ESTIMATING THE MEAN FUNCTION OF A GAUSSIAN PROCESS AND THE STEIN EFFECT

by

James Berger*, Purdue University
and
Robert Wolpert**, Duke University and
University of North Carolina at Chapel Hill

Mimeograph Series #81-17

Statistics Department
Mathematical Sciences Division
Purdue University

June 1981

*Research supported by the Alfred P. Sloan Foundation and by the National Science Foundation under Grants # MCS-7802300A3 and MCS-8101670.

**Research supported by the National Science Foundation under Grant MCS-78-01737 and by the Air Force office of Scientific Research grant #AFOSR 80-0080.

ESTIMATING THE MEAN FUNCTION OF A GAUSSIAN PROCESS AND THE STEIN EFFECT

by

James Berger*, Purdue University

and

Robert Wolpert**, Duke University

*Research supported by the Alfred P. Sloan Foundation and by the National Science Foundation under Grants #MCS-7802300A3 and MCS-8101670.

**Research supported by the National Science Foundation under Grant MCS-78-01737.

Short Title: ESTIMATING THE MEAN FUNCTION

AMS 1970 subject classifications. Primary 62M99; Secondary 62C99, 62F10.

Key Words and Phrases. Estimation, mean function, Gaussian process, Stein effect, integrated quadratic loss, risk function, minimax, Karhunen-Loéve expansion, incorporation of prior information.

Send correspondence to: James Berger
Department of Statistics
Purdue University
W. Lafayette, IN 47907

SUMMARY

The problem of global estimation of the mean function $\theta(\cdot)$ of a quite arbitrary Gaussian process is considered. The loss function in estimating θ by a function $a(\cdot)$ is assumed to be of the form $L(\theta,a) = \int[\theta(t) - a(t)]^2\mu(dt)$, and estimators are evaluated in terms of their risk function (expected loss). The usual minimax estimator of θ is shown to be inadmissible via the Stein phenomenon; in estimating the function θ we are trying to simultaneously estimate a large number of normal means. Estimators improving upon the usual minimax estimator are constructed, including an estimator which allows the incorporation of prior information about θ. The analysis is carried out by using a version of the Karhunen-Loéve expansion to represent the original problem as the problem of estimating a countably infinite sequence of means from independent normal distributions.

1. Introduction

The problem of global estimation of the mean function of a quite arbitrary continuous Gaussian process will be considered. Before presenting the general setup we will outline the problem in an important special case, that of estimating the mean function of a Wiener process.

Suppose we observe, for $t \in [0,T]$,

$$X(t) = \theta(t) + Z(t),$$

where $\theta(t)$ (the signal) is an unknown function and $Z(t)$ (the noise) is zero mean Brownian motion. It is desired to estimate the function $\theta(\cdot)$ under the global loss

$$L(\theta,a) \equiv \int_0^T [\theta(t) - a(t)]^2 dt.$$

Of course, $\{Z(t), t \in [0,T]\}$ has an underlying probability structure, and so it is natural to evaluate an estimator $\delta[X](\cdot)$ (which for each X is a function on $[0,T]$) by its risk (or expected loss)

$$R(\theta,\delta) = E \int_0^T [\theta(t) - \delta[X](t)]^2 dt.$$

The usual estimator of $\theta(\cdot)$ is $\delta^0[X](\cdot) \equiv X(\cdot)$, and under appropriate conditions this is best invariant and minimax. Since we are trying to simultaneously estimate an uncountably infinite number of means (the $\theta(t)$ for all $t \in [0,T]$) one would suspect that the usual estimator can be improved upon via the Stein phenomenon.

Stein estimation (in its simplest and original setting - see Stein (1955) and James and Stein (1960)) deals with estimation of the mean $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^t$ of a p-variate normal random variable $\underset{\sim}{X} = (X_1, \ldots, X_p)^t$ (identity covariance matrix) under sum of squared errors loss

$$L(\underset{\sim}{\theta},\underset{\sim}{\delta}) = \sum_{i=1}^{p} (\theta_i - \delta_i)^2.$$

Here, also, the usual (best invariant and minimax) estimator of $\underset{\sim}{\theta}$ is $\underset{\sim}{\delta}^o(\underset{\sim}{X}) = \underset{\sim}{X}$. This estimator cannot be uniformly improved upon for p = 1 or 2, but for $p \geq 3$ the James-Stein estimator

$$\delta^{JS}(\underset{\sim}{X}) = (1 - \frac{p-2}{|\underset{\sim}{X}|^2})\underset{\sim}{X}$$

has risk

$$R(\underset{\sim}{\theta},\underset{\sim}{\delta}^{JS}) < R(\underset{\sim}{\theta},\underset{\sim}{\delta}^o)$$

for all $\underset{\sim}{\theta}$. (Here again $R(\underset{\sim}{\theta},\underset{\sim}{\delta}) = EL(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{X}))$.)

The stochastic process setting discussed earlier is the obvious infinite dimensional analogue of the finite dimensional situation, and so one would hope that improved estimators analogous to δ^{JS} could be found. This will be done in Section 3. Indeed obvious analogues of δ^{JS} such as

(1.1) $$\delta[X](\cdot) = (1 - \frac{a}{\int |X(t)|^2 \mu(dt)})X(\cdot)$$

will be considered (along with more sophisticated estimators) and shown, under certain conditions, to satisfy

$$R(\theta,\delta) < R(\theta,\delta^o)$$

for all $\theta(\cdot)$.

The analysis will be carried out in a fairly general framework. The "noise" $Z(\cdot)$ discussed above will be allowed to be a quite arbitrary path-continuous Gaussian process, defined on an index set $I \subset R^1$. (Thus the finite dimensional problem of estimating a normal mean will be included as a special case.) Also, the global loss will be allowed to have a "weighting" measure; i.e.,

$$L(\theta,a) = \int [\theta(t) - a(t)]^2 \mu(dt)$$

will be considered. The formal statement of the problem, with needed assumptions, follows.

Let $\mathcal{X}$ be the complete metric space of continuous real-valued functions on a closed set $I \subset R^1$ with the supremum norm. Let Θ be a subset of $\mathcal{X}$ and let Z be a Borel-measurable $\mathcal{X}$-valued Gaussian process on some probability space $(\Omega, \mathfrak{F}, P)$ with mean $EZ(t) = 0$ and known continuous covariance function $\gamma(s,t) = EZ(s)Z(t)$. Here (as usual) we suppress the ω-dependence of functions $Y \in L^1(\Omega, \mathfrak{F}, P)$ and write the integral of such a random variable indifferently as $\int_\Omega Y dP$ or EY.

Let A be a subset of the Borel-measurable real-valued functions on I. We consider the problem of estimating the unknown mean $\theta \in \Theta$ of the Gaussian process $X(t) = Z(t) + \theta(t)$ on the basis of a single observation of $X \in \mathcal{X}$ by an estimate $a \in A$, under the quadratic loss

$$(1.2) \qquad L(\theta,a) \equiv \int_I [\theta(s) - a(s)]^2 \mu(ds) \qquad \theta \in \Theta, \ a \in A .$$

Here μ is an arbitrary but specified nonnegative Borel measure on I satisfying

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

A2) $A \supset L^2(I, d\mu)$;

A3) $C \equiv \int_I \gamma(s,s)\mu(ds) < \infty$.

(Note that μ may be a singular measure.)

If several observations $X^{(1)}(\cdot), \ldots, X^{(n)}(\cdot)$ are available then their average $\bar{X}(\cdot) \equiv \frac{1}{n} \sum_{i=1}^{n} X^{(i)}(\cdot)$ satisfies our conditions and is sufficient for θ, so the restriction to a single observation is harmless.

