1987

Admissibility as a Touchstone

Avishai Mandelbaum

Larry A. Shepp
University of Pennsylvania

Recommended Citation

Mandelbaum, A., & Shepp, L. A. (1987). Admissibility as a Touchstone. *The Annals of Statistics, 15* (1), 252-268. http://dx.doi.org/10.1214/aos/1176350264

Admissibility as a Touchstone

Abstract

Consider the problem of estimating simultaneously the means θ_i of independent normal random variables x_i with unit variance. Under the weighted quadratic loss $L(\theta,a)=\Sigma_i\lambda_i(\theta_i-a_i)^2$ with positive weights it is well known that:

(1) An estimator which is admissible under one set of weights is admissible under all weights.

(2) Estimating individual coordinates by proper Bayes estimators results in an admissible estimator.

(3) Estimating individual coordinates by admissible estimators may result in an inadmissible estimator, when the number of coordinates is large enough. A dominating estimator must link observations in the sense that at least one θ_i is estimated using observations other than x_i.

We consider an infinite model with a countable number of coordinates. In the infinite model admissibility does depend on the weights used and by linking coordinates it is possible to dominate even estimators which are proper Bayes for individual coordinates. Specifically, we show that when θ_i are square summable, the estimator $\delta_i(x)\equiv 1$ is admissible for $\lambda_i=e^{-ic},c>1/2$, but inadmissible for $\lambda_i=1/i^{1+c},c>0$. In the latter case, a dominating estimator $\pi=(\pi_1,\pi_2,\cdots)$ is of the form $\pi_i(x)=1-\varepsilon_i(x)$, where ε_i links all the observations $x_1,x_2,\cdots$.

Infinite models frequently arise in estimation problems for Gaussian processes. For example, in estimating the drift function θ of the Wiener process W under the loss $L(\theta,a)=\int[\theta(t)-a(t)]^2dt$, the transformation $x_i=\int\Phi_i dW$ with Φ_i an appropriate complete orthonormal sequence gives rise to a model which is equivalent to an infinite model with $\lambda_i = 1/i^2$.

Keywords

admissibility, James-Stein estimation, Gaussian processes, Karhunen-Loeve expansion

Disciplines

Statistics and Probability

The Annals of Statistics
1987, Vol. 15, No. 1, 252–268

ADMISSIBILITY AS A TOUCHSTONE

BY AVI MANDELBAUM AND L. A. SHEPP

Stanford University, AT & T Bell Laboratories, Inc. and Stanford University

Consider the problem of estimating simultaneously the means θ_i of independent normal random variables x_i with unit variance. Under the weighted quadratic loss $L(\theta, a) = \Sigma_i \lambda_i(\theta_i - a_i)^2$ with positive weights it is well known that:

(1) An estimator which is admissible under *one* set of weights is admissible under all weights.
(2) Estimating individual coordinates by proper Bayes estimators results in an admissible estimator.
(3) Estimating individual coordinates by admissible estimators may result in an *inadmissible* estimator, when the number of coordinates is *large enough*. A dominating estimator must link observations in the sense that at least one θ_i is estimated using observations other than x_i.

We consider an infinite model with a countable number of coordinates. In the infinite model admissibility *does* depend on the weights used and by linking coordinates it is possible to dominate even estimators which are *proper Bayes* for individual coordinates. Specifically, we show that when θ_i are square summable, the estimator $\delta_i(x) \equiv 1$ is admissible for $\lambda_i = e^{-ic}$, $c > \frac{1}{2}$, but inadmissible for $\lambda_i = 1/i^{1+c}$, $c > 0$. In the latter case, a dominating estimator $\pi = (\pi_1, \pi_2, \dots)$ is of the form $\pi_i(x) = 1 - \varepsilon_i(x)$, where ε_i links *all* the observations $x_1, x_2, \dots$.

Infinite models frequently arise in estimation problems for Gaussian processes. For example, in estimating the drift function θ of the Wiener process W under the loss $L(\theta, a) = \int[\theta(t) - a(t)]^2\, dt$, the transformation $x_i = \int \Phi_i\, dW$ with Φ_i an appropriate complete orthonormal sequence gives rise to a model which is equivalent to an infinite model with $\lambda_i = 1/i^2$.

1. Introduction and summary of results.

1.1. Consider the problem of simultaneously estimating the coordinate means $\theta = (\theta_1, \theta_2, \dots)$ of independent normal observations $x_1, x_2, \dots$, all with variance one. The concept of admissibility of an estimator allows one to delimit the class of reasonable estimators: All inadmissible estimators are judged unreasonable on the grounds that each can be outperformed uniformly by a dominating estimator. One is left with the admissible estimators, from which a recommended one is to be chosen. However, the admissibility–inadmissibility criterion leads to rather unexpected consequences: While estimators that are clearly unreasonable in most circumstances turn out to be admissible, estimators which are natural and

Received August 1985; revised July 1986.
AMS 1980 *subject classifications*. Primary 62C15; secondary 62H12.
Key words and phrases. Admissibility, James–Stein estimation, Gaussian processes, Karhunen–Loève expansion.

widely used are inadmissible. More specifically, under the weighted square error loss

$$L(\theta, a) = \sum_i \lambda_i (\theta_i - a_i)^2, \qquad \lambda_i > 0, \tag{1.1}$$

the constant estimator $\delta^c(x) \equiv (c, c, \dots)$, c constant, which does not even use the observations, is admissible while the estimator $\delta(x) = (x_1, x_2, \dots)$ is inadmissible if the number of coordinates is three or more.

The inadmissibility of $\delta(x) = x$ was first discovered by Stein [13] for the case $\lambda_i \equiv 1$ and further extended to a larger class of loss functions by Brown [4]. The class of estimators that dominate $\delta(x) = x$ is very wide (some are described in Berger [1]). However, all share a common feature: Coordinates must be linked in the sense that the estimation of at least one θ_i involves observations other than x_i.

1.2. Stein's result has inspired a great deal of research, as demonstrated by the large amount of work devoted to the subject. A recent textbook survey of the literature can be found in Lehmann [8]. Most research efforts have been devoted to the phenomenon that estimators, which at first glance seem good, are actually improvable. We are dealing, roughly speaking, with the dual phenomenon in which estimators that are clearly unreasonable turn out to be unimprovable. An example of such estimators are the constant estimators δ^c. A less trivial example, due to Brown, is described in Makani [9].

