s^{-2}\} = \frac{s^{-4}}{12}\int_0^1 \alpha(t)\frac{\phi^2(t)}{h^2(t)}dt + \frac{s^{-4}}{12}\kappa, \tag{1.8}$$

where

$$\kappa = \int_0^1 \left[\frac{K^{0,1}(t,0)}{h(0)} - \frac{K^{0,1}(t,1)}{h(1)}\right] h(t)dt$$

and $K(t,s) = (\phi/h)(t)R(t,s)(\phi/h)(s)$ (see Eq. (51) in Bucklew and Cambanis (1988) which corrects Eq. (5.14′) in Cambanis (1985)), so the median scheme is not generally asymptotically optimal.

Modified median sampling designs. Here we modify the median sampling design and the simpler estimator coefficients, so that the resulting sequence of estimators $\beta_n(mm,h)$ becomes asymptotically optimal, i.e. satisfies

$$\lim_{n\to\infty} n\,\mathrm{Bias}(\beta_n) = 0, \tag{1.9}$$

$$\lim_{n\to\infty} n^2\{\mathrm{Var}(\beta_n) - s^{-2}\} = \frac{s^{-4}}{12}\int_0^1 \alpha(t)\frac{\phi^2(t)}{h^2(t)}dt, \tag{1.10}$$

or equivalently,

$$\lim_{n\to\infty} n^2\{\mathrm{MSE}(\beta_n) - s^{-2}\} = \frac{s^{-4}}{12}\int_0^1 \alpha(t)\frac{\phi^2(t)}{h^2(t)}dt. \tag{1.11}$$

The expression of κ in (1.8) suggests that proper use of the endpoints 0 and 1 in the design may improve the asymptotic constant. Since the values at the sampling points are used to represent the values over each partitioning interval, using the endpoints will require choosing the two end intervals shorter than the remaining ones. We found that with equal first and last partitioning intervals (of h-length $a/(n-1)$) and equal remaining $(n-2)$ intervals (of h-length $b_n/(n-1)$), and with the natural choice of estimator coefficients (cf. (1.14)), a unique choice

($a^2 = 1/12$) leads to the following asymptotically optimal design. We partition $[0, 1]$ into n intervals $0 = s_{n,0} < s_{n,1} < \cdots < s_{n,n-1} < s_{n,n} = 1$ determined by

$$\begin{aligned} &\int_0^{s_{n,1}} h(t)dt = \int_{s_{n,n-1}}^{1} h(t)dt = \frac{1}{2\sqrt{3}(n-1)}, \\ &\int_{s_{n,k}}^{s_{n,k+1}} h(t)dt = \frac{\rho_n}{n-1}, \qquad k = 1, \ldots, n-2, \end{aligned} \tag{1.12}$$

where $\rho_n = (n - 1 - 1/\sqrt{3})/(n-2)$. The modified median sampling points $T_{n,m} = \{0 = t_{n,1,m} < t_{n,2,m} < \cdots < t_{n,n-1,m} < t_{n,n,m} = 1\}$ are determined by

$$t_{n,k,m} = H^{-1}\left(\frac{1}{2\sqrt{3}(n-1)} + \frac{k - 3/2}{n-1}\rho_n\right), \qquad k = 2, \ldots, n-1, \tag{1.13}$$

i.e., each $t_{n,k+1,m}$ is the h-median of $(s_{n,k},\ s_{n,k+1})$ for $k = 1, \ldots, n-2$, and the two end points are included in the design (see Fig. 1). When $h \equiv 1$, then, with $\Delta t_{n,k,m} = t_{n,k+1,m} - t_{n,k,m}$,

$$\begin{aligned} \Delta t_{n,1,m} &= \Delta t_{n,n-1,n} = (\rho_n + 1/\sqrt{3})/[2(n-1)], \\ \Delta t_{n,k,m} &= \rho_n/(n-1) \qquad \text{for} \quad k = 2, \ldots, n-2, \end{aligned}$$

and thus as $n \uparrow \infty$, $\Delta t_{n,1,m}/\Delta t_{n,2,m} \uparrow (1 + 1/\sqrt{3})/2 \approx 79\%$.

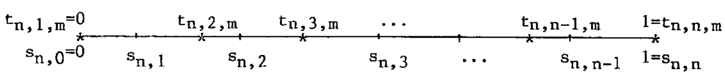


Fig. 1. The partitioning points $\{s_{n,k}, k = 0, 1, \ldots, n\}$ and the modified median sampling points $\{t_{n,k,m}, k = 1, \ldots, n\}$.

