

# On Perturbations of Strongly Admissible Prior Distributions

Morris L. Eaton  
School of Statistics  
University of Minnesota  
Minneapolis, MN 55455 USA  
eaton@stat.umn.edu

James P. Hobert  
Department of Statistics  
University of Florida  
Gainesville, FL 32611 USA  
jhobert@stat.ufl.edu

Galin L. Jones  
School of Statistics  
University of Minnesota  
Minneapolis, MN 55455 USA  
galin@stat.umn.edu

December 2005 (revised August 2006)

## Abstract

Consider a parametric statistical model,  $P(dx|\theta)$ , and an improper prior distribution,  $\nu(d\theta)$ , that together yield a (proper) formal posterior distribution,  $Q(d\theta|x)$ . The prior is called *strongly admissible* if the generalized Bayes estimator of every bounded function of  $\theta$  is admissible under squared error loss. Eaton [5] used the Blyth-Stein Lemma to develop a sufficient condition, call it  $\mathcal{C}$ , for strong admissibility of  $\nu$ . Our main result says that, under mild regularity conditions, if  $\nu$  satisfies  $\mathcal{C}$  and  $g(\theta)$  is a bounded, non-negative function, then the *perturbed prior distribution*  $g(\theta)\nu(d\theta)$  also satisfies  $\mathcal{C}$  and is therefore strongly admissible. Our proof has three basic components: (i) Eaton's [5] result that the condition  $\mathcal{C}$  is equivalent to the *local recurrence* of the Markov chain whose transition function is  $R(d\theta|\eta) = \int Q(d\theta|x)P(dx|\eta)$ ; (ii) a new result for general state space Markov chains giving conditions under which local recurrence is equivalent to recurrence; and (iii) a new generalization of Hobert and Robert's [10] result that says Eaton's Markov chain is recurrent if and only if the chain with transition function  $\tilde{R}(dx|y) = \int P(dx|\theta)Q(d\theta|y)$  is recurrent. One important application of our results involves the construction of strongly admissible prior distributions for estimation problems with restricted parameter spaces.

---

*AMS 2000 subject classifications.* Primary 62C15; secondary 60J05

*Abbreviated title.* Perturbations of Strongly Admissible Priors

*Key words and phrases.* Admissibility, Dirichlet form, Estimation, Formal Bayes rule, Formal posterior distribution, Improper prior distribution, Irreducibility, Symmetric Markov chain, Recurrence, Restricted parameter space

# 1 Introduction

## 1.1 The problem

Consider a statistical decision problem involving a parametric statistical model  $P(dx|\theta)$  on a sample space  $\mathcal{X}$  and a  $\sigma$ -finite *improper* prior distribution  $\nu(d\theta)$  on a parameter space  $\Theta$ . Both  $\mathcal{X}$  and  $\Theta$  are assumed to be Polish spaces equipped with their Borel  $\sigma$ -algebras  $\mathcal{B}(\mathcal{X})$  and  $\mathcal{B}(\Theta)$ . Since the prior is improper, the existence of a proper posterior distribution is not guaranteed. However, when the marginal measure  $M$ , defined as

$$M(dx) = \int_{\Theta} P(dx|\theta)\nu(d\theta) ,$$

is  $\sigma$ -finite, a formal posterior distribution  $Q(d\theta|x)$  exists and is characterized by

$$P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx) . \quad (1)$$

This equality means that the two joint measures on  $\mathcal{X} \times \Theta$  given by the left and right-hand sides agree. For each  $x \in \mathcal{X}$ ,  $Q(\cdot|x)$  is a probability measure and for each  $C \in \mathcal{B}(\Theta)$ ,  $Q(C|\cdot)$  is a measurable function. See Eaton [4, 5] and the references therein for details regarding the existence and uniqueness of  $Q(d\theta|x)$ . Throughout this paper,  $M$  is assumed to be  $\sigma$ -finite.

The formal posterior distribution can be used to solve statistical decision problems by choosing actions to minimize posterior expected loss. In particular, the formal Bayes solution to the problem of estimating a bounded, real-valued function  $\gamma(\theta)$  under squared error loss is the estimator

$$\hat{\gamma}(x) = \int_{\Theta} \gamma(\theta)Q(d\theta|x) , \quad (2)$$

and the risk function of a generic estimator, say  $\delta$ , is its mean squared error; i.e.,

$$R(\delta, \theta) = \int_{\mathcal{X}} (\delta(x) - \gamma(\theta))^2 P(dx|\theta) .$$

We will judge estimators using the notion of almost-admissibility, which is now defined. Let  $\alpha$  be a non-trivial,  $\sigma$ -finite measure on  $(\Theta, \mathcal{B}(\Theta))$ .

**Definition 1.** *An estimator  $\delta$  is almost- $\alpha$ -admissible if for any estimator  $\delta'$  such that*

$$R(\delta', \theta) \leq R(\delta, \theta) \quad \forall \theta \in \Theta ,$$

*the set  $\{\theta \in \Theta : R(\delta', \theta) < R(\delta, \theta)\}$  has  $\alpha$ -measure zero.*

If the statistical model  $P(dx|\theta)$  and the improper prior  $\nu(d\theta)$  combine to yield a formal posterior distribution that generates (almost) admissible estimators (via (2)) for a large class of functions of  $\theta$ , then we might be willing to endorse  $\nu$  as a good “all purpose” prior to use in conjunction with  $P(dx|\theta)$ . This idea provides motivation for the following definition.

**Definition 2.** *The improper prior  $\nu(d\theta)$  (or, equivalently, the corresponding formal posterior  $Q(d\theta|x)$ ) is called strongly admissible if, for every bounded, measurable function  $\gamma : \Theta \rightarrow \mathbb{R}$ , the estimator  $\hat{\gamma}$  is almost- $\nu$ -admissible.*

Eaton [5] used the Blyth-Stein Lemma to develop a sufficient condition for strong admissibility of  $\nu$  and then went on to show that this sufficient condition is, in fact, equivalent to a recurrence property of an associated Markov chain defined in terms of  $P(dx|\theta)$  and  $\nu(d\theta)$ . Our results build on this theory, but this paper is not meant to be a review. For an expository look at this area, see Eaton [6, 8]. We begin with the sufficient condition for strong admissibility. First, define a (symmetric) measure on  $\Theta \times \Theta$  as

$$S(d\theta, d\eta) = \int_{\mathcal{X}} Q(d\theta|x)Q(d\eta|x)M(dx) .$$

Let  $L_2(\nu)$  be the set of  $\nu$ -square integrable functions defined on  $\Theta$ . For  $h \in L_2(\nu)$ , define

$$\Delta(h) = \frac{1}{2} \int_{\Theta} \int_{\Theta} (h(\theta) - h(\eta))^2 S(d\theta, d\eta) .$$

It's not hard to show that  $\Delta(h) < \infty$  for all  $h \in L_2(\nu)$ . A measurable set  $C \subset \Theta$  is called  $\nu$ -proper if  $0 < \nu(C) < \infty$ . For each  $\nu$ -proper set  $C$ , define

$$V(C) = \{h \in L_2(\nu) : h \text{ is bounded, } h(\theta) \geq I_C(\theta)\} .$$

Here is a sufficient condition for strong admissibility.

**Theorem 1.** *(Eaton, 1992) The prior  $\nu$  is strongly admissible if*

$$\inf_{h \in V(C)} \Delta(h) = 0 \text{ for each } \nu\text{-proper set } C . \quad (3)$$

Theorem 1 provides a single condition involving  $P$  and  $\nu$  that simultaneously implies the (almost) admissibility of a large class of formal Bayes estimators. (For an extension that can handle unbounded  $\gamma$ , see [7].) Moreover, if (3) holds, then it follows that a large class of *perturbations* of  $\nu$  also satisfy (3) and are therefore also strongly admissible. To be specific, consider a perturbation of  $\nu$  given by

$$\nu_g(d\theta) = g(\theta) \nu(d\theta) ,$$

where  $g : \Theta \rightarrow [0, \infty)$ . Assume that  $g$  is bounded away from 0 and away from  $\infty$ ; i.e., assume that there exists an  $\varepsilon > 0$  such that  $\varepsilon < g(\theta) < 1/\varepsilon$  for all  $\theta \in \Theta$ . Under this assumption,  $\nu_g$  is a  $\sigma$ -finite, improper prior and the corresponding marginal  $M_g(B) = \int_{\Theta} P(B|\theta) g(\theta) \nu(d\theta)$  is also  $\sigma$ -finite. Hence, the corresponding posterior  $Q_g(d\theta|x)$  exists and satisfies the following analogue of (1)

$$P(dx|\theta) g(\theta) \nu(d\theta) = Q_g(d\theta|x) M_g(dx) . \quad (4)$$

Let  $\Delta_g$  denote the analogue of  $\Delta$  for the perturbed problem; i.e., for  $h \in L_2(\nu_g) = L_2(\nu)$ , define

$$\Delta_g(h) = \frac{1}{2} \int_{\Theta} \int_{\Theta} (h(\theta) - h(\eta))^2 S_g(d\theta, d\eta) ,$$

where  $S_g(d\theta, d\eta) := \int_{\mathcal{X}} Q_g(d\theta|x)Q_g(d\eta|x)M_g(dx)$ . Note that  $C \subset \Theta$  is  $\nu$ -proper if and only if it is  $\nu_g$ -proper. Finally, the inequality  $\Delta_g(h) \leq \varepsilon^{-3}\Delta(h)$ , which is established in the Appendix, shows that the analogue of (3) (for the perturbed problem) is satisfied and we have the following result.

**Corollary 1.** (*Eaton, 1992*) *Suppose that (3) holds so that  $\nu$  is strongly admissible. If  $g : \Theta \rightarrow [0, \infty)$  is bounded away from 0 and  $\infty$ , then the prior  $\nu_g(d\theta) = g(\theta) \nu(d\theta)$  is also strongly admissible.*

Our proof that  $\Delta_g(h) \leq \varepsilon^{-3}\Delta(h)$  breaks down when  $g$  is not bounded away from zero (and when  $g$  is unbounded). The question that motivated our research is “Does Corollary 1 hold true *without* the assumption that  $g$  is bounded away from 0?” We will show that (under mild regularity conditions) the answer to the question is “yes.” This result has important implications for estimation problems with restricted parameter spaces. Indeed, if  $g$  is the indicator function of  $\Theta^* \subset \Theta$ , then  $\nu_g(d\theta) = I_{\Theta^*}(\theta)\nu(d\theta)$  and the use of this prior is effectively the same as restricting the parameter space to be  $\Theta^*$  from the start. Further details concerning this application are provided later in this section. It is important to recognize that, in the absence of the assumption that  $g$  is bounded away from 0, there is no guarantee that  $\nu_g$  is improper.

## 1.2 Eaton’s Markov chain

Our arguments rely heavily on a result of Eaton [5] showing that the sufficient condition (3) is equivalent to a recurrence property of a Markov chain associated with the decision problem. A precise description of this equivalence involves a few basic concepts from the theory of general state space Markov chains. Let  $Z$  be a Polish space with Borel  $\sigma$ -algebra  $\mathcal{B}(Z)$ . Suppose that  $K : \mathcal{B}(Z) \times Z \rightarrow [0, 1]$  is a Markov transition function; i.e., for each  $z \in Z$ ,  $K(\cdot|z)$  is a probability measure and for each  $A \in \mathcal{B}(Z)$ ,  $K(A|\cdot)$  is a measurable function. The transition function  $K$  defines a Markov chain  $\Phi = \{\Phi_n\}_{n=0}^{\infty}$  on the infinite product space  $Z^{\infty}$  such that, given  $\Phi_n = z$ , the probability distribution of  $\Phi_{n+1}$  is  $K(\cdot|z)$ . Let  $\Pr_z(\cdot)$  denote the overall law governing  $\Phi$  on  $Z^{\infty}$  assuming that  $\Phi_0 = z$ .