From A1) and (1.2) it follows that $L(\theta,a) = \infty$ for any $a \notin L^2(I, d\mu)$; thus we may restrict our attention to estimates $a \in A \cap L^2(I, d\mu) = L^2(I, d\mu)$.

Let $\mathcal{D}$ denote the decision space of all Borel-measurable mappings $\delta\colon \mathcal{X} \to L^2(I,d\mu)$. As usual in decision theory we will evaluate an estimator $\delta \in \mathcal{D}$ by considering its risk function

(1.3) $$R(\theta,\delta) \equiv EL(\theta,\delta)$$

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

The usual estimator for the mean θ of a Gaussian process is

$$\delta^o[X] \equiv X,$$

with constant risk

(1.4) $$R(\theta,\delta^o) = E \int_I |\theta(s) - X(s)|^2 \mu(ds)$$

$$= \int_I E|Z(s)|^2 \mu(ds)$$

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

$$= C.$$

This is the best invariant estimator and will be shown to be minimax, but (except in some trivial cases) it is not admissible.

Indeed we will derive estimators $\delta^* \in \mathcal{D}$ which satisfy

(1.5) $$R(\theta,\delta^*) < R(\theta,\delta^o)$$

for every $\theta \in \Theta$.

Rather than working directly in $L^2(I,d\mu)$, we will employ a generalization of the Karhunen-Loéve expansion of a stochastic process to transform the problem into that of estimating a countably infinite sequence of normal means, $\theta_0,\theta_1,\theta_2,\ldots$, based on independent normal observations. This will enable some of the theory of Stein estimation to be brought more directly to bear on the problem. This transformation of the problem (carried out in Section 2) should be useful in other statistical analyses.

Section 3 contains the analysis of the transformed countably infinite Stein estimation problem. Although some relatively simple estimators improving upon δ^0 will be presented, it is observed, as in Berger (1980a) and Berger (1982) that in intelligently selecting among the many possible improved estimators, it is necessary to incorporate prior information concerning the θ_i. An estimator allowing incorporation of prior means ξ_i and variances τ_{ii} for the θ_i, and yet having risk better than δ^0, is developed. This estimator can also be viewed as a robust Bayesian estimator of the θ_i. It is also indicated how the ξ_i and τ_{ii} can be obtained from prior information concerning $\theta(\cdot)$.

Section 4 discusses several examples (introduced in Section 2); namely the finite dimensional situation, the original Brownian motion example and an example concerning the Brownian bridge with a weighted global loss. Section 5 presents some concluding remarks.

2. The Transformed Problem

Denote the norm of an element $u \in L^2(I,d\mu)$ by $||u||_\mu \equiv (\int_I |u(s)|^2\mu(ds))^{\frac{1}{2}}$. The Schwartz inequality guarantees that

$$(2.1)\qquad |\gamma(s,t)|^2 = |EZ(s)Z(t)|^2 \le \gamma(s,s)\gamma(t,t)$$

for every $s,t \in I$, and hence that for each fixed $s \in I$ the continuous function $\gamma(s,\cdot)$ satisfies

$$||\gamma(s,\cdot)||_\mu^2 \le \gamma(s,s)C.$$

It follows that, for each $f \in L^2(I,d\mu)$, the function

$$\Gamma f(s) \equiv \int_I \gamma(s,t)f(t)\mu(dt)$$

is bounded by

(2.2) $$|\Gamma f(s)| \leq ||f||_\mu \sqrt{C\gamma(s,s)}$$

and so has $L^2(I,d\mu)$ norm

$$||\Gamma f||_\mu \leq C||f||_\mu \; .$$

The function Γf is also continuous, since

(2.3) $$|\Gamma f(s) - \Gamma f(t)| \leq ||f||_\mu ||\gamma(s,\cdot) - \gamma(t,\cdot)||_\mu$$

and (by A3), (2.1), and Lebesgue's dominated convergence theorem) the mapping $s \to \gamma(s,\cdot)$ is continuous from I to $L^2(I,d\mu)$. In fact, (2.2) and (2.3) show that $\{\Gamma f: ||f||_\mu \leq 1\}$ is uniformly bounded and uniformly equicontinuous on compact sets, so Γ is a compact operator from $L^2(I,d\mu)$ to $\mathcal{X}$. It is also nonnegative definite and Hilbert-Schmidt as an operator on $L^2(I,d\mu)$ since $\gamma(\cdot,\cdot)$ is nonnegative definite and satisfies (by 2.1)

$$\iint |\gamma(s,t)|^2 \mu(ds)\mu(dt) \leq C^2 < \infty \; .$$

It follows that there exists an orthonormal family $\{e_0,e_1,\ldots\} = \{e_i\}_{i<p}$ of $p \leq \infty$ continuous functions $e_i \in \mathcal{X} \cap L^2(I,d\mu)$ and p numbers $v_0 \geq v_1 \geq v_2 \geq \ldots > 0$ satisfying

P0) $\int_I e_i(s)e_j(s)\mu(ds) = 1$ if $i = j$, 0 else;

P1) $\Gamma[e_i](s) \equiv \int_I \gamma(s,t)e_i(t)\mu(dt)$

$= v_i e_i(s)$ for every $s \in I$;

P2) $\Gamma[f](s) = 0$ for every $s \in I$ if and only if $\int_I f(s)e_i(s)\mu(ds) = 0$ for each $0 \leq i < p$;

P3) $\sum_{0\leq i<p} v_i = \int_I \gamma(s,s)\mu(ds) = C$;

P4) $\sum_{0\leq i<p} v_i^2 = \int_I \int_I |\gamma(s,t)|^2 \mu(ds)\mu(dt) \leq C^2$.

Here $p \leq \infty$ is the dimension of the range of Γ in $L^2(I,d\mu)$. If an assertion A depending on $\omega \in \Omega$ is true for all ω outside of a set $N \in \mathfrak{F}$ with $P(N) = 0$, say A holds "a.s.[P]"; similarly an assertion B depending on $t \in I$ holds "a.s.[μ]" if it is true for all t outside a Borel set N' with $\mu(N') = 0$.

The Hilbert-Schmidt property and nonnegative-definiteness of Γ guarantee the existence of a family $\{e_i\} \subset L^2(I,d\mu)$ satisfying P0), P2), and (a.s.[μ])P1 . It follows from P0) and 2.3) that we may redefine $\{e_i\}$ (if necessary) to ensure that $\{e_i\} \subset \mathcal{X}$ and that P1) holds for <u>every</u> $s \in I$. It follows from A1) and A3) that

(2.4) $$E||X(\cdot)||_\mu^2 = C + ||\theta||_\mu^2 < \infty$$

and hence that $X \in L^2(I,d\mu)$ a.s.[P]. Thus we can define for $0 \leq i < p$

(2.5) $$X_i \equiv \int_I e_i(s)X(s)\mu(ds)$$

$$\theta_i \equiv \int_I e_i(s)\theta(s)\mu(ds).$$

<u>Lemma 1.</u> The random variables $\{X_i\}_{i<p}$ are independent and normally distributed with means $\{\theta_i\}_{i<p}$ and variances $\{v_i\}_{i<p}$.

<u>Proof</u>. Apply Fubini's theorem and P0), P1) to (2.5). ||

In general we cannot be sure that $\{e_i\}$ are complete, for there may be $f \in L^2(I,d\mu)$ satisfying $\Gamma f(s) = 0$ for all $s \in I$ and yet $||f||_\mu > 0$. In that case f would be orthogonal to each e_i by P2).

