Probab. Th. Rel. Fields 71, 231–269 (1986)

Probability Theory and Related Fields



# Admissible Estimation, Dirichlet Principles and Recurrence of Birth-Death Chains on $\mathbb{Z}_+^p$

Iain Johnstone*

Department of Statistics, Stanford University*, Stanford, California 94305, USA

**Summary.** We establish a connection between admissible simultaneous estimation and recurrence of reversible Markov chains on $\mathbb{Z}_+^p$. Specifically, to each generalized Bayes estimator of the mean of a vector of $p$ independent Poisson variables for a weighted quadratic loss, we associate a variational problem and a reversible birth and death chain on $\mathbb{Z}_+^p$. The variational problem is closely related to the Dirichlet principle for reversible chains studied recently by Griffeath, Liggett and Lyons. Under side conditions, admissibility of the estimator is equivalent to zero infimal energy in the variational problem and to recurrence of the Markov chain. This yields analytic and probabilistic criteria for inadmissibility which are applied to establish a broad class of results and previous conjectures.

## § 1. Introduction

Consider the idealized statistical problem of estimating unknown means $\lambda_i$ of each of $p$ Poisson populations based on a single observation $X_i$, $i=1,\dots,p$ taken independently from each population. This paper develops a connection between this problem and the recurrence properties of certain reversible Markov chains defined on the sample space $\mathbb{Z}_+^p$ of the observations $\{X_i\}$ and applies analytic and probabilistic methods to discuss optimality of statistical estimators.

Our formulation of the statistics problem is decision theoretic: an estimator $d(x)=(d_1(x),\dots,d_p(x))$ incurs a loss $L_{-1}(d(x),\lambda)=\sum_{i=1}^{p}\lambda_i^{-1}(d_i(x)-\lambda_i)^2$ if $\lambda=(\lambda_1,\dots,\lambda_p)>0$ is the (unknown) value of the parameters and $x=(x_1,\dots,x_p)$. The estimator $d(x)$ is evaluated by studying its risk function $R(\lambda,d)$

* Research supported by an Australian National University Scholarship and A.D. White Fellowship at Cornell University and by NSF at Mathematical Sciences Research Institute, Berkeley and at Stanford

$=E_\lambda L_{-1}(d(X),\lambda)$, where the expectation is taken assuming independent Poisson $(\lambda_i)$ distributions for each $X_i$. A (weak) optimality property of an estimator is that there not exist another estimator $d'(x)$ with $R(\lambda, d') \leqq R(\lambda, d)$ for all $\lambda > 0$ and strict inequality for at least some $\lambda$: such estimators $d$ are termed *admissible.* In principle, one would not use an inadmissible estimator, since a uniformly better rule exists. The aim of this work is to provide probabilistic descriptions of the class of admissible estimators and explicit criteria for determining (in)admissibility in the Poisson problem.

C. Stein's (1956b) celebrated discovery that sample averages are inadmissible for estimating $p \geqq 3$ normal means under quadratic loss established admissibility as a significant qualitative concept. Subsequently James and Stein (1960), Efron and Morris (1973) and others showed that very substantial savings in risk over sample averages were attainable by "shrinkage" estimators having intuitive interpretations. The situation is the same for simultaneous estimation of parameters of independent distributions with (infinite) *discrete* sample spaces: the typical and simplest example being the Poisson problem introduced above. Thus Peng (1975) and Clevenson and Zidek (1975) showed that the simplest (and maximum likelihood estimator) $d(x)=(x_1, \ldots, x_p)$ was inadmissible in dimensions $p \geqq 3$ and $p \geqq 2$ for losses $L_0$ (discussed later) and $L_{-1}$ respectively. Much recent frequentist work in developing improved estimators and measuring the resultant savings in risk for the Poisson and other discrete problems is surveyed by Ghosh et al. (1983). Morris (1983) catalogues some significant practical applications of shrinkage methods, including a number based on discrete data. Berger (1985) gives a comprehensive survey of shrinkage theory from both frequentist and Bayesian perspectives.

The program of the paper is as follows: estimators that are Bayes relative to a finite prior distribution on $\lambda$ are typically admissible. The converse is almost true: admissible rules can be described in terms of generalized Bayes rules - rules obtained from a possibly infinite prior measure on $\lambda$. Roughly put, the search for admissible rules may thus be confined to the class of generalized Bayes procedures. A reversible Markov chain $\{X_t\}$ on $\mathbb{Z}_p^+$ is then associated with each generalized Bayes procedure, $d_P$, say. The main step is to recast the question of admissibility of $d_P$ as a variational problem familar in the probabilistic potential theory associated with the chain $\{X_t\}$. Finally one shows that admissibility of $d_P$ is equivalent to recurrence of $\{X_t\}$.

This program was formulated and executed in the multivariate normal estimation problem by Brown (1971, 1973), building in part on some heuristic ideas of Stein (1965). His most striking example is the association of Brownian motion with the maximum likelihood estimator $d(x)=x$ and the identification of Stein's inadmissibility phenomenon with the transience of Brownian motion in $p \geqq 3$ dimensions. These remarkable results were complemented by Brown (1979a), Srinivasan (1981) and Johnstone and Lalley (1984).

That such a theory could be developed for estimation of a single Poisson mean was detailed in Johnstone (1984, referred to as *I*). The one-dimensional case has the virtue of technical simplicity. However, several essential features arise only in the multivariate setting that forms the focus of this work. These include (i) Stein's phenomenon: the inadmissibility of natural estimators in combined problems, (ii) the clearer roles of probabilistic elements such as reversibility and potential theory, including general tests for recurrence and transience, (iii) the shift in analytic methods from those of ordinary differential equations to those of elliptic partial d.e.'s, notably maximum principles and (simple) apriori estimates. At the level of statement and discussion of results, this paper may be read independently of *I*. For details in proofs, we occasionally refer to *I* to avoid duplication and to indicate which one dimensional arguments do (and do not) generalize.

**Outline of Results.** We begin with the reduction to generalized Bayes estimators. That admissible estimators of the natural parameter of an exponential family in $\mathbb{R}^p$ must be generalized Bayes goes back to Sacks (1963), Brown (1971) and Berger and Srinivasan (1978). The corresponding result for estimation of the parameters of independent power series distributions such as the Poisson on $\mathbb{Z}_+^p$ is less clear cut (Brown and Farrell, 1983a). In order to avoid complications inherent in their general theory, we consider only estimators satisfying

$$d_i(x)=0 \Leftrightarrow x_i=0. \tag{1.1}$$

This assumption covers most estimators of practical interest. Section 5.7 discusses some interesting cases in which (1.1) fails. To state the complete class theorem, we employ multi-index notation: Set $\lambda^x=\prod_1^p \lambda_i^{x_i}$, $x!=\prod_1^p x_i!$, and $\Lambda=\sum_1^p \lambda_i$ so that the joint Poisson density may be written as $P_\lambda(X=x)=p_\lambda(x)$ $=e^{-\Lambda}\lambda^x/x!$. If $P(d\lambda)$ is a measure on $\mathbb{R}_+^p=[0,\infty)^p$, define transforms

$$p_x=\int e^{-\Lambda}\lambda^x P(d\lambda), \qquad \pi_x=p_x/x!=\int p_\lambda(x)\,P(d\lambda).$$

$\pi_x$ is the marginal density of the prior $P$. Let $e_i=(0,\ldots,1,0,\ldots,0)$.

**1.1. Proposition.** (*Brown, Farrell*). *Suppose that $d(x)$ is admissible and satisfies* (1.1). *Then there exists a prior $P(d\lambda)$ on $[0,\infty)^p$ such that $d(x)$ is generalized Bayes for $P$: $p_x<\infty$ for $x\in\mathbb{Z}_+^p$, and*

$$d_i(x)=p_x/p_{x-e_i} \quad \text{if } x_i>0.$$

To characterize admissible rules, therefore, we need only investigate estimators having the above quotient representation. A direct proof of this result is outlined in Sect. 2.

We turn now to the association of a variational problem with a generalized Bayes estimator. Stein (1955) and LeCam (1955) have shown for general statistical decision problems that a rule $d(x)$ is admissible if and only if it can be approximated arbitrarily closely (in the sense of Bayes regret) by rules which are Bayes with respect to a proper (i.e. finite) prior distribution. This result (described in Sect. 2) connects an abstract minimization problem with each estimator $d(x)$. In Sect. 2, we exploit the quadratic loss structure, the exponential family form of the Poisson densities and the quotient representation above to derive from the Stein-LeCam result a much more concrete minimization problem familiar in potential theory. Specifically, let $d_P$ be the generalized Bayes estimator whose admissibility is in question. With $d_P$ associate coefficients $a_{i,x}=d_{P,i}(x)\pi(x)$, for $x$ such that $x_i>0$ and $i=1,\ldots,p$. Here $\pi(x)=\pi_x$ is, as before, the marginal density of the prior $P$. Let $\mathscr{U}=\{u: \mathbb{Z}_+^p\to\mathbb{R}: u_0=1, u_x\to 0$ as $|x|\to\infty\}$ be the class of feasible solutions, and write $D_i u_x$ for the backward difference $u(x)-u(x-e_i)$. Here and throughout, it is convenient to define $|x| =\sum_i^p |x_i|$ for points $x\in\mathbb{Z}^p$. For simplicity assume also that

$$\operatorname{supp} P=[0,\infty)^p. \tag{1.2}$$

(Remark 3.7 comments on relaxation of this condition.) The first main result is a necessary condition for admissibility.

**1.2. Theorem.** *If $d_P(x)$ satisfies* (1.1) *and* (1.2) *and is admissible, then*

$$\inf_{u\in\mathscr{U}}\sum_i \sum_{x:x_i\geqq 1}' (D_i u_x)^2 a_{i,x}=0. \tag{1.3}$$

As discussed in $I$ and references listed there, the double sum in (1.3) has a "physical" interpretation as the power dissipated by a system of voltages $u_x$ at sites $x\in\mathbb{Z}_+^p$, when neighboring sites are connected by resistors parallel to the coordinate axes with conductances $a_{i,x}$ for the resistor connecting $x$ and $x-e_i$. For this reason, we shall, with slight abuse of terminology, call (1.3) an energy condition.

The major part of this work is devoted to establishing the converse and applications of this theorem under suitable side conditions. Consider first an important special case. The simplest results hold for those estimators $d_P(x) =\Phi(\sum x_i)x$ which are generalized Bayes for "simplex symmetric" priors of the form $P(d\lambda)=M(d\Lambda)\,d\theta_1\ldots d\theta_p$, where $\Lambda=\sum\lambda_i$ and $\theta_i=\lambda_i/\Lambda$. These priors are the analogues for the Poisson problem of spherically symmetric priors in the normal case: $M$ is uniform on each fixed multiple of the unit simplex. For $L_{-1}$ the analogy is surprisingly strong, in view of the lack of any natural large group leaving the sets $\{\lambda\geqq 0: \sum\lambda_i=\Lambda\}$ invariant. The resulting theory turns out to be essentially one dimensional, and Sect. 4 applies the univariate characterization of admissibility derived in $I$ to obtain sharp results in this simplest of multivariate contexts. Of course, the maximum likelihood estimator, and those considered by Clevenson and Zidek are included in this setting.

The converse to Theorem 1.2 in the general case is proved in Sect. 5 under the following assumption. Subject to the earlier caveat concerning (1.1), this condition holds, to my knowledge, for all estimators proposed for this problem. The significance of its components and their analogy with the Gaussian case is discussed in the heuristics part of Sect. 5. Here we note only that the counterexample in § 7 of $I$ shows that some form of growth condition on $d_i$ (or $d_{P,i}$) is needed. However, assumption (1.2), that $\operatorname{supp} P=[0,\infty)^p$, is not required for the converse.

**Assumption A.** *Suppose $d_{P,i}(x)$ is Lipschitz and that there exist increasing Lipschitz functions $d_i: \mathbb{Z}^+\to\mathbb{R}^+$* satisfying $d_i(x_i)\leqq x_i+Mx_i^{1/2}$ *such that*

a) $|d_{P,i}(x)-d_i(x_i)| \leqq M(x_i \vee 1)^{1/2}$,

b) $d_{P,i}(x)/d_i(1) \geqq \varepsilon > 0$ if $x_i = 1$.

**1.3. Theorem.** *Under* (1.1) *and assumption A, if the zero energy condition* (1.3) *holds, then $d_P$ is admissible.*

Methods of difference equations and probabilistic potential theory enable (1.3) to be translated into more perspicuous and/or checkable conditions. We shall, therefore, discuss these before turning to the applications of the main results. The first step is to set up the Euler-Lagrange difference equation belonging to (1.3).

Regard $\mathbb{Z}_+^p$ as a lattice with sites connected by bonds parallel to the coordinate axes. Denote by $\partial x$ the collection of the neighbors of $x$ which lie in $\mathbb{Z}_+^p$. Define connection coefficients for $x, y \in \mathbb{Z}_+^p$ by

$$\alpha_{x,y} = \begin{cases} a_{i,x+e_i} & \text{if } y = x + e_i \\ a_{i,x} & \text{if } y = x - e_i \in \mathbb{Z}_+^p \\ 0 & \text{for other } y \neq x \\ -\sum\limits_{z \neq x} \alpha_{x,z} & \text{if } y = x. \end{cases} \tag{1.4}$$

Assumption (1.1) implies for all $x \in \mathbb{Z}_+^p$, $y \in \partial x$ that $\alpha_{x,y}$ is strictly positive. Note that $\alpha_{x,y}$ is *symmetric*: $\alpha_{xy} = \alpha_{yx}$. Let $\|u\|^2$ stand for the double sum in (1.3); it clearly also has the more symmetric form

$$2\|u\|^2 = \sum_{x,y} \alpha_{xy}(u_y - u_x)^2, \tag{1.5}$$

where the sum need only be taken over those $y \in \partial x$.

Now suppose that $\inf_{\mathscr{U}} \|u\|^2$ is attained by some function $u$. If $v$ is an arbitrary function such that $u \pm \varepsilon v \in \mathscr{U}$ for small $\varepsilon$, then on letting $\varepsilon \to 0$ and using symmetry of $\alpha_{x,y}$, it follows that $\sum_x v_x L u_x = 0$, where

$$L u_x = \sum_{y \in \partial x} \alpha_{x,y}(u_y - u_x). \tag{1.6}$$

Thus $u$ satisfies the Euler-Lagrange equation $Lu = 0$ on $\mathbb{Z}_+^p \setminus \{0\}$.

As is well known, a (continuous time parameter) Markov process $\{X_t; t \geqq 0, P^x\}$ is associated with the difference operator $L$. The process $X_t$ has state space $\mathbb{Z}_+^p$ (the sample space in the estimation problem), and may jump from the point $x$ only to one of its neighbors $y \in \partial x$, these transitions occurring at rates $\alpha_{x,y}$. It may therefore be thought of as a multidimensional birth and death process. Thus, for $h \downarrow 0$, $x \neq 0$ and $y \in \mathbb{Z}_+^p$, $P^x(X_h = y) = \alpha_{x,y} h + o(h)$. Thus $\{X_t\}$ is as close to being a diffusion (such as occurs in the Gaussian estimation problem) as the state space permits. Since $\alpha_{x,y} > 0$ for all $y \in \partial x$, the process is irreducible. The transition rates are symmetric in $x$ and $y$, so the process is time reversible. Should the rates be such that an explosion occurs ($|X_t| \to \infty$ in finite time), then the process is banished to a coffin state thereafter. The existence of a right continuous, strong Markov process with these properties follows from Markov chain theory (cf. for example, Freedman, 1971, especially §§6.4, 6.5 and 7.4). It follows also for $x \in \mathbb{Z}_+^p \setminus \{0\}$ that $L$ agrees with the infinitesimal generator of $\{X_t\}$ when applied to functions belonging to the domain of the latter.

An alternative and perhaps conceptually more direct approach is to observe that the bilinear form associated with (1.5) is a regular Dirichlet form in the sense of Fukushima (1980). For this one takes as domain the set of all square summable (for counting measure on $\mathbb{Z}_+^p$) functions $u$ for which also $\|u\|^2$ is finite. The general potential theory of Fukushima could then be applied to construct a symmetric Hunt process $\{X_t\}$ having the properties described above. This is outlined for the Gaussian case in Johnstone and Lalley (1984).

The plan now is to recast the energy condition (1.3) in terms of the Euler Lagrange equation and the recurrence of $X_t$. Let the hitting probability function $\bar{u}_x = P^x(\exists t \geqq 0: X_t = 0)$. The process $\{X_t\}$ is recurrent if $\bar{u}_x \equiv 1$ and transient otherwise. Standard arguments show that if $\{X_t\}$ is transient, then $\bar{u}_x < 1$ for all $x$ and $\liminf \bar{u}_x = \lim_{r\to\infty} \inf_{x:|x| \geqq r} \bar{u}_x = 0$. Given Theorems 1.2 and 1.3, the next result provides the promised alternate characterizations of admissibility.

**1.4. Theorem.**

$$\min_{\mathscr{U}} \|u\|^2 = \|\bar{u}\|^2, \tag{1.7}$$

*and the minimum is attained iff $\bar{u} \in \mathscr{U}$. Consequently, the following are equivalent:*

(i) $\min_{\mathscr{U}} \|u\|^2 = 0$,

(ii) $\{X_t, P^x\}$ *is recurrent,*

(iii) *There is no bounded solution to the exterior boundary value problem:*

$$Lu = 0 \quad on \quad \mathbb{Z}_+^p \setminus \{0\}; \quad u_0 = 1; \quad \liminf_{|x| \to \infty} u_x = 0. \tag{$\mathscr{P}$}$$

*In the transient case, $\bar{u}$ is a solution to $\mathscr{P}$.*

Results of this genre are known in probabilistic potential theory (e.g. Griffeath-Liggett, 1982; Fukushima, 1980) and differential equations (in the continuous case). In view of the technical simplicity of the discrete setting, and relative completeness of the results, a self-contained proof is given in Sect. 3.