Two different notions of the *recurrence* of a Markov chain are used in this article. The definition used by Meyn and Tweedie [15, Chapter 8], which we refer to simply as “recurrence,” applies only to Markov chains that satisfy certain irreducibility properties. In order to postpone a discussion of the technical matter of irreducibility, we defer a formal statement of this definition until Section 2. The Markov chains studied by Eaton [5] are not necessarily irreducible, so he used a different notion called “local recurrence” that does not require any irreducibility. Let  $\sigma_A$  denote the first return to  $A$ ; that is,  $\sigma_A = \min \{n \geq 1 : \Phi_n \in A\}$  with the understanding that  $\sigma_A = \infty$  if  $\Phi_n \in A^c$  for all  $n \geq 1$ . Also, let  $\alpha$  be a non-trivial,  $\sigma$ -finite measure on  $Z$ . Recall that a set  $A$  is  $\alpha$ -proper if  $0 < \alpha(A) < \infty$ .

**Definition 3.** *The Markov chain  $\Phi$  is called locally- $\alpha$ -recurrent if for each  $\alpha$ -proper set  $A$ , the set*

$$\{z \in A : \Pr_z(\sigma_A < \infty) < 1\}$$

*has  $\alpha$ -measure 0.*

The definition says that, for every  $\alpha$ -proper set  $A$ , aside from a set of starting values in  $A$  that has  $\alpha$ -measure 0, the chain returns to the set  $A$  with probability 1.

If  $\pi$  is a non-trivial,  $\sigma$ -finite measure on  $Z$  such that  $K(dw|z)\pi(dz) = K(dz|w)\pi(dw)$ , then the Markov chain  $\Phi$  is called  $\pi$ -symmetric (or reversible with respect to  $\pi$ ). Suppose that  $\Phi$  is  $\pi$ -symmetric and define a Dirichlet form associated with  $\Phi$  as

$$D(h) = \frac{1}{2} \int_Z \int_Z (h(z) - h(w))^2 K(dz|w) \pi(dw) ,$$

for  $h \in L_2(\pi)$ . For each  $\pi$ -proper set  $A$ , define

$$U(A) = \{h \in L_2(\pi) : h \text{ is bounded, } h(z) \geq I_A(z)\} .$$

Here is a characterization of local- $\pi$ -recurrence for a  $\pi$ -symmetric Markov chain.

**Theorem 2.** *(Eaton, 1992) Suppose the Markov chain  $\Phi$  is  $\pi$ -symmetric. Then  $\Phi$  is locally- $\pi$ -recurrent if and only if*

$$\inf_{h \in U(A)} D(h) = 0 \text{ for each } \pi\text{-proper set } A . \quad (5)$$

To see the connection between the sufficient condition (3) and local recurrence, define

$$R(d\theta|\eta) = \int_{\mathcal{X}} Q(d\theta|x) P(dx|\eta) ,$$

which is a Markov transition function on  $\mathcal{B}(\Theta) \times \Theta$ , and note that, by using (1), we can write  $S(d\theta, d\eta) = R(d\theta|\eta)\nu(d\eta)$ . Let  $W = \{W_n\}_{n=0}^\infty$  denote the Markov chain corresponding to  $R$ . Since  $S$  is a symmetric measure, we have  $R(d\theta|\eta)\nu(d\eta) = R(d\eta|\theta)\nu(d\theta)$ ; that is,  $W$  is a  $\nu$ -symmetric chain. Therefore, to each decision problem; i.e., to each  $(P, \nu)$  pair, there corresponds a Markov chain  $W$ . With Theorem 2 in mind, Theorem 1 can be reinterpreted as saying that a sufficient condition for the strong admissibility of  $\nu$  is the local- $\nu$ -recurrence of  $W$ . Not only is this an interesting connection between admissibility and recurrence similar to those established by Brown [1] and Johnstone [12, 13], it has practical implications as well. Indeed, while direct verification of (5) can be very difficult, there are alternative techniques for establishing recurrence that are often effective in practice [see; e.g., 5, Example 3.1 & p.1163].

We now return to the original perturbed prior problem where it is assumed that  $g$  is bounded away from 0 and  $\infty$ . Mimicking what was done earlier, define

$$R_g(d\theta|\eta) = \int_{\mathcal{X}} Q_g(d\theta|x) P(dx|\eta) ,$$

which is a Markov transition function on  $\mathcal{B}(\Theta) \times \Theta$ , and note that  $S_g(d\theta, d\eta) = R_g(d\theta|\eta)\nu_g(d\eta)$ . Let  $W^g = \{W_n^g\}_{n=0}^\infty$  denote the Markov chain corresponding to  $R_g$ . Using Theorem 2, we can now state the key result underlying Corollary 1 in the language of Markov chains as follows.

**Theorem 3.** *If  $W$  is locally- $\nu$ -recurrent and  $g : \Theta \rightarrow [0, \infty)$  is bounded away from zero and  $\infty$ , then  $W^g$  is locally- $\nu_g$ -recurrent.*

Figure 1 depicts the implications used in Corollary 1. Note that everything follows from the local- $\nu$ -recurrence of  $W$ .

$$\begin{aligned} W \text{ is locally-}\nu\text{-recurrent} &\Rightarrow \nu(d\theta) \text{ is strongly admissible} \\ &\Downarrow \\ W^g \text{ is locally-}\nu_g\text{-recurrent} &\Rightarrow g(\theta) \nu(d\theta) \text{ is strongly admissible} \end{aligned}$$

Figure 1: Implications holding under the assumption that  $g$  is bounded away from 0 and  $\infty$ .

### 1.3 A closely related Markov chain on $\mathcal{X}$

In the same way that  $R(d\theta|\eta)$  defines the Markov chain  $W$  on the parameter space  $\Theta$ , the function

$$\tilde{R}(dy|x) = \int_{\Theta} P(dy|\theta) Q(d\theta|x)$$

defines a Markov chain, say  $\tilde{W} = \{\tilde{W}_n\}_{n=0}^\infty$ , on the sample space  $\mathcal{X}$ . Moreover, it's easy to see that  $\tilde{W}$  is  $M$ -symmetric; that is,  $\tilde{R}(dy|x)M(dx) = \tilde{R}(dx|y)M(dy)$ . So now each decision problem has associated with it two Markov chains:  $W$  on the parameter space, and  $\tilde{W}$  on the sample space. Despite the fact that  $W$  and  $\tilde{W}$  live on different spaces, they have similar stability properties due to the structural similarity of their transition functions and we will exploit this later.

There are two decision problems under consideration here: the original problem with model  $P(dx|\theta)$  and prior  $\nu(d\theta)$  and the perturbed problem with model  $P(dx|\theta)$  and prior  $g(\theta)\nu(d\theta)$ , where it is assumed that  $g$  is bounded away from 0 and  $\infty$ . There are a total of four Markov chains associated with these two problems and the only one that has not been introduced yet is  $\tilde{W}^g$ , the Markov chain on  $\mathcal{X}$  with Markov transition function

$$\tilde{R}_g(dy|x) = \int_{\Theta} P(dy|\theta) Q_g(d\theta|x) .$$

The key to the proof of Corollary 1 is the inequality  $\Delta_g(h) \leq \varepsilon^{-3}\Delta(h)$ . A quick glance at the Appendix reveals that this inequality is proved by establishing a simple relationship between  $S(d\theta, d\eta)$  and  $S_g(d\theta, d\eta)$ . The Dirichlet forms corresponding to  $\tilde{W}$  and  $\tilde{W}^g$  involve the (symmetric) measures  $\tilde{S}(dy, dx) = \tilde{R}(dy|x)M(dx)$  and  $\tilde{S}_g(dy, dx) = \tilde{R}_g(dy|x)M_g(dx)$ , respectively. It turns out that the

relationship between  $\tilde{S}$  and  $\tilde{S}_g$  is actually even simpler than the one between  $S$  and  $S_g$ . Indeed, (1) can be used to show that

$$\tilde{S}(dy, dx) = \int_{\Theta} P(dy|\theta) P(dx|\theta) \nu(d\theta) ,$$

and

$$\tilde{S}_g(dy, dx) = \int_{\Theta} P(dy|\theta) P(dx|\theta) g(\theta) \nu(d\theta) ,$$

and hence the two measures differ by just a single factor of  $g(\theta)$ . This fact is used in the Appendix to prove the following result.

**Theorem 4.** *If  $\tilde{W}$  is locally- $M$ -recurrent and  $g : \Theta \rightarrow [0, \infty)$  is bounded, then  $\tilde{W}^g$  is locally- $M_g$ -recurrent.*

In contrast to Theorem 3, it is *not* assumed in Theorem 4 that  $g$  is bounded away from 0. Thus, Theorem 4 is much more than just the analogue of Theorem 3 for the chains on the sample space. On the other hand, to this point we have not mentioned any connections between the chains on the sample space ( $\tilde{W}$  &  $\tilde{W}^g$ ) and strong admissibility. Thus, something more is needed before we can take advantage of Theorem 4.

*Remark 1.* Theorem 4 fails without the assumption that  $g$  is bounded above (even if we add the assumption that  $g$  is bounded away from 0). Indeed, the results in [10] can be used to construct examples where  $\tilde{W}$  is locally- $M$ -recurrent and  $g$  is unbounded (and bounded away from zero), but  $\tilde{W}^g$  is *not* locally- $M_g$ -recurrent. Thus, an extension of our perturbation results to the case of unbounded  $g$  would require a more delicate analysis.

#### 1.4 Exploiting a relationship between $W$ and $\tilde{W}$

The final piece of the puzzle is a stability relationship between  $W$  and  $\tilde{W}$  that allows us to prove Corollary 1 using Theorem 4 instead of Theorem 3 and hence to dispense with the assumption that  $g$  is bounded away from 0. Hobert and Robert [10] established a stability relationship between  $W$  and  $\tilde{W}$  and one of the main results in this paper is a substantial generalization of their result. To describe this generalization, we need to introduce a notion of irreducibility for general state space Markov chains.

Consider again our generic Markov chain  $\Phi$  with state space  $Z$  and suppose that  $\alpha$  is a non-trivial,  $\sigma$ -finite measure on  $Z$ . The chain  $\Phi$  is called  $\alpha$ -irreducible if for every set  $A$  with  $\alpha(A) > 0$  and every  $z \in Z$ , there exists an  $n \geq 1$  (which may depend on  $z$  and  $A$ ) such that  $\Pr_z(\Phi_n \in A) > 0$ . Simply put,  $\Phi$  is  $\alpha$ -irreducible if every set with positive  $\alpha$ -measure is accessible from every point in the state space. Here is our result.

