INCORPORATING PRIOR INFORMATION IN MINIMAX ESTIMATION OF THE MEAN OF A GAUSSIAN PROCESS

by

Robert Wolpert[1]
Department of Mathematics
Duke University
Durham, North Carolina, U.S.A.

James Berger[2]
Department of Statistics
Purdue University
W. Lafayette, Indiana, U.S.A.

Mimeograph Series #81-37

Department of Statistics
August 1981

[1]Research supported by the National Science Foundation under Grant No. MCS-78-01737.

[2]Research supported by the Alfred P. Sloan Foundation and by the National Science Foundation under Grants No. MCS-7802300A3 and MCS-8101670.

INCORPORATING PRIOR INFORMATION IN MINIMAX ESTIMATION OF THE MEAN OF A GAUSSIAN PROCESS

by

Robert Wolpert[1]
Department of Mathematics
Duke University
Durham, North Carolina, U.S.A.

James Berger[2]
Department of Statistics
Purdue University
W. Lafayette, Indiana, U.S.A.

I. INTRODUCTION

Let $\mathcal{X}$ be the complete metric space of continuous real-valued functions on a closed set $I \subset \mathbb{R}^1$, and let Θ be a subspace of $\mathcal{X}$. Let Z be a Borel-measurable $\mathcal{X}$-valued Gaussian process on some probability space $(\Omega, \mathfrak{F}, P)$ with zero mean $0 = EZ(t)$ and known covariance $\gamma(s,t) = EZ(s)Z(t)$ for $s,t \in I$. Denote by $\gamma(s) = \gamma(s,s)$ the variance of $Z(s)$. Here (as usual) we suppress the ω-dependence of functions $Y \in L^1(\Omega, \mathfrak{F}, P)$ and denote $\int Y dP$ by EY when convenient.

We consider the problem of estimating the mean $\theta \in \Theta$ of the Gaussian process $X(t) = \theta(t) + Z(t)$, based upon the observation of one or more sample paths $\{X_1, \ldots, X_n\} \in \mathcal{X}$, under a quadratic loss function L. The usual estimator in this situation is $\delta^0[\vec{X}](t) = \bar{X}(t)$; in Section 2 we develop an estimator δ^M which incorporates prior information about θ in an intelligent manner and whose risk function $R(\theta, \delta^M) = EL(\theta, \delta^M[\vec{X}])$ satisfies

(1.1) $$R(\theta, \delta^M) < R(\theta, \delta^0) \text{ for every } \theta \in \Theta .$$

[1] Research supported by the National Science Foundation under Grant No. MCS-78-01737.

[2] Research supported by the Alfred P. Sloan Foundation and by the National Science Foundation under Grants No. MCS-7802300A3 and MCS-8101670.

It was shown in Berger and Wolpert [3] that (except in trivial cases) δ^0 is minimax but inadmissible. Indeed broad classes of estimators improving upon δ^0 were found. In selecting an alternative estimator, it was pointed out that prior information concerning $\theta(\cdot)$ must be taken into account to ensure that the region of significant risk improvement over δ^0 coincides with a set in which $\theta(\cdot)$ is believed to lie. (No estimator δ can have uniformly large risk improvement over δ^0, since δ^0 is minimax.) A simple type of prior information concerning $\theta(\cdot)$ is specification of a "best guess" $\xi(\cdot)$ for $\theta(\cdot)$ and specification of a subjective "variance function" $\lambda(\cdot)$ representing the expected squared error in the guess $\xi(\cdot)$ for $\theta(\cdot)$. Although specification of other features of the prior distribution may sometimes be possible, it would be useful to be able to proceed making use only of $\xi(\cdot)$ and $\lambda(\cdot)$. Of course, sometimes $\theta(\cdot)$ may really be random with a known distribution. In such a case one would want to use the optimal Bayes estimator (or optimal filter) for the problem. If, however, the distribution of $\theta(\cdot)$ is only approximately known, then one might well wish to use a minimax estimator employing the known features of the distribution of $\theta(\cdot)$ (as developed here), since this ensures robustness against misspecification of the distribution of $\theta(\cdot)$.

