IMS Collections
Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman
Vol. 8 (2012) 117–130

DOI: 10.1214/11-IMSCOLL809

Asymptotic admissibility of priors and elliptic differential equations

John A. Hartigan

Department of Statistics, Yale University

Abstract: We evaluate priors by the second order asymptotic behaviour of the corresponding estimators. Under certain regularity conditions, the risk differences between efficient estimators of parameters taking values in a domain D, an open connected subset of R^d, are asymptotically expressed as elliptic differential forms depending on the asymptotic covariance matrix V. Each efficient estimator has the same asymptotic risk as a "local Bayes" estimate corresponding to a prior density p. The asymptotic decision theory of the estimators identifies the smooth prior densities as admissible or inadmissible, according to the existence of solutions to certain elliptic differential equations. The prior p is admissible if the quantity pV is sufficiently small near the boundary of D. We exhibit the unique admissible invariant prior for $V = I$, $D = R^d - \{0\}$. A detailed example is given for a normal mixture model.

Contents

1 Introduction 117
2 Risk matching priors 118
3 Explicit matching priors using the Feynman-Kac integral 120
4 Decision theory for asymptotic risks 121
5 When is pV Bayes on R^d? 122
6 When is V Bayes on bounded D? 124
7 One dimension 125
8 Invariant admissible priors 125
9 A mixture model 126
10 A prior beating the uniform 127
Acknowledgement 129
References 129

1. Introduction

A parameter x takes values in a *domain* D, an open connected subset of R^d. I use the partial differential equation symbol x rather than the usual statistical symbol θ because the evaluation of asymptotic risk reduces to existence problems in the theory of partial differential equations. The parameter indexes a probability density $p(y_n|x)$ with respect to some measure μ_n, say, for data y_n.

P.O. Box 208290 New Haven CT 06520-8290 USA

AMS 2000 subject classifications: Primary 62C20; secondary 62F10, 62P30

Keywords and phrases: Birge ratio, consensus value, jackknife estimator, matrix weighted means, meta-analysis, normal mean, shrinkage estimator

We use the Kullback-Leibler loss function

$$\hat{x}, x \in D : L_n(\hat{x}, x) = \int \log[p(y_n|x)/p(y_n|\hat{x})]p(y_n|x)d\mu_n(y_n) \tag{1.1}$$

to define the risk of the estimator $\hat{x}_n$, a function of y_n taking values in D:

$$R_n(\hat{x}_n, x) = \int L_n(\hat{x}_n(y_n), x)p(y_n|x)d\mu_n. \tag{1.2}$$

For a prior density p, the posterior Bayes estimator $\hat{x}(y_n, p)$ minimizes the posterior Kullback-Leibler risk

$$R(\hat{x}|y_n) = \int L(\hat{x}, x)p(y_n|x)p(x)dx / \int p(y_n|x)p(x)dx. \tag{1.3}$$

Define

$$V_n(x) = -1/\int \tfrac{\partial}{\partial x}\tfrac{\partial}{\partial x'} \log[p(y_n|x)]p(y_n|x)d\mu_n(y_n). \tag{1.4}$$

We will assume $nV_n(x) \to V(x)$, the *asymptotic covariance matrix.*

Following Brown [Br79], and letting U denote the prior uniform over D, we consider estimators of form $\hat{x}(y_n, U) + V_n b(\hat{x}(y_n, U))$ for fixed *decision functions* $b : D \to D$. Under smoothness conditions [Ha98] requiring smooth variation of the data density and the prior with x, the asymptotic risks for different decision functions differ only by terms of order n^{-2}; therefore we define the *asymptotic risk* for decision function b, relative to the decision function $b = 0$ corresponding to the uniform prior U, by the assumed limit

$$R(b, x) = \lim_{n\to\infty} n^2[R_n(b, x) - R_n(U, x)] = \sum\nolimits_{i,j} \{\partial_i(V_{ij}b_j) + \tfrac{1}{2}b_i b_j V_{ij}\} \tag{1.5}$$

where ∂_i denotes the partial derivative $\frac{\partial}{\partial x_i}$.

For a prior with density p, the posterior bayes estimate corresponds asymptotically to the decision function $b_i^p = \partial_i \log p$, and then the risk may be expressed in elliptic operator form

$$R(b^p) = 2\sum\nolimits_{ij} \partial_i(V_{ij}\partial_j\sqrt{p})/\sqrt{p}. \tag{1.6}$$

It turns out that, under certain conditions of smoothness and boundedness for b and V, there is a *risk matching* prior density p for which $R(b^p) = R(b)$. Thus the behaviour of asymptotic risk for all smooth decisions is captured in the theory of elliptic differential equations, equations whose relevance to decision theory were first indicated in Stein [St56], but which were extensively elucidated for the normal location problem in Brown [Br71]. See also Strawderman and Cohen [SC71]. The asymptotic behavior of Bayes estimators near maximum likelihood estimators have been studied for loss functions of form $L_n(\hat{x}, x) = W(\hat{x} - x)$ by Levit in [Le82, Le83, Le85]; in particular, he shows that the Bayes estimators form a complete class under certain regularity conditions.