Modified simpler estimator coefficients. The modified coefficients $\{C_{n,k,m}\}$ are chosen by

$$C_{n,k,m} = s^{-2}\frac{\phi}{h}(t_{n,k,m}) \int_{s_{n,k-1}}^{s_{n,k}} h(t)dt, \qquad k = 1, \ldots, n. \tag{1.14}$$

The resulting modified median estimator can be written as

$$\beta_n(mm, h) = s^{-2} \sum_{k=1}^{n} W_{n,k} \left(\frac{\phi}{h} X\right)(t_{n,k,m}) \tag{1.15}$$

where the weights $\{W_{n,k}\}_1^n$ are given by

$$\begin{aligned} W_{n,1} &= W_{n,n} = \frac{1}{2\sqrt{3}(n-1)}, \\ W_{n,k} &= \frac{1}{n-2}\left\{1 - \frac{1}{\sqrt{3}(n-1)}\right\}, \qquad k = 2, \ldots, n-1, \end{aligned}$$

and sum up to one. Thus the modified median estimator $\beta_n(mm,h)$ takes a much simpler form than the BLUE $\hat{\beta}_{T_n}$ (cf. (1.3)) in those cases when the function ϕ in (1.2) has an explicit and simple expression.

THEOREM 1.1. *Assume that ϕ/h is twice continuously differentiable and R satisfies the assumptions stated at the beginning of the section. The estimator* (1.15) *with sampling points given by* (1.13) *satisfies*

$$\lim_{n\to\infty} n^2\{\mathrm{MSE}[\beta_n(mm,h)] - s^{-2}\} = \frac{s^{-4}}{12}\int_0^1 \alpha(t)\frac{\phi^2(t)}{h^2(t)}dt, \tag{1.16}$$

and thus the estimator $\beta_n(mm,h_o)$ is asymptotically optimal.

More general regression functions. As in Sacks and Ylvisaker (1966), Theorem 1.1 can be extended to more general regression functions of the form

$$f(t) = \int_0^1 R(t,s)\phi(s)ds + \sum_{\ell=1}^{L} b_\ell R(t,a_\ell), \tag{1.17}$$

where ϕ is a known continuous function, the b_ℓ's are known nonzero constants and the a_ℓ's are known points in $[0,1]$. For this model, the BLUE of β based on observations over the entire interval and its variance are

$$\hat{\beta} = s_L^{-2}\left\{\int_0^1 X(s)\phi(s)ds + \sum_{\ell=1}^{L} b_\ell X(a_\ell)\right\}, \qquad \mathrm{Var}(\hat{\beta}) = s_L^{-2} \tag{1.18}$$

where

$$s_L^2 \triangleq \iint \phi R \phi + 2\sum_{\ell=1}^{L} b_\ell \int_0^1 \phi(s)R(s,a_\ell)ds + \sum_{\ell=1}^{L}\sum_{j=1}^{L} b_\ell R(a_\ell,a_j)b_j.$$

The modified median estimator of β is then

$$\beta_n(mm,h,L) = s_L^{-2}\left\{\sum_{k=1}^{n} W_{n,k}\left(\frac{\phi}{h}X\right)(t_{n,k,m}) + \sum_{\ell=1}^{L} b_\ell X(a_\ell)\right\}$$

where the sampling design $\{t_{n,k,m}\}$ determined by (1.13) is augmented by the L fixed sampling points $\{a_\ell\}$. As in Theorem 1.1, under the same conditions, we have

$$\lim_{n\to\infty} n^2\left\{\mathrm{MSE}[\beta_n(mm,h,L)] - s_L^{-2}\right\} = \frac{s_L^{-4}}{12}\int_0^1 \alpha(t)\frac{\phi^2(t)}{h^2(t)}dt. \tag{1.19}$$

Multiple regression. We extend Theorem 1.1 to the multiple regression model

$$X(t) = \beta' f(t) + N(t), \qquad t \in [0,1], \tag{1.20}$$

where $\beta' = (\beta_1, \ldots, \beta_q)$ is a vector of unknown parameters and $f'(t) = (f_1(t), \ldots, f_q(t))$ is a vector of known regression functions. The error process $N(t)$ and each regression function $f_i(t)$ are as in (1.1). Specifically, each f_i satisfies (1.2) for some ϕ_i. If $X(t)$ is available over the entire interval $[0, 1]$ the BLUE of β and its MSE are

$$\hat{\beta}' = S^{-1} \int_0^1 X(t)\phi'(t)dt, \qquad \mathrm{MSE}(\hat{\beta}) = E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) = \mathrm{trace}(S^{-1}),$$

where $S = (s_{ij} = \iint \phi_i R \phi_j)_{q\times q}$ is assumed to be invertible and $\phi'(t) = (\phi_1(t), \ldots, \phi_q(t))$ (Parzen (1961), p. 484).

