The Annals of Statistics
1993, Vol. 21, No. 2, 903–923

AN ANALYSIS OF BAYESIAN INFERENCE FOR NONPARAMETRIC REGRESSION[1]

BY DENNIS D. COX

Rice University

The observation model $y_i = \beta(i/n) + \varepsilon_i$, $1 \le i \le n$, is considered, where the ε's are i.i.d. with mean zero and variance σ^2 and β is an unknown smooth function. A Gaussian prior distribution is specified by assuming β is the solution of a high order stochastic differential equation. The estimation error $\delta = \beta - \hat{\beta}$ is analyzed, where $\hat{\beta}$ is the posterior expectation of β. Asymptotic posterior and sampling distributional approximations are given for $\|\delta\|^2$ when $\|\cdot\|$ is one of a family of norms natural to the problem. It is shown that the frequentist coverage probability of a variety of $(1-\alpha)$ posterior probability regions tends to be larger than $1-\alpha$, but will be infinitely often less than any $\varepsilon > 0$ as $n \to \infty$ with prior probability 1. A related continuous time signal estimation problem is also studied.

1. Introduction. In this article we consider Bayesian inference for a class of nonparametric regression models. Suppose we observe

$$Y_{ni} = \beta(t_{ni}) + \varepsilon_i, \qquad 1 \le i \le n, \tag{1.1}$$

where $t_{ni} = i/n$, β: $[0,1] \to \mathbb{R}$ is an unknown smooth function, and $\varepsilon_1, \varepsilon_2, \dots$ are i.i.d. random errors with mean 0 and known variance $\sigma^2 < \infty$. The ε_i are modeled as $N(0, \sigma^2)$. A Gaussian prior for β will now be specified. Let $m \ge 2$ and for some constants $a_0, \dots, a_m$ with $a_m \ne 0$ let

$$L = \sum_{i=0}^{m} a_i D^i$$

be a constant coefficient linear differential operator ($D = d/dt$). Let ω denote a standard Gaussian white noise on $[0,1]$, which is formally the derivative of a standard Brownian motion $\langle W(t)\colon 0 \le t \le 1\rangle$. Let Σ be an $m \times m$ positive definite matrix and let $B_1, \dots, B_m$ be boundary value operators of order $m-1$ or less, that is, $B_i(\beta)$ is a linear form in

$$\beta(0), \beta(1), D\beta(0), D\beta(1), \dots, D^{m-1}\beta(0), D^{m-1}\beta(1).$$

Assume that $Lg \equiv 0$ and $B_i(g) = 0$, $1 \le i \le m$ imply $g \equiv 0$. Then β is

Received February 1987; revised June 1992.
[1]Research partially supported by the Office of Naval Research under contract N00014-84-C-0169 and NSF Grant DMS-820-2560.
AMS 1991 *subject classifications.* Primary 62A15; secondary 62G15, 62J99, 62E20, 62M99, 60G35.
Key words and phrases. Bayesian inference, nonparametric regression, confidence regions, signal extraction, smoothing splines.

assumed to be the solution of the stochastic differential equation

$$L\beta = \omega, \qquad \mathbf{B}(\beta) \sim N(0, \Sigma), \tag{1.2}$$

where $\mathbf{B}(\beta) = (B_1(\beta), \dots, B_m(\beta))$. One may avoid the use of a white noise ω by writing an $m \times m$ first order system. Perhaps the simplest model of this type is the integrated Brownian motion with random initial conditions. This is obtained from $D^2\beta = \omega$, $(\beta(0), D\beta(0)) \sim N(0, \Sigma)$, so that $\beta(t) = \beta(0) + (D\beta(0))t + \int_0^t W(s)\,ds$, with $\beta(0)$, $D\beta(0)$ bivariate normal. In general, one may express β through a simple stochastic integral

$$\beta(t) = \int_0^1 G(t, s)\, dW(s) + \beta_1(t),$$

where G is the Green's function for the operator L with homogeneous boundary conditions $B_i(\beta) = 0$, $1 \le i \le m$ [see Section 3.3 and 3.4 of Naimark (1967)] and $\beta_1(t)$ is a solution to the homogeneous differential equation ($L\beta_1 \equiv 0$) with boundary conditions that agree with β ($B_i(\beta_1) = B_i(\beta)$, $1 \le i \le m$).

This specifies the prior for β. The Bayesian model thus defined is very similar to others that have appeared in the literature [Kimeldorf and Wahba (1970a, b), Wahba (1978) and Wecker and Ansley (1983)]. Related models and applications are discussed by Diaconis (1988). More abstract models are discussed in Section 2.

One of the attractive features of the Bayesian approach is that in principle one can solve virtually any statistical decision or inference problem. In particular, one can provide an accuracy assessment for $\hat{\beta}_n = E[\beta | Y_n]$ using posterior probability regions. For instance, letting $\|\cdot\|$ denote $L_2[0, 1]$ norm, one can in principle determine a number Δ_n such that

$$P\left[\|\hat{\beta}_n - \beta\|^2 \le \Delta_n | Y_n\right] = 0.95,$$

thus giving a 95% posterior probability bound on the L_2-norm of the estimation error. A useful large sample approximation for Δ_n follows from results given below, namely

$$\Delta_n \sim \mu_n + 1.645\tau_n,$$

where

$$\mu_n \sim \left(2\pi m a_m^{-1/m}\right)^{-1} B(1/(2m), 1 - 1/(2m))(\sigma^2/n)^{1 - 1/(2m)},$$

$$\tau_n^2 \sim \left(2\pi m a_m^{-1/m}\right)^{-1} B(1/(2m), 2 - 1/(2m))(\sigma^2/n)^{2 - 1/(2m)}$$

and $B(x, y)$ is the beta function. [See (3.18), Theorem 3.1(a), (3.3), (3.1) and the remarks after Theorem 4.1.]

Non-Bayesians often find such Bayesian procedures attractive because as $n \to \infty$, the frequentist coverage probability of the Bayesian region tends to the posterior coverage probability in "typical" cases. It was my hope that this

TABLE 1

m	2	3	4	10	∞
$1-\alpha$	89.4%	90.5%	91.2%	92.7%	95%

would hold in the nonparametric regression setting, thus providing a methodology for constructing large sample confidence regions. Unfortunately, the hoped for result is false in about the worst possible way, viz.,

$$\liminf_{n\to\infty} P\Big[\|\hat{\beta}_n - \beta\|^2 \le \Delta_n|\beta\Big] = 0, \quad \text{a.s.} \tag{1.3}$$

Thus, if one fixes a sample path from the Gaussian prior, then the frequentist coverage probability of the region $\{\beta\colon \|\beta - \hat{\beta}_n\|^2 \le \Delta_n\}$ will infinitely often be arbitrarily small as $n \to \infty$, for almost all sample paths.

Nonetheless, for fixed n (large), the frequentist coverage probability is at least 95% for "most" sample paths in the following sense:

$$\lim_{n\to\infty} P\Big[P\Big[\|\hat{\beta}_n - \beta\|^2 \le \Delta_n|\beta\Big] \ge 0.95\Big] = 1 - \alpha,$$

where $1 - \alpha$ depends on m as indicated in Table 1 (see Corollary 3.4).

The procedure analyzed here differs from that advocated by Wahba (1983) and Wecker and Ansley (1983) in the following ways:

1. These authors use a model wherein $\beta = \beta_0 + \beta_1$ and β_1 is given a proper prior of the type above but β_0 is finite dimensional and given a Lebesgue prior.
2. We have assumed σ^2 is known.
3. These authors assume $\beta = b\beta_0$ where β_0 has a given prior and $b > 0$ is an unknown scale factor which is estimated.
4. Wahba (1983) considers the true function β to be fixed and smoother than one generated by the Gaussian prior.

I conjecture that 1, 2 and 4 do not change the negative result (1.3). See Cox (1989). Concerning point 3, (1.3) depends on the law of the iterated logarithm fluctuations of the bias $E[\hat{\beta}_n - \beta|\beta]$ about its mean (Lemma 3.2). Such fluctuations undoubtedly impact the smoothing parameter estimation procedure of Wahba (1983), known as generalized cross validation [see also Craven and Wahba (1979) and Speckman (1983)], so (1.3) may not hold when b is estimated.

In Section 2 we consider a general Gaussian prior on an abstract space. One of the essential facts about Gaussian measures is the existence of the Karhunen–Loeve expansion, which gives in our setup

$$\beta(t) = \sum_{\nu=1}^{\infty} \alpha_\nu \beta_\nu \eta_\nu(t),$$

where $\alpha_1, \alpha_2, \ldots$ are constants, $\beta_1, \beta_2, \ldots$ are i.i.d. $N(0,1)$ and $\eta_1(t), \eta_2(t), \ldots$

is an orthonormal basis for $L_2[0, 1]$. Our model for β as a solution of an mth order stochastic differential equation allows fairly precise determination of α_ν as $\nu \to \infty$ [(3.2), noting that $\gamma_\nu = \alpha_\nu^{-2}$]. Many of the results cited below depend on the particular stochastic differential equation model (through the limiting behavior of α_ν), although some generalization is clearly possible. In Section 3, a continuous time analog of (1.1) is investigated, and in Section 4 it is shown that the discrete time model (1.1) can be well approximated by the continuous time model. In Sections 2, 3 and 4 we obtain results on the estimation error in a family of norms indexed by a parameter ρ, denoted $\|\cdot\|_\rho$. When $\rho = 0$, we obtain L_2 norm (and hence the claims above), and when $\rho > 0$ the norms are equivalent to certain Sobolev norms (basically L_2 norms on derivatives). These norms allow us to show, for example, that (1.3) holds more generally than just for L_2 norms. Section 5 contains some concluding remarks.