2. Risk matching priors

For each decision function b we will find a risk matching prior p for which $R(b^p) = R(b)$. Then we need only consider decision functions of form $b_i^p = \partial_i \log p$ and risks of form $R(b^p) = 2\sum_{ij} \partial_i(V_{ij}\partial_j\sqrt{p})/\sqrt{p}$ in the asymptotic decision theory. This result will be proved under boundedness and smoothness assumptions using some standard tools from Pinsky [Pi95].

For the domain D with closure $\bar{D}$, a function f is uniformly Holder continuous with exponent $\alpha, 0 < \alpha \leq 1$, in $\bar{D}$ if

$$||f||_{0,\alpha,\bar{D}} = \sup_{x,y\in\bar{D},x\neq y} \frac{|f(x) - f(y)|}{|x-y|^\alpha} < \infty. \tag{2.1}$$

The Holder spaces $C^{k,\alpha}(\bar{D})$ consist of functions whose k-th order partial derivatives are uniformly Holder continuous with exponent α in $\bar{D}$. Say $D' \subset\subset D$ if D' is bounded and properly included in D. The Holder spaces $C^{k,\alpha}(D)$ consist of functions that lie in $C^{k,\alpha}(\bar{D}')$ for each $D' \subset\subset D$.

The domain D has a $C^{k,\alpha}$ boundary ∂D if for each point $x_0 \in \partial D$, there is a ball B centered at x_0 and a 1-1 mapping $\psi : B \to A \subset R^d$, such that $\psi(B \cap D) \subset \{x \in R^d, x_n > 0\}, \psi(B \cap \partial D) \subset \{x \in R^d, x_n = 0\}, \psi \in C^{k,\alpha}(B), \psi^{-1} \in C^{k,\alpha}(A)$.

Condition $\bar{A}$: D bounded, $\partial D \in C^{2,\alpha}, V_{ij} \in C^{1,\alpha}(\bar{D}), V$ positive definite in $\bar{D}$.
Condition A : $V_{ij} \in C^{1,\alpha}(D), V$ positive definite in D.

We follow a standard approach which first proves results under the strong condition $\bar{A}$, and then extends the results to the weak condition A by approximating D by an increasing sequence of bounded subdomains D_n.

Theorem 1. *Suppose $b_i = \partial_i b \in C^{1,\alpha}(\bar{D})$ and assumption $\bar{A}$ holds. Then there exists a prior $p, \sqrt{p} \in C^{2,\alpha}(\bar{D}), p > 0$ in D, such that*

$$R(b) = R(b^p) = 2\sum_{ij} \partial_i(V_{ij}\partial_j\sqrt{p})/\sqrt{p}. \tag{2.2}$$

Proof. From [Pi95, theorem 5.5], for some eigenvalue λ, there exists $u \in C^{2,\alpha}(\bar{D}), u > 0$ in $D, u = 0$ in ∂D, satisfying

$$2\sum_{ij} \partial_i(V_{ij}\partial_j u) - R(b)u = \lambda u. \tag{2.3}$$

Since

$$\int_D \lambda u^2 = \int_D \{2\sum_{ij} \partial_i(V_{ij}\partial_j u) - R(b)u\}u \tag{2.4}$$

$$= \int_D -\tfrac{1}{2}\textstyle\sum_{ij}(ub_i - 2\partial_i u)(ub_j - 2\partial_j u)V_{ij} \leq 0, \tag{2.5}$$

it follows that $\lambda \leq 0$. If $\lambda = 0$, the corresponding eigenvector u provides the b-*matching* prior $p = u^2$ with $b_i = \partial_i \log p$. If $\lambda < 0$, from [Pi95, Theorem 6.5], for each $\phi \in C^{2,\alpha}(\bar{D}), \phi > 0$ there exists a unique b-*matching* solution $\sqrt{p} = u \in C^{2,\alpha}(\bar{D}), u > 0$ in $D, u = \phi$ on ∂D, to the equation $R(b) = R(b^p)$. □

Theorem 2. *Suppose $b_i = \partial_i b \in C^{1,\alpha}(D)$ and assumption A holds.*

Then there exists a prior $p, \sqrt{p} \in C^{2,\alpha}(D), p > 0$ in D, such that

$$R(b) = R(b^p) = 2\sum_{ij} \partial_i(V_{ij}\partial_j\sqrt{p})/\sqrt{p}. \tag{2.6}$$

Proof. Select an increasing sequence of bounded domains $\bar{D}_n \subset\subset D_{n+1}, \cup D_n = D$. Within each domain, assumption $\bar{A}$ holds, so there exists a sequence of solutions $\sqrt{p_n} = u_n \in C^{2,\alpha}(\bar{D}_n), u_n > 0$ in D_n, such that $R(b) = R(b^{p_n}), x \in D_n$.

For some $x_0 \in D_1$, without loss of generality set $u_n(x_0) = 1$, all n. Note that any solution $u_N, N > n$ is also a solution to $R(b) = R(b^{p_n})$, $x \in D_n$. A Harnack inequality [Pi95, p. 124] implies that for $N > n, c_n \leq u_N(x) \leq C_n, x \in D_n$ for some constants c_n, C_n.

