The Annals of Statistics
1984, Vol. 12, No. 4, 1173–1198

ADMISSIBILITY, DIFFERENCE EQUATIONS AND RECURRENCE IN ESTIMATING A POISSON MEAN[1]

BY IAIN JOHNSTONE

Stanford University and Mathematical Sciences Research Institute, Berkeley

Consider estimation of a Poisson mean λ based on a single observation x, using estimator $d(x)$ and loss function $(d(x) - \lambda)^2/\lambda$. The goal is to decide (in)admissibility of $d(x)$. To every generalized Bayes estimator there corresponds a unique reversible birth and death process $\{X_t\}$ on $\mathbb{Z}_+$. Under side conditions $d(x)$ is admissible if and only if it is generalized Bayes and $\{X_t\}$ is recurrent. Explicit equivalent conditions exist in terms of difference equations and minimization problems. The theory is a discrete, univariate counterpart to Brown's (1971) diffusion characterization of admissibility in estimation of a multivariate normal mean. A companion paper discusses simultaneous estimation of several Poisson means.

1. Introduction. Consider the problem of simultaneously estimating the unknown parameters of several independent discrete distributions such as Poisson or negative binomial. Recent work (discussed and extended in Ghosh et al., 1983), shows that substantial savings in frequentist risk can be attained by constructing alternatives to the usual maximum likelihood (MLE) and minimum variance estimators. This complements insights originally obtained in the multivariate Gaussian problem by Stein, and later Efron and Morris and others.

In comparing risk functions of competing estimators, the qualitative issue of admissibility has proved a useful prelude to the quantitative problem of constructing and evaluating better estimators. Thus Stein (1956) first established the unexpected inadmissibility of the ML estimate of $p \geq 3$ Gaussian means under squared error loss. With James he then showed (1961) that an estimator of simple form could attain risk improvements of up to $(1 - 2/p)100\%$.

Admissibility alone is a weak optimality property, so a finding of inadmissibility against an otherwise plausible estimator is of greater statistical significance. Inadmissibility is often most conveniently established by solving differential inequalities (Stein 1973, 1981, Ghosh et al., 1983), especially as this process explicitly constructs a better estimator. A classical approach, that of explicitly describing all admissible estimators (the minimal complete class), can still be of use if the differential inequality is intractable or a quick qualitative answer is desired. The latter method, in the context of simultaneous estimation of Poisson means, is the primary subject of this paper and its companion. Roughly, to each "potentially admissible" estimator is associated a birth-death Markov chain, and

Received July 1983; revised June 1984.

[1] Research supported by an Australian National University Traveling Scholarship and A. D. White Fellowship at Cornell University and by NSF at Mathematical Sciences Research Institute, Berkeley.

AMS 1980 *subject classifications.* Primary 62C15; secondary 60J10, 62F10.

Key words and phrases. Admissibility, birth and death process, Poisson mean, difference equations, Dirichlet problem, recurrence.

admissibility corresponds exactly to recurrence of the chain. The apparatus of probabilistic potential theory and difference equations (including recent results of Griffith-Liggett, 1982, and Lyons, 1983) is thus available for testing admissibility.

L. Brown (1971) discovered and elaborated this phenomenon in the multivariate Gaussian means problem. The simplest and most striking instance is the association of Brownian motion (B.M.) with the MLE and the interpretation of Stein's result via the transience of B.M. and existence of superharmonic functions on $\mathbb{R}^p$ for $p \geq 3$. His idea, also used in the present paper, was to reformulate the admissibility question for a particular estimator in terms of a minimization problem familiar in the calculus of variations. The chief step, conceptually and technically, is in establishing this equivalence—the translation into differential equations and probability is then accomplished via more familiar methods.

This paper gives an account of the form this theory takes in the simplest discrete context: estimation of a single Poisson mean for a normalized squared error loss function. The restriction to one dimension affords explicit tests for recurrence/admissibility, simplicity of notation and technique, and more complete results. The companion paper (Johnstone, 1984; referred to as II) treats the statistically more interesting but technically harder question of simultaneous estimation of several Poisson means, and presents new, general (in)admissibility results for this problem.

OUTLINE. Suppose that X has a Poisson distribution with mean $\lambda \in (0, \infty)$. The problem is to estimate λ using an estimator $d(x) \in [0, \infty)$ and loss function $L_{-1}(d, \lambda) = (d - \lambda)^2/\lambda$. (Of course, the case of an i.i.d. sample of size n is included, by sufficiency.) For this loss function, the "usual" estimator $d(x) = x$ has constant risk equal to 1 and is minimax. A multiplicative factor λ^{-m} in the loss function has no effect on admissibility considerations in one dimension, but decisively affects admissibility of the MLE in two or more dimensions (Clevenson/Zidek, 1975, Peng, 1975). Clevenson and Zidek argue that the choice $m = +1$ also has some practical appeal—errors of a given magnitude are more heavily penalized when λ is small.