**Theorem 5.** *Assume that either  $W$  is  $\nu$ -irreducible or  $\tilde{W}$  is  $M$ -irreducible. Then the Markov chain  $W$  is recurrent if and only if the Markov chain  $\tilde{W}$  is recurrent.*

Note that Theorem 5 involves (Meyn and Tweedie’s definition of) recurrence. Thus, in order to use this result in conjunction with the previous results, which involve local recurrence, we need some results relating the two forms of recurrence. Indeed, we will show that, under certain irreducibility conditions, recurrence and local recurrence are equivalent. Hence, when the chains in question ( $W$ ,  $\widetilde{W}$ ,  $W^g$  and  $\widetilde{W}^g$ ) are all suitably irreducible and  $g$  is bounded, the set of implications in Figure 2 holds. Therefore, at the expense of adding some mild irreducibility conditions to Corollary 1, we can remove the assumption that  $g$  is bounded away from zero. Corollary 2 is a formal statement of our extension of Corollary 1.

$$\begin{array}{c} \widetilde{W} \text{ is recurrent} \Leftrightarrow W \text{ is recurrent} \Rightarrow \nu(d\theta) \text{ is strongly admissible} \\ \Downarrow \\ \widetilde{W}^g \text{ is recurrent} \Leftrightarrow W^g \text{ is recurrent} \Rightarrow g(\theta) \nu(d\theta) \text{ is strongly admissible} \end{array}$$

Figure 2: Implications holding under (irreducibility and) the assumption that  $g$  is bounded.

**Corollary 2.** *Suppose that (3) holds so that  $\nu$  is strongly admissible. Suppose that  $g : \Theta \rightarrow [0, \infty)$  is bounded and that the following two conditions hold*

1. *Either  $W$  is  $\nu$ -irreducible or  $\widetilde{W}$  is  $M$ -irreducible.*
2. *Either  $W^g$  is  $\nu_g$ -irreducible or  $\widetilde{W}^g$  is  $M_g$ -irreducible.*

*Then the prior  $\nu_g(d\theta) = g(\theta) \nu(d\theta)$  is strongly admissible.*

There is an important subtle difference between Corollary 1 and Corollary 2. In order to apply either result, one must show that (3) holds, or equivalently, that  $W$  is locally- $\nu$ -recurrent. If Corollary 2 is applicable, that is, if the irreducibility conditions hold, then there are two different Markov chains,  $W$  and  $\widetilde{W}$ , whose recurrence leads to the desired result. Thus, if  $W$  happens to be analytically intractable, one can operate on  $\widetilde{W}$  instead. This idea has proved useful in practice. Indeed, experience has shown that, of the two chains,  $\widetilde{W}$  is often more amenable to analysis [see; e.g. 9, 10, 11]. This fact may make it slightly less surprising that we are able to make a stronger statement about the relationship between  $\widetilde{W}$  and  $\widetilde{W}^g$  (Theorem 4) than we can about that between  $W$  and  $W^g$  (Theorem 3).

We now describe an application to estimation problems with restricted parameter spaces. Recall our original decision problem with parameter space  $\Theta$ , sample space  $\mathcal{X}$  and statistical model  $P(dx|\theta)$ . Call this “Problem A.” Consider a slightly different problem where everything is the same except that the parameter space is now  $\Theta^* \in \mathcal{B}(\Theta)$ , a proper subset of  $\Theta$ . Call this “Problem B.” Suppose that  $\nu$  is an improper prior on  $\Theta$  that satisfies (3) so  $\nu$  is strongly admissible in Problem A. Let  $\nu^*$  denote the measure  $\nu$  restricted to  $\Theta^*$ . Since  $\nu$  is strongly admissible when the parameter space is  $\Theta$ , it seems plausible that  $\nu^*$  might be strongly admissible for Problem B. (This does not seem to follow immediately from the



definition of almost admissibility.) In Section 4, we show that, under mild irreducibility conditions,  $\nu^*$  is strongly admissible for Problem B. We illustrate this result using an example concerning independent and identically distributed random vectors from a translation family.

The rest of the paper is organized as follows. Section 2 contains a series of results for general state space Markov chains that culminates in conditions under which local recurrence is equivalent to recurrence. The results in Section 2 are used in Section 3 to prove Theorem 5 and Corollary 2. Finally, Section 4 describes the application of our results to estimation problems with restricted parameter spaces.

## 2 Reconciling two different notions of recurrence

In this section, we will establish general conditions under which recurrence and local recurrence are equivalent. After stating some definitions and results from [15], we proceed to state and prove several new results that lead to the connection between recurrence and local recurrence.

Consider again our generic Markov chain  $\Phi$  with state space  $Z$  and Markov transition function  $K(\cdot|\cdot)$ . For  $n \in \mathbb{N} := \{1, 2, 3, \dots\}$ , let  $K^n : \mathcal{B}(Z) \times Z \rightarrow [0, 1]$  denote the  $n$ -step Markov transition function defined inductively by

$$K^{n+1}(A|z) = \int_Z K^n(A|y) K(dy|z) ,$$

where  $K^1 \equiv K$ . Of course,  $K^n(A|z) = \Pr_z(\Phi_n \in A)$ . Here is a formal definition of  $\alpha$ -irreducibility.

**Definition 4.** [15, p.87]. *The chain  $\Phi$  is called  $\alpha$ -irreducible if  $\alpha$  is a non-trivial,  $\sigma$ -finite measure on  $(Z, \mathcal{B}(Z))$  such that for every  $A \in \mathcal{B}(Z)$  with  $\alpha(A) > 0$  and every  $z \in Z$ , there exists an  $n \geq 1$  (which may depend on  $z$  and  $A$ ) with  $K^n(A|z) > 0$ . The measure  $\alpha$  is called an irreducibility measure for  $\Phi$ .*

Recall that two measures,  $\mu_1$  and  $\mu_2$ , defined on the same measurable space are *equivalent* (written  $\mu_1 \equiv \mu_2$ ) if they have the same sets of measure 0; that is, if  $\mu_1 \succ \mu_2$  and  $\mu_2 \succ \mu_1$ . If  $\Phi$  is  $\alpha$ -irreducible, then there exists a *maximal irreducibility measure*,  $\psi(\cdot)$ , that satisfies the following properties [15, Section 4.2.2]:

MX1:  $\Phi$  is  $\psi$ -irreducible;

MX2:  $\psi$  is a probability measure;

MX3: if  $\alpha'$  is any other measure, the chain  $\Phi$  is  $\alpha'$ -irreducible if and only if  $\psi \succ \alpha'$ ;

MX4: if  $\psi(A) = 0$ , then  $\psi\left\{z \in Z : \Pr_z(\sigma_A < \infty) > 0\right\} = 0$  .

While maximal irreducibility measures are not unique, it follows from MX3 that any two maximal irreducibility measures are equivalent. When we say “ $\Phi$  is a  $\psi$ -irreducible Markov chain” we mean that  $\Phi$  is  $\alpha$ -irreducible for some  $\alpha$  and that  $\psi$  is a maximal irreducibility measure for the chain. It is usually possible to analyze a  $\psi$ -irreducible chain without knowing  $\psi$  as long as the set  $\mathcal{B}^+(Z) := \{A \in \mathcal{B}(Z) : \psi(A) > 0\}$  is known. In fact, some of our effort in this section is dedicated to establishing that, if  $\Phi$  is  $\alpha$ -symmetric and  $\alpha$ -irreducible, then  $\alpha \equiv \psi$ .

We now state Meyn and Tweedie’s definition of recurrence as well as some results that will be required in the sequel. For  $A \in \mathcal{B}(Z)$ , let  $\eta_A$  denote the number of times the chain visits the set  $A$  after time zero; i.e.,  $\eta_A := \sum_{n=1}^{\infty} I_A(\Phi_n)$ . Let  $E_z(\cdot)$  denote the expectation operator corresponding to  $\text{Pr}_z(\cdot)$ .

**Definition 5.** [15, Chapter 8] *Let  $\Phi$  be a  $\psi$ -irreducible Markov chain. The chain  $\Phi$  is recurrent if for each  $z \in Z$  and each  $A \in \mathcal{B}^+(Z)$ ,*

$$E_z(\eta_A) = \sum_{n=1}^{\infty} K^n(A|z) = \infty .$$

*Remark 2.* Despite the fact that the name “ $\psi$ -recurrent” is probably more consistent with the general terminology in this paper, we will stick to the more standard “recurrent” with its implicit reference to the measure  $\psi$ .

For a measurable set  $A$ , define the event

$$\{\Phi \in A \text{ i.o.}\} := \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \{\Phi_k \in A\} .$$

**Theorem 6.** [15, p.201] *If the  $\psi$ -irreducible Markov chain  $\Phi$  is recurrent, then there exists a set  $H \in \mathcal{B}(Z)$  satisfying  $\psi(Z \setminus H) = 0$  and for each  $z \in H$  and each  $A \in \mathcal{B}^+(Z)$ ,*

$$\text{Pr}_z(\Phi \in A \text{ i.o.}) = 1 .$$

*The set  $H$  is called the maximal Harris set.*

The next result is a useful characterization of recurrence.

**Theorem 7.** [15, p.187] *The  $\psi$ -irreducible Markov chain  $\Phi$  is recurrent if and only if there do not exist two sets  $A, B$  in  $\mathcal{B}^+(Z)$  with*

$$\text{Pr}_z(\sigma_A < \infty) < 1 \quad \forall z \in B .$$

Define the  $n$ -step taboo probabilities as

$${}_A K^n(B|z) = \text{Pr}_z(\Phi_n \in B, \sigma_A \geq n) \quad \text{for } z \in Z \text{ and } A, B \in \mathcal{B}(Z) .$$

This is simply the probability that the Markov chain started at the point  $z$ , ends up in the set  $B$  after  $n$  steps having avoided the set  $A$  along the way. Here is our first result concerning the general state space Markov chain  $\Phi$ .

**Lemma 1.** Suppose that  $\Phi$  is  $\psi$ -irreducible and that  $A$  and  $B$  are disjoint sets in  $\mathcal{B}^+(Z)$ . Then  $\psi(S) > 0$  where

$$S = \{z \in A : {}_A K^n(B|z) > 0 \text{ for some } n \in \mathbb{N}\}.$$

*Proof.* Suppose that  $\psi(S) = 0$ . It follows from MX3 that the set

$$D = \{z \in Z : K^n(S|z) > 0 \text{ for some } n \in \mathbb{N}\}$$

also has  $\psi$ -measure zero. Hence, the set  $A \setminus (D \cup S)$  has positive  $\psi$ -measure and if  $z^* \in A \setminus (D \cup S)$ , then  ${}_A K^n(B|z^*) = 0$  for all  $n \in \mathbb{N}$  and  $K^n(S|z^*) = 0$  for all  $n \in \mathbb{N}$ . Now for any  $n \in \mathbb{N}$ , the *last-exit decomposition* of  $K^n(B|z^*)$  [15, p.180] yields

$$\begin{aligned} K^n(B|z^*) &= {}_A K^n(B|z^*) + \sum_{j=1}^{n-1} \int_A K^j(dw|z^*) {}_A K^{n-j}(B|w) \\ &= {}_A K^n(B|z^*) + \sum_{j=1}^{n-1} \left[ \int_{A \setminus S} K^j(dw|z^*) {}_A K^{n-j}(B|w) + \int_S K^j(dw|z^*) {}_A K^{n-j}(B|w) \right] \\ &= 0. \end{aligned}$$

Hence,  $K^n(B|z^*) = 0$  for all  $n \in \mathbb{N}$ ; that is, the set  $B$  is not accessible from the point  $z^*$ . This contradicts the  $\psi$ -irreducibility of  $\Phi$  since  $\psi(B) > 0$ .  $\square$

**Lemma 2.** Suppose the Markov chain  $\Phi$  is  $\psi$ -irreducible and locally- $\pi$ -recurrent and that  $\pi \equiv \psi$ . Then for each  $\pi$ -proper set  $A$ , the set

$$\{z \in Z : \Pr_z(\sigma_A < \infty) < 1\}$$

has  $\pi$ -measure 0.