Let us turn now to some statistical applications of Theorems 1.2 through 1.4. Although some generality is lost, comparison tests provide very simple methods of checking admissibility since they are based on easily computed functions of the candidate rule $d(x)$. To illustrate, suppose $d_i(x) \geqq \eta x_i$ for some positive $\eta$ and that

$$\sum_i d_i(x) \geqq z - (p-1) + \delta \qquad \text{for large } z = \sum x_i, \tag{1.8}$$

for some $\delta > 0$: then $d$ is inadmissible (Corollary 6.3). Conversely, if $d$ is generalized Bayes, satisfies condition $A$ and

$$\sum_i d_i(x + e_i) \leqq z + 1 \qquad \text{for large } z, \tag{1.9}$$

then $d$ is admissible. These results are proved together with sharper comparison tests in Sect. 6, either directly from Theorems 1.2 and 1.3, or via 1.4 and the Nash-Williams and Royden-Lyons tests for recurrence and transience respectively (cf. Griffeath-Liggett, 1982; Lyons, 1983).

Here is an example of the use of comparison tests. Brown (1979) gives detailed heuristics to arrive at his conjectures that generalized Bayes estimators of the form $d(x)=x+\phi(x)$, where

$$\phi(x)=\frac{bx}{(\sum x_i)}+O((\sum x_i)^{-1/2}) \tag{1.10}$$

are inadmissible if $b>1-p$ and admissible for $b\leqq 1-p$ (cf. his 2.3.9, p. 984). Perturbations of the form (1.10) correspond to generalized priors of the form $P(d\lambda)\sim(\sum\lambda_i)^b d\lambda$ as $\sum\lambda_i\to\infty$. The conjectures follow immediately from (1.8) and (1.9) above (assuming only that the specific estimator $d(x)$ in question has, for small $x$, values compatible with (1.1) (or 5.27 and Assumption A).

Section 6 also discuss the connection of our results with the semi-tail upper bounds for inadmissibility of Hwang (1982) – another simple method for checking inadmissibility. It is further easy to read off an admissibility classification of linear estimates of the form $d(x)=Mx+\gamma$ for $M$ non-singular. Indeed for admissibility $M$ must be diagonal, with diagonal entries lying in (0, 1), and the sum of the $\gamma_i$ corresponding to the $J$ unit eigenvalues must be bounded by $1-J$.

The two closing comments of the introduction to $I$ extend to the multiparameter case considered here. Thus the results of Sects. 2 and 3 will likely extend to more general power series distributions, including negative binomial and logarithmic. Secondly, to recover the difference operator occurring in the unbiased-risk-estimate approach to inadmissiblity (cf. for example Ghosh et al., 1983), we need to take the Euler Lagrange equation of the 'original' minimization problem (second line of (2.3) below) rather than the simpler, linearized version that appears in the energy condition (1.3).

Only partial results on the extension of the theory to other loss functions such as $L_0$ are currently available (cf. Remarks 2.4, 4.4). The full force of Brown and Farrell's stepwise Bayes complete class theorem is needed: even natural admissible estimators (such as the MLE for $p=2$) correspond to several recurrent processes on disjoint subsets of $\mathbb{Z}^p_+$. In the case of simplex symmetric estimators, the reduction to a one-dimensional problem is less clean than for $L_{-1}$. It seems that the recurrence and variational theory is most natural for $L_{-1}$ because the component problems are equally balanced as $\lambda$ varies in $\mathbb{R}^p_+$: each component $x_i$ of the MLE has constant risk in $\lambda_i$.

## § 2. The Variational Condition for Admissibility

We begin with the Stein-Le Cam characterization of admissibility. Fix $\lambda_0\in(0,\infty)^p$, and let $\mathcal{Q}$ be the set of finite measures $Q$, supported on a finite set in $(0,\infty)^p$ with $Q(\{\lambda_0\})\geqq 1$. Let $R(d,\lambda)=E_\lambda L_{-1}(d(X),\lambda)$ denote the risk function of an estimator $d$ and write $B(d, Q) = \int R(d, \lambda)\, Q(d\lambda)$. It is easy to check that the estimator attaining $B(Q) = \inf_d B(d, Q)$ (the Bayes estimator) is given by $d_{Q,i}(x) = q_x / q_{x-e_i}$, where $q_x = \int e^{-\Lambda} \lambda^x Q(d\lambda)$. Then $d(x)$ is admissible if and only if

$$\inf_{Q \in \mathcal{Q}} B(d, Q) - B(Q) = 0. \tag{2.1}$$

Now the (standard) calculation given at $I$ (2.2) shows that for any prior $Q(d\lambda)$ with $B(d, Q) < \infty$,

$$B(d, Q) - B(Q) = \sum_{i=1}^{p} \sum_{x \geqq 0} [d_i(x) - d_{Q,i}(x)]^2 q_{x-e_i}/x!. \tag{2.2}$$

The previous two displays and the availability of quotient representations for (generalized) Bayes estimators are basic to this study. An elementary proof of the characterization (2.1) is given by Brown and Farrell (1985c). Given (2.2) however, the sufficiency half is sufficiently simple that it will be useful to give a proof here. Indeed, suppose $B(d, Q) - B(Q_n) \to 0$ and that $R(\lambda, d') \leqq R(\lambda, d)$ for all $\lambda$. Then

$$\begin{aligned}\sum_i \sum_x [d_i(x) - d_i'(x)]^2 q_{x-e_i}^{(n)}/x! &\leqq 2\{B(d, Q_n) - B(Q_n) + B(d', Q_n) - B(Q_n)\} \\ &\leqq 4\{B(d, Q_n) - B(Q_n)\} \to 0.\end{aligned}$$

Since $Q_n(\{\lambda_0\}) \geqq 1$, $q_x^{(n)} \geqq e^{-\Lambda_0} \lambda_0^x$, we have $d_i(x) = d_i'(x)$ and hence that $d$ is admissible.

It is clear from this argument that the $Q_n$ need not be discrete: any sequence such that $\inf_n Q_n(N) > 0$, where $N$ is a neighborhood of $\lambda_0$, will suffice. In § 5, we use $Q_n(d\lambda) = u_n^2(\lambda)\, P(d\lambda)$, with $u_n^2 \nearrow 1$.

*2.1. Remark.* Let $\mathcal{D}_0 = \{d: x_i = 0 \Rightarrow d_i(x) = 0\}$. To prove admissibility of rules in $\mathcal{D}_0$, it is convenient to modify (2.1) slightly. First note that if $d(x)$ is admissible in $\mathcal{D}_0$, then it is unconditionally admissible. This is proved by showing that $\lim_{\lambda_i \to 0} R(\lambda, d) < \infty$ for all $\lambda$ iff $d_i(x)$ vanishes on $\{x: x_i = 0\}$, which in turn follows by examining the terms in the Laurent series expansion of $R(\lambda, d)$ about 0.

Now to prove admissibility in $\mathcal{D}_0$, it is enough to establish (2.1) with $B(d_Q, Q)$ replaced by $B_{\mathcal{D}_0}(Q)$, the infimum of $B(d, Q)$ over rules in $\mathcal{D}_0$. But this infimum is attained by $\tilde{d}_Q$, where $\tilde{d}_{Q,i}(x) = d_{Q,i}(x)\, I\{x_i > 0\}$. Thus the $i^{\text{th}}$ sum over $x$ in (2.2) can be restricted to the set on which $x_i > 0$.

*2.2. Remark.* Let $d^{(1)}$ be admissible for $\lambda^{(1)}$ and $d^{(2)}$ be Bayes for a proper prior $H$ on $\lambda^{(2)}$ in independent problems (which may each involve more than one coordinate). It is a general consequence of (2.1) that $d = (d^{(1)}, d^{(2)})$ is admissible for $\lambda = (\lambda^{(1)}, \lambda^{(2)})$ if the loss functions are added. Indeed, if $Q_n^{(1)}$ is a sequence of priors provided by (2.1) from the admissibility of $d^{(1)}$, then $Q_n = Q_n^{(1)} \times H$ suffice to show admissibility of $d$.

*Proof of Proposition 1.1.* Since $d$ is admissible, it is a pointwise limit of Bayes procedures with respect to proper priors $Q_n$. This may be seen from (2.1) and

(2.2) using the condition $q_x^{(n)} \geqq e^{-\Lambda_0} \lambda_0^x$ as before. Defining $\xi_n(d\lambda) = e^{-\Lambda} dQ_n(\lambda) / \int e^{-\Lambda} dQ_n(\lambda)$, one obtains a probability measure $\xi(d\lambda)$ such that $\int \lambda^x \xi_n(d\lambda) \to \int \lambda^x \xi(d\lambda)$ for all $x \in \mathbb{Z}_+^p$. That the latter integral is positive and finite for all $x \geqq 0$ follows from (1.1). Defining $P(d\lambda) = e^{\Lambda} \xi(d\lambda)$ and using $d_{P_n,i}(x) \to d_i(x)$ for all $x$, we conclude that $d_i(x) = p_x / p_{x-e_i}$ whenever $x_i > 0$. Furthermore, if $x_i > 0$, it is easy to check that $E[L(d_i, \lambda_i) \mid x]$ is minimized by the choice $d_{P,i}(x) = p_x / p_{x-e_i} = d_i(x)$, so that $d_i(x)$ is generalized Bayes on $\{x : x_i > 0\}$.

*Proof of Theorem 1.2.* Let $\Delta_i$ denote the $i^{\text{th}}$ sum in (2.2). On $T_i = \{x \in \mathbb{Z}_+^p : x_i > 0\}$, $d_i(x) = p_x / p_{x-e_i}$ with both numerator and denominator finite and positive. We now proceed with the analogue of *I*. (3.1)–(3.4) but ignore the sets $T_i^c$. Below, $u_x^2 = q_x / p_x$.

$$\Delta_i \geqq \sum_{T_i} \left( \frac{p_x}{p_{x-e_i}} - \frac{q_x}{q_{x-e_i}} \right)^2 \frac{q_{x-e_i}}{x!} = \sum_{T_i} (u_x - u_{x-e_i})^2 (1 + u_x / u_{x-e_i})^2 a_{i,x}$$
$$\geqq \sum_{T_i} (D_i u_x)^2 a_{i,x}. \tag{2.3}$$

It follows that the infimum of $\|u\|^2$ over $\mathcal{U}_{\mathcal{Q}} = \{u : u_x^2 = q_x / p_x$ for some prior $Q \in \mathcal{Q}\}$ is zero. As in *I* 3, the condition $Q([\lambda_0, \lambda_0 + 1]) \geqq 1$ for $Q$ in $\mathcal{Q}$ allows us to replace the infimum over $\mathcal{U}_{\mathcal{Q}}$ by an infimum over $\mathcal{U}_{\mathcal{Q}} \cap \{u : u_0 \geqq 1\}$.

The proof is completed by showing that $\mathcal{U}_{\mathcal{Q}} \subset \{u : \lim_{|x| \to \infty} u_x = 0\}$. We state first an appropriate form of the Birnbaum-Stein theorem for exponential families (Birnbaum, 1955; Stein, 1956a). If $S(d\theta)$ is a measure on $\mathbb{R}^p$, let $K(S)$ denote the convex hull of its support. Suppose $S$ and $R$ are measures on $\mathbb{R}^p$ and that there exists a point $w \in [\operatorname{Int} K(R)] \setminus K(S)$. Let $U_w = \{y : y \cdot w \geqq \sup\{y \cdot \theta : \theta \in K(S)\}\}$. Then there exist constants $B$, $\varepsilon > 0$ depending on $w$, such that for $y \in U_w$,

$$\frac{\int e^{\theta \cdot y} S(d\theta)}{\int e^{\theta \cdot y} R(d\theta)} \leqq B e^{-\varepsilon |y|}.$$

In the Poisson case, $\operatorname{supp} P = (0, \infty)^p$ and $Q \in \mathcal{Q}$ has compact support, so we may apply the result to $S$ and $R$ defined by putting $\theta_i = \log \lambda_i (i = 1, \ldots, k)$ and

$$S(d\theta) = e^{-\Lambda} Q(d\lambda) \qquad R(d\theta) = e^{-\Lambda} P(d\lambda). \tag{2.4}$$

Now choose $w = \alpha \mathbf{1}$ with $\alpha = \sup\{\|\theta\| : \theta \in K(S)\}$ and $\mathbf{1} = (1, 1, \ldots, 1)$. Since $\operatorname{supp} P = (0, \infty)^p$, we can assume (by increasing $\alpha$ if necessary) that $w \in [\operatorname{int} K(R)] \setminus K(S)$. If $y \geqq 0$, then $y \cdot w = \alpha \sum y_i \geqq \alpha \|y\| \geqq y \cdot \theta$ for all $\theta \in K(S)$, where here $\|y\|$ denotes Euclidean distance. Hence $\mathbb{Z}_+^p \subset U_{\alpha \mathbf{1}}$ and so there exist constants $\varepsilon$ and $B$ positive such that for all $y \geqq 0$,

$$u_y^2 = \frac{\int e^{\theta \cdot y} S(d\theta)}{\int e^{\theta \cdot y} R(d\theta)} \geqq B e^{-\varepsilon |y|} \to 0$$

as $|y| \to \infty$. □

*2.3. Remark.* If $P(d\lambda) = M(d\Lambda) d\theta_1, \ldots, d\theta_k$ (where $\theta_i = \lambda_i / \Lambda$), as is the case in Sect. 4, then we need only assume that $M$ has unbounded support. Indeed, if $S$

has compact support, then there will always exist an $\alpha$ sufficiently large that $\alpha\mathbf{1}\in[\operatorname{int} K(R)]\setminus K(S)$.

*2.4. Remark.* Under squared error loss $L_0$ there is no convenient analog of Proposition 1.1, which for $L_{-1}$ simplifies the analysis by ensuring that a potentially admissible $d_i(x)$ has a representation $p_x/p_{x-e_i}$ whenever it is non-zero. The full force of the stepwise (generalized) Bayes representation of Brown and Farrell (1985 a) is needed. For simplicity, consider estimates satisfying (1.1) and suppose also that $p=2$. If $d$ is admissible, then it follows from Brown and Farrell's Theorem 4.1 that there exists *finite* measures $w_{10}$, $w_{01}$ and $w_{11}$ supported on $\mathbb{R}_+\times\{0\}$, $\{0\}\times\mathbb{R}_+$ and $\mathbb{R}_+^2$ respectively (here $\mathbb{R}_+=[0,\infty)$) for which

$$d_i(x)=\begin{cases} p_{\alpha(x),x+e_i}/p_{\alpha(x),x} & \text{if } x_i>0 \\ 0 & \text{if } x_i=0, \end{cases}$$

where $p_{\alpha,x}=\int\prod_i \lambda_i^{\alpha_i(x_i-1)}w_\alpha(d\lambda)$; $\alpha$ takes the values (1, 0), (0, 1) or (1, 1) and $\alpha(x)=(I(x_1>0), I(x_2>0))$.

The analogue of (2.2) breaks naturally into a sum over three regions $S_{10}$, $S_{01}$ and $S_{11}$, where $S_\alpha=\{x\in\mathbb{Z}_+^2: \alpha(x)=\alpha\}$. After going through the analogue of (2.3), the sum over region $S_\alpha$ may be bounded below by

$$\sum_{i:\alpha_i>0}\sum_{\substack{x\in S_\alpha\\ x_i>1}}(D_i u_x)^2\,\tilde{\alpha}_{i,x} \tag{2.5}$$

where $u_x^2=q_{\alpha(x),x}/p_{\alpha(x),x}$, and $\tilde{\alpha}_{i,x}=p_{\alpha(x),x}^2/p_{\alpha(x),x-e_i}(x-e_i)!$ has a different structure from that of the $\alpha_{i,x}$ occurring for $L_{-1}$. Thus, the sums over $S_{10}, S_{01}, S_{11}$ are one, one and two dimensional respectively, and each can be made arbitrarily small if $d_P$ is admissible. To anticipate the discussion of Sect. 3, in addition to the two dimensional birth and death process on $S_{11}$ that is naturally associated with the coefficients $\tilde{\alpha}_{i,x}$, there are two *further* one dimensional processes on the mutually disjoint state spaces $S_{10}$ and $S_{01}$. Admissibility then entails recurrence of all three processes.

## § 3. Difference Equations and Recurrence

This section begins with the proof of the probabilistic and analytic characterizations of the admissibility condition (1.3). Examples follow which illuminate the hypotheses of Theorem 1.4 and apply it to the MLE and priors with bounded support. Conditions for null and positive recurrence and their statistical interpretation are discussed. Finally we describe modifications needed to allow for $\operatorname{supp} P\subsetneqq(0,\infty)^p$, and some probabilistic facts for later use.

*Proof of Theorem 1.4.* (iii)$\Rightarrow$(ii). This follows from the Strong Markov property by a standard argument. Suppose that $\{X_t\}$ is transient: then $\bar{u}_0=1$ and

$\liminf_{|x|\to\infty} \bar{u}_x = 0$. If $\sigma = \inf\{t \geqq 0: X_t \neq X_0\}$ is the first jump time of a path, then for $x \neq 0$,

$$\bar{u}_x = E^x \bar{u}(X_\sigma) = \sum_{y \in \partial x} \alpha_{x,y} \bar{u}_y / \sum_{y \in \partial x} \alpha_{x,y},$$

which is equivalent to $L\bar{u}_x = 0$.

*Maximum principle for L.* A function $u$ on $\mathbb{Z}_+^p$ cannot attain a strict local maximum (or minimum) on $S = \{x: Lu_x = 0\}$. Indeed, if $u$ attains a local maximum at $x \in S$, then $u_x \geqq u_y$ for $y \in \partial x$, and

$$\sum_{y \in \partial x} \alpha_{x,y}(u_x - u_y) = 0,$$

so that $u_y = u_x$ for all $y \in \partial x$ (since all $\alpha_{x,y} > 0$ for $y \in \partial x$).

*Remark.* Simple examples show that the maximum principle as stated here does not hold for non self-adjoint 'elliptic' difference operators.