Denote by ℓ^2 the space of square-summable sequences $\underset{\sim}{u} = \{u_0, u_1, \ldots\}$ of real numbers with finite norm

$$|\underset{\sim}{u}| \equiv \left(\sum_{0 \leq i < p} u_i^2 \right)^{\frac{1}{2}} .$$

The mapping $u(\cdot) \to \underset{\sim}{u}$ from $L^2(I,d\mu)$ to ℓ^2 determined by $\underset{\sim}{u} = \{u_i\}_{i<p}$, $u_i \equiv \int_I e_i(s)\, u(s)\mu(ds)$, is a contraction (Bessel's inequality) and an isometry on the span L^{2*} of $\{e_i\}_{i<p}$ in $L^2(I,d\mu)$. Denote by u^* the orthogonal projection of $u \in L^2(I,d\mu)$ onto L^{2*}, and by $u^{\#}$ the coprojection $u - u^*$. Then

$$\int_I |u(s)|^2 \mu(ds) = \int_I |u^{\#}(s)|^2 \mu(ds) + \sum_{0 \le i < p} |u_i|^2 .$$

In many cases of interest Γ is positive definite (i.e. $\Gamma[f] = 0$ entails $||f||_\mu = 0$) and hence $L^{2*} = L^2(I,d\mu)$ and

$$||u||_\mu = |\underset{\sim}{u}| \quad . \tag{2.6}$$

In any case we have (2.6) for $u \in L^{2*}$. The paths of $Z(\cdot)$ lie in L^{2*} a.s. [P], so

$$E|\underset{\sim}{X}|^2 = E||\theta^* + Z||_\mu^2 = C + ||\theta^*||_\mu^2 < \infty$$

and $||X^{\#} - \theta^{\#}||_\mu = 0$. Also, for any $a \in A \cap L^2(I,d\mu)$ we have

$$\begin{aligned} L(\theta,a) &= ||\theta - a||_\mu^2 \\ &= ||\theta^{\#} - a^{\#}||_\mu^2 + ||\theta^* - a^*||^2 \\ &= ||X^{\#} - a^{\#}||_\mu^2 + \sum_{0 \le i < p} |\theta_i - a_i|^2 . \end{aligned} \tag{2.7}$$

Let $\delta \in \mathcal{D}$ be any measurable mapping from $\mathcal{X}$ to $L^2(I,d\mu)$, and set

$$\delta^+[X] = (\delta[X])^* + X^{\#} . \tag{2.8}$$

Then $\delta^+ \in \mathcal{D}$ and

$$\begin{aligned} L(\theta,\delta^+) &= ||\theta^{\#} - X^{\#}||_\mu^2 + ||\theta^* - (\delta[X])^*||_\mu^2 \\ &= ||(\theta - \delta[X])^*||_\mu^2 \end{aligned}$$

$$\leq ||\theta - \delta[X]||_{\mu}^{2}$$

$$= L(\theta,\delta)$$

for every θ, so we may restrict our attention to estimators

$$\delta \in \mathcal{D}^{+} = \{\delta \in \mathcal{D}:\ \delta = \delta^{+}\} = \{\delta:\ (\delta[X])^{\#} = X^{\#}\}.$$

For such an estimator, set $\underset{\sim}{\delta}[X] = \{\delta_i[X]\}$, and

$$\delta_i[X] = \int_I e_i(s)\delta[X](s)\mu(ds).$$

Then

$$L(\theta,\delta[X]) = \sum_{0\leq i<p} |\theta_i - \delta_i[X]|^2.$$

We conclude this section with the introduction of the three examples we will study in detail in Section 4.

Example 1. Multivariate Normal

Let $I = \{t_1,\ldots,t_m\}$ be any finite set in R^1, and let $\mathcal{X}$, Θ, A, μ, γ, and X satisfy A1) - A4). Each function u on I may be uniquely identified with the vector $\tilde{u} \in R^m$ with coordinates $\tilde{u}_i \equiv u(t_i)$; under this identification $\mathcal{X}$, A, and ℓ^2 are each identified with R^m, Θ with a subset of R^m, and X with an m-variate normal random vector $\tilde{X}$ with unknown mean $\tilde{\theta}$ and known covariance matrix $\ddagger_{ij} \equiv \gamma(t_i,t_j)$.

Let Q be the $m\times m$ diagonal matrix with entries $Q_{ii} \equiv \mu(\{t_i\})$; $p \leq m$ is the rank of $\ddagger Q$. There exists a $p\times p$ diagonal matrix D with $D_{11} \geq \cdots \geq D_{pp} > 0$ and an $m\times p$ matrix $\tilde{E}$ with $\tilde{E}^t Q\tilde{E} = I_p$, $\ddagger Q\tilde{E} = \tilde{E}D$; let $e_i \in \mathcal{X}$ be determined by $e_i(t_j) = \tilde{E}_{j,i+1}$. Then $\int e_i(s)e_j(s)\mu(ds) = (\tilde{E}^t Q E)_{i+1,j+1} = 1$ if $i = j$, 0 else and $\Gamma[e_i](t_j) = (\ddagger Q\tilde{E})_{j,i+1} = (\tilde{E}D)_{j,i+1} = v_i e_i(t_j)$. In this case $C = \Sigma v_i = \mathrm{tr}\,\tilde{E}^t Q\ddagger Q\tilde{E} = \mathrm{tr}\,\ddagger Q$.

With this choice of $\{e_i\}$, $\underset{\sim}{X} = \tilde{E}^t Q\tilde{X} \in R^p$ has expectation $\underset{\sim}{\theta} = \tilde{E}^t Q\tilde{\theta}$ and $p\times p$ covariance $E(\underset{\sim}{X} - \underset{\sim}{\theta})(\underset{\sim}{X} - \underset{\sim}{\theta})^t = \tilde{E}^t Q \ddagger Q\tilde{E} = D$.

Indeed <u>any</u> m-variate normal random variable $\tilde{X}$ with covariance $\ddagger$ determines a continuous path Gaussian process on (e.g.) $I = \{1,2,\ldots,n\}$ as above. The transformation described above changes the problem of estimating the mean $\tilde{\theta} \in R^m$ of the m-variate normal vector $\tilde{X}$ with covariance $\ddagger$ under weighted loss $\sum_{i=1}^{m} |\tilde{\theta}_i - \tilde{a}_i|^2 \mu(\{i\})$ into that of estimating the mean vector $\underset{\sim}{\theta} \in R^p$ of the p independent normal random variables $\{X_i\}$ with variances $\{v_i\}$ under squared-error loss $\sum_{i=1}^{p} |\theta_i - a_i|^2$.

<u>Example 2.</u> Brownian Motion

Let $I = [0,T]$ and let μ be Lebesgue measure on I. Let $X(t) = \theta(t) + Z(t)$ where $Z(\cdot)$ is a Wiener process with mean $EZ(t) = 0$ and covariance function $\gamma(s,t) \equiv EZ(s)Z(t) = \sigma^2 \min(s,t)$ for some constant $\sigma^2 > 0$. Then the eigenvalue equation

$$\Gamma[f] = vf \tag{2.9}$$

has no solution $f \neq 0$ when $v = 0$, while (2.9) with $v \neq 0$ implies that f is twice differentiable and satisfies

$$vf'' + \sigma^2 f = 0 \tag{2.10a}$$

with boundary conditions

$$f(0) = f'(T) = 0. \tag{2.10b}$$

The well-known complete set of normalized solutions to (2.10) are

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

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

with

$$C = \sum_{0 \le i < \infty} v_i = \sigma^2 T^2/2.$$

See Wong (1971) for a detailed account.