The Schauder interior estimate [Pi95, p. 86] implies, for some constant C_n, for all $N > n$,

$$(2.7) \qquad ||u_N||_{2,\alpha,D_n} = \sup_{\substack{x, y \in D_n \\ x \neq y}} \sum_{i,j} \frac{|\partial_i \partial_j (u_N(x) - u_N(y))|}{|x-y|^\alpha} \leq C_n.$$

Thus, for each n, the sequence u_N is precompact in the $||u_N||_{2,0,D_n}$ norm. By diagonalization, there exists a subsequence, say u_N, that converges to u in $C^{2,0}(D)$ for which

$$(2.8) \qquad R(b) = R(b^p) = 2\sum_{ij} \partial_i (V_{ij} \partial_j u)/u, x \in D.$$

Finally, we need to show that $u \in C^{2,\alpha}(D)$. First note that $u \in C^{2,\alpha}(D_n)$, since

$$(2.9) \qquad ||u_N||_{2,\alpha,D_n} \leq c_n \text{ for } N > n \Rightarrow ||u||_{2,\alpha,D_n} \leq c_n.$$

For any $D' \subset\subset D$, the compact set $\bar{D}'$ is covered by $\cup D_n = D$, and therefore by a finite subcovering, and so by a particular D_n. Thus $||u||_{2,\alpha,D'} \leq c_n(D')$ for every $D' \subset\subset D$ which implies $u \in C^{2,\alpha}(D)$. □

3. Explicit matching priors using the Feynman-Kac integral

When $\lambda = 0$ in equation (9), the decision function b is the posterior bayes decision corresponding to the prior $p : b_i = \partial_i \log p$. We could determine the ratio $p(x)/p(y)$ by the integral $\int_\rho \sum_i b_i dx_i$ over *any* path ρ connecting x and y.

When $\lambda < 0$, so that b is no longer a gradient, it is plausible to attempt to find an approximating p by averaging these integrals over *all* paths between x and y. Consider the stochastic differential equation with initial condition $X_i(0) = x$:

$$(3.1) \qquad dX_i(t) = \sum_{ij} V_{ij}^{1/2} dW_j(t) + \tfrac{1}{2}(b_i - \textstyle\sum_j \partial_j V_{ij})dt$$

We propose the Feynman Kac integral formula to specify a risk matching prior:

$$(3.2) \qquad u(x) = E[\exp(-\tfrac{1}{2} \textstyle\int_0^T \sum_i b_i \circ dX_i) \sqrt{p(X(T))}]$$

where $\int_0^T \sum_i b_i \circ dX_i$ is the Stratonovitch stochastic integral, and T is the time to reach the boundary of D. I suspect that the condition $\lambda < 0$ is sufficient for the existence of the stochastic process and the integral when assumption $\bar{A}$ holds. When the formula is valid, we see that $u(x)$ is determined as a weighted combination of its boundary values, with the weight at each boundary point determined by the path integral $\exp(-\frac{1}{2} \int_\rho \sum_i b_i dx_i)$ over the various paths that reach that particular boundary point. Many different priors risk matching b are available corresponding to different smooth assignments to the boundary values.

4. Decision theory for asymptotic risks

We now apply decision theoretic classifications to the asymptotic risk formula. Our conclusions about the prior p will depend only on the domain D and the asymptotic variance V. From now on we will drop the term asymptotic. We consider a particular variance V and the set of risks, real valued functions on the domain D, corresponding to priors satisfying

$$\text{(4.1)} \qquad \text{Assumption } B : p \in C^{2,\alpha}(D), V_{ij} \in C^{1,\alpha}(D); p, V > 0 \text{ in } D.$$

The risks will be written $R(p) = R(b^p)$.

The posterior bayes decision b^p is *locally Bayes*: for any alternative decision $b^p + v$ where $v \in C^{1,\alpha}(D)$ and $v = 0$ in $D - D'$, some $D' \subset\subset D$, integration by parts shows

$$\text{(4.2)} \qquad \int_D [R(b^p + v) - R(b^p)]p \geq 0.$$

Theorem 3. *Under assumption B, with D bounded, the following conditions are equivalent:*

b^p *is Bayes: there exists no* $p^* \neq p$ *with* $\int (R(p^*) - R(p))p \leq 0$.

b^p *is Admissible: there exists no* $p^* \neq p$ *with* $R(p^*) \leq R(p), x \in D$.

b^p *is Unique Risk: there exists no* $p^* \neq p$ *with* $R(p^*) = R(p), x \in D$.

b^p *is Brown: no non-trivial positive* h *solves* $\sum_{ij} \partial_i (pV_{ij} \partial_j h) = 0$.

Proof. *Bayes* implies *Admissible* because $u* \neq u$ violating *Admissible* also violates *Bayes*. *Admissible* implies *Unique Risk* because $u* \neq u$ violating *Unique Risk* also violates *Admissible*. *Brown [Br71, 1.3.9]* and *Unique Risk* are equivalent, because $R(p\sqrt{h}) = R(p)$ if and only if $\sum_{ij} \partial_i (pV_{ij} \partial_j h) = 0$.

It only remains to show that failure of *Bayes* implies failure of *Unique Risk*.