To incorporate $\xi(\cdot)$ and $\lambda(\cdot)$ in an improved estimator, it is convenient to <u>pretend</u> that $\theta(\cdot)$ is itself a Gaussian process (independent of $Z(\cdot)$) with a mean function $\xi(\cdot)$ and a variance function $\lambda(\cdot)$. Actually, we will assume that the entire prior covariance function $\lambda(s,t) = E\{[\theta(s)-\xi(s)][\theta(t)-\xi(t)]\}$ has been specified, although in Section 3 it will be shown that knowledge solely of $\lambda(\cdot)$ will suffice in many applications.

In Berger and Wolpert [3], a version of the Karhunen-Loéve expansion of $X(\cdot)$ was used to reduce the estimation problem to that of estimating a countable sequence of normal means $\{\theta_i\}$. The prior information concerning $\theta(\cdot)$ was also transformed into prior information about the θ_i, but in selecting a minimax estimator using the prior information, the covariances among the θ_i were ignored. This could potentially lead to a serious misrepresentation of the prior information. In this paper a more complicated expansion of the process is considered, one which allows use of all the prior information in selecting a minimax estimator. This expansion is developed in Section 2, which the desired minimax estimator is also derived. The implementation of this expansion is particularly easy when $\gamma(t,s)$ and $\lambda(t,s)$ commute in an appropriate sense, as discussed in Section 3.

II. THE MINIMAX ESTIMATOR

Let $\mathcal{A}$ (the "action space") be a subset of the Borel-measurable real-valued functions on I. The loss incurred in estimating $\theta \in \Theta$ by $a \in \mathcal{A}$ will be

(2.1) $$L(\theta,a) = \int|\theta(s)-a(s)|^2\mu(ds).$$

Here μ is an arbitrary but specified non-negative Borel measure satisfying

A1) $L^2(I,d\mu) \supset \Theta$,

A2) $L^2(I,d\mu) \subset \mathcal{A}$,

A3) $\gamma(\cdot) \in L^1(I,d\mu)$,

A4) $\gamma(\cdot,\cdot)$ is continuous on $I \times I$.

As in Berger and Wolpert [3] it suffices to take $\mathcal{A} = L^2(I,d\mu)$ and to consider only the case of a single observation of X.

Let $\mathcal{D}$ denote the <u>decision space</u> of all Borel-measurable mappings $\delta: \mathcal{X} \to \mathcal{A}$, and $R: \Theta \times \mathcal{D} \to \mathbb{R}_+$ denote the risk function

(2.2) $$R(\theta,\delta) \equiv EL(\theta,\delta[X]) = \int_\Omega \int_I |\theta(s)-\delta[X](s)|^2\mu(ds)dP.$$

The usual estimator $\delta^0[X] = X$ has constant risk

(2.3)
$$\begin{aligned} C &\equiv R(\theta,\delta^0) \\ &= E \int_I |\theta(s)-X(s)|^2 \mu(ds) \\ &= \int_I \gamma(s,s)\mu(ds) \\ &< \infty \quad \text{by A3)}. \end{aligned}$$

Since δ^0 is minimax (if, e.g., Θ is dense in $L^2(I,d\mu)$), any estimator δ^M satisfying (1.1) must also be minimax and, for each $\varepsilon > 0$,

(2.4)
$$A_\varepsilon^\delta \equiv \{\theta\colon R(\theta,\delta) < C-\varepsilon\}$$

must be a proper subset of Θ. When prior information about the location of θ is available it is desirable to use an estimator δ for which A_ε^δ is quite likely to contain θ. As discussed in Section 1, we will assume that prior information is available and is modeled as a Gaussian process with mean function $\xi(\cdot)$ and covariance function $\lambda(\cdot,\cdot)\colon I\times I \to \mathbb{R}^1$. Assume that $\xi \in \mathcal{X}$ and that $\lambda(\cdot,\cdot)$ is a positive-definite function satisfying

A5) $\iint \lambda(s,t)\mu(ds)\mu(dt) < \infty$.

Denote by Γ (respectively Λ) the integral operator on $L^2(I,d\mu)$ with kernel $\gamma(\cdot,\cdot)$ (resp. $\lambda(\cdot,\cdot)$), i.e.

