Markov Chain Conditions for Admissibility in Estimation Problems with Quadratic Loss

Morris L. Eaton
School of Statistics
206 Church Street SE
Minneapolis, MN 55455, USA

ABSTRACT
Consider the problem of estimating a parametric function when the loss is quadratic. Given an improper prior distribution, there is a formal Bayes estimator for the parametric function. Associated with the estimation problem and the improper prior is a symmetric Markov chain. It is shown that if the Markov chain is recurrent, then the formal Bayes estimator is admissible. This result is used to provide a new proof of the admissibility of Pitman's estimator of a location parameter in one and two dimensions.

1991 Mathematics Subject Classification: Primary 62C15, 62C05. Secondary 62C10, 62C99.
Keywords and Phrases: Admissibility, Formal Bayes rules, Quadratic loss, Symmetric Markov chains, Pitman's estimator.
Note: Research was supported in part by National Science Foundation Grant DMS 96-26601. Research was supported in part by grants from CWI and NWO in the Netherlands.

1. Introduction

In this paper we consider a classical parametric estimation problem when the loss is quadratic. Here attention is restricted to the so-called formal Bayes estimators – that is, estimators obtained as minimizers of the posterior risk calculated via a formal posterior distribution. Because the loss is quadratic, admissibility questions regarding such estimators are typically attacked using the explicit representation of the estimator as the posterior mean of the function to be estimated. Examples can be found in Karlin (1958), Stein (1959), Zidek (1970), Portnoy (1971), Berger and Srinivasan (1978), Brown and Hwang (1982), Eaton (1992), and Hobert and Robert (1999).

To describe the problem of interest here, let $P(dx|\theta)$ be a statistical model on a sample space $\mathcal{X}$ where the parameter $\theta \in \Theta$ is unknown. That is, for each θ, $P(\cdot|\theta)$ is a probability measure on the Borel sets of $\mathcal{X}$. Both $\mathcal{X}$ and Θ are assumed to be Polish spaces with the natural σ-algebra. Given a real valued function $\phi(\theta)$ that is to be estimated, consider the loss function

$$L(a, \theta) = (a - \phi(\theta))^2, \qquad a \in R^1. \tag{1.1}$$

In order to define a formal Bayes estimator of $\phi(\theta)$, let ν be a σ-finite improper prior distribution defined on the Borel sets of Θ, so $\nu(\Theta) = +\infty$. The marginal measure on $\mathcal{X}$ is defined by

$$M(B) = \int_{\Theta} P(B|\theta)\nu(d\theta) \tag{1.2}$$

for Borel subsets of $\mathcal{X}$. When M is σ-finite (assumed throughout this paper), then a formal posterior $Q(d\theta|x)$ exists and is characterized by

$$P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx). \tag{1.3}$$

The equality in (1.3) means that the measures on $\mathcal{X} \times \Theta$ defined by the left and right side of (1.3) are equal. The formal posterior $Q(\cdot|x)$ is a probability measure for each $x \in \mathcal{X}$. For a discussion of the existence of Q and uniqueness (up to sets of M-measure zero), see Johnson (1991).

When the loss is (1.1) and the improper prior is ν, the formal Bayes estimator of $\phi(\theta)$ is defined to be the point $a(x)$ which miminizes (over a's)

$$\int (a - \phi(\theta))^2 Q(d\theta|x). \tag{1.4}$$

Of course, the minimizer is

$$\hat{\phi}(x) = \int \phi(\theta) Q(d\theta|x). \tag{1.5}$$

For the present, questions concerning the existence of integrals will be ignored. The risk function of this estimator is

$$R(\hat{\phi}, \theta) = E_\theta(\hat{\phi}(X) - \phi(\theta))^2 \tag{1.6}$$

where E_θ denotes expectation under $P(\cdot|\theta)$. The main focus of this paper concerns the admissibility of $\hat{\phi}$ and the relationship of this admissibility to a Markov chain associated with the estimation problem.

For our purposes, the relevant notion of admissibility is the following (Stein (1965)).

Definition 1.1. For any estimator $t(X)$ of $\phi(\theta)$, let $R(t, \theta) = E_\theta(t(X) - \phi(\theta))^2$ be the risk function of t. The estimator $\hat{\phi}$ is *almost-ν-admissible* $(a - \nu - a)$ if for every estimator t which satisfies

$$R(t, \theta) \le R(\hat{\phi}, \theta) \text{ for all } \theta, \tag{1.7}$$

the set

$$B = \{\theta | R(t, \theta) < R(\hat{\phi}, \theta)\} \tag{1.8}$$

has ν-measure zero.

In other words, $\hat{\phi}$ is $a - \nu - a$ if there is no estimator t which is at least as good as $\hat{\phi}$ everywhere (i.e., (1.7) holds) and which beats $\hat{\phi}$ on a set of positive ν-measure.

An important technical tool for establishing $a - \nu - a$ is the so-called Blyth-Stein condition (Blyth (1951), Stein (1955)). To describe this condition, let C be a measureable subset of Θ with $0 < \nu(C) < +\infty$. Consider the following class of real valued functions defined on Θ:

$$\begin{aligned} U(C) = \{g | g \ge 0,\ & g \text{ is bounded },\\ & g(\theta) \ge 1 \text{ for } \theta \in C, \quad \int g(\theta)\nu(d\theta) < +\infty\}. \end{aligned} \tag{1.9}$$

For $g \in U(C)$, think of $g(\theta)\nu(d\theta)$ as defining a proper prior distribution (it has not been normalized to integrate to one) and consider the marginal measure on $\mathcal{X}$ given by

$$M_g(B) = \int P(B|\theta) g(\theta) \nu(d\theta). \tag{1.10}$$

Because the measure M_g is finite, we can write (as in (1.3)),

$$P(dx|\theta) g(\theta) \nu(d\theta) = Q_g(d\theta|x) M_g(dx) \tag{1.11}$$

where $Q_g(d\theta|x)$ now is a proper posterior distribution corresponding to the proper prior $cg(\theta)\nu(d\theta)$ where c is the normalizing constant. Thus, the Bayes solution to the estimation problem is the Bayes estimator

$$\hat{\phi}_g(x) = \int \phi(\theta) Q_g(d\theta|x) \tag{1.12}$$

which is the posterior mean of $\phi(\theta)$. Next, consider the *integrated risk difference*

$$IRD(g) = \int [R(\hat{\phi}, \theta) - R(\hat{\phi}_g, \theta)] g(\theta)\nu(d\theta). \tag{1.13}$$

Roughly (subject to some regularity described precisely in later sections), one version of the Blyth-Stein condition is:

$$\begin{cases} \text{For sufficiently many sets } C, \\ \qquad \inf\limits_{g \in U(C)} IRD(g) = 0. \end{cases} \tag{1.14}$$

When (1.14) holds, then $\hat{\phi}$ is a $a - \nu - a$ (for example, see Stein (1965)). In typical examples, a direct verification of (1.14) is not routine.