2. The abstract Gauss–Bayes linear model. In this section, we formulate and solve a general version of the abstract linear model when the error vector is modeled as Gaussian with known covariance and a Gaussian prior is used for the parameter vector. Well known results about Gaussian measures on Banach spaces will be used, for which Kuo (1975) is an excellent reference; see also Kuelbs (1970, 1971) and Kallianpur (1971). *We will drop the subscript n throughout this section* as we only consider a fixed prior and observation.

Suppose ε is (modeled as) a mean 0 Gaussian random vector on a real separable Banach space $\overline{\mathscr{Y}}$, and $\mathscr{Y}$ is the generating Hilbert space (GHS) for the Gaussian measure $\mathscr{L}(\varepsilon)$ = the distribution of ε on $\overline{\mathscr{Y}}$. [$\mathscr{Y}$ is also known as the reproducing kernel Hilbert space for $\mathscr{L}(\varepsilon)$; see Kuelbs (1970), Kallianpur (1971) or Kuo (1975).] Let β be a mean zero Gaussian random vector on a real separable Banach space $\overline{\Theta}$ with GHS Θ, and assume β and ε are independent. Now let

$$\overline{X}: \overline{\Theta} \to \overline{\mathscr{Y}}$$

be a bounded linear operator, called the design operator, and suppose we observe

$$Y = \overline{X}\beta + \sigma\varepsilon,$$

where $\sigma > 0$ is known. We wish to estimate β from Y. To do this, we will show that a posterior distribution exists and characterize it.

As a remark, we note that $\overline{\mathscr{Y}}$ plays little role in what follows—it is $\mathscr{Y}$ that is important. If it were not for the technical problem that $\mathscr{L}(\varepsilon)$ will not live on $\mathscr{Y}$ when $\dim \mathscr{Y} = \infty$, we could carry out the analysis using only $\mathscr{Y}$ with no mention of $\overline{\mathscr{Y}}$. Kallianpur and Karandikar (1985) have an elegant approach to problems in this vein. Our approach here is in the more classical style using ordinary measure theory, for which it is necessary to keep $\overline{\mathscr{Y}}$ around.

PROPOSITION 2.1. *There is a regular conditional posterior distribution $\mathscr{L}(\beta|Y)$, and in fact $\mathscr{L}(\beta|Y)$ is absolutely continuous w.r.t. the prior $\mathscr{L}(\beta)$, written $\mathscr{L}(\beta|Y) \ll \mathscr{L}(\beta)$.*

PROOF. The assumption that the range $\mathscr{R}(\overline{X}) \subseteqq \mathscr{Y}$ guarantees the sampling distribution $\mathscr{L}(Y|\beta) \ll \mathscr{L}(\sigma\varepsilon)$ by Theorem 3.1, page 118 of Kuo (1975). An elementary Fubini argument completes the proof. $\square$

As the inclusion map $\iota\colon \Theta \to \overline{\Theta}$ is continuous, it follows that the restriction X of $\overline{X}$ to Θ is in $\mathscr{B}(\Theta, \mathscr{Y})$, the set of all bounded linear operators from $\Theta \to \mathscr{Y}$. Let $X^* \in \mathscr{B}(\mathscr{Y}, \Theta)$ denote the adjoint and put

$$U = X^*X \in \mathscr{B}(\Theta),$$

where $\mathscr{B}(\Theta) = \mathscr{B}(\Theta, \Theta)$. The description of the posterior is most easily accomplished using the singular value decomposition of X, given next.

LEMMA 2.2. *There exist complete orthonormal systems (abbreviated c.o.n.s.) $\langle \psi_\nu\colon \nu = 1,2,\dots \rangle$ for Θ and $\langle \eta_\nu\colon \nu = 1,2,\dots \rangle$ for $\overline{\mathscr{R}}(X)$, the closure of $\mathscr{R}(X)$ in $\mathscr{Y}$, and constants $\langle \alpha_\nu^2\colon \nu = 1,2,\dots \rangle$ such that $U\psi_\nu = \alpha_\nu^2\psi_\nu$ for all ν, and*

$$X\psi_\nu = \alpha_\nu \eta_\nu, \quad \forall\, \nu, \tag{2.1}$$

$$X^*\eta_\nu = \alpha_\nu \psi_\nu, \quad \forall\, \nu, \tag{2.2}$$

$$\sum \alpha_\nu^2 < \infty. \tag{2.3}$$

PROOF. $X = \overline{X} \circ \iota$ and ι is a compact operator. Hence X, U and $V = XX^*$ are compact. U is also self-adjoint and nonnegative definite on Θ, so there are nonnegative eigenvalues $\alpha_1^2, \alpha_2^2, \dots$ and eigenvectors $\psi_1, \psi_2, \dots$ such that $\langle \psi_\nu \rangle$ is a c.o.n.s. for Θ, and $U\psi_\nu = \alpha_\nu^2 \psi_\nu$ [Theorem 1.8, page 8 of Kuo (1975)]. We assume w.l.o.g. that $\alpha_1^2 \ge \alpha_2^2 \ge \cdots$. Put $\eta_\nu = \alpha_\nu^{-1} X\psi_\nu$ for $\alpha_\nu > 0$ and then (2.1) and (2.2) hold. Now we have $\beta = \sum_\nu \langle \beta, \psi_\nu \rangle_\Theta \psi_\nu$ with convergence $\mathscr{L}(\beta)$ a.s. in $\overline{\Theta}$, so $\overline{X}\beta = \sum_\nu \alpha_\nu \langle \beta, \psi_\nu \rangle_\Theta \eta_\nu$. As this latter series converges $\mathscr{L}(\beta)$ a.s. in $\mathscr{Y}$, and $\langle \beta, \psi_1 \rangle_\Theta, \langle \beta, \psi_2 \rangle_\Theta, \dots$ are i.i.d. $N(0,1)$, it follows that $\sum \alpha_\nu^2 < \infty$. Alternatively, one checks that V is the covariance of $X\beta$ in $\mathscr{Y}$, so trace $V = \sum \alpha_\nu^2 < \infty$. $\square$

Even though $\beta \in \Theta$ with probability 0, the "stochastic linear functional" $\beta \to \langle \theta, \beta \rangle_{\tilde{\Theta}}$ is defined a.s. as a measurable map $\overline{\Theta} \to \mathbb{R}$, for each fixed $\theta \in \Theta$. The properties of these maps are given by Kuelbs (1971) and Kuo (1975), Lemma 4.7, page 78. Similar remarks hold for ε and $\langle \varepsilon, \eta \rangle_{\tilde{\mathscr{Y}}}$, $\eta \in \mathscr{Y}$. Note that $\langle Y, \eta \rangle_{\tilde{\mathscr{Y}}} = \langle X\beta, \eta \rangle_{\tilde{\mathscr{Y}}} + \sigma \langle \varepsilon, \eta \rangle_{\tilde{\mathscr{Y}}}$, a.s. Since $X\beta \in \mathscr{Y}$, the term $\langle X\beta, \eta \rangle_{\mathscr{Y}}$ is defined as an ordinary linear function. We also have Karhunen–Loeve expansions: if $e_1, e_2, \dots$ is any c.o.n.s. for Θ, then $\beta = \sum_\nu \langle \beta, e_\nu \rangle_{\tilde{\Theta}} e_\nu$ with convergence in $\overline{\Theta}$, almost surely. See Kuelbs (1971).

LEMMA 2.3. *Let $Y_\nu = \langle Y, \eta_\nu \rangle_{\tilde{\mathscr{Y}}}$ and $\beta_\nu = \langle \beta, \psi_\nu \rangle_{\tilde{\Theta}}$, $\nu = 1,2,\dots$. Then*

$$\mathscr{L}(\beta_1, \beta_2, \dots | Y) = \bigotimes_\nu L(\beta_\nu | Y_\nu) = \bigotimes_\nu N\Big(\big(\sigma^2 + \alpha_\nu^2\big)^{-1} \alpha_\nu Y_\nu, \big[1 + (\alpha_\nu/\sigma)^2\big]^{-1} \Big),$$

that is, the posterior distribution of $\beta_1, \beta_2, \ldots$ is that of independent normal random variables with the indicated means and variances.