(2.5)
$$\Gamma[f](s) = \int_I \gamma(s,t)f(t)\mu(dt)$$

$$\Lambda[f](s) = \int_I \lambda(s,t)f(t)\mu(dt).$$

Let $\mathcal{N}$ and $\mathcal{N}^\perp$ represent the null space of Γ and its orthogonal complement, $f^{\mathcal{N}}$ and $f^\perp$ the orthogonal projections of an element $f \in L^2(I,d\mu)$ onto $\mathcal{N}$ and $\mathcal{N}^\perp$, respectively. Since $(X-\theta)^{\mathcal{N}} = 0$ almost surely and since

$$L(\theta,a) = L(\theta^{\perp},a^{\perp}) + L(\theta^{\mathcal{N}},a^{\mathcal{N}})$$
$$\geq L(\theta^{\perp},a^{\perp})$$
$$= L(\theta,a^{\perp}+X^{\mathcal{N}}),$$

we can restrict our attention without loss of generality to estimators δ satisfying

(2.6) $$\delta[X] = (\delta[X]^{\perp}) + (X^{\mathcal{N}}).$$

We will in fact restrict attention to the smaller class of estimators satisfying (2.6) and also $\delta[X]^{\perp} = \delta[X^{\perp}]$, i.e., to the problem of estimating $\theta^{\perp}$ by observing $X^{\perp}$. This entails no serious loss of generality (once the prior mean $\xi(\cdot)$ and covariance $\lambda(\cdot,\cdot)$ are updated by the observation of $\theta^{\mathcal{N}} = X^{\mathcal{N}}$) and permits us to simplify notation by assuming that $\mathcal{N} = \{0\}$, i.e.

A6) Γ is positive definite.

It follows from A3) and A5) that Γ is positive definite and trace class, Λ is nonnegative definite and Hilbert Schmidt, and hence that $(\Gamma+\Lambda)$ is positive-definite and Hilbert-Schmidt; thus

(2.7) $$Q \equiv (\Gamma+\Lambda)^{-\frac{1}{2}}\Gamma^{2}(\Gamma+\Lambda)^{-\frac{1}{2}}$$

is positive-definite and trace class, with a complete orthonormal set of $p \leq \infty$ eigenfunctions $\{e_i\}_{0\leq i<p} \subset L^2(I,d\mu)$ with corresponding eigenvalues $q_0 \geq q_1 \geq \ldots > 0$ satisfying

(2.8) $$\operatorname{tr}(Q) = \sum_{i<p} q_i \leq \operatorname{tr}(\Gamma) = C.$$

Here $p \leq \infty$ is the dimension of the range of Q; in most interesting cases $p = \infty$. Define $B = \Gamma(\Gamma+\Lambda)^{-\frac{1}{2}}$ and set (for $0 \leq i < p$)

(2.9) $$e_i^* \equiv Be_i$$

$$X_i^* = \frac{1}{q_i} \int_I X(s)e_i^*(s)\mu(ds),$$

$$\theta_i^* = \frac{1}{q_i} \int_I \theta(s) e_i^*(s) \mu(ds),$$

$$\xi_i^* = \frac{1}{q_i} \int_I \xi(s) e_i^*(s) \mu(ds).$$

The random variables $\{X_j^*\}_{j<p}$ are a Gaussian family with means $EX_j^* = \theta_j^*$ and covariances

$$\begin{aligned} \sigma_{ij}^* &= E(X_i^*-\theta_i^*)(X_j^*-\theta_j^*) \\ &= \frac{1}{q_i q_j} \int_I \int_I e_i^*(s) \gamma(s,t) e_j^*(t) \mu(ds) \mu(dt) \\ &= \frac{1}{q_i q_j} \langle e_i, B^t \Gamma B e_j \rangle_\mu . \end{aligned}$$

Here $\langle f,g \rangle_\mu = \int_I f g d\mu$ is the inner-product in $L^2(I,d\mu)$ and B^t represents the adjoint of B with respect to $\langle \cdot,\cdot \rangle_\mu$. The $\{e_i^*\}$ are a complete orthogonal family since

$$\begin{aligned} \langle e_i^*, e_j^* \rangle_\mu &= \langle e_i, B^t B e_j \rangle_\mu \\ &= \langle e_i, Q e_j \rangle_\mu \\ &= q_j \langle e_i, e_j \rangle_\mu \\ &= q_i \text{ if } i=j, \text{ 0 else.} \end{aligned}$$