A main result in this paper provides an upper bound for $IRD(g)$ which allows us to use results from Markov chain theory to establish a sufficient condition for (1.14). This result, established in Section 3 under regularity conditions, is the following:

$$\begin{cases} \text{For } \; g \in U(C),\; IRD(g) \leq \Delta(\sqrt{g}) \text{ where} \\ \Delta(h) = \int\limits_{\Theta}\int\limits_{\Theta}\int\limits_{\mathcal{X}} (h(\theta) - h(\eta))^2 (\phi(\theta) - \phi(\eta))^2 Q(d\theta|x) Q(d\eta|x) M(dx) \\ \text{is defined for real valued functions } h. \end{cases} \tag{1.15}$$

Although the function $\Delta(h)$ looks rather complicated, there is a Markov chain associated with Δ lurking in the background. To see this, recall (1.3) and let

$$R(d\theta|\eta) = \int\limits_{\mathcal{X}} Q(d\theta|x) P(dx|\eta). \tag{1.16}$$

Then $R(\cdot|\eta)$ is the expected value of the formal posterior $Q(\cdot|x)$ when the model is $P(\cdot|\eta)$. Obviously, $R(\cdot|\eta)$ is a transition function (see Eaton (1992, 1997) for further discussion; see Hobert and Robert (1999) for some related material) and we can write

$$\Delta(h) = \int\limits_{\Theta}\int\limits_{\Theta} (h(\theta) - h(\eta))^2 (\phi(\theta) - \phi(\eta))^2 R(d\theta|\eta)\nu(d\eta). \tag{1.17}$$

Then, with

$$\begin{cases} \psi(\eta) &= \int\limits_{\Theta} (\phi(\theta) - \phi(\eta)^2 R(d\theta|\eta) \\ T(d\theta|\eta) &= \psi^{-1}(\eta)(\phi(\theta) - \phi(\eta))^2 R(d\theta|\eta) \\ \xi(d\eta) &= \psi(\eta)\nu(d\eta) \end{cases} \tag{1.18}$$

it follows that

$$\Delta(h) = \int_\Theta \int_\Theta (h(\theta) - h(\eta))^2 T(d\theta|\eta)\xi(d\eta). \tag{1.19}$$

By definition, $T(d\theta|\eta)$ is a transition function and hence defines a discrete time Markov chain, $W = (W_0 = \eta, W_1, W_2, \ldots)$ whose state space is Θ and whose path space is Θ^∞. That is, under $T(\cdot|\eta)$, the chain starts at $W_0 = \eta$ and the successive states of the chain W_{i+1} have distribution $T(\cdot|W_i)$, $i = 0, 1, 2, \ldots$. Under some regularity conditions to be specified later, when the chain W is "recurrent", it follows from results in Eaton (1992, Appendix 2) that

$$\begin{cases} \text{for each set } C \text{ with } 0 < \nu(C) < +\infty, \\ \inf_{g \in U(C)} \Delta(\sqrt{g}) = 0 \end{cases} \tag{1.20}$$

Therefore, the recurrence of the chain W implies that (1.14) holds and hence $a - \nu - a$ for $\hat{\phi}$ obtains.

In summary, the above argument runs as follows:

(i) The Blyth-Stein condition (1.14) is sufficient for $a - \nu - a$.
(ii) The integrated risk difference is bounded above by $\Delta(\sqrt{g})$ as in (1.15).
(iii) When the Markov chain associated with Δ is recurrent, then (1.20) implies (1.14) holds and we have $a - \nu - a$.

Step (i) is a well known technique in decision theory and has appeared in many application such as those listed at the beginning of this section. Step (iii) was used in Eaton (1992) and is a direct consequence of general results concerning symmetric Markov chains. What is new in this paper is step (ii) as expressed in (1.15). Inequalities like (1.15) were used in Eaton (1992) but only for bounded functions ϕ. Thus the advance here is the extension of the Markov chain arguments to cover cases of estimating unbounded functions such as mean values.

The following is a simple, but not so trivial, example which shows how the results described above can be applied.

Example 1.1 Let f be a symmetric density with an absolute third moment on R^1 and assume one observation X is made from $f(x - \theta)dx$ where θ is an unknown translation parameter, $\theta \in R^1$. The loss function is $(a - \theta)^2$ so the parameter θ is to be estimated. Consider the improper prior distribution $d\theta$ so the formal posterior is $Q(d\theta|x) = f(x - \theta)d\theta$. Thus the formal Bayes estimator is

$$\int \theta Q(d\theta|x) \; = \; x$$

and the risk function is just the constant $E_0 X^2$ where E_0 denotes expectation when $\theta = 0$. The Markov chain associated with this problem has transition function T given in (1.18). A routine calculation shows that the transition function $R(d\theta|\eta)$ of (1.16) is

$$R(d\theta|\eta) \; = \; r(\theta - \eta)d\theta$$

where

$$r(u) \; = \; r(-u) \; = \; \int f(x - u) f(x) dx$$

is a density on R^1. Thus

$$\psi(\eta) \; = \; \int (\theta - \eta)^2 R(d\theta|\eta) \; = \; \int \theta^2 r(\theta) d\theta = c_2$$

is constant. From (1.18) we have

$$T(d\theta|\eta) \;=\; \frac{(\theta-\eta)^2 r(\theta-\eta)}{c_2} d\theta \;=\; t(\theta-\eta)d\theta$$

and

$$\xi(d\eta) \;=\; c_2 d\eta.$$

Therefore T is a translation kernel with density t, so the Markov chain associated with T is a random walk on R^1. Thus the existence of a first moment for t implies this random walk is recurrent (CHUNG-FUCHS (1951)).
Using the definition of t and the third moment assumption for f yields

$$\int |u| t(u) du \;=\; \frac{1}{c_2}\int |u|^3 r(u) du \;=\;$$
$$\frac{1}{c_2}\int\int |u|^3 f(x-u) f(x) du dx \leq \frac{8}{c_2} E_0|X|^3 < +\infty.$$

Hence the random walk is recurrent and the estimator x is almost admissible (relative to Lebesque measure). Of course, this example is just a very special case of the admissibility of Pitman's estimator on R^1 when third moments exist. This was first established by STEIN (1959) using the Blyth-Stein method directly.