PROOF. Put $\varepsilon_\nu = \langle \varepsilon, \eta_\nu \rangle_{\tilde{\mathscr{Y}}}$. Then $\beta_1, \beta_2, \ldots, \varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. $N(0,1)$. Also $Y_\nu = \alpha_\nu \beta_\nu + \sigma \varepsilon_\nu$. The result follows from this and elementary calculations. □

As the inclusion $\iota\colon \Theta \to \overline{\Theta}$ is continuous, every $\zeta \in \overline{\Theta}^*$, the dual of $\overline{\Theta}$, defines by restriction an element of Θ^*. If we identify $\Theta = \Theta^*$, then $\iota^* \in \mathscr{B}(\overline{\Theta}^*, \Theta)$ satisfies $\langle \zeta, \iota\theta \rangle_{\overline{\Theta}^*, \overline{\Theta}} = \langle \iota^* \zeta, \theta \rangle_\Theta$, $\forall\ \theta \in \Theta$. An inner product type notation is being used for the $(\overline{\Theta}^*, \overline{\Theta})$ duality pairing.

THEOREM 2.4

(a) *The posterior $\mathscr{L}(\beta|Y)$ on $\overline{\Theta}$ is Gaussian with mean*

$$\hat{\beta} = (\sigma^2 I + U)^{-1} X^* Y, \qquad \mathscr{L}(\beta, \varepsilon)\ a.s. \tag{2.4}$$

and covariance

$$V = \iota (I + \sigma^{-2} U)^{-1} \iota^*.$$

(b) *$\delta = \beta - \hat{\beta}$ is independent of Y, and has a Gaussian distribution on $\overline{\Theta}$ with mean* 0 *and covariance V.*

REMARK. For (2.4) to make sense, we need to show X^*Y is defined a.s. and an element of Θ. Put

$$X^* Y = \sum_\nu Y_\nu X^* \eta_\nu = \sum_\nu \alpha_\nu Y_\nu \psi_\nu.$$

Note that $\Sigma E[(\alpha_\nu Y_\nu)^2] = \Sigma \alpha_\nu^2 (\alpha_\nu^2 + \sigma^2) < \infty$, so the series on the right converges a.s. in Θ. Furthermore, we have for all $\theta \in \Theta$ that

$$\langle \theta, X^* Y \rangle_\Theta = \langle X\theta, Y \rangle_{\tilde{\mathscr{Y}}} \quad \text{a.s.},$$

as one can check from series expansions of both sides. The covariance operator for a $\overline{\Theta}$ valued mean 0 random vector δ is a linear operator $V\colon \overline{\Theta}^* \to \overline{\Theta}$ satisfying $E\langle \zeta_1, \delta \rangle_{\overline{\Theta}^*, \overline{\Theta}} \langle \zeta_2, \delta \rangle_{\overline{\Theta}^*, \overline{\Theta}} = \langle \zeta_1, V\zeta_2 \rangle_{\overline{\Theta}^*, \overline{\Theta}}$ for all $\zeta_1, \zeta_2 \in \overline{\Theta}^*$.

PROOF OF THEOREM 2.4. $\beta = \Sigma_\nu \psi_\nu$ with a.s. convergence in $\overline{\Theta}$. Hence, if $\zeta \in \overline{\Theta}^*$,

$$\langle \zeta, \beta \rangle_{\overline{\Theta}^*, \overline{\Theta}} = \sum_\nu \beta_\nu \langle \zeta, \iota\psi_\nu \rangle_{\overline{\Theta}^*, \overline{\Theta}} = \sum \beta_\nu \langle \iota^* \zeta, \psi_\nu \rangle_\Theta.$$

As $\Sigma_\nu \langle \iota^* \zeta, \psi_\nu \rangle_\Theta^2 = \|\iota^* \zeta\|_\Theta^2 < \infty$, it follows from Lemma 2.3 that $\mathscr{L}(\langle \zeta, \beta \rangle_{\overline{\Theta}^*, \overline{\Theta}} | Y)$ is Gaussian with mean

$$\mu = \sum_\nu (\sigma^2 + \alpha_\nu^2)^{-1} \alpha_\nu Y_\nu \langle \iota^* \zeta, \psi_\nu \rangle_\Theta$$

and variance

$$\tau^2 = \sum_\nu \left[1 + (\alpha_\nu/\sigma)^2\right]^{-1} \langle \iota^*\zeta, \psi_\nu\rangle_\Theta^2.$$

Note that

$$\begin{aligned}\left\langle \zeta, \iota(\sigma^2 I + U)^{-1} X^* Y\right\rangle_{\overline{\Theta}^*, \Theta} &= \left\langle (\sigma^2 I + U)^{-1} \iota^*\zeta, X^* Y\right\rangle_\Theta \\ &= \sum_\nu \left\langle (\sigma^2 I + U)^{-1} \iota^*\zeta, \psi_\nu\right\rangle_\Theta (\alpha_\nu Y_\nu) \\ &= \sum_\nu \alpha_\nu \left(\sigma^2 + \alpha_\nu^2\right)^{-1} \langle \iota^*\zeta, \psi_\nu\rangle_\Theta Y_\nu = \mu.\end{aligned}$$

Also $\langle \zeta, \iota(I + \sigma^2 U)^{-1}\iota^*\zeta\rangle_{\overline{\Theta}^*, \Theta} = \tau^2$. This completes the proof of (a).

Put $\delta_\nu = \langle \delta, \psi_\nu\rangle_{\tilde{\Theta}} = \beta_\nu + (\sigma^2 + \alpha_\nu^2)^{-1}\alpha_\nu Y_\nu$. It is easy to check that δ_ν is Gaussian and independent of Y_ν, hence Y. The rest of (b) is easy. □

Now we introduce a parametrized family of Hilbert spaces which is natural for the problem. For convenience, assume

$$\mathscr{R}(X^*) \text{ is dense in } \Theta, \tag{2.5}$$

where $\mathscr{R}(X^*) = \mathscr{R}(U)$ is the range of X^*. Also, put

$$\gamma_\nu = \alpha_\nu^{-2}, \qquad \phi_\nu = \alpha_\nu^{-1}\psi_\nu.$$

For $\rho \in \mathbb{R}$ let Θ_ρ be the Hilbert space obtained by completing the set of finite-norm elements of Θ under the norm given by $\|\theta\|_\rho = \langle \theta, \theta\rangle_\rho^{1/2}$, where

$$\langle \theta_1, \theta_2\rangle_\rho = \sum_\nu \gamma_\nu^\rho \langle \theta_1, U\phi_\nu\rangle_\Theta \langle \theta_2, U\phi_\nu\rangle_\Theta.$$

We collect some elementary facts about these spaces.

PROPOSITION 2.5. *Suppose* (2.5) *holds, then*:

(a) $\langle \theta, \zeta\rangle_0 = \langle \theta, U\zeta\rangle_\Theta$, *and* $\langle \phi_\nu \colon \nu = 1, 2, \dots\rangle$ *is a c.o.n.s. for* Θ_0.

(b) $\Theta_1 = \Theta$.

(c) *X extends to a Hilbert space isomorphism from Θ_0 to the $\mathscr{Y}$ closure of $\mathscr{R}(X)$.*

(d) *If $\rho > \tau$, then $\Theta_\rho \subseteqq \Theta_\tau$ with continuous inclusion.*

(e) *Letting ι_ρ denote the inclusion $\Theta \to \Theta_\rho$ where $\rho \le 1$, and identifying Θ_ρ^* with Θ_ρ in the usual way, we have $\iota_\rho^* = U^{1-\rho}$.*

(f) *Θ_ρ supports $\mathscr{L}(\beta)$ if and only if $\sum_\nu \gamma_\nu^{\rho-1} < \infty$.*

PROOF. (a) and (b) are elementary. For (c) note that $X\phi_\nu = \eta_\nu$ and $\langle \theta, U\phi_\nu\rangle_\Theta = \langle X\theta, \eta_\nu\rangle_{\mathscr{Y}}$. For (d) note that $\|\theta\|_\rho^2 \le (\max_\nu \gamma_\nu^{\rho-\tau})\|\theta\|_\tau^2$, and the $\max_\nu \gamma_\nu^{\rho-\tau}$ is achieved at a finite, positive value by (2.3). For (e), note that $U^{1-\rho}\phi_\nu = \gamma_\nu^{-(1-\rho)}\phi_\nu$ and by comparison of the series defining the two sides of $\langle \iota_\rho\theta, \eta\rangle_\rho = \langle \theta, \iota_\rho^*\eta\rangle_1$, one obtains that $\iota_\rho^*\eta = \sum_\nu \gamma_\nu^{-(1-\rho)}\langle \eta, U\phi_\nu\rangle_1 \phi_\nu$ which gives the result.