Without loss of generality, assume $p = U$ is uniform, so that $R(p) = 0$. If p is not *Bayes*, there exists $p^* \neq p$ with

$$\text{(4.3)} \qquad 0 \geq \int R(p^*) = \tfrac{1}{2} \int_D \sum_{ij} V_{ij} b_i^* b_j^* + \int_D \sum_{ij} \partial_i (V_{ij} b_j^*)$$

Since $p^* \neq p$ and the middle integral is positive, then

$$\text{(4.4)} \qquad \int_D \sum_{ij} \partial_i (V_{ij} b_j^*) = C < 0.$$

Let $|\partial D|$ be the lebesgue measure of the boundary ∂D when D is smooth and bounded, let τ denote the outward pointing normals on the boundary, and note that

$$\text{(4.5)} \qquad \int_{\partial D} \sum_{ij} \{\tau_i b_i^* V_{ij}\} = \int_D \sum_{ij} \partial_i (V_{ij} b_j^*) = C < 0.$$

Applying Theorem 6.31 from [GT97], for $D_n \subset\subset D$, since $C < 0, \sum_{ij} \{\tau_i \tau_j V_{ij}\} > 0$, there exists a solution $p_n = u_n^2 \in C^{2,\alpha}(\bar{D}_n)$ to the oblique derivative problem

$$\text{(4.6)} \qquad R(p_n) = 0 \text{ in } D_n, u_n = 2 * C \sum_{ij} \{\tau_i \partial_i u V_{ij}\} / |\partial D_n| \text{ in } \partial D_n$$

so that $\int_D \sum_{ij} \partial_i (V_{ij} \partial_i p_n / p_n) = C < 0$.

Repeating the compactness argument of theorem 2 on $D_n \subset\subset D_{n+1}, \cup D_n = D$, there exists $p_0 \in C^{2,\alpha}(D)$ with

$$R(p_0) = 0, \int_D \sum\nolimits_{ij} \partial_i(V_{ij}\partial_i p_0/p_0) = C < 0. \tag{4.7}$$

The first condition states that p_0 and $p = U$ have the same risk, and the second condition guarantees that $p_0 \neq p = U$, so that the unique risk condition fails, as required. □

5. When is pV Bayes on $\mathbf{R}^d$?

Brown's condition shows that the locally Bayes estimate b^p is Bayes or not depending only on the product pV. For example, b^p is Bayes with V, if and only if b^U is Bayes with pV. We will therefore rephrase the admissibility question in terms of the product pV: the prior-scaled covariance matrix pV is Bayes on D if and only if there is no non-trivial solution to Brown's equation.

Theorem 4. *Let $D = R^d$, $pV \in C^{1,\alpha}(D), r = |x|, x = rs$ where s ranges over the surface S of the unit sphere. Define $W(R,s) = [\int_1^R \frac{1}{p}V^{-1}r^{1-d}dr]^{-1}$. Suppose that, uniformly over $s \in S$,*

$$\lim_{R\to\infty} W(R,s) = W(s), \lim_{R\to\infty} W(s)W^{-1}(R,s)W(s) = W(s). \tag{5.1}$$

Then pV is Bayes on R^d only if

$$\int_{s\in S} \sum_{ij} s_i s_j W_{ij}(s)ds = 0. \tag{5.2}$$

Proof. Let $Q = \{Q_i, 1 \leq i \leq d\} \in C^{1,\alpha}(D)$ be a *test* function with relative risk

$$R(b^p + Q) - R(b) = \sum_{ij} \{\partial_i(pV_{ij}Q_j) + \tfrac{1}{2}Q_iQ_jpV_{ij}\}. \tag{5.3}$$

From theorem 2 for every Q there exists a prior q with $R(b^p + Q) = R(b^q)$. Thus pV is Bayes if and only if

$$\int_D \{R(b^p + Q) - R(b^p)\}p = \int_D \sum_{ij} \{\partial_i(pV_{ij}Q_j) + \tfrac{1}{2}Q_iQ_jpV_{ij}\} \geq 0 \tag{5.4}$$

for every *test* Q where the integral is defined. Equivalently, with the test $(pV)^{-1}Q$,

(5.5)
$$t(Q) = \int_D \{R(b^p + (pV)^{-1}Q) - R(b^p)\}p = \int_D \{\sum_i \partial_i Q_i + \tfrac{1}{2}\sum\nolimits_{ij} Q_iQ_jV_{ij}^{-1}/p\} \geq 0$$

for every test Q where the integral is defined.

The possible negative term $\int_D \sum_i \partial_i Q_i$ is determined by values in the neighbourhood of infinity, so pV being Bayes is determined by the behaviour of pV near the infinite boundary. In particular if two functions pV are identical outside a compact subset of D, they have the same admissibility classification.

We therefore consider a test function that is zero inside the unit sphere:

$$Q(rs) = g(r)r^{1-d}q(s), q(s) = -Ws \tag{5.6}$$

where g is twice differentiable, $g(r) = 0$ for $0 \leq r \leq 1, 0 < g(r) \leq 1$ for $1 < r < 2, g(r) = 1$ for $r \geq 2$.