*Proof.* Of course, local- $\pi$ -recurrence implies that  $\pi\{z \in A : \Pr_z(\sigma_A < \infty) < 1\} = 0$ . Thus, it suffices to show that  $B = \{z \in \bar{A} : \Pr_z(\sigma_A < \infty) < 1\}$  also has  $\pi$ -measure zero. Assume to the contrary that  $\pi(B) > 0$ . Then there exists an  $\varepsilon > 0$  such that the set

$$B_\varepsilon = \{z \in \bar{A} : \Pr_z(\sigma_A < \infty) < 1 - \varepsilon\}$$

has positive  $\pi$ -measure. Lemma 1 then implies that the set

$$S = \{z \in A : {}_A K^n(B_\varepsilon|z) > 0 \text{ for some } n \in \mathbb{N}\}$$

has positive  $\pi$ -measure. If  $z \in S$ , there exists an  $n = n(z)$  such that  ${}_A K^{n(z)}(B_\varepsilon|z) > 0$ . Moreover, once the chain enters the set  $B_\varepsilon$ , there is a probability of at least  $\varepsilon$  that the chain never returns to  $A$ . Thus,  $\Pr_z(\sigma_A = \infty) \geq \varepsilon {}_A K^{n(z)}(B_\varepsilon|z) > 0$ . This contradicts the local- $\pi$ -recurrence of  $\Phi$  since  $0 < \pi(S) < \infty$ .  $\square$

Here is the main result of this section.

**Theorem 8.** *Suppose the Markov chain  $\Phi$  is  $\psi$ -irreducible and that  $\pi$  is a measure on  $(Z, \mathcal{B}(Z))$  such that  $\pi \equiv \psi$ . Then recurrence and local- $\pi$ -recurrence are equivalent.*

*Proof.* Assume that  $\Phi$  is recurrent and let  $A$  be a  $\pi$ -proper set. The equivalence of  $\pi$  and  $\psi$  implies that  $\psi(A) > 0$  and from Theorem 6 it follows that

$$\Pr_z(\Phi \in A \text{ i.o.}) = 1$$

for all  $z \in H$ . From this we deduce that  $\Pr_z(\sigma_A < \infty) = 1$  for all  $z \in H$ . Therefore,

$$\{z \in A : \Pr_z(\sigma_A < \infty) < 1\} \subseteq \{z \in Z : \Pr_z(\sigma_A < \infty) < 1\} \subseteq Z \setminus H .$$

Thus, the set  $\{z \in A : \Pr_z(\sigma_A < \infty) < 1\}$  has  $\psi$ -measure zero and, since  $\pi$  and  $\psi$  are equivalent, it also has  $\pi$ -measure 0 and hence  $\Phi$  is locally- $\pi$ -recurrent.

Conversely, assume that  $\Phi$  is locally- $\pi$ -recurrent. Lemma 2 implies that for every  $\pi$ -proper set  $A$ , we have

$$\pi\{z \in Z : \Pr_z(\sigma_A < \infty) < 1\} = 0 . \quad (6)$$

Now suppose that  $\Phi$  is not recurrent so that by Theorem 7 there exist sets  $C_1$  and  $C_2$  with  $\psi(C_1) > 0$  and  $\psi(C_2) > 0$  such that

$$\Pr_z(\sigma_{C_1} < \infty) < 1 \quad \forall z \in C_2 . \quad (7)$$

The equivalence of  $\pi$  and  $\psi$  implies that  $\pi(C_1) > 0$  and  $\pi(C_2) > 0$ . There are two possible cases: (i)  $\pi(C_1) < \infty$  and (ii)  $\pi(C_1) = \infty$ . In case (i),  $C_1$  is a  $\pi$ -proper set so (7) directly contradicts (6). In case (ii), because  $\pi$  is  $\sigma$ -finite, there exists a set  $C'_1 \subset C_1$  such that  $C'_1$  is  $\pi$ -proper. Then

$$\Pr_z(\sigma_{C'_1} < \infty) \leq \Pr_z(\sigma_{C_1} < \infty) < 1 \quad \forall z \in C_2 ,$$

which again contradicts (6). Thus,  $\Phi$  must be recurrent. □

*Remark 3.* An obvious implication of Theorem 8 is that for a  $\psi$ -irreducible Markov chain, recurrence and local- $\psi$ -recurrence are equivalent.

We now develop a specialized version of Theorem 8 for symmetric Markov chains that will be used to analyze the chains described in the Introduction.

**Definition 6.** [15, p.229]. *Let  $\pi$  be a non-trivial,  $\sigma$ -finite measure on  $(Z, \mathcal{B}(Z))$ . The measure  $\pi$  is said to be invariant for  $\Phi$  if  $\pi(A) = \int_Z K(A|w) \pi(dw)$  for all  $A \in \mathcal{B}(Z)$ .*

*Remark 4.* Suppose  $\Phi$  is a  $\psi$ -irreducible Markov chain and that  $\pi$  is invariant for  $\Phi$ . If  $\pi(Z) < \infty$ , then  $\Phi$  is recurrent and is called *positive recurrent*. If  $\pi(Z) = \infty$ , then  $\Phi$  may or may not be recurrent. If  $\pi(Z) = \infty$  and  $\Phi$  is recurrent, then it is called *null recurrent*.

**Lemma 3.** *Suppose  $\pi$  is invariant for the Markov chain  $\Phi$ . If  $\pi(A) = 0$ , then the set*

$$\{z \in Z : K^n(A|z) > 0 \text{ for some } n \in \mathbb{N}\}$$

*has  $\pi$ -measure zero.*

*Proof.* Define  $D_n = \{z \in Z : K^n(A|z) > 0\}$  for  $n \in \mathbb{N}$  and note that

$$\{z \in Z : K^n(A|z) > 0 \text{ for some } n \in \mathbb{N}\} = \bigcup_{n=1}^{\infty} D_n .$$

Thus, it suffices to show that for each  $n \in \mathbb{N}$ ,  $\pi(D_n) = 0$ . We will accomplish this using an induction argument. First, invariance yields

$$0 = \pi(A) = \int_Z K(A|z) \pi(dz) .$$

Hence,  $\pi\{z \in Z : K(A|z) > 0\} = 0$ ; i.e.,  $\pi(D_1) = 0$ . Now assume that  $\pi(D_n) = 0$ . Then

$$K^{n+1}(A|z) = \int_Z K^n(A|u) K(du|z) = \int_{D_n} K^n(A|u) K(du|z) .$$

Now,  $D_n$  has  $\pi$ -measure 0, so it follows (from what we have already shown) that  $\{z \in Z : K(D_n|z) > 0\}$  has  $\pi$ -measure 0. Hence,  $\pi(D_{n+1}) = 0$ .  $\square$

**Lemma 4.** *Suppose that the measure  $\pi$  is invariant for the Markov chain  $\Phi$  and that  $\Phi$  is  $\pi$ -irreducible. Then  $\pi$  is equivalent to the maximal irreducibility measure,  $\psi$ ; that is,  $\psi \equiv \pi$ . Consequently,  $\mathcal{B}^+(Z) := \{A \in \mathcal{B}(Z) : \psi(A) > 0\} = \{A \in \mathcal{B}(Z) : \pi(A) > 0\}$ .*

*Proof.* Because  $\psi$  is the maximal irreducibility measure, we know that  $\psi \succ \pi$ . Thus, to show that  $\psi \equiv \pi$ , we need only establish that  $\pi \succ \psi$ . If  $\pi(A) = 0$ , then we know from Lemma 3 that the set  $\{z \in Z : K^n(A|z) > 0 \text{ for some } n \in \mathbb{N}\}$  has  $\pi$ -measure zero. Hence, there exists a  $z^* \in Z$  such that  $K^n(A|z^*) = 0$  for all  $n \in \mathbb{N}$  and since  $\psi$  is an irreducibility measure for  $\Phi$ , it follows that  $\psi(A) = 0$ .  $\square$

Suppose  $\Phi$  is  $\pi$ -symmetric and that  $A \in \mathcal{B}(Z)$ . Then

$$\pi(A) = \pi(A) \int_Z K(du|v) = \int_Z K(du|v) \int_A \pi(dv) = \int_Z \int_A K(dv|u) \pi(du) = \int_Z K(A|u) \pi(du) ;$$

i.e.,  $\pi$  is an invariant measure for  $\Phi$ . Combining Theorem 8, Lemma 4 and the relationship between symmetry and invariance yields the following result.

**Corollary 3.** *Suppose the Markov chain  $\Phi$  is  $\pi$ -symmetric and  $\pi$ -irreducible. Then recurrence and local- $\pi$ -recurrence are equivalent.*

### 3 Relationships between $W$ and $\widetilde{W}$

Recall that we have a statistical model  $P(dx|\theta)$  and a  $\sigma$ -finite, improper prior  $\nu(d\theta)$  such that the resulting marginal  $M(dx) = \int_{\Theta} P(dx|\theta)\nu(d\theta)$  is also  $\sigma$ -finite. Thus, a formal posterior distribution  $Q(d\theta|x)$  exists and is characterized by (1). We have defined two Markov chains:  $W$  is the  $\nu$ -symmetric Markov chain on  $\Theta$  whose Markov transition function is  $R(d\theta|\eta) = \int_{\mathcal{X}} Q(d\theta|x)P(dx|\eta)$  and  $\widetilde{W}$  is the  $M$ -symmetric Markov chain on  $\mathcal{X}$  whose Markov transition function is  $\tilde{R}(dy|x) = \int_{\Theta} P(dy|\theta)Q(d\theta|x)$ . In this section, we will prove Theorem 5 and Corollary 2. We start with a result concerning irreducibility.