1.3. Stein [13] supported his inadmissibility proof by a heuristic argument which applies when the number of coordinates is large. We take the notion of a large number of coordinates to the extreme, and consider *a countable* number: An infinite sequence $x = (x_1, x_2, \dots)$ is observed and the sequence of means $\theta = (\theta_1, \theta_2, \dots)$ is estimated under the loss (1.1). For brevity, the model with a countable number of coordinates will be called the *infinite* model, to distinguish it from the usual multivariate model which we call the *finite* model.

Assume that the coordinate mean vector $\theta = (\theta_1, \theta_2, \dots)$ is square summable ($\theta \in \ell^2$). Denote the risk function of an estimator $\delta(x) = (\delta_1(x), \delta_2(x), \dots)$ associated with the loss (1.1) by

$$R(\theta, \delta) = E_\theta \sum_{i=1}^{\infty} \lambda_i [\delta_i(x) - \theta_i]^2, \qquad \theta \in \ell^2, \tag{1.2}$$

where E_θ expresses the fact that $x_i \sim N(\theta_i, 1)$. Let 1 stand for the estimator $\delta^1(x) \equiv (1, 1, \dots)$. For 1 to have finite risk, we must and shall restrict attention to summable weights:

$$\sum_i \lambda_i < \infty.$$

The situation where only a finite number, say N, of coordinates is of interest can be captured by assuming that $\lambda_i \equiv 0$ for $i > N$. In view of the fact that, in the finite model 1 is trivially admissible, the following result (proved in Section

3) may not be too surprising:

1.3.A. If the weights λ_i decrease to 0 sufficiently fast, the estimator 1 is admissible.

However, our next result has no analogue in the finite model. We prove in Section 2:

1.3.B. If the weights λ_i decrease to 0 sufficiently slowly, the estimator 1 is inadmissible. An estimator π that dominates it over ℓ^2 is of the form

$$\pi_i(x) = 1 - \varepsilon(1 - x_i)e^{-q_i(x)}, \tag{1.3}$$

where ε is positive, chosen small enough, and $q_i(x)$ are positive quadratic forms in x.

Some remarks are now in order.

1.4. In the finite model, the estimator 1 is trivially admissible because it has zero risk at the point $\theta = 1$. However, in the infinite model, $\theta = 1$ has been excluded from the parameter space since 1 is not square summable. The square summability of coordinates is a natural restriction for modeling purposes. Indeed, suppose that θ^1 and θ^2 are such that their difference $\theta^1 - \theta^2$ is not square summable. Then the distributions of the sequence x under θ^1 and θ^2 have essentially disjoint support (Shepp [11]). So, in principle at least, θ^1 and θ^2 are distinguishable almost surely. Restricting attention to parameters in ℓ^2 is an idealization of the following situation. Suppose that the parameter space is centered around a *known* vector θ^0; parameters θ for which $\theta - \theta^0$ is not square summable are considered outliers and can be detected with certainty. Then, considering ℓ^2 amounts to choosing $\theta^0 = 0$ for convenience.

1.5. The estimator π given in (1.3) links coordinates in the strongest sense: Each θ_i is estimated using *all* the coordinates of $x = (x_1, x_2, \dots)$. Indeed, any estimator $\delta = (\delta_1, \delta_2, \dots)$ which dominates 1 must link *infinitely* many coordinates of x in at least one δ_i. This is a consequence of the fact that if δ dominates 1 and if $\delta_j(x) = f(x_1, \dots, x_k)$ for some j and $k < \infty$, then $f(x_1, \dots, x_k) \equiv 1$ a.s. To see that, let $e_i^N = 1$ or 0 according to $i \le N$ or $i > N$. Since δ dominates 1, we have for all $N > k$

$$\lambda_j E_0[f(x_1 + 1, \dots, x_k + 1) - 1]^2 \le R(\delta, e^N) \le R(1, e^N) = \sum_{i=N+1}^{\infty} \lambda_i. \tag{1.4}$$

Letting $N \to \infty$ in (1.4), we conclude that $f(x_1, \dots, x_k) \equiv 1$ a.s.

1.6. In the finite model, an estimator that is admissible under square error loss with respect to one set of positive weights is admissible with respect to all possible weights (Bhattacharya [3] and Shinozaki [12]). The results 1.3.A and 1.3.B demonstrate that this fact does not carry over to the infinite model.

1.7. In addition to being a limit of the finite model, the infinite model arises in inference problems for Gaussian processes. Via a Karhunen–Loève type expansion, inference problems for Gaussian processes are reduced to inference problems for the infinite model. This approach is well illustrated by Grenander [6] where both testing and estimation problems are treated. In particular, the approach is useful in reducing admissibility questions for estimators of the means of general Gaussian distributions to admissibility questions for sequences of means in the infinite model. Berger and Wolpert [2] considered James–Stein estimators for mean functions of Gaussian processes along these lines. In [10], linear estimators for the mean of a Gaussian distribution on a Hilbert space were characterized under a very special loss function. The model in [10] reduces to an infinite model with weights $\lambda_i \equiv 1$ which is treated there in details. With weights $\lambda_i \equiv 1$ the estimator 1 is trivially inadmissible, having a risk that is identically infinite. Also the characterization of admissible linear estimators in both the finite and infinite models are similar. The present paper is a first attempt to extend the class of loss functions beyond the simplest loss function considered in [10]. Only here, the striking differences between the finite and infinite models become apparent.

For a recent textbook treatment of the general Gaussian model and its specialization to the Brownian motion, the reader is referred to Chapter 8 in Farrell [5]. Since we know of no reference in which the reduction to an infinite model sufficiently emphasizes the role of the loss function chosen, we have included an appendix that addresses this issue. The relation of the infinite model to the important Brownian motion model is described there as well. It is worth mentioning that the Gaussian model described in the Appendix includes essentially all Gaussian processes of practical importance, not only the Brownian motion. Further examples are described in [2] and [6].

1.8. We restrict the parameter space to ℓ_2 and show that 1 can be inadmissible. The point 1 is not in the parameter space. However, for summable weights, 1 is in the closure of the parameter space with respect to the norm that defines the loss function. With respect to norms with weights that decrease to 0 sufficiently slowly, the point 1 is so "far" from the parameter space that it is inadmissible. This is somewhat analogous to the finite dimensional situation where 1 is inadmissible if θ is restricted to $|\theta - 1| \geq \delta > 0$ (the convex combination $\alpha x + (1 - \alpha) \cdot 1$ dominates 1 for α positive small enough).