*Approximating Problems.* Let $B_n$ be an increasing sequence of finite subsets of $\mathbb{Z}_+^p$ with $\mathbb{Z}_+^p = \bigcup_n B_n$. Let

$$\langle u, v \rangle_n = \sum_{x, y \in B_n} \alpha_{x,y}(u_y - u_x)(v_y - v_x),$$

and $\|u\|_n^2 = \langle u, u \rangle_n$ be the corresponding seminorm. Write $\partial_n x = \partial x \cap B_n$ for the neighbors in $B_n$ of $x$, and set $\partial B_n = \{x: \partial_n x \neq \partial x\} \cup \{0\}$. Let $\mathcal{U}_n^\phi$ be the class of functions on $B_n$ that agree with a function $\phi$ defined on $\partial B_n$. Finally, let $\sigma_n = \inf\{t \geqq 0: X_t \in \partial B_n\}$.

**Proposition** (Dirichlet principle). *Let $\phi$ be a function defined on $\partial B_n$. The function $u_x^\phi = E^x \phi(X_{\sigma_n})$ is the unique solution to*

$$Lu = 0 \quad on \quad B_n \setminus \partial B_n, \qquad u = \phi \quad on \quad \partial B_n. \tag{$\mathcal{P}_n$}$$

*Further, $u^\phi$ is the unique function in $\mathcal{U}_n^\phi$ which minimizes $\|u\|_n^2$.*

*Proof.* It follows as before from the Strong Markov property that $u^\phi$ satisfies ($\mathcal{P}_n$), and uniqueness is clear from the maximum principle. A calculation using symmetry of $\alpha_{x,y}$ shows that

$$-\langle u, v \rangle_n = \sum_{B_n \setminus \partial B_n} v_x Lu_x + \sum_{\partial B_n \cap B_n} v_x L^{(n)} u_x,$$

where $L^{(n)}$ is defined by (1.6), but with the sum taken over $y \in \partial_n x$. Now write $u \in \mathcal{U}_n^\phi$ in the form $u = u^\phi + \psi$, so that $\psi = 0$ on $\partial B_n$. Since $Lu^\phi = 0$ on $B_n \setminus \partial B_n$,

$$\|u\|^2 = \langle u^\phi, u^\phi \rangle_n + 2\langle u^\phi, \psi \rangle_n + \langle \psi, \psi \rangle_n = \|u^\phi\|_n^2 + \|\psi\|_n^2.$$

Hence $\|u\|^2$ is uniquely minimized on $\mathcal{U}_n^\phi$ by setting $\psi = 0$. $\square$

Let $u^{(n)}(x) = P^x(X_{\sigma_n} = 0)$ be the unique solution of ($\mathcal{P}_n$) corresponding to $\phi = \delta_{\{0\}}$. By the maximum principle, $\{u^{(n)}\}$ form an increasing sequence on $\mathbb{Z}_+^p$ with limit $\bar{u}_x = P^x\{\exists t \geqq 0: X_t = 0\}$. Clearly $\bar{u}_0 = 1$ and $L\bar{u}_x = 0$ for all $x \in \mathbb{Z}_+^p$.

(ii)$\Rightarrow$(iii). Let $w$ be a bounded solution to $(\mathscr{P})$ of Theorem 1.4. Choose $\alpha>0$ sufficiently small that $\tilde{w}=\alpha w+1-\alpha$ satisfies $0\leqq w\leqq 2$. By the maximum principle, $u^{(n)}\leqq\tilde{w}\leqq 2-u^{(n)}$ on $B_n$. Letting $n\to\infty$, we find that $\tilde{w}\equiv 1$ and hence $w\equiv 1$.

(i)$\Leftrightarrow$(ii). This is an immediate consequence of (1.7), to whose proof we now turn. In what follows, $B_n=\{x:\sum x_i\leqq n\}$.

*Proof of* (1.7). 1°. Let $m=\min_{\mathscr{U}}\|u\|^2$. We show first that $\|\bar{u}\|^2\leqq m$. Let $\{k_n\}\subset\mathscr{U}$ be a minimizing sequence: $\|k_n\|^2\to m$ as $n\to\infty$. Since $k_n\in\mathscr{U}$, there exists an integer $m_n$ for which $|k_n(x)|\leqq 1/n$ for $|x|=m_n$. Let $v_n$ be the solution to $(\mathscr{P}_{m_n})$ for boundary data $\phi=k_n$ on $\partial B_{m_n}$. On $\partial B_{m_n}\cap B_{m_n}$,

$$|v_n-u^{(m_n)}|=|k_n-\delta_{\{0\}}|\leqq 1/n,$$

and the maximum principle implies that this inequality is valid on all $B_n$. It follows that $\{v_n\}$ converges pointwise to $\bar{u}$. We conclude from this and the Dirichlet principle that for any fixed integer $p$:

$$\|\bar{u}\|_p^2=\lim_n\|v_n\|_p^2\leqq\lim_n\|v_n\|_{m_n}^2\leqq\lim_n\|k_n\|_{m_n}^2\leqq\lim_n\|k_n\|^2=m.$$

On letting $p\uparrow\infty$, we find that $\|\bar{u}\|^2\leqq m$.

2°. We show that $\|u^{(n)}\|^2\to m$, and then that $\|\bar{u}\|^2=m$. It follows from the Dirichlet principle that

$$\|u^{(n)}\|^2=\|u^{(n)}\|_{n+1}^2\geqq\|u^{(n+1)}\|_{n+1}^2=\|u^{(n+1)}\|^2.$$

Thus the sequence $\|u^{(n)}\|^2$ decreases to a limit $\tilde{m}\geqq m$. Fix two integers $n>p$ and set $k=(u^{(n)}+u^{(p)})/2$. Using the facts that $u_x^{(n)}=0$ for $|x|\geqq n$ and $u_x^{(p)}=0$ for $|x|\geqq p$ together with the Cauchy-Schwartz inequality and the Dirichlet principle,

$$\begin{aligned}\|u^{(n)}\|^2&\leqq\|k\|_n^2=\tfrac{1}{4}\|u^{(n)}\|_n^2+\tfrac{1}{4}\|u^{(p)}\|_p^2+\tfrac{1}{2}\langle u^{(p)},u^{(n)}\rangle_p\\&\leqq\tfrac{1}{2}\|u^{(p)}\|_p^2+\tfrac{1}{2}\|u^{(p)}\|_p\|u^{(n)}\|_p.\end{aligned}$$

Letting $n\to\infty$, $\|u^{(n)}\|_p^2\to\|\bar{u}\|_p^2$, and then letting $p\to\infty$, we get

$$\tilde{m}\leqq\tfrac{1}{2}\tilde{m}+\tfrac{1}{2}(\tilde{m}\|\bar{u}\|^2)^{1/2}.$$

Combine with previous results to get $m\leqq\tilde{m}\leqq\|\bar{u}\|^2\leqq m$, from which our claim and hence (1.7) are obvious. This completes the proof of (1.7).

Finally, suppose that the minimum in (1.7) is attained in $\mathscr{U}$ by some function $w$. Since $w\in\mathscr{U}$, $\lim_{|x|\to\infty} w_x=0$, and so it follows from the maximum principle that $0\leqq w\leqq 1$ for all $x\in\mathbb{Z}_+^p$, and then, as above, that $w\geqq\bar{u}$. Consequently $\lim_{|x|\to\infty}\bar{u}_x=0$, and hence $\bar{u}\in\mathscr{U}$. This completes the proof of Theorem 1.4. □

*3.1. Example.* (i) (L.D. Brown) This example shows that $\bar{u}$ need not lie in $\mathscr{U}$. Let $p=2$, and define transition rates for $X_t=(X_t^1,X_t^2)$ by $a_{2,x}\equiv 1$, $a_{1,x}=x_1^2$, (for $x_1>1$) and $a_{1,x}=\frac{1}{2}(1+2^{x_2})^{-1}$ (for $x_1=1$). On $\{x:x_1>1\}$, the transition rates of $X_t^i$ depend only on $x_i$, so that $X_t^1$ is transient, and hence $X_t$ itself is transient.

We show however that $\lim_{x_2\to\infty} \bar{u}_{0,x_2} \geqq 1/2$. Let $\sigma_1 = \inf\{t \geqq 0\colon X_t = 0 \text{ or } (1, x_2) \text{ for some } x_2 \geqq 0\}$. Since $X_t^2$ moves up or down with equal probabilities, $P^{(0,x_2)}(\sigma_1 < \infty) = 1$. The function $v_x = \frac{1}{2}(1 + 2^{-x_2}) I\{x_1 = 0\}$ satisfies $Lv_x = 0$ for $x = (0, x_2)$ with $x_2 > 0$, so $v(X_{t\wedge\sigma_1})$ is a bounded $P^{(0,x_2)}$ martingale. From the optional sampling theorem,

$$v_{0,x_2} = P^{(0,x_2)}(X_{\sigma_1} = 0) \leqq P^{(0,x_2)}\{\exists t \geqq 0\colon X_t = 0\} = \bar{u}_{0,x_2},$$

from which the claim follows.

(ii) If $\mathscr{U}$ is extended to include functions (such as $\bar{u}$) with $\liminf_{|x|\to\infty} u_x = 0$, then the theorem is no longer valid. To see this, let $p = 2$ and $P = P_1 \times P_2$, where $P_1$ is a *proper* prior on $\mathbb{R}^1_+$, and $P_2$ yields an inadmissible estimator. The jumps in each component of $X_t = (X_t^{(1)}, X_t^{(2)})$ occur independently of the position of the other component, and according to the probabilities induced by each $P_i$. Hence $X_t$ is transient since $X_t^{(2)}$ is. Now let $\{v^{(n)}(x_1)\}$ be a sequence of functions on $\mathbb{Z}_+$ with $v_0^{(n)} = 1$, $v_\infty^{(n)} = 0$ and energy decreasing to zero in the $P_1$-problem. Then it is easy to check, using the finiteness of $P_1$, that the energy of the functions on $\mathbb{Z}^2_+$ defined by

$$u_x^{(n)} = \begin{cases} 1 & \text{if } x_2 > 0 \\ v^{(n)}(x_1) & \text{if } x_2 = 0 \end{cases}$$

decreases to zero.

As a first application of Theorem (1.4), we derive a result of Clevenson and Zidek (1975). New results are given in Sects. 4 and 6.

**3.2. Corollary.** *$d(x) = x$ is inadmissible for $L_{-1}$ if $p \geqq 2$.*

*Proof.* The estimator $d(x) = x$ is generalized Bayes for the prior $P(d\lambda) = d\lambda$, yielding $a_{i,x} = x_i$ for $x$ such that $x_i > 0$. Put $z = \sum x_i$; it is easy to check that

$$u_x = \prod_{i=1}^{p-1} \frac{i}{i+z} \tag{3.1}$$

satisfies $Lu_x = 0$ if $x \in \mathbb{Z}^p_+ \setminus \{0\}$, and hence is a solution to $(\mathscr{P})$. It follows that $\{X_t\}$ is transient and (from Theorems 1.2 and 1.4) that $d(x)$ must be inadmissible. (The form of the solution was noticed by L.D. Brown.) □

*Remarks.* The hitting probability $\bar{u}_x$ is also given by (3.1). This follows from the observation that

$$k_x^{(n)} = \frac{\prod (i+z)^{-1} - \prod (i+n)^{-1}}{\prod i^{-1} - \prod (i+n)^{-1}}$$

(where the products range over $i = 1, \ldots, p-1$) is the unique solution to $(\mathscr{P}_n)$ for $B = \{\sum x_i \leqq n\}$ and $\phi = \delta_{\{0\}}$.

The second application shows that priors with bounded support lead to recurrent birth and death processes.

**3.3. Corollary.** *If* $\operatorname{supp} P$ *is bounded and* (1.1) *holds, then* $\{X_t\}$ *is recurrent.*

*Proof.* Suppose that $\operatorname{supp} P \subset \{\lambda\colon |\lambda| \leqq M\}$ for some $M$. Hence $d_{P,i}(x) = E(\lambda_i | x - e_i) \leqq M$ for all $i$ and $x \geqq 0$. Since $p_0 < \infty$, $C = P\{|\lambda| < M\} < \infty$, and using Stirling's formula,

$$\frac{p_x}{x!} \leqq C \prod_i \left(\frac{Me}{x_i \vee 1}\right)^{x_i}.$$

Consequently $a_{i,x} \leqq M_1 \prod_i (M_1/x_i \vee 1)^{x_i}$. Define $u^{(n)} \in \mathscr{U}$ by 1 for $|x| \leqq n$, and $1/|x|$ for $|x| > n$. Clearly

$$\|u^{(n)}\|^2 \leqq M_2 \sum_{|x| \geqq n} \prod_i [M_1/(x_i \vee 1)]^{x_i} \to 0,$$

as $n \to \infty$, since the summand lies in $L_1(\mathbb{Z}_+^p)$. Hence $\inf_{\mathscr{U}} \|u\|^2 = 0$. □

*3.4. Remark.* The above assumptions imply finiteness of the integrated risk of $d_P$, and hence its admissibility. Thus the corollary also follows from Theorems (1.2) and (1.4) combined.

In the Gaussian case, the distinction between positive and null recurrence has an important statistical interpretation: namely the distinction between priors of finite and infinite total mass (the so-called 'proper' and 'improper' priors respectively). This was noted by Brown (1971), and the (apparently non-trivial) proof given in Johnstone and Lalley (1985), where the phenomenon was applied to discuss 'immunity' (in the sense of Gutmann 1982, 1983) of Generalized Bayes estimators to the Stein effect. There is a corresponding interpretation of the positive/null recurrence dichotomy in the Poisson situation, but its details depend on the specific choice of loss function. This was discussed for the one dimensional case in *I*, Sect. 4, which notes also that to recover exactly the proper/improper prior case, one needs the loss function $\lambda^{-2}(d-\lambda)^2$. The results, examples and proofs given in *I* generalize to the multivariate situation: we will be content to simply state the results. Let $\{Y_n\}$ be the embedded discrete-time chain associated with $\{X_t\}$: it clearly has transition probabilities

$$p_{x,y} = \alpha_{x,y}/\mu_x \qquad \mu_x = \sum_{y \neq x} \alpha_{x,y}.$$

**3.5. Lemma.** *Suppose that $\{Y_n\}$ is recurrent. It is positive recurrent iff $\sum_{x \in \mathbb{Z}_p^+} \mu_x < \infty$; in which case the invariant probability measure is proportional to $\{\mu_x\}_{x \in \mathbb{Z}_p^+}$.*

We now give a (nearly) sharp condition on the prior $\pi(d\lambda)$ for positive recurrence.

**3.6. Lemma.** *If $\int_{\mathbb{R}_+^p} \lambda_i \, d\pi(\lambda) < \infty$ for $i = 1, \ldots, p$, then $\sum_{\mathbb{Z}_+^p} \mu_x < \infty$. Conversely, if $\int_{\mathbb{R}_+^p} \lambda_i \, d\pi(\lambda) = \infty$ for some $i = 1, \ldots, p$, and for that $i$ either*

$$d_{P,i}(k+e_i) \leqq M d_{P,i}(k) \qquad \text{for all } k \in \mathbb{Z}_+^p$$

or

$$\inf_k d_{P,i}(k+e_i)/(k_i+1) > 0$$

then

$$\sum \mu_x = \infty.$$

**3.7.** Priors with $\operatorname{supp} P \subsetneqq (0,\infty)^p$ can arise in situations where there is prior information on the ordering of the means: for example $\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_k$ (Cohen and Sackrowitz (1970) discuss this problem in the Gaussian case). Theorems 1.2 and 1.4 will remain valid if $\mathscr{U}$ is redefined as

$$\{u_x : u_0 = 1, \lim_r \sup_{\{|x|=r,\, x\in N(P)\}} u_x = 0\},$$

where $N(P)$ is an appropriate neighbourhood of the support of $P$ with the properties

(i) $\sup_{|x|=r,\, x\in N(P)} q_x/p_x \to 0$ if $Q$ has compact support in $(0,\infty)^p$,

(ii) if $\{X_t\}$ is the process corresponding to $P$, then for each large $r$

$$P^x\{\exists t : X_t = 0 \quad or \quad X_t \in N(P) \cap \{|x| \geqq r\}\} = 1.$$

Given (i) and (ii), the only substantial change needed in the proofs is the analogue of Brown's Lemma 4.2.2.

In important special cases, such as $\operatorname{supp} P = \{\lambda : \lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_k\}$, we may set $N(P) = \operatorname{supp} P$. In general, however $N(P)$ cannot be taken as a neighborhood of the convex hull of $\operatorname{supp} P$ as in the Gaussian case: if $P(d\lambda) = I\{\lambda \in \mathbb{R}^2_+ : \lambda_1^2 \lambda_2 \wedge \lambda_1 \lambda_2^2 \leqq 1\}\, d\lambda$, then $\lim_{r\to\infty} q(r,r)/p(r,r)$ need not be zero. We will not go further into the existence and description of $N(P)$ in the general case, except to remark that the transformation (2.4) and the Birnbaum-Stein theorem can be used to established (i) for any $\delta$-neighborhood of $\operatorname{supp} P$ and for certain cones in $\mathbb{Z}^p_+$.

**3.8.** While explicit formulae for $\bar{u}$ are not available in general even for problems in which $P = P^1 \times P^2$ (but cf. Sect. 4), some useful bounds are possible. If $\bar{u}^i_{x_i}$ are the hitting probabilities of zero for the marginal problems induced by $P^i$, then $\bar{u}_x \leqq \bar{u}^1_{x_1} \bar{u}^2_{x_2}$. This is a probabilistic version of the statement "an estimator in a product problem is inadmissible if any component is". It follows from the maximum principle and the observation that $L\bar{u}^1 \bar{u}^2 \leqq 0$ on the boundary of $\mathbb{Z}^p_+$, and is zero elsewhere. Alternatively it may be seen probabilistically from independence of the co-ordinate processes.

**3.9.** We give a sufficient condition for recurrence of the image of a recurrent Markov chain under a transformation of the state space. This will be used in the proof of Theorem 1.3. Let $G$ be a graph with (symmetric) transition rates $\alpha_{x,y}$ between neighboring vertices $x \sim y$. If $g: G \to G$ is a "graph respecting" function: $x \sim y$ implies $gx = gy$ or $gx \sim gy$, then $g$ induces new transition rates $(g\alpha)_{x,y} = \alpha(gx, gy)$, which are interpreted as zero if $gx = gy$. Let $\mathscr{U}$ be a class of real valued functions on $G$, and for $u \in \mathscr{U}$, set $(gu)(x) = u(gx)$.