Example 3. Brownian Bridge with Weighted Loss

Suppose $I = [0,T]$, $\mu(dt) = t^{-1}(1 - t/T)^{-1}dt$, and $X(t) = \theta(t) + Z(t)$, where $\{Z(t), t \in I\}$ is the Brownian bridge, i.e. the Gaussian process with mean zero and covariance function

$$\gamma(s,t) = \sigma^2[\min\{s,t\} - st/T].$$

As in Example 2, the e_i and v_i satisfy a differential equation

(2.11a) $$v f''(s) = \frac{-\sigma^2}{s(1-s/T)} f(s)$$

(2.11b) $$f(0) = f(T) = 0.$$

Defining $h(s) = f(s)/\{s(1 - s/T)\}$, (2.11a) becomes (after dividing through by v)

$$s(1 - \frac{s}{T})h''(s) + 2(1 - \frac{2s}{T})h'(s) + (\frac{\sigma^2}{v} - \frac{2}{T})h(s) = 0.$$

The solutions to this equation are the Jacobi polynomials (on [0,T])

$$p_i(s) = \sum_{k=0}^{i} \binom{i+1}{k}\binom{i+1}{i-k}(\frac{s}{T} - 1)^{i-k}(\frac{s}{T})^k, \quad 0 \le i < \infty,$$

providing $v = \sigma^2 T/[(i+1)(i+2)]$. Multiplying by $s(1 - \frac{s}{T})$ and normalizing, gives a complete set of orthonormal solutions to the original differential equation, namely

$$e_i(s) = -\left\{\frac{(2i+3)(i+2)}{(i+1)}\right\}^{\frac{1}{2}} \sum_{k=0}^{i} \binom{i+1}{k}\binom{i+1}{i-k}\left(\frac{s}{T}-1\right)^{i-k+1}\left(\frac{s}{T}\right)^{k+1},$$

with corresponding eigenvalues

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

Here

$$C = \sum_{0\le i<\infty} v_i = \sigma^2 T \sum_{0\le i<\infty} \frac{1}{(i+1)(i+2)} = \sigma^2 T.$$

Comment. Whenever $Z(\cdot)$ is a Markov process (as in the above examples), $\gamma(\cdot,\cdot)$ will be a Greens function for some differential operator L with specified boundary conditions, and the $\{e_i\}$ will satisfy $v_i L e_i(\cdot) = e_i(\cdot)$ for $v_i \neq 0$ as well as the boundary conditions.

3. Minimax and Stein Estimation

We begin by showing that $\delta^0[X] \equiv X$ (respectively $\delta^0[\underset{\sim}{X}] \equiv \underset{\sim}{X}$) is a minimax estimator of $\theta \in \Theta$ (respectively $\underset{\sim}{\theta} \in \underset{\sim}{\Theta}$), i.e.

$$\sup_{\theta\in\Theta} R(\theta,\delta^0) = C = \inf_{\delta\in\mathcal{D}} \sup_{\theta\in\Theta} R(\theta,\delta).$$

Theorem 1. Let X be a Gaussian process in $\mathcal{X}$ (as in Section 1) with continuous covariance $\gamma(\cdot,\cdot)$ and loss measure μ satisfying A1) - A3). Then if Θ contains finite linear combinations of $\{e_i\}$, δ^0 is a minimax estimator of $\theta \in \Theta$ and $\underset{\sim}{\delta}^0$ is a minimax estimator of $\underset{\sim}{\theta} \in \underset{\sim}{\Theta}$.

Proof. Our original proof, based on a limiting Bayesian argument, was more complicated than the following simple proof suggested by Larry Brown.

Suppose that δ^0 were not minimax. Then for some $\varepsilon > 0$ and $\delta' \in \mathcal{D}$,

$$R(\theta,\delta') \le C - \varepsilon \quad \text{for every } \theta \in \Theta .$$

By P4) there is a finite $N \leq p$ satisfying

$$\sum_{i \geq N} v_i < \varepsilon/2$$

$$\sum_{i<N} v_i > C - \varepsilon/2.$$

The proof proceeds by constructing an estimator $\tilde{\delta}'$ of the N-variate normal mean $\tilde{\theta} = (\theta_0, \theta_1, \ldots, \theta_{N-1})$ with squared-error risk $R(\tilde{\theta}, \tilde{\delta}') \leq R(\tilde{\theta}, \tilde{\delta}^o) - \varepsilon/2$, contradicting the known minimaxity of the estimator $\tilde{\delta}^o[\tilde{X}] = \tilde{X}$ $= (X_0, X_1, \ldots, X_{N-1})$ in the N-variate case.

Let $\varphi : R^N \to \mathcal{X}$ be the mapping $\varphi[\tilde{u}](\cdot) \equiv \sum_{i<N} u_i e_i(\cdot)$, where $\tilde{u} = (u_0, \ldots, u_{N-1}) \in R^N$, and denote by $\psi : \mathcal{X} \cap L^2(I, d\mu) \to R^N$ its left inverse given by $\psi[u] = \tilde{u}$, where $u_i = \int_I u(t) e_i(t) \mu(dt)$. Then when $\theta = \varphi(\tilde{\theta})$, $\tilde{\theta} \in R^N$, the random variable $\tilde{X} \equiv \psi(X)$ has an N-variate normal distribution with mean $\tilde{\theta}$ and diagonal covariance $\tilde{\Sigma}$ with $\tilde{\Sigma}_{ii} = v_i$, $0 \leq i < N$. The invariant estimator $\tilde{\delta}^o[\tilde{x}] = \tilde{x}$ has constant risk $R(\tilde{\theta}, \tilde{\delta}^o) = \sum_{i<N} v_i > C - \varepsilon/2$ while the estimator

$$\tilde{\delta}'[x] = E[\psi(\delta'[X]) | \psi(X) = \tilde{x}]$$

has risk

$$\begin{aligned} R(\tilde{\theta}, \tilde{\delta}') &\leq \sum_{i<N} E|\theta_i - \delta_i'[X]|^2 \\ &\leq \sum_{i<p} E|\theta_i - \delta_i'[X]|^2 \\ &= R(\theta, \delta') \\ &\leq C - \varepsilon \\ &< R(\tilde{\theta}, \tilde{\delta}^o) - \varepsilon/2. \quad | \end{aligned}$$

Note that the sufficiency of $\tilde{X}$ for $\tilde{\theta}$ (see Comment 4 in Section 5) is needed in ensuring that $\tilde{\delta}'$ is a valid estimator of $\tilde{\theta}$. ||

We now turn to the problem of improving upon $\tilde{\delta}^o$ in case $p > 2$. The following theorem provides a starting point.

Theorem 2. For uniformly bounded constants $a_i \geq 0$, $b_i \geq 0$, and $d \geq 0$, $0 \leq i < p$, define an estimator $\delta^*[X]$ by

(3.1) $$\delta_i^*[X] = \left(1 - \frac{a_i}{d + \sum_{j<p} b_j X_j^2}\right) X_i \, .$$

Then if $p > 2$, $R(\theta,\delta^*) \leq R(\theta,\delta^0)$ for all θ if and only if

(3.2) $$2 \sum_{i<p} a_i v_i \geq \sup_{j<p} \{4a_j v_j + a_j^2/b_j\}$$

(where $a_j^2/b_j \equiv 0$ if $a_j = b_j = 0$).