We seek necessary and sufficient conditions for an estimator $d(x)$ to be admissible. A complete class theorem due to Farrell and Brown (1983) provides a reduction of the problem—all admissible estimators are conditionally Bayes (in a sense elaborated in Section 2) and have a basic "logarithmic difference" representation: if $d(x)$ is admissible and its first nonzero value occurs at $x = r + 1 \geq 0$, then there is a finite measure $P(d\lambda)$ on $[0, \infty)$ such that

$$d(x) = d_P(x) = p_x/p_{x-1}, \quad x \geq r + 1, \tag{1.1}$$

where $p_x = \int \lambda^{x-r} P(d\lambda)$. We may therefore restrict attention to such estimators in characterizing admissibility.

When $r = -1$ or $P(\{0\}) = 0$ (the cases of greatest statistical interest), d_P may be regarded as a generalized Bayes estimator with respect to L_{-1} and the prior $\pi(d\lambda) = \lambda^{-r} e^{\lambda} P(d\lambda)$ supported on $(0, \infty)$. That not every admissible estimator is generalized Bayes follows from noncompactness of the parameter space $(0, \infty)$.

The mathematical connection between d_P and its associated birth and death process is obtained in a discrete minimization problem in the calculus of variations. To arrive at the minimization problem, we use a version of the necessary and sufficient condition for admissibility due to Stein, Le Cam and Farrell (see Farrell, 1968, Section 3, or Brown, 1976, Chapter 5). Write $R(\lambda, d) = E_\lambda L(\lambda, d(x))$ and $B(d, P) = \int R(\lambda, d)P(d\lambda)$ for the risk and integrated risk of d respectively, with respect to a loss function L.

Then d is admissible iff the difference $B(d, Q) - B(d_Q, Q)$ can be made arbitrarily small as Q ranges over a certain class $\mathscr{Q}$ of measures with finite total mass. For an appropriate choice of quadratic loss function, and for $d(x) = d_P(x)$ for $x \geq r + 1$, it follows (Section 3) that

$$B(d_P, Q) - B(d_Q, Q) \geq \sum_{r+1}^{\infty} (Du_x)^2 a_x. \tag{1.2}$$

Here $u_x^2 = q_x/p_x$, $Du_x = u_x - u_{x-1}$, and $a_x = p_x^2/p_{x-1}x!$. These relations lie at the root of the connection between admissibility and both difference equations and birth-death processes.

The expression $\sum (Du_x)^2 a_x$ depends on Q only through u, which ranges through a class of functions $\mathscr{U}_{\mathscr{Q}}$ determined by $\mathscr{Q}$. Section 3 shows that $\mathscr{U}_{\mathscr{Q}}$ may be replaced by $\mathscr{U}_r = \{u_x : \mathbb{Z} \to \mathbb{R} : u_r = 1, u_\infty = \lim_{x\to\infty} u_x = 0\}$, which yields a more familiar variational problem with boundary conditions and the following result.

THEOREM 1.1. *If $d(x)$ is admissible then*

$$\inf_{\mathscr{U}_r} \sum_{r+1}^{\infty} (Du_s)^2 a_s = 0. \tag{1.3}$$

By formal analogy with electrical circuits, condition (1.3) has a physical interpretation as a "zero energy" condition. Let pairs of succeeding sites $s - 1$ and s be connected by resistors with conductance a_s, with sites r and n being held at fixed voltages 1 and 0 respectively. As a consequence of Ohm's law, the resultant voltage configuration v_s $(r \leq s \leq n)$ leads to energy dissipation at the rate $\sum_{r+1}^{n} (Dv_s)^2 a_s$. Thomson's principle states that $\{v_s\}$ minimizes $\sum_{r+1}^{n} (Du_s)^2 a_s$ amongst all functions u satisfying $u_r = 1$, $u_n = 0$. Condition (1.3) refers to the (physically unrealizable) system on $\mathbb{Z}_r = \{r, r + 1, \cdots\}$; alternatively it describes the behavior of the minimal energy dissipation rate as $n \to \infty$. These viewpoints are treated more explicitly in II. For historical remarks and further details from a probabilistic perspective we refer to Griffeath-Liggett (1982) and an entertaining monograph by Doyle and Snell (1981).

The main result of this paper is that under two mild statistical conditions, the "zero energy" property (1.3) is *equivalent* to admissibility of d_P. The first condition

$$d_P(x) - x \leq M(1 + \sqrt{x}), \tag{1.4}$$

corresponds to $x + (d_P(x) - x)^+$ having bounded risk, while the second,

$$d_P(x + 1) - d_P(x) \in 0(1), \tag{1.5}$$

is of yet milder character, and amounts to d_P having bounded posterior risk for each x. (Proposition 6.1). In the situation of greatest interest, when d_P has

bounded risk, (1.5) is unnecessary. The conditions correspond to those of Srinivasan (1981) in the normal case, and are discussed further in Section 6.

