A Festschrift for Herman Rubin
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 45 (2004) 5–20

Evaluating improper priors and the recurrence of symmetric Markov chains: An overview

Morris L. Eaton[1]

University of Minnesota

Abstract: Given a parametric statistical model, an improper prior distribution can often be used to induce a proper posterior distribution (an inference). This inference can then be used to solve decision problems once an action space and loss have been specified. One way to evaluate the inference is to ask for which estimation problems does the above formal Bayes method produce admissible estimators. The relationship of this problem to the recurrence of an associated symmetric Markov chain is reviewed.

Appreciation

Near the end of my graduate study at Stanford, Carl Morris and I had a conversation which lead us to ask whether or not the usual χ^2-test for a point null hypothesis in a multinomial setting was in fact a proper Bayes test. After a few months of struggle, we eventually reduced the problem to one involving La Place transforms. At this point it was clear we needed help, and even clearer whose assistance we should seek – namely Herman Rubin. Herman's stature as a researcher, problem solver and font of mathematical knowledge was well known to the Stanford students.

Within a few days of having the problem described to him, Herman had sketched an elegant solution minus a few "obvious" details that Carl and I were able to supply in the next month or so. This eventually led to an Eaton–Morris–Rubin publication in the *Journal of Applied Probability*. During this collaboration, I was struck with Herman's willingness to share his considerable gifts with two fledgling researchers. In the succeeding years it has become clear to me that this is an essential part of his many contributions to our discipline. Thank you Herman.

1. Introduction

This expository paper is concerned primarily with some techniques for trying to evaluate the formal Bayes method of solving decision problems. Given a parametric model and an improper prior distribution, the method has two basic steps:

1. Compute the formal posterior distribution (proper) for the parameter given the data (assuming this exists)

2. Use the formal posterior to solve the "no data" version of the decision problem.

This two step process produces a decision rule whose properties, both desirable and undesirable, can be used in the assessment of the posterior distribution and hence

[1]School of Statistics, University of Minnesota, 224 Church Street S. E., Minneapolis, MN 55455, USA. e-mail: eaton@stat.umn.edu
Keywords and phrases: formal Bayes rules, admissibility, Markov chains.
AMS 2000 subject classifications: 62A01, 62C15, 62F15.

the improper prior. Typically, when frequentist measures of assessment are proposed, they often include some discussion of admissibility (or almost admissibility) for the formal Bayes rules obtained from the posterior. However, there is a delicate balance that arises immediately. If only a few decision problems are considered in the assessment, then the evidence may not be very convincing that the posterior is suitable since admissibility is, by itself, a rather weak optimality property. On the other hand, even in simple situations with appealing improper prior distributions, it is certainly possible that there are interesting decision problems where formal Bayes solutions are inadmissible (for example, see Blackwell (1951), Eaton (1992, Example 7.1), and Smith (1994)).

One approach to the above problem that has yielded some interesting and useful results is based on estimation problems with quadratic loss. In this case, formal Bayes decision rules are just the posterior means of the functions to be estimated and risk functions are expected mean squared error. Conditions for admissibility, obtained from the Blyth–Stein method (see Blyth (1951) and Stein (1955)), involve what is often called the integrated risk difference (IRD). In the case of quadratic loss estimation, various techniques such as integration by parts or non-obvious applications of the Cauchy–Schwarz inequality applied to the IRD, sometimes yield expressions appropriate for establishing admissibility (for example, see Karlin (1958), Stein (1959), Zidek (1970), and Brown and Hwang (1982)). These might be described as "direct analytic techniques."

In the past thirty years or so, two rather different connections have been discovered that relate quadratic loss estimation problems to certain types of "recurrence problems." The first of these appeared in Brown (1971) who applied the Blyth–Stein method to the problem of establishing the admissibility of an estimator of the mean vector of a p-dimensional normal distribution with covariance equal to the identity matrix. The loss function under consideration was the usual sum of squared errors. In attempting to verify the Blyth–Stein condition for a given estimator δ, Brown showed that there corresponds a "natural" diffusion process, although this connection is far from obvious. However, the heuristics in Section 1 of Brown's paper provide a great deal of insight into the argument. A basic result in Brown (1971) is that the estimator δ is admissible iff the associated diffusion is recurrent. This result depends on some regularity conditions on the risk function of δ, but holds in full generality when the risk function of δ is bounded. The arguments in Brown's paper depend to some extent on the underlying multivariate normal sampling model. Srinivasan (1981) contains material related to Brown (1971). The basic approach in Brown has been extended to the Poisson case in Johnstone (1984, 1986) where the diffusion is replaced by a birth and death process. A common feature of the normal and Poisson problems is that the associated continuous time stochastic process whose recurrence implies admissibility, are defined on the sample space (as opposed to the parameter space) of the estimation problem. In addition the inference problems under consideration are the estimation of the "natural" parameters of the model. Brown (1979) describes some general methods for establishing admissibility of estimators. These methods are based on the ideas underlying the admissibility–recurrence connection described above.

Formal Bayes methods are the focus of this paper. Since the posterior distribution is the basic inferential object in Bayesian analysis, it seems rather natural that evaluative criteria will involve this distribution in both proper and improper prior contexts. As in Brown (1971), just why "recurrence problems" arise in this context is far from clear. Briefly, the connection results from using admissibility in quadratic loss estimation problems to assess the suitability of the posterior distribution. In particular, if the posterior distribution of θ given the data x is $Q(d\theta|x)$ (depending, of course, on a model and an improper prior), then the formal Bayes estimator of any bounded function of θ, say $\phi(\theta)$, is the posterior mean of $\phi(\theta)$, say

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

It was argued in Eaton (1982, 1992) that the "admissibility" of $\hat{\phi}$ for all bounded ϕ constituted plausible evidence that the formal posterior might be suitable for making inferences about θ. To connect the admissibility of $\hat{\phi}$ to recurrence, first observe that when $\phi_A(\theta) = I_A(\theta)$ is an indicator function of a subset A of the parameter space, then the formal Bayes estimator

$$\hat{\phi}_A(x) = Q(A|x)$$

is the posterior probability of A. If η denotes the "true value of the model parameter" from which X was sampled, then the expected value (under the model) of the estimator $Q(A|X)$ is

$$R(A|\eta) = \mathcal{E}_\eta Q(A|X). \tag{1.1}$$

Next, observe that R in (1.1) is a transition function defined on the parameter space Θ of the problem. Thus, R induces a discrete time Markov chain whose state space is Θ. The remainder of this paper is devoted to a discussion of the following result.