**Theorem 9.** *The  $\nu$ -symmetric Markov chain  $W$  is  $\nu$ -irreducible if and only if the  $M$ -symmetric Markov chain  $\widetilde{W}$  is  $M$ -irreducible.*

*Proof.* Suppose that  $W$  is not  $\nu$ -irreducible. Then there exists a set  $C$  with  $\nu(C) > 0$  and a point  $\theta^* \in \Theta$  such that  $R^n(C|\theta^*) = 0$  for all  $n \in \mathbb{N}$ . We will demonstrate that  $\widetilde{W}$  is not  $M$ -irreducible. Since  $\nu(C) = \int_{\mathcal{X}} Q(C|x)M(dx) > 0$ , we have  $M\{x \in \mathcal{X} : Q(C|x) > 0\} > 0$ . Therefore, there exists an  $\varepsilon > 0$  such that  $M\{x \in \mathcal{X} : Q(C|x) > \varepsilon\} > 0$ . Set  $B_\varepsilon = \{x \in \mathcal{X} : Q(C|x) > \varepsilon\}$ . Define

$$\overline{D} = \{x \in \mathcal{X} : \tilde{R}^n(B_\varepsilon|x) = 0 \text{ for all } n \in \mathbb{N}\}.$$

We will show that  $P(\overline{D}|\theta^*) = 1$  and hence there exists an  $x^* \in \mathcal{X}$  such that  $\tilde{R}^n(B_\varepsilon|x^*) = 0$  for all  $n \in \mathbb{N}$ , which implies that  $\widetilde{W}$  is not  $M$ -irreducible. For each  $n \in \mathbb{N}$ , define

$$D_n = \{x \in \mathcal{X} : \tilde{R}^n(B_\varepsilon|x) > 0\}.$$

Note that

$$\cup_{n=1}^{\infty} D_n = \{x \in \mathcal{X} : \tilde{R}^n(B_\varepsilon|x) > 0 \text{ for some } n \in \mathbb{N}\} = D.$$

Hence, it suffices to show that  $P(D_n|\theta^*) = 0$  for each  $n \in \mathbb{N}$ . A straightforward induction argument shows that

$$R^{n+1}(C|\theta^*) = \int_{\mathcal{X}} \int_{\mathcal{X}} Q(C|x) \tilde{R}^n(dx|y) P(dy|\theta^*).$$

Hence, if for some  $n \in \mathbb{N}$ ,  $P(D_n|\theta^*) > 0$ , then

$$\begin{aligned} R^{n+1}(C|\theta^*) &= \int_{\mathcal{X}} \int_{\mathcal{X}} Q(C|x) \tilde{R}^n(dx|y) P(dy|\theta^*) \\ &\geq \int_{B_\varepsilon} \int_{D_n} Q(C|x) \tilde{R}^n(dx|y) P(dy|\theta^*) \\ &\geq \varepsilon \int_{B_\varepsilon} \int_{D_n} \tilde{R}^n(dx|y) P(dy|\theta^*) \\ &= \varepsilon \int_{D_n} \tilde{R}^n(B_\varepsilon|y) P(dy|\theta^*) > 0, \end{aligned}$$

which contradicts the assumption regarding  $W$ . Hence,  $P(D_n|\theta^*) = 0$  for all  $n \in \mathbb{N}$ . The other direction is analogous and is left to the reader.  $\square$

*Remark 5.* Note that  $W$  and  $\widetilde{W}$  have invariant measures  $\nu(d\theta)$  and  $M(dx)$ , respectively. Suppose both chains are  $\psi$ -irreducible. Since  $\nu(\Theta) = M(\mathcal{X}) = \infty$ ,  $W$  and  $\widetilde{W}$  cannot be positive recurrent.

The key to proving that  $W$  and  $\widetilde{W}$  have the same stability is that they can be represented as the “marginal” Markov chains associated with a “joint” Markov chain on  $\mathcal{X} \times \Theta$ . Indeed, consider the Markov transition function

$$R^*(d\theta, dx|\eta, y) = Q(d\theta|x) P(dx|\eta), \quad (8)$$

defined on the measurable space  $(\mathcal{X} \times \Theta, \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\Theta))$ . Let  $\{\widetilde{W}_n, W_n\}_{n=0}^\infty$  denote the corresponding Markov chain. Due to the special structure of the transition function  $R^*$  (that is, the fact that the right-hand side of (8) is free of  $y$ ), given  $W_n$ , the future of the process; i.e.,  $\{\widetilde{W}_i, W_i\}_{i=n+1}^\infty$ , is conditionally independent of  $\widetilde{W}_n$ . Analogously, given  $\widetilde{W}_n$ , the future of the process is conditionally independent of  $W_{n-1}$ . We shall denote the overall probability law governing this chain by  $\text{Pr}^*(\cdot)$ . Using the Markov property and the other property described above, we have

$$\begin{aligned} \text{Pr}^*(W_{n+1} \in C | W_n = w_n, W_{n-1} = w_{n-1}, \dots, W_0 = w) &= \text{Pr}^*(W_{n+1} \in C | W_n = w_n) \\ &= \int_{\mathcal{X}} Q(C|x) P(dx|w_n). \end{aligned}$$

Thus, the marginal sequence  $W = \{W_n\}_{n=0}^\infty$  is itself a Markov chain on  $\Theta$  with Markov transition function

$$R(d\theta|\eta) = \int_{\mathcal{X}} Q(d\theta|x) P(dx|\eta).$$

A similar argument shows that  $\widetilde{W} = \{\widetilde{W}_n\}_{n=1}^\infty$  is a Markov chain on  $\mathcal{X}$  with Markov transition function

$$\tilde{R}(dy|x) = \int_{\Theta} P(dy|\theta) Q(d\theta|x).$$

We now prove Theorem 5, which is restated here for convenience.

**Theorem 5.** *Assume that either  $W$  is  $\nu$ -irreducible or  $\widetilde{W}$  is  $M$ -irreducible. Then the Markov chain  $W$  is recurrent if and only if the Markov chain  $\widetilde{W}$  is recurrent.*

*Proof.* Theorem 9 implies that  $W$  is  $\nu$ -irreducible and  $\widetilde{W}$  is  $M$ -irreducible. Hence, Lemma 4 implies that  $\nu$  and  $M$  are equivalent to the maximal irreducibility measures of  $W$  and  $\widetilde{W}$ , respectively.

Now assume that  $\widetilde{W}$  is recurrent. Fix  $\theta \in \Theta$  and suppose  $C \in \mathcal{B}(\Theta)$  is such that  $\nu(C) > 0$ . We need to show that  $\sum_{n=1}^\infty R^n(C|\theta) = \infty$ . Since  $\nu(C) = \int_{\mathcal{X}} Q(C|x) M(dx) > 0$ , we have  $M\{x \in \mathcal{X} : Q(C|x) > 0\} > 0$ . Therefore, there exists an  $\varepsilon > 0$  such that  $M\{x \in \mathcal{X} : Q(C|x) > \varepsilon\} > 0$ . Set  $B_\varepsilon = \{x \in \mathcal{X} : Q(C|x) > \varepsilon\}$ . Let  $E_\theta^*(\cdot)$  denote the expectation operator corresponding to

$\Pr^*(\cdot|W_0 = \theta)$ . Now

$$\begin{aligned}
\sum_{n=2}^{\infty} R^n(C|\theta) &= \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^*[I_C(W_n)] \\
&= \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^*\left[\mathbf{E}_{\theta}^*[I_C(W_n)|\widetilde{W}_n]\right] \\
&\geq \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^*\left[I_{B_{\varepsilon}}(\widetilde{W}_n)\mathbf{E}_{\theta}^*[I_C(W_n)|\widetilde{W}_n]\right] \\
&= \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^*\left[I_{B_{\varepsilon}}(\widetilde{W}_n)Q(C|\widetilde{W}_n)\right] \\
&\geq \varepsilon \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^*[I_{B_{\varepsilon}}(\widetilde{W}_n)] \\
&= \varepsilon \sum_{n=2}^{\infty} \Pr^*(\widetilde{W}_n \in B_{\varepsilon}|W_0 = \theta) \\
&= \varepsilon \sum_{n=2}^{\infty} \left[ \int_{\mathcal{X}} \Pr^*(\widetilde{W}_n \in B_{\varepsilon}|\widetilde{W}_1 = x, W_0 = \theta) P(dx|\theta) \right] \\
&= \varepsilon \int_{\mathcal{X}} \left[ \sum_{n=1}^{\infty} \tilde{R}^n(B_{\varepsilon}|x) \right] P(dx|\theta) = \infty,
\end{aligned}$$

where the last equality follows from the fact that  $\widetilde{W}$  is recurrent and  $M(B_{\varepsilon}) > 0$ .

The other direction is similar, but there are some important differences so we provide the details. Now assume that  $W$  is recurrent. Fix  $x \in \mathcal{X}$  and suppose  $B \in \mathcal{B}(\mathcal{X})$  is such that  $M(B) > 0$ . We need to show that  $\sum_{n=1}^{\infty} \tilde{R}^n(B|x) = \infty$ . Since  $M(B) = \int_{\Theta} P(B|\theta) \nu(d\theta) > 0$ , we have  $\nu\{\theta \in \Theta : P(B|\theta) > 0\} > 0$ . Therefore, there exists an  $\varepsilon > 0$  such that  $\nu\{\theta \in \Theta : P(B|\theta) > \varepsilon\} > 0$ . Set  $C_{\varepsilon} = \{\theta \in \Theta : P(B|\theta) > \varepsilon\}$ . Again,  $\mathbf{E}_{\theta}^*(\cdot)$  denotes the expectation operator corresponding to  $\Pr^*(\cdot|W_0 = \theta)$ . Now

$$\begin{aligned}
\sum_{n=1}^{\infty} \tilde{R}^n(B|x) &= \sum_{n=2}^{\infty} \Pr^*(\widetilde{W}_n \in B|\widetilde{W}_1 = x) \\
&= \sum_{n=2}^{\infty} \left[ \int_{\Theta} \Pr^*(\widetilde{W}_n \in B|W_1 = \theta) Q(d\theta|x) \right] \\
&= \int_{\Theta} \left[ \sum_{n=1}^{\infty} \mathbf{E}_{\theta}^*[I_B(\widetilde{W}_n)] \right] Q(d\theta|x).
\end{aligned}$$

Thus, it is enough to show that, for each  $\theta \in \Theta$ ,

$$\sum_{n=2}^{\infty} \mathbf{E}_{\theta}^*[I_B(\widetilde{W}_n)] = \infty.$$



Now, for fixed  $\theta \in \Theta$ , we have

$$\begin{aligned}
\sum_{n=2}^{\infty} \mathbf{E}_{\theta}^* [I_B(\widetilde{W}_n)] &= \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^* \left[ \mathbf{E}_{\theta}^* [I_B(\widetilde{W}_n) | W_{n-1}] \right] \\
&\geq \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^* \left[ I_{C_{\varepsilon}}(W_{n-1}) \mathbf{E}_{\theta}^* [I_B(\widetilde{W}_n) | W_{n-1}] \right] \\
&= \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^* \left[ I_{C_{\varepsilon}}(W_{n-1}) P(B | W_{n-1}) \right] \\
&\geq \varepsilon \sum_{n=2}^{\infty} \mathbf{E}_{\theta}^* [I_{C_{\varepsilon}}(W_{n-1})] \\
&= \varepsilon \sum_{n=1}^{\infty} R^n(C_{\varepsilon} | \theta) = \infty,
\end{aligned}$$

where the last equality follows from the fact that  $W$  is recurrent and  $\nu(C_{\varepsilon}) > 0$ .  $\square$

*Remark 6.* It is clear that, instead of  $R^*(d\theta, dx | \eta, y) = Q(d\theta | x) P(dx | \eta)$ , we could just as easily have used the Markov transition function

$$\tilde{R}^*(dx, d\theta | y, \eta) = P(dx | \theta) Q(d\theta | y)$$

as the basis of our joint chain. Moreover, the measure  $P(dx | \theta) \nu(d\theta)$  is invariant for both  $R^*$  and  $\tilde{R}^*$ .

Theorem 5 is a substantial generalization of Hobert and Robert's [10] Theorem 1 which assumes that  $\mathcal{X}$  and  $\Theta$  are both subsets of  $\mathbb{R}$  and that the collection of sets  $\{B \in \mathcal{B}(\mathcal{X}) : P(B | \theta) > 0\}$  is the same for all  $\theta \in \Theta$ ; that is, that the support of  $P(dx | \theta)$  is the same for all  $\theta \in \Theta$ . We will see below that this second condition actually implies that  $W$  and  $\widetilde{W}$  are irreducible. We now provide a proof of Corollary 2.