Thus any $f \in L^2(I,d\mu)$ may be expanded in an L^2-convergent series

(2.10a) $$f(\cdot) = \sum_{i<p} f_i e_i^*(\cdot) ,$$

where the $f_i \equiv \frac{1}{q_i} \langle f, e_i^* \rangle_\mu$ satisfy

(2.10b) $$\langle f,f \rangle_\mu = \sum_{i<p} q_i |f_i|^2 < \infty .$$

If $\theta(\cdot)$ were regarded as a sample path of a Gaussian process independent of $Z(\cdot)$, with mean ξ and covariance $\lambda(\cdot,\cdot)$, then the θ_i^* would themselves be Gaussian random variables with means ξ_i^* and covariances

$$\lambda_{ij}^* = \frac{1}{q_i q_j} \langle e_i, B^t \Lambda B e_j \rangle_\mu .$$

Nevertheless in the expectations in the sequel, θ will be regarded as constant.

The following estimator will be considered. Define

$$\delta^M[X](\cdot) = \sum_{i \geq 0} \delta_i^{*M}[X] e_i^*(\cdot), \tag{2.11}$$

where for $0 \leq i < p$,

(2.12) $\delta_i^{*M}[X]$

$$= X_i^* - \frac{1}{q_i} \sum_{j \geq i} (q_j - q_{j+1}) \min\{1, \frac{2(j-1)^+}{||\underset{\sim}{X}^* - \underset{\sim}{\xi}^*||_j^2}\} [\Sigma_{(j)}^{*-1}(\underset{\sim}{X}_{(j)}^* - \underset{\sim}{\xi}_{(j)}^*)]_i ,$$

$$||\underset{\sim}{X}^* - \underset{\sim}{\xi}^*||_j^2 = (\underset{\sim}{X}_{(j)}^* - \underset{\sim}{\xi}_{(j)}^*)^t \Sigma_{(j)}^{*-2} (\underset{\sim}{X}_{(j)}^* - \underset{\sim}{\xi}_{(j)}^*),$$

$$\underset{\sim}{X}_{(j)}^* = (X_0, X_1, \ldots, X_j)^t, \quad \underset{\sim}{\xi}_{(j)}^* = (\xi_0, \xi_1, \ldots, \xi_j)^t ,$$

and $\Sigma_{(j)}^*$ is the $(j+1)\times(j+1)$ matrix with entries $\sigma_{k\ell}^*$.

<u>Theorem.</u> δ^M is well defined and (if $p \geq 3$) $R(\theta, \delta^M) < R(\theta, \delta^0)$.

<u>Proof.</u> To show that δ^M is well defined, it is first necessary to prove that the summation in (2.12) converges. To see this, let

$$Z_{(j)} = \Sigma_{(j)}^{*-1}(\underset{\sim}{X}_{(j)}^* - \underset{\sim}{\xi}_{(j)}^*),$$

so that the sum in (2.12) can be written

$$\sum_{j=i}^{\infty} (q_j - q_{j+1}) \min\{1, \frac{2(j-1)^+}{|Z_{(j)}|^2}\} Z_{(j)i} . \tag{2.13}$$

Clearly each term in the series is bounded by

$$(q_j-q_{j+1})\min\left\{1, \frac{2(j-1)^+}{|Z_{(j)}|^2}\right\}|Z_{(j)}| \leq (q_j-q_{j+1}) \sqrt{2(j-1)^+} \ .$$

Also, summation by parts gives that

$$\sum_{j=i}^{\infty} (q_j-q_{j+1}) \sqrt{2(j-1)^+} = \sqrt{2} \left\{ q_i \sqrt{(i-2)^+} + \sum_{j=i}^{\infty} q_j\left[\sqrt{(j-1)^+} - \sqrt{(j-2)^+}\right]\right\}$$

$$\leq \sqrt{2} \left\{ q_i \sqrt{(i-2)^+} + \sum_{j=i}^{\infty} q_j[1] \right\} .$$

By (2.8), this sum is bounded by

$$\sqrt{2} \ (q_i \sqrt{(i-2)^+} + \sum_{j<p} q_j) < \sqrt{2} \ (q_i\sqrt{i} + C) < \infty .$$

and (2.12) converges uniformly.