**3.10. Lemma.** *Suppose that $g\mathscr{U} \subset \mathscr{U}$ and that for some integer $M$,*

$$\sup_{y\sim y'} |\{(x,x') \in G \times G : x \sim x', gx = y, gx' = y'\}| \leqq M.$$

Then

$$\inf_{\mathscr{U}} \sum_{x,x'} (g\alpha)_{x,x'} [u(x')-u(x)]^2 \leqq M \inf_{\mathscr{U}} \sum_{x,x'} \alpha_{x,x'} [u(x')-u(x)]^2.$$

The proof is immediate. For the later application, let $X = \mathbb{Z}_p^+$ with the usual lattice structure and, given functions $a_i(x)$: $\mathbb{Z}_p^+ \cap \{x_i \geqq 1\} \to \mathbb{R}^+$, define rates $\alpha_{x,y}$ as at (1.4) and $\mathscr{U}$ as in §1. Suppose that $gx = (g_1 x_1, \ldots, g_p x_p)$, where $g_i$: $\mathbb{Z}^+ \to \mathbb{Z}^+$ are increasing functions with steps of unit height and length bounded above by $M$. Let $b_i(x) = a_i(g_1 x_1, \ldots, g_i x_i \vee 1, \ldots, g_p x_p)$. Since $b_i(x)[D_i(gu)(x)]^2 \leqq \alpha(gx, g(x-e_i))[D_i(gu)(x)]^2$, it follows from the proof of the Lemma that if $g\mathscr{U} \subset \mathscr{U}$, then

$$\inf_{\mathscr{U}} \sum_i \sum_{x:x_i \geqq 1} b_i(x)[D_i u(x)]^2 \leqq M \inf_{\mathscr{U}} \sum_i \sum_{x:x_i \geqq 1} a_i(x)[D_i u(x)]^2. \tag{3.2}$$

## § 4. Admissibility Results for Symmetric Cases

This section is devoted to estimators $d_P(x) = \Phi(z)\,x$, $z = \sum x_i$, which are generalized Bayes for "simplex symmetric" priors of the form $P(d\lambda) = M(d\lambda)\,d\theta_1 \ldots d\theta_p$, where $\Lambda = \sum \lambda_i$, $\theta_i = \lambda_i/\Lambda$. We show the equivalence of this admissibility problem to that of the one-dimensional Poisson case studied in *I* and derive the consequences for the Clevenson-Zidek estimator amongst others.

Assume for simplicity that $m_z = \int e^{-\Lambda} \Lambda^z M(d\Lambda) \in (0, \infty)$ for $z \geqq 0$. Clevenson and Zidek show (1975, Theorem 2.2) that for $x \geqq 0$ and $z = \sum x_i$

$$d_P(x) = (m_z/m_{z-1})(x/(z+p-1)), \qquad \pi_P(x) = m_z/(z+p-1)! \tag{4.1}$$

We write $d_M(z)$ for $m_z/m_{z-1}$ and set $d_M(0) = 0$. Note that $d_P$ satisfies (1.1).

**4.1. Theorem.** *$d_P$ is admissible for $\lambda$ if and only if $d_M$ is admissible for $\Lambda$ in the one-dimensional problem with loss function $(d-\Lambda)^2/\Lambda$.*

**4.2. Corollary.** *Let $\beta_z = m_z^2/m_{z-1}\,z!$. If $d_P(x)$ is admissible, then $\sum_1^\infty 1/\beta_z = \infty$. The converse holds if $d_M$ satisfies for $z \geqq 1$*

$$\begin{aligned} d_M(z) - z &\leqq c_1 z^{1/2} \\ d_M(z+1) - d_M(z) &\leqq c_2. \end{aligned} \tag{4.2}$$

*Proofs.* 1°. The risk function of $d(x) = \Phi(z)\,x/(z+p-1)$ under $L_{-1}$ is given by

$$\begin{aligned} R(d,\lambda) &= \sum_i E\left[\left(\frac{\Phi(z)}{z+p-1}\right)^2 E(x_i^2/\lambda_i \mid Z=z) - 2\frac{\Phi(z)}{z+p-1} z + \Lambda\right] \\ &= E_\Lambda\left[\frac{z}{z+p-1}\,\frac{(\Phi(z)-\Lambda)^2}{\Lambda} + \frac{p-1}{z+p-1}\Lambda\right], \end{aligned}$$

since $\mathscr{L}(X_i \mid Z=z)$ is binomial with parameters $z$ and $\lambda_i/\Lambda$. We now show that admissibility of $d(x)$ is equivalent to that of $\Phi(z)$ for the loss function $\bar{L}(d, \Lambda, z)$

$=[z/(z+p-1)](d-\Lambda)^2/\Lambda$. First, if $\Phi'$ is strictly better than $\Phi$, it is obvious that $R(d',\lambda)-R(d,\lambda)\leqq 0$, with strict inequality for some $\lambda$. Conversely, for any finite prior $N(d\Lambda)$ charging a neighborhood in $\mathbb{R}^1$ of $\Lambda=1$, let $Q(d\Lambda)=N(d\Lambda)\prod_1^p d\theta_i$. Writing $d_Q(x)=d_N(z)\,x/(z+p-1)$, it is easy to check that

$$B(d,Q)-B(d_Q,Q)=\frac{1}{(p-1)!}\int[\bar{R}(\Phi,\Lambda)-\bar{R}(d_N,\Lambda)]\,N(d\Lambda), \tag{4.3}$$

where $\bar{R}$ is the risk function corresponding to $\bar{L}$. Thus, from the Stein-LeCam characterization, admissibility of $\Phi$ implies that the right side of (4.3) can be made arbitrarily small, which in turn implies admissibility of $d$.

2°. By the usual argument of I. §2.2, the integral on the right side of (4.3) is equal to

$$\sum_{z\geqq 1}(\Phi(z)-d_N(z))^2\,\frac{n_{z-1}}{z!}\,\frac{z}{z+p-1}. \tag{4.4}$$

The factor $z/(z+p-1)$ is bounded between $1/p$ and 1, and can thus be ignored in admissibility considerations. Thus we reduce to the one-dimensional conditional problem $\mathscr{P}_0$ discussed in I §2, and it follows from I, Lemma 2.1 that $\Phi$ is admissible in $\mathscr{P}_0$ iff it is admissible in the original (unconditional) problem. This establishes the Theorem, and the Corollary may now be read off from Theorems 1.1 to 1.3 of I. □

None of the above argument depends on the $p$-dimensional theory of the two preceeding sections. It is instructive therefore to give an alternative derivation of the necessity of Corollary 4.2 by specializing Theorems 1.3 and 1.4 to the 'simplex symmetric' case. This will be applied in Sect. 6 to give explicit tests for admissibility. Let $\mathscr{U}$ and $\|u\|$ be as defined in §2, and $\mathscr{V}=[u\in\mathscr{U}\colon u(x)=v(\sum x_i)$ for some $v\colon\mathbb{Z}_+\to\mathbb{R}]$. Now if $u(x)=v(z)\in\mathscr{V}$, then a computation shows that

$$\|u\|^2=\frac{1}{(p-1)!}\sum_z(Dv_z)^2\,b_z, \tag{4.5}$$

where $b_z=m_z^2/m_{z-1}(z-1)!\,(z+p-1)\sim\beta_z$ as $z\to\infty$, and one uses the fact that there are $\binom{z+p-1}{p-1}$ points $x$ in $\mathbb{Z}_+^p$ with $\sum x_i=z$. We now show that $\min_{\mathscr{U}}\|u\|^2=\min_{\mathscr{V}}\|u\|^2$. Let $B_n=[x\in\mathbb{Z}_+^p\colon\sum x_i\leqq n]$, and $u_x^{(n)}=P^x\{|X_t|$ hits 0 before $n\}$. Noting that for $u\in\mathscr{V}$, $Lu(x)=z!\,\bar{L}v_z/(z+p-1)!$, where $\bar{L}v_z=D^+(b_z D^- v_z)$, it follows as in I §4 that

$$u_x^{(n)}=\left[\sum_1^n 1/b_s\right]^{-1}\sum_{z+1}^n 1/b_s,\qquad x\in B_n,$$

so that $u^{(n)}\in\mathscr{V}$. As in 2° of the proof of Theorem 1.4, $\|u^{(n)}\|^2\searrow\min_{\mathscr{U}}\|u\|^2$, and since $\mathscr{V}\subset\mathscr{U}$, this suffices to show equality of the two minima.

Suppose now that $d_P(x)$ is admissible. If $\operatorname{supp}P$ is unbounded, divergence of $\sum 1/b_s$ follows from 1°, Theorem 1.2, Remark 2.3 and Lemma I.3.1. If $\operatorname{supp}P$ and hence $\operatorname{supp}M$ is bounded then the argument of Lemma I.3.2 applies to $M(d\Lambda)$ and $m_z$.

The following comparison test provides an easily checked criterion for (in-) admissibility in many applications.

**4.3. Corollary.** *Let P be a planar symmetric prior.*

(i) *If for some* $\delta>0$, $d_P(x)\cdot 1 \geqq z-(p-1)+\delta$ *for large* $z$, *then* $d_P$ *is inadmissible;*

(ii) *If* (4.2) *holds and* $d_P(x)\cdot 1 \leqq z-(p-1)$ *for large* $z$, *then* $d_P$ *is admissible.*

*Proof.* Theorem 4.1 allows an appeal to Corollary 5.2 of I, which in fact gives the stronger result that $d_M(z) \geqq z+\alpha/\log z$ for $\alpha>1$ and large $z$ implies inadmissibility, and $d_M(z) \leqq z+\alpha/\log z$ for $\alpha \leqq 1$, $z$ large implies admissibility, under (4.2). □

*Examples.* 1. $M(d\Lambda)=\Lambda^{p-1}d\Lambda$, $m_z=(z+p-1)!$, and $d_P(x)=x$ - the "usual" estimator. Clearly, $d$ is admissible iff $p=1$ (Clevenson and Zidek, 1975).

2. $M(d\Lambda)=\int_0^\infty (1+\Lambda t)^{-\beta} t^{-p} \exp(-t^{-1})\,dt\,d\Lambda$. These priors were used by Cleveson and Zidek who showed that $d_P(x)=\left(1-\dfrac{\beta+p-1}{z+\beta+p-1}\right)x$ is admissible for $\beta>1$. Clearly, $d_P$ is admissible iff $\beta \geqq 0$. Brown and Hwang (1982) used their unified admissibility method to establish the admissibility half, and Hwang (1982) treated the case $\beta<0$ via difference inequalities.

3. $M(d\Lambda)=\Lambda^{p-1}\int_0^\infty g(t)\,t e^{-t\Lambda}\,dt\,d\Lambda$. Priors of this form were studied by Ghosh and Parsian (1981), who showed that they included the Clevenson-Zidek family. If $g(t)=Ct^{m-1}(1+t)^{-m-n}$, then Ghosh and Parsian show admissibility for $d_P(x)=\left(1-\dfrac{m+p}{z+m+n+p-1}\right)x$ for $m>0$. It is obvious from Corollary 4.3 that $d_P$ is admissible iff $m \geqq -1$.

*4.4. Remark.* What happens for planar symmetric priors under squared error loss $L_0$? Suppose $P(d\lambda)=M(d\Lambda)\prod_{i=1}^{p} d\theta_i/\theta_i$, so that $p_x=\int e^{-\Lambda}\Lambda^z M(d\lambda)\int \theta^{x-1}d\theta$ $=m_z(x-1)!/(z-1)!$. This definition is chosen so that under $L_0$, $d_{P,i}(x)=p_{x+e_i}/p_x$ $=(m_{z+1}/m_z)(x_i/z)$. In particular, the MLE $d(x)=x$ arises from $M(d\Lambda)=\Lambda^{-1}d\Lambda$. It turns out that the $p$-dimensional admissibility problem for $d_P$ under $L_0$ is *not* isomorphic to the 1-dimensional question for $d_M(z)=m_{z+1}/m_z$. In this case, the analogue of (4.3) and (4.4) is found to be

$$B(d_P,Q)-B(d_Q,Q)=\sum_{z\geqq 0}\left(\frac{m_{z+1}}{m_z}-\frac{n_{z+1}}{n_z}\right)^2 \frac{n_z}{z\cdot z!}\,\gamma(z),$$

where for large $z$

$$C_p z(\log z)^{p-1} \leqq \gamma(z)=\sum_{|x|=z}\left(\sum x_i^2\right)/\prod_j (x_j \vee 1) \leqq C_p^1 z(\log z)^{p-1}, \tag{4.6}$$

and $|x|=\sum x_i$ (see Appendix A.3 for proof).

Only a sufficient condition seems to be available: $d_P$ is admissible if $d_M$ satisfies (4.2) and

$$\sum 1/(\bar{b}_z \log^{p-1} z)=\infty, \tag{4.7}$$

where $\tilde{b}_z = m_z^2/m_{z-1}(z-1)!$. As an example, for the MLE $\tilde{b}_z = z-1$, which yields admissibility for $p=1$ and 2 (Peng, 1975). The proof that (4.7) entails admissibility is obtained by modifying the argument of I § 5, working with the density function $\tilde{p}_\lambda(z) = p_\lambda(\log z)^{p-1}$ (not a probability density!). The proof of the Poincaré inequalities of I Lemma 5.1 is no longer valid: instead a one dimensional argument on the lines of the present § 5 is needed. (The $(\log z)^{p-1}$ factor, being slowly varying, causes no problems.)

I do not know if finiteness of (4.7) is necessary for admissibility. This would be true if finiteness were equivalent to a zero infimal energy condition for the terms (2.5) in the setting of Remark 2.4. However, the elliptic difference operators associated with (2.5) are no longer one-dimensional on $\mathscr{V}$ (compare the discussion around (4.5) above).

## § 5. The General Admissibility Theorem

In this section we prove Theorem 1.3. As the argument is long and technical, it is split over subsections 5.1 through 5.6 and is prefaced by a discussion of the plan of the proof. Subsections 5.1 and 5.2 contain material on Poincaré inequalities and tail behavior of Poisson densities that may be of independent interest. § 5.3 provides bounds on the growth of marginal densities and Bayes estimators that flow from assumption A. The proof proper of Theorem 1.3 is spread over § 5.4–5.6. Finally § 5.7 addresses estimator $d_i(x)$ which may violate (1.1) by being positive on $\{x: x_i = 0\}$, for later use in Sect. 6.

First, some general comments on the nature of Assumption A. Essentially, the generalized Bayes rule $d_P(x)$ is required to be approximable by a 'product rule' $d^*(x) = (d_1(x_1), \ldots, d_p(x_p))$, each of whose components $d_i(x_i)$ is a function of the $i^{\text{th}}$ observation $x_i$ alone and satisfies the conditions of the corresponding univariate result (Theorem 1.2 in I.). Conditions a) and b) refer to approximation of $d_{P,i}(x_i)$ by $d_i(x_i)$ at the boundaries for $x_i$ at $\infty$ and 0 respectively. Brown's (1971) condition in the normal case that $\delta_F(y) - y$ be bounded is analogous to assumption a), for the special choice $d_i(y_i) = y_i$. The analogy may be seen through the (variance stabilizing) transformation $y_i = \sqrt{x_i}$, $\theta_i = \sqrt{\lambda_i}$, $\delta_i(y) = \sqrt{d_i}(x)$ used by Brown (1979b). The boundary at 0 does not occur in the Gaussian setting so condition b) has no counterpart there.

We shall now describe, with some deliberate lack of precision, the thrust of the argument that follows. In partial contrast with Brown's (1971) method, the approach is almost entirely analytic, rather than probabilistic. The heuristics described in Brown (1971, § 1) and I, § 5 do still provide useful guidance. Regrettably, the argument is complicated by two things: the lack of invariance properties of the Poisson family – necessitating separate estimates for $\lambda \to 0$, moderate $\lambda$ and $\lambda \to \infty$, and the discreteness of the sample space.

The aim of the proof is to bound the difference in integrated risks $B(d_P, Q) - B_{\mathscr{D}_0}(Q)$ in terms of the energy condition (1.3) or (1.5). The prior measure $P$ is fixed throughout. The function $u$ in the energy norm corresponds to $Q$ as follows: $\sqrt{\frac{dQ}{dP}}(x)$ equals $u(x)$ smoothed by a uniform distribution having position dependent bandwidth $r(x)$. The choice of $dQ/dP=u^2$ in the normal case is motivated by Brown (1971, p. 861), while the technical value of smoothing here is explained below.

The difference in integrated risks $B(d_P, Q)-B_{\mathscr{D}_0}(Q)$ is an $L^2$ distance between the Bayes estimates $d_P$ and $d_Q$ (cf. 5.13). The first step in relating this to the (discrete) Sobolev norm of $u$ of (1.3) is to express $(d_{Q,i}(x)-d_{P,i}(x))^2$ in terms of a (weighted) posterior variance of $\bar{u}(\lambda)$ (cf. 5.14). By triangle inequality methods and integrating now over $x$ also this can be estimated by a (weighted) $L^2$ norm

$$\int \omega_{\bar{u}}^2(x,\lambda)\, M(dx, d\lambda)$$

of the oscillation $\omega_{\bar{u}}(x,\lambda)=\bar{u}(x)-\bar{u}(\lambda)$(cf. (5.16)).

The task is then to convert the oscillation bounds to bounds on the derivatives (and then differences) of $u$. For this we use Poincaré inequalities of the form

$$\int_{B_1} (\bar{v}(x)-\bar{v}(0))^2\, dx \leqq C \int_{B_1} |\nabla v(x)|^2\, dx \tag{5.1}$$

where $v$ is $C^1$, $\bar{v}=v*I_{B_1}$, $I_{B_1}$ is the indicator function of the unit ball $B_1$ in $\mathbb{R}^p$ and $*$ denotes convolution. This inequality is demonstrably false for $p\geqq 2$ if $\bar{v}$ is replaced by $v$ on the left side (but is valid if $v$ is a solution of an elliptic differential equation - see Johnstone and Shahshahani (1983)).