Proof. Suppose (3.2) holds and set $||u||^2 = d + \Sigma b_i u_i^2$ for $u \in \ell^2$; then define

$$\begin{aligned}\Delta(\theta) &= R(\theta,\delta^0) - R(\theta,\delta^*)\\ &= \sum_{i<p} \left\{ E\ (\theta_i - X_i)^2 - \left(\theta_i - X_i + \frac{a_i X_i}{||X||^2}\right)^2 \right\}\\ &= \sum_{i<p} E \left\{ 2a_i \frac{(X_i - \theta_i) X_i}{||X||^2} - \frac{a_i^2 X_i^2}{||X||^4} \right\} \, .\end{aligned}$$

Integrating the first term by parts (with respect to X_i) yields

$$2a_i E \frac{(X_i - \theta_i) X_i}{||X||^2} = E \left\{ \frac{2a_i v_i}{||X||^2} - \frac{4a_i b_i v_i X_i^2}{||X||^4} \right\}$$

so

(3.3) $$\begin{aligned}\Delta(\theta) &= \sum_{i<p} E \frac{1}{||X||^2} \left\{ 2a_i v_i - \frac{X_i^2(4a_i b_i v_i + a_i^2)}{||X||^2} \right\}\\ &\geq E \frac{1}{||X||^2} \left\{ \sum_{i<p} 2a_i v_i - \sup_{j<p}(4a_j v_j + a_j^2/b_j) \right\}\end{aligned}$$

since $\sum_{i<p} u_i b_i X_i^2 \leq ||\underset{\sim}{X}||^2 \sup_{i<p} |u_i|$ for any bounded sequence $\{u_i\}$. It follows that $\Delta(\underset{\sim}{\theta}) \geq 0$ when (3.2) is satisfied. This proves the "if" half of the theorem.

Now for the "only if" half. Suppose (3.2) fails, so there exists $0 \leq j < p$ for which $a_j^2/b_j + 4a_j v_j > 2 \sum_{i<p} a_i v_i$. Consider $\underset{\sim}{\theta} = \{\theta_i\}_{i<p}$ where $\theta_i = 0$ for $i \neq j$ and $\theta_j = x \in R^1$; then it is straightforward to show from the first equality in (3.3) that $\Delta(\underset{\sim}{\theta}) < 0$ for large enough x. For a detailed similar argument, see Berger (1976). ||

<u>Comment.</u> Theorem 2 is an extension of Theorem 1 of Berger (1976) and could be generalized as in that paper.

<u>Corollary 1.</u> If $\{a_i\}$ and $\{b_i\}$ satisfy (3.2) and $d \geq 0$, the "positive part" estimator $\underset{\sim}{\delta}^+$ determined by

$$\delta_i^+[\underset{\sim}{X}] = \left(1 - \frac{a_i}{d + \sum_{j<p} b_j X_j^2}\right)^+ X_i$$

satisfies

$$R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^+) \leq R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^*) \leq R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^0)$$

for every $\underset{\sim}{\theta}$.

<u>Proof.</u> Just as in the $p < \infty$ case described in Berger and Bock (1976). ||

<u>Corollary 2.</u> Suppose $v_j < \sum_{i>j} v_i$ for some j. Choose any $0 < a \leq 2(-v_j + \sum_{i>j} v_i)$. Then the estimator $\underset{\sim}{\delta}^*$ determined by

$$\delta_i^*[\underset{\sim}{X}] = \begin{cases} X_i & i < j \\ \left(1 - \frac{a}{\sum_{k \geq j} X_k^2}\right) X_i & i \geq j \end{cases} \tag{3.4}$$

and its positive-part version each have smaller risk than $\underset{\sim}{\delta}^0$.

Proof. Set $d = 0$, $a_i = 0$ for $i < j$, $a_i = a$ for $i \geq j$, and $b_i = 0$ for $i < j$, $b_i = 1$ for $i \geq j$ in Theorem 2. ||

In particular, when $c = -v_0 + \sum_{i>0} v_i > 0$ then $\underset{\sim}{\delta}^*[\underset{\sim}{X}] = \left(1 - \frac{a}{|\underset{\sim}{X}|^2}\right)\underset{\sim}{X}$ is minimax for any $0 \leq a \leq 2c$. This is the analogue (1.1) of the James-Stein estimator; indeed the case $p < \infty$, $v_0 = \ldots = v_{p-1} = 1$, and $a = c$ is the James-Stein estimator exactly.

In examples 2 and 3 it is not the case that $v_0 < \sum_{i>0} v_i$, so that (3.4) with $j = 0$ does not outperform $\underset{\sim}{\delta}^0$. In each case $v_1 < \sum_{i>1} v_i$, however, so the $j = 1$ case improves upon $\underset{\sim}{\delta}^0$.

As discussed in Berger (1980a) and Berger (1982), such simple estimators as those above are rarely optimal. Since $\underset{\sim}{\delta}^0$ is minimax, any better estimator can only have significantly smaller risk in a fairly small region of $\underset{\sim}{\Theta}$. It is thus important to specify the region in which significant improvement is desired, and choose an improved estimator tailored to this region. The problem is best phrased in Bayesian terms, since the region in which significant improvement is desired will be the region in which $\underset{\sim}{\theta}$ is apriori thought likely to lie.

We will assume a very simple type of prior input; namely the specification of prior means ξ_i and variances τ_{ii} for the θ_i. In some situations it may happen that the θ_i are meaningful quantities concerning which prior information is available. For example, the $\{e_i\}$ might be the possible frequencies of a given signal and the θ_i their amplitudes. Often, however, it will be the case that $\theta(\cdot)$ itself is the only real quantity for which one has prior information. The problem then becomes that of transforming prior information about $\theta(\cdot)$ into suitable means ξ_i and variances τ_{ii} for the θ_i.

An obvious avenue to follow is to model the prior information about $\theta(\cdot)$ by pretending that $\{\theta(t), t \in I\}$ is itself a stochastic process with mean function $\xi(t)$ and covariance function $\tau(s,t)$. Apriori determination of $\xi(t)$ is straightforward, since $\xi(t)$ can just be considered to be the "best guess" for $\theta(t)$. Determination of $\tau(s,t)$ is harder, though intuitively it should just reflect the covariance of the "error" in the guess $\xi(t)$. In this light it is reasonable to expect fairly accurate specification of $\rho(t) = \tau(t,t)$, since this could just be chosen to be the square of the expected error in the guess $\xi(t)$. Specifying covariances is harder, but will usually not greatly affect the values τ_{ii}. In later examples we will just choose convenient covariances.

Since ξ_i is supposed to be the mean of θ_i and $\xi(t)$ the mean of $\theta(t)$, we must have

$$\xi_i = E\theta_i = E \int_I \theta(t)e_i(t)\mu(dt) = \int_I \xi(t)e_i(t)\mu(dt).$$

Likewise,

$$\tau_{ij} = E[(\theta_i - \xi_i)(\theta_j - \xi_j)] = \int\int_{I\times I} \tau(s,t)e_i(s)e_j(t)\mu(dt)\mu(ds).$$

Unfortunately, working with all the τ_{ij} is much more complicated than working just with the τ_{ii}, so we will ignore the prior covariances. For a method of incorporating these covariances into the analysis, see Wolpert and Berger (1982). (The method is based on a considerably more involved Karhunen-Loéve type expansion, although some special situations in which the expansion considered here will suffice are given.)

The estimator that is recommended, for given ξ_i and τ_{ii}, is the generalization to infinite dimensions of the estimator given in Berger (1982), and is given as follows: define $q_i = v_i^2/(v_i + \tau_{ii})$, and relabel (if necessary) so that $q_1 \geq q_2 \geq \ldots$; and let

(3.5) $$\delta_i^M(\underset{\sim}{X}) = X_i - \frac{v_i}{(v_i+\tau_{ii})}(X_i - \xi_i) r_i(\underset{\sim}{X}),$$

where

$$r_i(\underset{\sim}{X}) = \frac{1}{q_i} \sum_{j=i}^{\infty} (q_j - q_{j+1}) \min \left\{ 1, \frac{2(j-1)^+}{\sum_{\ell=0}^{j} (X_\ell - \xi_\ell)^2/(v_\ell+\tau_{\ell\ell})} \right\}.$$

Theorem 3. The estimator $\underset{\sim}{\delta}^M(\underset{\sim}{X}) = (\delta_0^M(\underset{\sim}{X}), \delta_1^M(\underset{\sim}{X}), \ldots)$ has smaller risk than $\underset{\sim}{\delta}^o$.

