The Annals of Statistics
1992, Vol. 20, No. 3, 1147–1179

SPECIAL INVITED PAPER

A STATISTICAL DIPTYCH: ADMISSIBLE INFERENCES—RECURRENCE OF SYMMETRIC MARKOV CHAINS

BY MORRIS L. EATON[1]

University of Minnesota

Given a parametric model and an improper prior distribution, the formal posterior distribution induces decision rules in any decision problem. The results here provide conditions under which this formal Bayes method produces admissible decision rules for all quadratically regular decision problems. The conditions derived are shown to be equivalent to the recurrence of a natural symmetric Markov chain (on the parameter space) generated by the model and the improper prior. The results are also used to give conditions under which formal predictive distributions are admissible decision rules in certain prediction problems.

1. Introduction. The formal Bayes method for deriving inferential procedures occupies a significant portion of both the decision theoretic and Bayesian literature. The formal Bayes representation of estimators is a standard strategy for attempting to establish admissibility; for example, see Karlin (1958), Stein (1959, 1965), Židek (1970), Portnoy (1971), Clevenson and Židek (1977), Berger and Srinivasan (1978) and Brown and Hwang (1982). In the Bayesian world arguments abound which attempt to justify the use of "flat," "uninformative" or "reference" prior distributions (typically improper), and implicitly, the posterior distributions these generate; see Berger (1985) for a discussion and references. Of course, any posterior distribution allows a Bayesian to solve decision problems; one just chooses actions to minimize posterior risk.

A mathematical formulation of the formal Bayes method requires some care. Given a statistical model $P(dx|\theta)$ on a sample space $\mathscr{X}$ and a σ-finite improper prior distribution ν on the parameter space Θ, the marginal measure on $\mathscr{X}$,

$$M(dx) = \int P(dx|\theta)\nu(d\theta), \tag{1.1}$$

may be badly behaved (i.e., not σ finite). However, when M is σ finite ($\mathscr{X}$ and

Received March 1990; revised February 1992.

[1]This work was supported in part by NSF Grants DMS-88-00284 and DMS-89-22607. Portions of this work were presented in a Special Invited Paper given at an IMS Regional Meeting in Davis, California in 1989.

AMS 1980 *subject classifications*. Primary 62C05; secondary 62C10, 62C15.

Key words and phrases. Improper prior distributions, formal posterior distributions, formal Bayes rules, almost admissibility, recurrence of symmetric Markov chains, prediction, predictive distributions.

Θ are assumed to be Polish with their Borel σ algebras), the formal posterior distribution on Θ, $Q(d\theta|x)$, exists and satisfies

$$P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx); \tag{1.2}$$

the equality means that the two measures on $\mathscr{X} \times \Theta$ agree. That is, $Q(\cdot|x)$ is a probability measure for each x, and for each measurable subset $B \subseteq \Theta$, $Q(B|\cdot)$ is measurable. In addition, Q is unique in the sense that if $\tilde{Q}$ also satisfies (1.2), then there is an M-null set B_0 such that $x \notin B_0$ implies $Q(\cdot|x) = \tilde{Q}(\cdot|x)$. For a discussion, see Eaton (1982); an attempt to circumvent the σ-finiteness assumption on M occurs in Hartigan (1983). Throughout this paper both ν and M are assumed to be σ finite, so a formal posterior exists.

Given an action space A and a nonnegative loss function L, a formal Bayes solution to the decision problem is any function $a(x) \in A$, which for each x satisfies

$$\int L(a,\theta,x)Q(d\theta|x) \geq \int L(a(x),\theta,x)Q(d\theta|x) \tag{1.3}$$

for all $a \in A$ (we are ignoring existence and measurability issues here). For example, if $A = R^n$ and $\varphi(\theta, x) \in R^n$ is any bounded measurable function of θ and x, then

$$\hat{\varphi}(x) = \int \varphi(\theta,x)Q(d\theta|x) \tag{1.4}$$

is a formal Bayes estimator of φ when the loss function is

$$L(a,\theta,x) = (a - \varphi(\theta,x))'(a - \varphi(\theta,x)). \tag{1.5}$$

Here, the prime denotes the transpose of vectors in R^n.

The results in this paper focus on the general question:

(1.6) Under what conditions on the model and the improper prior will the formal Bayes method produce "reasonable" decision rules?

The notion of reasonable adopted here is Stein's notion of almost-ν-admissibility (a-ν-a), which is defined formally in Section 2. The class of decision problems to which our results apply includes the *quadratically regular* problems introduced in Section 2. The quadratically regular problems include:

1. The problem of estimating a bounded measurable function of θ and x when the loss is quadratic [as in (1.4) and (1.5)].
2. A wide class of prediction problems when the loss is a quadratic measure of distance between distributions (see Section 6 for a precise formulation).

The precise mathematical problem studied in this paper is:

(1.7) For quadratically regular problems, what conditions on the model and ν will imply that the formal Bayes decision rules defined by ν are a-ν-a?

A main result in this paper, described in Theorem 1, answers (1.7) in the present generality. This result and its connection with Markov chains is described in Sections 2 through 5.

The conditions for a-ν-a are expressed in terms of the behavior of the transition function

$$R(d\theta|\eta) = \int_{\mathscr{X}} Q(d\theta|x)P(dx|\eta). \tag{1.8}$$

For $B \subseteq \Theta$, $R(B|\eta)$ is the average (over $\mathscr{X}$) probability assigned to B by the formal posterior $Q(\cdot|x)$ when X is sampled from $P(\cdot|\eta)$. Thus from the Bayesian viewpoint, it seems reasonable to call R the *pre-posterior transition function* since averaging over the sample space has taken place. It should be observed that R is defined only in terms of the model and the improper prior ν. Conditions on the behavior of R supply one answer to (1.7). For a measurable subset $C \subseteq \Theta$ satisfying $0 < \nu(C) < +\infty$, consider the class of real valued functions on Θ:

$$V(C) = \left\{h \middle| \int h^2\, d\nu < +\infty, h \geq 0, h(\theta) \geq 1 \text{ for } \theta \in C\right\}. \tag{1.9}$$

For $h \in V(C)$, set

$$\Delta(h) = \iint (h(\theta) - h(\eta))^2 R(d\theta|\eta)\nu(d\eta). \tag{1.10}$$

Here is a key result.

THEOREM 1.1. *If for each C satisfying* $0 < \nu(C) < +\infty$,

$$\inf_{h \in V(C)} \Delta(h) = 0, \tag{1.11}$$

then for all quadratically regular decision problems, the formal Bayes rules induced by ν are a-ν-a.

The proof of Theorem 1, given in Sections 2 and 3, uses Blyth's method [Blyth (1951)] and an application of the Cauchy–Schwarz inequality described in Appendix 1. Obviously, the function $h^* = 1$ yields $\Delta(h^*) = 0$, but h^* is not in $V(C)$. Thus, condition (1.11) can be interpreted as the extent to which h^* can be approximated by functions in $V(C)$. An application of Theorem 1 to random samples from one-dimensional translation families which have means appears in Section 3. This application shows that (1.11) holds when ν is Lebesgue measure, so the formal Bayes rules are all a-ν-a, thus providing some justification for using Lebesgue measure as an improper prior distribution in this problem.

Because the inf in (1.11) is typically not achieved by functions in $V(C)$, the successful application of Theorem 1 depends on describing "approximate" minimizers of Δ in (1.10). This leads to the introduction of discrete time Θ-valued Markov chain whose transition function is $R(\cdot|\eta)$ defined in (1.8). To see this connection, let $K \supseteq C$ satisfy $\nu(K) < +\infty$ and let

$$V(C, K) = \{h \in V(C) | h(\theta) = 0 \text{ for } \theta \in K^c\}. \tag{1.12}$$

In Appendix 2, a minimizer of Δ over $V(C, K)$ is characterized as a certain "hitting probability" of the Markov chain defined by R. Using this result and letting K increase to Θ yields a connection between recurrence properties of the chain and (1.11). To be precise, let $W = (\eta, W_1, W_2, \dots)$ be the Markov chain which starts at η and evolves according to R. Consider the stopping time

$$\sigma_C = \begin{cases} \text{first } n \geq 1, & \text{with } W_n \in C, \\ +\infty, & \text{if } W_n \notin C \text{ for all } n \geq 1. \end{cases}$$

THEOREM 1.2. *For each C with* $0 < \nu(C) < +\infty$,

$$\inf_{h \in V(C)} \Delta(h) = \int_C [1 - \Pr\{\sigma_C < +\infty | W_0 = \eta\}] \nu(d\eta), \tag{1.13}$$

where W_0 is the initial state of the chain.

Now, if (1.11) holds, then for each C, the integral over C in (1.13) is zero. This means that for each $\eta \in C$ (except for a ν-null set), when the chain starts at η, it returns to C w.p.1. Therefore (1.11) is equivalent to a recurrence property of W (called *local ν recurrence* here).

The technical details involving the connection between Δ and the chain W are given in Appendix 2. The connection established there is valid for any ν-symmetric chain (see Appendix 2 for the definition of ν symmetry), and not just for chains whose transition functions have the form (1.8). The arguments proceed from first principles and are valid for any Polish space. Fortunately, a discussion of the rather technical matter of irreducibility has been avoided [see Nummelin (1983) for such matters]. For the countable state space case when the chain is irreducible, some similar-looking results appear in Griffeath and Liggett (1982) [also see Lyons (1983)].

Even though the minimizers of Δ over $V(C, K)$ can be characterized, they remain elusive. For the case $\Theta = [0, \infty)$, a heuristic method for finding "approximate" minimizers of Δ appears in Section 5. The method is successfully applied to the one-dimensional Poisson case.

The results given here provide one possible criterion for the evaluation of the improper prior ν and hence the induced posterior distribution. An alternative formulation, described in Example 2.1, which concentrated directly on the formal posterior $Q(\cdot|x)$ was introduced in Eaton (1982). The idea there was to regard $Q(\cdot|x)$ as a decision rule [also called an *inference* in Eaton (1982)], and ask for conditions under which this decision rule is a-ν-a. This approach led to the introduction of fair Bayes loss functions. The admissibility of $Q(\cdot|x)$ for a variety of such loss functions is then regarded as evidence that the improper prior ν leads to sensible inferences. It was pointed out that the prediction problem could also be viewed this way. This viewpoint is developed further here.

The problem of predicting the value of some future observable random quantity on the basis of available data has received considerable attention in the statistical literature. The time series literature is replete with derivations of minimum mean squared error predictions, while the prediction of a future response, given values of covariates, is a classical problem in linear model theory which is ordinarily attacked via mean squared error considerations. No less attention is afforded the prediction problem in the Bayesian world, although the emphasis is somewhat different. Given a probabilistic model and a prior distribution, the Bayesian solution to the prediction problem is just the conditional distribution of the quantity to be predicted, given the data and the prior. This conditional distribution is called the *predictive distribution* and is discussed at length in the basic text by Aitchison and Dunsmore (1975). It has been argued in the literature that prediction, as opposed to say parametric estimation, is the proper activity of statisticians, partly because prediction is often the scientific question of interest and partly because the ability of statisticians to predict can actually be checked, unlike the popular parametric estimation-confidence set activity. For an introduction to this point of view and further references, see Geisser (1980).