Turning to (f), suppose $\Sigma_\nu \gamma_\nu^{\rho-1} < \infty$. Then the operator $D\colon \Theta \to \Theta$ given by $D\theta = \Sigma_\nu \gamma_\nu^{(\rho-1)/2}\langle\theta, \phi_\nu\rangle_0 \phi_\nu$ is Hilbert–Schmidt, so the norm $|||\cdot|||$ on Θ given by $|||\theta||| = \|D\theta\|_1$ is a measurable norm [Exercise 17, page 59 of Kuo (1975)] and hence the completion of Θ under $|||\cdot|||$ supports $\mathscr{L}(\beta)$ [Theorem 4.1, page 63 of Kuo (1975)]. One easily checks that $|||\cdot||| = \|\cdot\|_\rho$. Conversely, suppose Θ_ρ supports $\mathscr{L}(\beta)$, and let us calculate the covariance operator V for $\mathscr{L}(\beta)$ on Θ_ρ. For $\theta \in \Theta$,

$$\begin{aligned}\langle\theta, V\theta\rangle_\rho &= E\langle\theta,\beta\rangle_\rho^2 = E\left[\left(\sum_\nu \gamma_\nu^\rho \langle\theta,\phi_\nu\rangle_0 \langle\beta,\phi_\nu\rangle_0\right)^2\right]\\ &= E\left[\left(\sum_\nu \gamma_\nu^{\rho-1/2}\langle\theta,\phi_\nu\rangle_0 \beta_\nu\right)^2\right] = \sum_\nu \gamma_\nu^{2\rho-1}\langle\theta, U\phi_\nu\rangle_1^2\\ &= \sum_\nu \gamma_\nu^\rho \langle\theta,\phi_\nu\rangle_0 \langle U^{\rho-1}\theta,\phi_\nu\rangle_0 = \langle\theta, U^{\rho-1}\theta\rangle_\rho,\end{aligned}$$

and since Θ is dense in Θ_ρ, we can identify $V = U^{\rho-1} = \iota_\rho\iota_\rho^*$. Now V must be a trace class operator [Theorem 2.3(a), page 29, and Definition 2.2, page 16 of Kuo, (1975)], and the eigenvalues of V are $\{\gamma_\nu^{\rho-1}\}$, so $\Sigma\gamma_\nu^{\rho-1} < \infty$. □

3. Continuous time estimation. In this section we consider a continuous time analog of the problem from Section 1. It will be seen in the next section that the sequence of experiments with discrete observations can be sufficiently well approximated by a sequence of experiments with continuous observations that the limit theory of the latter is inherited by the former.

Suppose $\langle\beta(t)\colon 0 \le t \le 1\rangle$ is the stochastic process of (1.2) and let $\langle\varepsilon(t)\colon 0 \le t \le 1\rangle$ be a standard Gaussian white noise. Consider the sequence of experiments with observations

$$Y_n(t) = \beta(t) + n^{-1/2}\sigma\varepsilon(t), \qquad 0 \le t \le 1,$$

where $\sigma > 0$ is a constant. We write

$$\lambda_n = \sigma_n^2 = \sigma^2/n. \tag{3.1}$$

Now the GHS for $\mathscr{L}(\varepsilon)$ is

$$\mathscr{Y} = L_2[0,1].$$

Θ, the GHS for $\mathscr{L}(\beta)$, is equal as a set to the Sobolev space

$$\begin{aligned}W_2^m &= W_2^m[0,1]\\ &= \{f\colon f \text{ maps } [0,1] \to \mathbb{R}, \text{ and}\\ &\qquad f, Df, \dots, D^{m-1}f \text{ are absolutely continuous with } D^m f \in L_2[0,1]\}.\end{aligned}$$

The Θ inner product is

$$\langle\theta,\zeta\rangle_\Theta = \mathbf{B}(\theta)'\Sigma^{-1}\mathbf{B}(\zeta) + \int_0^1 (L\theta)(L\zeta).$$

One can show that $\|\cdot\|_\Theta$ is equivalent to $\|\cdot\|_{W_2^m}$, where

$$\|\theta\|_{W_2^m}^2 = \|D^m\theta\|_{L_2}^2 + \|\theta\|_{L_2}^2.$$

Now let $\overline{\Theta}$ be a Banach space which will support $\mathscr{L}(\beta)$ and is continuously imbedded in L_2 (i.e., $\overline{\Theta} \subseteqq L_2$ and has a stronger norm), for example, $\overline{\Theta} = C^{m-1}$ or W_2^{m-1}. Let $\overline{X}$: $\overline{\Theta} \to L_2$ be the imbedding operator ($\overline{X}\theta = \theta$), so $Y = \overline{X}\beta + \sigma_n\varepsilon$. Then X, the restriction of $\overline{X}$ to Θ, is the imbedding of Θ in L_2. To identify X^*, start with

$$\int_0^1 (L\theta)(LX^*\eta) + \mathbf{B}(\theta)'\Sigma^{-1}\mathbf{B}(X^*\eta) = \int_0^1 \theta\eta$$

valid for $\theta \in \Theta$ and $\eta \in \mathscr{Y}$. Apply integration by parts to the integral [as in Section 1.5 of Naimark (1967)] and express the l.h.s. as

$$\int_0^1 (L^*L)(X^*\eta)\theta + \sum_{j=0}^{m-1} \left[C_{j0}(X^*\eta)D^j\theta(0) + C_{j1}(X^*\eta)D^j\theta(1)\right].$$

Here L^* is the formal adjoint of L. Also, $C_{ji}(\zeta)$ is a boundary value operator of order $2m-j$ whose highest order term is of the form $\pm a_m^2 D^{2m-j}\zeta(i)$. Thus $\zeta = X^*\eta$ can be obtained as a solution of the differential equation

$$L^*L\zeta = \eta,$$
$$C_{ji}(\zeta) = 0, \qquad 0 \le j < m, i = 1, 2.$$

To derive $\hat{\beta}_n = E[\beta|Y_n] = (\lambda_n I + U)^{-1}X^*Y_n$, it is convenient to use the following characterization: $\hat{\beta}_n$ is the unique $\beta \in \Theta$ such that for all $\theta \in \Theta$,

$$\lambda_n\langle\beta, \theta\rangle_\Theta + \langle X\beta, X\theta\rangle_{\mathscr{Y}} = \langle Y_n, X\theta\rangle_{\mathscr{Y}}.$$

If one writes this out and uses integration by parts, there results that $\hat{\beta}_n$ is the solution of the stochastic differential equation

$$(\lambda_n L^*L + 1)\hat{\beta} = Y_n,$$
$$C_{ji}(\hat{\beta}) = 0, \qquad 0 \le j < m, i = 1, 2.$$

The ϕ_ν's and γ_ν's are the eigenvectors and eigenvalues of the differential operator L^*L with boundary conditions $C_{ji} = 0$, $0 \le j < m$, $i = 0, 1$. Since the highest order term of $C_{ji}(\zeta)$ is $\pm a_m^2 D^{2m-j}\zeta(i)$, the boundary conditions are regular by the argument on pages 60–61 of Naimark (1967). By Theorem 2, pages 64–65 of Naimark,

$$\gamma_\nu = a_m^2(\pi\nu)^{2m}\left(1 + O(\nu^{-3/2})\right), \qquad \nu \to \infty. \tag{3.2}$$

(Note that the eigenvalues have to be real and positive.) Put

$$C_k(\lambda, \rho) = \sum_\nu \gamma_\nu^\rho(1 + \lambda\gamma_\nu)^{-k}.$$

These quantities will have many uses. In particular note that

$$E\|\hat{\beta}-\beta\|_\rho^2 = \lambda_n C_1(\lambda_n,\rho),$$

$$E\left\|\hat{\beta}-E\left[\hat{\beta}|\beta\right]\right\|_\rho^2 = \lambda_n C_2(\lambda_n,\rho).$$

Using (3.2), one can show that if $-1/(2m)<\rho<1-1/(2m)$, then as $\lambda\to 0$

$$\begin{aligned} C_k(\lambda,\rho) &\sim \kappa\lambda^{-(\rho+1/(2m))}\int_0^\infty x^{2m\rho}\left[1+x^{2m}\right]^{-k}dx \\ &= \kappa\lambda^{-(\rho+1/(2m))}B(\rho+1/(2m),k-\rho-1/(2m)), \end{aligned} \tag{3.3}$$

where

$$\kappa = \left(2\pi m a_m^{1/m}\right)^{-1}$$

and $B(u,\nu)=\Gamma(u)\Gamma(\nu)/\Gamma(u+\nu)$ is the beta function. The derivation for the asymptotic formula for $C_k(\lambda,\rho)$ follows as in Corollary 5.4 of Speckman (1981). The evaluation of the integral is given in Gradshteyn and Ryzhik (1965), Formula 3.194.3, page 285.

Now $\Theta_0=L_2$, so our results for $\rho=0$ will imply the claims made in the Introduction. Also, $\Theta_1\cong W_2^m$, where $\cong$ means equal as sets and with equivalent norms. From this it follows that

$$\Theta_\rho \cong W_2^{m\rho}, \qquad 0\le\rho\le 1. \tag{3.4}$$

The proof is most easily accomplished with the K-method of interpolation; see Theorem 3.4 of Cox (1988).