THEOREM 1.2. *Under* (1.4) *and* (1.5), *if* $\min_{\mathscr{U}_r} \sum_{r+1}^{\infty} (Du_s)^2 a_s = 0$, *then* d_P *is admissible.*

Thus, from the statistical perspective of characterizing admissibility, nothing is lost at the inequality in (1.2) (essentially a linearization of the problem), or in the extension from $\mathscr{U}_\varnothing$ to $\mathscr{U}_r$. A heuristic argument and then the detailed proof are given in Section 5. A necessarily bizarre counterexample is given in Section 7 to show that condition (1.4) cannot be dispensed with.

Condition (1.3) can be translated into checkable probabilistic and analytic criteria for admissibility, given in Theorem 1.3 below. Let $\| u \|^2 = \sum_{r+1}^{\infty} (Du_s)^2 a_s$. The Euler-Lagrange equation of the variational problem $\inf_{\mathscr{U}_r} \| u \|^2 = 0$ gives a necessary condition satisfied by any local optimum, if such exists. Indeed, if $\tilde{u}$ attains the minimum, then $\inf_{\mathscr{U}_r} \| u \|^2 > 0$, and $\tilde{u}$ satisfies the second order difference equation

$$Au_s := D(a_{s+1} Du_{s+1}) = 0, \quad s > r, \quad u_r = 1, \quad u_\infty = 0. \tag{1.6}$$

Associated with the second order difference operator A is an irreducible continuous time Markov chain $(X_t, P^x, x \in \mathbb{Z}_r)$ on $\mathbb{Z}_r$ whose infinitesimal generator is A. Since the difference operator is local—Au_x depends only on u_x, u_{x+1} and u_{x-1}—the corresponding Markov chain is a birth and death process. Hence transitions from x can occur only to $x + 1$ and $x - 1$, at rates a_{x+1} and a_x respectively. These processes are discrete state space analogues of diffusions.

It is a familiar consequence of the Markov property that hitting probabilities of a set B satisfy $Au_s = 0$ for $s \notin B$. In particular, $k_x = P^x(\exists\, t \geq 0 : X_t = r)$ satisfies $Ak_s = 0$ for $s > r$ and $k_r = 1$. The process is recurrent iff $k_x \equiv 1$ for all x, while in the transient case $k_x \to 0$ as $x \to \infty$. Hence k_x solves (1.6) iff $\{X_t\}$ is transient. The results indicated in this discussion are summarized and extended in the next theorem: a short proof is given in Section 4, together with admissibility applications in Sections 4 and 5.

THEOREM 1.3. *The following are equivalent*:

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

(ii) *the boundary value problem* (1.6) *has no solution*,

(iii) $\{X_t\}$ *is recurrent*,

(iv) $\sum_{r+1}^{\infty} 1/a_x = \infty$.

In the transient case, the hitting probability k_x *is the unique solution of* (1.6) *and is the unique function in* $\mathscr{U}_r$ *minimizing* $\| u \|^2$. *It is given by*

$$k_x = [\textstyle\sum_{r+1}^{\infty} 1/a_s]^{-1} \sum_{x+1}^{\infty} 1/a_s, \quad x \geq r.$$

In the recurrent case, the embedded discrete-time chain $\{X_n\}$ *is null recurrent iff* $\sum a_x = \infty$.

Most, if not all, parts of this theorem are well known in various contexts in probability and differential equations (e.g. Fukushima, 1980, Section 4.4; Brown,

1971; Ichihara, 1978). We present the results in this form because of their statistical relevance, simplicity and completeness.

The connection between admissibility and recurrence is thus indirect, going via a variational problem of analysis. Whether the link can be made more direct is an interesting open question. Nevertheless, the Markov process viewpoint already provides useful intuition and constructions, especially in the multiparameter case (cf. Johnstone and Lalley, 1984 and II, Section 3). This is perhaps obscured in our one-dimensional setting by the availability of explicit analytic solutions to (1.6).

As a result, the special condition (iv) above yields the easiest test for admissibility. For example the linear estimators $d_{\alpha,\beta}(x) = (x + \alpha)/(1 + \beta)$ are, at least for large x, generalized Bayes for the conjugate priors $\pi(d\lambda) = \lambda^{\alpha}e^{-\beta\lambda}$. It is easy to check that $a_x \sim Cx^{\alpha+1}/(1 + \beta)^x$ and hence that $d_{\alpha,\beta}(x) \vee 0$ is admissible iff $\beta > 0$, or $\beta = 0$ and $\alpha \leq 0$.

Just as the normal distribution is the prototype of a continuous exponential family from the point of view of the recurrence characterization, the Poisson density is the prototype of discrete exponential families on $\mathbb{Z}_+$. All results stated above extend easily to "power series distributions" except for the admissibility theorem (1.3), which will be studied elsewhere. Hence important distributions such as the negative binomial and logarithmic may also be included in the theory.