Proof. The proof is exactly the proof of minimaxity of the finite dimensional version of the above estimator given in Berger (1982). ||

The motivation for $\underset{\sim}{\delta}^M$ is interesting. Note that if the prior information is correct, $E^*[X_\ell] = \xi_\ell$ and

$$E^*[X_\ell - \xi_\ell]^2 = E^*[(X_\ell - \theta_\ell) + (\theta_\ell - \xi_\ell)]^2 = v_\ell + \tau_{\ell\ell},$$

where E* denotes expectation over both X and θ(and the prior and sample information are assumed to be independent). Hence for $j \geq 2$

$$E^*\left[\frac{j-1}{\sum_{\ell=0}^{j} (X_\ell - \xi_\ell)^2/(v_\ell+\tau_{\ell\ell})}\right] \cong 1$$

and so for $i \geq 2$ we "expect" that

$$r_i(\underset{\sim}{X}) = \frac{1}{q_i} \sum_{j=i}^{\infty} (q_j - q_{j+1})(1) = 1.$$

Thus (at least with substantial probability)

$$\delta_i^M(X) \cong X_i - \frac{v_i}{v_i+\tau_{ii}}(X_i - \xi_i),$$

which is the best linear estimate of θ_i for the given prior information. The actual form of $r_i(\underset{\sim}{X})$ protects against prior misspecification by going to zero if X seems too far from ξ. Indeed, by Theorem 2, δ^M is better than δ^0 always (but will only be significantly better if θ turns out to be near $\xi = (\xi_0, \xi_1, \ldots)$ as expected). For smaller i (particularly i = 0 and i = 1), r_i can be substantially smaller than 1 even for X near ξ, but this is the price that must be paid to ensure dominance over δ^0. (The "best" linear estimator for the given prior information has a risk function which increases quadratically in the $(\theta_i - \xi_i)$, and can hence be infinitely worse than δ^0.) In the next section, the above results are applied to Examples 1, 2, and 3 (introduced in Section 2.)

Since δ^M involves an infinite sum, it can, of course, never be calculated exactly. The somewhat delicate point thus arises as to whether any approximation used for δ^M dominates δ^0. The natural approximation to use would result in the following estimator for $\theta(\cdot)$:

$$\delta^m[X](\cdot) = \{X(\cdot) - \sum_{i=0}^{m} X_i e_i(\cdot)\} + \sum_{i=0}^{m} \delta_i^m(X) e_i(\cdot), \tag{3.6}$$

where

$$\delta_i^m(X) = X_i - \frac{v_i}{v_i + \tau_{ii}} (X_i - \xi_i) r_i^m(X),$$

$$r_i^m(X) = \frac{1}{q_i} \sum_{j=i}^{m} (q_j - q_{j+1}) \min\left\{1, \frac{2(j-1)^+}{\sum_{\ell=0}^{j} (X_\ell - \xi_\ell)^2/(v_\ell + \tau_{\ell\ell})}\right\},$$

and q_{m+1} is here defined to be zero. This estimator avoids the infinite sum, and furthermore (and realistically) necessitates only the calculation of the relevant quantities through the $m^{\underline{th}}$ coefficient. Here m could be chosen as the largest number for which the Fourier analysis is calculationally feasible. It can be shown by arguments virtually identical to the preceeding ones that δ^m is minimax (and dominates δ^o) as desired.

We end this section with a lemma which is useful in indicating the improvement in risk obtainable by using the new estimators. It is, unfortunately, very difficult to calculate the risk of the estimator (3.5). A related estimator, for which the risk can be explicitly calculated, is given componentwise by

$$\delta_i^c(\underset{\sim}{X}) = X_i - \frac{v_i}{(v_i+\tau_{ii})}(X_i - \xi_i)r_i^c(\underset{\sim}{X}), \tag{3.7}$$

where

$$r_i^c(\underset{\sim}{X}) = \frac{c}{q_i}\sum_{j=i}^{\infty}(q_j - q_{j+1})\frac{(j-1)^+}{\sum_{\ell=0}^{j}(X_\ell-\xi_\ell)^2/(v_\ell+\tau_{\ell\ell})}.$$

(Recall that $q_i = v_i^2/(v_i + \tau_{ii})$.) This estimator is minimax for $0 \leq c \leq 2$, but is most likely inferior to the estimator δ^M.

Lemma 2. If $\underset{\sim}{\theta}$ has an $\mathcal{N}_p(\underset{\sim}{\xi},\underset{\sim}{\tau})$ prior distribution π, where $\underset{\sim}{\tau}$ is the diagonal matrix with diagonal elements τ_{ii}, then

$$r(\pi,\underset{\sim}{\delta}^c) = \sum_{i=0}^{\infty} v_i - 2cA_1 + c^2A_2,$$

where

$$A_1 = \sum_{i=2}^{\infty} q_i$$

and

$$A_2 = \sum_{i=0}^{\infty} \frac{1}{q_i} \sum_{j=i}^{\infty} \frac{(j-1)^+(q_j^2-q_{j+1}^2)}{(j+1)}$$

$$= \sum_{i=2}^{\infty} q_i - 2 \sum_{i=2}^{\infty} \frac{q_i}{i+1} [1 - \frac{q_i}{i} \sum_{j=0}^{i-1} \frac{1}{q_j}] .$$

The optimal minimax choice of c is

(3.8) $$c^* = \min\{2, A_1/A_2\} ,$$

and the corresponding Bayes risk is

$$r(\pi, \underset{\sim}{\delta}^{c^*}) = \sum_{i=0}^{\infty} v_i - 2c^*A_1 + c^{*2}A_2.$$

Proof. The proof is exactly analogous to the proof of Lemma 2 in Berger (1982) and will be omitted. ||

4. Examples

Example 1. Finite Dimensional Problem: The complete analysis is carried out in Berger (1982), to which the reader is referred.

Example 2. Brownian Motion: In applying the minimax results in Section 3 to Example 2, a slight modification of the problem seems indicated. Our original intuition behind this modification was that Bayesian estimators discussed in Section 3 shift the X_i towards the ξ_i, and hence might shift $X(0)$ towards $\xi(0)$, an undesirable feature in that

$X(0)=\theta(0)$ exactly. Referees pointed out that it is unclear whether or not $X(0)$ gets shifted, and we agree. An alternate intuition for the modification we propose is that explicitly incorporating the knowledge that $X(0)=\theta(0)$ seems to result in a problem with smaller prior variance and hence a minimax estimator with smaller Bayes risk. We give an example to illustrate this phenomenon after discussing the modification.

To explicitly incorporate the information that $X(0)=\theta(0)$, define

$$X^*(\cdot) = X(\cdot) - X(0),$$

$$\theta^*(\cdot) = \theta(\cdot) - \theta(0),$$

and, for any estimator δ,

$$\delta^*[X](\cdot) = \delta[X](\cdot) - X(0).$$

It is easy to check that X^* is itself a Wiener process with mean θ^* and covariance function $\gamma^*(s,t) = \sigma^2\min\{s,t\} = \gamma(s,t)$, and furthermore that

$$\begin{aligned} L(\theta,\delta[X]) &= \int_0^T [\theta(t) - \delta[X](t)]^2 dt \\ &= \int_0^T [\theta^*(t) - \delta^*[X](t)]^2 dt \\ &= L(\theta^*,\delta^*[X]). \end{aligned}$$

Thus the "*" problem is formally the same as the original problem, and all previous results apply to it. The key difference is in the fact that the prior information transforms into

$$\xi^*(\cdot) = \xi(\cdot) - \xi(0)$$

and

$$\begin{aligned} (4.1)\quad \tau^*(t,s) &= E\{[\theta^*(t) - \xi^*(t)][\theta^*(s) - \xi^*(s)]\} \\ &= E\{[(\theta(t) - \xi(t)) - (\theta(0) - \xi(0))][(\theta(s) - \xi(s))(\theta(0) - \xi(0))]\} \\ &= \tau(t,s) - \tau(t,0) - \tau(0,s) + \tau(0,0). \end{aligned}$$

Determination of ξ^* and τ^* can thus either be done directly through subjective consideration of θ^*, or from the prior knowledge ξ and τ about θ.