1.9. The paper is structured as follows. The inadmissibility of 1 under weights $\lambda_i = 1/i^{1+c}$, $c > 0$, is proved in Section 2. The admissibility of 1 under weights $\lambda_i = e^{-ai}$, $a > \frac{1}{2}$, is proved in Section 3. Section 4 includes concluding remarks and some intriguing open questions. The paper ends with an Appendix whose role was described above.

2. Inadmissibility of 1 under polynomially decreasing weights.

2.1. We now prove that if the weights in (1.2) are given by

$$\lambda_i = \frac{1}{i^{1+c}}, \quad \text{where } c > 0, \tag{2.1}$$

then the estimator 1 is inadmissible. An estimator $\pi(x) = (\pi_1(x), \pi_2(x), \dots)$ which dominates 1 at all $\theta \in \ell^2$ is of the form

$$\pi_i(x) = 1 - \varepsilon(1 - x_i)e^{-q(x^i)}, \qquad i \geq 1, \tag{2.2a}$$

where $q(\cdot)$ is the positive quadratic form

$$q(x) = \gamma \sum_{j=1}^{\infty} \frac{1}{j^{2+\alpha}}(x_1 + \cdots + x_j)^2 + \delta \sum_{j=1}^{\infty} \frac{1}{j^{1+\beta}} x_j^2, \tag{2.2b}$$

and x^i [the argument of q in (2.2a)] is just x with the ith coordinate made zero, i.e., $x_i^i = 0$ and $x_j^i = x_j$ for $j \neq i$. The positive constants $0 < \alpha < \beta < 1$ and γ, δ, ε will be chosen small enough so that

$$R(\theta, \pi) < R(\theta, 1), \qquad \forall\, \theta \in \ell^2,$$

or, equivalently,

$$E_\theta \sum_{i=1}^{\infty} \lambda_i(\pi_i(x) - \theta_i)^2 < \sum_{i=1}^{\infty} \lambda_i(1 - \theta_i)^2, \qquad \forall\, \theta \in \ell^2. \tag{2.3}$$

2.2. The quadratic form in (2.2b) converges a.s. under all $\theta \in \ell^2$. Indeed, $E_0 q(x^i) \leq E_0 q(x)$ for all i and

$$E_0 q(x) = \gamma \sum_{j=1}^{\infty} \frac{1}{j^{2+\alpha}} j + \delta \sum_{j=1}^{\infty} \frac{1}{j^{1+\beta}} 1 < \infty.$$

Thus, $q(x) < \infty$ a.s. if $\theta = 0$. But "almost sure" statements under $\theta = 0$ are "almost sure" statements under $\theta \in \ell^2$ (Shepp [11]), implying that $q(x^i) < \infty$ a.s. under all $\theta \in \ell^2$. In particular, π differs from 1 in all coordinates.

2.3. We now start proving (2.3). The first step is to expand squares, divide by $\varepsilon > 0$, and use the independence of x_i and x^i and the identities

$$E_\theta(1 - x_i) = 1 - \theta_i, \qquad E_\theta(1 - x_i)^2 = 1 + (1 - \theta_i)^2,$$

to get an equivalent inequality

$$\varepsilon \sum_i \lambda_i\left[1 + (1 - \theta_i)^2\right] E_\theta e^{-2q(x^i)} < 2 \sum_i \lambda_i (1 - \theta_i)^2 E_\theta e^{-q(x^i)}, \qquad \forall\, \theta \in \ell^2. \tag{2.4}$$

Clearly, (2.4) holds for some $\varepsilon > 0$ if and only if

$$\inf_{\theta \in \ell^2} \frac{2\Sigma\lambda_i(1 - \theta_i)^2 E_\theta e^{-q(x^i)}}{\Sigma\lambda_i\left[1 + (1 - \theta_i)^2\right] E_\theta e^{-2q(x^i)}} > 0. \tag{2.5}$$

2.4. The hope that (2.5) holds is based on the heuristic that the numerator is small only when $\theta_i \approx 1$ for many i's. But then $e^{-q(x^i)}$ must be small and $e^{-2q(x^i)}$ even smaller. This heuristic is even more founded if we replace in (2.5) the expectations $E_\theta[\cdot]$ by the expectants evaluated at θ. We now prove two lemmas that justify such a step.

LEMMA 1. *Let Q be an $n \times n$ symmetric positive definite matrix. Denote the quadratic form associated with Q by $Q(x) = x'Qx$. Then*

$$E_0 Q(x) = \operatorname{trace}(Q)$$

and

$$e^{-Q(\theta)/2} \le \frac{E_\theta e^{-Q(x)/2}}{E_0 e^{-Q(x)/2}} \le e^{-(1-\varepsilon)Q(\theta)/2}, \tag{2.6}$$

for all $\varepsilon \ge \operatorname{trace}(Q)/[1 + \operatorname{trace}(Q)]$ *and all* $\theta \in R^n$.

PROOF. By the multidimensional Gauss formula

$$E_\theta e^{-Q(x)/2} = E_0 e^{-Q(x)/2} \exp\left\{\tfrac{1}{2}\theta'\left[(I+Q)^{-1} - I\right]\theta\right\}.$$

Let $\{\nu_1, \dots, \nu_n\}$ be the positive eigenvalues of Q.

For $\varepsilon \ge \max_i \nu_i/(1+\nu_i)$

$$(1-\varepsilon)Q \le I - [I+Q]^{-1} \le Q,$$

since

$$(1-\varepsilon)\nu_i \le 1 - (1+\nu_i)^{-1} \le \nu_i, \quad \text{for all } i.$$

Finally $\operatorname{trace}(Q) \ge \max_i(\nu_i)$ and the fraction $x/(1+x)$ is increasing in x, hence (2.6). □

REMARK. We shall apply (2.6) to infinite matrices or, more precisely, to continuous positive quadratic forms Q on ℓ^2. As long as $E_0 e^{-Q(x)/2}$ is strictly positive, (2.6) generalizes without difficulties.