One of the reasons the norms $\|\cdot\|_\rho$ are so useful is because we can derive explicit limiting distribution results as $\lambda_n\to 0$ for the norm of the error vector $\delta_n=\beta-\hat{\beta}_n$.

THEOREM 3.1. *Let* $-1/(2m)<\rho<1-1/(2m)$.

(a) *Put*

$$\mu_n(\rho) = \lambda_n C_1(\lambda_n,\rho),$$

$$\tau_n^2(\rho) = 2\lambda_n^2 C_2(\lambda_n,2\rho).$$

Then

$$\mathscr{L}\left(\left[\|\delta_n\|_\rho^2-\mu_n\right]/\tau_n\right) \quad\Rightarrow\quad N(0,1)$$

as $n\to\infty$.

(b) *Put*

$$\Omega_n(\rho,\beta) = \lambda_n^2\sum_\nu \gamma_\nu^{\rho+1}(1+\lambda_n\gamma_\nu)^{-2}(\beta_\nu^2-1),$$

$$\eta_n^2(\rho) = 2\lambda_n^2\left[C_3(\lambda_n,2\rho)+\lambda_n C_4(\lambda_n,1+2\rho)\right].$$

Then

$$\mathscr{L}\Big(\big[\|\delta_n\|_\rho^2 - \mu_n - \Omega_n\big]/\eta_n|\beta\Big) \quad\Rightarrow\quad N(0,1)$$

as $n \to \infty$, $\mathscr{L}(\beta)$ *a.s.*

REMARKS. In view of Theorem 2.4(b), part (a) gives the asymptotic posterior distribution for $\|\delta_n\|_\rho^2$. Part (b) gives the asymptotic sampling distribution of $\|\delta_n\|_\rho^2$. Note that Ω_n is random as it depends on β.

PROOF OF THEOREM 3.1. For (a), note that

$$\|\delta_n\|_\rho^2 = \sum_\nu \gamma_\nu^\rho \langle \delta_n, \phi_\nu\rangle_0^2 = \sum_\nu \gamma_\nu^{\rho-1}\langle \delta_n, \psi_\nu\rangle_\Theta^2$$

and $\langle \delta_n, \psi_1\rangle_\Theta, \langle \delta_n, \psi_2\rangle_\Theta, \dots$ are independent normal with mean 0 and variance $\operatorname{Var}\langle \delta_n, \psi_\nu\rangle_\Theta = \lambda_n\gamma_\nu(1+\lambda_n\gamma_\nu)^{-1}$. The result follows from Lindeberg's central limit theorem after suitable truncation of the infinite series for $\|\delta_n\|_\rho^2$.

Part (b) is somewhat more difficult. The conditional distribution of $\langle \delta_n, \psi_1\rangle_\Theta, \langle \delta_n, \psi_2\rangle_\Theta, \dots$ given β is that of independent normals with mean $\lambda_n\gamma_\nu\beta_\nu(1+\lambda_n\gamma_\nu)^{-1}$ and variance $\lambda_n\gamma_\nu(1+\lambda_n\gamma_\nu)^{-2}$. From this it follows that $E[\|\delta_n\|_\rho^2|\beta] = \mu_n + \Omega_n$ and

$$\operatorname{Var}\Big[\|\delta_n\|_\rho^2|\beta\Big] = \eta_n^2 + 4\lambda_n^3\sum_\nu \gamma_\nu^{1+2\rho}\big(\beta_\nu^2 - 1\big)\big(1+\lambda_n\gamma_\nu\big)^{-4}.$$

From (3.3) we have $\eta_n^2 \approx \lambda_n^{2-2\rho-1/(2m)}$ while from Lemma 3.2, the second term above is $o(\lambda_n^{2-2\rho-1/(2m)})$, a.s. Now the proof follows as in part (a). □

LEMMA 3.2. *Let* $\tau > 1/(2m)$ *and* k *be such that*

$$k > \tau + 1/(4m). \tag{3.5}$$

Put

$$\omega(n) = \sum_{\nu=1}^{\infty} (\lambda_n\gamma_\nu)^\tau (1+\lambda_n\gamma_\nu)^{-k}\big(\beta_\nu^2 - 1\big).$$

Then for some $C \in (0,\infty)$,

$$\limsup_{n\to\infty} \lambda_n^{1/(4m)}(\log\log n)^{-1/2}\omega(n) = C, \quad \text{a.s.},$$

$$\liminf_{n\to\infty} \lambda_n^{1/(4m)}(\log\log n)^{-1/2}\omega(n) = -C, \quad \text{a.s.}$$

PROOF. The first step is to show that there is a probability space carrying a probabilistic replica of $\omega(n)$, also denoted $\omega(n)$, and a standard Wiener process $\{W(t)\colon t \ge 0\}$ such that the process given by

$$V(s) = -\int_0^\infty W(st)g(t)\,dt, \qquad s \ge 0,$$

$$g(t) = (d/dt)h(t^{2m}), \qquad h(x) = x^\tau/(1+x)^k$$

satisfies

$$(3.6)\qquad \left|\omega(n) - 2^{1/2}V\left((c\lambda_n)^{-1/(2m)}\right)\right| = o\left(\lambda_n^{-1/(4m)}(\log\log n)^{1/2}\right)$$

almost surely, where

$$c = \left(2\pi m a_m^{1/m}\right)^{2m}.$$

Note that (3.5) guarantees convergence of the integral defining V.

Let $S_\nu = \sum_{j=1}^{\nu} 2^{-1/2}(\beta_j^2 - 1)$ and

$$D_1(n) = 2^{-1/2}\omega(n) = -\sum_\nu S_\nu[h(\lambda_n\gamma_{\nu+1}) - h(\lambda_n\gamma_\nu)],$$

where the last formula follows by partial summation. By Theorem 2.6.1, page 107 of Csörgő and Révész (1981), there is a probability space carrying a version of $\{S_\nu: \nu = 1, 2, \dots\}$ and a Wiener process $\{W(t): t \geq 0\}$ such that for all $\delta > 0$, as $\nu \to \infty$

$$(3.7)\qquad |S_\nu - W(\nu)| = o(\nu^\delta) \quad \text{a.s.}$$

Let

$$D_2(n) = -\sum_\nu W(\nu)[h(\lambda_n\gamma_{\nu+1}) - h(\lambda_n\gamma_\nu)].$$

Note that

$$|h(\lambda\gamma_{\nu+1}) - h(\lambda\gamma_\nu)| = \lambda[\gamma_{\nu+1} - \gamma_\nu]|h'(\lambda\tilde{\gamma}_\nu)|,$$

where $\gamma_\nu \leq \tilde{\gamma}_\nu \leq \gamma_{\nu+1}$. Utilizing (3.2) one can show that

$$(3.8)\qquad \gamma_{\nu+1} - \gamma_\nu = c2m\nu^{2m-1}\left(1 + O(\nu^{-1/2})\right).$$

Also,

$$|h'(\lambda\gamma_\nu)| \leq C_1\left[(\lambda\nu^{2m})^{\tau-1}(1 + c\lambda\nu^{2m})^{-k} + (\lambda\nu^{2m})^\tau(1 + c\lambda\nu^{2m})^{-(k+1)}\right]$$

for some $C_1 \in (0, \infty)$. Collecting things together we have

$$(3.9)\qquad \begin{aligned} |D_1(n) - D_2(n)| \leq{}& C_2\lambda_n^{-\delta/(2m)}\sum_\nu \left(\lambda_n\nu^{2m}\right)^{1+(\delta-1)/(2m)} \\ &\times\left[\left(\lambda_n\nu^{2m}\right)^{\tau-1}\left(1 + c\lambda_n\nu^{2m}\right)^{-k}\right. \\ &\left.+\left(\lambda_n\nu^{2m}\right)^\tau\left(1 + c\lambda_n\nu^{2m}\right)^{-(k+1)}\right]\lambda_n^{1/(2m)}, \end{aligned}$$

where $C_2 \in (0, \infty)$ is random [from (3.7)] and depends on δ. Note that the summation above tends to a finite integral as $n \to \infty$. Hence, taking $\delta < 1/2$ gives that

$$|D_1(n) - D_2(n)| = o\left(\lambda_n^{-1/(4m)}(\log\log n)^{1/2}\right).$$

Next we show that $D_2(n)$ is suitably close to

$$D_3(n) = -\sum_\nu W(\nu)(c\lambda_n)^{1/(2m)}g\left((c\lambda_n)^{1/(2m)}\nu\right).$$

Two applications of the mean value theorem along with (3.8) yield

$$|D_2(n) - D_3(n)|$$
$$\leq C\sum_\nu |W(\nu)|\left(\lambda_n \nu^{2m-1}\right)^2(1 + O(\nu^{-1/2}))\left|h''\left(c\lambda_n \nu^{2m}\left[1 + O(\nu^{-1})\right]\right)\right|$$

for some $C \in (0, \infty)$, almost surely. If one uses the fact that $|W(\nu)| = O(\nu)$ a.s. and applies the same argument that was used on (3.9), it can be shown that the last expression is $O(1)$ a.s., which is $o(\lambda_n^{-1/(4m)}(\log\log n)^{1/2})$. [The assumption $\tau > 1/(2m)$ is needed here to guarantee that a summation converges to a finite integral.]