Theorem 1.1. *If the Markov chain on Θ defined by R in* (1.1) *is "recurrent," then $\hat{\phi}$ is "admissible" for each bounded measurable ϕ when the loss is quadratic.*

Because Θ is allowed to be rather general, the recurrence of the Markov chain has to be defined rather carefully – this is the reason for the quotes on recurrent. As in Brown (1971), what connects the decision theoretic aspects of the problem to the Markov chain is the Blyth–Stein technique – and this yields what is often called "almost admissibility." Thus, the quotes on admissibility.

The main goal of this paper is to explain why Theorem 1.1 is correct by examining the argument used to prove the result. The starting point of the argument is that the Blyth–Stein condition that involves the IRD provides a sufficient condition for admissibility. Because this condition is somewhat hard to verify directly, it is often the case that a simpler condition is provided via an application of the Cauchy–Schwarz Inequality. In the development here, this path leads rather naturally to a mathematical object called a Dirichlet form. Now, the connection between the resulting Dirichlet form, the associated chain with the transition function R in Theorem 1.1, and the recurrence of the chain is fairly easy to explain.

In brief, this paper is organized as follows. In Section 2, the Blyth–Stein condition is described and the basic inequality that leads to the associated Dirichlet form is presented. In Section 3 the background material (mainly from the Appendix in Eaton (1992)) that relates the Markov chain to the Dirichlet form is described. The conclusion of Theorem 1.1 is immediate once the connections above are established.

The application of Theorem 1.1 in particular examples is typically not easy – primarily because establishing the recurrence of a non-trivial Markov chain is not easy. Examples related to the Pitman estimator of a translation parameter are discussed in Section 4. The fact that the Chung-Fuchs (1951) Theorem is used here supports the contention that interesting examples are not routine applications of general theory. Also in Section 4, a recent result of Lai (1996) concerning the multivariate normal translation model is described.

A detailed proof of Theorem 3.2 is given in an appendix to this paper. The conclusion of Theorem 3.2 is hinted at in Eaton (1992), but a rigorous proof is rather more involved than I originally believed it would be. Thus the careful proof here.

Although the Markov chain of interest here has the parameter space as its state space, some interesting related work of Hobert and Robert (1999) use a related chain on the sample space in some examples where the two spaces are both subsets of the real line.

2. The Blyth–Stein condition

Here are some basic assumptions that are to hold throughout this paper. The sample space $\mathcal{X}$ and the parameter space Θ are both Polish spaces with their respective σ-algebras of Borel sets. All functions are assumed to be appropriately measurable. The statistical model for $X \in \mathcal{X}$ is $\{P(\cdot|\theta)|\theta \in \Theta\}$ and the improper prior distribution ν is assumed to be σ-finite on the Borel sets of Θ. The *marginal measure* M on $\mathcal{X}$ is defined by

$$M(B) = \int_\Theta P(B|\theta)\nu(d\theta). \tag{2.1}$$

Because $\nu(\Theta) = +\infty$ it is clear that $M(\mathcal{X}) = +\infty$. However, in some interesting examples, the measure M is not σ-finite and this prevents the existence of a formal posterior distribution [For example, look at $\mathcal{X} = \{0, 1, \cdots, m\}$, the model is Binomial (m, θ) and $\nu(d\theta) = [\theta(1-\theta)]^{-1}d\theta$ on (0,1). No formal posterior exists here]. In all that follows the measure M is assumed to be σ-finite. In this case, there exists a proper conditional distribution $Q(d\theta|x)$ for θ given $X = x$ which satisfies

$$P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx). \tag{2.2}$$

Equation (2.2) means that the two joint measures on $\mathcal{X} \times \Theta$ agree. Further, $Q(\cdot|x)$ is unique almost everywhere M. For more discussion of this, see Johnson (1991).

Given the formal posterior, $Q(\cdot|x)$, the formal Bayes estimator for any bounded function $\phi(\theta)$ when the loss is quadratic is the posterior mean

$$\hat{\phi}(x) = \int \phi(\theta)Q(d\theta|x). \tag{2.3}$$

The risk function of this estimator is

$$R(\hat{\phi}, \theta) = \mathcal{E}_\theta\big[\hat{\phi}(X) - \phi(\theta)\big]^2 \tag{2.4}$$

where $\mathcal{E}_\theta$ denotes expectation with respect to the model. Because ϕ is bounded, $\hat{\phi}$ exists and $R(\hat{\phi}, \theta)$ is a bounded function of θ. The bounded assumption on ϕ simplifies the discussion enormously and allows one to focus on the essentials of the admissibility-recurrence connection. For a version of this material that is general enough to handle the estimation of unbounded ϕ's, see Eaton (2001).

The appropriate notion of "admissibility" for our discussion here is captured in the following definition due to C. Stein.

Definition 2.1. The estimator $\hat{\phi}$ is *almost-ν-admissible* if for any other estimator $t(X)$ that satisfies

$$R(t, \theta) \leq R(\hat{\phi}, \theta) \quad \text{for all } \theta, \tag{2.5}$$

the set

$$B = \big\{\theta|R(t, \theta) < R(\hat{\phi}, \theta)\big\} \tag{2.6}$$

has ν-measure zero.

Definition 2.2. The formal posterior $Q(\cdot|x)$ is *strongly admissible* if the estimator $\hat{\phi}$ is almost-ν-admissible for every bounded function ϕ.