2.5. Before stating Lemma 2, let us introduce some convenient notation:

$$\Delta(\theta) = \sum_j \frac{1}{j^{2+\alpha}}(\theta_1 + \cdots + \theta_j)^2, \qquad \sigma(\theta) = \sum_j \frac{1}{j^{1+\beta}}\theta_j^2; \tag{2.7a}$$

$$\Delta_i(\theta) = \Delta(\theta^i), \qquad \sigma_i(\theta) = \sigma(\theta^i); \tag{2.7b}$$

$$q(\theta) = \gamma\Delta(\theta) + \delta\sigma(\theta), \qquad q_i(\theta) = q(\theta^i). \tag{2.7c}$$

The q in (2.7c) is the quadratic form in (2.2b). Clearly, $\sigma(\theta)$ and $\sigma_i(\theta)$ are continuous positive quadratic forms on ℓ^2. The inequality

$$(\theta_1 + \cdots + \theta_j)^2 \le j(\theta_1^2 + \cdots + \theta_j^2)$$

implies that the rest of the positive quadratic forms in (2.7) are continuous as well.

Now fix α, β and ε, $0 < \varepsilon < \frac{1}{2}$.

LEMMA 2. *There exists a positive $B = B(\gamma, \delta)$ such that for all γ, δ sufficiently small, all $\theta \in \ell^2$ and all i*

$$Be^{-q(\theta^i)} \le E_\theta e^{-q(x^i)} \le Be^{-(1-\varepsilon)q(\theta^i)}. \tag{2.8}$$

NOTE. The dependence on γ, δ in (2.8) is through B and q.

PROOF. First note that for γ, δ small enough

$$E_0 e^{-q(x^i)} \ge 1 - E_0 q_i(x) \ge 1 - E_0 q(x) > 0, \tag{2.9}$$

since

$$E_0 q(x) = \gamma \sum_j \frac{1}{j^{1+\alpha}} + \delta \sum_j \frac{1}{j^{1+\beta}}.$$

Applying Lemma 1 to $Q = 2q_i$, we get that for all i

$$e^{-q_i(\theta)} \le \frac{E_\theta e^{-q_i(x)}}{E_0 e^{-q_i(x)}} \le e^{-(1-\varepsilon)q_i(\theta)}, \qquad \theta \in \ell^2, \tag{2.10}$$

whenever

$$\varepsilon \ge \frac{2\,\mathrm{trace}(q_i)}{1 + 2\,\mathrm{trace}(q_i)}.$$

Since $\mathrm{trace}(q_i) = E_0 q_i(x) \le E_0 q(x) = \mathrm{trace}(q)$, for

$$\varepsilon \ge \frac{2\,\mathrm{trace}(q)}{1 + 2\,\mathrm{trace}(q)}, \tag{2.11}$$

the inequalities (2.10) hold simultaneously for all i. Finally, given $\alpha, \beta, \varepsilon > 0$, choose γ, δ small enough to satisfy both (2.9) and (2.11), and conclude (2.8) from (2.10). □

2.6. We now apply (2.8) to both $q(x^i)$ and $2q(x^i)$. Substituting the result into (2.5) yields a stronger inequality

$$\inf_{\theta \in \ell^2} \frac{2\Sigma_i \lambda_i (1-\theta_i)^2 e^{-q(\theta^i)}}{\Sigma_i \lambda_i \left[1 + (1-\theta_i)^2\right] e^{-2(1-\varepsilon)q(\theta^i)}} > 0, \tag{2.12}$$

stronger in the sense that (2.12) implies (2.5) [which, in turn, implies (2.2)]. Now, note that $e^{-q(\theta^i)} \ge e^{-2(1-\varepsilon)q(\theta^i)}$, so we can ignore the smaller $(1-\theta_i)^2 e^{-2(1-\varepsilon)q(\theta^i)}$ terms in the denominator of (2.12) [for positive numbers, if $a/b \ge \zeta$ then

$a/(b+c) \geq \zeta/(1+\zeta)$ for all $c \leq a$] and only show that

$$\inf_{\theta \in \ell^2} \frac{\sum_i \lambda_i (1-\theta_i)^2 e^{-q(\theta^i)}}{\sum_i \lambda_i e^{-2(1-\varepsilon)q(\theta^i)}} > 0. \tag{2.13}$$

2.7. We now fix a number $\eta > 0$ throughout the rest of the proof. Separate the sum in the denominator of (2.13) to two sums: One sum is over all i with $|\theta_i - 1| \geq \eta$ and the other is over the rest. The i's for which $|\theta_i - 1| \geq \eta$ satisfy $(1-\theta_i)^2 e^{-q(\theta^i)} \geq \eta^2 e^{-2(1-\varepsilon)q(\theta^i)}$. Again, one may ignore the smaller terms in the denominator of (2.13) to reduce it to

$$\inf_{\theta \in \ell^2} \frac{\eta^2 \sum_{|\theta_i - 1| \geq \eta} \lambda_i e^{-q(\theta^i)}}{\sum_{|\theta_i - 1| < \eta} \lambda_i e^{-2(1-\varepsilon)q(\theta^i)}} > 0. \tag{2.14}$$

REMARK. It is assumed that there are some i's for which $|\theta_i - 1| < \eta$. If there are none, the infimum in (2.13) is greater than η^2 and we are done.

Multiplying numerator and denominator by $e^{2(1-\varepsilon)q(\theta)}$, using the notation in (2.7) and omitting the arguments θ and θ^i for notational convenience, (2.14) reads

$$\inf_{\theta \in \ell^2} \frac{\eta^2 \sum_{|\theta_i - 1| \geq \eta} \lambda_i e^{\gamma(\Delta - \Delta_i) + \delta(\sigma - \sigma_i) + (\tilde{\gamma} - \gamma)\Delta}}{\sum_{|\theta_i - 1| < \eta} \lambda_i e^{\tilde{\gamma}(\Delta - \Delta_i) + \tilde{\delta}(\sigma - \sigma_i)}} e^{(\tilde{\delta} - \delta)\sigma} > 0, \tag{2.15}$$

where $\tilde{\gamma} = 2(1-\varepsilon)\gamma$, $\tilde{\delta} = 2(1-\varepsilon)\delta$. Since $\varepsilon < \frac{1}{2}$ is used, $\tilde{\gamma} > \gamma$ and $\tilde{\delta} > \delta$. The idea now is to choose a new parameter $\zeta < \tilde{\gamma} - \gamma$ so that, up to $\zeta\Delta$, the exponent in the numerator of (2.15) will be nonnegative, namely,

$$\gamma(\Delta - \Delta_i) + \delta(\sigma - \sigma_i) + (\tilde{\gamma} - \gamma - \zeta)\Delta \geq 0. \tag{2.16}$$