Here is a brief summary of this paper. Section 2 contains the formal problem statement, basic assumptions, and a statement of the Blyth-Stein condition. The basic inequality is proved in Section 3 while Section 4 contains some background material on symmetric Markov chains. The main theorem connecting recurrence and admissibility is proved in Section 5, while some useful extensions are described in Section 6.

The results are then applied in Section 7 to provide an alternative proof of the admissibility of the Pitman estimator of a location parameter in one and two dimensions.

BROWN (1971) considered the problem of estimating the mean vector of a multivariate normal distribution when the loss is quadratic. Under regularity conditions, he established a close connection between admissibility and the recurrence of an associated diffusion process defined on the sample space. The relationship between Brown's work and the results here remain quite obscure. For further discussion, see EATON (1992, 1997).

2. NOTATION AND ASSUMPTIONS:

Certain integrability assumptions are needed to justify the arguments sketched in Section 1. Some of these assumption are stated here.

The two spaces $\mathcal{X}$ and Θ are assumed to be Polish spaces with the natural σ-algebras. The model $P(dx|\theta)$ is a Markov kernel and the improper prior distribution ν is σ-finite. The marginal measure $M(dx)$ defined in (1.2) is assumed to be σ-finite so that equation (1.3) holds for the formal posterior $Q(d\theta|x)$.

Let ϕ be a real valued function defined on Θ such that

$$\int \phi^2(\theta) Q(d\theta|x) < +\infty \quad \text{for all } x.. \tag{A.1}$$

Then the formal Bayers estimator $\hat{\phi}(x)$ given in (1.5) is well defined. The risk function defined by (1.6) is assumed to satisfy the following local integrability condition:

$$\begin{cases} \text{There exists an increasing sequence of} \\ \text{sets } \{K_i\} \text{ such that } \bigcup K_i \;=\; \Theta,\; 0 < \nu(K_i) < \infty, \\ \int\limits_{K_i} R(\hat{\phi}, \theta)\nu(d\theta) < \infty, \text{ for each } i. \end{cases} \tag{A.2}$$

Observe that if $g \in U(K_i)$ (as defined in (1.9)) and g vanishes outside some K_j with $j > i$, then the integrated risk

$$\int R(\hat{\phi}, \theta) g(\theta) \nu(d\theta). \tag{2.1}$$

is finite.

Now, recalling (1.9), let $g \in U(C)$ and consider

$$\hat{g}(x) \;=\; \int g(\theta) Q(d\theta|x). \tag{2.2}$$

Recall that the marginal measure M_g is

$$M_g(B) \;=\; \int\limits_{\Theta} \int\limits_{\mathcal{X}} I_B(x) P(dx|\theta) g(\theta) \nu(d\theta). \tag{2.3}$$

Using (1.3), we see

$$M_g(dx) \;=\; \hat{g}(x) M(dx) \tag{2.4}$$

so that $\hat{g}$ is the Radon-Nikodym derivative of M_g with respect to M. Hence the set $A_0 = \{x|\hat{g}(x) = 0\}$ has M_g measure zero. Now, define $Q_g(d\theta|x)$ as follows:

$$Q_g(d\theta|x) \;=\; \begin{cases} \dfrac{g(\theta)}{\hat{g}(x)} Q(d\theta|x) & \text{if } x \notin A_0 \\ \quad Q(d\theta|x) & \text{if } x \in A_0. \end{cases} \tag{2.5}$$

It is then easy to verify that

$$P(dx|\theta) g(\theta) \nu(d\theta) \;=\; Q_g(d\theta|x) M_g(dx). \tag{2.6}$$

Therefore the Bayes estimator

$$\hat{\phi}_g(x) \;=\; \int \phi(\theta) Q_g(d\theta|x) \tag{2.7}$$

is well defined because (A.1) and the boundedness of g imply

$$\int \phi^2(\theta) Q_g(d\theta|x) \;<+\infty \text{ for all } x. \tag{2.8}$$

A rigorous statement of the Blyth-Stein Lemma follows. Given a K_i in (A.2), let

$$U^*(K_i) \;=\; \{g|g \in U(K_i),\; \int R(\hat{\phi}, \theta) g(\theta) \nu(d\theta) < +\infty\}. \tag{2.9}$$

Theorem 2.1 (Blyth-Stein Lemma). *For each i, assume that*

$$\inf_{g \in U^*(K_i)} IRD(g) \;=\; 0. \tag{2.10}$$

Then $\hat{\phi}$ is $a - \nu - a$.

PROOF. The proof of this well known condition is by contradiction. The details are left to the reader.

THEOREM 2.2 *For* $g \in U^*(K_i)$,

$$IRD(g) \;=\; \int_{A_0^c} (\hat{\phi}(x) - \hat{\phi}_g(x))^2 \hat{g}(x) M(dx). \tag{2.11}$$

PROOF. The proof of (2.11) is routine algebra coupled with the earlier observation that A_0 has M_g measure zero.

3. The Basic Inequality

In this section, the inequality described in (1.15) is established for $g \in U^*(K_i), i = 1, 2, \ldots$. Here is a basic lemma which may be of independent interest.