It is natural to seek a connection between this approach and that of Hudson (1978), Hwang (1982) and Ghosh et al (1983), who for multiparameter discrete exponential families construct a difference inequality $\mathscr{D}_d(x) \leq 0$ and employ its solutions (if any) to prove inadmissibility of d and exhibit dominating estimators. It is shown in Section 8 that the operator $\mathscr{D}_d$ is precisely the (nonlinear, second order) operator that arises in the Euler-Lagrange equation for the *original* minimization problem defined by (1.1). We conclude that the "linearized" problem (1.3) suffices to resolve the qualitative property of admissibility but not for the quantitative problem of constructing a better estimator.

2. Decision theoretic preliminaries. Since the loss function is strictly convex in d, nonrandomized decision rules $d(x)$: $\mathbb{Z}_+ \to [0, \infty)$ form a complete class. Moreover, procedures with everywhere finite risk functions form a complete class. This is a fairly general fact in one-dimensional estimation problems (Brown, 1976, Theorem V.2.4), but in our case it suffices to observe that if $R(\lambda, d) = \infty$, then $R(\lambda', d) = \infty$ for all $\lambda' \geq \lambda$. If $\lambda_0 = \inf\{\lambda: R(\lambda, d) = \infty\}$, then $d'(x) = \min\{d(x), \lambda_0\}$ has everywhere finite risk and dominates $d(x)$.

Let $\mathscr{D}_r = \{d(x): d(x) = 0 \text{ for } 0 \leq x \leq r\}$ and let d be a nonrandomized rule in $\mathscr{D}_r \setminus \mathscr{D}_{r+1}$. To discuss the admissibility of $d(x)$ in the original problem, $\mathscr{P}$ say, it is convenient to consider a corresponding Poisson decision problem $\mathscr{P}_r$ *conditional* on observing $x \geq r + 1$. Thus the sample space is $\mathbb{Z}_{r+1}$, supporting conditional densities $P_\lambda(X = x \mid X > r) = \phi_r^{-1}(\lambda)\lambda^x/x!$, where of course $\phi_r(\lambda) = \sum_{x>r} \lambda^x/x!$. The parameter space is extended to $[0, \infty)$ and for loss function we take $g_r(\lambda)(d - \lambda)^2$, where the factor $g_r(\lambda)$:$[0, \infty) \to (0, \infty)$, although not affecting admissibility conclusions, will be specified below for technical convenience. The rationale for bringing in the conditional problem and compactifying the param-

eter space is explained by Brown (1981) and Brown and Farrell (1983), who construct complete classes of stepwise Bayes procedures in finite sample space and discrete exponential family settings respectively. Here we note only the Brown-Farrell example: $d(x) = 0, \frac{1}{2}, 1$ according as x is 0, 1 or ≥ 2. This estimator is admissible, but not generalized Bayes for any prior concentrated on $(0, \infty)$. It is however Bayes in $\mathscr{P}_0$ for the prior $P(d\lambda) = \frac{1}{2}\delta_{\{0\}} + \frac{1}{2}\delta_{\{1\}}$.

LEMMA 2.1. *Let $d \in \mathscr{D}_r \backslash \mathscr{D}_{r+1}$. Then d is admissible in $\mathscr{P}$ iff it is admissible in $\mathscr{P}_r$.*

PROOF. Suppose that d' dominates d in $\mathscr{P}_r$. Since, by assumption, d has finite risk in $\mathscr{P}$, d' and d have risk functions in $\mathscr{P}_r$ that are continuous as functions from $[0, \infty) \to [0, \infty]$. Defining $d'(x)$ to be zero for $0 \leq x \leq r$, it is then clear that d' dominates d in $\mathscr{P}$.

Conversely, suppose that d is admissible in $\mathscr{P}_r$. It is enough to show that d is admissible in $\mathscr{P}_{r-1}$, and then iterate. If $\tilde{d} \in \mathscr{D}_{r-1} \backslash \mathscr{D}_r$, then $\tilde{d}_r > 0$. Denote the risk function of a rule d in $\mathscr{P}_r$ by $R_r(d, \lambda)$: clearly

$$R_{r-1}(\tilde{d}, 0) - R_{r-1}(d, 0) = g_{r-1}(0)\tilde{d}_r^2 > 0,$$

so that $\tilde{d}$ cannot dominate d. □

Let $P(d\lambda)$ be a finite measure on $[0, \infty)$ having all moments finite. Consider P as a prior measure on λ in problem $\mathscr{P}_r$: the posterior density is then well defined for $x \geq r + 1$ and is proportional to $\phi_r^{-1}(\lambda)\lambda^x P(d\lambda)$. Choose $g_r(\lambda) = \phi_r(\lambda)\lambda^{-r-1} \in (0, \infty)$ for $\lambda \in [0, \infty)$. Minimizing posterior expected loss to evaluate the Bayes rule requires d to minimize