The notion of strong admissibility is intended to capture a robustness property of the formal Bayes method across problems – at least for quadratic loss estimation problems when ϕ is bounded. The soft argument is that $Q(\cdot|x)$ cannot be too badly behaved if $\hat{\phi}$ is almost-ν-admissible for all bounded ϕ.

To describe a convenient version of the Blyth–Stein conditions for almost-ν-admissibility, consider a bounded function $g \geq 0$ defined on Θ and satisfying

$$0 < \int g(\theta)\nu(d\theta) \equiv c < +\infty. \tag{2.7}$$

Now $\nu_g(d\theta) = g(\theta)\nu(d\theta)$ is a finite measure on Θ so we can write

$$P(dx|\theta)\nu_g(d\theta) = \tilde{Q}_g(d\theta|x)M_g(dx) \tag{2.8}$$

where M_g is the marginal measure defined by

$$M_g(dx) = \int P(dx|\theta)\nu_g(d\theta). \tag{2.9}$$

Of course, $\tilde{Q}_g(d\theta|x)$ is a version of the conditional distribution of θ given $X = x$ when the proper prior distribution of θ is $c^{-1}\nu_g$. Setting

$$\hat{g}(x) = \int g(\theta)Q(d\theta|x), \tag{2.10}$$

it is not hard to show that

$$M_g(dx) = \hat{g}(x)M(dx). \tag{2.11}$$

Since the set $\{x|\hat{g}(x) = 0\}$ has M_g-measure zero, it follows that a version of $\tilde{Q}_g(d\theta|x)$ is

$$Q_g(d\theta|x) = \begin{cases} \dfrac{g(\theta)}{\hat{g}(x)}Q(d\theta|x), & \text{if} \quad \hat{g}(x) > 0, \\ Q(d\theta|x), & \text{if} \quad \hat{g}(x) = 0. \end{cases} \tag{2.12}$$

In all that follows, (2.12) is used as the conditional distribution of θ given $X = x$ when the prior distribution is ν_g.

Now, the Bayes estimator for $\phi(\theta)$, given the posterior (2.12), is

$$\hat{\phi}_g(x) = \int \phi(\theta)Q_g(d\theta|x) \tag{2.13}$$

whose risk function is

$$R(\hat{\phi}_g, \theta) = \mathcal{E}_\theta\big[\hat{\phi}_g(X) - \phi(\theta)\big]^2. \tag{2.14}$$

The so called *Integrated Risk Difference*,

$$IRD(g) = \int \big[R(\hat{\phi}, \theta) - R(\hat{\phi}_g, \theta)\big]g(\theta)\nu(d\theta) \tag{2.15}$$

plays a key role in the Blyth–Stein condition for the almost-ν-admissibility of $\hat{\phi}$. To describe this condition, consider a measurable set $C \subseteq \Theta$ with $0 < \nu(C) < +\infty$ and let

$$U(C) = \left\{ g \geq 0 \;\middle|\; \begin{array}{l} g \text{ is bounded}, \quad g(\theta) \geq 1 \quad \text{for } \theta \in C, \\ \text{and } \displaystyle\int g(\theta)\nu(d\theta) < +\infty \end{array} \right\} \tag{2.16}$$

.

Theorem 2.1 (Blyth–Stein). *Assume*

$$\begin{cases} \text{There is a sequence of sets } C_i \subseteq C_{i+1} \subseteq \Theta, \quad i = 1, \cdots \quad \text{with} \\ 0 < \nu(C_i) < +\infty \quad \text{and} \quad C_i \nearrow \Theta \quad \text{so that} \\ \inf_{g \in U(C_i)} IRD(g) = 0 \quad \text{for } i = 1, 2, \cdots \end{cases} \tag{2.17}$$

Then $\hat{\phi}$ is almost-ν-admissible.

The proof of this well known result is not repeated here. The usual interpretation of Theorem 2.1 is that when $\hat{\phi}$ is "close enough to a proper Bayes rule $\hat{\phi}_g$" then $\hat{\phi}$ is almost-ν-admissible, but the notion of closeness is at best rather vague.

A possible first step in trying to apply Theorem 2.1 is to find a tractable (and fairly sharp) upper bound for $IRD(g)$ in (2.15). Here is the key inequality that allows one to see eventually why "recurrence" implies strong-admissibility.

Theorem 2.2. *For a real valued measurable function h defined on Θ, let*

$$\Delta(h) = \iiint (h(\theta) - h(\eta))^2 Q(d\theta|x)Q(d\eta|x)M(dx). \tag{2.18}$$

Then for each bounded function ϕ, there is constant K_ϕ so that

$$IRD(g) \le K_\phi \Delta\left(\sqrt{g}\,\right), \tag{2.19}$$

for all bounded non-negative g satisfying $\int g(\theta)\nu(d\theta) < +\infty$.

Proof. A direct proof of (2.19) using the Cauchy–Schwarz Inequality follows. First, let $A = \{x|\hat{g}(x) > 0\}$ and recall that A^c has M_g measure zero. Thus,

$$\begin{aligned} IRD(g) &= \int_{\mathcal{X}}\int_\Theta \left[\left(\hat{\phi}(x) - \phi(\theta)\right)^2 - \left(\hat{\phi}_g(x) - \phi(\theta)\right)^2\right] P(dx|\theta)g(\theta)\nu(d\theta) \\ &= \int_{\mathcal{X}}\int_\Theta \left[\left(\hat{\phi}(x) - \phi(\theta)\right)^2 - \left(\hat{\phi}_g(x) - \phi(\theta)\right)^2\right] Q_g(d\theta|x)M_g(dx) \\ &= \int_A \left(\hat{\phi}(x) - \hat{\phi}_g(x)\right)^2 \hat{g}(x)M(dx) \\ &= \int_A \left[\int_\Theta \phi(\theta)\left(1 - \frac{g(\theta)}{\hat{g}(x)}\right)Q(d\theta|x)\right]^2 \hat{g}(x)M(dx) \\ &= \int_A \frac{1}{\hat{g}(x)}\left[\int_\Theta \phi(\theta)\left(g(\theta) - \hat{g}(x)\right)Q(d\theta|x)\right]^2 M(dx). \end{aligned} \tag{2.20}$$