(5.7)

Let $t(Q,R) = \int_{|x|<R} \{\sum_i \partial_i Q_i + \frac{1}{2}\sum_{ij} Q_i Q_j V_{ij}^{-1}/p\} r^{d-1} dr ds = \int_{s\in S} I(R,s)ds.$

Consider the contribution $I(s,R)$ to the test integral for a particular s:

$$I(s,R) = \sum_i s_i Q_i(Rs)R^{d-1} + \int_0^R \{\tfrac{1}{2}\sum_{ij} Q_i Q_j V_{ij}^{-1}/p\} r^{d-1} dr \tag{5.8}$$

$$= \sum_i s_i q_i(s) + \frac{1}{2}\sum_{ij} q_i q_j \int_0^R g(r)^2 r^{1-d} V_{ij}^{-1}/p dr \tag{5.9}$$

$$\leq -\sum_{ij} s_i s_j W_{ij} + \frac{1}{2}\sum_{ijkl} s_i s_j W_{ik} W_{jl} W_{kl}^{-1}(R,s) \tag{5.10}$$

$$\to -\frac{1}{2}\sum_{ij} s_i s_j W_{ij} \text{ uniformly in } s \text{ as } R \to \infty \tag{5.11}$$

Thus

$$t(Q) = \lim_{R\to\infty} \int_{s\in S} I(s,R)ds < 0 \tag{5.12}$$

unless

$$\int_{s\in S} \sum_{ij} s_i s_j W_{ij}(s) ds = 0. \tag{5.13}$$

which shows that the condition in the theorem is necessary for p to be Bayes. □

Failure of the condition in the theorem allows construction of an explicit test function for showing p to be not Bayes. I suspect that the weaker condition $\lim_{R\to\infty}\{\sum_{ij} s_i s_j [\int_1^R \frac{1}{p} V^{-1} r^{1-d} dr]^{-1} ds = 0$ is also necessary. It may be that the condition is also sufficient. A similar condition for the recurrence of diffusion processes is given in [Ic78].

Brown [Br71] studies the admissibility of estimates for the normal location problem in d dimensions in which it is assumed that the data x are gaussian with unknown mean and identity covariance matrix. He shows that an estimate corresponding to the marginal density of the data $p(x)$ is admissible if $p_i = \partial_i \log p = \frac{\partial}{\partial x_i} \log p$ is bounded and if

$$\int_1^\infty [\int_{s\in S} p ds]^{-1} r^{1-d} dr = \infty. \tag{5.14}$$

Brown, Theorem 6.4.4, also shows that an estimate corresponding to the marginal density $p(x)$ is admissible only if

$$\int_{s\in S} [\int_1^\infty p^{-1} r^{1-d} dr]^{-1} ds = 0. \tag{5.15}$$

The asymptotic version requires data $X_n \sim N(\theta, I/n)$, with $n \to \infty$. Theorem 4 implies (40): a prior p is Bayes only if

$$(5.16) \qquad \int_{s \in S} [\int_1^\infty p^{-1} r^{1-d} dr]^{-1} ds = 0$$

or equivalently, almost everywhere on S

$$(5.17) \qquad \int_1^\infty p^{-1} r^{1-d} dr = \infty.$$

If the prior density p is expressed as a density ρ on the polar co-ordinates $x = rs$, the condition simplifies to $\int_1^\infty \rho^{-1}(r,s) dr = \infty$ almost everywhere on S. See Strawderman and Cohen [SC71], theorem 4.4.1. For example, the prior corresponding to r^α being uniformly distributed is Bayes in every dimension for $\alpha \leq 2$ but not Bayes for $\alpha > 2$.

6. When is V Bayes on bounded D?

Let D be a bounded domain with boundary in $C^{2,\alpha}$. Let $\nu(s), s \in \partial D$ denote the outward pointing normal at a point s on the boundary of D, assumed defined almost everywhere in ds, lebesgue measure on the boundary. It will be assumed that, for almost all $s \in \partial D$, the inward pointing normal$\{s - u\nu(s), 0 < u < \varepsilon\}$ lies in D for ε small enough.

Theorem 5. *The covariance matrix pV is Bayes only if*

$$(6.1) \qquad \lim_{\varepsilon \to 0} \int_{s \in \partial D} \nu_i \nu_j [\int_0^\varepsilon \frac{1}{p} V_{ij}^{-1}(s - u\nu) du]^{-1} ds = 0.$$

This is proved similarly to theorem 4, using test functions that are constant on the inward pointing normal segments.

It may happen that, for each s, the normal vector $\nu(s)$ at s is the limit of some eigenvector of pV as $x \in D \to$s$\in \partial D$, (the *normal eigenvector* case) in which case the condition in the theorem simplifies to

$$(6.2) \qquad \int_0^\varepsilon \frac{\nu_i \nu_j}{p} V_{ij}^{-1}(s - u\nu) du = \infty \text{ almost all } s.$$

We will say that the integral condition *fails* at s if $\int_0^\varepsilon \frac{\nu_i\nu_j}{p} V_{ij}^{-1}(s - u\nu) du < \infty$. The theorem now states that pV is admissible only if the integral condition fails on a set of measure 0.

The admissible pV are those where $pV \to 0$ fast enough near the boundary. If pV is inadmissible, we can render it admissible by attenuating p near the boundary. In the *normal eigenvector* case, let D_ε consist of those points $x \in D$ within ε of the boundary, and suppose that each such point is closest to a unique boundary point $s(x)$. Each such point may be written $x = s(x) - u(x)\nu(x)$ for some u.