LEMMA 3.1 *Let W and Y be real valued random variables such that $EW^2 < +\infty, Y \geq 0$, and $\mu = EY < +\infty$. Also let $(\tilde{W}, \tilde{Y})$ be an independent and identically distributed copy of (W, Y). Then*

$$[\mathrm{Cov}(W, Y)]^2 \leq \mu E(W - \tilde{W})^2 (\sqrt{Y} - \sqrt{\tilde{Y}})^2. \tag{3.1}$$

PROOF. A direct calculation shows that

$$\mathrm{Cov}(W, Y) \;=\; \frac{1}{2} E(W - \tilde{W})(Y - \tilde{Y}).$$

Writing

$$(Y - \tilde{Y}) \;=\; (\sqrt{Y} - \sqrt{\tilde{Y}})(\sqrt{Y} + \sqrt{\tilde{Y}})$$

and using the Cauchy-Schwarz inequality yields

$$[\mathrm{Cov}(W, Y)]^2 \leq \frac{1}{4} E(\sqrt{Y} + \sqrt{\tilde{Y}})^2 E(W - \tilde{W})^2 (\sqrt{Y} - \sqrt{\tilde{Y}})^2.$$

But $(\sqrt{Y} + \sqrt{\tilde{Y}})^2 \leq 2(Y + \tilde{Y})$ so that $\frac{1}{4} E(\sqrt{Y} + \sqrt{\tilde{Y}})^2 \leq \mu$. This completes the proof.

THEOREM 3.1 *For* $g \in U^*(K_i)$,

$$IRD(g) \leq \Delta(\sqrt{g}) \tag{3.2}$$

where Δ is defined in (1.15).

PROOF. For each $x \in A_0^c \;=\; \{x | \hat{g}(x) > 0\}$,

$$\begin{aligned} \hat{\phi}(x) - \hat{\phi}_g(x) \quad &= \int \phi(\theta) Q(d\theta|x) - \int \phi(\theta) Q_g(d\theta|x) \\ &= \frac{1}{\hat{g}(x)} \int \phi(\theta)(\hat{g}(x) - g(\theta)) Q(d\theta|x) \\ &= -\frac{1}{\hat{g}(x)} \mathrm{Cov}_x(\phi, g) \end{aligned} \tag{3.3}$$

where Cov_x denotes covariance under the probability measure $Q(\cdot|x)$. The last equality follows since $\hat{g}(x)$ is the mean of $g(\theta)$ under $Q(\cdot|x)$. Applying inequality (3.1) with $W = \phi$ and $Y = g$, we have

$$\begin{aligned} &(\hat{\phi}(x) - \hat{\phi}_g(x))^2 = \frac{1}{\hat{g}(x))^2} (\mathrm{Cov}_x(\phi, g))^2 \leq \\ &\frac{1}{\hat{g}(x)} \int_\Theta \int_\Theta (\phi(\theta) - \phi(\eta))^2 (\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 Q(d\theta|x) Q(d\eta|x). \end{aligned} \tag{3.4}$$

Subsituting this inequality into the rightside of (2.11) clearly yields (3.2). This completes the proof.

The upper bound $\Delta(\sqrt{g})$ in (3.2) depends only on the three essential components of the original problem – namely the model, the improper prior and the function ϕ to be estimated. Of course this statement assumes that the loss is quadratic. When the function ϕ is bounded, say $|\phi(\theta| \leq c$, then obviously

$$\Delta(\sqrt{g}) \leq 4c^2 \Delta_1(\sqrt{g}) \tag{3.5}$$

where

$$\Delta_1(\sqrt{g}) = \int \int \int (\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 Q(d\theta|x) Q(d\eta|x) M(dx). \tag{3.6}$$

The function Δ_1 appeared in Eaton (1992) and was used to relate Markov chain recurrence to admissibility questions regarding the estimation of bounded functions. Not only is the argument here more general, it is far more transparent than the original in the case when ϕ is bounded.

4. Symmetric Markov chains

Some basic theory concerning symmetric Markov chains with values in a Polish space is described here. Of course, the emphasis is on those aspects of the theory which are most directly related to the admissibility questions under consideration here. The discussion follows Eaton (1992, Appendix 2) quite closely.

Let $(\mathcal{Y}, \mathcal{B})$ be a measurable space where $\mathcal{Y}$ is Polish and $\mathcal{B}$ is the usual Borel σ-algebra. Consider a Markov kernel $S(du|v)$ defined on $\mathcal{B} \times \mathcal{Y}$ so that $S(\cdot|v)$ is a probability measure for each $v \in \mathcal{Y}$ and $S(B|\cdot)$ is $\mathcal{B}$-measureable for each $B \in \mathcal{B}$. Let ξ be a σ-finite measure defined on $\mathcal{B}$ with $\xi(\mathcal{Y}) > 0$.

Definition 4.1 The Markov kernel $S(du|v)$ is ξ-*symmetric* if the measure

$$m(du, dv) \ = \ S(du|v)\xi(dv) \tag{4.1}$$

defined on $\mathcal{B} \times \mathcal{B}$ is a symmetric measure.

In all that follows, $S(du|v)$ is assumed to be ξ-symmetric. The assumption that ξ is σ-finite is important (see the development in Appendix 2 in Eaton (1992)). The symmetry of m implies that m has marginal measures ξ-that is,

$$m(\mathcal{Y} \times B) \ = \ m(B \times \mathcal{Y}) \ = \ \xi(B). \tag{4.2}$$

Of course, (4.2) implies that ξ is a stationary measure for $S(du|v)$ since

$$\int_{\mathcal{Y}} S(B|v)\xi(dv) \ = \ \xi(B). \tag{4.3}$$

Now, each Markov kernel defines a Markov chain, and conversely, to specify a Markov chain one needs, at least implicity, a Markov kernel. A Markov chain is called symmetric if this Markov kernel is symmetric with respect to some σ-finite measure. For finite and countable state spaces, symmeteric Markov chains are also called reversible chains, but that terminology is not used here (see Kelly (1979) or Lawler (1995)).

According to the above terminology, a symmetric Markov chain on $\mathcal{Y}$ gives rise to a symmetric measure (as in (4.1)) on $\mathcal{B} \times \mathcal{B}$ and this symmetric measure has a σ-finite marginal measure as defined in (4.2). Conversely, suppose $n(du, dv)$ is a symmetric measure on $\mathcal{B} \times \mathcal{B}$ and suppose its marginal measure

$$\mu(B) \ = \ n(B \times \mathcal{Y}) \tag{4.4}$$

is σ-finite. This implies that there is a unique (up to sets of μ-measure zero) Markov kernel $T(du|v)$ such that

$$n(du, dv) \;=\; T(du|v)\mu(dv). \tag{4.5}$$

This result seems to be well known but I do not know a reference with an explicit statement. A slightly more general result can be found in Johnson (1991). The above discussion shows there is a one to one correspondence between symmetric Markov chains and symmetric measures with σ-finite marginals. This observation is what allows us to associate a Markov chain with the function Δ appearing in (1.15). More about this in the next section.