In Section 6, we formulate the prediction problem as a decision theory problem with a fair Bayes loss function. The decision rules are the predictive distributions. It is then an easy matter to show the decision problem is quadratically regular and thus Theorem 1 applies when (1.11) holds. In particular, the same condition which establishes the a-ν-a of $\hat{\varphi}$ in (1.4) establishes the a-ν-a of predictive distributions. The fair Bayes estimation problems [as described in Eaton (1982)] are special cases of the prediction problems, so (1.11) is also a sufficient condition for the a-ν-a of formal posterior distributions on Θ—a result of some interest to Bayesians, since one has an explicit justification for the use of some improper priors.

Section 7 contains discussion concerning open problems and the relationship of the results here with other work on admissibility. In addition, it is pointed out that certain common groups which arise in invariant statistical problems, do not support any recurrent random walks. Therefore the sufficient condition for a-ν-a given in Theorem 1 cannot hold in these problems when an invariant prior is used. This fact strongly suggests that the routine use of invariant prior distributions for invariant problems is suspect in such cases. However, improving decision rules by modifying invariant priors remains an open problem.

2. Preliminaries.

2.1. *Blyth's lemma*. Here we review a standard technique for establishing admissibility—commonly known as Blyth's method [Blyth (1951)]. Consider a statistical decision problem with a risk function $R(\delta, \theta)$ where δ is a decision rule and $\theta \in \Theta$ is a parameter. Let ν denote a σ-finite measure on the measurable space $(\Theta, \mathscr{B})$ with $\nu \not\equiv 0$.

DEFINITION 2.1 [Stein (1965)]. A decision rule δ_0 is *almost-ν-admissible* (a-ν-a) if for any decision rule δ_1 which satisfies

$$R(\delta_1, \theta) \le R(\delta_0, \theta) \quad \text{for all } \theta,$$

the set

$$\{\theta | R(\delta_1, \theta) < R(\delta_0, \theta)\}$$

has ν measure zero.

Now, let U be the set of all nonnegative functions g defined on Θ such that

$$0 < \int g(\theta)\nu(d\theta) < +\infty.$$

For each $g \in U$, we assume there is a Bayes rule for the prior measure $g(\theta)\nu(d\theta)$; that is, a decision rule δ_g exists such that

$$\int R(\delta, \theta) g(\theta)\nu(d\theta) \ge \int R(\delta_g, \theta) g(\theta)\nu(d\theta) \tag{2.1}$$

for all δ.

A measurable subset $C \subseteq \Theta$ is ν *proper* if $0 < \nu(C) < +\infty$. For such a subset, let

$$U(C) = \{g \in U | g(\theta) \ge 1 \text{ for } \theta \in C\}. \tag{2.2}$$

The following sufficient condition for a-ν-a is a variation of Blyth's condition [Blyth (1951); also see Stein (1955), Židek (1970), Brown and Hwang (1982) and Berger (1985)].

PROPOSITION 2.1. *Let δ_0 be a decision rule. If for each ν-proper set C,*

$$\inf_{g \in U(C)} \int [R(\delta_0, \theta) - R(\delta_g, \theta)] g(\theta)\nu(d\theta) = 0, \tag{2.3}$$

then δ_0 is a-ν-a.

PROOF. The well-known proof by contradiction is omitted. □

Because ν is σ finite, (2.3) need only be verified for a countable number of C's. Here is a precise statement.

COROLLARY 2.1. *Let $\{C_n | n = 1, 2, \dots\}$ be any collection of ν-proper sets with $C_n \subseteq C_{n+1}$ and $\bigcup C_n = \Theta$. If* (2.3) *holds for each C_n, then δ_0 is* a-ν-a.

PROOF. A minor variation of the proof of Proposition 2.1 establishes this result. □

REMARK 2.1. In concrete problems, one usually establishes almost admissibility, and then uses a separate argument to try to obtain admissibility. For example, if one can show all finite valued risk functions are continuous and if ν assigns positive measure to all nonempty open sets, it is clear admissibility follows from almost admissibility. This technique is often applicable in examples dealing with exponential families.

In what follows, upper bounds on the integrated risk difference in (2.3) involve the variation distance between probability measures. Recall that if α_1 and α_2 are probabilities defined on the same measurable space, then

$$\|\alpha_1 - \alpha_2\| = 2\sup_B |\alpha_1(B) - \alpha_2(B)| \tag{2.4}$$

is the variation distance between α_1 and α_2. Here, the sup ranges over the relevant σ algebra. Further, if λ is any σ-finite measure which dominates α_1 and α_2, then

$$\|\alpha_1 - \alpha_2\| = \int |p_1 - p_2|\, d\lambda, \tag{2.5}$$

where $p_i = d\alpha_i/d\lambda$ is the Radon–Nikodym derivative.

2.2. *Quadratically regular decision problems.* The main class of decision problems to which our a-ν-a results apply are described below. To set notation, the sample space $\mathscr{X}$ and the parameter space Θ are assumed to be Polish spaces (complete separable metric spaces) equipped with the usual σ algebras $\mathscr{B}_1$ and $\mathscr{B}_2$. The available data $X \in \mathscr{X}$ are assumed to have a distribution belonging to the parametric family $\{P(\cdot|\theta)|\theta \in \Theta\}$. Throughout this paper, ν denotes a σ-finite improper prior measure on Θ $[\nu(\Theta) = +\infty]$. The sets U and $U(C)$ are as defined in Section 2.1.

The marginal measure on $\mathscr{X}$ defined by (1.1) is assumed to be σ finite. Thus, the formal posterior $Q(\cdot|x)$ exists (such objects are sometimes called transition functions) and is characterized by (1.2). Further, for each $g \in U$, the finite measure $g(\theta)\nu(d\theta)$ defines a finite marginal measure

$$M_g(dx) = \int P(dx|\theta)g(\theta)\nu(d\theta) \tag{2.6}$$

on $\mathscr{X}$, and thus a posterior distribution $Q_g(\cdot|x)$ exists, which is characterized by

$$P(dx|\theta)g(\theta)\nu(d\theta) = Q_g(d\theta|x)M_g(dx). \tag{2.7}$$

Now, consider a decision problem with an action space A and a nonnegative loss function $L(a, \theta, x)$ defined on $A \times \Theta \times \mathscr{X}$. It is assumed that for $Q(\cdot|x)$ and $Q_g(\cdot|x)$, there exist measurable functions $a_0(x)$ and $a_g(x)$ from $\mathscr{X}$ to A such that

$$\int L(a, \theta, x)Q(d\theta|x) \geq \int L(a_0(x), \theta, x)Q(d\theta|x) \tag{2.8}$$

and

$$\int L(a, \theta, x) Q_g(d\theta|x) \geq \int L\big(a_g(x), \theta, x\big) Q_g(d\theta|x) \tag{2.9}$$

for all $a \in A$. Let δ_0 and δ_g denote the decision rules defined by a_0 and a_g, respectively. Thus, by definition, δ_0 is a formal Bayes rule defined by ν [via $Q(\cdot|x)$] and δ_g is a Bayes rule when the prior measure is $g(\theta)\nu(d\theta)$. For any decision rule δ, $R(\delta, \theta)$ denotes the risk function of δ.

DEFINITION 2.2. Given the model and ν, the preceding decision problem is *quadratically regular* if there is a constant $K \in [0, \infty)$ such that for all $g \in U$,

$$\begin{aligned} &\int_\Theta \big[R(\delta_0, \theta) - R(\delta_g, \theta)\big] g(\theta)\nu(d\theta) \\ &\qquad \leq K \int_{\mathscr{X}} \big\|Q(\cdot|x) - Q_g(\cdot|x)\big\|^2 M_g(dx). \end{aligned} \tag{2.10}$$

Here, $\|\cdot\|$ denotes variation distance.

In the preceding definition, the constant K is allowed to depend on the model, on ν and on the loss function, but not on $g \in U$. Before discussing (2.10) further, we give two examples which provide some motivation for the definition. Of course, the upper bound in (2.10) is to be used in verifying the sufficient condition (2.3) for a-ν-a.

EXAMPLE 2.1. In this example we show that the estimation of bounded measurable functions with quadratic loss is a quadratically regular decision problem. Let $\varphi(\theta, x)$ take values in R^n and suppose each coordinate of φ is bounded in absolute value by K_1. With the action space $A = R^n$, consider a loss function L given by

$$L(a, \theta, x) = (a - \varphi(\theta, x))' B (a - \varphi(\theta, x)), \tag{2.11}$$

where B is a nonnegative definite matrix. Then

$$\hat{\varphi}_0(x) = \int \varphi(\theta, x) Q(d\theta|x) \tag{2.12}$$

defines a formal Bayes estimator of φ and

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

is a Bayes estimator for the finite prior measure $g(\theta)\nu(d\theta)$. The risk function of any (nonrandomized) estimator $\hat{\varphi}$ is

$$R(\hat{\varphi}, \theta) = \int (\hat{\varphi}(x) - \varphi(\theta, x))' B (\hat{\varphi}(x) - \varphi(\theta, x)) P(dx|\theta). \tag{2.14}$$

A routine calculation shows that

$$\int\left[R(\hat{\varphi}_0,\theta)-R(\hat{\varphi}_g,\theta)\right]g(\theta)\nu(d\theta) \tag{2.15}$$
$$=\int(\hat{\varphi}_0(x)-\hat{\varphi}_g(x))'B(\hat{\varphi}_0(x)-\hat{\varphi}_g(x))M_g(dx).$$

However, for each x, the integrand in the right-hand side of (2.15) is bounded above by

$$\lambda n K_1\|Q_0(\cdot|x)-Q_g(\cdot|x)\|^2, \tag{2.16}$$

where λ is the maximum eigenvalue of B. With $K=\lambda n K_1$, this decision problem is thus quadratically regular.

Two easy generalizations of this example are possible. First, the matrix B can depend on x as long as the maximum eigenvalue of $B(x)$ remains uniformly bounded. Second, $\varphi(\theta,x)$ can take values in a separable Hilbert space as long as the Hilbert space norm of φ is uniformly bounded in (θ,x). The easy details are omitted.