Let a be an *attenuation* factor defined at each point in D by:

$$\begin{aligned}
a(x) &= 1 \text{ for } x \in D - D_\varepsilon, \\
a(x) &= 1 \text{ for } x \in D_\varepsilon, \text{ integral condition holds at } x(s), \\
g(x) &= \frac{\partial}{\partial u} (\int_0^{u(x)} [\nu_i \nu_j \frac{1}{p} V_{ij}^{-1}(s(x) - w\nu(x))]^{1/2} dw)^2, \\
a(x) &= min[1, 1 - (1 - \frac{g(x)}{g(\varepsilon})^3] \text{ for } x \in D_\varepsilon, \text{ integral condition fails at } x(s).
\end{aligned}$$

The proposed attenuation factor will be 1 except near boundary points s where the integral condition fails, where it will approach zero. With the prior ap, the integral condition becomes

(6.3)
$$\int_0^1 \nu_i\nu_j \frac{1}{ap} V_{ij}^{-1}(s-u\nu)du \geq \frac{1}{6}g(\varepsilon)\int_0^{\varepsilon} \frac{\partial}{\partial u}[log(\int_0^u [\nu_i\nu_j \frac{1}{p} V_{ij}^{-1}(s-w\nu)dw]du = \infty$$

7. One dimension

For the one dimensional parameter x on $D = (a,b)$ with variance V, Brown's condition implies that pV is admissible if and only if

$$\int_a (Vp)^{-1} = \int^b (Vp)^{-1} = \infty. \tag{7.1}$$

Since a smooth monotone transformation renders pV equal to 1 on $D = (a,b)$, an equivalent result is that there exists a non-zero differentiable test function w on D such that $\int_a^b (w' + \frac{1}{2}w^2) \leq 0$ if and only if either a or b are finite.

Jeffreys' density $J = V^{-\frac{1}{2}}$ is admissible on $D = (a,b)$ if and only if

$$\int_a J = \int^b J = \infty, \tag{7.2}$$

which means that Jeffreys must be "improper" in both tails to be admissible. I take a certain delight in this impropriety, because although "improper" priors abound in decision theory and in Bayesian analysis, they remain objects of suspicion. See for example the excellent review in [KW96]. However, in decision theory, the prior appears only when multiplied with a loss function, which may be arbitrarily scaled, so improper priors form a natural part of the range of procedures we need to study. Asymptotically, the prior appears only as a product with the covariance matrix in admissibility questions, and again it makes no sense to constrain priors to be improper. In the Jeffreys' case, the admissibility of the product pV requires that Jeffreys be improper in the tails.

The pearson correlation coefficient computed for n bivariate normal observations with true correlation ρ has asymptotic variance $1/(1-\rho^2)^2$.

Thus a prior p on $D = (-1,1)$ is admissible if and only if

$$\int_{-1} ((1-\rho^2)^2 p)^{-1} = \int^1 ((1-\rho^2)^2 p)^{-1} = \infty. \tag{7.3}$$

For priors of form $p = 1/(1-\rho^2)^{\alpha}$, the prior p is admissible if and only if $\alpha \leq 1$. Thus if we wish to skirt the edge of inadmissibility, we might use $p = 1/(1-\rho^2)$.

8. Invariant admissible priors

Since the Kullback-Leibler loss function does not change under smooth transformation of the parameter space, differences between the asymptotic risk functions for two priors are also invariant under such transformations. We are free to transform to a convenient p, V, D in deciding admissibility problems. If a transformation T takes p, V, D into say $T(p), T(V), T(D)$, then the admissibility of PV in D equals the admissibility of $T(p)V(p)$ in $T(D)$. A prior p is *relatively invariant* if

$p(Tx)J = cp(x), x \in D$, where J is the Jacobian of the the transformation $x \to Tx$ and when T is one to one $D \to D$ such that $TV(Tx) = CV(x)$.

For example if $D = R^d - 0, V = I$, arbitrary rotations and scalings leave D invariant, and change the covariance by a constant, so the only invariant priors are of form $p = r^\alpha, r^2 = \sum_i x_i^2$. From Brown's condition, the prior p is admissible if

$$\int_0 r^{1-d-\alpha} dr = \int^\infty r^{1-d-\alpha} = \infty, \tag{8.1}$$

which occurs only when $\alpha = d-2$. In this case there is a single admissible invariant prior $p = r^{d-2}$. This prior, discussed in [Br71] and [SC71], corresponds to r^2 being uniform over D.

If $D = \{x | R_1 < r = |x| < R_2\}$, the invariant transformations are rotations, which require that an invariant p depends only on r. Admissibility requires

$$\int_{R_1} \frac{1}{p} dr = \int^{R_2} \frac{1}{p} dr = \infty. \tag{8.2}$$

Admissibility is achieved by $p(x) = \min_{y \in \partial D} |x - y|$.

Although invariance considerations no longer always apply, the above solution can be extended to general bounded D with $V = I$, namely $p(x) = \min_{y \in \partial D} |x-y|$. For general D, V, define $|x - y|$ as the path length between points $x, y \in \bar{D}$ in the metric $d(x, y) = (x - y)' V^{-1} (x - y)$. Then D is bounded in this metric if all paths have finite length, and we again define $p(x) = \min_{y \in \partial D} |x - y|$. We offer this merely as a suggestion for an admissible prior that flirts with inadmissiblity near the boundaries, and is consistent under transformations of the data.

9. A mixture model

Suppose that y_n is a sample of size n from the normal mixture

$$Y = Z + (1 - B(q))x_1 + B(q)x_2 \tag{9.1}$$

where $Z \sim N(0, 1), B(q) \sim$ Bernoulli with mean $q, x_1 > 0, x_2 > 0, q = \frac{x_1}{x_1 + x_2}$.