As a specific example of the above considerations suppose it is determined that $\tau^*(t,s) = \lambda\min\{t,s\}$. This is sensible, in that $\tau^*(t,t) = \lambda t$, reflecting the likely fact that the uncertainty in the guess $\xi^*(t)$ is increasing in t (it being known that $\xi^*(0) = 0$). The constant λ would reflect the degree of uncertainty in ξ^*. Then

$$\xi_i^* = \int_0^T \xi^*(s)e_i(s)ds = \int_0^T \xi^*(s)(\tfrac{2}{T})^{\frac{1}{2}} \sin[(\tfrac{1}{2}+i)\pi s/T]ds$$

and

$$\tau_{ii}^* = \int_0^T \tau^*(t,s)e_i(t)e_i(s)dtds = \lambda T^2/[\pi^2(\tfrac{1}{2}+i)^2],$$

the last calculation being like those in Section 2. Some algebra then gives that the estimator δ^M in (3.5) is given coordinatewise (for the "*" problem) by

$$\delta_i^{M*}(\underset{\sim}{X}^*)$$

$$= X_i^* - 2(i+\tfrac{1}{2})^2(X_i^* - \xi_i^*) \sum_{j=i}^{\infty} \frac{(j+1)}{(j^2+2j+\frac{3}{4})^2} \min\left\{ \frac{\sigma^2}{(\sigma^2+\lambda)}, \frac{2(j-1)^+}{||\underset{\sim}{X}^*-\underset{\sim}{\xi}^*||_j^2} \right\},$$

where $||\underset{\sim}{X}^* - \underset{\sim}{\xi}^*||_j^2 = \sum_{\ell=0}^{j} (X_\ell^* - \xi_\ell^*)^2\pi^2(\ell+\frac{1}{2})^2/(\sigma^2T^2)$, and

$$X_i^* = \int_0^T X^*(t)e_i(t)dt = \int_0^T [X(t)-X(0)]e_i(t)dt$$

$$= X_i - X(0)\int_0^T (\tfrac{2}{T})^{\frac{1}{2}} \sin[(\tfrac{1}{2}+i)\pi s/T]ds$$

$$= X_i - \frac{(2T)^{\frac{1}{2}}X(0)}{\pi(\frac{1}{2}+i)} .$$

Also, $\xi_i^* = \xi_i - (2T)^{\frac{1}{2}}X(0)/[\pi(\frac{1}{2}+i)]$, and the estimator δ^M for the original problem is $\delta^M[X](\cdot) = \delta^{M*}[X](\cdot) + X(0)$ or, equivalently, $\delta_i^M(\underset{\sim}{X}) = \delta_i^{M*}(\underset{\sim}{X}^*) + (2T)^{\frac{1}{2}}X(0)/[\pi(\frac{1}{2}+i)]$. Hence we can finally write the desired estimator as

$$\delta_i^M(\underset{\sim}{X}) = X_i - 2(i+\tfrac{1}{2})^2[(X_i-\xi_i) - \frac{(2T)^{\frac{1}{2}}}{\pi(\frac{1}{2}+i)}(X(0)-\xi(0))]$$

$$\times \sum_{j=i}^{\infty} \frac{(j+1)}{(j^2+2j+\frac{3}{4})^2} \min\left\{\frac{\sigma^2}{\sigma^2+\lambda}, \frac{2(j-1)^+}{||\underset{\sim}{X}-\underset{\sim}{\xi}||_j^2}\right\},$$

where $||\underset{\sim}{X}-\underset{\sim}{\xi}||_j^2 = \sum_{\ell=0}^{j} [(X_i-\xi_i) - \frac{(2T)^{\frac{1}{2}}}{\pi(\frac{1}{2}+\ell)}(X(0)-\xi(0))]\pi^2(\ell+\tfrac{1}{2})^2/(\sigma^2T^2)$.

Observing that

$$q_i = v_i^2/(v_i+\tau_{ii}^*) = \frac{\sigma^4T^2}{(\sigma^2+\lambda)\pi^2(\frac{1}{2}+i)^2},$$

a calculation (see Lemma 2) gives that

$$A_1 \cong \frac{\sigma^4T^2}{(\sigma^2+\lambda)}(.0497),$$

$$A_2 \cong \frac{\sigma^4T^2}{(\sigma^2+\lambda)}(.0331),$$

and

$$c^* = \min\{2, A_1/A_2\} \cong 1.503.$$

The minimax estimator δ^{c^*} (see (3.7) and (3.8)) thus has Bayes risk

$$r(\pi,\underset{\sim}{\delta}^{c^*}) = \sum_{i=0}^{\infty} v_i - 2c^*A_1 + c^{*2}A_2$$

$$= \frac{\sigma^2T^2}{2} - \frac{\sigma^4T^2}{(\sigma^2+\lambda)}(.0746)$$

$$= \frac{\sigma^2T^2}{2}\left(1 - \frac{\sigma^2}{(\sigma^2+\lambda)}(.149)\right).$$

An analysis of the original problem (i.e., without attempting to explicitly incorporate the information $X(0) = \theta(0)$) can be done along the same lines using $\tau(t,s)$. We were unable to prove that the modified analysis is always better, and so merely considered an example, namely,

$$\tau(t,s) = \lambda(K - \tfrac{1}{2}\,|t-s|), \text{ where } K \geq T/4.$$

For the indicated range of K this is a valid covariance function, and when used in (4.1) gives the $\tau^*(t,s)$ evaluated earlier. (This $\tau(t,s)$ is also sensible intuitively, the variance function $\tau(t,t)=\lambda K$ indicating a constant estimated variance for the guess $\xi(\cdot)$ of $\theta(\cdot)$.) A calculation of the Bayes risk of δ^{c*} gives

$$r(\pi,\delta^{c*}) = \frac{\sigma^2 T^2}{2}\left(1 - \frac{\sigma^2}{(\sigma^2+\lambda)}\, h\left(\frac{\lambda}{\sigma^2}, \frac{K}{T}\right)\right),$$

where h is a decreasing function of its arguments, going to zero as λ/σ^2 or K/T go to infinity, and going to .149 (the answer for the modified problem) as λ/σ^2 goes to zero. For a numerical example, when $\lambda/\sigma^2=1$ and $K/T=\frac{1}{2}$ then h=.095. For prior covariance functions of this particular form, therefore, the modified minimax estimator appears to be better than the unmodified minimax estimator. Again, however, there is no theory to support this conclusion in general.

Example 3. Brownian Bridge with Weighted Loss: As in Example 2, the initial difficulty is encountered that $X(0) = \theta(0)$ and $X(T) = \theta(T)$ exactly. To ensure that the minimax estimators estimate $\theta(0)$ and $\theta(T)$ correctly, define, for $0 \leq t \leq T$,

$$X^*(t) = X(t) - [X(0)(1-\frac{t}{T}) + X(T)\frac{t}{T}],$$

$$\theta^*(t) = \theta(t) - [\theta(0)(1-\frac{t}{T}) + \theta(T)\frac{t}{T}],$$

$$\delta^*[X](t) = \delta[X](t) - [X(0)(1-\frac{t}{T}) + X(T)\frac{t}{T}],$$

and

$$\xi^*(t) = \xi(t) - [\xi(0)(1-\frac{t}{T}) + \xi(T)\frac{t}{T}].$$

Again the "*" problem is the same as the original problem, in that $X^*(t) = \theta^*(t) + Z(t)$, where $\{Z(t), t \in [0,T]\}$ is a Brownian bridge.