Three special cases of this decision problem deserve special mention. Assume $n=1$ and $B=1$. When φ is the indicator function of a subset of Θ, say D, then $\hat{\varphi}_0$ is the posterior probability of D under $Q(\cdot|x)$. Thus in a Bayesian context, the problem is to assess the a-ν-a of formal posterior probabilities. When φ is the indicator of a subset of $\Theta\times\mathscr{X}$, say H, then one can interpret φ as the coverage function; that is, the x section of H is a confidence set for θ. Then, $\hat{\varphi}(x)$ is an estimate of the coverage probability for the confidence set. This approach to the evaluation of confidence set procedures is discussed in Hwang and Brown (1991) as well as the references therein. In a testing situation, the decision theoretic formulation in Hwang, Casella, Robert, Wells and Farrell (1992) also results in a quadratically regular decision problem.

EXAMPLE 2.2. For this example, the action space is the set $\mathscr{M}_1(\Theta)$ of all probability measures on Θ. After seeing the data x, one is to announce a probability distribution on Θ which presumably reflects one's posterior opinions about θ. It was argued in Eaton (1982) that fair Bayes loss functions are most relevant for this problem. A class of such loss functions can be constructed as follows. Consider a bounded measurable kernel $k(\theta_1,\theta_2,x)$ defined on $\Theta\times\Theta\times\mathscr{X}$ taking values in R (complex valued kernels are also relevant for this problem, but we leave the extension to the reader). It is assumed that for each x, the kernel $k(\cdot,\cdot,x)$ is nonnegative definite and symmetric on bounded signed measures. In other words, the bilinear form $\langle\cdot,\cdot\rangle$ defined on bounded signed measures by

$$\langle\xi_1,\xi_2\rangle=\iint k(\theta_1,\theta_2,x)\xi_1(d\theta_1)\xi_2(d\theta_2) \tag{2.17}$$

is symmetric and $\langle\xi,\xi\rangle\geq 0$ for all ξ. Examples of such k's are easy to construct [see Eaton (1982)]. Note that the dependence of $\langle\cdot,\cdot\rangle$ on x is suppressed notationally.

Define a loss function by

$$L(a, \theta, x) = \langle a - \varepsilon_\theta, a - \varepsilon_\theta \rangle, \tag{2.18}$$

where ε_θ denotes the probability measure with mass 1 at θ. In effect, the bilinear form is used to define a squared "distance" between elements of $\mathscr{M}_1(\Theta)$ via the formula

$$\langle a_1 - a_2, a_1 - a_2 \rangle. \tag{2.19}$$

This squared distance is then used to define the loss function. The bilinear form $\langle \cdot , \cdot \rangle$ is not assumed to be positive definite, so it is possible that (2.19) is zero but $a_1 \neq a_2$.

Given any finite prior measure $g(\theta)\nu(d\theta)$ on Θ, it is verified in Eaton (1982) that a Bayes solution to the decision problem with loss function (2.18) is just the posterior $Q_g(\cdot|x)$ (this is the fair Bayes property). Using the arguments in Eaton (1982), it is easy to show that the decision problem with loss function (2.18) is a quadratically regular decision problem where the constant K in (2.10) is an upper bound on the absolute value of the kernel k. Thus our results on a-ν-a apply to this example.

Expressions of the form

$$\int \| Q(\cdot|x) - Q_g(\cdot|x) \|^2 M_g(dx) \tag{2.20}$$

have appeared elsewhere in work dealing with the approximation of formal posteriors by proper posteriors and the application of such ideas in decision theory [see Stein (1965)]. In some related work, Stone (1965) used the expression

$$\int \| Q(\cdot|x) - Q_g(\cdot|x) \| M_g(dx) \tag{2.21}$$

(when g is a density) to measure closeness of proper to improper posteriors [see Heath and Sudderth (1989) for a relationship of this to coherence]. The difference between (2.20) and (2.21) appears to be rather important. For example, it is not hard to construct cases where the inf over $U(C)$ of (2.20) is zero but the inf over $U(C)$ of (2.21) is positive. In most cases, it is difficult to show directly that the inf over $U(C)$ of (2.20) is zero. Rather, one first bounds (2.20) above by a more analytically tractable expression (in g), and then attempts to show the inf over $U(C)$ of the upper bound is zero. This we do in the next section.

3. The condition for almost admissibility. A main theorem, which provides a useful sufficient condition for the a-ν-a of formal Bayes rules in quadratically regular problems, is stated here. Let L_2 be the set of ν-square integrable functions. For a ν-proper set C, recall that

$$V(C) = \{ h \in L_2 | h \geq 0, h(\theta) \geq 1 \text{ for } \theta \in C \}. \tag{3.1}$$

The transition function $R(\cdot|\eta)$ defined in (1.8) appears here via the measure

$$T(d\theta, d\eta) = R(d\theta|\eta)\nu(d\eta) = \int_{\mathscr{X}} Q(d\theta|x)P(dx|\eta)\nu(d\eta) \tag{3.2}$$
$$= \int_{\mathscr{X}} Q(d\theta|x)Q(d\eta|x)M(dx)$$

defined on $\Theta \times \Theta$. Equation (3.2) shows that T is symmetric and has ν as its marginals. The discussion in Appendix 2 implies that

$$\Delta(h) = \iint (h(\theta) - h(\eta))^2 T(d\theta, d\eta) \tag{3.3}$$

described in the Introduction is well defined and finite for $h \in L_2$.

THEOREM 3.1. *The condition*

$$\inf_{h \in V(C)} \Delta(h) = 0 \quad \textit{for each } \nu\textit{-proper } C \tag{3.4}$$

is sufficient for the a-ν-a *of formal Bayes rules in quadratically regular problems.*

PROOF. We verify condition (2.3) of Proposition 2.1. Since the decision problem under consideration is quadratically regular, (2.3) will hold if

$$\inf_{g \in U(C)} \int \left\| Q(\cdot|x) - Q_g(\cdot|x) \right\|^2 M_g(dx) = 0 \tag{3.5}$$

for each ν-proper C. However, for each $g \in U(C)$, Corollary A.1 in Appendix 1 yields

$$\int \left\| Q(\cdot|x) - Q_g(\cdot|x) \right\|^2 M_g(dx) \le 2\,\Delta\left(\sqrt{g}\right), \tag{3.6}$$

where Δ is defined in (3.3). Setting $h = \sqrt{g}$, when (3.4) holds, (3.5) holds and thus (2.3) holds. This completes the proof. □

COROLLARY 3.1. *Let* $\{C_n | n = 1, 2, \dots\}$ *be any sequence of* ν*-proper sets satisfying* $C_n \subseteq C_{n+1}$ *and* $\bigcup C_n = \Theta$. *If* (3.4) *holds for each* C_n, *then the conclusion of Theorem* 3.1 *holds.*

PROOF. Use Corollary 2.1. □

REMARK 3.1. The converse of Corollary 3.1 is valid. If (3.4) holds for all C_n, then (3.4) holds for all ν-proper C.

EXAMPLE 3.1 (One-dimensional translation). This example concerns the additive group $R^1 = \Theta$ and a one-dimensional translation family model when the improper prior distribution is Lebesgue measure $d\theta$. Under very mild conditions, we show (3.4) holds, so a-ν-a obtains.

For ease of exposition, we write the model in the invariant (Pitman) form. Suppose the sample space is $R^1 \times \mathscr{Y}$ and the model $P(dx, dy|\theta)$ has a density $f(x-\theta, y)$ with respect to $dx\,\lambda(dy)$. Obviously the marginal density of $Y \in \mathscr{Y}$ with respect to λ is

$$m(y) = \int_{-\infty}^{\infty} f(u, y)\, du. \tag{3.7}$$

Then the marginal measure on the sample space is

$$M(dx, dy) = m(y)\, dx\, \lambda(dy) \tag{3.8}$$

which is clearly σ finite. Define $q(\theta|x, y)$ by

$$q(\theta|x,y) = \begin{cases} f(x-\theta, y)/m(y), & \text{if } 0 < m(y) < +\infty, \\ q_0(\theta), & \text{otherwise,} \end{cases} \tag{3.9}$$

where q_0 is some fixed density on R^1. Then $q(\cdot|x, y)$ serves as a version of the conditional density of θ given (x, y); that is,

$$Q(d\theta|x,y) = q(\theta|x,y)\, d\theta. \tag{3.10}$$

For $v \in R^1$, consider

$$t(v) = \iint q(v|x,y) f(x,y)\, dx\, \lambda(dy) \tag{3.11}$$

and note that

$$t(v) = t(-v), \qquad \int t(v)\, dv = 1. \tag{3.12}$$

A routine calculation shows that the measure $T(d\theta, d\eta)$ in (3.2) is given by

$$T(d\theta, d\eta) = t(\theta - \eta)\, d\theta\, d\eta. \tag{3.13}$$

THEOREM 3.2. *Assume*

$$\int |u| t(u)\, du < +\infty. \tag{3.14}$$

Then (3.4) *holds.*

PROOF. It suffices to verify (3.4) for $C_m = [-m, m]$ where $m = 1, 2, \dots$. Define h_n, $n = 1, 2, \dots$, by

$$h_n(\theta) = a_n h(\theta/n),$$

where

$$h(\theta) = \frac{1}{1+\theta^2}, \qquad a_n = 1 + \frac{m^2}{n^2}.$$

Since $h_n \in V(C_m)$, it suffices to show that

$$\lim_{n\to\infty} \Delta(h_n) = 0. \tag{3.15}$$

Now, use the symmetry of t and a change of variables to obtain

$$\Delta(h_n) = 2a_n^2 \iint \frac{h(\eta + w/n) - h(\eta)}{w/n} h(\eta) w t(w)\, d\eta\, dw.$$

Since h has a bounded derivative and (3.14) holds, the dominated convergence theorem yields

$$\lim_{n\to\infty} \Delta(h_n) = 2\iint h'(\eta) h(\eta) w t(w)\, d\eta\, dw,$$

which is zero. This completes the proof. □

A sufficient condition for (3.14) to hold, expressed directly in terms of the model follows.

THEOREM 3.3. *If*

$$\mathscr{E}_0|X| = \iint |x| f(x,y)\, dx\, \lambda(dy) < +\infty, \tag{3.16}$$

then (3.14) *holds*.

PROOF. We sketch the proof. Given $Y = y$, let W and $\tilde{W}$ be i.i.d. with the distribution of X given $Y = y$ when $\theta = 0$. Then it can be shown that

$$\int |u| t(u)\, du = \mathscr{E}\mathscr{E}\big[|W - \tilde{W}|\,|Y = y\big].$$

This is clearly bounded above by

$$2\mathscr{E}\mathscr{E}\big[|W|\,|Y = y\big] = 2\mathscr{E}_0|X|,$$

and the proof is complete. □

These results apply directly to the case of a random sample $X_1, \dots, X_n$ from a one-dimensional translation family. Let $X_{(1)} \le \cdots \le X_{(n)}$ be the order statistic of the sample and take X to be $X_{(r)}$. A good choice for $Y \in R^{n-1}$ is

$$Y_i = \begin{cases} X_{(i)} - X_{(r)}, & i = 1, \dots, r-1, \\ X_{(i+1)} - X_{(r)}, & i = r, \dots, n-1. \end{cases}$$

If $\mathscr{E}_0|X_{(r)}| < +\infty$, then Theorem 3.3 applies directly and the use of Lebesgue measure as an improper prior produces a-ν-a decision rules in quadratically regular problems. It should be noted that the condition $\mathscr{E}_0|X_{(r)}| < +\infty$ is weaker than the condition $\mathscr{E}_0|X_1| < +\infty$ when $1 < r < n$. For example, consider $n \ge 3$ when $X_1, \dots, X_n$ is a random sample from a Cauchy translation family, and take $r = [(n+1)/2]$.