A bit of algebra shows that for each x,

$$\begin{aligned} &\int_\Theta \phi(\theta)\left(g(\theta) - \hat{g}(x)\right)Q(d\theta|x) \\ &\quad = \frac{1}{2}\iint \left(\phi(\theta) - \phi(\eta)\right)\left(g(\theta) - g(\eta)\right)Q(d\theta|x)Q(d\eta|x). \end{aligned}$$

Using the non-negativity of g and the Cauchy–Schwarz inequality we have

$$\begin{aligned} &\left|\iint \left(\phi(\theta) - \phi(\eta)\right)\left(g(\theta) - g(\eta)\right)Q(d\theta|x)Q(d\eta|x)\right| \\ &\quad \le W(x) \cdot \left[\iint \left(\sqrt{g(\theta)} - \sqrt{g(\eta)}\right)^2 Q(d\theta|x)Q(d\eta|x)\right]^{\frac{1}{2}} \end{aligned}$$

where

$$W^2(x) = \iint (\phi(\theta) - \phi(\eta))^2 \left(\sqrt{g(\theta)} + \sqrt{g(\eta)}\right)^2 Q(d\theta|x)Q(d\eta|x).$$

Since ϕ is bounded, say $|\phi(\theta)| \leq c_0$, and since $(\sqrt{g(\theta)} + \sqrt{g(\eta)}\,)^2 \leq 2(g(\theta) + g(\eta))$, we have

$$W^2(x) \leq 4c_0^2 \hat{g}(x).$$

Substituting these bounds into the final expression in (2.20) yields

$$\begin{aligned} IRD(g) &\leq 4c_0^2 \int_A \iint \left(\sqrt{g(\theta)} - \sqrt{g(\eta)}\right)^2 Q(d\theta|x)Q(d\eta|x)M(dx) \\ &\leq 4c_0^2 \Delta\left(\sqrt{g}\,\right). \end{aligned}$$

Setting $K_\phi = 4c_0^2$ yields the result. $\square$

Combining Theorem 2.1 and Theorem 2.2 gives the main result of this section.

Theorem 2.3. *Assume*

$$\begin{cases} \textit{There is a sequence of increasing sets } C_i \subseteq \Theta, \quad i = 1, 2, \ldots \\ \textit{with } 0 < \nu(C_i) < +\infty \quad \textit{and} \quad C_i \nearrow \Theta \quad \textit{so that} \\ \inf\limits_{g \in U(C_i)} \Delta\left(\sqrt{g}\right) = 0, \ \textit{for each } i. \end{cases} \tag{2.21}$$

Then $Q(d\theta|x)$ is strongly admissible.

Proof. When (2.21) holds, inequality (2.19) shows that (2.17) holds for each bounded measurable ϕ. Then $Q(d\theta|x)$ is strongly admissible. $\square$

It should be noted that the assumption (2.21) does not involve ϕ (as opposed to assumption (2.17)). Thus the conditions for strong admissibility involve the behavior of Δ. It is exactly the functional Δ that provides the connection between (2.21) and the "recurrence" of the Markov chain with transition function R in (1.1).

To put the material of the next section in perspective, it is now useful to isolate some of the essential features of the decision theory problem described above – namely, under what conditions on the given model $P(dx|\theta)$ and the improper prior $\nu(d\theta)$ with the formal posterior $Q(d\theta|x)$ be strongly admissible? A basic ingredient in our discussion will be the transition function

$$R(d\theta|\eta) = \int Q(d\theta|x)P(dx|\eta) \tag{2.22}$$

introduced in Section 1. A fundamental property of R is its symmetry with respect to ν – that is, the measure on $\Theta \times \Theta$ defined by

$$s(d\theta, d\eta) = R(d\theta|\eta)\nu(d\eta). \tag{2.23}$$

is a symmetric measure in the sense that

$$\begin{aligned} s(A \times B) &= \iint I_A(\theta) I_B(\eta) R(d\theta|\eta)\nu(d\eta) \\ &= s(B \times A) \end{aligned} \tag{2.24}$$

for Borel subsets A and B of Θ. This is easily established from the definition of R. It is this symmetry that drives the theory of the next section and allows us to connect the behavior of Δ, namely

$$\Delta(h) = \iint (h(\theta) - h(\eta))^2 \, R(d\theta|\eta)\nu(d\eta), \tag{2.25}$$

to the "recurrence" of the Markov chain defined by R. The expression (2.25) for Δ follows from (2.18) and the disintegration formula (2.2).

Also, note that ν is a stationary measure for R – that is,

$$\int R(A|\eta)\nu(d\eta) = \nu(A) \tag{2.26}$$

for all Borel sets A. This is an easy consequence of the symmetry of s in (2.23).

The discussion in the next section begins with an abstraction of the above observations. Much of the discussion is based on the Appendix in Eaton (1992).

Here is the standard Pitman example that gives a concrete non-trivial example of what the above formulation yields.