Now, let $S(du|v)$ be ξ-symmetric and let $Y = (Y_0 = v, Y_1, Y_2, \dots)$ be the corresponding Markov chain with values in $\mathcal{Y}$. The notation means the chain starts at v and the succesive Y_{i+1} have distribution $S(\cdot|Y_i)$ for $i = 0, 1, \dots$. The joint measure of the chain on $\mathcal{Y}^\infty$ is denoted by $\mathrm{Prob}(\cdot|v)$ where $Y_0 = v$ is the initial state of the chain.

Next, we turn to a discussion of recurrence when $S(du|v)$ is ξ-symmetric.

Definition 4.2 Let $B \in \mathcal{B}$ satisfy $0 < \xi(B) < +\infty$. The set B is *locally-ξ-recurrent* $(l - \xi - r)$ if the set

$$\{v | v \in B, \mathrm{Prob}(Y_j \in B \text{ for some } j \geq 1 | v) < 1\} \tag{4.6}$$

has ξ measure zero.

In other words, B is $l - \xi - r$ if except for a set of starting values of ξ-measure zero, the chain returns to B with probability one when it starts in B. A characterization of local-ξ-recurrence can be given in terms of a quadratic form. For $h \in L^2(\xi)$, the linear space of ξ square integrable functions, define $D(h)$ by

$$D(h) \;=\; \int\int (h(u) - h(v))^2 m(du, dv). \tag{4.7}$$

where m is the symmetric measure given by (4.1). For B such that $0 < \xi(B) < +\infty$, let

$$V(B) \;=\; \{h | h \geq 0, h \in L^2(\xi),\ h(u) \geq 1 \text{ for } u \in B\}. \tag{4.8}$$

Theorem 4.1 *The following are equivalent:*

(i) B is $l - \xi - r$

(ii) $\inf_{h \in V(B)} D(h) = 0$

Proof. This is a direct consequence of Theorem A.2 in Eaton (1992).

For our applications, a slight strengthening of Theorem 4.1 is needed. Let $C \in \mathcal{B}$ satisfy $C \supseteq B$ and $\xi(C) < +\infty$. Then set

$$V(B, C) \;=\; \{h | h \in V(B), h \text{ is bounded }, h(u) = 0 \text{ for } u \in C^c\}. \tag{4.9}$$

Theorem 4.2 *Consider $C_1 \subseteq C_2 \subseteq \cdots$ with $B \subseteq C_1$ and $\lim C_i = \mathcal{Y}$. The following are equivalent*

(i) B *is* $l - \xi - r$

(ii) $\lim_{i\to\infty} \inf_{h \in V(B, C_i)} D(h) \;=\; 0.$

Proof. This is a consequence of results in Eaton (1992, Appendix 2).

It is Theorem 4.2 which will be used to establish a connection between the Blyth-Stein condition and recurrence.

DEFINITION 4.3 The chain Y is *locally-ξ-recurrent* if for each set B with $0 < \xi(B) < +\infty$, B is $l-\xi-r$.

It is not too hard to show that Y is locally-ξ-recurrent iff there exists an increasing sequence of sets $C_1 \subseteq C_2 \subseteq \cdots$ with $0 < \xi(C_i) < +\infty$ and $\lim\ C_i = \mathcal{Y}$ such that each C_i is $l - \xi - r$. In applications one can often choose a convenient sequence of sets C_i in order to check $l - \xi - r$.

The quadratic form $D(h)$ in (4.7) is well known in the theory and applications of symmetric Markov chains. In the probability literature $\frac{1}{2}D(h)$ is known as the *Dirichlet form* associated with the symmetric measure m, or the symmetric transition S in (4.1). It is typical to write $\frac{1}{2}D(h)$ in terms of the linear transformation S^* defined on $L^2(\xi)$ as follows:

$$(S^*h)(v) \ = \ \int h(u)S(du|v). \tag{4.10}$$

Let (h_1, h_2) denote the standard inner product on $L^2(\xi)$ given by

$$(h_1, h_2) \ = \ \int h_1(u)h_2(u)\xi(du). \tag{4.11}$$

A routine calculation shows that

$$\frac{1}{2}D(h) \ = \ (h, (I - S^*)h) \tag{4.12}$$

where I is the identity. The operator $I - S^*$ is commonly called the LaPlacian. Further discussion and some applications can be found in DIACONIS and STROOK (1991) and LAWLER (1995).

5. RECURRENCE IMPLIES ADMISSIBILITY

It is argued here that, under an additional assumption, recurrence of the Markov chain associated with the quadratic form

$$\Delta(h) = \int_\Theta \int_\Theta \int_\mathcal{X} (h(\theta) - h(\eta))^2(\phi(\theta) - \phi(\eta))^2 Q(d\theta|x)Q(d\eta|x)M(dx) \tag{5.1}$$

will imply that the Blyth-Stein condition of Theorem 2.1 holds, so that $\hat{\phi}$ is $a - \nu - a$.