To show that (2.11) converges in $L^2(I,d\mu)$ it is enough to show that $\sum_{i<p} q_i(\delta_i^{*M}[X]-\theta_i)^2 < \infty$; we do this and prove minimaxity using techniques originated in Bhattacharya [4]. First note that by Berger [1] the finite-dimensional estimators

$$(2.14) \qquad \delta^{(j)}[X^*_{(j)}] \equiv X^*_{(j)} - \min\left\{1, \frac{2(j-1)^+}{||X^*-\xi^*||^2_j}\right\} [\ddagger^{*-1}_{(j)}(X^*_{(j)}-\xi^*_{(j)})]$$

are (for sum of squares error loss) minimax estimators of the mean $\theta^*_{(j)} \equiv (\theta^*_0,\ldots,\theta^*_j)^t$ of a multivariate normal $X^*_{(j)} = (X^*_0,\ldots,X^*_j)^t$ with covariance matrix $\ddagger^*_{(j)}$. It follows that the random variable

$$\delta_i^{*M}[X] = \frac{1}{q_i} \sum_{j\geq i} (q_j-q_{j+1})\delta_i^{(j)}[X^*_{(j)}]$$

satisfies

$$E(\delta_i^{*M}-\theta^*_i)^2 = E[\frac{1}{q_i} \sum_{j\geq i} (q_j-q_{j+1})(\delta_i^{(j)}-\theta^*_i)]^2$$

$$\leq E\,\frac{1}{q_i} \sum_{j\geq i} (q_j-q_{j+1})[\delta_i^{(j)}-\theta^*_i]^2 ,$$

so

$$\sum_{i\geq 0} q_i E(\delta_i^{*M}-\theta_i^*)^2 \leq \sum_{0\leq i\leq j} (q_j-q_{j+1})E[\delta_i^{(j)}-\theta_i^*]^2$$

$$\leq \sum_{0\leq i\leq j} (q_j-q_{j+1})\sigma_{ii}^*$$

$$= \sum_{0\leq i} q_i\sigma_{ii}^*$$

$$= \sum_{i\geq 0} \frac{1}{q_i} \iint e_i^*(s)e_i^*(t)\gamma(s,t)\mu(ds)\mu(dt)$$

$$= \int_I \gamma(s,s)\mu(ds) = C.$$

Since $C < \infty$ and (by A1)) $\theta \in L^2(I,d\mu)$, Parseval's identity (2.10) guarantees that the sum (2.11) converges in $L^2(I\times\Omega;d\mu\times dP)$ to an estimator δ^M in $\mathcal{D}$ with risk

(2.15) $$R(\theta,\delta^M) = E\sum q_i(\delta_i^{*M}-\theta_i)^2 \leq C.$$

Since $R(\theta,\delta^0) \equiv C$ and δ^0 is minimax, δ^M must be minimax too. The inequality (2.15) is strict (by Berger [1]) if $p \geq 3$. ||

The estimator δ^M is the infinite dimensional analog of the estimator δ^{MB} in Berger [2]. Indeed the decomposition induced by Q in Section 2 corresponds to the linear transformation induced by Q* in Berger [2]. The reader is referred to Berger [2] and Berger and Wolpert [3] for extensive discussion of the motivation for this estimator.

III. ANALYSIS WHEN Γ AND Λ COMMUTE

In general, it is difficult to work with Q and to determine the $\{e_i^*\}$ and $\{q_i\}$. When Γ and Λ commute, however, in the sense that

$$\Gamma\Lambda f(\cdot) = \Lambda\Gamma f(\cdot)$$

for all $f \in \mathcal{L}^2(I;d\mu)$, then the problem simplifies considerably. This is because a complete set $\{e_i\}$ of eigenfunctions of Γ with eigenvalues $\{v_i\}$ can be found which are also eigenfunctions of Λ with eigenvalues, say, $\{\lambda_i\}$, and hence

$$Qe_i(\cdot) = \frac{v_i^2}{v_i+\lambda_i} e_i(\cdot) ,$$

so that we can choose

$$e_i^* = e_i \text{ and } q_i = \frac{v_i^2}{v_i+\lambda_i} . \tag{3.1}$$

The estimator δ^M reduces in this case to the estimator considered in Berger and Wolpert [3] (letting $\lambda_i = \lambda_{ii}$).