To apply (5.1), we return to the $L^2$ norm of $\omega_{\bar{u}}(x,\lambda)$. The aim is to fix $\lambda$ and express the average (over $x$) of $\omega_{\bar{u}}^2$ as a mixture of (centered and scaled) integrals of the form of the left side of (5.1) with 0 and $B_1$ replaced by $\lambda$ and $B_r(\lambda)=\{u\colon |u-\lambda|<r\}$ respectively. To achieve this, it is convenient to replace $M(dx, d\lambda)$ by an upper bound $\tilde{M}(dx, d\lambda)$ for which the conditional density of $x$ given $\lambda$ is strictly unimodal about $\lambda$ (cf. 5.18). Applying (5.1) to each element of the mixture (Corollary 5.2) yields a bound of the form

$$\int [\bar{u}(x)-\bar{u}(\lambda)]^2\, \tilde{M}(dx\,|\,\lambda) \leqq C \sum_i \int dx\, [D_i u(x)]^2 a_i(x,\lambda). \tag{5.2}$$

From this point the line of argument is conceptually fairly clear, but technically cumbersome and non-trivial, due to the lack of invariance. Suppose that we were to average over the $M$-marginal measure of $\lambda$, namely $P(d\lambda)$. All would be essentially finished if $\int a_i(x,\lambda)\, P(d\lambda)$ could be bounded by the intensities $a_i(x)$ appearing in the energy expression (1.3).

To attempt this, the first step is to derive a simpler bound for the output $a_i(x,\lambda)$ of the integration by parts procedure of Corollary 5.2 (cf. 5.6). Now $\tilde{M}(dx\,|\,\lambda)$ involves the Poisson density $p_\lambda(x)$ and certain multiples by rational functions of $x$ and $\lambda$ (cf. 5.16), so $a_i(x,\lambda)$ involves integrals, $L$ say, of (weighted) tails of the Poisson distribution (in both directions away from each mean $\lambda_i$, $i=1,\ldots,p$) (cf. 5.19)). After some algebraic reduction (Appendix 2), it develops that these integrals of weighted Poisson tails may be assumed to be (products of) the forms

$$\int_t^\infty \quad \text{or} \quad \int_0^t \left|\frac{s-\lambda}{\sqrt{\lambda}}\right|^b |p'_\lambda(s)|\,ds$$

considered in Corollary 5.7. Corollary 5.7 follows directly from Lemma 5.6, which says that, because of the exponential tails of the Poisson density, the polynomial factor $|(s-\lambda)/\sqrt{\lambda}|^b$ can be ignored, at the expense of shifting the argument of $p'_\lambda(\cdot)$ a fixed number of standard deviations towards the mode $\lambda$ of $p_\lambda(\cdot)$. Corollary 5.7 is then applied in Appendix 2 to bound the integrals $L$ as in (5.19).

The upshot of all this is that $a_i(x,\lambda)$ can be bounded by $\tilde{a}_i(x,\lambda)$ which consists essentially of rational functions of $\lambda$ and $d_i(\lambda)$ multiplied by a smoothed version of $p_\lambda(x)$ with bandwidth $2c'\sqrt{x}$ (cf. 5.20). The penultimate step is to show that $\int \tilde{a}(x,\lambda)\,dP(\lambda)$ is bounded by not (alas) $a_i(x)$, but at least by $a_i(g(x))$, where $g_i(x_i)-x_i\in O(\sqrt{x_i})$. Since $\int p_\lambda(x)\,dP(\lambda)=\pi_x$, this part (cf. subsection 5.6) requires comparisons of marginal densities $\pi_{x+a}$ over ranges of order $c'\sqrt{x}$. These are accomplished with the aid of Lemma 5.8 and 5.9 which bound the growth rate of such marginal densities and Bayes rules. The various provisions of Assumption A are used to convert the rational functions of $\lambda$ into functions of $x$ when necessary, and thence to obtain the desired bounds (for example (5.21) and ff.). Finally, the transition from $a_i(g(x))$ to $a_i(x)$ in the energy norms (cf. 5.25) is handled by Lemma 3.10, a simple general condition for recurrence of the image of a recurrent Markov chain under a transformation of the state space.

We restrict attention throughout to rules having everywhere finite risk function.

### *5.1. Inequalities of Poincaré Type*

For this statistical application, it is useful to have inequalities of Poincaré type for functions which vanish at a specified interior point. To accomplish this one can smooth the function with an appropriate kernel. The Gaussian situation is simplest, in part because kernels of fixed width suffice, and the approach offers material simplifications of Brown's (1971) original proof (Johnstone, 1983b). To develop these ideas in the Poisson setting, kernels which are indicators of boxes having variable widths are needed. Let

$$I_a(x)=I\{x\in\mathbb{R}^p\colon |x_i|<a_i \text{ for all } i\}\Big/\prod_{i=1}^p a_i.$$

**5.1. Proposition.** *Let $r_i\colon \mathbb{R}\to\mathbb{R}_+$, $i=1,\dots,p$ be monotone functions such that $|r'_i(x)|\le\beta<1$. For a piecewise $C^1$ function $u\colon\mathbb{R}^p\to\mathbb{R}$, define $\bar{u}(x)=u*I_{r(x)}(x)$. Then*

$$\int_{[0,1]} [\bar{u}(x)-\bar{u}(0)]^2\,dx \le C_{p,\beta}\big(\sum_i r_i(0)^{1-p}\big) \int_{[0,1]_r} |Du(x)|^2\,dx, \tag{5.3}$$

*where* $[0,1]_r=[0,1]+r([0,1])$ *is the Hausdorff sum of* $[0,1]\subset\mathbb{R}^p$ *and its image under* $r=(r_1,\dots,r_p)$.

*Proof.* First note that if g is a piecewise $C^1$ function vanishing at 0, then

$$\int_{[0,1]} g^2(x)\,dx \leqq C_p \int_{[0,1]} \frac{|Dg(x)|^2}{|x|^{p-1}}\,dx. \tag{5.4}$$

Write $\bar{u}(x)=\int_{[-1,1]} u(x+sr_x)\,ds$, employ (5.4) and the bound $|r_i'(x)|\leqq 1$ to get

$$\begin{aligned}\int_{[0,1]}[\bar{u}(x)-\bar{u}(0)]^2\,dx &\leqq 2^p \int_{[-1,1]} ds \int_{[0,1]} dx[u(x+sr_x)-u(sr_0)]^2\\ &\leqq C_p \int_{[0,1]} dx \int_{[-r_x,r_x]} \frac{dy}{\Pi r_i(x_i)}\frac{|Du(x+y)|^2}{|x|^{p-1}}\\ &\leqq C_p \int_{[0,1]_r} dz|Du(x)|^2 \int_{|t_i|\leqq r_i(z_i-t_i)} \frac{dt}{\Pi r_i(z_i-t_i)|z-t|^{p-1}}.\end{aligned}$$

It remains to bound the inner integral, denoted $m(z)$. Since $\beta<1$, one can define $t_i^{\pm}(z_i)=t_{z_i}^{\pm}$ by the equations $t_{z_i}^{\pm}=r_i(z_i\pm t_{z_i}^{\pm})$. At this point we assume, for sake of definiteness only, that each $r_i$ is decreasing. It follows that

$$m(z)\leqq\prod_i \frac{t_i^-(z_i)}{t_i^+(z_i)}\int_{[-1,1]}\frac{ds}{|z-st_z^-|^{p-1}}. \tag{5.5}$$

Notice from the bound on $r'(z)$ that $t_{z_i}^-/t_{z_i}^+\leqq(1+\beta)/(1-\beta)=\gamma$, say. Let $z^0$ be the closest point in $[z-t_z^-,z+t_z^-]$ to 0. If $|z_i^0|>r_i(0)$ for some $i$, then $m(z)\leqq(2\gamma)^p r_i^{1-p}(0)$. In the contrary case, replace $z$ by 0 in the integral in (5.5) to obtain the bound

$$m(z)\leqq\gamma^p\Pi'\,1/t_{z_i}^-\leqq\gamma^p\sum_i\left(\frac{r_i(0)}{t_{z_i}^-}\right)^{p-1}\frac{1}{r_i^{p-1}(0)},$$

where $\Pi'$ means that the smallest term in the product is omitted. It is then easy to check that $r_i(0)/r(z-t_{z_i}^-)\leqq(1-\beta)^{-1}$, and this completes the proof. □

Proposition 5.1 will be applied to integrals with respect to unimodal density functions by using partial integration.

**5.2. Corollary.** *Suppose that* $Q_i:\mathbb{R}_+\to\mathbb{R}_+$ *is increasing on* $[0,\lambda)$, *decreasing on* $(\lambda,\infty)$ *and zero off* $\mathbb{R}_+$. *Set* $Q(y)=\Pi_i Q_i(y_i)$. *Fix* $\lambda\in\mathbb{R}_+^p$ *and for* $s\in\mathbb{R}^p$, *let* $A_s$ *be the section at* $s$ *of* $A=\{(s,x)\in\mathbb{R}_+^p\times\mathbb{R}_+^p:\ s_i<x_i<\lambda_1$ *or* $\lambda_i<x_i<s_i,\ i=1,\dots,p\}$. *Suppose* $u$, $r_i$ *and* $\bar{u}$ *are as in Proposition* 5.1. *Then*

$$\begin{aligned}&\int[\bar{u}(x)-\bar{u}(\lambda)]^2 Q(x)\,dx\\ &\quad\leqq C_{p,\beta}\sum_i\int dx[D_i u(x)]^2\int_{B_x} dQ(s)(s_i-\lambda_i)^2\sum_j\left|\frac{s_j-\lambda_j}{r_j(\lambda_j)}\right|^{p-1}\end{aligned} \tag{5.6}$$

*where* $B_x$ *is the section at* $x$ *of* $B=\{(s,x)\in\mathbb{R}_+^p\times\mathbb{R}^p: x\in A_s+r(A_s)\}$ ($C_{p,\beta}$ *does not depend on* $\lambda$, $u$ *or* $Q$).

*Proof.* Write (5.6) as an integral on $A$ and apply Proposition 5.1 on the sections $A_s$ (after making the appropriate changes of variable). Then use Fubini's theorem.

5.3. *Remark.* If all $r_i(x)=(x\vee 1)^{1/2}$ for $x\geqq 0$, then straighforward calculations shows that on setting $s(x)=1/2+(x+1/4)^{1/2}$, we have if $\lambda\geqq 1$,

$$B_x\subset\begin{cases}[x+1-s(x),\infty) & \text{if } x>\lambda+\lambda^{1/2}\\ [0,\infty) & \text{if } |x-\lambda|\leqq\lambda^{1/2}\\ [0,x+s(x)] & \text{if } 0\leqq x\leqq\lambda-\lambda^{1/2}\end{cases}$$

and if $\lambda<1$,

$$B_x\subset\begin{cases}[x+1-s(x),\infty) & \text{if } x>2\\ [x-1,\infty) & \text{if } \lambda+1\leqq x<2\\ [0,\infty) & \text{if } \lambda-1<x<\lambda+1\\ [0,x+1] & \text{if } -1\leqq x\leqq\lambda-1.\end{cases}$$

(Here $x, t, \lambda$ are scalars, and $B_x\subset\mathbb{R}^p$ in Corollary 5.2 is a product of sets of the above type.)

5.4. *Remark.* Suppose, as occurs in § 5.4, that $u(x)=u(|x_1|,\dots,|x_p|)$ is defined by reflection from the positive orthant. Since $B_x\subset B_{|x|}$, it follows that the integral on $x$ on the right side of (5.6) can be restricted to $\mathbb{R}_+^p$.

### 5.2. *Tails of Poisson Densities*

The results of this subsection express the exponential decline of the tails of the Poisson density. The Poisson density has no invariance properties, so the cases of large ($\lambda\geqq 1$) and small ($\lambda<1$) means are treated separately. Otherwise, the bounds are uniform with respect to $\lambda$. Throughout, the density function is extended to $x\in\mathbb{R}^+$ via the formula $p_\lambda(x)=e^{-\lambda}\lambda^x/\Gamma(x+1)$. For the rest of Sect. 5, *we abbreviate* $(x\vee 1)^{1/2}$ *by* $\sqrt{x}$. (However, $\lambda^{1/2}$ is still the usual square root of $\lambda$.)

**5.5. Proposition.** *Given an integer $k>0$, and $c\in\mathbb{R}^+$, there exist $M=M_{c,k}$ and $c'>c+k/2$ such that*

(i) *if $\lambda\geqq 1$; $|\delta|\leqq c\sqrt{x}$ and $x+\delta\geqq 0$, then*

$$\left|\frac{x-\lambda}{\lambda^{1/2}}\right|^k p_\lambda(x+\delta)\leqq M\{p_\lambda(x-c'\sqrt{x})+p_\lambda(x+c'\sqrt{x})\}, \tag{5.7}$$

*and*

(ii) *if $\lambda<1$; $|\delta|\leqq c\sqrt{x}$ and $x+\delta\geqq 1$, then*

$$((x-\lambda)^k/\lambda^{I(k)})\, p_\lambda(x+\delta)\leqq M p_\lambda(x-c'\sqrt{x}). \tag{5.8}$$

*(Here $I(k)=I\{k>0\}$, and if $k=0$, $c'$ may be set equal to $c$.)*

*Convention.* If $x-c'\sqrt{x}<0$, then it is replaced by 0 in the formulas above.

*Remark.* To motivate (5.7), consider a Gaussian analogue obtained as follows. Replace $\sqrt{x}$ by $\sqrt{\lambda}$ in the arguments of $p_\lambda(\cdot)$. Approximate the Poisson $(\lambda)$ density by $N(\lambda,\sqrt{\lambda})$, and then standardize to zero mean and unit variance. Then an analogue of (5.7) is the claim that if $|\delta|\leqq c$, then there exists $c'>c$ such that

$$|x|^k\phi(x+\delta)\leqq M_{c,k}[\phi(x-c')+\phi(x+c')],$$

where $\phi$ is the standard Gaussian density.

*Proof of Proposition* 5.5. The argument involves consideration of several cases and is outlined in Appendix 1. The ingredients are strict unimodality (log-concavity) of $x\to p_\lambda(x)$ and the bounds on Poisson probability ratios given in Proposition I.7.1.

We need also a more specific bound on the tails of the derivative $p'_\lambda(x)$ away from the mean $\lambda$.

**5.6. Lemma.** *Suppose $\lambda\geqq 1$. Given constants $b\geqq 0$, $c>b\vee 1$; if $|y|\geqq 2c\lambda^{1/2}$, then*

$$|y/\lambda^{1/2}|^b|p'_\lambda(\lambda+y)|\leqq M_{b,c}|p'_\lambda(\lambda+y-c\,\mathrm{sgn}(y)\,\lambda^{1/2})|.$$

**5.7. Corollary.** *Suppose $\lambda\geqq 1$, $c\geqq 2b\vee 2$. There exists $M=M_{b,c}$ such that:*

*If* $t\geqq\lambda+c\lambda^{1/2}$, $\displaystyle\int_t^\infty\left|\frac{s-\lambda}{\lambda^{1/2}}\right|^b|p'_\lambda(s)|\,ds\leqq Mp_\lambda(t-c\lambda^{1/2}/2)$.

*If* $0\leqq t\leqq\lambda-c\lambda^{1/2}$, $\displaystyle\int_0^t\left|\frac{\lambda-s}{\lambda^{1/2}}\right|^b p'_\lambda(s)\,ds\leqq Mp_\lambda(t+c\lambda^{1/2}/2)$.

*Proof* (of the lemma). Write $s$ for $y/\lambda^{1/2}$, then

$$\frac{s^b\dfrac{d}{ds}p_\lambda(\lambda+s\lambda^{1/2})}{\dfrac{d}{ds}p_\lambda(\lambda+(s\mp c)\,\lambda^{1/2})}=\frac{s^bp_\lambda(\lambda+s\lambda^{1/2})}{p_\lambda(\lambda+(s\mp c)\,\lambda^{1/2})}\cdot\frac{h(s)}{h(s\mp c)},\tag{5.9}$$

where

$$h_\lambda(s)=\psi(\lambda+s\lambda^{1/2}+1)-\log\lambda=\int_0^\infty(e^{-t\lambda}-(1+t)^{-(\lambda+s\lambda^{1/2}+1)})\frac{dt}{t}.$$

Here we have used integral representations of $\psi(z)=\dfrac{d}{dz}\log\Gamma(z)$ and $\log\lambda$ given in Lebedev (1972, § 1.3). It follows that $h(s)$ is log-concave and that $s\to h(s)/h(s-c)$ is decreasing for $s>c$. Thus, for $s>2c$, $h(s)/h(s-c)\leqq h_\lambda(2c)/h_\lambda(c)$, which may be bounded for $\lambda\in[1,\infty)$ using asymptotic expansions for $\psi$ (Abramowitz and Stegun, 1955, p. 259). A similar argument applies for $s<-2c$. The first term on the right side of (5.9) may be bounded using Proposition I.7.1 and the techniques of Proposition 5.5. □

### 5.3. *Growth of Marginal Densities and Bayes Rules*

The next result is used to show that the marginal density $\pi(x)$ of $P(d\lambda)$ grows by a bounded fraction over distances of order $\sqrt{x}$ (i.e. one standard deviation if $\lambda=x$). Lemma 5.9 has a similar objective, but explicitly includes a growth term for $d_P(x)$.

**5.8. Lemma.** *If $d_P$ is generalized Bayes for $P(d\lambda)$ and satisfies*

$$d_{P,i}(x) \leqq x_i + M(x_i \vee 1)^{1/2} \tag{5.10}$$

*then there exists a constant $C$ such that*

$$\pi(x+ae_j)/\pi(x) \leqq C \exp(Ma/(x_j+1)^{1/2}) \qquad a, x \in \mathbb{R}_+^p.$$

*Notational Conventions.* For $y \in \mathbb{R}_+^p$, $\Gamma(y) = \Pi\Gamma(y_i)$, $\pi(y) = p_y/\Gamma(y+1)$, $d_i(y) = p_y/p_{y-e_i}$, $1 = (1, \ldots, 1)$.