Naturally, the preceding results can be used to provide conditions for a-ν-a when the model is a one-dimensional scale parameter model and the improper prior is $d\theta/\theta$ on $(0,\infty)$. This is because both the formulation of the original problem and the condition (3.4) are invariant under one-to-one bimeasurable transformations of the parameter space. This ends Example 3.1.

REMARK 3.2. For any decision problem with risk function $R(\delta, \theta)$, if there exists a δ_1 such that

$$\int R(\delta_1, \theta)\nu(d\theta) < +\infty, \tag{3.17}$$

then a formal Bayes rule for ν is easily shown to be a-ν-a. In Example 2.1, take $n = 1$, $B = 1$ and suppose $\varphi(\theta, x) \in R^1$ satisfies

$$\int \varphi^2(\theta, x)P(dx|\theta)\nu(d\theta) < +\infty. \tag{3.18}$$

It is then obvious that (3.17) holds and hence

$$\hat{\varphi}(x) = \int \varphi(\theta, x)Q(d\theta|x)$$

is a-ν-a whether or not (3.4) holds. Thus, for Example 2.1, the real import of Theorem 3.1 is for functions which do not satisfy (3.18).

REMARK 3.3. Assume that (3.4) holds for a particular improper prior ν. Consider another improper prior ν_1 given by

$$\nu_1(d\theta) = \Psi(\theta)\nu(d\theta),$$

where Ψ is uniformly bounded away from zero and infinity; that is, there are constants c_1 and c_2 such that

$$0 < c_1 \le \Psi(\theta) \le c_2 < +\infty, \quad \text{all } \theta. \tag{3.19}$$

Let $\Delta(h)$ be given by (3.3) when the improper prior is ν and let $\Delta_1(h)$ be given by (3.3) when the improper prior is ν_1. It is not hard to show that

$$\Delta_1(h) \le \frac{c_2^2}{c_1}\Delta(h) \tag{3.20}$$

for $h \in V(C)$. Thus, when (3.4) holds for ν, it holds for ν_1. For example, in the translation problem of Example 3.1, we obtain a-ν-a for any prior $\Psi(\theta)\,d\theta$ as long as Ψ satisfies (3.19) and (3.4) holds for Lebesgue measure.

4. The Markov chain connection. The condition for a-ν-a given in Theorem 3.1 involves the behavior of the transition function $R(d\theta|\eta)$ defined by (3.1). This condition is

$$\inf_{h \in V(C)} \Delta(h) = 0 \quad \text{for each } \nu\text{-proper set } C. \tag{4.1}$$

Typically, the inf in (4.1) is not achieved by a function in $L_2(\nu)$, but the inf can be approximated in the following manner. Fix a ν-proper set C and let K be ν proper with $K \supseteq C$. With $V(C, K)$ given by (1.12), let

$$\delta_K = \inf_{h \in V(C, K)} \iint (h(\theta) - h(\eta))^2 R(d\theta|\eta)\nu(d\eta). \tag{4.2}$$

The main results in Appendix 2 provide both a formula for δ_K and a characterization of a function in $V(C, K)$ which achieves the inf in (4.2). A statement of these results is conveniently given in the language of Markov chains.

The transition function $R(\cdot|\eta)$ defines a Markov chain

$$W = (\eta, W_1, W_2, \dots)$$

on the infinite product space Θ^∞ [see Neveu (1964), Chapter 5]. The initial state of the chain is $W_0 = \eta$ and successive states, say W_{i+1}, are generated from the probability measure $R(\cdot|W_i)$, $i = 0, 1, \dots$. The probability measure of W on Θ^∞ is denoted by

$$S(\cdot|W_0 = \eta), \tag{4.3}$$

where W_0 is the initial state of the chain. Observe that the chain W is ν *symmetric*; that is, the measure

$$T(d\theta, d\eta) = R(d\theta|\eta)\nu(d\theta)$$

introduced in Section 3 is a symmetric measure on $\Theta \times \Theta$. This property underlies all of the results in Appendix 2.

To describe a minimizer in (4.2) introduce two stopping times:

$$\tau = \begin{cases} \text{first } n \geq 0, & \text{such that } W_n \in C \cup K^c, \\ +\infty, & \text{if no } n \text{ exists,} \end{cases}$$
$$\sigma = \begin{cases} \text{first } n \geq 1, & \text{such that } W_n \in C \cup K^c, \\ +\infty, & \text{if no } n \geq 1 \text{ exists,} \end{cases} \tag{4.4}$$

and let $B_\tau = \{\tau < +\infty\}$, $B_\sigma = \{\sigma < +\infty\}$. Now, start the chain at $W_0 = \eta$ and let $h_0(\eta)$ be the probability that the stopped chain W_τ is in C (and it stops). In symbols,

$$h_0(\eta) = S\{(W_\tau \in C) \cap B_\tau | W_0 = \eta\}. \tag{4.5}$$

Since h_0 is 1 on C, 0 on K^c and is bounded by 1 on $C^c \cap K$, it is in $V(C, K)$.

THEOREM 4.1. *The function h_0 in* (4.5) *achieves the* inf *in* (4.2). *Furthermore,*

$$\delta_K = \int_C [1 - P\{(W_\sigma \in C) \cap B_\sigma | W_0 = \eta\}]\nu(d\eta). \tag{4.6}$$

PROOF. See Theorem A.1 in Appendix 2. □

Theorem 4.1 contains important qualitative information concerning the form of functions which are "approximate" minimizers of $\Delta(h)$. However, even in the simplest examples, the explicit calculation of h_0 seems hopeless, but the "rough method" described in the next section does provide some hope for finding reasonable approximations to h_0 in nontrivial examples. The next result shows that the approximate minimization problem (involving K) actually converges to the minimization problem of interest when K increases to Θ.

Given a ν-proper set C, define a stopping time σ_C by

$$\sigma_C = \begin{cases} \text{first } n \geq 1, & \text{such that } W_n \in C, \\ +\infty, & \text{if no } n \geq 1 \text{ exists}. \end{cases}$$

Let K_m be an increasing sequence of ν-proper sets such that $C \subseteq K_1$ and $K_m \to \Theta$.

THEOREM 4.2. *The following equalities hold*:

$$\text{(4.7)} \quad \begin{aligned} &\text{(i)} \qquad \lim_{m\to\infty} \delta_{K_m} = \inf_{h\in V(C)} \Delta(h), \\ &\text{(ii)} \qquad \inf_{h\in V(C)} \Delta(h) = \int_C [1 - P\{\sigma_C < +\infty | W_0 = \eta\}] \nu(d\eta). \end{aligned}$$

PROOF. These are proved in Appendix 2. □

Now, we interpret (4.7)(ii) when the condition (4.1) for a-ν-a holds; that is, when

$$\text{(4.8)} \quad \int_C [1 - P\{\sigma_C < +\infty | W_0 = \eta\}] \nu(d\eta) = 0 \quad \text{for each } \nu\text{-proper set } C.$$

Given the definition of local ν recurrence in Appendix 2, we have:

THEOREM 4.3. *The condition* (4.1) *for a-ν-a holds if and only if the symmetric chain W is locally ν recurrent.*

EXAMPLE 4.1. Take $\mathscr{X} = \Theta = \{0, 1, 2, \dots\}$ and let c denote counting measures on $\mathscr{X}$. Consider the model with density (with respect to c)

$$f(x|\theta) = \begin{cases} p(\theta), & \text{if } x = \theta, \\ 1 - p(\theta), & \text{if } x = \theta + 1, \\ 0, & \text{otherwise}, \end{cases}$$

where $0 < p(\theta) < 1$ for all $\theta \in \Theta$. Let the prior distribution on Θ be $\nu(d\theta) = \pi(\theta)c(d\theta)$ with $\pi(\theta) > 0$.

Setting

$$m(x) = \int f(x|\theta)\nu(d\theta)$$

and calculating the transition function

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

we find that the transition density r is given by

$$r(0|0) = \frac{p^2(0)}{m(0)} + \frac{(1 - p(0))^2}{m(1)},$$

$$r(1|0) = \frac{p(1)(1 - p(0))}{m(1)}$$

and for $\eta \geq 1$,

$$r(\eta - 1|\eta) = \frac{(1 - p(\eta - 1))p(\eta)}{m(\eta)},$$

$$r(\eta|\eta) = \frac{p^2(\eta)}{m(\eta)} + \frac{(1 - p(\eta))^2}{m(\eta + 1)},$$

$$r(\eta + 1|\eta) = \frac{p(\eta + 1)(1 - p(\eta))}{m(\eta + 1)}.$$

For other values of θ, $r(\theta|\eta) = 0$. Thus the chain is an irreducible random walk so recurrence and local ν recurrence are equivalent. Applying the well-known condition for recurrence in a random walk [see Karlin and Taylor (1975), page 108] we find that (4.1) holds if and only if

$$\sum_0^\infty \frac{1}{\pi(\theta)p(\theta)(1 - p(\theta))} = +\infty. \tag{4.9}$$

In particular, if the $p(\theta)$ are uniformly bounded away from 0 and 1, (4.8) holds if and only if the sum of the $\pi^{-1}(\theta)$'s diverges. This supports the well-known admonition that one should not use improper priors which "put too much mass on remote portions of the parameter space." However, given *any* sequence $\pi(\theta) > 1$ with $\pi(\theta) \to \infty$ as $\theta \to \infty$, the model with $p(\theta) = \pi^{-1}(\theta)$ satisfies (4.9). Thus conditions implying (4.1) will necessarily involve both the improper prior and the model.