Here, we estimate β from the observations of $X(t)$ at the sampling points $T_n = \{t_{n,k}\}_1^n$, using linear estimators β_{T_n} of the form $\beta_{T_n} = S^{-1}D_{T_n}X_{T_n}$, where $D_{T_n} = (D^i_{n,k})_{q\times n}$ are coefficient matrices. Then

$$\mathrm{MSE}(\beta_{T_n}) = \mathrm{trace}\{S^{-1}D_{T_n}R_{T_n}D'_{T_n}S^{-1}\} + \|\,\mathrm{Bias}(\beta_{T_n})\|^2,$$

where

$$\begin{aligned} \|\,\mathrm{Bias}(\beta_{T_n})\|^2 &= \{E(\beta_{T_n}) - \beta\}'\{E(\beta_{T_n}) - \beta\} \\ &= \beta'\{D_{T_n}f_{T_n} - S\}'S^{-2}\{D_{T_n}f_{T_n} - S\}\beta \end{aligned} \tag{1.21}$$

with $f_{T_n} = \{f_i(t_{n,k})\}_{i=1,\ldots,q}^{k=1,\ldots,n} = (f_{1,T_n}, \ldots, f_{q,T_n}) = (f(t_{n,1}), \ldots, f(t_{n,n}))'$. For the modified median estimator, we choose $D^i_{n,k}$ as in (1.14), i.e.,

$$D^i_{n,k} = \frac{\phi_i}{h}(t_{n,k,m}) \int_{s_{n,k-1}}^{s_{n,k}} h(t)dt, \qquad k = 1, \ldots, n, \qquad i = 1, \ldots, q,$$

where $\{s_{n,k}\}_1^n$ and $T_{n,m} = \{t_{n,k,m}\}_1^n$ are as in (1.12) and (1.13), respectively. Then, the estimator can be written as

$$\beta'_n(mm, h) = S^{-1} \sum_{k=1}^{n} W_{n,k} \left(\frac{\phi'}{h}X\right)(t_{n,k,m}), \tag{1.22}$$

where the weights $\{W_{n,k}\}_1^n$ are as in (1.15). We have the following result.

THEOREM 1.2. *If ϕ_i/h, $i = 1, \ldots, q$, are twice continuously differentiable and $R(t, s)$ is as in Theorem 1.1, then*

$$\lim_{n\to\infty} n^2\{\mathrm{MSE}[\beta_n(mm, h)] - \mathrm{trace}(S^{-1})\} = \frac{1}{12} \int_0^1 \alpha(t) \frac{\phi'(t)S^{-2}\phi(t)}{h^2(t)} dt, \tag{1.23}$$

and the estimator $\beta_n(mm, h_o)$ is asymptotically optimal where

$$h_o(t) = \{\alpha(t)\phi'(t)S^{-2}\phi(t)\}^{1/3} \Big/ \int_0^1 \{\alpha(u)\phi'(u)S^{-2}\phi(u)\}^{1/3} du.$$

In Section 2, we establish (1.23) by showing

$$\lim_{n\to\infty} n\|\mathrm{Bias}[\beta_n(mm,h)]\| = 0, \tag{1.24}$$

$$\lim_{n\to\infty} n^2 \,\mathrm{trace}\{S^{-1}D_{T_{n,m}}R_{T_{n,m}}D'_{T_{n,m}}S^{-1} - S^{-1}\} = \frac{1}{12}\int_0^1 \alpha(t)\frac{\phi'(t)S^{-2}\phi(t)}{h^2(t)}dt. \tag{1.25}$$

Example 1. Quadratic regression in Wiener process error. Consider the model (1.1) with $R(t,s) = \min(t,s)$ and the quadratic regression function $f(t) = t^2$. Other power regression functions (along with BLUE estimators) were considered in Eubank *et al.* (1982). In this case $\alpha(t) = 1$ and t^2 can be written in the form (1.17): $t^2 = -2\int_0^1 \min(t,s)ds + 2t$, $t \in [0,1]$, with $\phi(s) = -2$, $L = 1$, $b_1 = 2$ and $a_1 = 1$. By direct calculations, we obtain $s_L^2 = 4/3$ and $h_o(t) \equiv 1$.

For this example, we are able to obtain the optimal sampling design corresponding to the BLUE for every sample size n as follows. By (1.3), minimizing $\mathrm{Var}(\hat\beta_{T_n})$ is equivalent to maximizing $\sigma_n^2 \triangleq f'_{T_n}R_{T_n}^{-1}f_{T_n}$. From the general expression in Sacks and Ylvisaker (1966), it follows that $\sigma_n^2 = \sum_{k=0}^{n-1}(t_{k+1}-t_k)(t_{k+1}+t_k)^2$ where $t_0 = 0$. It is easy to show that the maximum of σ_n^2 is achieved by the optimal design $t_k = k/n$, $k = 1,\dots,n$, which is periodic with corresponding value $\sigma_n^2 = 4/3 - 1/(3n^2)$. The corresponding BLUE and its variance are

$$\hat\beta_{T_n,\mathrm{opt}} = \frac{3n}{2n+1}\left\{-\frac{2}{2n-1}\sum_{k=1}^{n-1} X\left(\frac{k}{n}\right) + X(1)\right\},$$

$$\mathrm{Var}(\hat\beta_{T_n,\mathrm{opt}}) = \frac{3n^2}{4n^2-1} = \frac{3}{4} + \frac{3}{4(4n^2-1)}.$$