Now let $D_4(n) = V((c\lambda_n)^{-1/(2m)})$. Then

$$|D_3(n) - D_4(n)| \leq (c\lambda_n)^{-1/(2m)}\sum_\nu Z_\nu d_{n\nu}, \tag{3.10}$$

where

$$Z_\nu = \sup_{\nu \leq t \leq \nu+1} |W(t) - W(\nu)|$$

is a sequence of i.i.d. random variables and

$$d_{n\nu} = \left|g'\left([c\lambda_n]^{1/(2m)}[\nu + \alpha_{n\nu}]\right)\right|$$

for some $\alpha_{n\nu} \in [0, 1]$. Now $P[Z_\nu \geq y] = 4P[W(1) \geq y]$, so as $\nu \to \infty$,

$$Z_\nu = O(\log \nu) \quad \text{a.s.}$$

Plugging this into (3.10) and using a familiar argument shows that for any $\delta > 0$

$$|D_3(n) - D_4(n)| = O\left(\lambda_n^{-\delta}\right) \quad \text{a.s.}$$

as $n \to \infty$. This completes the proof of (3.6).

Now we show that if $s_n = \pi^{-1}\lambda_n^{-1/(2m)}$, then for some $C \in (0, \infty)$

$$\limsup_n (s_n \log\log s_n)^{-1/2} V(s_n) = C \quad \text{a.s.} \tag{3.11}$$

Since $\log\log s_n = \log\log(\pi^{-1}\sigma^{-1/m} n^{1/(2m)}) \sim \log\log n$, the lemma will follow from (3.6) and (3.11). (Note that the corresponding lim inf in (3.11) will be $-C$ by symmetry.) Put

$$Y(u) e^{-1/2u} V(e^u), \qquad -\infty < u < \infty,$$
$$u_n = \log s_n = A + (2m)^{-1}\log n, \qquad n = 1, 2, \ldots,$$

where $A = -\log \pi\sigma^{1/m}$ is a constant. One can check that $Y(u)$ is a stationary mean zero Gaussian process with covariance

$$\Gamma(u) = e^{-1/2u}\int_0^\infty G(e^{-u}t)G(t)\,dt,$$
$$G(t) = h(t^m).$$

Equation (3.11) will follow by showing

$$\limsup_n (\log u_n)^{-1/2} Y(u_n) = (2\Gamma(0))^{1/2} \quad \text{a.s.} \tag{3.12}$$

To prove this latter we will apply Case 2 of Theorem D of Qualls (1977) with $\alpha = 2$. Assuming this result applies, it states that for $c > 0$,

$$P\Big[Y(u_n) > c\big(\Gamma(0)\log u_n\big)^{1/2}\Big] = 0 \quad \text{or} \quad 1$$

according as the integral

$$\int_1^\infty t^{-c^2/2}(\log t)^{-c^2/4}\big(\log t - \tfrac{1}{2}\log\log t\big)^{-1/2}\,dt < \infty \quad \text{or} \quad = \infty.$$

This latter integral converges for $c > 2^{1/2}$ and diverges for $c < 2^{1/2}$, which establishes (3.12).

In order to apply the aforementioned result in Qualls (1977), it is necessary to check that Y has continuous paths, $\Gamma(u) = o(1/\log u)$ as $u \to \infty$, and $\Gamma''(0) < 0$. To show Y is a.s. continuous, it suffices to show that V is a.s. continuous on $[0, \infty)$, and for this it suffices to show V is a.s. continuous on any finite interval $[0, b]$ and then let $b \to \infty$ through some countable set. Now let A be an event of probability 1 on which W is continuous and for which there exists T_0 (depending on the path of W) such that for all $T \geq T_0$,

$$(2T\log\log T)^{-1/2} \sup_{\substack{0 \le t \le T/2 \\ 0 \le s \le T/2}} |W(t+s) - W(t)| \leq 2. \tag{3.13}$$

Such an event A exists by equation (1.2.4), page 30 of Csörgő and Révész (1981). Let $\varepsilon > 0$ be given. Pick $T_1 \geq T_0$ such that

$$4\int_{T_1/2b}^\infty 4(bt\log\log bt)^{1/2}|g(t)|\,dt < \varepsilon/2. \tag{3.14}$$

This is possible by (3.5). Now

$$\begin{aligned}
&\sup_{0 \le s_1 \le s_2 \le b} |V(s_1) - V(s_2)| \\
&\quad \leq \int_{T_1/2b}^\infty \sup_{0 \le s_1 \le s_2 \le b} |W(s_1 t) - W(s_2 t)|\,|g(t)|\,dt \\
&\qquad + \int_0^{T_1/2b} \sup_{0 \le s_1 \le s_2 \le b} |W(s_1 t) - W(s_2 t)|\,|g(t)|\,dt = I_1 + I_2.
\end{aligned}$$

Since

$$\sup_{0 \le s_1 \le s_2 \le b} |W(s_1 t) - W(s_2 t)| \leq \sup_{\substack{0 \le u \le bt \\ 0 \le v \le bt}} |W(u + v) - W(u)|$$

it follows from (3.13) and (3.14) that $I_1 < \varepsilon/2$. A straightforward uniform continuity argument shows that there is a $\delta > 0$ such that $|s_1 - s_2| \leq \delta$ implies $I_2 < \varepsilon/2$. This completes the proof of continuity.

To show $\Gamma(u) = o(1/\log u)$ as $u \to \infty$, we have for $u > 0$ and $\tau \neq (1/2)k - m$

$$\Gamma(u) \le e^{-(1/2+\tau/(2m))u}\left[\int_0^1 t^{t/m}\left[1 + \tau^{1/(2m)}\right]^{-k} dt + \int_1^{e^u} t^{(2\tau-k)/(2m)}\,dt + e^{ku/(2m)}\int_{e^u}^{\infty} t^{(\tau-k)/m}\,dt\right]$$
$$= O\left(e^{-(1/2+\tau/(2m))u}\right) + O\left(e^{((\tau-k)/(2m)+1/2)u}\right).$$

As all the exponents are negative, the desired result follows. If $\tau = (1/2)k - m$, then the middle integral is treated differently but the result still holds.

To show that $\Gamma''(0) < 0$, first note that

$$\Gamma''(0) = \tfrac{1}{4}\int_0^{\infty} G^2(t)\,dt + \int_0^{\infty}\left[2tg(t) + t^2 g'(t)\right]G(t)\,dt. \tag{3.15}$$

The calculation is carried out by differentiating under the integral sign which is justified by the dominated convergence theorem and by the fact that there exists a constant c such that for all $u \in [-1, 1]$, $|G(e^{-u}t)|$, $|tg(e^{-u}t)|$ and $|t^2 g'(e^{-u}t)|$ are all bounded by $c \min\{1, t\}^{2m(\tau-k)}$, which is square integrable on $(0, \infty)$ by (3.5). Noting that the quantity in brackets on the r.h.s. of (3.15) is $(d/dt)[t^2 g(t)]$, an integration by parts will yield

$$\Gamma''(0) = \tfrac{1}{4}\int_0^{\infty} G^2(t)\,dt - \int_0^{\infty} t^2 g^2(t)\,dt.$$

Let I_1 denote the first integral in this last expression and I_2 the second. Then integration by parts gives

$$I_1 = -2\int_0^{\infty} tg(t)G(t)\,dt$$

and Cauchy–Schwarz applied to this integral gives $I_1 < 2I_1^{1/2}I_2^{1/2}$ which implies $I_1 < 4I_2$ and hence $\Gamma''(0) = (1/4)I_1 - I_2 < 0$. □

Corollary 3.3. *Let* $-1/(2m) < \rho < 1 - 1/(2m)$. *Suppose that* $\Delta_n > 0$ *is chosen so that*

$$P\left[\|\hat{\beta}_n - \beta\|_\rho^2 \le \Delta_n | Y_n\right] = 1 - \alpha \tag{3.16}$$

for some $\alpha \in (0, 1)$. *Then*

$$\liminf_{n\to\infty} P\left[\|\hat{\beta}_n - \beta\|_\rho^2 \le \Delta_n | \beta\right] = 0, \quad a.s. \tag{3.17}$$