The connection with Markov chains has direct implications for the use of Haar measure as an improper prior distribution when the parameter space in question is a group. For example, suppose $\mathscr{X} = \Theta = R^p$ and the model is

$$P(dx|\theta) = f(x - \theta)\,dx.$$

Thus, we have one observation X from a translation family on R^p. Taking the improper prior to be the translation invariant measure on R^p, namely $\nu(d\theta) = d\theta$, a routine calculation shows that the transition function is

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

where

$$r(v) = r(-v) = \int f(x - v) f(x)\,dx.$$

In this case, the Markov chain $\{W_n|n = 0, 1, \dots\}$ is just a classical random walk on R^p. That is, let $U_1, U_2, \dots$ be i.i.d. with density r. Given $W_0 = \eta$, the chain is equal in distribution to

$$\left\{\tilde{W}_n = \eta + \sum_1^n U_i \middle| n = 0, 1, 2, \dots\right\}.$$

Example 3.1 shows that when $p = 1$ and

$$\int |v| r(v)\,dv < +\infty,$$

then the Markov chain (random walk) on R^1 defined by $R(d\theta|\eta)$ is recurrent (recurrence and almost ν recurrence are equivalent). When $p = 2$, it is known that if

$$\int \|v\|^2 r(v)\, dv < +\infty,$$

then the random walk on R^2 is recurrent. Hence the improper prior $d\theta$ on R^2 produces a-ν-a decision rules for quadratically regular problems. But, for $p \geq 3$, there are no nontrivial recurrent random walks on the group R^p, and thus (4.1) must fail to hold [see Guivarc'h, Keane and Roynette (1977)]. Other invariant problems are discussed briefly in Section 7.

REMARK 4.1. If X is a single observation from a Cauchy translation family on R^1, then condition (3.14) does not hold (with $d\theta$ as the prior), so the results of Theorem 3.1 do not apply. However, it is known that the Cauchy random walk on R^1 is recurrent and thus (4.1) holds. Hence the formal Bayes rules in this example are a-ν-a for quadratically regular problems.

REMARK 4.2. Using the Markov chain results and the recurrence of random walks on R^2 when the transition density has second moments, it is easy to extend the results of Example 3.1 from R^1 to R^2. The details are left to the reader.

5. A heuristic method. The results of Section 4 characterize the minimizer of $\Delta(h)$ over the class $V(C, K)$. Typically one can calculate neither the minimizer nor the minimum explicitly. The method presented here, for the special case that $\Theta = [0, \infty)$, consists of:

(5.1)
- (i) trying to bound $\Delta(h)$ by a constant times $\rho(h)$ [which is defined in (5.4)];
- (ii) obtaining an explicit minimizer of $\rho(h)$ over a subclass of $V(C, K)$ for nice sets C and K;
- (iii) using (ii) to derive verifiable conditions that drive $\rho(h)$ [and we hope $\Delta(h)$] to zero.

Until further notice, $\Theta = [0, \infty)$, $C = [0, a]$ and $K = [0, b]$ with $b > a > 0$. Assume h is differentiable and write

$$(h(\theta) - h(\eta))^2 = (h'(\xi))^2(\theta - \eta)^2, \tag{5.2}$$

where ξ is between θ and η. Next, replace $(h'(\xi))^2$ by what one hopes is an upper bound, namely

$$D\left[(h'(\theta))^2 + (h'(\eta))^2\right], \tag{5.3}$$

where D is a constant (not depending on b). Then, set

$$\rho(h) = \iint \left[(h'(\theta))^2 + (h'(\eta))^2\right](\theta - \eta)^2 R(d\theta|\eta)\nu(d\eta). \tag{5.4}$$

The symmetry of the measure $R(d\theta|\eta)\nu(d\eta)$ yields

$$\rho(h) = 2\int (h'(\eta))^2 \sigma(\eta)\nu(d\eta), \tag{5.5}$$

where

$$\sigma(\eta) = \int (\theta - \eta)^2 R(d\theta|\eta). \tag{5.6}$$

Now, assume $\nu(d\eta) = \pi(\eta)\,d\eta$ and define h_b as

$$h_b(\theta) = \begin{cases} 1, & \text{if } \theta \in [0, a], \\ 0, & \text{if } \theta \in [b, \infty), \\ \dfrac{\int_\theta^b [\sigma(\eta)\pi(\eta)]^{-1}\,d\eta}{\int_a^b [\sigma(\eta)\pi(\eta)]^{-1}\,d\eta}, & \theta \in (a, b). \end{cases} \tag{5.7}$$

Of course, it is assumed that for sufficiently large a,

$$0 < \int_\theta^b [\sigma(\eta)\pi(\eta)]^{-1}\,d\eta < +\infty$$

for all $\theta \in (a, b)$ and all $b > a$. This choice of h_b is prompted by the fact that h_b minimizes $\rho(h)$ over those h's in $V(C, K)$ which are a.e. differentiable and satisfy $h(a) = 1$, $h(b) = 0$. Further,

$$\rho(h_b) = \frac{2}{\int_a^b [\sigma(\eta)\pi(\eta)]^{-1}\,d\eta}. \tag{5.8}$$

The preceding discussion yields the following.

THEOREM 5.1. *With h_b defined by* (5.7) *assume that*:

(i) $\Delta(h_b) \le D\rho(h_b)$ *for all sufficiently large b where D is a fixed constant.*
(ii) $\lim_{b\to\infty} \int_a^b [\sigma(\eta)\pi(\eta)]^{-1}\,d\eta = \infty$.

Then the condition (4.1) *for* a-ν-a *holds.*

PROOF. Obvious from (5.8). □

EXAMPLE 5.1. Take $\mathscr{X} = \{0, 1, 2, \dots\}$, $\Theta = [0, \infty)$ and suppose X is Poisson with parameter θ. Consider priors of the form $\nu(d\theta) = \theta^\alpha\,d\theta$, where α is a parameter. In order that the marginal measure $M(dx)$ be σ finite it is necessary and sufficient that $\alpha \in (-1, \infty)$, which we assume. The transition function $R(d\theta|\eta)$ is

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

where

$$r(\theta|\eta) = \exp[-\theta - \eta] \sum_{j=0}^{\infty} \frac{(\theta\eta)^j}{j!\,\Gamma(j + \alpha + 1)}. \tag{5.9}$$

From this, we calculate that

$$\sigma(\eta) = \int (\theta - \eta)^2 R(d\theta|\eta) = 2\eta + (\alpha + 1)(\alpha + 2). \tag{5.10}$$

Thus condition (ii) of Theorem 5.1 holds for $\alpha \in (-1, 0]$, but not for $\alpha > 0$. Condition (i) holds with $D = 1$, but is more difficult to verify. However, the argument is little more than calculus and the fact that for $\alpha \in (-1, 0]$, $\theta^\alpha[\theta + (\alpha + 1)(\alpha + 2)]$ is increasing on $[a, \infty)$ for a large enough. The details are omitted. Thus for the Poisson, the argument shows that for $\alpha \in (-1, 0]$, the improper prior $\theta^\alpha \, d\theta$ yields a-ν-a decision rules for quadratically regular problems. This range for α coincides with that in Johnstone (1984), who considered admissibility of formal Bayes estimators of θ.

REMARK 5.1. Conditions for the recurrence of Markov chains on $[0, \infty)$ were discussed in Lamperti (1960). His conditions involved

$$\mu(\eta) = \int_0^\infty (\theta - \eta) R(d\theta|\eta)$$

and

$$\sigma(\eta) = \int_0^\infty (\theta - \eta)^2 R(d\theta|\eta).$$

Lamperti showed that if

$$\mu(\eta) \le \frac{\sigma(\eta)}{2\eta} + O(\eta^{-1-\delta}) \tag{5.11}$$

for some $\delta > 0$, then the chain generated by R is recurrent. For the Poisson example, $\mu(\eta) = \alpha + 1$ when the prior is $\theta^\alpha \, d\theta$, and $\sigma(\eta)$ is given in (5.10). Thus for $\alpha \in (-1, 0]$, (5.11) holds.

Conditions resembling (ii) in Theorem 5.1 have appeared elsewhere in the decision theoretic literature, typically in papers dealing with estimation of unbounded functions when the loss is quadratic [see Karlin (1958) and Brown and Hwang (1982)]. However, the explicit use of $\sigma(\eta)$ in this condition appears to be new. Two-sided versions of the condition when $\Theta = R^1$ also appear in some of these works.

Stein (1965) also indicated that some multidimensional problems might be amenable to arguments similar to those above. This we illustrate with a simple example when $\Theta = R^p$ (such as is the case for p-dimensional translation problems). Given a model $P(dx|\theta)$, consider a prior of the form

$$\nu(d\theta) = \xi(du)\pi(t) \, dt,$$

where $\theta = tu$ with $t \ge 0$ and u a unit vector in R^p, so $\|\theta\| = t$. Here, ξ is assumed to be a probability measure on unit vectors, so the "improper part" of the prior ν is $\pi(t) \, dt$ on $[0, \infty)$. (It is possible to let ξ depend on t in what follows, but we eschew that generalization.) Define the new probability model

$$\tilde{P}(dx|t) = \int P(dx|tu)\xi(du)$$

with parameter space $[0, \infty)$. It is a routine argument to show that if (4.1) holds for the model $\tilde{P}(dx|t)$ and prior $\pi(t)\,dt$, then (4.1) holds for the model $P(dx|\theta)$ and prior $\nu(d\theta)$. Of course, Theorem 5.1 may apply to the $\tilde{P}$-π problem.

An alternative approach to multidimensional problems is the following obvious extension of the heuristic method described for $[0, \infty)$. Just replace (5.2) to (5.5) with the obvious multidimensional versions to obtain

$$\tilde{\rho}(h) = 2\int \|\nabla h(\eta)\|^2 \sigma(\eta)\nu(d\eta), \tag{5.12}$$

where ∇h denotes the gradient vector and

$$\sigma(\eta) = \int \|\theta - \eta\|^2 R(d\theta|\eta). \tag{5.13}$$

Next attempt to minimize $\tilde{\rho}(h)$ over a suitable class of h's (for nice sets C and K). Expressions similar to (5.12) have arisen elsewhere; see Brown (1971) and Srinivasan (1981), for example.

It should be mentioned that both (1.7) and the sufficient condition (1.11) are invariant under one-to-one bimeasurable transformations of Θ, as is the condition for local ν recurrence of the Markov chain given in Section 4. However, the heuristics proposed above are not invariant under such transformations, so it becomes relevant to ask for a "good" coordinate system in which to try the heuristics. The following remark is based on comments from a referee/Associate Editor.

REMARK 5.2. First observe that condition (ii) in Theorem 5.1 is not affected by our $\sigma(\eta)$ if there exists an $\varepsilon > 0$ such that

$$\varepsilon \le \sigma(\eta) \le 1/\varepsilon \quad \text{for all } \eta. \tag{5.14}$$

Basically, condition (5.14) is that $\sigma(\eta)$ is essentially a constant, at least for the problem at hand. When (5.14) holds, the convergence or divergence of the integral in (5.8) is thus determined by the prior π. In cases where (5.14) does not hold, the transformation

$$\xi(\theta) = \int_0^\theta \frac{1}{\sigma(u)}\,du$$

yields a new parametrization (namely ξ) where (5.14) seems to hold. This is based on a rather imprecise heuristic and would have to be checked in particular examples. This transformation is related to material in Brown (1979).

REMARK 5.3. In some cases, the assumptions in Theorem 5.1 hold for h_b defined in (5.7), even though (5.3) is not an upper bound on $[h_b'(\xi)]^2$. Thus, the method of this section is to use h_b as an "approximate" minimizer of $\Delta(h)$, and then try to verify the assumptions of Theorem 5.1.