The parameter $x = (x_1, x_2)$ lies in the domain $D = \{x_1 > 0, x_2 > 0\}$.

The density of a single observation y is

$$f(y) = \{x_2 \phi(y + x_1) + x_1 \phi(y - x_2)\}/(x_1 + x_2). \tag{9.2}$$

The asymptotic variance V is the inverse of the information matrix of expected values of products of the score functions:

$$l_1 = -\frac{1}{x_1 + x_2} + \frac{\phi(y - x_2)}{(x_1 + x_2)f} - \frac{(y + x_1)x_2 \phi(y + x_1)}{(x_1 + x_2)f} \tag{9.3}$$

$$l_2 = -\frac{1}{x_1 + x_2} + \frac{\phi(y + x_1)}{(x_1 + x_2)f} + \frac{(y - x_2)x_1 \phi(y - x_2)}{(x_1 + x_2)f} \tag{9.4}$$

$$L_{ij} = \int l_i l_j f dy \tag{9.5}$$

$$V = L^{-1} \tag{9.6}$$

Asymptotic admissibility for the prior p is determined by behaviour of Vp near the boundaries. Let $x_1 = rs_1, x_2 = rs_2, s_1^2 + s_2^2 = 1$.

$$x_1 \to 0 : L_{11} \to (\exp(x_2^2) - 1 - x_2^2)/x_2^2, L_{12} \to 0, L_{22} \to 0, \tag{9.7}$$

$$x_2 \to 0 : L_{22} \to (\exp(x_1^2) - 1 - x_1^2)/x_1^2, L_{12} \to 0, L_{11} \to 0, \tag{9.8}$$

$$r \to \infty : L_{11} \to s_2/(s_1 + s_2), L_{12} \to 0, L_{22} \to s_1/(s_1 + s_2). \tag{9.9}$$

At the boundary $x_1 = 0$ the normal is an eigenvector at all points $(0, x_2)$, and the integral condition for admissibility for that boundary is $\int_0^1 \frac{1}{p} L_{11} dx_1 = \infty$ almost all x_2 which reduces to $\int_0^1 \frac{1}{p} dx_1 = \infty$ almost all x_2. Similarly, the condition for admissibility on the boundary $x_2 = 0$ is $\int_0^1 \frac{1}{p} dx_2 = \infty$ almost all x_1.

For the infinite "boundary" $r \to \infty$, the integral condition for admissibility is

$$\lim_{R\to\infty} \int_{s \in S} s_i s_j [\int_1^R \frac{1}{p} V_{ij}^{-1}(rs) r^{-1} dr]^{-1} ds = 0, \tag{9.10}$$

where S is the intersection of the boundary of the unit circle and the upper right quadrant. Using the behavior of L as $r \to \infty$, this condition becomes

$$\int_1^R \frac{1}{rp(sr)} dr \to \infty \text{ almost all } s \in S. \tag{9.11}$$

Choosing a prior p to make pV admissible requires that

$$\int_0^1 \frac{1}{p} dx_1 = \int_0^1 \frac{1}{p} dx_2, \int_1^\infty \frac{1}{pr} dr = \infty. \tag{9.12}$$

Roughly, we need that p be of order x_1 near $x_1 = 0$, of order x_2 near $x_2 = 0$, and of order $log(r)$ near $r = \infty$. For example, $p = \frac{x_1 x_2}{(x_1 + x_2)^2}$ will do the job, as will many other priors with the correct behavior near the boundary. The uniform is inadmissible because it fails at $x_1 = 0$ and $x_2 = 0$.

The plot of confidence ellipses when 1000 points are sampled from the mixture model shows how the boundaries affect asymptotic admissibility. For the boundaries $x1 = 0$ and $x2 = 0$, the asymptotic variances orthogonal to the boundary in fact approach a positive limit; thus the integral of the inverse variances up to the boundary is positive rather than infinite, and the uniform density is therefore inadmissible. For the boundary at infinity, the variances are bounded away from infinity, so the integral of the inverse variance is infinite, and this boundary is admissible for a uniform prior.

10. A prior beating the uniform

It is of interest to exhibit a prior with asymptotic risk everywhere smaller than an inadmissible prior such as the the uniform in this problem. Brown's condition exhibits a prior satisfying $\frac{\partial}{\partial x'}[V \frac{\partial p}{\partial x}) = 0$. The asymptotic risk of p, relative to the uniform, is $-\frac{1}{2}(\frac{\partial p}{\partial x})' V \frac{\partial p}{\partial x}/p$. There are many solutions to the elliptic differential equation, depending on boundary values of p. The solutions are not necessarily admissible.