Note that

$$\sigma - \sigma_i = \frac{1}{i^{1+\beta}} \theta_i^2 \tag{2.17}$$

and

$$\Delta - \Delta_i = A_i \theta_i^2 + 2B_i \theta_i, \tag{2.18}$$

where

$$A_i = \sum_{j=i}^{\infty} \frac{1}{j^{2+\alpha}}, \qquad B_i = \sum_{j=i}^{\infty} \frac{1}{j^{2+\alpha}} \left(\theta_1^i + \cdots + \theta_j^i\right).$$

By the Cauchy–Schwarz inequality

$$B_i^2 = \left\{ \sum_{j=i}^{\infty} \left[\frac{1}{j^{(2+\alpha)/2}} \right] \left[\frac{1}{j^{(2+\alpha)/2}} \left(\theta_1^i + \cdots + \theta_j^i\right) \right] \right\}^2 \leq A_i \Delta_i. \tag{2.19}$$

Hence,

$$\inf_{\theta_i}\{\gamma(\Delta-\Delta_i)+\delta(\sigma-\sigma_i)+(\tilde\gamma-\gamma-\zeta)\Delta\}$$

$$=\inf_{\theta_i}\left\{\theta_i^2\left[(\tilde\gamma-\zeta)A_i+\delta\frac{1}{i^{1+\beta}}\right]+\theta_i 2B_i(\tilde\gamma-\eta)+(\tilde\gamma-\gamma-\zeta)\Delta_i\right\}$$

$$=(\tilde\gamma-\gamma-\zeta)\Delta_i-\frac{(\tilde\gamma-\zeta)^2B_i^2}{(\tilde\gamma-\zeta)A_i+\delta\frac{1}{i^{1+\beta}}}$$

$$\geq\left[(\tilde\gamma-\gamma-\zeta)-\frac{(\tilde\gamma-\zeta)^2A_i}{(\tilde\gamma-\zeta)A_i+\delta\frac{1}{i^{1+\beta}}}\right]\Delta_i. \tag{2.20}$$

The expression in (2.16) is nonnegative if (2.20) is. To show the latter, observe that if $\alpha>\beta$,

$$A_i\leq\int_{i-1}^{\infty}\frac{1}{x^{2+\alpha}}\,dx\leq\frac{1}{i^{1+\beta}}\left(\frac{i}{i-1}\right)^{1+\beta}\frac{1}{1+\alpha}\leq\frac{c}{i^{1+\beta}},\quad\text{for } i\geq 2.$$

Consequently,

$$(\tilde\gamma-\gamma-\zeta)\geq\frac{(\tilde\gamma-\zeta)^2c}{(\tilde\gamma-\zeta)c+\delta}\geq\frac{(\tilde\gamma-\zeta)^2A_i}{(\tilde\gamma-\zeta)A_i+\delta\frac{1}{i^{1+\beta}}},$$

if we choose

$$\delta\geq\frac{\tilde\gamma\gamma}{\tilde\gamma-\gamma}c=\gamma\frac{1-\varepsilon}{\frac{1}{2}-\varepsilon}c. \tag{2.21}$$

To summarize, we first choose $\alpha>\beta>0$ and $0<\varepsilon<\frac{1}{2}$. Then we choose $\gamma,\delta>0$ small enough so that (2.8) holds for 2γ and 2δ. If necessary, we make γ even smaller so that (2.21) holds. Then (2.16) holds, and (2.15) reduces to

$$\inf_{\theta\in\ell^2}\frac{\sum_{|\theta_i-1|\geq\eta}\lambda_i}{\sum_{|\theta_i-1|<\eta}\lambda_i e^{\tilde\gamma(\Delta-\Delta_i)+\tilde\delta(\sigma-\sigma_i)}}e^{\zeta\Delta}>0. \tag{2.22}$$

2.8. The last step involves bounding above the exponents in the denominator of (2.22) by a term of the order of $\frac{1}{2}\zeta\Delta$. The idea is that either $\Delta-\Delta_i$ or $\sigma-\sigma_i$ are large only when θ_i is large, which is excluded since $|\theta_i-1|\leq\eta$. We shall use the following:

LEMMA 3. *Let Q be a symmetric positive semidefinite matrix Q. Then*

$$Q(x)-bQ(y)\geq-\frac{b}{1-b}Q(y-x), \tag{2.23}$$

where $Q(x) = x'Qx$ *is the quadratic form associated with* Q, *and* b *is any number between* 0 *and* 1.

PROOF. Just use the identity

$$Q(x) - bQ(y) + \frac{b}{1-b}Q(y-x) = \frac{1}{1-b}Q(x-by) \geq 0. \qquad \square$$

Obviously, (2.23) extends to continuous positive quadratic forms on ℓ^2, such as Δ defined in (2.7b). We can now prove:

LEMMA 4. *There exists a constant c such that*

$$\tilde{\gamma}(\Delta - \Delta_i) + \tilde{\delta}(\sigma - \sigma_i) \leq \tfrac{1}{2}\zeta\Delta + c \tag{2.24}$$

uniformly over all i for which $|\theta_i - 1| < \eta$.

PROOF. By (2.17), (2.18) and (2.19), when $|\theta_i - 1| < \eta$ we have

$$\tilde{\gamma}(\Delta - \Delta_i) + \tilde{\delta}(\sigma - \sigma_i) \leq c_1 + c_2\sqrt{\Delta_i}\,.$$

By (2.23) with $b = \frac{1}{2}$,

$$\Delta - \tfrac{1}{2}\Delta_i = \Delta(\theta) - \tfrac{1}{2}\Delta(\theta^i) \geq -\Delta(\theta - \theta^i) = -\theta_i^2 A_i,$$

implying that, when $|\theta_i - 1| < \eta$, $\Delta_i \leq 2\Delta + c_3$ or $(\sqrt{x+y} \leq \sqrt{x} + \sqrt{y})$,

$$\sqrt{\Delta_i} \leq c_4 + c_5\sqrt{\Delta}\,.$$

Thus

$$\tilde{\gamma}(\Delta - \Delta_i) + \tilde{\delta}(\sigma - \sigma_i) \leq c_6 + c_7\sqrt{\Delta}\,.$$

To conclude (2.24), just note that the function (in x) $\frac{1}{2}\zeta x - c_7\sqrt{x}$ is bounded below for $x \geq 0$. $\square$