REMARK 5.4. Brown (1979) considers expressions like (5.12) in situations much more general than the normal location model treated in Brown (1971) and Srinivasan (1981).

6. The prediction problem. Here we formulate the prediction problem as a fair Bayes decision problem, much in the manner of Example 2.2. For loss functions of the type used in Example 2.2, it is shown that the corresponding decision problem is quadratically regular. Thus, Theorem 3.1 provides a sufficient condition that formal predictive distributions are a-ν-a decision rules.

The prediction problem consists of data $X \in \mathscr{X}$, a variable to be predicted $Z \in \mathscr{Z}$ and an unknown parameter $\theta \in \Theta$, which indexes the probability model describing the joint distribution of X and Z. The spaces $\mathscr{X}$, $\mathscr{Z}$ and Θ are assumed to be Polish and the σ algebras are those generated by the open sets. The probability model is written

$$P(dx|z,\theta)S(dz|\theta), \tag{6.1}$$

where $P(\cdot|z,\theta)$ is the conditional distribution of X given z and θ, and $S(\cdot|\theta)$ is the conditional distribution of Z given θ. The marginal distribution of X given θ is then

$$P(dx|\theta) = \int_{\mathscr{Z}} P(dx|z,\theta)S(dz|\theta). \tag{6.2}$$

Our formulation of the prediction problem resembles that in Aitchison and Dunsmore (1975). After seeing the data $X = x$, one wants to specify a distribution for Z. Thus a "decision" consists of a distribution $\delta(\cdot|x)$ defined on the Borel sets of $\mathscr{Z}$. In a decision theoretic framework, this means that the appropriate action space for the prediction problem is the set of all probability measures on the Borel sets of $\mathscr{Z}$, say $\mathscr{M}_1(\mathscr{Z})$. [The σ algebra for $\mathscr{M}_1(\mathscr{Z})$ is that generated by the weak topology on $\mathscr{M}(\mathscr{Z})$; see Eaton (1982) for some discussion.]

Let ν be an improper σ-finite prior distribution on Θ. Then

$$\tilde{\nu}(dz,d\theta) = S(dz|\theta)\nu(d\theta) \tag{6.3}$$

is σ finite on $\mathscr{Z} \times \Theta$. The marginal measure on $\mathscr{Z}$,

$$M(dx) = \int P(dx|z,\theta)\tilde{\nu}(dz,d\theta) \tag{6.4}$$

is assumed to be σ finite. Thus the formal posterior distribution of (Z,θ) given X, say $\tilde{Q}(dz,d\theta|x)$, exists. The marginal probability measures on $\mathscr{Z}$ and Θ obtained from $\tilde{Q}(\cdot|x)$ are denoted by $Q^*(dz|x)$ and $Q(d\theta|x)$, respectively.

With U as defined in Section 2, each $g \in U$ induces a finite measure

$$\tilde{\nu}_g(dz,d\theta) = g(\theta)S(dz|\theta)\nu(d\theta) \tag{6.5}$$

on $\mathscr{Z} \times \Theta$. In turn, this induces a conditional distribution $\tilde{Q}_g(dz, d\theta|x)$ and two marginals $Q_g^*(dz|x)$ and $Q_g(d\theta|x)$. Of course $Q_g^*(dz|x)$ is the *predictive distribution* of Z given x obtained from $\tilde{\nu}_g$. The corresponding $Q^*(dz|x)$ obtained from $\tilde{\nu}$ is the *formal predictive distribution* of Z given x.

To introduce the loss structure for the prediction problem, let $k(z_1, z_2, x)$ be a bounded real valued function defined on $\mathscr{Z} \times \mathscr{Z} \times \mathscr{X}$. As in Example 2.2, consider the bilinear function defined on bounded signed measure on $\mathscr{Z}$:

$$\langle \xi_1, \xi_2 \rangle = \iint k(z_1, z_2, x) \xi_1(dz_1) \xi_2(dz_2). \tag{6.6}$$

Assume that $\langle \cdot , \cdot \rangle$ is symmetric and nonnegative definite. Given the model (6.1), let $H(\cdot|x, \theta)$ denote the conditional distribution of Z given x and θ. Next, define the loss function by

$$L(a, \theta, x) = \langle a - H(\cdot|x, \theta), a - H(\cdot|x, \theta) \rangle \tag{6.7}$$

for $a \in \mathscr{M}_1(\mathscr{Z})$. If x and θ were known, $H(\cdot|x, \theta)$ is the predictive distribution one would use for Z. Thus, the loss function (6.7) is a measure of squared distance between the action taken and the "best" action. Note that $H(\cdot|x, \theta)$ is just $S(\cdot|\theta)$ when X and Z are conditionally independent given θ. However, in situations such as time series analysis, $H(\cdot|x, \theta)$ is not $S(\cdot|\theta)$.

Now, consider $g \in U$ so $\tilde{\nu}_g$ is a finite prior measure. With the loss function (6.7), it is easy to show that a Bayes solution to the decision problem is just the predictive distribution $Q_g^*(dz|x)$. This is the fair Bayes property discussed in Eaton (1982).

For any nonrandomized decision rule $\delta(\cdot|x) \in \mathscr{M}_1(\mathscr{Z})$, let $R(\delta, \theta)$ denote the risk function of δ. Note that Q^* and Q_g^* are decision functions.

PROPOSITION 6.1. *The above decision problem is quadratically regular; that is, for each $g \in U$,*

$$\begin{aligned} &\int \left[R(Q^*, \theta) - R(Q_g^*, \theta) \right] g(\theta) \nu(d\theta) \\ &\qquad \le K \int \left\| Q(\cdot|x) - Q_g(\cdot|x) \right\|^2 M_g(dx), \end{aligned} \tag{6.8}$$

where $\|\cdot\|$ denotes variation distance and K is an upper bound on k.

PROOF. Minor variations on arguments in Eaton (1982) suffice. The details are omitted. □

The marginal distribution of X given θ in (6.2) together with the improper prior distribution ν define the quadratic form Δ used in Theorem 3.1. Proposition 6.1 shows that Theorem 3.1 applies to the predictive problem. In summary we have:

THEOREM 6.1. *For each ν-proper set C, assume that* (3.4) *holds. Then the formal predictive distribution $Q^*(\cdot|x)$ is an* a-ν-a *decision rule when the loss function is* (6.7).

7. Discussion. It is somewhat surprising that the connection between admissibility conditions and recurrence criteria is as complete as described in Section 4, particularly given the technical issues which arose in Brown (1971) and Johnstone (1984, 1986) enroute to establishing an admissibility-recurrence connection in the normal and Poisson cases. The relationship between these two approaches is very far from clear, especially since in our approach the natural space for the Markov chain is the parameter space, while in Brown and Johnstone, the associated process is constructed on the sample space. For a discussion of related issues including an admissibility-boundary value problem tie, see Srinivasan (1981) and Johnstone (1986). Of course, the types of problems are different for at least two reasons. First, the results here give sufficient conditions for a-ν-a for formal Bayes rules in quadratically regular problems, while other authors have concentrated on the estimation (quadratic loss) of a fixed "natural" parametric functions (typically unbounded). Second, the sufficient conditions for admissibility in Brown (1971), Srinivasan (1981), Johnstone (1986) and others appear to be fairly close to necessary, while the necessity question is wide open here. Presumably, one natural way to phrase this necessity question is:

(7.1) Suppose (1.11) does not hold. Can one find a quadratically regular problem for which the formal Bayes rule is not a-ν-a?

REMARK 7.1. Even in the "simplest" cases, question (7.1) is interesting. For example, suppose X is $N_p(\theta, I_p)$ and $\nu(d\theta)$ is Lebesgue measure on R^p. Then (4.1) holds for $p = 1, 2$ but (4.1) fails for $p \geq 3$. This corresponds exactly to the Stein shrinkage phenomenon. However, consider the problem of estimating a bounded function of θ, say $\varphi(\theta)$, which has compact support. When the loss is quadratic, Remark 3.2 implies that the formal Bayes estimator of φ is a-ν-a no matter what p is. On the other hand, for $p \geq 3$, I have been able to construct a prediction problem (as described in Section 6) so that the formal Bayes predictive distribution is not admissible.

Now, suppose the problem is to estimate

$$\varphi(\theta, x) = \begin{cases} 1, & \text{if } \|x - a\| \leq r, \\ 0, & \text{otherwise,} \end{cases}$$

with quadratic loss. For $1 \leq p \leq 4$, the formal Bayes estimator of ϕ is admissible, but for $p \geq 5$, this estimator is not admissible [see Hwang and Brown (1991) for a discussion and references]. Thus, the failure of (4.1) cannot correspond exactly to inadmissibility even in interesting problems where Remark 3.2 does not apply.

In the context of Example 4.1, assume $p(\theta) = 1/2$ for all θ. Then the corresponding Markov chain is recurrent if and only if

$$\sum (\pi(\theta))^{-1} = +\infty.$$

When the sum is finite, one suspects there should be a quadratically regular decision problem with an inadmissible formal Bayes rule, thus providing a complete connection between admissibility and recurrence. My attempts to do this problem have failed thus far.

Given what is now known, it seems plausible that the earlier admissibility-reference connections are related to the behavior of a Markov chain on the sample space whose transition function is

$$\tilde{R}(dx|y) = \int P(dx|\theta)Q(d\theta|y). \tag{7.2}$$

This chain is M symmetric where M is the marginal measure on $\mathscr{X}$ induced by ν.

A connection between admissibility and a Markov process on the parameter space is suggested in Brown (1979, 1988). In this work, Brown explores a relationship between differential inequalities (on the sample space) and admissibility. A "dual" differential inequality on the parameter space then results, which in turn suggests looking at a Markov process on the parameter space, but the process is not introduced explicitly. Concerning the admissibility-recurrence connection, the similarity in structure between our Theorems A.1 and A.2 and Theorem 4.3.1 in Brown (1971) should be noted. Our results in Appendix 2 can be viewed as Markov chain analogs of Brown's Markov process results.

The results established in this paper do not bode well for the use of relatively invariant prior distributions when the parameter space is a noncompact Lie group, except in special circumstances. Consider a model $P(dx|\theta)$ where the parameter space Θ is a group G [e.g., R^p; GTp (group of $p \times p$ lower triangular matrices with positive diagonal elements); the affine group generated by GTp and R^p], and assume the model is invariant under G [we are using the terminology and notation in Eaton (1989)]. Take ν to be any relatively invariant prior distribution on G. It is fairly easy to show that the induced transition function $R(d\theta|\eta)$ on G corresponds to a random walk on G. For example, the case $G = R^p$ was discussed in Section 4. For many groups G of interest in statistics (e.g., R^p, $p \geq 3$; GTp, $p \geq 2$; the affine group generated by GTp and R^p, $p \geq 1$), the results in Guivarc'h, Keane and Roynette (1977) show that there are no nontrivial recurrent random walks on G. Hence for these cases, (1.11) must fail and the corresponding formal posterior becomes less attractive. In invariant problems when the parameter space is a homogeneous space (rather than a group), the situation concerning random walks is less clear [see Varoupolis (1988) and Schott (1984, 1986)].