Example 2.1. Consider $X_1, \ldots, X_n$ that are independent and identically distributed random vectors in R^p with a density $f(x - \theta)$ (with respect to Lebesgue measure). Thus $\Theta = R^p$ and the model for $X = (X_1, \ldots, X_n)$ is

$$P(dx|\theta) = \prod_{i=1}^{n} f(x_i - \theta) dx_i$$

on the sample space $\mathcal{X} = R^{pn}$. With dx as Lebesgue measure on $\mathcal{X}$, the density of $P(dx|\theta)$ with respect to dx is

$$p(x|\theta) = \prod_{i=1}^{n} f(x_i - \theta).$$

Next take $\nu(d\theta) = d\theta$ on $\Theta = R^p$ and assume, for simplicity, that

$$m(x) = \int_{R^p} p(x|\theta) d\theta$$

is in $(0, \infty)$ for all x. Then a version of "$Q(d\theta|x)$" is

$$Q(d\theta|x) = \frac{p(x|\theta)}{m(x)} \, d\theta.$$

Thus the transition function R is given by

$$R(d\theta|\eta) = \left(\int_{\mathcal{X}} \frac{p(x|\theta) p(x|\eta)}{m(x)} \, dx \right) d\theta.$$

Therefore,

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

where the density $r(\cdot|\eta)$ is

$$r(\theta|\eta) = \int_{\mathcal{X}} \frac{p(x|\theta) \; p(x|\eta)}{m(x)} \, dx.$$

Now, it is easy to show that for each vector $u \in R^p$,

$$r(\theta + u|\eta + u) = r(\theta|\eta)$$

so that r is only a function of $\theta - \eta$, say

$$t(\theta - \eta) = r(\theta - \eta|0).$$

Further routine calculations give

$$\begin{cases} t(u) = t(-u) & \text{for } u \in R^p \\ \int t(u)du = 1 \end{cases}$$

In summary then, for the translation model with $d\theta$ as the improper prior distribution, the induced transition function is

$$R(d\theta|\eta) = t(\theta - \eta)d\theta$$

and t is a symmetric density function on R^p. We will return to this example later.

3. Symmetric Markov chains

Here, a brief sketch of symmetric Markov chain theory, recurrence and Dirichlet forms is given. The purpose of this section is two-fold – first to explain the relationship between recurrence and the Dirichlet form and second to relate this to the strong admissibility result of Theorem 2.3.

Let Y be a Polish space with the Borel σ-algebra $\mathcal{B}$ and consider a Markov Kernel $R(dy|z)$ on $\mathcal{B} \times Y$. Also let λ be a non-zero σ-finite measure on $\mathcal{B}$.

Definition 3.1. The kernel $R(dy|z)$ is λ-*symmetric* if the measure

$$\alpha(dy, dz) = R(dy|z)\lambda(dz) \tag{3.1}$$

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

Typically, R is called symmetric without reference to λ since λ is fixed in most discussions. As the construction in Section 2 shows, interesting examples of symmetric kernels abound in statistical decision theory. In all that follows, it is assumed that R is λ-symmetric. Note that the assumption of σ-finiteness for λ is important.

Given a λ-symmetric R, consider a real valued measurable function h and let

$$\Delta(h) = \iint (h(y) - h(z))^2 R(dy|z)\lambda(dz). \tag{3.2}$$

The quadratic form Δ (or sometimes $\frac{1}{2}\Delta$) is often called a Dirichlet form. Such forms are intimately connected with continuous time Markov Process Theory (see Fukushima et al (1994)) and also have played a role in some work on Markov chains (for example, see Diaconis and Strook (1991)). A routine calculation using the symmetry of R shows that

$$\Delta(h) \leq 4 \int h^2(y)\lambda(dy) \tag{3.3}$$

so Δ is finite for $h \in L_2(\lambda)$, the space of λ-square integrable functions.

Now, given $R(dy|z)$, there is a Markov chain with state space Y and transition function $R(dy|z)$. More precisely, consider the path space $\mathcal{W} = Y^{\infty} = Y \times Y \times \cdots$ with the usual product σ-algebra. Given the initial value w_0, there is a Markov chain $W = (w_0, W_1, W_2, \ldots)$ so that $R(dw_{i+1}|w_i)$ is the conditional distribution of W_{i+1} given $W_i = w_i$, for $i = 0, 1, 2, \ldots$. The unique probability measure on path space that is consistent with this *Markov specification*, is denoted by $S(\cdot|w_0)$.

Because the space Y is rather general, the definition of recurrence has to be selected with some care. The reader should note that neither irreducibility nor periodicity occur in the discussion that follows (see Meyn and Tweedie (1993) for a discussion of such things in the general state space case). Let $C \subseteq Y$ satisfy $0 < \lambda(C) < +\infty$. Such measurable sets are called λ-*proper*. Define the random variable T_C on W as follows:

$$T_C = \begin{cases} +\infty & \text{if } W_i \notin C \quad \text{for } i = 1, 2, \ldots \\ 1 & \text{if } W_1 \in C \\ n & \text{if } W_n \in C \quad \text{for some } n \geq 2 \quad \text{and} \\ & \quad W_i \notin C \quad \text{for } i = 1, \ldots, n-1 \end{cases} \tag{3.4}$$

Then T_C ignores the starting value of the chain and records the first hitting time of C for times greater than 0. The set

$$B_C = \{T_C < +\infty\} \tag{3.5}$$

is the event where the chain hits C at some time after time 0.

Definition 3.2. A λ-proper set $C \subseteq Y$ is called *locally* -λ-*recurrent* if the set

$$B_0 = \{w_0 \in C | S(B_C|w_0) < 1\}$$

has λ-measure zero.

Definition 3.3. A λ-proper set $C \subseteq Y$ is called λ-*recurrent* if the set

$$B_1 = \{w_0 | S(B_C|w_0) < 1\}$$

has λ-measure zero.

In other words, C is locally-ν-recurrent if whenever the chain starts in C, it returns to C w.p.1, except for a set of starting values of λ-measure zero. It is this notion of recurrence that is most relevant for admissibility considerations. Of course, C is λ-recurrent if the chain hits C no matter where it starts, except for a set of starting values of λ-measure zero. This second notion is closer to traditional ideas related to recurrence.

To describe the connection between the Dirichlet form Δ and local-λ-recurrence, consider

$$V(C) = \left\{ h \in L_2(\lambda) \,\middle|\, \begin{array}{l} h \geq 0, \quad h(y) \geq 1 \quad \text{for } y \in C, \\ h \text{ is bounded} \end{array} \right\}. \tag{3.6}$$

Note that $U(C)$ in (2.16) and $V(C)$ are in one-to-one correspondence via the relation $h(y) = \sqrt{g(y)},\ y \in Y$.