$$\int g_r(\lambda)(d - \lambda)^2 \phi_r^{-1}(\lambda)\lambda^x \, dP(\lambda) = p_{x-1}\left(d - \frac{p_x}{p_{x-1}}\right)^2 + p_x\left(\frac{p_{x+1}}{p_x} - \frac{p_x}{p_{x-1}}\right),$$

where $p_x = \int \lambda^{x-r} P(d\lambda)$. Thus the Bayes rule $d_P(x) = p_x/p_{x-1}$ for $x \geq r + 1$, and substitution into the previous display yields an identity which is important in Section 5:

$$(2.1) \quad \frac{1}{p_x}\int (\lambda - d_P(x))^2 \lambda^{x-r-1} \, dP(\lambda) = d_P(x + 1) - d_P(x), \quad x \geq r + 1.$$

To make precise the discussion of Section 1, define the integrated risk in $\mathscr{P}_r$ by $B_r(d, P) = \int R_r(d, \lambda)P(d\lambda)$, and note that $B_r(d_P, P) = B_r(P) = \inf_d B_r(d, P)$. For $\lambda_0 \in (0, \infty)$, let $\mathscr{Q}$ be the set of measures Q with finite total mass and compact support in $[0, \infty)$ such that $B(d_Q, Q) < \infty$ and $Q([\lambda_0, \lambda_0 + 1]) \geq 1$. The Stein-LeCam-Farrell theorem says in this context that an estimator $d(x) \in \mathscr{P}_r$ is admissible if and only if

$$\inf_{Q \in \mathscr{Q}} B_r(d, Q) - B_r(d_Q, Q) = 0.$$

Now we record a basic and well-known identity valid for any statistical estimation problem with a loss function of quadratic type. Suppose that Q is a

finite measure with support compact in $(0, \infty)$. For any decision rule $d(x)$ with finite risk, $B_r(d, Q) < \infty$ and

$$B_r(d, Q) - B_r(Q)$$

$$= \sum_{x>r} \frac{1}{x!} \int_0^\infty \{d^2(x) - d_Q^2(x) - 2[d(x) - d_Q(x)]\lambda\}\lambda^{x-r-1}\, dQ(\lambda) \tag{2.2}$$

$$= \sum_{x>r} [d(x) - d_Q(x)]^2 q_{x-1}/x!.$$

Here we have used Fubini's theorem and the definition of $d_Q(x)$.

Brown and Farrell (1983) have studied "Sacks-type" complete class theorems for simultaneous estimation of the means of independent Poisson laws. Their Theorem 5.1 specializes their results to one dimension, which for convenience we repeat here together with a direct proof.

THEOREM 2.2 (Brown, Farrell). *Let $d(x)$ be a nonrandomized decision rule with $r + 1 = \inf\{x \in \mathbb{Z}_+ : d(x) \neq 0\}$. If d is admissible, there is a finite measure $P(d\lambda)$ on $[0, \infty)$ such that $p_x = \int \lambda^{x-r} dP(\lambda) < \infty$ for $x \geq r$, and*

$$d(x) = p_x/p_{x-1} \quad \text{for} \quad x \geq r + 1, \tag{2.3}$$

so that d is Bayes in problem $\mathscr{P}_r$.

REMARK. Note that $d(x)$ is thus monotone nondecreasing in x. This follows from the (strict) monotone likelihood ratio property of the kernel $f(x, \lambda) = \lambda^x$ in λ and x (Lehmann, 1959, page 74).

NOTATION. Let $\mathscr{C}_r = \{d(x)\colon d(x) = 0,\ 0 \leq x \leq r$ and $\exists\ P(d\lambda)$ such that $p_x < \infty$ for $x \geq r$ and (2.3) holds$\}$. Thus the class of admissible rules in $\mathscr{P}$ is contained in $\cup_{r=1}^\infty \mathscr{C}_r$.

PROOF (Farrell). Since any measure $Q \in \mathscr{Q}$ satisfies $Q([\lambda_0, \lambda_0 + 1]) \geq 1$ (see Section 1), it follows that $\inf_{Q \in \mathscr{Q}} q_x > 0$ for each integer $x \geq r + 1$. So if $d(x)$ is admissible, we deduce from (2.2) the existence of a sequence of probability measures $\{Q_n\}$ with compact support in $(0, \infty)$ for which $d_{Q_n}(x) \to d(x)$ for $x \geq r + 1$.

Let $P_n(d\lambda) = Q_n(d\lambda)/Q_n([0, \infty))$. Since

$$\int \lambda^k\, dP_n(\lambda) = \prod_{i=1}^k d_{Q_n}(r + i), \quad k \geq 1,$$

all positive integral moments of the probability measures P_n, concentrated on $(0, \infty)$, converge to finite, positive limits. Standard weak convergence arguments produce a unique limiting p.m. P on $[0, \infty)$, which is easily seen to have the desired properties.