*Proof.* Suppose that $y \in \mathbb{Z}_p^+$, and $0<c<1$. We first show that $\pi(y+ce_1)/\pi(y)$ and $\pi(y)/\pi(y-ce_1)$ are bounded. By Hölder's inequality, for $y \geqq 1$,

$$\frac{\pi(y)}{\pi(y-ce_1)} \leqq C \frac{\Gamma(y+1-ce_1)}{\Gamma(y+1)} \left(\frac{d_{P,1}(y+(1-c)e_1)}{y_1+1-c}\right)^c \leqq C' \left(\frac{d_{P,1}(y+e_1)}{y_1+1}\right)^c \leqq C'',$$

since $a \to d_{P,1}(y+ae_1)$ is increasing. A similar argument applies for $\pi(y+ce_1)/\pi(y)$.

It remains to prove the result for $x \in \mathbb{Z}_p^+$, $a \in \mathbb{Z}^+$. Using (5.10), we have

$$\begin{aligned}\pi(x+ae_1)/\pi(x) &= \prod_{i=1}^{a} \frac{d_{P,1}(x+ie_1)}{x_1+i} \\ &\leqq (1+M/(x_1+1)^{1/2})^a \leqq \exp(Ma/(x_1+1)^{1/2}). \quad \square\end{aligned}$$

**5.9. Lemma.** *Let $d$ be generalized Bayes for $P(d\lambda)$, Lipschitz and satisfy* (5.10). *If $0 \leqq a < b \leqq x_i$ are such that $b/(x_i-b+1)^{1/2} \leqq M_0$ and $(b-a)/(x_i-b+1)^{1/2} \geqq \varepsilon$, then*

$$\frac{(d(x)-x)^2}{x_i \vee 1} \frac{\pi(x-ae_i)}{\pi(x-be_i)} \leqq C(M_0, M, \varepsilon). \tag{5.11}$$

*Proof.* It suffices to prove this in the one-dimensional case ($p=1$) for $x \geqq 1$. Introduce $\alpha(x) = \sqrt{x}\{d(x)/x - 1\} \leqq M$ in virtue of (5.10). Since $d_P$ is Lipschitz $\alpha(x+1) - \alpha(x) \leqq M_1/\sqrt{x}$, where $M_1 = M_1(M)$ and hence the left side of (5.11) is bounded by

$$\begin{aligned}&\alpha^2(x) \prod_{y=x-b+1}^{x-a} \left(1+\frac{\alpha(y)}{y^{1/2}}\right) \\ &\leqq \alpha^2(x) \exp\left\{\left[\alpha(x) + \frac{bM_1}{(x-b+1)^{1/2}}\right] \frac{b-a}{(x-b+1)^{1/2}}\right\} \leqq C(M_0, M, \varepsilon). \quad \square\end{aligned}$$

*5.10. Remark.* The posterior risk of $d_P$ satisfies some useful identities (proved by direct calculation exactly as in (I.2.1)):

$$E[(\lambda_i - d_{P,i}(x))^2/\lambda_i \mid x] = D_i^+ d_{P,i}(x),$$
$$E[(\lambda_i - d_{P,i}(x))^2 \mid x - e_i] = d_{P,i}(x)\, D_i^+ d_{P,i}(x). \tag{5.12}$$

*5.4.*

We now begin the proof that $d_P$ is admissible in $\mathcal{D}_0$ (this suffices for Theorem 1.3, by Remark 2.1). Let $u \in \mathcal{U}$ and extend $u$ to be defined on $\mathbb{R}_+^p$ by linear interpolation and then to $\mathbb{R}^p$ by reflection in the co-ordinate axes. Let $r(x) = \sqrt{x}$ $(=(x \vee 1)^{1/2})$, and define a smoothed version $\bar{u}$ of $u$ by $\bar{u}(x) = (u * \iota_{r(x)})(x)$, where $\iota_r$ denotes the uniform probability density on $\prod_{i=1}^{p} [-r_i, r_i]$. Finally let $Q(d\lambda) = \bar{u}^2(\lambda) P(d\lambda)$.

In view of Remark (2.1), we write (2.2) in the form

$$B(d_P, Q) - B_{\mathcal{D}_0}(Q) = \sum_i \sum_{x \in T_i} [d_{P,i}(x) - d_{Q,i}(x)]^2 q_{x-e_i}/x!, \tag{5.13}$$

where as before $T_i = \{x \in \mathbb{Z}_+^p : x_i > 0\}$. The argument initially parallels that of 1° of I § 5, so we give here only an outline of this part. The first task is to express $d_Q - d_P$ in terms of the prior density $\bar{u}$, with result

$$[d_{P,i}(x) - d_{Q,i}(x)]^2 q_{x-e_i}$$
$$\leq 4 \int (\lambda_i - d_{P,i}(x))^2 \lambda_i^{-1} [\bar{u}(\lambda) - E(\bar{u}(\lambda) \mid x - e_i)]^2 e^{-\Lambda} \lambda^x dP(\lambda). \tag{5.14}$$

To express the right side in terms of oscillations of $\bar{u}$, we employ the bound

$$\bar{u}(\lambda) - E(\bar{u}(\lambda) \mid \underset{\sim}{x} - e_i) \leq 2(\bar{u}(\lambda) - \bar{u}(x))^2 + 2E[(\bar{u}(\lambda) - \bar{u}(x))^2 \mid \underset{\sim}{x} - e_i]. \tag{5.15}$$

It is convenient for the Poincaré inequalities to allow $x$ to range continuously in $T_i$. As a result, replace the sums on $x \in T_i$ in (5.15) by integrals on $\{x \in \mathbb{R}_+^p : x_i \geq 1\}$, which is denoted $S_i$. Thus, in (5.15), $x \in S_i$ and $\underset{\sim}{x}$ is shorthand for $[x]$. The "surplus" conditional expectation on the right side of (5.15) is removed via the following identity, which flows from (5.12) and Fubini's theorem:

$$E[(\bar{u}(\lambda) - \bar{u}(x))^2 \mid \underset{\sim}{x} - e_i] \int dP(\lambda)(\lambda_i - d_{P,i}(\underset{\sim}{x}))^2 \lambda_i^{-1} p_\lambda(\underset{\sim}{x})$$
$$= \frac{d_{P,i}(\underset{\sim}{x})}{\underset{\sim}{x}_i} D_i^+ d_{P,i}(\underset{\sim}{x}) \int (\bar{u}(\lambda) - \bar{u}(x))^2 \frac{\underset{\sim}{x}_i}{\lambda_i} p_\lambda(\underset{\sim}{x})\, dP(\lambda),$$

where the coefficient in front of the integral is uniformly bounded for $x \in S_i$, in view of Assumption A. Combining now the pieces, we finally obtain the following analog of I (5.4) as a bound for (5.13):

$$M \sum_i \int dP(\lambda) \left\{ \frac{(\lambda_i - d_i(\lambda))^2}{\lambda_i} \int_{S_i} dx [\bar{u}(\lambda) - \bar{u}(x)]^2 p_\lambda(\underset{\sim}{x}) \right.$$
$$\left. + \int_{S_i} dx [\bar{u}(\lambda) - \bar{u}(x)]^2 \left[ \frac{(\lambda - x)^2}{\lambda_i} + \frac{x_i \vee 1}{\lambda_i} \right] p_\lambda(\underset{\sim}{x}) \right\}. \tag{5.16}$$

5.5.

Fix $i=1$. We now apply the Poincaré inequality of Corollary 5.2 to bound

$$I^{(1)}=\int_{S_1} dx[\bar{u}(x)-\bar{u}(\lambda)]^2 p_\lambda(\underline{x}), \quad \text{and}$$
$$I^{(2)}=\int_{S_1} dx[\bar{u}(x)-\bar{u}(\lambda)]^2 \left[\frac{(\lambda_1-\underline{x}_1)^2+\underline{x}_1}{\lambda_1}\right] p_\lambda(\underline{x}) \tag{5.17}$$

in terms of integrals of $|Du|^2$.

To discuss these simultaneously, it is convenient to introduce the unimodal functions $Q_j\colon \mathbb{R}_+\to\mathbb{R}_+$ as follows. The first column applies for all $j$ in $I^{(1)}$ and for $j\geqq 2$ in $I^{(2)}$ and the second column for $j=1$ in $I^{(2)}$. For $\lambda\geqq 1$,

$$Q_j(x)=\begin{cases} p_\lambda(\underline{x}), \dfrac{(\underline{x}-\lambda)^2+\underline{x}}{\lambda} p_\lambda(\underline{x}) & \text{if } (x-\lambda)^2\geqq c\lambda, \\ p_\lambda(\underline{\lambda}), 6p_\lambda(\underline{\lambda}) & \text{if } (x-\lambda)^2<c\lambda. \end{cases} \tag{5.18}$$

while for $\lambda<1$,

$$Q_j(x)=\begin{cases} p_\lambda(\underline{x}), (\underline{x}+2)\,p_\lambda(\underline{x}-1) & \text{if } x\geqq 2, \\ p_\lambda(\underline{\lambda}), 4p_\lambda(0) & \text{if } x<2, \end{cases}$$

*except* that in both cases, we set $Q_j(x)=0$ whenever $x\in\Pi_j S_1$ ($\Pi_j$ is projection on the $j^{\text{th}}$ coordinate axis). Note that in (5.18), $c>2$ is chosen to ensure the desired unimodality. Setting $Q(x)=\Pi_j Q_j(x_j)$, it follows that $I^{(1)}$, and $I^{(2)}$ now are both bounded by expressions of the form

$$\int[\bar{u}(x)-\bar{u}(\lambda)]^2 Q(x)\,dx,$$

and hence are primed for application of Corollary 5.2.

Let $r_j(\lambda)=(\lambda\vee 1)^{1/2}$ for each $j$. The goal now is to obtain bounds for quantities on the right side of (5.6) in Corollary 5.2. The claim is that

$$L_j^{(k)}(x)=\int_{B_x}\left|\frac{s-\lambda}{(\lambda\vee 1)^{1/2}}\right|^b dQ_j(s)$$
$$\leqq C[p_\lambda^{(j,k)}(x-c_1\sqrt{x})+p_\lambda(x+c_1\sqrt{x})], \qquad x,\lambda\in\mathbb{R}^+. \tag{5.19}$$

for some constant $c_1>1$. Here $p_\lambda^{(j,k)}(z)=p_\lambda(z\vee I(j=k=1))$ and the superscript $k$ refers to $I^{(k)}$. *Note:* Throughout Sect. 5, we make the convention that a negative argument $z$ causes $p_\lambda(z)$ to be replaced by $p_\lambda(0)$.

The proof of (5.19) necessarily splits into cases dictated by the various forms of $B_x$ given at Remark 5.3. These are sketched in Appendix 2. For $\lambda\geqq 1$, the basic technique is to express $Q_j'(s)$ in terms of a polynomial in $(s-\lambda)/\lambda^{1/2}$ multiplied by $p_\lambda'(s)$ and then apply Proposition 5.5. For $\lambda<1$, the Poisson tails decrease "super-exponentially", so that the desired bounds may be obtained directly, in combination with Lemma I.A.1.

5.6.

From (5.19), Corollary 5.2 and Remark 5.4, it follows that the $i=1$ term in (5.16) is bounded by

$$C\int dP(\lambda)\sum_j \int_{\mathbb{R}_+^p} dx[D_j u(x)]^2(\lambda_j \vee 1)\sum_\sigma p_\lambda(x+\sigma c'\sqrt{x})$$
$$+\lambda_1^{-1}(\lambda_1-d_1(\lambda_1))^2\bar{p}_\lambda(x+\sigma c'\sqrt{x}), \tag{5.20}$$

where $\sigma$ ranges over $\{-1,1\}^p$, and $\bar{p}_{\lambda_j}(z)=p_\lambda(z\vee I\{i=1\})$ for $z\in\mathbb{R}_+$. Since $d_1$ is Lipschitz, $(\lambda_1-d_1(\lambda_1))^2\leqq M[(\lambda_1-x_1)^2+(d_1(x_1)-x_1)^2]$. It follows from Proposition 5.5 and the definition of $\bar{p}_{\lambda_1}$ that by increasing $c$, (5.20) may be bounded by

$$C\int dx\sum_j [D_j u(x)]^2\left\{1+\frac{(d_1(x_1)-x_1)^2}{x_1\vee 1}\right\}\cdot\int dP(\lambda)(\lambda_j+1)\sum_\sigma p_\lambda(x+\sigma c\sqrt{x}). \tag{5.21}$$

For definiteness we consider $j=p$. The inner integral equals

$$\sum_\sigma\{\pi(x+c\sigma_p\sqrt{x})+(x_p+1+c\sigma_p\sqrt{x_p})\,\pi(x+e_p+c\sigma\sqrt{x})\}. \tag{5.22}$$

Repeated application of Lemma 5.8 over successive co-ordinate directions bounds (5.22) by $M\pi(x-c\sqrt{x})+M(x_{p+1}-c\sqrt{x_p})\,\pi(x+e_p-c\sqrt{x})$. Insert this into (5.21) and appeal to Lemma 5.9 to obtain the bound (after increasing $c$)

$$C\int dx[D_p u(x)]^2\{\pi(x-c\sqrt{x})+(x_p+1-c\sqrt{x_p})\,\pi(x+e_p-c\sqrt{x})\}. \tag{5.23}$$

It is now convenient to discretize (5.23): at the expense of increasing $C$ (by Lemma 5.8), replace each term $x_i-c(x_i\vee 1)^{1/2}$ (when positive) by the smaller integer $\{[x_i]+[-c([x_j]\vee 1)^{1/2}]\}\vee 0$, although we will not show this explicitly in the notation. For assumption A(b), it follows that $\pi(x)+(x_p+1)\,\pi(x+e_p)\leqq M\bar{a}_p(x)$, where $\bar{a}_p(x)=a_p(x',x_p\vee 1)$. This bounds (5.23) by

$$C\int dx[D_p u(x)]^2\,\bar{a}_p(x-c\sqrt{x}). \tag{5.24}$$

Since $u$ is a linear interpolation of a lattice function, $D_p u(x)$ is a convex combination of forward differences $D_p^+ u([x]+\tau_{(p)})$ where the entries of $\tau_{(p)}=\sum_1^{p-1}\tau_i e_i$ each range in $\{0,1\}$. Here $D_p^+ u(x)=u(x+e_p)-u(x)$, while $D_p^- u(x)=u(x)-u(x-e_p)$. Since $b_p(x)=\bar{a}_p(x-c'\sqrt{x})$ is constant on lattice squares we may estimate (5.24) by

$$C\sum_{x\in\mathbb{Z}_p^+}\sum_{\tau(p)}[D_p^+ u(x+\tau_{(p)})]^2 b_p(x)=C\sum_{x\in T_p}(D_p^- u(x))^2\sum_{\tau(p)} b_p(x-\tau_{(p)}-e_p),$$
$$\leqq C\sum_{x\in T_p}(D_p^- u(x))^2 b_p(x-1),$$

by Lemma 5.8 and Assumption A, where $1=(1,\ldots,1)$ and $b_p(x)$ is extended to $b_p(x\vee 0)$ if $x_i<0$ for any $i$.

Repeating the argument from (5.22) onwards for each $i=1,\dots,p$, we find that (5.21) is dominated by

$$C\sum_i \sum_{x\in T_i} (D_i^- u(x))^2 \bar{a}_i(g(x)), \tag{5.25}$$

where $g(x)=(g_1(x_1),\dots,g_p(x_p))$ and $g_1(x)=\{x-1+[-c'((x-1)\vee 1)^{1/2}]\}\vee 0$.

In conclusion, we apply Lemma 3.10 and (3.2) to $g(x)$ to conclude that (5.25) can be made arbitrarily small under the recurrence hypothesis. It follows that $d_P$ is admissible. □

## 5.7. *Boundary-positive Estimators*

We discuss briefly the modifications needed when assumption (1.1) fails by virtue of $d_i(x)$ being positive for some $x$ with $x_i=0$. Such estimators can arise as Bayes estimates for certain conjugate priors, or for priors the convex hull of whose support does not intersect $\{\lambda: \lambda_i=0\}$. The results will be applied to the admissibility classification of linear estimators in the next section. For simplicity we write $H_i=\{x: x_i=0\}$ and consider only rules satisfying

$$d_i(x)\not\equiv 0 \quad \text{on} \quad H_i \Rightarrow \inf_{H_i} d_i(x)>0. \tag{5.26}$$

For definiteness, renumber the coordinates so that $d_1(0)=\dots=d_J(0)>0=d_{J+1}(0)=\dots=d_P(0)$. Replace (1.1) by

$$d_i(x)=0 \Rightarrow x_i=0. \tag{5.27}$$

Theorem 1.2 remains valid if $a_{i,x}$ is defined as in the following paragraph. If in Assumption A, condition b) is enforced only for $i\geqq J$, then Theorem 1.3 remains valid also.

In outline, the proofs are modified as follows. Consider first the effect on the quotient representation of Proposition 1.1. In the proof of Proposition 1.1, we may take $\xi_{1n}(d\lambda)$ as the probability measure proportional to $e^{-A}\lambda_1^{-1}dQ_n(\lambda)$ and deduce weak convergence to a p.m. $\xi_1(d\lambda)$ as before. Defining now $w_1(d\lambda)=e^{A}\xi_1(d\lambda)$ and $p_{1,x}=\int e^{-A}\lambda^x\lambda_1 w_1(d\lambda)$, one obtains the representation $d_1(x)=p_{1,x}/p_{1,x-e_1}$ for *all* $x$ in $\mathbb{Z}_+^p$. For other $i\leqq J$, one can proceed analogously to obtain $w_i(d\lambda)$ and $p_{i,x}$, furthermore, the family $\{w_i: i\leqq J\}$ is *compatible* in the sense of Brown and Farrell (1985 a): for $x\in\mathbb{Z}_+^p$, $\lambda^{x+e_i}w_i(d\lambda)=\lambda^{x+e_j}w_j(d\lambda)$. For $i>J$, we may use the representation $d_i(x)=p_{1,x}/p_{1,x-e_i}$ whenever $d_i(x)>0$ (i.e. if $x\notin H_i$). The coefficients $a_{i,x}$ are now defined using $p_{i,x}$.