Theorem 3.1. *For a λ-proper set C,*

$$\inf_{h \in V(C)} \Delta(h) = 2 \int_C (1 - S(B_C|w))\lambda(dw). \tag{3.7}$$

A proof of this basic result can be found in Appendix 2 of Eaton (1992). From (3.7), it is immediately obvious that C is a locally-λ-recurrent set iff the inf over $V(C)$ of the Dirichlet form Δ is zero.

Definition 3.4. The Markov chain $W = (W_0, W_1, W_2, \ldots)$ is *locally-λ-recurrent* if each λ-proper set C is locally-λ-recurrent.

In applications, it is useful to have some conditions that imply local-λ-recurrence since the verification that every λ-proper C is locally-λ-recurrent can be onerous. To this end, we have

Theorem 3.2. *The following are equivalent:*

(i) *The Markov chain $W = (W_0, W_1, W_2, \ldots)$ is locally-λ-recurrent*

(ii) *There exists an increasing sequence of λ-proper sets C_i, $i = 1, \ldots$ such that $C_i \longrightarrow Y$ and each C_i is locally-λ-recurrent*

Proof. Obviously (i) implies (ii). The converse is proved in the appendix. □

In a variety of decision theory problems, it is often sufficient to find *one* set B_0 that is "recurrent" in order to establish "admissibility." For an example of the "one-set phenomenon," see Brown and Hwang (1982). In the current Markov chain context, here is a "one-set" condition that implies local-λ-recurrence for the chain W.

Theorem 3.3. *Suppose there exists a λ-proper set B_0 that is λ-recurrent* (*see Definition* 3.3). *Then the Markov chain W is locally-ν-recurrent.*

Proof. Since λ is σ-finite, there is a sequence of increasing λ-proper sets $B_i, i = 1, 2, \ldots$ such that $B_i \longrightarrow Y$. Let $C_i = B_i \cup B_0$, $i = 1, 2, \ldots$ so the sets C_i are λ-proper, are increasing, and $C_i \longrightarrow Y$. The first claim is that each C_i is locally-λ-recurrent. To see this, let N be the λ-null set where $S(T_{B_0} < +\infty | w) < 1$. Then for $w \notin N$, the chain hits B_0 w.p.1 after time 0 when $W_0 = w$. Thus, for $w \notin N$, the chain hits $B_0 \cup B_i$ w.p.1 after time 0 when $W_0 = w$. Therefore the set $C_i = B_0 \cup B_i$ is locally-λ-recurrent. By Theorem 3.2, W is locally-λ-recurrent. □

The application of the above results to the strong-admissibility problem is straightforward. In the context of Section 2, consider a given model $P(dx|\theta)$ and a σ-finite improper prior distribution $\nu(d\theta)$ so that the marginal measure M in (2.1) is σ-finite. This allows us to define the transition $R(d\theta|\eta)$ in (2.22) that is ν-symmetric. Therefore the above theory applies to the Markov chain $W = (W_0, W_1, W_2, \ldots)$ on Θ^∞ defined by $R(d\theta|\eta)$. Here is the main result that establishes "Theorem 1.1" stated in the introductory section of this paper.

Theorem 3.4. *Suppose the Markov chain W with state space Θ and transition function $R(d\theta|\eta)$ is locally-ν-recurrent. Then the posterior distribution $Q(d\theta|x)$ defined in* (2.2) *is strongly-admissible.*

Proof. Because W is locally-ν-recurrent, the infimum in (3.7) is zero for each ν-proper set C. This implies that condition (2.21) in Theorem 2.3 holds. Thus, $Q(d\theta|x)$ is strongly admissible. □

Of course Theorem 3.2 makes it a bit easier to show W is locally-ν-recurrent, while Theorem 3.3 provides an extremely useful sufficient condition for this property of W. An application is given in the next section.

4. Examples

Here we focus on two related examples. The first is based on the Pitman model introduced in Example 2.1. In this case, the induced Markov chain is a random walk on the parameter space, and as is well known, under rather mild moment conditions for dimensions $p = 1$ and $p = 2$, the random walks are recurrent. But for $p \geq 3$, there are no recurrent random walks on R^p that have densities with respect to Lebesque measure. Of course this parallels what decision theory yields for admissibility of the Pitman estimator of a mean vector – admissibility for $p = 1$ and $p = 2$ (under mild moment conditions) and inadmissibility in many examples when $p \geq 3$. The results here do not concern estimation of a mean vector, but rather involve the strong admissibility of the posterior, and again the dimension phenomenon prevails.

The second example is from the thesis of Lai (1996) and concerns the p-dimensional normal distribution with an unknown mean vector and the identity covariance matrix. In essence, Lai's results provide information regarding a class of improper priors that yield strong admissibility when the parameter space is R^p. Even in the case of the normal distribution there are many open questions concerning the "inappropriateness" of the posterior when the improper prior is $d\theta$ on R^p, $p \geq 3$.

Example 4.1 (continued). As shown in Section 2, the induced transition function on the parameter space $\Theta = R^p$ is given by

$$R(d\theta|\eta) = t(\theta - \eta)\, d\theta \tag{4.1}$$

where the density function t is defined in Example 2.1. The Markov chain induced by R is just a random walk on R^p. When $p = 1$, the results of Chung and Fuchs (1951) apply directly. In particular, if $p = 1$ and

$$\int_{-\infty}^{\infty} |u| t(u)\, du < +\infty, \tag{4.2}$$

then the Markov chain is recurrent and so the posterior distribution in this case is strongly admissible. A sufficient condition for (4.2) to hold is that the original density function f in Example 2.1 has a finite mean (see Eaton (1992) for details).

When $p = 2$, a Chung–Fuchs-like argument also applies (see Revuz (1984)). In particular, if

$$\int_{R^2} \|u\|^2 t(u)\, du < +\infty, \tag{4.3}$$

then the Markov chain on R^2 is recurrent so strong admissibility obtains. Again, it is not too hard to show that the existence of second moments under f in Example 2.1 imply that (4.3) holds. These results for $p = 1, 2$ are suggested by the work of Stein (1959) and James and Stein (1961).