3. The variational problem—a necessary condition for admissibility. In this section, we establish Theorem 1.1. Let $r \geq -1$ be an integer and

suppose that $d \in \mathscr{G}_r$, so that d is Bayes in problem $\mathscr{P}_r$. To apply the Stein-LeCam characterization given previously, we consider the difference in integrated risk $B_r(d, Q) - B_r(Q)$ for priors Q with finite total mass and compact support in $(0, \infty)$. Substituting (2.3) into (2.2) gives

$$B_r(d, Q) - B_r(Q) = \sum_{x=r+1}^{\infty} (p_x/p_{x-1} - q_x/q_{x-1})^2 q_{x-1}/x!. \tag{3.1}$$

Setting $h_x = u_x^2 = q_x/p_x$, we find that this

$$= \sum_{r+1}^{\infty} ((h_x - h_{x-1})^2/h_{x-1})(p_x^2/p_{x-1}x!) \tag{3.2}$$

$$= \sum_{r+1}^{\infty} (u_x - u_{x-1})^2(1 + u_x/u_{x-1})^2 a_x, \tag{3.3}$$

where $a_x = p_x^2/p_{x-1}x!$.

Since $u_x > 0$, it is clear that

$$(1 + u_x/u_{x-1})^2 \geq 1, \tag{3.4}$$

and this yields (1.2).

From Theorem 2.2 and the Stein-LeCam theorem, we conclude that if $d(x)$ is admissible (and belongs to $\mathscr{G}_r$), then

$$\inf_{u \in \mathscr{U}_{\mathscr{Q}}} \sum_{x=r+1}^{\infty} (Du_x)^2 a_x = 0. \tag{3.5}$$

REMARK 3.1.

a) If u_x is a decreasing function of x (for example if $dQ/dP = g(\lambda)$ is decreasing), then $(1 + u_x/u_{x-1})^2 \leq 4$. The "natural" priors for proving admissibility via Blyth's method have $g(\lambda)$ decreasing (c.f. Section 5). Hence the simplification at (3.4) (effectively a linearization—see Sections 4 and 7) preserves the essence of the statistical problem.

b) In the continuous case, the factor corresponding to a_x is just the marginal density of the prior. When $r = -1$ or $P(\{0\}) = 0$, $p_x/x!$ is the marginal density of the prior $\pi(d\lambda) = \lambda^{-r}e^{\lambda}P(d\lambda)$ in problem $\mathscr{P}$ so that $a_x = d_P(x)\pi p(x)$. The appearance of the factor p_x/p_{x-1} is forced by the discreteness of the problem.

The class of functions $\mathscr{U}_{\mathscr{Q}}$ appearing in (3.5) is somewhat artificial. We replace it with a class satisfying boundary conditions of a kind more commonly associated with variational problems. The sequence of measures occurring in the Stein-LeCam theorem all satisfy $Q([\lambda_0, \lambda_0 + 1]) \geq 1$. It follows that $u_r^2 = q_r/p_r \geq 1/p_r$. Inspection of (3.1) reveals that $B_r(d, Q) - B_r(Q)$ is unaffected if $P(d\lambda)$ is multiplied by an arbitrary constant b. Thus we can rescale u_r^2 by any desired value—in particular the infimum in (3.5) can be replaced by an infimum over $\mathscr{U}_{\mathscr{Q}} \cap \{u_x : u_r \geq 1\}$.

Assume temporarily that the support of P is unbounded. Since $Q \in \mathscr{Q}$ has compact support, it can easily be verified that $u_x^2 \to 0$ as $x \to \infty$. (In fact, the convergence is exponentially fast.) Thus, we may replace $\mathscr{U}_{\mathscr{Q}}$ in (3.5) by $\mathscr{U}_{\mathscr{Q}} \cap \mathscr{U}_r$, where $\mathscr{U}_r$ is as defined in Section 1. Dropping the restriction that $u \in \mathscr{U}_{\mathscr{Q}}$, it follows that if $d(x)$ is admissible, then (1.3) holds.

Condition (1.3) is purely analytical. It may be rephrased more perspicuously in terms of the constants $\{a_s\}$ alone.

LEMMA 3.1. $\inf_{\mathscr{U}_r} \sum_{r+1}^{\infty} (Du_s)^2 a_s = 0$ *iff* $\sum_{r+1}^{\infty} 1/a_s = \infty$.

PROOF. From the Cauchy-Schwartz inequality, if $n > r$ and $u \in \mathscr{U}_r$,

$$(1 - u_n)^2 = [\sum_{r+1}^{\infty} Du_s]^2 \leq (\sum_{r+1}^{n} (Du_s)^2 a_s)(\sum_{r+1}^{n} 1/a_s), \tag{3.6}$$

with equality iff $Du_s = c/a_s$ for $r + 1 \leq s \leq n$. For each n, choose c so that $u_n = 0$. Now let $n \to \infty$. □

Let us now suppose that supp P is bounded. The above approach could probably be adapted to this case, since all that is required for our application is sup(supp P) > sup(supp Q). Here is a more direct route.