To carry out this argument, first observe that the measure on $\Theta \times \Theta$ given by

$$\alpha(d\theta, d\eta) \ = \ \int_\mathcal{X} (\phi(\theta) - \phi(\eta))^2 Q(d\theta|x)Q(d\eta|x)M(dx) \tag{5.2}$$

is, by inspection, symmetric. Using (1.3) and (1.16), the measure α can be written

$$\alpha(d\theta, d\eta) \ = \ (\phi(\theta) - \phi(\eta))^2 R(d\theta|\eta)\nu(d\eta) \tag{5.3}$$

where $R(d\theta|\eta)$ is a transition function and ν is the improper prior used to defined the estimator $\hat{\phi}(x)$ in (1.15). Next, for $\eta \in \Theta$, let

$$\psi(\eta) \ = \ \int (\phi(\theta) - \phi(\eta))^2 R(d\theta|\eta). \tag{5.4}$$

The following assumption controls the behavior of ψ and is expressed in terms of the sets K_i appearing in assumption (A.2) of Section 2.

$$\begin{cases} 0 < \psi(\eta) < +\infty \text{ for all } \eta \in \Theta, \text{ and} \\ \int_{K_i} \psi(\eta)\nu(d\eta) < +\infty \text{ for all } i. \end{cases} \tag{A.3}$$

Theorem 5.1 *Assume (A.3) holds. Then the symmetric measure α has a σ-finite marginal measure*

$$\xi(d\eta) \;=\; \psi(\eta)\nu(d\eta). \tag{5.5}$$

Further, with

$$T(d\theta|\eta) \;=\; \psi^{-1}(\eta)(\phi(\theta) - \phi(\eta))^2 R(d\theta|\eta), \tag{5.6}$$

The measure α is given by

$$\alpha(d\theta, d\eta) \;=\; T(d\theta|\eta)\xi(d\eta). \tag{5.7}$$

Proof. That (5.7) holds in immediate from (5.3) and the definition of ξ and T. Since $T(d\theta|\eta)$ is a transition function by definition, integration of (5.7) over Θ shows that α has ξ as a marginal measure. Th σ-finiteness of ξ is immediate from assumption (A.3). This completes the proof.

Now, let $W = (W_0 = \eta, W_1, W_2, \dots)$ be the Markov chain on Θ with transition function T. The above discussion shows that T is ξ-symmetric (i.e. W is a symmetric Markov chain). Observe that the quadratic form associated with this chain as defined in (4.6) is exactly Δ given in (5.1). In other words, for $h \in L^2(\xi)$,

$$\Delta(h) \;=\; \int\int (h(\theta) - h(\eta))^2 \alpha(d\theta, d\eta) \tag{5.8}$$

so that the results described in Section 4 are directly applicable.

Here is the main result of this paper.

Theorem 5.2 *Assume (A.1), (A.2) and (A.3) hold. If the Markov chain W associated with the quadratic form Δ is locally-ξ-recurrent, then the formal Bayes estimator $\hat{\phi}(x)$ is almost-ν-admissible.*

Proof. It suffices to show that condition (2.10) holds for each $i, i = 1, 2, \dots$. Fix an index $j > i$ and consider the set $V(K_i, K_j)$ defined in (4.9). Assumptions (A.2) and (A.3) show that if $\sqrt{g} \in V(K_i, K_j)$ then $g \in U^*(K_i)$. This observation together with the basic inequality (3.2) yields

$$\inf_{g \in U^*(K_i)} IRD(g) \leq \inf_{\sqrt{g} \in V(K_i, K_j)} \Delta(\sqrt{g}). \tag{5.9}$$

By assumption the chain W is $l - \xi - r$ so the limit of the right side of (5.9) as $j \to \infty$ is zero. Thus for each i, (2.10) holds and the proof is complete.

6. An Extension

In this section, we extend the results of the previous sections to cover the case of estimating a vector valued function $\phi(\theta, x), \theta \in \Theta, x \in \mathcal{X}$. The model $P(dx|\theta)$ and the improper prior are as in Section 2. For vectors $w \in R^k, ||w||$ denotes the usual Euclidean norm. The loss function for the estimation problem is

$$L(a, \theta, x) \;=\; ||a - \phi(\theta, x)||^2, a \in R^k \tag{6.1}$$

so $\phi(\theta, x)$ is a k-dimensional vector and the loss function now depends on $x \in \mathcal{X}$. The following assumption is the appropriate analogue of (A.1) given in Section 2. Assume

$$\int ||\phi(\theta,x)||^2 Q(d\theta|x) < +\infty \text{ for all } x \tag{B.1}$$

where $Q(d\theta|x)$ is the formal posterior. Thus, the formal Bayes estimator is now the vector function

$$\hat{\phi}(x) \;=\; \int \phi(\theta,x)Q(d\theta|x). \tag{6.2}$$

Of course, the risk function is

$$R(\hat{\phi},\theta) \;=\; \int ||\hat{\phi}(x)-\phi(\theta,x)||^2 P(dx|\theta). \tag{6.3}$$

Assumption (B.2) is that the risk function satisfies the local integrability condition (A.2) given in Section 2.

Now, the Blyth-Stein Lemma given in Theorem 2.1 remains valid and the analogue of Theorem 2.2 is

Theorem 6.1 *For* $g \in U^*(K_i)$,

$$IRD(g) \;=\; \int_{A_0^c} ||\hat{\phi}(x)-\hat{\phi}_g(x)||^2 \hat{g}(x)M(dx). \tag{6.4}$$

Proof. Apply Theorem 2.2 one coordinate at a time to the problem of estimating $\phi_j(\theta,x)$ where $\phi_j(\theta,x)$ is the jth coordinate of $\phi(\theta,x)$. Then sum on j to obtain (6.4). This completes the proof.

The next step is to extend Theorem 3.1 to the case at hand. To this end, define Δ_2 for real valued functions $h(\theta)$ by

$$\begin{aligned}\Delta_2(h) \;=&\\ &\int_\Theta\int_\Theta\int_{\mathcal{X}} (h(\theta)-h(\eta))^2||\phi(\theta,x)-\phi(\eta,x)||^2 Q(d\theta|x)Q|d\eta|x)M(dx).\end{aligned} \tag{6.5}$$

Theorem 6.2 *For* $g \in U^*(K_i)$,

$$IRD(g) \le \Delta_2(\sqrt{g}) \tag{6.6}$$

where Δ_2 *is defined by (6.5).*

Proof. The argument used to prove Theorem 3.1 shows that for each $x \in A_0^c = \{x|\hat{g}(x) > 0\}$,

$$\begin{aligned}&(\hat{\phi}_j(x)-\hat{\phi}_{gj}(x))^2 \le\\ &\frac{1}{\hat{g}(x)}\int\int (\phi_j(\theta,x)-\phi_j(\eta,x))^2(\sqrt{g(\theta)}-\sqrt{g(\eta)})^2 Q(d\theta|x)Q(d\eta|x).\end{aligned} \tag{6.7}$$

Summing this on j, integrating with respect to $\hat{g}(x)M(dx)$, and using (6.4) shows that (6.6) holds. This completes the proof.