Theorem 1.2 is proved as before. For the analog of Theorem 1.3, we suppose that $d$ satisfies (5.26) and has quotient representations for measures $w_i$, $i\leqq J$ as described above. The main change in the proof is forced by the need in (5.13) to sum over all $x\in\mathbb{Z}_+^p$ rather than $T_i$ for $i\leqq J$: $\mathscr{D}_0$ is modified accordingly and throughout the proof, $dP(\lambda)$ is replaced by $\lambda_i w_i(d\lambda)$ in the $i^{\text{th}}$ inner sum. The argument then proceeds as before with the addition of appropriate bounds based on (5.22) to handle the extra sums over $H_i$.

## § 6. Applications

This section provides some simpler sufficient conditions for (in)admissibility and discusses applications of the main results. Lemma 6.1 is a particular form of the Nash-Williams test for recurrence, (Griffeath-Liggett 1982, Lyons 1983) which is appropriate for estimators with approximate simplicial symmetry.

This is then combined with the main theorems 1.2 and 1.3 to derive simple comparison tests for (in)admissibility (Proposition 6.2, Corollary 6.3). Corollary 6.3 is used to settle a number of conjectures and to give an admissibility classification for linear estimators. We go on to discuss the use of the Royden-Lyons transience test in deriving some of the results, and mention a connection with unpublished work of Charles Stein.

In the Nash-Williams terminology, the next lemma lumps together all sites in the set $A_n=\left\{x\colon \sum_{i=1}^{p} x_i=n\right\}$, $n\in\mathbb{Z}_+$. In what follows, $z=|x|=\sum_1^p x_i$.

**6.1. Lemma.** *Let* $\bar{a}(z)=\sum_{|x|=z}\sum_i a_i(x)$, *and assume conditions A.*

$$\text{If } \sum^{\infty} 1/\bar{a}(z)=\infty, \text{ then } d_P \text{ is admissible.} \tag{6.1}$$

*Proof.* Let $u_m(x)=c_m\sum_{|x|}^{m-1} 1/\bar{a}(z+1)$, where $c_m=\left[\sum_1^m 1/\bar{a}(z)\right]^{-1}$. Now

$$\sum_i \sum_{x\colon x_i\geqq 1} a_i(x)[D_i u_m(x)]^2=c_m\to 0$$

as $m\to\infty$, and $u_m\in\mathscr{U}$, so $\min_{\mathscr{U}}\|u\|^2=0$ and $d_P$ is admissible by Theorem 1.3. □

Of course, if $P(d\lambda)$ is exactly simplex symmetric, the condition (6.1) reduces to that of Corollary 4.2, but is then less general because of the conditions A needed for Theorem 1.3. We now apply Lemma 6.1 to derive a comparison test analogous to Corollary 4.3 which applies in situations of approximate simplicial symmetry and also in certain quite asymmetric cases (see the discussion of linear estimators below). For an arbitrary function

$$\varepsilon\colon \mathbb{Z}_+\to\mathbb{R}^+, \quad \text{let} \quad E(z)=\varepsilon(z+1)\prod_{s=1}^{z}\varepsilon(s)/s \quad \text{for } z\geqq 0.$$

**6.2. Proposition.** (i) *If there exists* $\eta>0$ *and a function* $\varepsilon(z)>0$ *such that* $d_i(x)\geqq\eta x_i$ *and* $\sum_i d_i(x)\geqq\frac{\varepsilon(z)\,z}{z+p-1}$, *with* $\sum_z 1/E(z)<\infty$, *then d is inadmissible.*

(ii) *If d is generalized Bayes, assumptions A hold and* $\sum_i d_i(x+e_i)\leqq\varepsilon(z+1)$, *with* $\sum_z 1/E(z)=\infty$, *then d is admissible.*

*Proof.* (i) For each $x$, $\pi(x)/\pi(x-e_i) \geqq \varepsilon(z)/z+p-1$ for some $i$. Thus $\pi(x) \geqq c_0\, g(z)$ and $a_i(x) \geqq c_0\, \eta\, x_i\, g(z)$, where $g(z) = \Pi_1^z \frac{\varepsilon(s)}{s+p-1}$. Thus

$$\sum_i \sum_x a_i(x)\,[D_i\, u(x)]^2 \geqq c \sum_i \sum_x g(z)\, x_i [D_i\, u(x)]^2. \tag{6.2}$$

As argued in §4, the infimum of the right side of (6.2) over $\mathscr{U}$ equals the infimum over $\mathscr{V}$ of

$$c \sum_{z=1}^{\infty} \binom{z+k-1}{k-1} z\, g(z) (Dv(z))^2 = c \sum_{z=1}^{\infty} E(z-1)(Dv(z))^2.$$

The result now follows from Theorems 1.2 and I.1.3.

(ii) To apply Lemma 6.1, note from the monotonicity of $d_i$ that

$$\bar{a}(z) = \sum_{|x|=z} \sum_i a_i(x) = \sum_{|x|=z} \pi(x) \sum_i d_i(x) \leqq \varepsilon(z+1) \sum_{|x|=z} \pi(x).$$

Use the lattice structure of the positive orthant to find

$$\sum_{|x|=z-1} \pi(x) \sum_i d_{P,i}(x+e_i) = \sum_{|x|=z-1} \sum_i \pi(x+e_i)(x_i+1) = z \sum_{|x|=z} \pi(x).$$

Iteration together with the hypothesis yields $\bar{a}(z) \leqq \pi(0)\, E(z)$, which suffices. □

*Remark.* A variant of the argument for (i) applies in case $d(x)$ satisfies

$$\left|\frac{d_i(x)}{x_i} - 1\right| \in O(1/z) \quad \text{as} \quad z \to \infty, \tag{6.3}$$

which obviates the need to study the product $E(z)$. If $x$ and $x'$ are points with $\sum x_i = \sum x_i' = z$, then there exists a path from $x$ to $x'$ of length at most $pz$ steps lying wholly within $\{x: z-1 \leqq \sum x_i \leqq z+1\}$. If (6.3) holds, it is easy to check that $[a_i(x)/x_i]/[a_j(x')/x_j']$ is uniformly bounded. Now if $c(z)$ is any selection of the multifunction $z \to \{a_i(x)/x_i;\ x \text{ s.t. } \sum x_i = z,\ i=1,\dots,p\}$, then by arguing as after (6.2), we conclude that $d(x)$ is inadmissible if

$$\sum_z 1/(z^p c(z)) < \infty. \tag{6.4}$$

**6.3. Corollary.** 1) *If $d_i(x) \geqq \eta x_i$ for some $\eta > 0$ and $\sum_i d_i(x) \geqq z-(p-1)+\delta$, for large $z$ and some $\delta > 0$, then $d(x)$ is inadmissible.*

2) *If $|d_i(x)/x_i - 1| \in O(1/z)$ and $\sum d_i(x+e_i) \geqq z+1+\delta$ for large $z$, then $d(x)$ is inadmissible.*

3) *If $d$ is generalized Bayes, assumptions A hold and $\sum_i d_i(x+e_i) \leqq z+1$ for large $z$, then $d$ is admissible.*

*Proof.* The proof of 3) of the Corollary is immediate from Proposition 6.2. To prove 1), note that for large $z$ and some $\delta' \in (0, \delta)$, $\varepsilon(z)/z \geqq (1+\delta'/z)$, and hence that $E(z) \geqq Cz\Pi_1^z(1+\delta'/w)$. Consequently, for some $\varepsilon > 0$, $E(z) \geqq Cz^{1+\varepsilon}$ for $z$ large, and this yields 1). The same argument, in conjunction with (6.4) establishes (2). □

*6.4. Remarks.* 1) All the conjectures of Brown (1979) relating to the simultaneous estimation of Poisson means for the present loss function may be easily verified using the Corollary.

2) Hwang (1982) calls an estimator $\delta^0$ a Semi Tail Upper Bound (STUB) on the class of admissible estimators in direction $\mathbf{1}=(1,\ldots,1)$ for the loss function $L_{-1}$ if every estimator satisfying

$$\sum_i d_i(x+e_i) \geqq \sum_i \delta_i^0(x+e_i)$$

for all large $x$ is inadmissible. He shows that estimators of the form considered by Clevenson and Zidek (1975): $\delta_\ell^{\mathrm{CZ}}(x)=[1-\ell/(p-1+z)]^+ x$, are such STUB's for $\ell<p-1$. Since $\sum_i \delta^{\mathrm{CZ}}(x+e_i)=z+1+\delta$ where $\delta=p-1-\ell$, statements 1) and 2) of Corollary 6.3 essentially say that $\delta_\ell^{\mathrm{CZ}}$ are STUB's. Hwang also conjectures (§4) that the STUB's approach a "dividing line" between admissibility and inadmissibility. This conjecture is established for the $\delta^{\mathrm{CZ}}$ STUB's by 3) of the Corollary.

*Linear Estimators.* We shall apply the preceding results to identify admissible linear estimators of the form $d(x)=Mx+\gamma$ under $L_{-1}$. As Brown and Farrell (1985 b) have given an exhaustive discussion, we shall for simplicity restrict attention to $M$ nonsingular and $\gamma>-1$, the latter condition being imposed by our restrictions (1.1) and (5.26).

As is seen in §2, admissible estimators are necessarily pointwise limits of Bayes estimators. In view of the representation of Proposition 1.1, admissible estimators must therefore satisfy

$$d_i(x+e_i+e_j)\,d_j(x+e_j)=d_j(x+e_i+e_j)\,d_i(x+e_i) \qquad \forall i,j;\ x\in\mathbb{Z}^+.$$

For linear estimators this forces $M^2=DM$, where $D$ is a diagonal matrix with entries $d_i=\sum_j m_{ji}$. So if $M$ is non-singular, it must be diagonal.

We therefore need only consider estimators of the form $d_i(x)=c_i(x+\alpha_i)$, with $c_i>0$. If any $\alpha_i<0$, then $d_i(x)$ is replaced by $d_i^+(x)=d_i(x)\vee 0$. These estimators are then (generalized) Bayes with respect to the conjugate priors $P(d\lambda)=\prod_i \lambda_i^{\alpha_i c_i} e^{-\lambda_i(1-c_i)/c_i} d\lambda_i$. Assume that the indices have been permuted so that $c_1\geqq c_2\geqq\ldots\geqq c_p>0$. If there exists a subset $I$ of indices such that the estimate formed from the components $d_i(x_i)$, $i\in I$ is inadmissible in the $|I|$-dimensional problem, then $d$ is inadmissible in the original problem (Remark 3.6). It follows then from Corollary I.5.2 that $d$ is inadmissible if $c_1>1$. Conversely, Remark 2.2 permits us to ignore components with $c_i<1$, since these are proper Bayes rules in the component problems. Now suppose that $c_1=\ldots=c_J=1$ and that $\alpha_1\geqq\alpha_2\geqq\ldots\geqq\alpha_J$. If $\alpha_J>0$, then inadmissiblity again follows from I.5.2. More generally Corollary 6.3 forces inadmissibility whenever $\sum_{i=1}^J \alpha_i>1-J$. Conversely, admissibility of $d(x)$ follows from Corollary 6.3 if $\sum_{i=1}^J \alpha_i\leqq 1-J$.

In this case one can also use the Nash-Williams criterion for recurrence of a reversible Markov chain (see for example Lyons (1983)) if $c_1 \leqq 1$, $c_2 < 1$; by taking $\Lambda_k = \{x: x_1 = k\}$ in Lyon's setup.

*Connections to Tests of Recurrence/Transience.* 1. It was mentioned earlier in this section that the admissibility test (6.1) could be viewed as a consequence of the Nash-Williams test for recurrence. A test (in fact a characterization) of transience for reversible Markov chains is given by Lyons (1983) in terms of flows. We show here that this leads to the inadmissiblity test of Corollary 4.2 in the simplex symmetric case, and more generally to a converse of Lemma 6.1.

Adopting Lyons' model, regard the lattice points of $\mathbb{Z}_+^p$ as being connected by tubes parallel to the coordinate axes, of length one and cross sectional area $a_i(x)$ for the pipe from $x$ to $x - e_i$. A flow on $\mathbb{Z}_+^p$ (emanating from 0) is a sequence $\{u_{xy},\ x, y \in \mathbb{Z}_+^p\}$ such that $u_{xy} = -u_{xy}$, $\sum_y u_{0y} \neq 0$, $\sum_y u_{xy} = 0$ if $x \neq 0$ and $u_{xy} = 0$ if $x$ and $y$ are not neighbors in $\mathbb{Z}_+^p$. Having in mind the interpretation of $u_{xy}$ as a volume rate of flow, define the (kinetic) energy of the flow by $\sum_{x,y} u_{x,y}^2/\alpha_{x,y}$, where $\alpha_{x,y}$ is defined from $\alpha_i(x)$ as in Sect. 3. The Royden-Lyons test states that the (discrete time) chain $X_n$ associated with $\alpha_{xy}$ is transient if (and only if) there exists a flow of finite energy.

Define a flow in the following concrete way. Suppose that 1 unit of fluid is introduced per second at 0. In each second, all fluid at a node $x$ is redistributed amongst nodes $x + e_i$, $i = 1, \ldots, p$ in proportion to the cross-sectional areas $\alpha_{x, x+e_i}$. Thus if $v(x)$ denotes the volume of fluid passing through $x$ per second, then $v$ satisfies

$$v(x) = \sum_i \left(\alpha_{x-e_i, x} \Big/ \sum_j \alpha_{x-e_i, x-e_i+e_j}\right) v(x - e_i). \tag{6.5}$$

The volume of fluid passing from $x$ to $x + e_i$ each second is

$$u_{x, x+e_i} = \left(\alpha_{x, x+e_i}\right) \Big/ \sum_j \alpha_{x, x+e_j}\Big) v_x,$$

and it is immediate from (6.5) that $u_{xy}$ defines a flow on $\mathbb{Z}_+^p$. Note that since the $\alpha_{xy}$ are known and $v_0 = 1$, (6.5) can be solved recursively and $u_{xy}$ and its energy explicitly evaluated. This approach is perhaps conceptually simpler than that of solving the boundary value problem $\mathscr{P}$ of Theorem 1.4.

Suppose temporarily that $a_j(x)$ is derived from a simplex symmetric prior. Then from (4.1) $a_i(x) = c(z) x_i$, where $c(z) = b(z)(z-1)!/(z+p-1)!$ and $b(z)$ is defined below (4.5), and (6.5) simplifies to

$$v(x) = \sum_i \frac{x_i}{z+p-1} v(x - e_i), \qquad x \in \mathbb{Z}_+^p - \{0\},$$

which is clearly solved by $v(x) = (p-1)!/(z+1)\ldots(z+p-1)$. The corresponding flow has $u_{x, x+e_i} = (x_i + 1) \Big/ \left[(z+1)\binom{z+p}{p-1}\right]$, and has energy

$$\sum_{x,y} \frac{u_{xy}^2}{\alpha_{xy}} = 2 \sum_{z=1}^{\infty} \sum_{|x|=z} \sum_i \frac{u_{x,x-e_i}^2}{a_i(x)} = 2(p-1)! \sum_z 1/b_z.$$

Thus, Lyons' test yields the same criterion for transience as did Theorem 4.1.

The flow $u_{xy}$ constructed above does not depend on the particular simplex symmetric coefficients $a_i(x)$ used in its derivation. It might therefore be used for more general $a_i(x)$: indeed it leads to an alternate proof of (1) of Corollary 6.3. The energy may be written as

$$\sum_z \operatorname*{ave}_{|x|=z} \sum_i x_i^2 \Big/ \left[ z^2 \binom{z+p-1}{p-1} a_i(x) \right],$$

which is bounded by $c \sum_z z^{-p} \operatorname*{ave}_{|x|=z} 1/\pi(x)$, since $d_i(x) \geqq \eta x_i$. Now $d(x) \cdot \mathbf{1} \geqq z - (p-1-\delta)$ implies that $\pi(x) \geqq \prod_{r=1}^{z} \left[ 1 - \left( \frac{p-1-\delta}{r} \right) \right] \geqq c z^{-(p-1-\delta')}$ for some $\delta' \in (0, \delta)$, which entails finiteness of the energy.

2. In unpublished notes from 1964, C. Stein used an abstract form of his necessary and sufficient condition for admissiblity (1955) to derive a characterization of recurrence for irreducible Markov chains on countable state spaces which in the reversible case reduces quite directly to the Griffeath-Liggett characterization. Thus our recurrence/transience tests for admissibility in a concrete statistical problem can be viewed as consequences of Stein's abstract characterization of admissibility.

## Appendix A.1

*Proof of Proposition* 5.5. Suppose first that $\lambda < 1$, and $k \geqq 1$. Writing $L$ for the left side of (5.7), (5.8), it is clear that $L \leqq M p_\lambda(0)$ if $x + \delta \leqq k + 1$. If $x + \delta > k + 1$, then

$$L \leqq \frac{x}{x+\delta} \prod_1^{k-1} \frac{x}{x+\delta-i} \lambda^{k-1} p_\lambda(x+\delta-k) \leqq M_c p_\lambda(x - c'\sqrt{x}).$$

Similar reasoning applies if $k = 0$.

Suppose now that $\lambda \geqq 1$ and fix $c'$. The numbers $x - c\sqrt{x}, x, x + c(x \vee 1)^{1/2}$ partition $\mathbb{R}^+$ into four (possibly degenerate) intervals, and we consider these in turn.