LEMMA 3.2. *If* supp P *is bounded, then* $\sum_{r+1}^{\infty} 1/a_x = \infty$.

PROOF. Let M be an upper bound for supp P. Now $d_P(s) \leq M$, while for $s \geq r + 1$,

$$p_s \leq P([0, 1]) + M^{s-r} P([1, M]) \leq CM^s.$$

Since $a_s = d_P(s)p_s/s!$

$$\sum_{r+1}^{\infty} 1/a_s \geq (1/CM) \sum_{r+1}^{\infty} s!/M^s = \infty. \quad \square$$

Combining the two lemmas, and the result for unbounded supp P, we obtain Theorem 1.1.

REMARK 3.2. The import of Theorem 1.1 is that extension of the class of "feasible" functions from $\mathscr{U}_2$ to $\mathscr{U}_r$ yields a familiar variational problem with boundary conditions. The connection to difference equations and Markov processes that results is described in the next section. Section 5 shows, modulo regularity conditions, that there is no harm from a *statistical* point of view in this change in the set of feasible directions.

REMARK 3.3. In Section 6.2, we show that all truncated Generalized Bayes procedures (i.e. in some $\mathscr{G}_r$) with supp P bounded are admissible. This is trivial whenever $B(P) < \infty$, but the latter does not follow from bounded support if $r > 1$.

REMARK 3.4. There is no difficulty in extending the results of this section to discrete exponential families (the so called "power series distributions"):

$$f(x \mid \theta) = \theta^x \phi(\theta) t_x, \quad x = 0, 1, \cdots,$$

where $t(x) > 0$ for all $x \geq 0$ and the parameter space

$$\Omega = \{\theta > 0: \sum_0^{\infty} \theta^x t_x < \infty\}$$

is an interval with left endpoint zero and right endpoint $\omega_0 \leq \infty$. Examples

include the Poisson, negative binomial ($f(x \mid \theta) = \binom{r+x-1}{r-1}\theta^x(1-\theta)^r$) and logarithmic distributions (the latter two having $\omega_0 = 1$). In general, a prior $P(d\theta)$ on $\overline{\Omega}$ leads to $p_x = \int \theta^{x-r}P(d\theta)$ in problem $\mathscr{P}$, and for the loss function $L(d, \theta) = (\sum_{x>r}^{\infty} \theta^x t_x)\theta^{-r-1}(d-\theta)^2$, $d_P(x) = p_x/p_{x-1}$ as before. The Brown-Farrell theorem applies in this generality, as does the argument that $u_x^2 = q_x/p_x \to 0$ as $x \to \infty$ (at least if $\sup(\text{supp } P) = \omega_0$). If $\sup(\text{supp } P) < \omega_0$, Lemma 3.2 remains in force, the only change in the proof being to observe that $\sum_s 1/M^s t_s = \infty$ since $M < \omega_0$ implies $\sum_s M^s t_s < \infty$. Consequently, Theorem 1.1 remains valid with $a_x = p_x^2 t_x/p_{x-1}$.

4. Difference equations and recurrence. This section considers equivalent analytic and probabilistic forms of condition (1.3).

Analytic versions. Let u, k be real-valued functions defined on

$$\mathbb{Z}_r = \{r, r+1, \cdots\}$$

which vanish at ∞. Since the coefficients a_s (corresponding to a generalized Bayes estimator $d_P \in \mathscr{C}_r$) are positive, the expression

$$\langle u, k \rangle = \sum_{r+1}^{\infty} a_s(Du_s)(Dk_s)$$

defines an inner product on the class of functions for which $\| u \|^2 < \infty$. (In analogy with differential equations and Markov processes, the linear space of such functions may be called the Dirichlet space corresponding to the generalized Bayes estimate d_P. This point of view is exploited in Johnstone and Lalley (1984).)

First we derive the Euler-Lagrange equation associated with the minimization problem (1.3). Let k be a function on $\mathbb{Z}_r$ with bounded support such that $k_r = 0$. Suppose u attains the minimum of $\| u \|^2$ over $\mathscr{U}_r$. Then $u + \varepsilon k \in \mathscr{U}_r$ for all ε, and hence

$$\begin{aligned} 0 &\le \lim_{\varepsilon \to 0} (1/\varepsilon)\{\| u + \varepsilon k \|^2 - \| u \|^2\} \\ &= 2 \sum_{r+1}^{\infty} a_s(Du_s)(Dk_s) \\ &= -2 \sum_{r+1}^{\infty} k_s D(a_{s+1}Du_{s+1}) - 2k_r a_{r+1}Du_{r+1}. \end{aligned}$$

Since the last term is zero, and k is arbitary, it follows that u satisfies (1.6) of Section 1.

It is easy to solve (1.6) and the minimization problem explicitly.