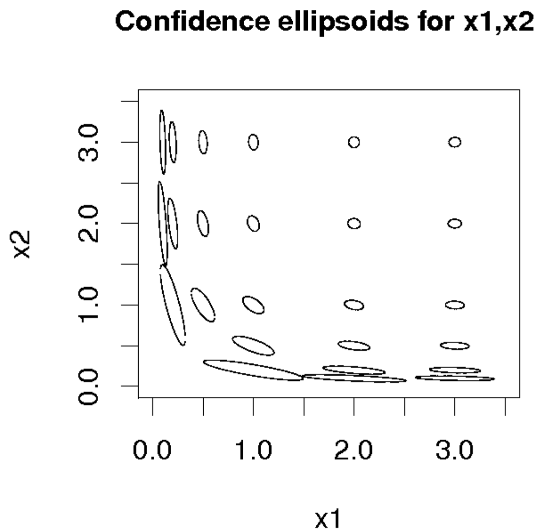

FIG 1. *Confidence Ellipses.*

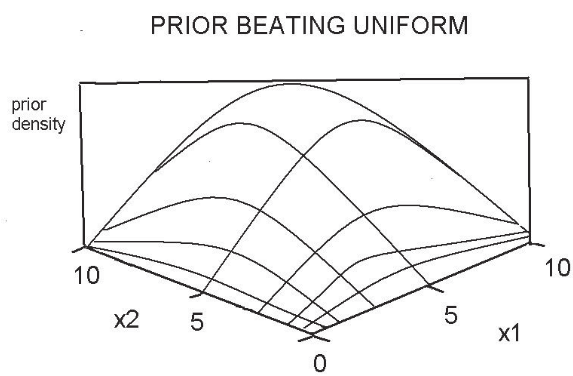

FIG 2. *Prior beating uniform.*

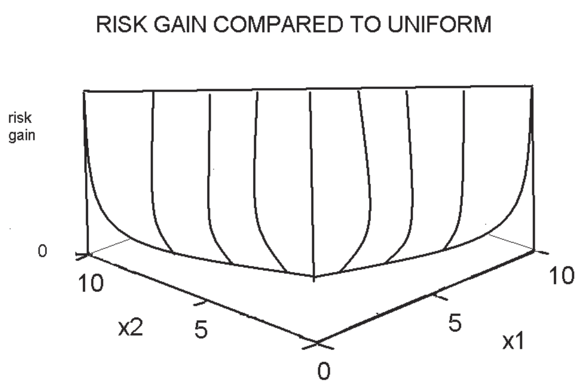

FIG 3. *Risk Gains compared to Uniform.*

We have computed solutions for the discrete approximation where x_1, x_2 each lie in the grid 0.1, 0.2,... 10. The boundary values for p are $p = \min(\frac{4x_1x_2}{(x_1+x_2)^2}, x_1x_2)$. We set these values so that p will satisfy the conditions for admissibility at the different boundaries. The following prior is obtained by using a relaxation method to solve the finite difference form of the differential equation; at the solution, the finite difference expressions are everywhere less than .01. A similar prior was developed in [Em02].

It will be noted that the prior density approaches zero at the lower and left boundary, but not at the other two boundaries, as required by the admissibility conditions.

The risk gains against the uniform are everywhere positive (as required by the theory), but are far greater near the low $x1$ and $x2$ boundaries. This is to be expected, because the prior is made admissible by changes near the boundaries, so that larger improvements in the risk should occur there.

Acknowledgement

Although Bill Strawderman has not seen this particular manuscript, he has over the years seen many an unsolicited manuscript from me on similar topics, and has always responded generously with his time, his experience, and his vast penetrating knowledge of this subject. Thanks Bill for the many kind assistances, for the numerous balloon-bursting counterexamples, and for the splendid life example of great imagination, clear thinking, clear exposition, and genial princely decency.

References

[Br71] Brown, L. D., *Admissible estimators, recurrent diffusions, and insoluble boundary value problems*, Ann. Math. Statist **42** (1971), 855–903.

[Br79] Brown, L. D., *A heuristic method for determining admissibility of estimators with applications*, Ann. Statist **7** (1979), 960–994.

[Em03] Emerson, J., *Asymptotic Admissibility and Bayesian Estimation* Ph.D. dissertation, Yale University, (2003) 180 pp. [Available from University Microfilm, 305 N. Zeeb Rd., Ann Arbor, MI 48106.]

[KW96] Kass, R. E. and Wasserman, L. A., *The selection of prior distributions by formal rules*, J. Amer. Statist. Assoc. **91**(1996), 1343–1370.

[GS98] Gilbarg, D. and Trudinger, N. S. *Elliptic Partial Differential Equations of Second Order*, Springer-Verlag, Berlin, 1998.

[Ha98] Hartigan, J. A., *The Maximum Likelihood Prior*, Ann. Statist. **35** (1998), 2083–2103.

[Ic78] Ichihara, K., *Some global properties of symmetric diffusion processes*, Publ. RIMS Kyoto University, **14** (1978), 441–486.

[Le82] Levit, B. Ya., *Minimax estimation and positive solutions of elliptic equations*, Theory Probab. Appl. **27** (1982), 563–586.

[Le83] Levit, B. Ya., *Second-order availability and positive solutions of the Schrodinger equation*, Lecture Notes in Math (**1021**) Springer, Berlin, 1983, pp. 372–385.

[Le85] Levit, B. Ya., Second-order asymptotic optimality and positive solutions of Schrodinger's equation, Theory Probab. Appl. **30** (1985), 333–363.

[Pi95] Pinsky, R. G., *Positive Harmonic Functions and Diffusion*, Cambridge University Press, Cambridge, 1995.

[St56] Stein, C., *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Proc. Third Berkeley Symp. Math. Statist. Probab.(**1**) Univ. California Press, Berkeley, 1956, pp. 197–206.

[SC71] Strawderman, W.E. and Cohen, A., *Admissibility of Estimators of the Mean Vector of a Multivariate Normal Distribution with Quadratic Loss*, Ann. Math. Statist. **42** (1971), 270–296.