The only remaining problem is that of determining when Γ and Λ commute. (In terms of $\lambda(s,t)$ and $\gamma(s,t)$ this means

$$g(t,s) \equiv \int \gamma(s,v)\lambda(t,v)\mu(dv)$$

must equal $g(s,t)$, so that we will also say $\lambda(s,t)$ and $\gamma(s,t)$ commute.) Since the eigenfunctions of Γ are often easy to determine (see Berger and Wolpert [3]), it will often suffice to merely check that these eigenfunctions are (or can be chosen to be) eigenfunctions of Λ.

If the $\{e_i\}$ are eigenfunctions of $\lambda(s,t)$, then it follows from A5) that

$$\lambda(s,t) = \sum_{i\geq 0} \lambda_i e_i(s) e_i(t). \tag{3.2}$$

(Although this sum is in general only an $\mathcal{L}^2(I\times I;d\mu\times d\mu)$ sum, if the λ_i are summable and $\gamma(\cdot,\cdot)$ bounded then the convergence is uniform.) The class of all such $\lambda(s,t)$ (with $\lambda_i \geq 0$, of course) is thus the class of prior covariance functions for which the analysis is particularly simple.

Finally, we can address the question of determination of suitable $\lambda(s,t)$ from knowledge of $\lambda(t) = \lambda(t,t)$. Using (3.2), it is clear that a suitable (i.e., commuting) $\lambda(s,t)$ can be found providing

$$\lambda(t) = \sum_{i \geq 0} \lambda_i e_i^2(t), \tag{3.3}$$

i.e., providing $\lambda(\cdot)$ is in the positive cone spanned by the $\{e_i^2\}$. We conclude with the application of these ideas to the situation of Example 2 in Berger and Wolpert [3].

<u>Example</u>. Suppose $X(\cdot)$ is Brownian motion with mean $\theta(\cdot)$ and covariance function $\gamma(s,t) = \sigma^2 \min\{s,t\}$ ($\sigma^2 > 0$ known), $I = [0,T]$, and μ = Lebesgue measure. In Berger and Wolpert [3] (or Wong [5]) it is shown that the eigenfunctions and eigenvalues of Γ are, for $i \geq 0$,

$$e_i(s) = (2/T)^{\frac{1}{2}} \sin[(i+\tfrac{1}{2})\pi s/T], \tag{3.4}$$

$$v_i = [\sigma T/\pi(i+\tfrac{1}{2})]^2.$$

For these eigenfunctions, using (3.2) and the multiple angle identity, we obtain the class of commuting $\lambda(s,t)$ as being those of the form (with $\lambda_i \geq 0$)

$$\lambda(s,t) = \sum_{i=0}^{\infty} \lambda_i \frac{1}{T} \{\cos[(i+\tfrac{1}{2})\pi(s-t)/T] - \cos[(i+\tfrac{1}{2})\pi(s+t)/T]\} \tag{3.5}$$

$$= h\left(\frac{|s-t|}{2}\right) - h\left(\frac{s+t}{2}\right),$$

where

$$h(y) = \sum_{i=0}^{\infty} \lambda_i \frac{1}{T} \cos[(2i+1)\pi y/T], \tag{3.6}$$

for $0 \leq y \leq T$. Noting that (for $j \geq 0$, $i \geq 0$)

$$\int_0^T \cos[js\pi/T]\cos[(2i+1)s\pi/T]ds = \begin{cases} 0 & \text{for} \quad j \neq 2i+1 \\ \frac{T}{2} & \text{for} \quad j = 2i+1 , \end{cases}$$

we obtain (for $j \geq 0$)

$$(3.7) \qquad \int_0^T h(s)\cos[js\pi/T]ds = \begin{cases} 0 & \text{if } j \text{ is even} \\ \frac{1}{2}\lambda_i & \text{if} \quad j = 2i+1 . \end{cases}$$