The median sampling scheme gives

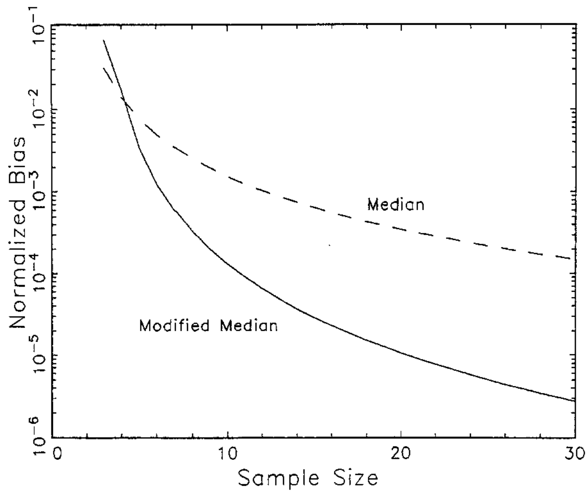
$$\beta_n(m,h_o) = \frac{3}{2}\left\{-\frac{1}{n-1}\sum_{k=1}^{n-1} X\left[\frac{2k-1}{2(n-1)}\right] + X(1)\right\},$$

$$\beta^{-1}\,\mathrm{Bias}\{\beta_n(m,h_o)\} = \frac{1}{8(n-1)^2}, \qquad \mathrm{Var}\{\beta_n(m,h_0)\} = \frac{3}{4} + \frac{3}{8(n-1)^2}.$$

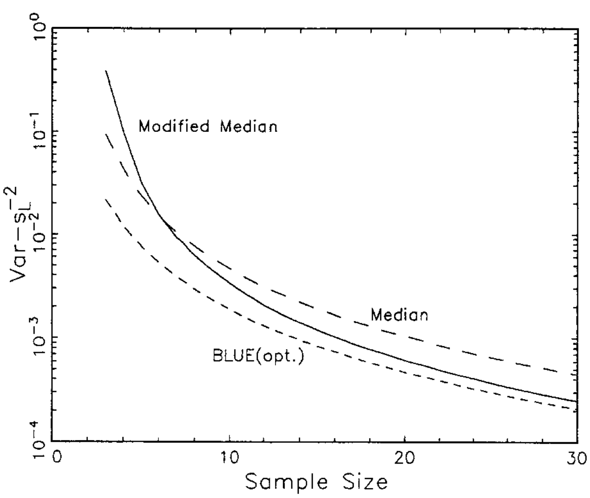
With $h_o \equiv 1$, the modified median sampling points and estimator are

$$t_{n,k,m} = \frac{1}{n-1}\left\{\frac{1}{2\sqrt{3}} + \rho_n\left(k - \frac{3}{2}\right)\right\}, \qquad k = 2,\dots,n-1,$$

$$\beta_n(mm,h_o) = \frac{3}{2}\left(-\frac{n-1-1/\sqrt{3}}{(n-1)(n-2)}\sum_{k=2}^{n-1} X(t_{n,k,m}) + \left\{1 - \frac{1}{2\sqrt{3}(n-1)}\right\}X(1)\right).$$



a) Normalized bias vs. sample size



b) Variance-s_L^{-2} vs. sample size

Fig. 2.

In this special case, $\hat{\beta}_{T_n,\text{opt}}$ takes the simplest form while $\beta_n(mm, h_o)$ involves complicated coefficients and sampling points. In order to compare the performance of $\hat{\beta}_{T_n,\text{opt}}$, $\beta_n(m, h_o)$ and $\beta_n(mm, h_o)$ for small and moderate sample sizes, we plot the normalized Bias, β^{-1} Bias, and the variance difference, $\text{Var}(\beta_n) - s_L^{-2}$, of these estimators versus the sample size $n = 3, \ldots, 30$ in Fig. 2.

Example 2. Linear regression in Gauss-Markov process error. We consider the model (1.1) with $R(t, s) = \sigma^2 e^{-\lambda|t-s|}$ and the linear regression function $f(t) =$

t. Then $\alpha(t) = 2\lambda\sigma^2$ and

$$t = \frac{\lambda}{2}\int_0^1 e^{-\lambda|t-s|} s ds - \frac{1}{2\lambda} e^{-\lambda t} + \frac{\lambda+1}{2\lambda} e^{-\lambda(1-t)},$$

which is of the form (1.17) with $\phi(s) = \lambda s/(2\sigma^2)$, $L = 2$, $a_1 = 0$, $a_2 = 1$, $b_1 = -1/(2\lambda\sigma^2)$ and $b_2 = (\lambda+1)/(2\lambda\sigma^2)$. By direct calculation, we obtain $s_L^2 = (\lambda^2 + 3\lambda + 3)/(6\lambda\sigma^2)$ and the asymptotically optimal sampling density $h_o(t) = (5/3)t^{2/3}$.