*Proof of Corollary 2.* According to Theorem 1, (3) implies the strong admissibility of  $\nu$ . What we need to show is that (3) in conjunction with the boundedness of  $g$  and the irreducibility assumptions implies that  $\nu_g(d\theta)$  is strongly admissible. We will accomplish this by following a string of implications shown in Figure 2. First, Theorem 9 implies that  $W$  is  $\nu$ -irreducible,  $\widetilde{W}$  is  $M$ -irreducible,  $W^g$  is  $\nu_g$ -irreducible, and  $\widetilde{W}^g$  is  $M_g$ -irreducible. Now, by Theorem 2, (3) implies that  $W$  is locally- $\nu$ -recurrent and hence recurrent by Corollary 3. Theorem 5 then shows that  $\widetilde{W}$  is recurrent and hence locally- $M$ -recurrent by Corollary 3. Theorem 4 implies that  $\widetilde{W}^g$  is locally- $M_g$ -recurrent and hence recurrent by Corollary 3. Another application of Theorem 5 shows that  $W^g$  is recurrent and hence locally- $\nu_g$ -recurrent by Corollary 3. Finally, Theorem 2 and Theorem 1 together show that  $\nu_g$  is strongly admissible.  $\square$

Here is a simple example illustrating the application of Corollary 2.

*Example 1.* Let  $X_1, \dots, X_n$  be independent and identically distributed random variables from a distribution with density

$$p(x|\theta) = \frac{1}{\theta} I_{(0,\theta)}(x),$$

where  $\theta \in \Theta = \mathbb{R}_+ := (0, \infty)$ . Take the prior distribution to be  $\nu(d\theta) = d\theta/\theta$ , where  $d\theta$  denotes Lebesgue measure on  $\Theta$ . It's easy to show that the marginal measure is

$$M(dx) = \frac{dx}{nx_{(n)}^n},$$

where  $dx$  denotes Lebesgue measure on  $\mathcal{X} = \mathbb{R}_+^n$  and  $x_{(n)} := \max\{x_1, \dots, x_n\}$ . The posterior is given by  $Q(d\theta|x) = q(\theta|x)d\theta$  where

$$q(\theta|x) = \frac{nx_{(n)}^n}{\theta^{n+1}} I_{(x_{(n)}, \infty)}(\theta).$$

It follows that  $R(d\theta|\eta) = r(\theta|\eta)d\theta$  where

$$r(\theta|\eta) = \frac{n\theta^{n-1}}{2\eta^n} I_{(0,\eta]}(\theta) + \frac{n\eta^n}{2\theta^{n+1}} I_{(\eta,\infty)}(\theta).$$

Note that  $r(\theta|\eta) > 0$  for all  $\theta, \eta \in \mathbb{R}_+$ . Now, if  $C$  is a set in  $\mathbb{R}_+$  such that  $\nu(C) > 0$ , then  $C$  must have positive Lebesgue measure. Thus, for any  $\eta > 0$ ,  $R(C|\eta) > 0$  so the Markov chain  $W$  is  $\nu$ -irreducible.

Because  $\eta^{-1}$  is a scale parameter in the density  $r(\theta|\eta)$ , the chain  $W = \{W_n\}_{n=0}^\infty$  can be represented as

$$W_{n+1} = W_n U_{n+1}, \tag{9}$$

where  $U_1, U_2, \dots$  are independent and identically distributed with density

$$f(u) = \frac{nu^{n-1}}{2} I_{(0,1]}(u) + \frac{n}{2u^{n+1}} I_{(1,\infty)}(u).$$

Taking logs in (9) yields a random walk with increment distribution equal to that of  $\log U_1$  and, since  $E[\log U_1] = 0$ , this random walk is recurrent [15, p.247]. It follows that  $W$  is also recurrent [10, Section 3] so the prior distribution  $\nu(d\theta) = d\theta/\theta$  is strongly admissible. (We note that strong admissibility of  $\nu$ , or equivalently, local- $\nu$ -recurrence of  $W$ , can also be deduced from the results in Example 3.1 of [5].) Since  $W$  is  $\nu$ -irreducible and recurrent, Corollary 2 is applicable. For instance, suppose that  $g(\theta) = I_{\Theta^*}(\theta)$  where  $\Theta^* = \cup_{i=0}^\infty (2i, 2i+1)$ . Let  $r_g(\theta|\eta)$  denote the Markov transition function of  $W_g$ . It's easy to see that  $r_g(\theta|\eta) > 0$  for all  $\theta \in \Theta^*$  and all  $\eta \in \Theta$  and it follows (using arguments like those above) that  $W_g$  is  $\nu_g$ -irreducible. Thus, by Corollary 2,  $\nu_g(d\theta) = g(\theta)\nu(d\theta)$  is strongly admissible. This ends Example 1.

In most applications, the  $n$ -step transition function,  $R^n$ , is intractable so irreducibility must be established through indirect methods. Notice, however, that in the preceding example, we were able to establish the  $\nu$ -irreducibility of  $W$  quite straightforwardly because every set  $C$  with  $\nu(C) > 0$  is

accessible from every point  $\theta$  in a *single step*. We now describe a simple condition that ensures such behavior. Suppose the collection of sets  $\{B \in \mathcal{B}(\mathcal{X}) : P(B|\theta) > 0\}$  is the same for all  $\theta \in \Theta$  and denote this collection by  $S$ . When this is the case, we say that the ‘‘SFP condition’’ holds, where SFP stands for ‘‘support free of parameter.’’ The following result shows that the SFP condition implies (among other things) that  $R(C|\theta) > 0$  for all  $\theta \in \Theta$  whenever  $\nu(C) > 0$ .

**Lemma 5.** *Suppose the SFP condition holds and let  $S = \{B \in \mathcal{B}(\mathcal{X}) : P(B|\theta) > 0\}$ . Then*

1.  $\{B \in \mathcal{B}(\mathcal{X}) : M(B) > 0\} = S$ ,
2. For  $C \in \mathcal{B}(\Theta)$ ,  $\nu(C) > 0 \Rightarrow R(C|\theta) > 0$  for all  $\theta \in \Theta$ ,
3. For  $C \in \mathcal{B}(\Theta)$ ,  $\nu(C) = 0 \Rightarrow R(C|\theta) = 0$  for all  $\theta \in \Theta$ ,
4. For  $B \in \mathcal{B}(\mathcal{X})$ ,  $M(B) > 0 \Rightarrow \tilde{R}(B|x) > 0$  for all  $x \in \mathcal{X}$ , and
5. For  $B \in \mathcal{B}(\mathcal{X})$ ,  $M(B) = 0 \Rightarrow \tilde{R}(B|x) = 0$  for all  $x \in \mathcal{X}$ .

Hence,  $W$  is  $\nu$ -irreducible and  $\tilde{W}$  is  $M$ -irreducible.

*Proof.* (1) This follows easily from the fact that  $M(B) = \int_{\Theta} P(B|\theta) \nu(d\theta)$ . (2) Of course,  $\nu(C) = \int_{\mathcal{X}} Q(C|x) M(dx)$ . Define  $A = \{x \in \mathcal{X} : Q(C|x) > 0\}$ . Since  $\nu(C) > 0$ ,  $M(A) > 0$  and it follows from (1) that  $P(A|\theta) > 0$  for all  $\theta$ . Thus, for any  $\theta \in \Theta$ , we have

$$R(C|\theta) = \int_{\mathcal{X}} Q(C|x) P(dx|\theta) = \int_A Q(C|x) P(dx|\theta) > 0.$$

(3) Since  $\nu(C) = 0$ ,  $M(A) = 0$  and hence  $P(A|\theta) = 0$  for all  $\theta$ . Thus, for any  $\theta \in \Theta$ ,  $R(C|\theta) = 0$ .

(4) Since  $M(B) > 0$ ,  $P(B|\theta) > 0$  for all  $\theta$ . Thus, for any  $x \in \mathcal{X}$ , we have

$$\tilde{R}(B|x) = \int_{\Theta} P(B|\theta) Q(d\theta|x) > 0.$$

(5) Since  $M(B) = 0$ ,  $P(B|\theta) = 0$  for all  $\theta$ . Thus, for any  $x \in \mathcal{X}$ ,  $\tilde{R}(B|x) = 0$ . □

While the SFP condition does not hold in Example 1, it does hold when  $P(dx|\theta)$  is an exponential family on  $\mathbb{R}^p$  [see; e.g., 2, Chapter 1]. Here is another example where the SFP condition holds.

*Example 2.* Let  $f : \mathbb{R}^p \rightarrow [0, \infty)$  be a density with respect to Lebesgue measure and suppose that  $X_1, \dots, X_n$  are independent and identically distributed random vectors from  $f(x_1 - \theta)$  where  $\theta \in \Theta \subset \mathbb{R}^p$ . In this case,  $\mathcal{X} = \mathbb{R}^{pn}$  and with  $dx$  denoting Lebesgue measure on  $\mathcal{X}$ ,  $P(dx|\theta) = p(x|\theta)dx$  where  $x = (x_1, \dots, x_n)$  and

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

Suppose that  $f > 0$ . Then it’s easy to see that the collection of sets  $\{B \in \mathcal{B}(\mathcal{X}) : P(B|\theta) > 0\}$  is the same for all  $\theta \in \Theta$ . Thus, Lemma 5 implies that for any  $\sigma$ -finite, improper prior  $\nu(d\theta)$ , the resulting  $W$  and  $\tilde{W}$  will be  $\nu$ -irreducible and  $M$ -irreducible, respectively. We return to this example later.

We now examine a more complicated example where there are sets  $C$  with  $\nu(C) > 0$  and  $R(C|\theta) = 0$ .

*Example 3.* Let  $\mathcal{X} = \Theta = \mathbb{R}$  and  $P(dx|\theta) = p(x|\theta)dx$  where

$$p(x|\theta) = I_{(\theta, \theta+1)}(x) ,$$

and  $dx$  denotes Lebesgue measure on  $\mathcal{X}$ . If we take the prior distribution to be  $\nu(d\theta) = d\theta$ , it's easy to show that the marginal is  $M(dx) = dx$ , which is clearly  $\sigma$ -finite. The posterior is given by  $Q(d\theta|x) = q(\theta|x)d\theta$  where

$$q(\theta|x) = I_{(x-1, x)}(\theta) .$$

It follows that  $R(d\theta|\eta) = r(\theta|\eta)d\theta$  where

$$r(\theta|\eta) = (1 + (\theta - \eta))I_{(-1, 0)}(\theta - \eta) + (1 - (\theta - \eta))I_{(0, 1)}(\theta - \eta) .$$

Since  $\eta$  is a location parameter in the density  $r(\theta|\eta)$ , the Markov chain  $W$  can be expressed as a random walk

$$W_{n+1} = W_n + Z_{n+1} ,$$

where  $Z_1, Z_2, \dots$  is an independent and identically distributed sequence of random variables with density given by

$$f(z) = (1 + z)I_{(-1, 0)}(z) + (1 - z)I_{(0, 1)}(z) .$$

Note that, if  $W_0 = w$ , then  $W_1 \in (w - 1, w + 1)$  with probability one, so there are many sets with positive Lebesgue measure that are not accessible from  $w$  in a single step. On the other hand, for any  $\delta \in (0, 1)$ , we have

$$P(Z_1 \in (0, \delta)) = P(Z_1 \in (-\delta, 0)) > 0 .$$

This implies that the chain  $W$  can make arbitrarily small jumps in either direction. Hence, it should be intuitively clear that any set  $C$  with positive Lebesgue measure is accessible from any point  $\theta \in \mathbb{R}$  and hence that  $W$  is  $\nu$ -irreducible. A formal proof of this irreducibility can be constructed using the ideas in Section 4.3.3 of [15]. This ends Example 3.