For $p \geq 3$, the Markov chain obtained from R in (4.1) can never be recurrent (see Guivarc'h, Keane, and Roynette (1977)) suggesting that the posterior distribution obtained from the improper prior $d\theta$ on $\Theta = R^p$ is suspect. At present, the question of "inadmissibility" of the posterior when $p \geq 3$ remains largely open. This ends our discussion of Example 2.1.

Example 4.2. The material in this example is based on the work of Lai (1996). Suppose X is a p-dimensional random vector with a normal distribution $N_p(\theta, I_p)$.

Here $\theta \in \Theta = R^p$ is unknown and the covariance matrix of X is the $p \times p$ identity. Consider an improper prior distribution of the form

$$\nu(d\theta) = (a + \|\theta\|^2)^\alpha \, d\theta \tag{4.4}$$

where the constant a is positive and α is a real number. In this setting Lai proved the following.

Theorem 4.1 (Lai (1996)). *If $\alpha \in (-\infty, -\frac{p}{2} + 1]$, then the posterior distribution for θ is strongly admissible.*

The above follows from the more general Theorem 5.3.3 in Lai (1996), but well illustrates the use of the Markov chain techniques. Lai's argument consists of proving that for the range of α indicated, the induced Markov chain on Θ is locally-ν-recurrent so strong admissibility obtains. In fact, the Markov chain techniques developed by Lai to handle this example include extensions of some recurrence criteria of Lamperti (1960) and an application of Theorem 3.3 given above. Although the class of priors in (4.4) is quite small, the extension of Theorem 4.1 to other improper priors can be verified via Remark 3.3 in Eaton (1992). In this particular example, Remark 3.3, coupled with Theorem 4.1, implies the following.

Theorem 4.2. *Consider a prior distribution ν of the form* (4.4) *with $\alpha \in (-\infty, -\frac{p}{2} + 1]$ and let $g(\theta)$ satisfy*

$$c \le g(\theta) \le \frac{1}{c} \quad \textit{for all } \theta$$

for some $c > 0$. Then the Markov chain induced by the prior distribution

$$\nu_g(d\theta) = g(\theta)\nu(d\theta) \tag{4.5}$$

is locally-ν-recurrent and the induced posterior distribution is strongly admissible.

For applications of Lai's ideas to the multivariate Poisson case, we refer the reader to Lai's thesis. This completes Example 4.1.

Appendix

Here we provide a proof of Theorem 3.2. To this end, consider a measurable subset $C \subseteq Y$ that is λ-proper and let

$$H(C) = \inf_{h \in V(C)} \Delta(h). \tag{A.1}$$

Also, let

$$V^*(C) = \{h | h \in V(C), \quad h(y) \in [0,1] \quad \text{for } y \in C^c\}.$$

The results in Appendix 2 of Eaton (1992) show that

$$H(C) = \inf_{h \in V^*(C)} \Delta(h). \tag{A.2}$$

Lemma 1.1. *Consider measurable subsets A and B of Y that are both λ-proper. If $A \subseteq B$, then*

$$H^{\frac{1}{2}}(A) \le H^{\frac{1}{2}}(B) \le H^{\frac{1}{2}}(A) + 2^{\frac{1}{2}} \lambda^{\frac{1}{2}}(B \cap A^c). \tag{A.3}$$

Proof. Since $V(A) \supseteq V(B)$, $H(A) \leq H(B)$ so the left hand inequality in (A.3) is obvious. For the right hand inequality, first note that $\Delta^{\frac{1}{2}}$ is a subadditive function defined on $L_2(\lambda)$– that is,

$$\Delta^{\frac{1}{2}}(h_1 + h_2) \leq \Delta^{\frac{1}{2}}(h_1) + \Delta^{\frac{1}{2}}(h_2). \tag{A.4}$$

A proof of (A.4) is given below. With $h_3 = h_1 + h_2$, (A.4) yields

$$\Delta^{\frac{1}{2}}(h_3) \leq \Delta^{\frac{1}{2}}(h_1) + \Delta^{\frac{1}{2}}(h_3 - h_1), \tag{A.5}$$

for h_1 and h_3 in $L_2(\lambda)$. Now consider $h \in V^*(A)$ and write

$$\tilde{h}(y) = h(y) + g(y)$$

where

$$g(y) = (1 - h(y)) I_{B \cap A^c}(y).$$

Then $\tilde{h} \in V^*(B)$ and (A.5) implies that

$$\Delta^{\frac{1}{2}}(\tilde{h}) \leq \Delta^{\frac{1}{2}}(h) + \Delta^{\frac{1}{2}}(g).$$

Thus,

$$H^{\frac{1}{2}}(B) \leq \Delta^{\frac{1}{2}}(h) + \Delta^{\frac{1}{2}}(g). \tag{A.6}$$

Because $g(y) \in [0, 1]$,

$$\begin{aligned} \Delta(g) &= \iint (g(y) - g(z))^2 R(dy|z)\lambda(dz) \\ &= 2\left[\int g^2(y)\lambda(dy) - \iint g(y)g(z)R(dy|z)\lambda(dz)\right] \\ &\leq 2\int_{B\cap A^c} g^2(y)\lambda(dy) \leq 2\lambda\big(B \cap A^C\big). \end{aligned}$$

Substituting this into (A.6) yields

$$H^{\frac{1}{2}}(B) \leq \Delta^{\frac{1}{2}}(h) + 2^{\frac{1}{2}}\lambda^{\frac{1}{2}}\big(B \cap A^c\big). \tag{A.7}$$

Since (A.7) holds for any $h \in V^*(A)$, the right hand inequality in (A.3) holds. This completes the proof. $\square$

The proof of (A.4) follows. For h_1 and h_2 in $L_2(\lambda)$, consider the symmetric bilinear form

$$< h_1, h_2 >= \int h_1(y)h_2(y)\lambda(dy) - \iint h_1(y)h_2(z)R(dy|z)\lambda(dz).$$

That $< \cdot, \cdot >$ is non-negative definite is a consequence of the symmetry of R and the Cauchy–Schwarz inequality:

$$\begin{aligned} &\left(\iint h_1(y)h_1(z)R(dy|z)\lambda(dz)\right)^2 \\ &\quad \leq \left(\iint h_1^2(y)R(dy|z)\lambda(dz)\right) \cdot \left(\iint h_1^2(z)R(dy|z)\lambda(dz)\right) \\ &\quad = \left(\int h_1^2(y)\lambda(dy)\right)^2. \end{aligned}$$

Thus $\|h\| =< h, h >^{\frac{1}{2}}$ is a semi-norm on $L_2(\lambda)$ and so is subadditive. Since $\Delta(h) = 2\|h\|^2$, inequality (A.4) holds.