Proof. By Theorem 3.1(a)

$$\Delta_n \sim \mu_n + z_\alpha \tau_n, \tag{3.18}$$

where z_α is the upper $100\alpha\%$ point of the $N(0,1)$ distribution. By Theorem 3.1(b)

$$\left|P\left[\|\hat{\beta}_n - \beta\|_\rho^2 \le \Delta_n|\beta\right] - \Phi((z_\alpha\tau_n - \Omega_n)/\eta_n)\right| \to 0$$

a.s. as $n \to \infty$, where Φ denotes the $N(0,1)$ distribution function. [Note that convergence in distribution to $N(0,1)$ implies uniform convergence of the distribution function to Φ.] Now τ_n/η_n tends to a finite limit and

$$\eta_n \approx \lambda_n^{1-\rho-1/(4m)}$$

by (3.3). Hence, by Lemma 3.2

$$\liminf_n (z_\alpha\tau_n - \Omega_n)/\eta_n = -\infty,$$

which proves (3.17). □

COROLLARY 3.4. *With the setup of the previous corollary, the prior probability that the* $(1-\alpha)$ *Bayesian posterior probability region* $\{\theta \in \Theta_\rho\colon \|\hat{\beta}_n - \theta\|_\rho^2 \le \Delta_n\}$ *is conservative in its frequentist coverage probability tends to* $\Phi(z_\alpha(1 - r^{1/2})(1-r^2)^{-1/2})$, *where*

$$r^2 = (1-\rho-1/(4m))(2\rho/3 - 1/(6m)).$$

PROOF. We wish to calculate

$$P\left\{\beta\colon P\left[\|\hat{\beta}_n - \beta\|_\rho^2 \le \Delta_n|\beta\right] \ge 1-\alpha\right\}.$$

The difference between this and

$$P\left[\Phi((z_\alpha\tau_n - \Omega_n)/\eta_n) \ge 1-\alpha\right] = P\left[\Omega_n \le (\tau_n - \eta_n)z_\alpha\right]$$

tends to 0. One can show as in Theorem 3.1(a) that $\mathscr{L}(\Omega_n/(\tau_n^2 - \eta_n^2)^{1/2}) \Rightarrow N(0,1)$. Using this and (3.3) one can show that the last displayed probability tends to a limit of the form $\Phi(z_\alpha R)$ where R is an algebraic form in some beta functions. If one substitutes $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ and uses the recurrence relation $\Gamma(x+1) = x\Gamma(x)$, then the claimed result follows after some algebraic manipulation. □

4. Discrete time estimation. In this section we consider the original discrete time problem in (1.1), where β is given the prior specified in (1.2). By virtue of the next result, it follows that Theorem 3.1 and Corollaries 3.3 and 3.4 hold in this setting as well, thus justifying the statements made in the introduction.

THEOREM 4.1. *Suppose that* $m \ge 2$ *and*

$$-1/(2m) < \rho < 1 - 1/(2m)$$

and that

$$E|\varepsilon_i|^{2+\delta} < \infty,$$

where

$$\delta \geq 6/(2m-3).$$

Then there exists a probability space carrying versions of β, $\varepsilon_1, \varepsilon_2, \dots$ and a sequence $\{\tilde{\varepsilon}_n(t)\colon 0 \leq t \leq 1\}$ of Gaussian white noises such that if

$$Y_{ni} = \beta(i/n) + \varepsilon_i, \qquad 1 \leq i \leq n,$$

$$\hat{\beta}_n = E[\beta | Y_n], \qquad \delta_n = \hat{\beta}_n - \beta,$$

$$\tilde{Y}_n(t) = \beta(t) + \sigma n^{-1/2}\tilde{\varepsilon}_n(t), \qquad 0 \leq t \leq 1,$$

$$\tilde{\beta}_n = E\big[\beta | \tilde{Y}_n\big], \qquad \tilde{\delta}_n = \tilde{\beta}_n - \beta,$$

then

$$\|\delta_n - \tilde{\delta}_n\|_\rho^2 = o_p\big(\lambda_n^{1-\rho}\big).$$

REMARKS. Note that $\|\tilde{\delta}_n\|_\rho^2 = O_p(\lambda_n^{1-\rho-1/(2m)})$ follows from $E\|\tilde{\delta}_n\|_\rho^2 = \mu_n(\rho)$. Since $\|\delta_n - \tilde{\delta}_n\|_\rho^2 = o_p(\lambda_n^{1-\rho-1/(2m)})$, it follows that $\|\delta_n\|_\rho^2 = O_p(\lambda_n^{1-\rho-1/(2m)})$. Hence

$$\Big| \|\delta_n\|_\rho^2 - \|\tilde{\delta}_n\|_\rho^2 \Big| \leq \|\delta_n - \tilde{\delta}_n\|_\rho \big[\|\delta_n\|_\rho + \|\tilde{\delta}_n\|_\rho \big]$$
$$= o_p\big(\lambda_n^{(1-\rho)/2}\big) O_p\big(\lambda_n^{(1-\rho-1/(2m))/2}\big) = o_p\big(\lambda_n^{1-\rho-1/(4m)}\big).$$

Now $\tau_n \approx \lambda_n^{1-\rho-1/(4m)}$ and $\eta_n \approx \lambda_n^{1-\rho-1/(4m)}$, so we may replace the continuous time estimation error (now denoted $\tilde{\delta}_n$) with the discrete time estimation error δ_n in Theorem 3.1, and the results still hold. Corollaries 3.3 and 3.4 depend only on these results and the properties of Ω_n, which depends only on β.

PROOF OF THEOREM 4.1. We have

$$\|\delta_n - \tilde{\delta}_n\|_\rho^2 \leq 2\|r_n - s_n\|_\rho^2 + 2\big\|\big(B_n - \tilde{B}_n\big)\beta\big\|_\rho^2,$$

where

$$r_n = \hat{\beta}_n - E\big[\hat{\beta}_n | \beta\big], \qquad s_n = \tilde{\beta}_n - E\big[\tilde{\beta}_n | \beta\big],$$

$$B_n\beta = E\big[\hat{\beta}_n - \beta | \beta\big] = \big[(\lambda_n I + U_n)^{-1} U_n - I\big]\beta,$$

$$\tilde{B}_n\beta = E\big[\tilde{\beta}_n - \beta | \beta\big] = \big[(\lambda_n I + U)^{-1} U - I\big]\beta.$$

We will show that each of the two quantities on the r.h.s. of the inequality above has the correct order as $n \to \infty$.

Now

$$r_n = (\lambda_n I + U_n)^{-1} X_n^* \underline{\varepsilon}_n$$

and following the notation of Cox (1984b) put

$$\bar{r}_n = (\lambda_n I + U)^{-1} X_n^* \underline{\varepsilon}_n.$$

Now we apply some results in Section 4 of Cox (1988). One can show that for all x_1, x_2 in Θ

$$\left|\langle (U - U_n)x_1, x_2\rangle_\Theta\right| \le Cn^{-1}\left[\|x_1\|_{1/m}\|x_2\|_0 + \|x_1\|_0\|x_2\|_{1/m}\right],$$

where $C \in (0, \infty)$ is a constant, so Assumption 4.1(f) of Cox (1988) holds with $k_n = Cn^{-1}$, $s = 1/m$, $j = 2$. See the remarks after this assumption for a proof of the above inequality. Now by equation (4.6) of Cox (1988),

$$E\|r_n - \bar{r}_n\|_\rho^2 \le Cn^{-2}\left[C_2(\lambda_n, \rho + 1/m)E\|r_n\|_0^2 + C_2(\lambda_n, \rho)E\|r_n\|_{1/m}^2\right].$$

This latter inequality can be inverted for $\rho = 0, 1/m$ to give bounds on $E\|r_n\|_0^2$ and $E\|r_n\|_{1/m}^2$ in terms of $E\|\bar{r}_n\|_0^2$ and $E\|\bar{r}_n\|_{1/m}^2$. When these are substituted back into this inequality and the estimates on C_2 are used there results

$$E\|r_n - \bar{r}_n\|_\rho^2 \le Cn^{-3}\lambda_n^{-\rho-2/m}.$$

It is easily seen that this last quantity is $o(\lambda_n^{1-\rho})$ as $m \ge 2$.

Now put

$$\bar{s}_n = \sigma n^{-1/2}\int_0^1 (\lambda_n I + U)^{-1}\xi(t)\, dB_n \circ F_n(t),$$

where $F_n(t) = n^{-1}[nt]$ as in (3.7) of Cox (1984b), then following (3.9) of that paper,

$$\|\bar{r}_n - \bar{s}_n\|_\rho^2 = n^{-1}C_2(\lambda_n, \rho + 1/m)o(n^{-\delta/(2+\delta)})$$

a.s. Using the estimate of C_2, the definition of λ_n, and the hypothesis that $\delta \ge 6/(2m-3)$, it follows that the last displayed quantity is $o(\lambda_n^{1-\rho})$, a.s.