For a fixed sample size n, it is not easy to find the optimal sampling design (see the discussion in Morrison (1970)). The regular sampling design specified by $h_o(t)$ uses $r_{n,k} = \{(k-1)/(n-1)\}^{3/5}$, $k = 1, \ldots, n$, and the median sampling scheme gives the following estimator

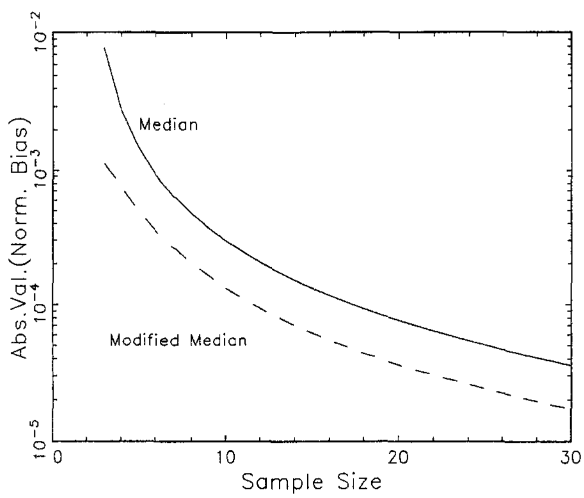
$$\beta_n(m, h_o) = s_L^{-2}\Bigg(-\frac{1}{2\lambda} X(0) + \frac{3\lambda}{10(n-1)} \sum_{k=1}^{n-2} \left[\frac{2k-1}{2(n-2)}\right]^{1/5} X\left(\left[\frac{2k-1}{2(n-2)}\right]^{3/5}\right) + \frac{\lambda+1}{2\lambda} X(1)\Bigg).$$

The modified median sampling design and estimator are

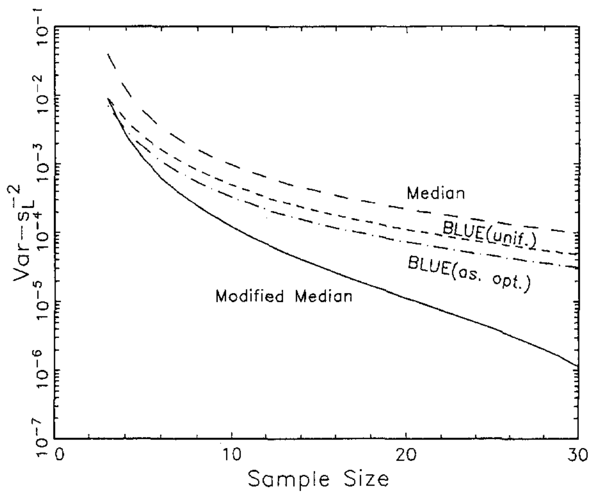
$$t_{n,k,m} = \left\{\left[\frac{1}{2\sqrt{3}} + \rho_n\left(k - \frac{3}{2}\right)\right]\frac{1}{n-1}\right\}^{3/5}, \qquad k = 2, \ldots, n-1,$$

$$\beta_n(mm, h_o) = s_L^{-2}\Bigg(-\frac{1}{2\lambda} X(0) + \frac{3\lambda\rho_n}{10(n-1)} \sum_{k=2}^{n-1} t_{n,k,m}^{1/3} X(t_{n,k,m}) + \left[\frac{\lambda+1}{2\lambda} + \frac{\sqrt{3}\lambda}{20(n-1)}\right] X(1)\Bigg).$$

We plot the absolute value of the normalized bias and the variance difference versus the sample size in Figs. 3(a) and (b), respectively. We have plotted these curves for a variety of values of the parameter λ with $\sigma^2 = 1$ and found that over the displayed range of sample size, the modified median estimator performs better than not only the median estimator but also the BLUE (as. opt.) when $\lambda < 5$, and that the improvement increases as λ becomes smaller. To illustrate this here, we choose $\lambda = 1.5$ and $\sigma^2 = 1$. Also we computed the variance difference of the modified median estimator up to the sample size 100 and found that it becomes negative at about $n = 34$, achieves a minimum at about $n = 55$ and then increases to zero from below. This is not surprising because the modified median estimator is biased, and thus it may have smaller variance than the BLUE $\hat{\beta}$.



a) Abs.Val.(norm. bias) vs. sample size



b) Variance-s_L^{-2} vs. sample size

Fig. 3.

2. Proofs

We adjust the analysis in Bucklew and Cambanis (1988) (hereafter referred to as BC) the modified median designs and coefficients to prove that (1.9), (1.10), (1.24) and (1.25) hold. For simplicity, we drop the subscripts n and m in $t_{n,k,m}$, $C_{n,k,m}$, etc.