The next example shows that  $W$  is not always  $\nu$ -irreducible.

*Example 4.* Let  $\mathcal{X} = \mathbb{R} \setminus \{0\}$ ,  $\Theta = \mathbb{R}$  and  $P(dx|\theta) = p(x|\theta)dx$  where

$$p(x|\theta) = e^{-(x-\theta)}I_{(\theta, \infty)}(x)I_{[0, \infty)}(\theta) + e^{x-\theta}I_{(-\infty, \theta)}(x)I_{(-\infty, 0)}(\theta) ,$$

and  $dx$  denotes Lebesgue measure on  $\mathcal{X}$ . If we take the prior distribution to be Lebesgue measure on  $\Theta$ ; that is,  $\nu(d\theta) = d\theta$ , then it's straightforward to show that the posterior distribution is given by  $Q(d\theta|x) = q(\theta|x)d\theta$  where

$$q(\theta|x) = \begin{cases} \frac{1}{1-e^{-x}}e^{-(x-\theta)}I_{(0, x)}(\theta) & x > 0 \\ \frac{1}{1-e^x}e^{x-\theta}I_{(x, 0)}(\theta) & x < 0 . \end{cases}$$

The Markov transition function of  $W$  is  $R(d\theta|\eta) = r(\theta|\eta)d\theta$  where  $r(\theta|\eta) = \int_{\mathcal{X}} q(\theta|x)p(x|\eta)dx$ . Now  $p(x|\eta) = 0$  whenever  $x < 0$  and  $\eta \geq 0$ . Also,  $q(\theta|x) = 0$  whenever  $\theta < 0$  and  $x > 0$ . Consequently, if  $\eta \geq 0$  and  $\theta < 0$ , we have

$$r(\theta|\eta) = \int_{\mathcal{X}} q(\theta|x)p(x|\eta)dx = \int_{-\infty}^0 q(\theta|x)p(x|\eta)dx + \int_0^{\infty} q(\theta|x)p(x|\eta)dx = 0 .$$

Thus, if  $\eta \geq 0$ , we have

$$R((-\infty, 0)|\eta) = \int_{-\infty}^0 r(\theta|\eta)d\theta = 0 ,$$

and a simple induction argument shows that  $R^n((-\infty, 0)|\eta) = 0$  for all  $\eta \geq 0$  and all  $n \in \mathbb{N}$ . Thus, the set  $(-\infty, 0)$ , which has positive  $\nu$ -measure, is not accessible from any  $\eta \geq 0$  and it follows that  $W$  is not  $\nu$ -irreducible. In fact, if  $W$  is started in  $[0, \infty)$ , it stays in  $[0, \infty)$  forever, and if  $W$  starts in  $(-\infty, 0)$ , it stays in  $(-\infty, 0)$  forever. This ends Example 4.

Several of the results developed in this paper are applicable only in situations where  $W$  and  $\widetilde{W}$  are irreducible. As shown above, establishing this irreducibility generally requires direct analysis of the Markov chains. One exception is when the (easily checked) SFP condition holds and the irreducibility conditions are automatically satisfied. In particular, we have the following simplified version of Corollary 2.

**Corollary 4.** *Suppose that  $P(dx|\theta)$  satisfies the SFP condition and that (3) holds so that  $\nu$  is strongly admissible. If  $g : \Theta \rightarrow [0, \infty)$  is bounded, then the prior  $\nu_g(d\theta) = g(\theta)\nu(d\theta)$  is also strongly admissible.*

*Example 5.* Let  $X$  be a random vector from a  $p$ -variate normal distribution with mean vector  $\theta \in \Theta = \mathbb{R}^p$  and identity covariance matrix; i.e.,  $X \sim N_p(\theta, I_p)$ . Assume that  $p \geq 3$ . Lai [14] considers priors of the form

$$\nu(d\theta) = (a + \|\theta\|^2)^{-b}d\theta ,$$

where  $a > 0$ ,  $b \in \mathbb{R}$ ,  $\|\theta\|^2 = \sum_{i=1}^p \theta_i^2$  and  $d\theta$  denotes Lebesgue measure on  $\mathbb{R}^p$ . It is straightforward to show that  $\nu(\Theta) = \infty$  if and only if  $b \leq p/2$  and that the marginal measure is  $\sigma$ -finite for any  $b$  in this range. Together with Wen-Lin Lai, the authors have recently shown that Eaton's Markov chain,  $\{W_n\}_{n=0}^{\infty}$ , is locally- $\nu$ -recurrent as long as  $b \geq \frac{p}{2} - 1$ . Thus,  $\nu(d\theta)$  is improper and strongly admissible for all  $b \in [\frac{p}{2} - 1, \frac{p}{2}]$ .

Now suppose that  $b \in [\frac{p}{2} - 1, \frac{p}{2}]$ ,  $g : \mathbb{R}^p \rightarrow [0, \infty)$  is bounded and  $\nu_g(d\theta) = \nu(d\theta)g(\theta)$  is improper. The SFP condition is clearly satisfied in this example. Hence, Corollary 4 is applicable and implies that  $\nu_g(d\theta)$  is strongly admissible. This ends Example 5.

Clearly, a general result that would allow one to check the irreducibility conditions without direct analysis of the Markov chains would be extremely useful and, as we now explain, may be attainable.

Note that in Example 4, there are sets  $B \subset \mathcal{X}$  and  $C \subset \Theta$  such that  $P(B|\theta) = 0$  for all  $\theta \in C$  and  $P(\overline{B}|\theta) = 0$  for all  $\theta \in \overline{C}$ . It is not hard to show that (under regularity conditions implying that the sets and their complements have positive measure), the existence of such sets implies that the chains are reducible. We believe that this existence is also necessary for reducibility, although we have been able to prove this only for the case where one or both of  $\mathcal{X}$  and  $\Theta$  have a countable number of elements. If our conjecture is correct, one could establish irreducibility simply by showing that sets  $B$  and  $C$  satisfying the conditions above do not exist.

## 4 Priors for estimation problems with restricted parameter spaces

Recall that “Problem A” is the one with parameter space  $\Theta$ , sample space  $\mathcal{X}$  and statistical model  $P(dx|\theta)$ , whereas “Problem B” is the same as Problem A except that the parameter space is  $\Theta^* \in \mathcal{B}(\Theta)$ , a proper subset of  $\Theta$ . We assume throughout this section that  $\nu$  is a  $\sigma$ -finite improper prior on  $\Theta$  that satisfies (3) so that  $W$  is locally- $\nu$ -recurrent (and  $\nu$  is strongly admissible). Let  $\nu^*$  be the measure  $\nu$  restricted to  $\Theta^*$  and assume  $\nu^*$  is improper. We now show that, under mild irreducibility conditions,  $\nu^*$  is strongly admissible for Problem B.

For the time being, we focus on Problem A and the perturbed prior  $\nu_g(d\theta) = I_{\Theta^*}(\theta)\nu(d\theta)$  with associated marginal  $M_g$ . Assume that Conditions 1 and 2 of Corollary 2 hold so that  $W^g$  and  $\widetilde{W}^g$  are both recurrent. To connect Problem A with Problem B we need to examine the Dirichlet form corresponding to the  $M_g$ -symmetric Markov chain  $\widetilde{W}^g$ . For an  $M_g$ -proper set  $B$ , define

$$\widetilde{V}_g(B) = \{h \in L_2(M_g) : h \text{ is bounded, } h(x) \geq I_B(x)\} . \quad (10)$$

The Dirichlet form associated with the Markov chain  $\widetilde{W}^g$  is defined for  $h \in L_2(M_g)$  as

$$\widetilde{\Delta}_g(h) = \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} (h(x) - h(y))^2 \widetilde{R}_g(dx|y) M_g(dy) . \quad (11)$$

Now, since  $\widetilde{W}^g$  is recurrent (and hence locally- $M_g$ -recurrent), we have

$$\inf_{h \in \widetilde{V}_g(B)} \widetilde{\Delta}_g(h) = 0 \quad \text{for each } M_g\text{-proper set } B . \quad (12)$$

Now, back to Problem B. Let  $W^*$  and  $\widetilde{W}^*$  denote the Markov chains associated with Problem B; that is, they are the analogues of  $W$  and  $\widetilde{W}$  from Problem A. (There are actually three different decision

problems and six different Markov chains under consideration here.) Note that

$$\begin{aligned}
\tilde{\Delta}_g(h) &= \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} (h(x) - h(y))^2 \tilde{R}_g(dx|y) M_g(dy) \\
&= \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} (h(x) - h(y))^2 \int_{\Theta} P(dx|\theta) P(dy|\theta) \nu_g(d\theta) \\
&= \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} (h(x) - h(y))^2 \int_{\Theta^*} P(dx|\theta) P(dy|\theta) \nu^*(d\theta) \\
&= \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} (h(x) - h(y))^2 \tilde{R}^*(dx|y) M^*(dy) ,
\end{aligned}$$

where  $\tilde{R}^*(dx|y)$  is the Markov transition function of  $\tilde{W}^*$  and  $M^*$  is the marginal corresponding to  $\nu^*$  for Problem B. Therefore,  $\tilde{\Delta}_g(h)$  is simultaneously the Dirichlet form for both  $\tilde{W}^g$  and  $\tilde{W}^*$ . Furthermore, since  $\nu_g(d\theta) = I_{\Theta^*}(\theta)\nu(d\theta)$ , we have

$$M_g(B) = \int_{\Theta} P(B|\theta)\nu_g(d\theta) = \int_{\Theta^*} P(B|\theta)\nu^*(d\theta) = M^*(B) .$$

Hence,  $B$  is  $M_g$ -proper if and only if  $B$  is  $M^*$ -proper and  $L_2(M_g) = L_2(M^*)$ . Therefore, (12) implies that  $\tilde{W}^*$  is recurrent. Finally, if  $W^*$  is  $\nu^*$ -irreducible (or if  $\tilde{W}^*$  is  $M^*$ -irreducible), then an argument similar to that used in the proof of Corollary 2 shows that  $\nu^*$  is strongly admissible in Problem B. We now continue with the translation example introduced in the previous section.

*Example 2 cont.* Recall that  $X_1, \dots, X_n$  are independent and identically distributed random vectors from  $f(x_1 - \theta)$  where  $f$  is a density on  $\mathbb{R}^p$  and  $\theta \in \Theta^* \subset \mathbb{R}^p$  is an unknown parameter to be estimated. We assume that  $f > 0$  so that the SFP condition holds. To preclude uninteresting cases, assume that  $\Theta^*$  has infinite Lebesgue measure and that  $\mathbb{R}^p \setminus \Theta^*$  has positive Lebesgue measure. We will use our results to show that (under regularity conditions), Lebesgue measure restricted to  $\Theta^*$  is strongly admissible when  $p \in \{1, 2\}$ .