Next, we have from (3.10) of Cox (1984b) that

$$\|\bar{s}_n - s_n\|_\rho^2 = O_p\left(n^{-1}(n^{-1}\log n)C_2(\lambda_n, \rho + 1/m)\right)$$

and this is $o_p(\lambda_n^{1-\rho})$. This completes the proof that

$$\|r_n - s_n\|_\rho^2 = o_p\left(\lambda_n^{1-\rho}\right).$$

Finally we must take care of $\|(B_n - \tilde{B}_n)\beta\|_\rho^2$, where B_n and $\tilde{B}_n$ denote the bias operators for the respective problems. Let $\max\{\rho, 1/m\} < \tau < 1 - 1/(2m)$, and let $\|\cdot\|_{\tau,\rho}$ denote the operator norm of a linear operator from Θ_τ to Θ_ρ. Then by equation (4.6) of Cox (1988),

$$\begin{aligned}\|B_n - \tilde{B}_n\|_{\tau,\rho} \le Cn^{-1}\Big\{&C_2^{1/2}(\lambda_n, \rho + 1/m)\|\tilde{B}_n\|_{\tau,0}\\ &+C_2^{1/2}(\lambda_n, \rho + 1/m)\lambda_n^{1/(2m)}\|\tilde{B}_n\|_{\tau,1/m}\\ &+C_2^{1/2}(\lambda_n, \rho)\lambda_n^{-1/(2m)}\|\tilde{B}_n\|_{\tau,0}\\ &+C_2^{1/2}(\lambda_n, \rho)\|\tilde{B}_n\|_{\tau,1/m}\Big\}.\end{aligned}$$

Now by Theorem 2.3(c) of Cox (1988), $\|\tilde{B}_n\|_{\tau,\rho} \le C\lambda^{(\tau-\rho)/2}$ as long as $\rho \le \tau \le \rho + 2$. Plugging this into the above expression and using the estimates on C_2

yields

$$\|B_n - \tilde{B}_n\|_{\tau,\rho} \le C\{\lambda_n^{2-\rho-3/(2m)+\tau}\}^{1/2}.$$

Now

$$\left\|(B_n - \tilde{B}_n)\beta\right\|_\rho^2 \le C\lambda_n^{2-\rho-3/(2m)+\tau}\|\beta\|_\tau^2.$$

As $\tau < 1 - 1/(2m)$, $\|\beta\|_\tau^2 < \infty$ a.s., by (3.2) and Proposition 2.5(f), and so the last expression is a.s.

$$O_p\left(\lambda_n^{2-\rho-3/(2m)+\tau}\right) = o_p\left(\lambda_n^{1-\rho}\right). \qquad \square$$

5. Concluding remarks. To provide more insight into the foregoing analysis, consider the case where $\dim \Theta = k < \infty$. Then it is easy to obtain the analog of Theorem 3.1. Calculating the asymptotic posterior distribution of $\|\delta_n\|_\rho^2$, we have

$$\mathscr{L}\left(\|\delta_n\|_\rho^2\right) = \mathscr{L}\left(\sum_{\nu=1}^{k} \lambda_n \gamma_\nu^\rho (1 + \lambda_n \gamma_\nu)^{-1} \varepsilon_\nu^2\right).$$

Since $(1 + \lambda_n\gamma_\nu)^{-1} \to 1$ as $n \to \infty$ for $\nu = 1, 2, \dots, k$, it follows that

$$\mathscr{L}\left(\lambda_n^{-1}\|\delta_n\|_\rho^2\right) \quad \Rightarrow \quad \mathscr{L}\left(\sum_{\nu=1}^{k} \gamma_\nu^\rho \varepsilon_\nu^2\right),$$

where the distribution on the right is a weighted sum of χ^2 random variables. The corresponding calculations for the sampling distribution are

$$\begin{aligned}\mathscr{L}\left(\|\delta_n\|_\rho^2 | \beta\right) = \mathscr{L}\Bigg(&\lambda_n \sum_{\nu=1}^{k} \gamma_\nu^\rho (1 + \lambda_n\gamma_\nu)^{-2} \varepsilon_\nu^2 \\ &+ \lambda_n^{3/2} \sum_{\nu=1}^{k} \gamma_\nu^{1/2+\rho} (1 + \lambda_n\gamma_\nu)^{-2} \varepsilon_\nu \beta_\nu \\ &+ \lambda_n^2 \sum_{\nu=1}^{k} \gamma_\nu^{1+\rho} (1 + \lambda_n\gamma_\nu)^{-2} \beta_\nu^2\Bigg).\end{aligned}$$

The second and third summations on the r.h.s. (which result from the bias) are each $o_p(\lambda_n^{-1})$, so

$$\mathscr{L}\left(\lambda_n^{-1}\|\delta_n\|_\rho^2 | \beta\right) \quad \Rightarrow \quad \mathscr{L}\left(\sum_{\nu=1}^{k} \gamma_\nu^\rho \varepsilon_\nu^2\right).$$

Thus, the asymptotic posterior and sampling distributions are identical, so a $(1 - \alpha)$ posterior probability region of the form in (3.16) has (asymptotically) the right coverage probability from the sampling point of view. This results from the fact that the random (conditional on β) variability dominates both the sampling and posterior distributions, with the bias $E[\hat{\beta}_n - \beta | \beta]$ being of smaller order.

In the setting of the foregoing analysis, the bias and random variability are always of the same order of magnitude. This is the driving force behind Corollaries 3.3 and 3.4.

Concerning extensions of the major results (Theorem 3.1, Lemma 3.2 and Theorem 4.1), one would like to look at more general Gaussian priors and more general quantities than $\|\delta_n\|_\rho^2$. I believe that analogous results can be obtained for $\|\delta_n\|_\rho^2$ in any setting wherein $\gamma_\nu \approx \nu^r$ for some $r > 0$ [see Cox (1988)]. The specific setting here allowed us to use the sharp eigenvalue asymptotics (3.3) at some points in Lemma 3.2 [see (3.8)]. The extension to other quadratic forms can also probably be accomplished. As in Wahba (1983), we would like to construct a confidence interval for $\beta(t)$ for some $t \in [0, 1]$. In the setup of Section 3, note that

$$\mathscr{L}(\delta_n(t)) = N\Big(0, \lambda_n \sum_\nu (1 + \lambda_n \gamma_\nu)^{-1} \phi_\nu^2(t)\Big),$$

$$\mathscr{L}(\delta_n(t)|\beta) = N\Big(M_n, \lambda_n \sum_\nu (1 + \lambda_n \gamma_\nu)^{-2} \phi_\nu^2(t)\Big),$$

where

$$M_n = \lambda_n \sum_\nu \gamma_\nu^{1/2} (1 + \lambda_n \gamma_\nu)^{-1} \beta_\nu \phi_\nu(t).$$

I conjecture that both variances above are approximately $\lambda_n^{1-1/(2m)}$, and also that

$$\limsup_{n \to \infty} \big[\lambda_n^{(1-1/(2m))} (\log \log 1/\lambda_n)\big]^{-1/2} M_n \in (0, \infty), \quad \text{a.s.}$$

This would imply analogs of Corollaries 3.3 and 3.4 for the "confidence" intervals for $\delta_n(t)$. The difficulty here is in obtaining results on the behavior of $\phi_\nu(t)$. In the setting of periodic smoothing splines, one can obtain concrete results [Cox (1989)].

REFERENCES

COX, D. (1984a). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

COX, D. (1984b). Gaussian approximation of smoothing splines. Unpublished manuscript.

COX, D. (1988). Approximation of method of regularization estimators. *Ann. Statist.* **16** 694–712.

COX, D. (1989). Coverage probability of Bayesian confidence intervals for smoothing splines. Unpublished manuscript.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403.

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics.* Academic, New York.

DIACONIS, P. (1988). Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics 4* (S. S. Gupta and J. O. Berger, eds.) **1** 163–175. Springer, Berlin.

GRADSHTEYN, I. and RYZHIK, I. (1965). *Tables of Integrals, Series, and Products.* Academic, New York.

KALLIANPUR, G. (1971). Abstract Wiener processes and their reproducing kernel Hilbert spaces. *Z. Wahrsch. Verw. Gebiete* **17** 113–123.

KALLIANPUR, G. and KARANDIKAR, R. (1985). White noise calculus and nonlinear filtering theory. *Ann. Probab.* **13** 1033–1107.

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.

KUELBS, J. (1970). Gaussian measures on a Banach space. *J. Funct. Anal.* **5** 354–367.

KUELBS, J. (1971). Expansions of vectors in a Banach space related to Gaussian measures. *Proc. Amer. Math. Soc.* **27** 364–370.

KUO, H. H. (1975). *Gaussian Measures on Banach Spaces. Lecture Notes in Math.* **463**. Springer, Berlin.

NAIMARK, M. A. (1967). *Linear Differential Operators*. Ungar, New York.

QUALLS, C. (1977). The law of the iterated logarithm on arbitrary sequences for stationary Gaussian processes and Brownian motion. *Ann. Probab.* **5** 724–739.

SPECKMAN, P. (1981). The asymptotic integrated mean square error for smoothing noisy data by splines. Technical report, Dept. Statistics, Univ. Missouri, Columbia.

SPECKMAN, P. (1983). Efficient nonparametric regression with cross validated smoothing splines. Technical report, Dept. Statistics, Univ. Missouri, Columbia.

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

WAHBA, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B* **45** 133–150.

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

DEPARTMENT OF STATISTICS
RICE UNIVERSITY
P.O. BOX 1892
HOUSTON, TEXAS 77251