By the mean value theorem of integrals and (1.12), we have

$$\frac{1}{2\sqrt{3}(n-1)} = \int_0^{s_1} h = h(w_1)s_1 = \int_{s_{n-1}}^1 h = h(w_n)(1-s_{n-1}), \tag{2.1}$$

$$(2.2)\qquad \frac{\rho_n}{n-1} = \int_{s_k}^{s_{k+1}} h = h(w_k)(s_{k+1} - s_k),$$

$$(2.3)\qquad \frac{\rho_n}{2(n-1)} = \int_{s_k}^{t_{k+1}} h = h(a_k)(t_{k+1} - s_k) = \int_{t_{k+1}}^{s_{k+1}} h = h(b_k)(s_{k+1} - t_{k+1}),$$

where $0 < w_1 < s_1$, $s_{n-1} < w_n < 1$, $s_k < w_k < s_{k+1}$ and $s_k < a_k < t_{k+1} < b_k < s_{k+1}$, $k = 1, \dots, n-2$. It is easy to see that for $k = 1, \dots, n-2$, as $n \to \infty$,

$$(2.4)\qquad \frac{t_{k+1} - s_k}{s_{k+1} - s_k} = \frac{h(w_k)}{2h(a_k)} \to \frac{1}{2} \quad \text{and} \quad \frac{s_{k+1} - t_{k+1}}{s_{k+1} - s_k} = \frac{h(w_k)}{2h(b_k)} \to \frac{1}{2}.$$

By the same arguments as given by BC ((1988), pp. 122–123), we have

$$(2.5)\qquad D_k = (s_{k+1} - t_{k+1}) - (t_{k+1} - s_k) = -\frac{1}{8h(t_k)}\left(\frac{h'(b'_k)}{h^2(b_k)} + \frac{h'(a'_k)}{h^2(a_k)}\right)\frac{\rho_n^2}{(n-1)^2},$$

where $h'(t) = dh(t)/dt$ and $s_k < a'_k < t_{k+1} < b'_k < s_{k+1}$, and then for $k = 1, \dots, n-2$,

$$\int_0^{s_1} th(t)dt = \frac{1}{24}\frac{h(\text{int.})}{h^2(w_1)}\frac{1}{(n-1)^2} = \frac{1}{24}\frac{1}{h(0)}\frac{1}{(n-1)^2} + o(n^{-2}),$$

$$(2.6)\qquad \int_{s_k}^{s_{k+1}} (t - t_{k+1})h(t)dt = -\frac{1}{24}\frac{h'(t_{k+1})}{h^2(t_{k+1})}(s_{k+1} - s_k)\frac{\rho_n^2}{(n-1)^2} + o(n^{-2}),$$

$$\int_{s_{n-1}}^{1} (t-1)h(t)dt = -\frac{1}{24}\frac{h(\text{int.})}{h^2(w_n)}\frac{1}{(n-1)^2} = -\frac{1}{24}\frac{1}{h(1)}\frac{1}{(n-1)^2} + o(n^{-2}),$$

and

$$\int_0^{s_1} t^2 h(t)dt = o(n^{-2}),$$

$$(2.7)\qquad \int_{s_{n-1}}^{1} (t-1)^2 h(t)dt = o(n^{-2}),$$

$$\int_{s_k}^{s_{k+1}} (t - t_{k+1})^2 h(t)dt = \frac{1}{12}\frac{1}{h(t_{k+1})}(s_{k+1} - s_k)\frac{\rho_n^2}{(n-1)^2} + o(n^{-2}).$$

Proof of (1.9) and (1.24). Let $g = \phi f/h$. By Taylor expansion, we have

$$g(t) = g(t_{k+1}) + g'(t_{k+1})(t - t_{k+1}) + \frac{1}{2}g''(x_k)(t - t_{k+1})^2,$$

where x_k is between t_{k+1} and t and depends continuously on t. Thus, using $s^2 = \int_0^1 gh$, we can write

$$\begin{aligned}
&\beta^{-1}s^2 \operatorname{Bias}[\beta_n(mm,h)] \\
&\quad = \sum_{k=0}^{n-1} g(t_{k+1}) \int_{s_k}^{s_{k+1}} h - s^2 = \sum_{k=0}^{n-1} \int_{s_k}^{s_{k+1}} [g(t_{k+1}) - g(t)]h(t)dt \\
&\quad = -\left\{ \sum_{k=0}^{n-1} g'(t_{k+1}) \int_{s_k}^{s_{k+1}} (t - t_{k+1})h(t)dt \right. \\
&\qquad \left. + \frac{1}{2} \sum_{k=0}^{n-1} \int_{s_k}^{s_{k+1}} g''(x_k)(t - t_{k+1})^2 h(t)dt \right\} \\
&\quad \triangleq -\{A_n + B_n\}.
\end{aligned}$$

By (2.1) and (2.6), we obtain

$$\begin{aligned}
A_n = &\frac{1}{24(n-1)^2} \left(\frac{g'}{h}\right)(0) - \frac{1}{24} \frac{\rho_n^2}{(n-1)^2} \sum_{k=2}^{n-2} \left(\frac{g'h'}{h^2}\right)(t_{k+1})(s_{k+1} - s_k) \\
&- \frac{1}{24(n-1)^2} \left(\frac{g'}{h}\right)(1) + o(n^{-2}).
\end{aligned}$$

For B_n we apply the mean value theorem to pull out $g''(y_k)$, $y_k \in (s_k, s_{k+1})$, and then use (2.7), to obtain

$$B_n = \frac{1}{2} \frac{\rho_n^2}{(n-1)^2} \sum_{k=1}^{n-2} g''(y_k) \frac{1}{12h(t_k)} (s_{k+1} - s_k) + o(n^{-2}).$$