2.9. In view of Lemma 4, (2.22) reduces to

$$\inf_{\theta \in \ell^2} \frac{\sum_{|\theta_i - 1| \geq \eta}\lambda_i}{\sum_{|\theta_i - 1| < \eta}\lambda_i} e^{\zeta\Delta/2} > 0$$

or, even simpler,

$$\inf_{\theta \in \ell^2} e^{\zeta\Delta(\theta)/2} \sum_{|\theta_i - 1| \geq \eta} \lambda_i > 0. \tag{2.25}$$

Let n be the first i such that $|\theta_i - 1| \geq \eta$. Note that $n < \infty$ since $\theta \in \ell^2$. Then

$$\begin{aligned}
\Delta(\theta) &\geq \sum_{j=1}^{n-1} \frac{1}{j^{2+\alpha}}(\theta_1 + \cdots + \theta_j)^2 \\
&\geq \sum_{j=1}^{n-1} \frac{1}{j^{2+\alpha}} j^2 (1-\eta)^2 \\
&= (1-\eta)^2 \sum_{j=1}^{n-1} \frac{1}{j^{\alpha}}\,.
\end{aligned}$$

For large n, the sum $\Sigma_{j=1}^{n-1} 1/j^{\alpha}$ behaves like $n^{1-\alpha}$, so (2.25) will follow from

$$\inf_n \left(\lambda_n e^{cn^{1-\alpha}}\right) > 0, \quad \text{for } c > 0. \tag{2.26}$$

Finally, note that $e^{cn^{1-\alpha}}$ grows faster than any polynomial. In particular, for the weights $\{\lambda_i\}$ in (2.1),

$$\lim_{n\to\infty} \lambda_n e^{cn^{1-\alpha}} = \infty.$$

Thus, (2.26) holds for polynomially decreasing weights and the proof of (2.3) is now complete.

3. Admissibility of 1 under exponentially decreasing weights.

3.1. We now show that if the weights are given by

$$\lambda_i = e^{-ai}, \quad \text{where } a > \tfrac{1}{2}, \tag{3.1}$$

then the estimator 1 is admissible. The proof uses Blyth's method in a way similar to the proof of Theorem 3 in [10]. Let G be a finite measure on ℓ^2. The following notation will be used: δ^G is the Bayes estimator with respect to G; $R(G, \delta)$ is the risk function $R(\theta, \delta)$ of an estimator δ integrated against G. The admissibility of 1 will follow if we exhibit a sequence G^n of finite measures on ℓ^2 such that:

3.1.A. The Bayes risk $R(G^n, \delta^{G^n})$ is finite, for $n = 1, 2, \dots$.

3.1.B. The G^n measure of the point set $\{0\}$ is greater than 1, for $n = 1, 2, \dots$.

3.1.C. The nonnegative sequence $\Delta_n = R(G^n, 1) - R(G^n, \delta^{G^n})$ converges to 0, as $n \to \infty$.

3.2. Define G^n on ℓ^2 by

$$G^n(d\theta) = I_{\{0\}}(d\theta) + c_n \prod_{i=1}^{n} I_{\{1\}}(d\theta_i) \prod_{i=n+1}^{\infty} I_{\{0\}}(d\theta_i),$$

where $I_{\{0\}}, I_{\{1\}}$ denote probability measures (on the appropriate spaces) which assign unit mass to the points 0, 1, respectively, and the sequence of constants $\{c_n\}$ will be chosen to satisfy 3.1.A–C. Note that $G^n(\{0\}) = 1$ and $G^n(\ell^2) = 1 + c_n$. After standard calculations (as in [10], page 1460) we get that the Bayes estimator δ^{G^n} is

$$\begin{aligned} \delta_i^{G^n} &= 1 - \frac{1}{g^n}, \qquad i \le n, \\ &= 0, \qquad i \ge n + 1, \end{aligned}$$

where

$$g^n(x) = 1 + c_n \exp\left[\sum_{i=1}^{n} x_i - \frac{n}{2}\right]$$

and

$$\Delta_n = E_0 \frac{1}{g^n} \sum_{i=1}^{n} \lambda_i + (1 + c_n) \sum_{i=n+1}^{\infty} \lambda_i.$$

For 3.1.A–C to hold, it suffices to choose $\{c_n\}$ so that

3.2.A. $$\lim_{n\to\infty} c_n \sum_{i=n+1}^{\infty} \lambda_i = 0,$$

3.2.B. $$\lim_{n\to\infty} c_n \exp\left[\sum_{i=1}^{n} x_i - \frac{n}{2}\right] = \infty \quad \text{a.s.},$$

where $x_1, x_2, \ldots$ are iid $N(0,1)$. Now choose

(3.2) $$c_n = e^{bn}, \quad \text{where } \tfrac{1}{2} < b < a.$$

By the strong law of large numbers, $\lim_{n\to\infty}(1/n)\sum_{i=1}^{n} x_i = 0$ a.s., hence

$$\left[\sum_{i=1}^{n} x_i - \frac{n}{2} + b \cdot n\right] = n\left[\frac{1}{n}\sum_{i=1}^{n} x_i + \left(b - \frac{1}{2}\right)\right] \to \infty, \quad \text{as } n \to \infty.$$

We conclude that $\{c_n\}$ as in (3.2) can be chosen to satisfy 3.2.A–B for $\{\lambda_i\}$ as in (3.1), implying the admissibility of 1.

4. Concluding remarks and open questions.

4.1. We have considered apparently one of the simplest estimators in the countable model. The analysis turned out to be rather delicate. The outcomes clearly indicate that the infinite model is very different from the finite one and much more work is needed to better understand these differences. In particular, it is interesting to know if results for the infinite model would lead to *useful* asymptotic results for the finite model. To this end, more interesting estimators than 1 must be considered. An important such estimator (important on its own merit) is the linear estimator $\delta(x) = \alpha x$ for $0 < \alpha < 1$. As with the estimator 1, individual coordinates of αx are proper Bayes estimators. The same procedure used in Section 3 can be used to prove that αx is admissible if the weights decrease to 0 sufficiently fast. However, the inadmissibility of αx for slowly decreasing weights is still an unresolved problem.

4.2. Our results illustrate the sensitivity of the infinite model to the weights being used. A mathematically intriguing question is still open.

We have shown that with weights $\lambda_i = e^{-ai}$, $a > \frac{1}{2}$, the estimator 1 is admissible. We have not been able to determine what holds when $0 < a \leq \frac{1}{2}$. Arguments for both admissibility and inadmissibility can be given. However, we are still not sure what the true answer is.