If $x - c\sqrt{x} < \lambda \leqq x$, simple algebra shows that $(x-\lambda)/\sqrt{\lambda} \leqq c^2/2 + (c^2 + c^4/4)^{1/2}$, and that $\lambda \leqq x + c'\sqrt{x} \leqq \lambda + M_{c,c'} \sqrt{\lambda}$. Note from Proposition I.7.1 that

$$\frac{p_\lambda(\lambda)}{p_\lambda(\lambda + m\sqrt{\lambda})} \leqq \left( \frac{\lambda + m\sqrt{\lambda}}{\lambda} \right)^{m\sqrt{\lambda}} \leqq e^{m^2},$$

so that

$$\left( \frac{x-\lambda}{\sqrt{\lambda}} \right)^k p_\lambda(x+\delta) \leqq M_{c,k} \frac{p_\lambda(\lambda)}{p_\lambda(\lambda + m_{c,c'}\sqrt{\lambda})} p_\lambda(\lambda + m_{c,c'}\sqrt{\lambda}) \leqq M_{c,c'} p_\lambda(x + c'\sqrt{x}).$$

If $x<\lambda<x+c\sqrt{x\vee 1}$, and $x>1$, then the argument runs parallel to the previous paragraph. If $x\leqq 1$, then

$$\left(\frac{x-\lambda}{\sqrt{\lambda}}\right)^k p_\lambda(x+\delta)\leqq c^k \frac{p_\lambda(\lambda)}{p_\lambda(\lambda+c')} p_\lambda(\lambda+c')\leqq c^k(1+c')^{c'} p_\lambda(x+c'\sqrt{x\vee 1}),$$

since $\lambda<x+c'\sqrt{x\vee 1}<\lambda+c'$.

If $\lambda\leqq x-c\sqrt{x}$, consider first the case in which $x\leqq 2c'\sqrt{x}$. Then $x\leqq 4c'^2$ and it is easy to check directly that $L\leqq M_{c,c',k}p_\lambda(x-c'\sqrt{x})$. If $x>2c'\sqrt{x}$, introduce $z=|x-\lambda|/\sqrt{\lambda}$, and note that for some $\varepsilon>0$

$$L/p_\lambda(x-c'\sqrt{x})=\left|\frac{x-\lambda}{\sqrt{\lambda}}\right|^k\left(\frac{\lambda}{x}\,\frac{x}{x-c'\sqrt{x}}\right)^{(c'-c)\sqrt{x}}\leqq e^{\varepsilon(c'-c)c'} z^k\left(1+\frac{z}{\sqrt{\lambda}}\right)^{-(c'-c)\sqrt{x}}.$$

Taking logarithms and appropriate derivatives shows that for $z\geqq c$, $\lambda\geqq 1$,

$$k\log z-(c'-c)(\lambda+z\sqrt{\lambda})^{1/2}\log\left(1+\frac{z}{\sqrt{\lambda}}\right)$$
$$\leqq k\log z-(c'-c)\sqrt{1+z}\log(1+z)\leqq M_{c,c',k}.$$

Finally, assume $x+c\sqrt{x\vee 1}<\lambda$. For $x>1$ we have

$$L/p_\lambda(x+c'\sqrt{x})\leqq\left|\frac{x-\lambda}{\sqrt{\lambda}}\right|^k\left[\frac{x}{\lambda}\left(1+\frac{c'}{\sqrt{x}}\right)\right]^{(c'-c)\sqrt{x}}. \tag{A.1}$$

Now if $1\leqq x\leqq\lambda/2$ and $c'-c>k/2$, this is bounded by

$$M_{c,c'}\left|1-\frac{x}{\lambda}\right|^k\lambda^{k/2}\left(\frac{x}{\lambda}\right)^{k/2}\left(\frac{x}{\lambda}\right)^{(c'-c)\sqrt{x}-\frac{k}{2}}$$
$$\leqq M_{c,c'}x^{1/2}2^{-(c'-c)\sqrt{x}+\frac{k}{2}}\leqq M_{c,c',k}$$

while if $\lambda/2<x\leqq\lambda$, and $z=(x-\lambda)/\sqrt{\lambda}$, (A.1) is bounded by

$$M_c|z|^k(1-|z|/\sqrt{\lambda})^{(c'-c)\sqrt{\lambda}}\leqq M_c|z|^k e^{-\varepsilon|z|}\leqq M_{c,c',k}.$$

If $x\leqq 1$, $\left|\frac{x-\lambda}{\sqrt{\lambda}}\right|^k\frac{p_\lambda(x+c)}{p_\lambda(x+c')}=\left|\frac{x-\lambda}{\sqrt{\lambda}}\right|^k\left(\frac{x+c'}{\lambda}\right)^{c'-c}\leqq M_{c,c',k}$, if $c'-c\geqq k/2$.

## Appendix A.2

*Proof of* (5.19).

i) $\lambda\geqq 1$, $x>\lambda+\lambda^{1/2}$.

Here $B_x = [t_x, \infty)$. Suppose first that $t_x > \lambda + c\sqrt{\lambda}$, so that

$$L_j(x) \leqq \int_{t_x}^{\infty} \left|\frac{s-\lambda}{\sqrt{\lambda}}\right|^b (-\bar{Q}_i'(s))\, ds,$$

where for $s \in \mathbb{R}^+ \setminus \mathbb{Z}^+$, $\bar{Q}_j(s)$ is defined from (5.19) by replacing $\underline{x}$ by $s$, and using for $p_\lambda(s)$ the form $e^{-\lambda}\lambda^s/\Gamma(s+1)$. Differentiation shows that for $s > \lambda + c\lambda^{1/2}$, $\bar{Q}_1'(s) \geqq \lambda^{-1}[(s-\lambda)^2 + s]\left[\frac{d}{ds} p_\lambda(s)\right]$, so that for all $i$ it suffices to bound expressions of the form

$$\int_{t_x}^{\infty} \left|\frac{s-\lambda}{\sqrt{\lambda}}\right|^b \left(-\frac{d}{ds} p_\lambda(s)\right) ds \leqq M_{b,c}\, p_\lambda\left(t_x - \frac{c}{2}\sqrt{\lambda}\right),$$

(Corollary 5.7). From Remark 5.3, $t_x - \frac{c}{2}\sqrt{\lambda} \geqq x + \frac{1}{2} - (x+\frac{1}{4})^{1/2} - \frac{c}{2}\sqrt{x} \geqq x - c'\sqrt{x}$. Thus if $x - c'\sqrt{x} \geqq \lambda$, then $L_j(x) \leqq M p_\lambda(x - c'\sqrt{x})$, while if $\lambda - c'\sqrt{\lambda} \leqq x - c'\sqrt{x} < \lambda$, then $L_j(x) \leqq M p_\lambda(\lambda) \leqq M p_\lambda^{(j)}(x - c'\sqrt{x})$, where we have used Proposition 5.5.

Suppose now that $x > \lambda + \sqrt{\lambda}$, but $t_x < \lambda + c\sqrt{\lambda}$. If $0 < t_x = x + \frac{1}{2} - (x+\frac{1}{4})^{1/2}$, then $t_x > \lambda - c\sqrt{\lambda}$. In this case, from Definition (5.18)

$$\begin{aligned} L_j(x) &\leqq \int_{\lambda + c\sqrt{\lambda}}^{\infty} \left|\frac{s-\lambda}{\sqrt{\lambda}}\right|^b [-\bar{Q}'(s)]\, ds + 6c^b p_\lambda(\lambda) \\ &\leqq M\left[p_\lambda\left(\lambda + \frac{c}{2}\sqrt{\lambda}\right) + p_\lambda(\lambda)\right] \leqq M p_\lambda^{(j)}(x - c\sqrt{x}), \end{aligned}$$

since $\lambda < x < \lambda + 3c\sqrt{\lambda}$.

ii) $\lambda \geqq 1$, $x \leqq \lambda - \lambda^{1/2}$.

Here $B_x = [0, t_x] \cap T$. Suppose first that $t_x \leqq \lambda - c\sqrt{\lambda}$ and $x \geqq \lambda/8$. Then as before, by extending $Q_j$ to $\bar{Q}_j$ defined for real $s$, and differentiating, we find

$$L_j(x) \leqq \int_0^{t_x} \left|\frac{\lambda - s}{\sqrt{\lambda}}\right|^b \bar{Q}_j'(s)\, ds,$$

which is in turn bounded by terms of the form

$$\int_0^{t_x} \left|\frac{\lambda - s}{\sqrt{\lambda}}\right|^b \left\{\frac{d}{ds} p_\lambda(s)\right\} ds \leqq p_\lambda\left(t_x + \frac{c}{2}\sqrt{\lambda}\right).$$

Since $x \geqq \lambda/8$, $t_x + \frac{c}{2}\sqrt{\lambda} \leqq x + \frac{1}{2} + (x+\frac{1}{4})^{1/2} + c\sqrt{2x} \leqq x + c''\sqrt{\lambda} \leqq \lambda + c''\sqrt{\lambda}$, so that $p_\lambda\left(t_x + \frac{c}{2}\sqrt{\lambda}\right) \leqq p_\lambda(x + c''\sqrt{x})$. The proof in the case $t_x > \lambda - c\sqrt{x}$ is similar to the corresponding part of (i) and so is omitted.

If $t_x \leqq \lambda - c\sqrt{\lambda}$ and $0 \leqq x < \lambda/8$, a different argument is used, for then $t_x = x + \frac{1}{2} + (x+\frac{1}{4})^{1/2} \leqq 2(x+1) < \lambda/2$, and $L_j(x) \leqq \lambda^{b/2} Q_j(t_x)$, which in turn is bounded by terms of the form $\lambda^\beta p_\lambda(x + 2\sqrt{x \vee 1})$. Now if $c > 2$

$$\lambda^\beta \frac{p_\lambda(x+2\sqrt{x})}{p_\lambda(x+c\sqrt{x})} \leqq \lambda^\beta \frac{(x+c\sqrt{x})^{(c-2)\sqrt{x}}}{\lambda^{(c-2)\sqrt{x}}}$$

$$\leqq x^\beta \left(\frac{x}{\lambda}\right)^{(c-2)\sqrt{x}-\beta} \left(1+\frac{c}{\sqrt{x}}\right)^{(c-2)\sqrt{x}} \leqq \frac{M x^\beta}{8^{(c-2)\sqrt{x}}} \leqq M'.$$

Finally, if $x=0$, $\beta_0 = [0, 1]$, and it is easy to check that $L_j(x) \leqq M p_p(c_0)$ for some positive constant $c_0$.

iii) $\lambda \geqq 1$, $\lambda - \lambda^{1/2} < x \leqq \lambda + \lambda^{1/2}$.

Here $B_x = \mathbb{Z}_+$, but a bound for $L_j(x)$ in terms of $p_\lambda(x + c''\sqrt{x}) + p_\lambda(x - c''\sqrt{x})$ is easily obtained by combining the methods of steps (i) and (ii).

iv) $\lambda < 1, j = 1$.

In what follows $t_x = x + \frac{1}{2} - (x+\frac{1}{4})^{1/2}$, and $S$ denotes the closed convex hull of the support of the measure defined by $Q_1$. From the definition of $Q_1$, it is clear that

$$L_1^{(2)}(x) \leqq 4p_\lambda(0)|\lambda|^b\{0 \in B_x \cap S\} + 4p_\lambda(0)\,|2-\lambda|^b\{2 \in B_x\} + \sum_{s \geqq t_x \vee 3} |s-\lambda|^b (s+1)\, p_\lambda(s-2).$$

If $t_x \leqq b+3$, then the above is trivially bounded by $M_b p_\lambda(0)$. If $t_x > b+3$, it is bounded (using I.A.1) by $M_b p_\lambda(t_x - 4 - b) \leqq M p_\lambda(x - c\sqrt{x})$ for appropriate $c$.

A similar argument is used for $L_1^{(1)}(x)$, except that if $r_1 = 1$, so that $S = [1, \infty)$, then for $t_x \leqq b+3$, a bound by $M_b p_\lambda(1)$ is possible.

The situation for $i \geqq 2$ is entirely similar to that of $L_1^{(1)}(x)$.

## Appendix A.3

*Proof* of (4.6). First,

$$\gamma_p(z) \leqq p\gamma_{p-1}(z) + p \sum_{\substack{|x|=z \\ \min x_i \geqq 1}} x_p / \Pi_1^{p-1} x_i$$

$$\leqq p\gamma_{p-1}(z) + pz\left(\sum_1^z 1/x\right)^{p-1} \leqq C_p z \log^{p-1} z.$$

For the converse,

$$\gamma_p(z) = p \sum_{|x|=z} x_p / \Pi_1^{p-1}(x_j \vee 1)$$

$$\geqq z \sum_{\substack{0 \leqq x_i \leqq z/p \\ 1 \leqq i \leqq p-1}} 1/\Pi_1^{p-1}(x_j \vee 1) \geqq z(\log z/p)^{p-1} \geqq c'_p z \log^{p-1} z. \quad \square$$

*Acknowledgements.* At Cornell, L. Brown was a constant source of insight and encouragement. At M.S.R.I. discussions with M. Shahshahani were a crucial stimulus for Sect. 5.1. Thanks go also to R. Arratia, P. Diaconis, E. Dynkin, R. Farrell, and H. Kesten for helpful conversations and references, to a referee and D. Donoho for suggestions on exposition and to N. Holmes for stoic manuscript preparation.

## References

Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions. New York: Dover 1965

Berger, J.: Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. Ann. Statist. **8**, 545–571 (1980)

Berger, J.: Multivariate Estimation – A Synthesis of Bayesian and Frequentist Approaches. Forthcoming SIAM monograph (1985)

Birnbaum, A.: Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. Ann. Math. Statist. **26**, 21–36 (1955)

Brown, L.D.: Admissible estimators, recurrent diffusions and insoluble boundary value problems. Ann. Math. Statist. **42**, 855–903 (1971). Correction, Ann. Statist. **1**, 594–596 (1973)

Brown, L.D.: Counterexample – An inadmissible estimator which is generalized Bayes for a prior with "light" tails. J. Multivariate Anal. **9**, 332–336 (1979 a)

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

Brown, L.D., Farrell, R.H.: Complete class theorems for estimation of multivariate Poisson means and related problems. Ann. Statist. **13**, 706–726 (1985 a)

Brown, L.D., Farrell, R.H.: All admissible linear estimators of the multivariate Poisson mean vector. Ann. Statist. **13**, 282–294 (1985 b)

Brown, L.D., Farrell, R.H.: Proof of a necessary and sufficient condition for admissibility in discrete multivariate problems. J. Multivarite Anal. To appear (1985 c)

Brown, L.D., Hwang, J.T.: A unified admissibility proof. Proc. 3rd Purdue Symposium Statistical Decision Theory **1**, 205–231 (1982)

Cohen, A., Sackrowitz, H.: Estimation of the last mean of a monotone sequence. Ann. Math. Statist. **41**, 2021–2034 (1970)

Clevenson, M., Zidek, J.: Simultaneous estimation of the mean of independent Poisson laws. J. Am. Stat. Assoc. **70**, 698–705 (1975)

Efron, B., Morris, C.: Stein's estimation rule and its competitors – An empirical Bayes approach. J. Am. Stat. Assoc. **68**, 117–130 (1973)

Farrell, R.H.: Towards a theory of generalized Bayes tests. Ann. Math. Statist. **39**, 1–22 (1968)

Freedman, D.: Markov Chains. San Francisco: Holden-Day 1971

Fukushima, M.: Dirichlet forms and Markov processes. Amsterdam-New York: North-Holland 1980

Ghosh, M., Hwang, J.T., Tsui, K.W.: Construction of improved estimators in multiparameter estimation for discrete exponential families (with discussion). Ann. Statist. **11**, 351–367 (1983)

Ghosh, M., Parsian, A.: Bayes minimax estimation of multiple Poisson parameters. J. Multivarite Anal. **11**, 280–288 (1981)

Griffeath, D., Liggett, T.M.: Critical phenomena for Spitzer's reversible nearest particle systems. Ann. Probab. **10**, 881–895 (1982)

Guttman, S.: Stein's paradox is impossible in a finite sample space. Ann. Statist. **10**, 1017–1020 (1982)

Guttman, S.: Decisions immune to Stein's effect. Sankhya, To appear (1983)

Hwang, J.T.: Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. Ann. Statist. **10**, 857–867 (1982)

Hwang, J.T.: Semi tail upper bounds on the class of admissible estimators with applications to Poisson and negative binomial families. Ann. Statist. **10**, 1137–1147 (1982)

Ichihara, K.: Some global properties of symmetric diffusion processes. Publ. Res. Inst. Math. Sci., Kyoto Univ., **14**, 441–486 (1978)

James, W., Stein, C.: Estimation with quadratic loss. Proc. Fourth Berkeley Sympos. Univ. Calif. Math. Statist. Probab. **1**, 361–379 (1960)

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

Johnstone, I.M.: Admissibility from recurrence via Poincaré inequalities in the Gaussian case. Manuscript notes (1983)

Johnstone, I.M., Lalley, S.P.: On independent statistical decision problems and products of diffusions. Zeitschrift für Wahrscheinlichkeitstheor. Verw. Geb. **68**, 29–47 (1984)

Johnstone, I.M., Shahshahani, M.: A Poincaré-type inequality for solutions of linear elliptic equations, with statistical applications. M.S.R.I. technical report (1983)

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

Lebedev, N.N.: Special Functions and Their Applications. New York: Dover 1972

LeCam, L.: An extension of Wald's theory of statistical decision functions. Ann. Math. Statist. **26**, 69–81 (1955)

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

Morris, C.N.: Parametric empirical Bayes inference: Theory and Applications (with discussion). J. Am. Statist. Assoc., 47–65 (1983)

Peng, J.C.M.: Simultaneous estimation of the parameters of independent Poisson distributions. Technical Report 78, Dept. of Statistics, Stanford University (1975)

Srinivasan, C.: Admissible generalised Bayes estimates and exterior boundary value problems. Sankhya **43**, Ser. A, 1–25 (1981)

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

Stein, C.: The admissibility of Hotelling's $T^2$-test. Ann. Math. Statist. **27**, 616–623 (1956 a)

Stein, C.: Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Sympos. Univ. Calif. Math. Statist. Probability **1**, 197–206 (1956 b)

Stein, C.: Approximation of improper prior measures by prior probability measures. In Bernouilli, Bayes, Laplace Anniversary Volume, pp. 217–241. Berlin Heidelberg New York: Springer 1965

Received November 21, 1983; in revised form February 26, 1985