Much more work needs to be done to understand the implications of the failure of (1.11). One interesting question is:

(7.3) Suppose (1.11) fails. Is there information in the Markov chain which tells one how to modify the prior (or estimators) to produce a better posterior?

For example, if X is $N(\theta, I_p)$ with $p \geq 3$ and $\nu(d\theta) = d\theta$ on R^p, the induced transition function $R(d\theta|\eta)$ corresponds to a $N(\eta, 2I_p)$ distribution. Can one use the transience of the Markov chain to construct "improved" posterior distributions?

The criterion adopted here for the evaluation of ν is (1.7). A more stringent requirement would be to ask that the formal posterior produce a-ν-a procedures for a much wider variety of decision problems than those in (1.7). The example due to Blackwell (1951) shows some care must be taken. The following example, related to Blackwell's shows that even in simple problems, the formal Bayes method may yield uniformly inadmissible estimators.

EXAMPLE 7.1. With $\mathscr{X} = \Theta = A =$ additive group of integers, suppose X given θ takes on the values θ or $\theta + 1$ each with chance $1/2$. Using the flat prior on Θ, the formal posterior puts mass $1/2$ at x and $x - 1$. It is easy to construct a bounded loss function L on $A \times \Theta$ with the following properties:

(i) $L(a, a) = 0$ for all a.
(ii) $L(\theta - 1, \theta) < L(\theta + 1, \theta)$ for all θ.
(iii) $L(\theta + 1, \theta) \leq L(a, \theta + 1) + L(a, \theta)$ for all a, with strict inequality for $a \neq \theta + 1$.

Consider the two estimators $t_0(x) = x$ and $t_1(x) = x - 1$. Using the above L, the formal Bayes method gives t_0 as the unique formal Bayes estimator, but $R(t_1, \theta) < R(t_0, \theta)$ for all θ. Of course, this problem is not quadratically regular.

Finally, it is natural to ask if the methods developed here can be adapted to give alternative proofs of standard results; for example, the exponential family results in Brown and Hwang (1982). The inequality (2.10) (quadratic regularity) is the key to our development. Some rough calculations indicate that for the problem of estimating an unbounded function of θ (quadratic loss), (2.10) will typically not hold. However, one can weaken the quadratic regularity assumptions and still use the methods here. For example, let $\{C_n | n = 1, 2, \dots\}$ be a sequence of ν-proper sets satisfying the assumptions of Corollary 2.1. For $g \in U(C_n)$, assume (2.10) holds where the constant K is allowed to depend on n. A minor modification of the arguments in Section 3 show that if (3.4) holds for each C_n, then the formal Bayes rule is a-ν-a. With some effort, I have been able to use these ideas to prove that the usual estimator of the mean in a univariate normal problem (known variance) is a-ν-a (quadratic loss). Even so, it is not clear how to adapt the results here to problems involving the estimation of unbounded parametric function—a standard statistical activity. However, in the prediction problem of Section 7 and the posterior distribution problem of Example 2.2, the boundedness of the kernel defining the loss function does not seem like such a bothersome assumption.

APPENDIX 1

A proof of inequality (3.6) follows. First, for probability measures α_1, α_2 with Radon–Nikodym derivatives $p_i = d\alpha_i / d\lambda$, apply the Cauchy–Schwarz inequality to obtain a bound on variation distance:

$$\text{(A.1)}\qquad \begin{aligned} \|\alpha_1 - \alpha_2\|^2 &= \left(\int |p_1 - p_2|\right)^2 \le \int\left(\sqrt{p_1} - \sqrt{p_2}\right)^2 \int\left(\sqrt{p_1} + \sqrt{p_2}\right)^2 \\ &= 4\left[1 - \left(\int \sqrt{p_1 p_2}\right)^2\right]. \end{aligned}$$

An alternative bound is given in Kraft (1955).

In the notation of Sections 2 and 3, let g be a density with respect to ν so $\int_g d\nu = 1$. It is easy to show M_g is absolutely continuous with respect to M. With

$$\text{(A.2)}\qquad m_g(x) = \frac{dM_g}{dM}(x),$$

the quantity we need to bound is

$$\text{(A.3)}\qquad \delta(g) = \int \|Q(\cdot|x) - Q_g(\cdot|x)\|^2 m_g(x) M(dx).$$

PROPOSITION A.1. *For each density* g,

$$\text{(A.4)}\qquad \delta(g) \le 2\Delta\left(\sqrt{g}\right).$$

PROOF. Obviously, the set

$$A_0 = \left\{x \middle| 0 < m_g(x) < +\infty\right\}$$

satisfies $M_g(A_0^c) = 0$. The equations

$$P(dx|\theta)g(\theta)\nu(d\theta) = g(\theta)Q(d\theta|x)M(dx) = Q_g(d\theta|x)m_g(x)M(dx)$$

imply that, except for a set of M_g measure zero,

$$\text{(A.5)}\qquad k(x,\theta) = \begin{cases} \dfrac{g(\theta)}{m_g(x)}, & \text{if } x \in A_0, \\ 1, & \text{if } x \notin A_0, \end{cases}$$

serves as a version of the Radon–Nikodym derivative of $Q_g(\cdot|x)$ with respect to $Q(\cdot|x)$. Now apply (A.1) with $\lambda = Q(\cdot|x)$ to get

$$\text{(A.6)}\qquad \|Q(\cdot|x) - Q_g(\cdot|x)\|^2 \le 4\left\{1 - \left[\int (k(x,\theta))^{1/2} Q(d\theta|x)\right]^2\right\}.$$

Integrating (A.6) with respect to M_g gives

$$\text{(A.7)}\qquad \begin{aligned} \frac{\delta}{4} &\le 1 - \int\left[\int \sqrt{g(\theta)}\, Q(d\theta|x)\right]^2 M(dx) \\ &= 1 - \iint (g(\theta)g(\eta))^{1/2} T(d\theta, d\eta) \\ &= \frac{1}{2}\Delta\left(\sqrt{g}\right). \end{aligned}$$

The last equality is a consequence of

$$\iint g(\theta)T(d\theta, d\eta) = \int g(\theta)\nu(d\theta) = 1.$$

The proof is complete. □

COROLLARY A.1. *For any nonnegative g which satisfies*

$$0 < \int g(\theta)\nu(d\theta) < +\infty,$$

inequality (A.4) *holds*.

PROOF. For any $a > 0$, $\delta(ag) = a\delta(g)$ and $\Delta(\sqrt{ag}) = a\,\Delta(\sqrt{g})$. □

APPENDIX 2

On symmetric Markov chains. In this Appendix, we establish a Dirichlet principle for symmetric Markov chains which provides proofs for the assertions in Section 4. Let $(\mathscr{W}, \mathscr{B})$ be a Polish space and let $R(\cdot|w)$ be a transition function on $\mathscr{B} \times \mathscr{W}$. The discrete time Markov chain on $(\mathscr{W}^\infty, \mathscr{B}^\infty)$ defined by $R(\cdot|w)$ with initial state w is denoted by $W = (w, W_1, W_2, \dots)$. The induced probability measure for W is $S(\cdot|W_0 = w)$, where W_0 denotes the initial state of the chain.

DEFINITION A.1. Let ν be a nonzero σ-finite measure on $(\mathscr{W}, \mathscr{B})$. The Markov chain is ν *symmetric* if the measure

$$T(dw_1, dw_2) = R(dw_1|w_2)\nu(dw_2) \tag{A.8}$$

is a symmetric measure on $(\mathscr{W} \times \mathscr{W}, \mathscr{B} \times \mathscr{B})$.

For a discussion of symmetric chains in the countable state space, see Kelly (1979), Griffeath and Liggett (1982) and Lyons (1983). The discussion in Section 4 provides many examples of ν-symmetric Markov chains. In all that follows, W is assumed to be a ν-symmetric chain.

The following definition, a modified notion of recurrence, allows us to circumvent a discussion of irreducibility issues while relating our previous admissibility results to the recurrence of W.

DEFINITION A.2. The chain W is *locally* ν *recurrent* (1-ν-r) if for each ν-proper set C, the set

$$\left[w \,\middle|\, S\{W_n \in C \text{ for some } n \ge 1 | W_0 = w\} < 1\right] \cap C$$

has ν measure zero.

In words, this means that for each ν proper C, given the chain starts in C, it returns to C w.p.1 except for a ν-null set. Of course, when $\mathscr{W}$ is countable and the chain is irreducible, 1-ν-r and recurrence are equivalent.

Let $L_2(\nu)$ denote the space of ν-square integrable functions. The symmetry of T implies that

$$\iint h^2(w_1)T(dw_1, dw_2) = \iint h^2(w_2)T(dw_1, dw_2) = \int h^2(w)\nu(dw)$$

for all $h \in L_2(\nu)$. Hence, the Cauchy–Schwarz inequality yields

$$\left|\iint h_1(w_1)h_2(w_2)T(dw_1, dw_2)\right|^2 \le \int h_1^2(w)\nu(dw)\int h_2^2(w)\nu(dw)$$

for $h_1, h_2 \in L_2(\nu)$. Thus the bilinear form $\langle \cdot , \cdot \rangle$ defined by

$$\langle h_1, h_2\rangle = \int h_1(w)h_2(w)\nu(dw) - \iint h_1(w_1)h_2(w_2)T(dw_1, dw_2) \tag{A.9}$$

is symmetric and nonnegative definite for $h_1, h_2 \in L_2(\nu)$. In most of what follows, $\langle \cdot , \cdot \rangle$ is written

$$\langle h_1, h_2\rangle = (h_1, (I - R)h_2), \tag{A.10}$$

where $(\cdot, \cdot)$ is the standard bilinear form on $L_2(\nu)$ given by

$$(h_1, h_2) = \int h_1(w)h_2(w)\nu(dw),$$

I is the identity transformation and Rh_2 is defined by

$$(Rh_2)(w) = \int h_2(w_1)R(dw_1|w). \tag{A.11}$$

The results in this Appendix relate l-ν-r of the chain to the behavior of the form $\langle \cdot , \cdot \rangle$. To this end, let C and K_0 be two ν-proper subsets of $\mathscr{W}$ such that $C \subseteq K_0$. Define the stopping times τ and σ and the sets B_τ and B_σ as in (4.4). Also let

$$V(C, K_0) = \{h \in L_2(\nu) | h(w) \ge 1 \text{ for } w \in C,\ h(w) = 0 \text{ for } w \in K_0^c\}$$

and observe that

$$h_0(w) = S\{(W_\tau \in C) \cap B_\tau | W_0 = w\} \tag{A.12}$$

is in $V(C, K_0)$. In fact, h_0 is 1 on C and is 0 on K_0^c.