The proof of Theorem 3.2 is now easy. Let C be any λ-proper set so $\lambda(C) < +\infty$ and let

$$E_i = C_i \cap C, \quad i = 1, 2, \ldots.$$

Since C_i is locally-λ-recurrent, $H(C_i) = 0$ so $H(E_i) = 0$ by Lemma A.1. Since $E_i \nearrow C$ and $\lambda(C) < +\infty$, we have

$$\lim_{i \longrightarrow \infty} \lambda(E_i) \longrightarrow \lambda(C)$$

and

$$\lim_{i \longrightarrow \infty} \lambda(C \cap E_i^c) \longrightarrow 0.$$

Applying (A.3) yields

$$H^{\frac{1}{2}}(C) \leq H^{\frac{1}{2}}(E_i) + 2^{\frac{1}{2}} \lambda(C \cap E_i^c).$$

The right hand side of this inequality converges to zero as $i \longrightarrow \infty$. Hence $H(C) = 0$. Since C was an arbitrary λ-proper set, the chain W is locally-ν-recurrent.

Acknowledgment

Many thanks to Jim Hobert, Tiefeng Jiang and Galin Jones for their valuable comments. Also special thanks to Anirban Das Gupta for his efforts on this Festschrift for Herman Rubin and his many comments on this contribution.

References

[1] Blackwell, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Statist.* **22**, 105–110. MR43418

[2] Blyth, C. R. (1951). On minimax statistical procedures and their admissibility. *Ann. Math. Statist.* **22**, 22–42. MR39966

[3] Brown, L. (1971). Admissible estimators, recurrent diffusions, and insolvable boundary value problems. *Ann. Math. Statist.* **42**, 855–904. MR286209

[4] Brown, L. (1979). A heuristic method for determining admissibility of estimators – with applications. *Ann. Statist.* **7**, 960–994. MR536501

[5] Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **1**, 205–230. New York: Academic Press. MR705290

[6] 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. MR40610

[7] Diaconis, P. and Strook, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Am. Appl. Probab.* **1**, 36–61. MR1097463

[8] Eaton, M. L. (1982). A method for evaluating improper prior distributions. In *Statistics Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **1**, 329–352. New York:Academic Press.

[9] Eaton, M. L. (1992). A statistical diptych: Admissible inferences- Recurrence of symmetric Markov chains. *Ann. Statist.* **20**, 1147–1179. MR1186245

[10] Eaton, M. L. (2001). Markov chain conditions for admissibility with quadratic loss. In *State of the Art in Statistics and Probability*, a Festschrift for Willem van Zwet edited by M. de Gunst, C. Klaassen, and A. van der Vaart, pp. 223–243. Volume **36** in the IMS Lecture Note Series, Beachwood, Ohio. MR1836563

[11] Fukushima, M., Oshima, Y., Takeda, M. (1994). *Dirichlet Forms and Symmetric Markov Processes.* New York: de Gruyter. MR1303354

[12] Guivarc'h, Y., Keane, M. and Roynette, B. (1977). *March aleatoires sur les Groupes de Lie. Lecture Notes in Math* **624**, New York: Springer. MR517359

[13] Hobert, J. and Robert, C. (1999). Eaton's Markov chain, its conjugate partner, and P-admissibility. *Ann. Statist.* **27**, 361–373. MR1701115

[14] James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1**, 361–380. Berkeley: University of California Press. MR133191

[15] Johnson, B. (1991). On the admissibility of improper Bayes inferences in fair Bayes decision problems. Ph.D. Thesis, University of Minnesota.

[16] Johnstone, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Ann. Statist.* **12**, 1173–1198. MR760682

[17] Johnstone, I. (1986). Admissible estimation, Dirichlet principles, and recurrence of birth-death chains on Z_+^p. *Probab. Theory Related Fields* **71**, 231–269. MR816705

[18] Karlin, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.* **29**, 406–436. MR124101

[19] Lai, W. (1996). Admissibility and the recurrence of Markov chains with applications. Ph.D. Thesis, University of Minnesota.

[20] Lamperti, J. (1960). Criteria for the recurrence or transience of stochastic processes I. *Journal of Math. Anal. Appl.* **1**, 314–330. MR126872

[21] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* New York: Springer-Verlag. MR1287609

[22] Revuz, D. (1984). *Markov Chains, 2nd Edition.* North Holland, Amsterdam. MR415773

[23] Smith, J. Q. (1994). The inadmissibility of certain Bayes decision rules under vague priors in location problems. *J. R. S. S. - B.* **56**, 543–548. MR1278226

[24] Srinivasan, C. (1981). Admissible generalized Bayes estimators and exterior boundary value problems. *Sankhya Ser. A.* **43**, 1–25. MR656265

[25] Stein, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Statist.* **76**, 518–522. MR70929

[26] Stein, C. (1959). The admissibility of Pitman's estimator of a single location parameter. *Ann. Math. Statist.,* **30**, 970–979. MR109392

[27] Zidek, J. V. (1970). Sufficient conditions for admissibility under squared error loss of formal Bayes estimators. *Ann. Math. Statist.* **41**, 446–456. MR261742