The final step in the argument here is to associate a symmetric Markov chain with Δ_2. To this end, define the measure α_2 on $\Theta \times \Theta$ by

$$\alpha_2(d\theta,d\eta) \;=\; \int_{\mathcal{X}} ||\phi(\theta,x)-\phi(\eta,x)||^2 Q(d\theta|x)Q(d\eta|x)M(dx). \tag{6.8}$$

Obviously, α_2 is symmetric. To formulate the analogue of assumption (A.3) of Section 5, define $\psi_2(\eta)$ by

$$\psi_2(\eta) \; = \; \int_{\mathcal{X}} \int_{\Theta} ||\phi(\theta, x) - \phi(\eta, x)||^2 Q(d\theta|x) P(dx|\eta). \tag{6.9}$$

Now, make the following assumption:

$$\text{(B.3)} \qquad \begin{cases} 0 < \psi_2(\eta) & < +\infty \quad \text{for all } \eta \in \Theta \text{ and} \\ \displaystyle\int_{K_i} \psi_2(\eta)\nu(d\eta) & < +\infty \quad \text{for all } i. \end{cases}$$

Setting

$$T_2(d\theta|\eta) \; = \; \psi_2^{-1}(\eta) \int_{\mathcal{X}} ||\phi(\theta, x) - \phi(\eta, x)||^2 Q(d\theta|x) P(dx|\eta), \tag{6.10}$$

and

$$\xi_2(d\eta) \; = \; \psi_2(\eta)\nu(d\eta), \tag{6.11}$$

it is clear that T_2 is a transition function, the measure ξ_2 is σ-finite since (B.3) holds, and

$$\alpha_2(d\theta, d\eta) \; = \; T_2(d\theta|\eta)\xi_2(d\eta). \tag{6.12}$$

Thus, α_2 has a σ-finite marginal measure ξ_2 and the results described in Section 5 apply directly to the Markov chain with transition function T_2. The extension of Theorem 5.2 is now immediate.

Theorem 6.3 *Assume (B.1), (B.2) and (B.3) hold. If the Markov chain associated with the quadratic form Δ_2 is locally-ξ-recurrent, then the formal Bayes estimator $\hat{\phi}(x)$ is almost-ν-admissble.*

Proof. The proof of Theorem 5.2 applies directly.

7. Admissibility of the Pitman Estimator

Here we provide an alternative proof of the almost admissibility of the Pitman estimator of a location parameter in one and two dimensions. The original proofs (Stein (1995) for one dimension and James and Stein (1961) for two dimensions) are based on a direct verification of the Blyth-Stein condition. The proof given here uses Markov chain arguments via Theorem 6.3.

For notational convenience, we consider a model in the so-called invariant Pitman form. A random quantity $X = (Y, Z)$ is to be observed where Y is a k-vector and Z takes value in a Polish space $\mathcal{Z}$. The parametric model for X is assumed to have the form

$$P(dx|\theta) \; = \; f(y - \theta, z) dy \lambda(dz) \tag{7.1}$$

where dy is Lebesque measure on R^k, λ is a σ-finite measure on the Borel sets of $\mathcal{Z}, \theta$ is an unknown vector in R^k and f is a density with respect to the product measure $dy\lambda(dz)$. The function to be estimated is the vector function $\phi(\theta) = \theta$, the loss is quadratic and the improper prior distribution is Lebesque measure $d\theta$ on R^k. It is clear that

$$m(z) \; = \; \int f(y - \theta, z) dy \; = \; \int f(y, z) dy \tag{7.2}$$

is the marginal density of Z with respect to λ. It is easy to see that the marginal measure on $R^k \times \mathcal{Z}$ is given by

$$M(dy, dz) \;=\; m(z) dy \lambda(dz) \tag{7.3}$$

and is σ-finite. Define $q(\theta|y,z)$ by

$$q(\theta|y,z) \;=\; \begin{cases} \dfrac{f(y-\theta,z)}{m(z)} & \text{if } 0 < m(z) < \infty \\ q_0(y-\theta) & \text{otherwise} \end{cases} \tag{7.4}$$

where q_0 is a density on R^k with finite second moments. A routine argument shows that

$$Q(d\theta|y,z) \;=\; q(\theta|y,z)d\theta \tag{7.5}$$

serves as a formal posterior so (1.3) holds. The Pitman estimator for θ is

$$\hat{\phi}(y,z) \;=\; \int \theta Q(d\theta|y,z), \tag{7.6}$$

which is the formal Bayes estimator for θ.

The formal statement regarding the almost admissibility of $\hat{\phi}$ is the following.

Theorem 7.1 *For $k = 1$ or $k = 2$ assume that*

$$\int\int ||y||^{2+k} f(y,z) dy \lambda(dz) < +\infty. \tag{7.7}$$

Then $\hat{\phi}$ in (7.6) is an almost admissible estimator for $\theta \in R^k, k = 1, 2$.

Proof. The arguments for $k = 1$ and $k = 2$ are essentially the same. The details are given for the case of $k = 1$. Assumption (7.7) implies that the set

$$N \;=\; \{z| \int |y|^3 f(y,z) dy \;=\; +\infty\} \tag{7.8}$$

has λ-measure zero. Thus the density f can be set equal to zero on this set without changing the problem,. In what follows, assume this has been done. Now, assumption (A.1) follows immediately. Since the estimator $\hat{\phi}$ is translation invariant, i.e.

$$\hat{\phi}(y-c,z) \;=\; c + \hat{\phi}(y,z) \;\;,\;\; c \in R^1, \tag{7.9}$$

and the model is invariant under translation, it follows that the risk function $R(\hat{\phi},\theta)$ is a constant given by

$$c_0 \;=\; \int\int (\hat{\phi}(y,z))^2 f(y,z) dy \lambda(dz). \tag{7.10}$$

That $c_0 < +\infty$ follows from (7.7). Therefore assumption (A.2) holds with $K_i \;=\; [-i,i] \subseteq R^1, i = 1, 2, \ldots.$

For the verification of (A.3), first observe that the transitition function defined in (1.16) is, in the present context, given by

$$R(d\theta|\eta) \;=\; r(\theta - \eta) d\theta \tag{7.11}$$

where

$$r(u) \;=\; r(-u) \;=\; \int\!\!\int q(u|y,z)f(y,z)dy\lambda(dz) \tag{7.12}$$

is a density on R^1. Therefore the function $\psi(\eta)$ defined in (5.4) is

$$\psi(\eta) \;=\; \int(\theta-\eta)^2 r(\theta-\eta)d\theta \;=\; \int u^2 r(u)du \tag{7.13}$$

which is a constant, say c_1. Again (7.7) implies that $c_1 < +\infty$ so (A.3) holds.