Since the prior information in this transformed problem is that $\theta^*(0) = \xi^*(0) = 0$ and $\theta^*(T) = \xi^*(T) = 0$ exactly, a reasonable prior covariance function is

$$\tau^*(t,s) = \lambda\left\{\min(s,t) - \frac{st}{T}\right\}.$$

Calculation gives

$$\tau_{ii} = \lambda T/[(i+1)(i+2)].$$

We won't bother to write out the estimator (3.5) in this case, but it is of interest that for the estimator δ^c in (3.7), the optimal choice of c^* is $c^* = 1.3269$, and the Bayes risk of δ^{c^*} is

$$r(\pi, \underset{\sim}{\delta}^{c^*}) = \sigma^2 T - \frac{\sigma^4 T}{(\sigma^2+\lambda)}(.4423)$$

$$= \sigma^2 T\left(1 - \frac{\sigma^2}{(\sigma^2+\lambda)}(.4423)\right).$$

Hence the estimator $\underset{\sim}{\delta}^{C*}$ can be up to 44% better than $\underset{\sim}{X}$ in terms of Bayes risk. Again, $\underset{\sim}{\delta}^{M}$ probably performs even better.

5. Conclusions and Generalizations

Comment 1. A comment is in order concerning the relationship of the results in this paper with standard "filtering theory". The classical estimator for the situation posed here is indeed δ^{0}, but more complicated situations are usually considered. In the first place, it is often assumed that $X(\cdot)$ is only observed at, say, n points or that $X(\cdot)$ is observed in a nonanticipative context (i.e., only $X(s)$ for $s \leq t$ is known when $\theta(t)$ is to be estimated) or both. Such models are outside the scope of this paper. Also, it is often assumed that $\theta(\cdot)$ belongs to some subset of Θ (say is an n^{th} degree polynomial), or that $\{\theta(t), t \in I\}$ really is a random process, perhaps jointly distributed with the "observational noise". In either case the classical problem is to filter out the noise, arriving at an estimate of $\theta(\cdot)$. Thus our analysis deals with what could be called the "noninformative" situation, in which we do not have (completely trustworthy) auxilliary information of this type.

To more clearly see the difference between the analyses, suppose it is felt that θ is itself a Gaussian process, independent of $X - \theta$, with mean function and covariance function as discussed in Section 4. Then the classical filter would be essentially the estimator determined by (3.5) with $r_i(\cdot) \equiv 1$, and this would be optimal. If, however, the beliefs about θ are uncertain, corresponding to vague prior information rather than concrete knowledge, this classical filter could be dangerous, in the sense that its risk can be very bad if $\theta(\cdot)$ is not what was anticipated (i.e., is not near $\xi(\cdot)$). The estimator proposed in (3.5) "hedges the bet", partially filtering out the suspected noise, but in a totally safe way; no matter how far the true $\theta(\cdot)$ is from $\xi(\cdot)$, the estimator in (3.5) will still be better than the unfiltered δ^{0}. These considerations can also be phrased in "robust Bayesian" terms. See Berger (1980b) for discussion. Of course, if one is really quite confident about the added information about $\theta(\cdot)$ (and usually one will know more than just that $\theta(\cdot)$ is in $L^2(I,d\mu)$), then classical filters should be used.

Comment 2. Even in the basic situation discussed here, certain generalizations would be very desirable. First of all, it would be nice to be able to carry out the analysis for an arbitrary complete system of functions $\{e_i(\cdot)\}$, so that a system could be chosen which is appropriate for the expected form of $\theta(\cdot)$. Also, it would be very useful to be able to handle more general types of prior information about $\theta(\cdot)$, particularly information relating to smoothness of the functions, or to knowledge that $\theta(\cdot)$ lies in some subspace of Θ. Finally, it is often not $\theta(\cdot)$ itself which is of interest, but some functional of $\theta(\cdot)$.

For instance, in Example 2 it is frequently the derivative of θ which is of interest, such as when a signal plus "white noise" is observed, in which case the standard mathematical treatment is to integrate and consider the integrated signal plus Brownian motion (which is the integral of white noise, in some sense).

Comment 3. The results here easily generalize to the situation in which $\theta(\cdot)$ (but not $Z(\cdot)$) is permitted to have jumps at a discrete set of points in I, by setting $\theta^J(t) = \sum_{s \le t} [\theta(t) - \lim_{s \uparrow t} \theta(s)]$ and estimating $\theta^c \equiv \theta - \theta^J$ upon observing $X^c = X - X^J$; the assumption that $Z(\cdot)$ has no jumps ensures that $X^J(\cdot) = \theta^J(\cdot)$. The problem of estimating θ^c by observing X^c can be solved as before.

<u>Comment 4.</u> The assumption that X and θ have continuous paths is convenient but not necessary; many other function spaces could be substituted for $\mathcal{X}$. However, in the proof of the minimaxity of the usual estimator δ^0 (Theorem 3.1) we use the sufficiency of $\tilde{X}$ for $\tilde{\theta}$, a consequence of the fact that the mapping $X \to \tilde{X}$ from $\mathcal{X} \cap L^2(I,d\mu)$ to ℓ^2 is one-to-one; δ^0 may not be minimax if $\mathcal{X}$ is replaced by a space for which this map is not injective.

As an example, consider the space $\mathcal{X}$ of bounded measurable functions with the supremum norm on I=[0,1], with Θ the continuous functions. Let B(·) be a Brownian motion, U a random variable independent of B and uniformly distributed on [0,1] and set

$$Z(t) = \begin{cases} 0 \text{ if } t-U \text{ is rational} \\ B(t) \text{ else,} \end{cases}$$

$$X(t) = Z(t) + \theta(t).$$

Then it is possible to estimate θ perfectly with a continuous estimator $\delta^*: \mathcal{X} \to L^2(I,d\mu)$ (here μ is Lesbesgue measure), while δ^0 has constant risk $\frac{1}{2}$.

The technically simple choice of $\mathcal{X} = L^2(I,d\mu)$ is unappealing in cases where the loss measure μ may have support smaller than all of I. The restriction of X to the support of μ is not (in general) a sufficient statistic for estimating the restriction of θ to the support of μ, so the statistical problem of estimating θ upon observing $X \in \mathcal{C}(I)$ may differ substantially from the problem of estimating θ upon observing $X \in L^2(I,d\mu)$.

Acknowledgements. We are grateful to Gopinath Kallianpur who suggested the problem to us and to Herman Rubin for helpful ideas.

REFERENCES

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

Berger, J. (1980a). A robust generalized estimator and confidence region for a multivariate normal mean. Ann. Statist. 8, 716-761.

Berger, J. (1980b). Statistical Decision Theory: Foundations, Concepts, and Methods. Springer-Verlag, New York.

Berger, J. (1982). Selecting a minimax estimator of a multivariate normal mean. Ann. Statist. 10, 81-92.

Berger, J. and Bock, M.E. (1976). Eliminating singularities of Stein-type estimators of location vectors. J. Roy. Statist. Soc. B, 38, 166-170.

James, W. and Stein, C. (1960). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1, 361-379. University of California Press.

Kallianpur, G. (1980). Stochastic Filtering Theory. Springer-Verlag, New York.

Lipster, E. Sh. and Shiryayev, A.N. (1977). Statistics of Random Processes. Springer-Verlag, New York.

Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. 1, 197-206. University of California Press.

Wolpert, R.L. and Berger, J. (1982). Incorporating prior information in minimax estimation of the mean of a Gaussian process. Statistical Decision Theory and Related Topics III, S.S. Gupta and J. Berger (eds.). Academic Press, New York.

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