Consider the unrestricted version of the problem where the parameter space is  $\Theta = \mathbb{R}^p$  and take the prior distribution to be Lebesgue measure on  $\mathbb{R}^p$ ; i.e.,  $\nu(d\theta) = d\theta$ . Assume, for simplicity, that

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

is in  $(0, \infty)$  for all  $x$ . Then a version of the posterior distribution is given by

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

It follows that  $R(d\theta|\eta) = r(\theta|\eta)d\theta$  where the density  $r(\theta|\eta)$  is given by

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

Eaton [8] shows that  $r(\theta|\eta)$  is only a function of  $\theta - \eta$  and that we can write  $r(\theta|\eta) = r(\theta - \eta|0) =: t(\theta - \eta)$  where  $t$  is a symmetric density with respect to Lebesgue measure; i.e.,  $t(u) = t(-u)$  for all

$u \in \mathbb{R}^p$  and  $\int_{\mathbb{R}^p} t(u) du = 1$ . It follows that the Markov transition function is  $R(d\theta|\eta) = t(\theta - \eta)d\theta$ , which implies that  $W$  can be expressed as a random walk

$$W_{n+1} = W_n + U_{n+1} ,$$

where  $U_1, U_2, \dots$  is an independent and identically distributed sequence of random vectors with density  $t(u)$ . A sufficient condition for recurrence of  $W$  when  $p = 1$  is  $\int_{\mathbb{R}} |u|t(u)du < \infty$  [3, 15]. Eaton [5] shows that this condition holds if

$$\int_{\mathbb{R}} |x|f(x)dx < \infty . \quad (13)$$

Thus, if  $p = 1$  and (13) holds, we can appeal to the arguments above (the SFP condition ensures that all of the irreducibility conditions are satisfied) to conclude that Lebesgue measure on  $\Theta^* \subset \mathbb{R}$  is strongly admissible for the restricted parameter problem. Similar arguments show that if  $p = 2$  and  $\int_{\mathbb{R}^2} \|x\|^2 f(x)dx < \infty$ , then Lebesgue measure on  $\Theta^* \subset \mathbb{R}^2$  is strongly admissible for the restricted parameter problem. This argument cannot be extended beyond  $p = 2$  because there are no recurrent random walks on  $\mathbb{R}^p$  (driven by increment distributions having densities with respect to Lebesgue measure on  $\mathbb{R}^p$ ) when  $p > 2$ . As Eaton [6] says, ‘‘The parallel with Stein estimation is obvious.’’ This ends Example 2.

## A Appendix: Relationships between Dirichlet forms

The following result was stated but not proved in Eaton [5, Remark 3.3].

**Proposition 1.** *If there exists an  $\varepsilon > 0$  such that  $\varepsilon < g(\theta) < 1/\varepsilon$  for all  $\theta \in \Theta$ , then  $\Delta_g(h) \leq \varepsilon^{-3}\Delta(h)$ .*

*Proof.* Define  $\hat{g}(x) = \int_{\Theta} g(\theta)Q(d\theta|x)$ . Note that  $\hat{g}(x) \geq \varepsilon$  and hence  $1/\hat{g}(x) \leq 1/\varepsilon$ . It is clear from (1) that  $\hat{g}(x)$  is the Radon-Nikodym derivative of  $M_g$  with respect to  $M$ ; i.e.,  $M_g(dx) = \hat{g}(x)M(dx)$ . It follows that the set  $\{x \in \mathcal{X} : \hat{g}(x) = 0\}$  has  $M_g$ -measure 0. Now, following [8], a version of  $Q_g(d\theta|x)$  is given by

$$\begin{cases} \frac{g(\theta)}{\hat{g}(x)}Q(d\theta|x) & \text{if } \hat{g}(x) > 0 \\ Q(d\theta|x) & \text{if } \hat{g}(x) = 0 . \end{cases}$$

Recall that

$$\Delta_g(h) = \int_{\Theta} \int_{\Theta} (h(\theta) - h(\eta))^2 S_g(d\theta, d\eta) .$$

Now using (4) we have

$$\begin{aligned} S_g(d\theta, d\eta) &= \int_{\mathcal{X}} Q_g(d\theta|x)P(dx|\eta)g(\eta)\nu(d\eta) \\ &= \int_{\mathcal{X}} Q_g(d\theta|x)Q_g(d\eta|x)M_g(dx) \\ &= g(\theta)g(\eta) \int_{\mathcal{X}} \frac{Q(d\theta|x)Q(d\eta|x)}{\hat{g}(x)}M(dx) . \end{aligned}$$



Putting all of this together, we have

$$\begin{aligned}
\Delta_g(h) &= \int_{\Theta} \int_{\Theta} (h(\theta) - h(\eta))^2 S_g(d\theta, d\eta) \\
&= \int_{\Theta} \int_{\Theta} g(\theta)g(\eta)(h(\theta) - h(\eta))^2 \int_{\mathcal{X}} \frac{Q(d\theta|x)Q(d\eta|x)}{\hat{g}(x)} M(dx) \\
&\leq \varepsilon^{-3} \int_{\Theta} \int_{\Theta} (h(\theta) - h(\eta))^2 \int_{\mathcal{X}} Q(d\theta|x)Q(d\eta|x) M(dx) \\
&= \varepsilon^{-3} \int_{\Theta} \int_{\Theta} (h(\theta) - h(\eta))^2 S(d\theta, d\eta) \\
&= \varepsilon^{-3} \Delta(h) .
\end{aligned}$$

□

We now prove Theorem 4, which is restated here for convenience.

**Theorem 4.** *If  $\widetilde{W}$  is locally- $M$ -recurrent and  $g : \Theta \rightarrow [0, \infty)$  is bounded, then  $\widetilde{W}^g$  is locally- $M_g$ -recurrent.*

*Proof.* Since  $g$  is nonnegative and bounded, there exists a  $K > 0$  such that  $g(\theta) < K$  for all  $\theta \in \Theta$ . Let  $\tilde{\Delta}$  denote the Dirichlet form associated with the Markov chain  $\widetilde{W}$ ; that is, for  $h \in L_2(M)$ , define

$$\tilde{\Delta}(h) = \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} (h(x) - h(y))^2 \tilde{R}(dx|y) M(dy) .$$

Also, for each  $M$ -proper set  $B$ , define

$$\tilde{V}(B) = \{h \in L_2(M) : h \text{ is bounded, } h(x) \geq I_B(x)\} .$$

Since  $\widetilde{W}$  is locally- $M$ -recurrent, Theorem 2 implies that

$$\inf_{h \in \tilde{V}(B)} \tilde{\Delta}(h) = 0 \quad \text{for each } M\text{-proper set } B .$$

Let  $\tilde{\Delta}_g$  and  $\tilde{V}_g$  be the analogues of  $\tilde{\Delta}$  and  $\tilde{V}$  for the perturbed problem. These quantities were defined at (10) and (11). Clearly,  $M(B) < \infty \Rightarrow M_g(B) < \infty$  and  $L_2(M) \subseteq L_2(M_g)$ . Hence, if  $B$  is  $M$ -proper and  $M_g(B) > 0$ , then  $B$  is  $M_g$ -proper and  $\tilde{V}(B) \subseteq \tilde{V}_g(B)$ . It's also easy to see that, for  $h \in L_2(M)$ ,  $\tilde{\Delta}_g(h) \leq K \tilde{\Delta}(h)$ . Therefore, if  $B$  is  $M$ -proper and  $M_g(B) > 0$ , we have

$$\inf_{h \in \tilde{V}_g(B)} \tilde{\Delta}_g(h) \leq \inf_{h \in \tilde{V}(B)} \tilde{\Delta}_g(h) \leq K \inf_{h \in \tilde{V}(B)} \tilde{\Delta}(h) .$$

Now since  $M$  is  $\sigma$ -finite, there exists a sequence of  $M$ -proper sets  $B_1 \subset B_2 \subset \dots$  such that  $\cup_{i=1}^{\infty} B_i = \mathcal{X}$ . The local- $M$ -recurrence of  $\widetilde{W}$  implies that for each  $i \in \mathbb{N}$ ,

$$\inf_{h \in \tilde{V}(B_i)} \tilde{\Delta}(h) = 0 .$$

Without loss of generality, we may assume that  $M_g(B_i) > 0$  for all  $i \in \mathbb{N}$ . Putting all of this together, we have a sequence of  $M_g$ -proper sets,  $B_1 \subset B_2 \subset \dots$  such that  $\cup_{i=1}^{\infty} B_i = \mathcal{X}$  and such that for each  $i \in \mathbb{N}$ , we have

$$\inf_{h \in \tilde{V}_g(B_i)} \tilde{\Delta}_g(h) = 0 .$$

The result now follows from Theorem 3.2 of Eaton [8]. □

## Acknowledgments

The authors are grateful to an anonymous referee for helpful comments and suggestions. Hobert's research partially supported by NSF Grants DMS-00-72827 and DMS-05-03648.

## References

- [1] Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Annals of Mathematical Statistics*, 42:855–904.
- [2] Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications to Statistical Decision Theory*. Institute of Mathematical Statistics, Hayward, California.
- [3] Chung, K. L. and Fuchs, W. H. J. (1951). On the distribution of values of sums of random variables. *Memoirs of the American Mathematical Society*, 6:1–12.
- [4] Eaton, M. L. (1982). A method for evaluating improper prior distributions. In Gupta, S. S. and Berger, J. O., editors, *Statistical Decision Theory and Related Topics III*, volume 1, New York. Academic.
- [5] Eaton, M. L. (1992). A statistical diptych: Admissible inferences-recurrence of symmetric Markov chains. *Annals of Statistics*, 20:1147–1179.
- [6] Eaton, M. L. (1997). Admissibility in quadratically regular problems and recurrence of symmetric Markov chains: Why the connection? *Journal of Statistical Planning and Inference*, 64:231–247.
- [7] Eaton, M. L. (2001). Markov chain conditions for admissibility in estimation problems with quadratic loss. In de Gunst, M., Klaassen, C., and van der Vaart, A., editors, *State of the Art in Probability and Statistics - A Festschrift for Willem R. van Zwet*, volume 36 of *The IMS Lecture Notes Series*. IMS, Beachwood, Ohio.

- [8] Eaton, M. L. (2004). Evaluating improper priors and the recurrence of symmetric Markov chains: An overview. In Dasgupta, A., editor, *A Festschrift to Honor Herman Rubin*, volume 45 of *The IMS Lecture Notes Series*. IMS, Beachwood, Ohio.
- [9] Hobert, J. P., Marchev, D., and Schweinsberg, J. (2004). Stability of the tail Markov chain and the evaluation of improper priors for an exponential rate parameter. *Bernoulli*, 10:549–564.
- [10] Hobert, J. P. and Robert, C. P. (1999). Eaton’s Markov chain, its conjugate partner and  $\mathcal{P}$ -admissibility. *Annals of Statistics*, 27:361–373.
- [11] Hobert, J. P. and Schweinsberg, J. (2002). Conditions for recurrence and transience of a Markov chain on  $\mathbb{Z}^+$  and estimation of a geometric success probability. *Annals of Statistics*, 30:1214–1223.
- [12] Johnstone, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Annals of Statistics*, 12:1173–1198.
- [13] Johnstone, I. (1986). Admissible estimation, Dirichlet principles and recurrence of birth-death chains on  $Z_+^p$ . *Probability Theory and Related Fields*, 71:231–269.
- [14] Lai, W.-L. (1996). *Admissibility and the Recurrence of Markov Chains with Applications*. PhD thesis, University of Minnesota.
- [15] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London.