The final step in the proof requires us to show that the Markov chain with the transition function

$$T(d\theta|\eta) \;=\; c_1^{-1}(\theta-\eta)^2 r(\theta-\eta)d\theta \tag{7.14}$$

is almost-ξ-recurrent. Since the transition function T has the form

$$T(d\theta|\eta) \;=\; t(\theta-\eta)d\theta \tag{7.15}$$

where t is a symmetric density on R^1, a sufficient condition for recurrence is

$$\int |u|t(u)du < +\infty. \tag{7.16}$$

(see Chung-Fuchs (1951)). Substituting the expressions for t and r into (7.16) shows that

$$\begin{aligned}
&\int |u|t(u)du \;=\; c_1^{-1}\int |u|^3 r(u)du \\
&= c_1^{-1}\int\!\!\int\!\!\int |u|^3 q(u|y,z)f(y,z)dy\lambda(dz)du \\
&= c_1^{-1}\int\!\!\int\!\!\int |u|^3 f(y-u,z)f(y,z)\frac{1}{m(z)}dy\lambda(dz)du \\
&= c_1^{-1}\int\!\!\int\!\!\int |y-w|^3 f(w,z)f(y,z)\frac{1}{m(z)}dydw\lambda(dz) \le \\
&4c_1^{-1}\int\!\!\int\!\!\int |y|^3 f(w,z)f(y,z)\frac{1}{m(z)}dydw\lambda(dz) + \\
&4c_1^{-1}\int\!\!\int\!\!\int |w|^3 f(w,z)f(y,z)\frac{1}{m(z)}dydw\lambda(dz) \;=\; \\
&8c_1^{-1}\textstyle\int |y|^3 f(y,z)dy\lambda(dz).
\end{aligned} \tag{7.17}$$

The final expression is finite by assumption (7.7) so the random walk associated with T in (7.14) is recurrent. By Theorem 5.2, ϕ is almost admissible. This completes the proof for dimension $k=1$.

When $k=2$, the argument proceeds as above until the final step. On R^2, the existence of a first moment for the transition density t in (7.15) is not sufficient for recurrence. However the existence of second moments is sufficient (see Revuz (1984, Chapter 3) for example). This is the reason condition (7.7) depends on the dimension parameter k. The details of the argument are left to the reader. This completes the proof.

Of course the above argument fails completely for $k \ge 3$ since $R^k (k \ge 3)$ does not support any non-trivial recurrent random walks (see Guivarch'h, Keane and Roynette (1977)). Appropriate shrinkage estimators on $R^k, k \ge 3$, provide explicit dominators of Pitman estimators in many translation problems.

The results in Perng (1970) show that in the case of $k = 1$, failure of the third moment assumption can lead to inadmissibility of the Pitman estimator. It is encouraging that the Markov chain arguments used here reproduce results which are known to be fairly sharp. At present, very little more is known concerning the sharpness of the Markov chain argument in Theorem 6.3. Work in this direction is underway.

References

1. Berger, J.O. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.* **6**. 783–801.
2. Blyth, C.R. (1951). On minimax statistical procedures and their admissibility. *Ann. Math. Statist.* **22**, 22–42.
3. Brown, L. (1971). Admissible estimators, recurrent diffusions, and insolvable boundary value problems. *Ann. Math. Statist.* **42**. 855–904.
4. Brown, L. and Hwang, J.T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics* IV. (S.S. Gupta and J.O. Berger, eds.). **1**, 299–324. Academic Press, New York.
5. Chung, K.L. and Fuchs, W.H. (1951). On the distribution of values of sums of random variables. *Mem. Amer. Math. Soc.* **6** 1–12.
6. Diaconis, P. and Strook, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1**, 36–61.
7. Eaton, M.L. (1992). A statistical diptych: Admissible inferences-Recurrence of symmetric Markov chains. *Ann. Statist.* **20**, 1147–1179.
8. Eaton, M.L. (1997). Admissibility in quadratically regular problems and recurrence of symmeteric Markov chains: Why the connection? *Jour. Statist. Planning and Inference.* **64**, 231–247.
9. Guivarc'h, Y. Keane, M. and Roynette, B. (1977). *Marches aleatoires sur les Groupes de Lie.* Lecture Notes in Math. **624**. Springer Verlag, New York.
10. Hobert, J.P. and Robert, C.P. (1999). Eaton's Markov chain, its conjugate partner and P-admissibility. To appear in *Ann. Statist.*
11. James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1**, 361–380. University of California Press, Berkeley.
12. Johnson, B. (1991). On the admissibility of improper Bayes inferences in fair Bayes decision problems. Ph.D. Thesis, University of Minnesota.
13. Karlin, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.* **29**, 406–436.
14. Kelly, F.P. (1979). *Reversibility and Stochastic Networks.* Wiley, New York.

15. Lawler, G.F. (1995). *Introduction to Stochastic Processes*. Chapman-Hall, London.

16. Perng, S.K. (1970). Inadmissibility of various "good" statistical procedures which are translation invariant. *Ann. Math. Statist.* **41**, 1311–1321.

17. Portnoy, S. (1971). Formal Bayes estimation with applications to a random effects model. *Ann. Math. Statist.* **42**, 1379–1402.

18. Revuz, D. (1984). *Markov chains, Second Edition*. North-Holland, Amsterdam.

19. Stein, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Statist.* **30**, 970–979.

20. Stein, C. (1959). The admissiblity of Pitman's estimator of a single location parameter. *Ann. Math. Statist.* **30**, 970–979.

21. Stein, C. (1965). Lecturer notes on decision theory. Unpublished manuscript.

22. Zidek, J.V. (1970). Sufficient conditions for admissibility under squared error loss of formal Bayes estimators. *Ann Math. Statist.* **41**, 446–456.