THEOREM A.1. *For a ν-symmetric chain W,*

(i) $$\inf_{h \in V(C, K_0)} \langle h, h\rangle = \langle h_0, h_0\rangle$$

and

(ii) $$\langle h_0, h_0\rangle = \int_C [1 - S\{(W_\sigma \in C) \cap B_\sigma | W_0 = w\}]\nu(dw).$$

PROOF. For $h \in V(C, K_0)$, write

$$h = h_0 + \phi.$$

The symmetry and nonnegative definiteness of $\langle \cdot , \cdot \rangle$ yields

$$\langle h, h \rangle = \langle h_0, h_0 \rangle + 2\langle \phi, h_0 \rangle + \langle \phi, \phi \rangle \geq \langle h_0, h_0 \rangle + 2\langle \phi, h_0 \rangle.$$

With $Q = I - R$, (A.11) yields

$$\langle \phi, h_0 \rangle = \int \phi(w)(Qh_0)(w)\nu(dw)$$

$$= \int \phi(Qh_0) = \left(\int_C + \int_{K_0^c} + \int_{K_0 \cap C^c} \right) [\phi(Qh_0)].$$

The integral over K_0^c is zero because ϕ is zero on K_0^c. The integral over C is nonnegative because $\phi \geq 0$ and C and $(Qh_0)(w) = 1 - (Rh_0)(w) \geq 0$ for $w \in C$. Thus

$$\langle \phi, h_0 \rangle \geq \int_{K_0 \cap C^c} \phi(Qh_0).$$

However, a standard Markov chain argument shows that

$$(Qh_0)(w) = 0 \quad \text{for } w \in K_0 \cap C^c \tag{A.13}$$

(that is, h_0 is harmonic for $w \in K_0 \cap C^c$). Thus (i) is established. For assertion (ii), use (A.13) and the fact that $h_0 \in V(C, K_0)$ to obtain

$$\langle h_0, h_0 \rangle = \int h_0(Qh_0) = \int_C (Qh_0) = \int_C [1 - (Rh_0)(w)]\nu(dw).$$

Again a standard Markov chain argument yields

$$(Rh_0)(w) = S\{(W_\sigma \in C) \cap B_\sigma | W_0 = w\} \quad \text{for all } w. \tag{A.14}$$

This completes the proof. □

Again let C be a ν-proper set and let

$$V(C) = \{h \in L_2(\nu) | h \geq 0,\ h(w) \geq 1 \text{ for } w \in C\},$$

$$\sigma_C = \begin{cases} \text{first } n \geq 1, & \text{such that } W_n \in C, \\ +\infty, & \text{if } W_n \notin C \text{ for all } n \geq 1. \end{cases}$$

THEOREM A.2. *For ν-symmetric chains,*

$$\inf_{h \in V(C)} \langle h, h \rangle = \int_C [1 - P\{\sigma_C < +\infty | W_0 w\}]\nu(dw). \tag{A.15}$$

PROOF. Let $\{K_m | m = 1, 2, \dots\}$ be a sequence of ν-proper sets with $C \subseteq K_1$, $K_m \subseteq K_{m+1}$ and $\mathscr{W} = \bigcup_1^\infty K_m$. With $K_0 = K_m$ in Theorem A.1,

$$\begin{aligned} \inf_{h \in V(C, K_m)} \langle h, h \rangle &= \langle h_m, h_m \rangle \\ &= \int_C \left[1 - S\left\{(W_{\sigma_m} \in C) \cap B_{\sigma_m} | W_0 = w\right\}\right]\nu(dw), \end{aligned} \tag{A.16}$$

where h_m and σ_m are the K_m counterparts of h_0 and σ defined for K_0. Our first task is to show that

$$\text{(A.17)} \qquad \lim_{m \to \infty} \langle h_m, h_m \rangle = \int_C [1 - P\{\sigma_C < +\infty | W_0 = w\}] \nu(dw).$$

To this end, let

$$E_m = \{(W_{\sigma_m} \in C)\} \cap B_{\sigma_m}$$

and let

$$E = \{\sigma_C < +\infty\}.$$

Clearly $E_m \subseteq E_{m+1}$. Further, it is not hard to show that

$$E_m \to E.$$

From this and (A.16), (A.17) follows from the dominated convergence theorem. Thus, the left side of (A.15) is bounded above by the right side of (A.15) because $V(C) \supseteq V(C, K_m)$ for all m.

Now, let $h \in V(C)$ and set

$$u_m = hI_{K_m} \in V(C, K_m).$$

Applying the dominated convergence theorem, Theorem A.1 and (A.17) in that order yields

$$\begin{aligned} \langle h, h \rangle &= \lim_{m \to \infty} \langle u_m, u_m \rangle \geq \lim_{m \to \infty} \langle h_m, h_m \rangle \\ &= \int_C [1 - P\{\sigma_C < +\infty | W_0 = w\}] \nu(dw). \end{aligned}$$

Thus (A.15) holds. □

THEOREM A.3. *The chain W is* 1-ν-r *if and only if for each ν-proper set C,*

$$\text{(A.18)} \qquad \inf_{h \in V(C)} \langle h, h \rangle = 0.$$

PROOF. This is immediate from Theorem A.2. □

It is not necessary to verify (A.18) for all ν-proper C to show W is 1-ν-r. Remark 3.1 shows that (A.18) need only be verified for some increasing sequence of ν-proper sets C_n, $n = 1, 2, \dots,$ such that $C_n \to \Theta$.

Acknowledgments. Thanks to my colleagues and friends at Minnesota, Brad Johnson, David Lane, Ashok Maitra and Bill Sudderth for comments and suggestions regarding various aspects of this work. In 1988, lectures on this material were given in Holland where conversations with Kobus Oosterhof, Frank den Hollander and Michael Keane were most helpful. Special thanks to Iain Johnstone for providing the essence of the proof of my original conjecture on the admissibility–Markov chain connection. The constructive remarks of the referees and an Associate Editor of this paper are here gratefully acknowledged.

REFERENCES

AITCHISON, J. and DUNSMORE, I. R. (1975). *Statistical Prediction Analysis*. Cambridge Univ. Press.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BERGER, J. O. and SRINIVASAN, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.* **6** 783–801.

BERNARDO, J. (1979). Expected information as expected utility. *Ann Statist.* **7** 686–690.

BLACKWELL, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Statist.* **22** 105–110.

BLYTH, C. R. (1951). On minimax statistical procedures and their admissibility. *Ann. Math. Statist.* **22** 22–24.

BROWN, L. D. (1971). Admissible estimators, recurrent diffusioins, and insolvable boundary value problems. *Ann. Math. Statist.* **42** 855–904.

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

BROWN, L. D. (1988). The differential inequality of a statistical estimation problem. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 299–324. Academic, New York.

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. Academic, New York.

CLEVENSON, M. and ŽIDEK, J. (1977). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc.* **72** 698–705.

EATON, M. L. (1982). A method for evaluating improper prior distributions. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **1** 329–352. Academic, New York.

EATON, M. L. (1989). *Group Invariance Applications in Statistics. Regional Conference Series in Probability and Statistics* **1**. IMS, Hayward, Calif.

GEISSER, S. (1980). A predictivistic primer. In *Bayesian Analysis in Econometrics and Statistics* (A. Zellner, ed.) 363–381. North-Holland, Amsterdam.

GRIFFEATH, D. and LIGGETT, T. (1982). Critical phenomena for Spitzer's reversible nearest particle systems. *Ann. Probab.* **10** 881–895.

GUIVARC'H, Y., KEANE, M. and ROYNETTE, B. (1977). *Marches aleatoires sur les Groupes de Lie. Lecture Notes in Math*. **624**. Springer, New York.

HARTIGAN, J. (1983). *Bayes Theory*. Springer, New York.

HEATH, D. and SUDDERTH, W. (1989). Coherent inference from improper priors and from finitely additive priors. *Ann. Statist.* **17** 907–919.

HENDRICKSON, A. D. and BUEHLER, R. J. (1971). Proper scores for probability forecasters. *Ann. Math. Statist.* **42** 1916–1921.

HWANG, J. and BROWN, L. (1991). Estimated confidence under the validity constraint. *Ann. Statist.* **19** 1964–1977.

HWANG, J. T., CASELLA, G., ROBERT, C., WELLS, M. and FARRELL, R. (1992). Estimation of accuracy in testing. *Ann. Statist.* **20** 490–509.

JOHNSTONE, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Ann. Statist.* **12** 1173–1198.

JOHNSTONE, I. (1986). Admissibile estimation, Dirichlet principles, and recurrence of birth–death chains on Z_{+}^{p}. *Probab. Theory Related Fields* **71** 231–269.

KARLIN, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.* **29** 406–436.

KARLIN, S. and TAYLOR, H. (1975). *A First Course in Stochastic Processes*. Academic, New York.

KELLY, F. P. (1979). *Reversibility and Stochastic Networks*. Wiley, New York.

KRAFT, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2** 125–141.

LAMPERTI, J. (1960). Criteria for the recurrence or transience of stochastic processes, I. *J. Math. Anal. Appl.* **1** 314–330.

LYONS, T. (1983). A simple criterion for transience of a reversible Markov chain. *Ann. Probab.* **11** 393–402.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day, San Francisco.

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge Univ. Press.

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

SAVAGE, L. J. (1971). The elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801.

SCHOTT, R. (1984). Random walks on homogeneous spaces. *Probability Measures on Groups VII. Lecture Notes in Math.* **1064** 564–575. Springer, New York.

SCHOTT, R. (1986). Recurrent random walks on homogeneous spaces. *Probability Measures on Groups VIII. Lecture Notes in Math.* **1210** 146–152. Springer, New York.

SRINIVASAN, C. (1981). Admissible generalized Bayes estimators and exterior boundary value problems. *Sankhyā Ser. A.* **43** 1–25.

STEIN, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Statist.* **26** 518–522.

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

STEIN, C. (1965a). Approximation of improper prior measures by prior probability measures. In *Bernoulli-Bayes-Laplace Festschrift* (J. Neyman and L. Le Cam, eds.) 217–240. Springer, New York.

STEIN, C. (1965b). Lecture notes on decision theory. Unpublished manuscript.

STONE, M. (1965). Right Haar measure for convergence in probability to quasi-posterior distributions. *Ann Math. Statist.* **36** 440–453.

VAROPOULOS, N. TH. (1988). Green's function and harmonic functions on manifolds. In *Geometry of Random Motion* (R. Durrett and M. Pinsky, eds.) 333–337. *Contemp. Math.* **73**. Amer. Math. Soc., Providence, R.I.

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

SCHOOL OF STATISTICS
270 VINCENT HALL
UNIVERSITY OF MINNESOTA
206 CHURCH STREET S.E.
MINNEAPOLIS, MINNESOTA 55455