4.3. Y. Vardi has asked whether it is possible to dominate 1 by an estimator which takes values in ℓ^2. More generally, the question whether the ℓ^2-valued estimators constitute a complete class remains open.

4.4. It is clear from our proofs that only the tail behavior of the sequence of weights is relevant. Thus, 1.3.B still holds for λ_i with $\lambda_i \cdot i^{1+c}$ converging to a positive number as i increases to infinity. Similarly, 1.3.A holds if $\lambda_i e^{ia}$, $a > \frac{1}{2}$, converges to a positive number.

5. Appendix. Decision theoretic models for Gaussian distributions on a Hilbert space.

5.1. Let H be a separable Hilbert space with inner product $(\cdot, \cdot)$ and norm $|\cdot|$. Denote by $N(\theta, C)$ the Gaussian measure on the Borel sets of H with mean θ and covariance operator C. The operator C: $H \to H$ must be linear compact operator which is positive semidefinite, self-adjoint and trace class. The support of the measure $N(\theta, C)$ is all H if and only if C is injective [7]. Assuming C is injective, the subspace $\Theta = C^{1/2}(H)$, as a range of the strictly positive operator $C^{1/2}$, is dense in H. Moreover, the measure $N(\theta, C)$ is equivalent to $N(0, C)$ if and only if $\theta \in \Theta$. One can equip Θ with a Hilbert space structure via the inner product

$$\langle \theta, \eta \rangle = \left(C^{-1/2}\theta, C^{-1/2}\eta\right), \qquad \theta, \eta \in \Theta.$$

Denote the corresponding norm in Θ by $\|\theta\| = |C^{-1/2}\theta|$, $\theta \in \Theta$. The subspace Θ can alternatively be described as

$$\Theta = \left\{x \in H \colon \sum_i \frac{1}{\gamma_i}(x, e_i)^2 < \infty\right\}, \tag{5.1}$$

where $\{e_i\}$ is the orthonormal basis of H consisting of eigenvalues of C and $\{\gamma_i\}$ are the corresponding positive eigenvalues. Note that

$$\text{trace}(C) = \sum_i \gamma_i < \infty,$$

since C is trace class. Thus, the faster γ_i converge to 0 the smaller Θ is.

5.2. Consider the problem of estimating the mean θ (C known) from a single observation $x \in H$ under two possible loss functions

$$L_H(\theta, a) = |\theta - a|^2, \qquad \theta \in \Theta,\ a \in H,$$

and

$$L_\Theta(\theta, a) = \|\theta - a\|^2, \qquad \theta \in \Theta,\ a \in \Theta.$$

The work in [10] focuses on the loss L_Θ while L_H is mentioned there only briefly. The estimation models under both losses are statistically isomorphic to an infinite model, namely estimating the mean sequence $(\theta_1, \theta_2, \dots) \in \ell^2$ from an observation $(x_1, x_2, \dots)$ under a loss

$$L(\theta, a) = \sum_i \lambda_i (\theta_i - a_i)^2,$$

where x_i are independent normal random variables with variance one and θ_i is the mean of x_i, $i = 1, 2, \dots$. We shall demonstrate below that the model with L_Θ

gives rise to an infinite model with $\lambda_i = 1$, while L_H reduces to $\lambda_i = \gamma_i$, the eigenvalues of C.

There is a seemingly minor difference between the two infinite models, namely, the tail behavior of the associated weights. This difference is, however, a reflection of a deep qualitative difference between the original models which is well illustrated by the important special case of the Brownian motion.

5.3. Let $H = L^2[0,1]$ be the space of functions on $[0,1]$ which are square integrable with respect to the Lebesgue measure. The standard Wiener measure on $L^2 = L^2[0,1]$ is the Gaussian measure with mean 0 and covariance C given by

$$Cx(t) = \int_0^1 (s \wedge t)x(s)\, ds, \qquad 0 \le t \le 1,\ x \in L^2.$$

The Wiener measure with mean $\theta \in L^2$ is a translate of the standard Wiener measure by θ. The parameter space Θ turns out to consist of absolutely continuous functions θ on $[0,1]$ with $\theta(0) = 0$ and square integrable derivatives θ'. The inner products from subsection 5.1 are

$$(x, y) = \int_0^1 x(t)y(t)\, dt, \qquad x, y \in L^2,$$

and

$$\langle \theta, \eta \rangle = \int_0^1 \theta'(t)\eta'(t)\, dt, \qquad \theta, \eta \in \Theta,$$

while

$$L_H(\theta, a) = \int_0^1 [\theta(t) - a(t)]^2\, dt \tag{5.2}$$

and

$$L_\Theta(\theta, a) = \int_0^1 [\theta'(t) - a'(t)]^2\, dt. \tag{5.3}$$

The eigenfunctions of C with their corresponding eigenvalues are given by

$$e_i(t) = \sqrt{2}\sin\left[\left(i - \tfrac{1}{2}\right)\pi t\right], \qquad \gamma_i = \left[\left(i - \tfrac{1}{2}\right)\pi\right]^{-2}, \tag{5.4}$$

for $i = 1, 2, \dots$. To appreciate how different is the model that uses (5.2) from that with (5.3), note that the "natural" estimator $\delta(x) = x$ is not even permissible under (5.3) because an observation from the Wiener measure (equivalently, a sample path of the Brownian motion) is almost surely nowhere differentiable.

5.4. We now reduce the model with loss L_θ to an infinite model with weights $\lambda_i = 1$ using the concept of a *white noise*. Let V be a subspace of H. A white noise over V is a collection of mean 0 random variables $\{W_h,\ h \in V\}$ such that

$$EW_hW_k = (h, k), \qquad h, k \in V,$$

and every finite subcollection has a multivariate normal distribution. In particular, each W_h is normal with mean 0 and variance $|h|^2$. Also, the mapping $h \to W_h$ is an isometry from H into the space of square integrable random variables over the probability space on which the white noise is defined. When V is dense in H, one can construct a unique white noise over H from a white noise over V by extending the isometry $h \to W_h$ from V to all H. The special white noise that will be used here is the white noise over H, defined over the dense subspace Θ by

$$W_h(x) = (x, C^{-1/2}h), \qquad h \in \Theta,\ x \in h.$$

Here W_h is thought of as a random variable on the probability space (H, Borel sets, $N(0, C)$). We write symbolically $(C^{-1/2}x, h)$ for the random variable W_h, $h \in H$.