It follows from Riemann integrability and $\rho_n \to 1$ that as $n \to \infty$,

$$\begin{aligned}
&n^2 \beta^{-1} s^2 \operatorname{Bias}[\beta_n(mm, h_0)] \\
&\quad \to -\frac{1}{24} \left(\left(\frac{g'}{h}\right)(0) - \left(\frac{g'}{h}\right)(1) + \int_0^1 \left(\frac{g''}{h} - \frac{g'h'}{h^2}\right) \right) = 0.
\end{aligned}$$

This proves (1.9). Then (1.24) follows from (1.9) and (1.21).

PROOF OF (1.10) AND (1.25). Suppose we can show that for $i, j = 1, \ldots, q$, as $n \to \infty$,

$$\begin{aligned}
(2.8) \qquad n^2 E_{ij}(n) &\triangleq n^2 \left\{ D_{T_n}^{i}{}' R_{T_n} D_{T_n}^{j} - \int_0^1 \int_0^1 \phi_i R \phi_j \right\} \\
&\to \frac{1}{12} \int_0^1 \frac{\alpha(t)}{h^2(t)} \phi_i(t) \phi_j(t) dt.
\end{aligned}$$

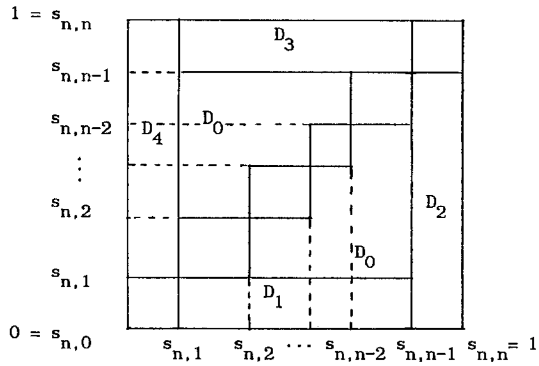


Fig. 4. The off-diagonal rectangles D_0, D_1, D_2, D_3, D_4 of the unit square.

Then, putting $S^{-2} = (\nu_{ij})_{q\times q}$, the left hand side of (1.25) equals

$$\begin{aligned}\lim_{n\to\infty} n^2 \mathrm{trace}(D'_{T_n} R_{T_n} D_{T_n} - S)S^{-2} &= \sum_{i=1}^{q}\sum_{j=1}^{q} \lim_{n\to\infty} n^2 E_{ij}(n)\nu_{ji} \\ &= \frac{1}{12}\sum_{i=1}^{q}\sum_{j=1}^{q}\int_0^1 \frac{\alpha(t)}{h^2(t)}\phi_i(t)\phi_j(t)dt\nu_{ji} \\ &= \frac{1}{12}\int_0^1 \frac{\alpha(t)}{h^2(t)}\phi'(t)S^{-2}\phi(t)dt\end{aligned}$$

and (1.25) follows. Note that (1.10) follows by letting $i = j$ in (2.8). Therefore, it remains to show (2.8) and it suffices to show it for $i = 1$ and $j = 2$. Putting $K(s,t) = (\phi_1/h)(s)R(s,t)(\phi_2/h)(t)$, the left hand side of (2.8) becomes

$$n^2\sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1}\int_{s_k}^{s_{k+1}}\int_{s_\ell}^{s_{\ell+1}}[K(t_{k+1},t_{\ell+1}) - K(s,t)]h(s)h(t)dsdt \triangleq n^2\sum_{k=0}^{n-1}\sum_{\ell=0}^{n-1} I_{n;k,\ell}.$$

Now split the integral over the unit square into the integrals over the diagonal squares and the off-diagonal rectangles in the regions D_r, $r = 0,1,2,3,4$ as shown in Fig. 4. Repeating the steps in BC ((1988), p. 123), i.e., splitting each diagonal square into the six regions shown in Fig. 5 of that paper and then using Taylor expansions and (2.1)–(2.4), we obtain

$$\begin{aligned}(2.9)\qquad n^2\sum_{k=0}^{n-1} I_{n;k,k} \to \frac{1}{12}\int_0^1 [&K^{0,1}(t,t-) - K^{0,1}(t,t+) \\ &+ K^{1,0}(t-,t) - K^{1,0}(t+,t)]dt.\end{aligned}$$

Repeating the procedures on p. 123 of BC (1988) and using (2.5)–(2.7) we obtain

$$(2.10)\qquad n^2\sum_{D_0} I_{n;k,\ell} \to -\frac{1}{24}\Big\{\int_0^1 [K^{0,1}(t,t-) - K^{0,1}(t,t+)]dt$$

$$+ \int_0^1 \left[\frac{K^{0,1}(t,0)}{h(1)} - \frac{K^{0,1}(t,0)}{h(0)} \right] h(t)dt$$
$$+ \int_0^1 [K^{1,0}(t-,t) - K^{1,0}(t+,t)]dt$$
$$+ \int_0^1 \left[\frac{K^{0,1}(t,0)}{h(1)} - \frac{K^{0,1}(t,0)}{h(0)} \right] h(t)dt \bigg\}.$$