Since $\{\cos[is\pi/T], i=0,1,\ldots\}$ is a complete orthogonal system in $\mathcal{L}^2(I;d\mu)$, the fact that all even Fourier coefficients are zero means that h must be an odd function about $\frac{T}{2}$, i.e.,

$$h(s) = h(T-s).$$

All odd functions can be represented as in (3.6), but the subclass for which the λ_i are nonnegative is, of course, smaller. Although this subclass is hard to describe in general, the following lemma describes an important special case.

Lemma. Suppose that

(i) $h(y)$ is continuous and nonincreasing;

(ii) $h(y)$ is convex on $[0, \frac{T}{2}]$; and

(iii) $h(y)$ is odd about $\frac{T}{2}$.

Then $h(y)$ is of the form (3.1) (and hence $\gamma(s,t)$ commutes with $\lambda(s,t)$), with

$$(3.8) \qquad \lambda_i = 2\int_0^T h(y)\cos[(2i+1)y\pi/T]dy \geq 0.$$

Proof. By (3.5), it is only necessary to show that (3.8) holds. This can be done analytically by dividing the integral up into regions of size $T/(4i+2)$, changing variables so all integrals are from 0 to $\frac{\pi}{2}$, using the periodicity of cosine to collect terms, and employing convexity and monotonicity of h to prove that the resulting integrand is positive. The details will be omitted. ||

The above observations also solve the problem of determining appropriate (i.e. commuting) $\lambda(s,t)$ from the variance function $\lambda(t)$. Indeed, (3.5) implies that

$$\lambda(t) = h(0) - h(t), \tag{3.9}$$

so, in particular, any function h satisfying the conditions of the Lemma will result in a suitable variance function via (3.9).

In Berger and Wolpert [3], the choice $h(t) = -\rho t$ $(\rho > 0)$ was considered, i.e., the variance function

$$\lambda(t) = \rho t$$

was investigated. This, however, corresponds to

$$\lambda(t,s) = h\left(\frac{|s-t|}{2} \right) - h\left(\frac{s+t}{2} \right) = \rho \min\{t,s\} ,$$

which is simply a multiple of $\gamma(s,t)$, and hence a rather trivial example of a commuting γ. Many other suitable variance (or covariance) functions can clearly be developed using the Lemma. For example, choosing

$$h(y) = \left(\frac{T}{2} - y\right)^3$$

(which clearly satisfies the conditions of the Lemma), results in

$$\gamma(t) = \left(\frac{T}{2} \right)^3 - \left(\frac{T}{2} - t\right)^3$$

and

$$\gamma(s,t) = \tfrac{1}{4} \min\{t,s\}[3(\max\{t,s\}-T)^2+\min\{t,s\}^2] .$$

(The above variance function (or a multiple of it) might be reasonable in a situation where the "expected error" in the prior guess $\xi(t)$ for $\theta(t)$ is more sharply increasing near the endpoints of [0,T] than near the middle.) An easy calculation yields

$$\lambda_i = \frac{6T^4}{(2i+1)^2\pi^2}\left[\tfrac{1}{2} - \frac{4}{(2i+1)^2\pi^2}\right],$$

which can be used with (3.4) and (3.1) to define δ^M. (In the commuting situation it is probably easier to use the expression in Berger and Wolpert [3] for δ^M than to use (2.11) and (2.12).)

REFERENCES

[1] Berger, J. (1976). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Multivariate Anal. 6, 256-264.

[2] Berger, J. (1982). Selecting a minimax estimator of a multivariate normal mean. Ann. Statist. 10, No. 1.

[3] Berger, J. and Wolpert, R. L. (1981). Estimating the mean function of a Gaussian process and the Stein effect. Mimeograph Series #81-17, Statistics Department, Purdue University.

[4] Bhattacharya, P. K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. Ann. Math. Statist. 37, 1819-1824.

[5] Wong, E. (1971). Stochastic Processes in Information and Dynamical Systems. McGraw-Hill, New York.