Now let $\{v_i\}$ be an *arbitrary* orthonormal basis of H. Define a measurable mapping $\mathscr{D}$: $H \to R^\infty$, the space of sequences equipped with the Kolmogorov σ-field, by

$$\mathscr{D}x = (x_1, x_2, \dots), \qquad x \in H,$$

where

$$x_i = \left(c^{-1/2}x, v_i\right), \qquad i = 1, 2, \dots .$$

Define an isometry D from H onto ℓ^2 by

$$D\theta = (\theta_1, \theta_2, \dots), \qquad \theta \in \Theta,$$

where

$$\theta_i = \left(c^{-1/2}\theta, v_i\right), \qquad i = 1, 2, \dots .$$

Under $N(\theta, C)$, the sequence $(x_1, x_2, \dots)$ consists of independent normal random variables with variance one and a corresponding sequence of means $(\theta_1, \theta_2, \dots)$. The loss function L_Θ can be expressed as

$$L_\Theta(\theta, a) = \sum_i (\theta_i - a_i)^2, \tag{5.5}$$

where $(a_1, a_2, \dots)$ equals Da. Finally, there is a one to one correspondence between estimators $(\delta_1(x_1, x_2, \dots), \delta_2(x_1, x_2, \dots), \dots)$ given by

$$\delta(x) = \sum_i \delta_i(\mathscr{D}x) C^{1/2} v_i,$$

where

$$\delta_i(\mathscr{D}x) = \left(C^{-1/2}\delta(x), v_i\right). \tag{5.6}$$

The relation (5.5) extends in an obvious manner to a relation between risk functions of estimators. Moreover, δ is admissible under L_Θ if and only if $(\delta_1, \delta_2, \dots)$ is admissible in the infinite model with weights $\lambda_i \equiv 1$, completing the description of the statistical isomorphism. Note that the basis $\{v_i\}$ is arbitrary and every basis leads to the same infinite model. The situation is not as simple under L_H.

5.5. We now reduce the model with loss L_H to an infinite model with weights $\lambda_i = \gamma_i$. Here we choose the specific orthonormal basis $\{e_i\}$ consisting of the eigenvectors of C. The statistical isomorphism is completely described by the relations

$$\theta_i = \frac{1}{\sqrt{\gamma_i}}(\theta, e_i), \qquad x_i = \frac{1}{\sqrt{\gamma_i}}(x, e_i), \qquad i = 1, 2, \dots,$$

$$L_\Theta(\theta, a) = \sum_i \gamma_i(\theta_i - a_i)^2,$$

and

$$\delta(x) = \sum_i \sqrt{\gamma_i}\,\delta_i(x_1, x_2, \dots)e_i,$$

where

$$\delta_i(x_1, x_2, \dots) = \frac{1}{\sqrt{\gamma_i}}(\delta(x), e_i). \tag{5.7}$$

Note that, by (5.1), $\delta(x)$ takes values in Θ if and only if $(\delta_1, \delta_2, \dots)$ takes values in ℓ^2. While (5.6) does not make sense if δ takes values outside Θ, (5.7) is well defined for all δ: $H \to H$.

5.6. The orthonormal basis (5.4) will now be used to illustrate the process described above on the Brownian motion model under both losses (5.2) and (5.3). The coordinates in the infinite model are

$$\theta_i = \int_0^1 \Phi_i(t)\theta'(t)\,dt, \qquad x_i = \int_0^1 \Phi_i(t)\,dx(t), \tag{5.8}$$

where $\Phi_i(t) = \sqrt{\gamma_i}\,e_i'(t)$ and the second integral in (5.8) is the usual Wiener integral. Moreover,

$$\left(\Phi_i(t), \Phi_j(t)\right) = \left\langle \sqrt{\gamma_i}\,e_i, \sqrt{\gamma_j}\,e_j \right\rangle = (e_i, e_j), \quad \text{for all } i, j.$$

It follows that $\{\Phi_i\}$ is an orthonormal basis of L^2. Thus, via the transformations (5.8), the problem of estimating the mean function of the Wiener process under the loss (5.2) reduces to the infinite model with weights λ_i that exhibit the tail behaviour of $1/i^2$. On the other hand, the loss (5.3) gives rise to $\lambda_i \equiv 1$.

5.7. A discussion the first author had with Iain Johnstone led to question whether the estimator 1 is admissible under the loss (5.2). The answer is still unknown, but the question led eventually to the present paper. Note that the estimator for the mean of the Wiener measure that corresponds to the estimator 1 in the infinite model is the constant estimator

$$\delta(x)_t = \sum_i \sqrt{\gamma_i}\,e_i(t) = \sum_i \frac{\sqrt{2}}{\left(i - \frac{1}{2}\right)\pi}\sin\left[\left(i - \tfrac{1}{2}\right)\pi t\right], \qquad 0 \le t \le 1.$$

REFERENCES

[1] BERGER, J. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* 4 223–226.

[2] BERGER, J. and WOLPERT, R. (1982). Incorporating prior information in minimax estimation of the mean of a Gaussian process. In *Statistical Decision Theory and Related Topics III* (S. Gupta and J. Berger, eds.) 2 451–464. Academic, New York.

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

[4] BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

[5] FARRELL, R. (1985). *Multivariate Calculation, Use of the Continuous Groups.* Springer, New York.

[6] GRENANDER, U. (1981). *Abstract Inference.* Wiley, New York.

[7] ITÔ, K. (1970). The topological support of a Gauss measure on Hilbert space. *Nagoya Math. J.* **38** 181–183.

[8] LEHMANN, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

[9] MAKANI, S. M. (1977). A paradox in admissibility. *Ann. Statist.* **5** 544–546.

[10] MANDELBAUM, A. (1984). All admissible linear estimators of the mean of a Gaussian distribution on a Hilbert space. *Ann. Statist.* **12** 1448–1466.

[11] SHEPP, L. A. (1965). Distinguishing a sequence of random variables from a translate of itself. *Ann. Math. Statist.* **36** 1107–1112.

[12] SHINOZAKI, N. (1975). A study of generalized inverse of matrix and estimation with quadratic loss. Ph.D. thesis, Keio University, Japan.

[13] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press.

GRADUATE SCHOOL OF BUSINESS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-5015

MATHEMATICAL SCIENCES RESEARCH CENTER
AT & T BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974