By Taylor expansion, we have

$$K(s,t) = K(t_{k+1},0) + K^{1,0}(t_k,0)(s-t_{k+1}) + K^{0,1}(t_k,0)t \\ + K^{1,1}(a_k,b_0)(s-t_{k+1})t + \frac{1}{2}K^{2,0}(a_k,b_0)(s-t_k)^2 + \frac{1}{2}K^{0,2}(a_k,b_0)t^2,$$

where (a_k, b_0) is on the line joining $(t_k, 0)$ to (s,t). Thus,

$$\sum_{D_1} I_{n;k,\ell} = \sum_{k=1}^{n-2} \int_{s_k}^{s_{k+1}} \int_0^{s_1} [K(t_{k+1},0) - K(s,t)]h(s)h(t)dsdt$$
$$= -\sum_{k=1}^{n-2} \int_{s_k}^{s_{k+1}} \int_0^{s_1} \Big[K^{1,0}(t_{k+1},0)(s-t_{k+1})$$
$$+ K^{0,1}(t_{k+1},0)t + K^{1,1}(a_k,b_0)(s-t_{k+1})t$$
$$+ \frac{1}{2}K^{2,0}(a_k,b_0)(s-t_{k+1})^2$$
$$+ \frac{1}{2}K^{0,2}(a_k,b_0)t^2 \Big] h(s)h(t)dsdt.$$

By (2.6), the first term of the right hand side equals

$$= -\sum_{k=1}^{n-2} K^{1,0}(t_k,0)\left(-\frac{1}{24}\right)\frac{h'(t_{k+1})}{h^2(t_{k+1})}(s_{k+1}-s_k)\frac{\rho_n^2}{(n-1)^2}\frac{1}{2\sqrt{3}(n-1)} = o(n^{-2})$$

and the second term equals

$$= -\sum_{k=1}^{n-2} K^{0,1}(t_{k+1},0)h(w_k)(s_{k+1}-s_k)\frac{1}{h(0)}\frac{1}{24(n-1)^2} + o(n^{-2}).$$

Similarly, by using (2.1) and (2.2), one can verify that the third, fourth and fifth terms are all of the order $o(n^{-2})$. Therefore,

$$\sum_{D_1} I_{n;k,\ell} = -\frac{1}{24}\frac{1}{(n-1)^2}\sum_{k=1}^{n-2} K^{0,1}(t_{k+1},0)\frac{h(w_k)}{h(0)}(s_{k+1}-s_k) + o(n^{-2})$$

which yields

$$n^2 \sum_{D_1} I_{n;k,\ell} \to -\frac{1}{24} \int_0^1 K^{0,1}(t,0) \frac{h(t)}{h(0)} dt. \tag{2.11}$$

In a similar way, we obtain

$$n^2 \sum_{D_2} I_{n;k,\ell} \to \frac{1}{24} \int_0^1 K^{1,0}(1,t) \frac{h(t)}{h(1)} dt, \tag{2.12}$$

$$n^2 \sum_{D_3} I_{n;k,\ell} \to \frac{1}{24} \int_0^1 K^{0,1}(t,1) \frac{h(t)}{h(1)} dt, \tag{2.13}$$

$$n^2 \sum_{D_4} I_{n;k,\ell} \to -\frac{1}{24} \int_0^1 K^{1,0}(0,t) \frac{h(t)}{h(0)} dt. \tag{2.14}$$

Adding (2.9)–(2.14) up, we obtain (2.8):

$$\begin{aligned} n^2 \sum_{k,\ell} I_{n;k,\ell} &\to \frac{1}{24} \int_0^1 \{K^{0,1}(t,t-) - K^{0,1}(t,t+) + K^{1,0}(t-,t) - K^{1,0}(t+,t)\} dt \\ &= \frac{1}{12} \int_0^1 \frac{\alpha(t)}{h^2(t)} \phi_1(t)\phi_2(t) dt. \end{aligned}$$

References

Bucklew, J. A. and Cambanis, S. (1988). Estimating random integrals from noisy observations: sampling designs and their performance, *IEEE Trans. Inform. Theory*, **34**, 111–127.

Cambanis, S. (1985). Sampling designs for time series, *Handbook of Statistics, 5: Time Series in Time Domain* (eds. E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 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.

Eubank, R. L., Smith, P. L. and Smith, P. W. (1982). On the computation of optimal designs for certain time series models with applications to optimal quantile selection for location or scale parameter estimation, *SIAM J. Sci. Statist. Comput.*, **3**, 238–249.

Morrison, D. F. (1970). The optimal spacing of repeated measurements, *Biometrics*, **26**, 281–290.

Parzen, E. (1961). Regression analysis of continuous parameter time series, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 469–489, University of California Press, Berkeley.

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

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