THEOREM 4.1. $\min_{\mathscr{U}_r} \| u \|^2 = [\sum_{r+1}^{\infty} 1/a_x]^{-1}$, *and* $\min_{\mathscr{U}_r} \| u \|^2 > 0$ *iff there is a solution* u_0 *to* (1.6).

In the case $\min_{\mathscr{U}_r} \| u \|^2 > 0$, *the minimum is uniquely attained by* u_0. *The latter is the unique solution to* (1.6) *and has representation*

$$u_0(s) = [\sum_{r+1}^{\infty} 1/a_x]^{-1} \sum_{s+1}^{\infty} 1/a_x. \tag{4.1}$$

PROOF. Clearly the general solution to $Au = a_{s+1}(Du_{s+1}) - a_s(Du_s) = 0$

satisfies $Du_s = c/a_s$ and has the form $u_s = c\sum_{r+1}^{\infty} 1/a_x + d$, for arbitrary constants c and d, and also attains equality in (3.6) for $r+1 \le s \le n$.

Now let $n \to \infty$. If $\sum_{r+1}^{\infty} 1/a_s < \infty$, we can choose c and d uniquely so that the resulting solution u_0 satisfies the boundary conditions imposed by $\mathscr{U}_r$. Indeed, $u_0(s) = c_0 \sum_{r+1}^{\infty} 1/a_x$, where $c_0 = [\sum_{r+1}^{\infty} 1/a_x]^{-1}$, and it is clear from (3.6) and strict convexity of $u \to \| u \|^2$ on $\mathscr{U}_r$ that u_0 uniquely minimizes $\| u \|^2$, with the claimed minimum value.

If $\sum_{r+1}^{\infty} 1/a_s = \infty$, then the functions

$$k_s^{(n)} = [\textstyle\sum_{r+1}^{n} 1/a_x]^{-1}[\sum_{s+1}^{n} 1/a_x]\chi_{\{r \le s \le n\}}$$

are in $\mathscr{U}_r$ and have $\| k^{(n)} \|^2 = [\sum_{r+1}^{n} 1/a_x]^{-1} \to 0$. It is also clear that no choice of c and d will satisfy the boundary conditions. □

Probabilistic version. It is well known that one can associate a diffusion (a continuous strong Markov process) with a nice elliptic second order differential operator, which then becomes the infinitesimal generator of the process. In the discrete problem (1.6), we encounter the second order *difference* operator A. This can be represented as a tridiagonal matrix $Q = (q_{st})$, $s, t, \ge r$. For $s > r$,

$$q_{s,s-1} = a_s, \quad q_{s,s+1} = a_{s+1}, \quad q_{ss} = -(a_s + a_{s+1}), \quad q_{st} = 0, \quad \text{otherwise.} \tag{4.2}$$

For $s = r$, we put

$$q_{rr} = -a_r, \quad q_{r,r+1} = a_r, \quad q_{rt} = 0 \text{ otherwise.} \tag{4.3}$$

On countable state spaces, the infinitesimal generator of a Markov process is an infinite matrix. The matrix Q is just the generator of an irreducible birth and death process on the state space $\mathbb{Z}_r = \{r, r+1, \cdots\}$.

The boundary condition (4.3) is chosen to preserve the symmetry at Q (and hence the time reversibility of the process). Thus, on hitting the lower boundary r, the process waits for an exponentially distributed time (mean $1/a_r$) and then jumps to $r+1$.

Technical remark. The construction problem (existence and uniqueness) for a birth and death process $\{X_t, t \ge 0, P^x\}$ with prescribed transition rates is discussed for example in Dynkin and Yushkevich (DY, 1969, Chapter 4).

If the transition rates a_s increase sufficiently fast as $s \to \infty$, the process will "explode". Let τ_n be the time of the nth jump of the process, and $T = \lim_{n\to\infty} \tau_n$. It is shown in DY that either $P^x(T = \infty) = 1$ for all $x \in \mathbb{Z}_r$ or $P^x(T < \infty) = 1$ for all $x \in \mathbb{Z}_r$. The latter event corresponds to explosion, and occurs iff (DY page 173)

$$\sum_{r+1}^{\infty} s/a_s < \infty.$$

In this case, we adjoin the point $\{\infty\}$ as an absorbing state to the state space, and set $X_t = \infty$ for $t \ge T$. With this modification, the rates (4.2) and (4.3) correspond to a unique right continuous strong Markov process of $\mathbb{Z}_r \cup \{\infty\}$.

For our purposes, it will usually suffice to consider the embedded discrete time Markov chain $\{X_n, n \ge 0, P^s\}$ with transition probabilities $P = (p_{st})$, $s, t \ge r$

given by

$$
\begin{aligned}
&p_{s,s-1} = \frac{a_s}{a_s + a_{s+1}}, \quad p_{s,s+1} = \frac{a_{s+1}}{a_s + a_{s+1}}, \quad s > r \\
&p_{r,r+1} = 1, \qquad\qquad p_{s,t} = 0 \qquad\qquad \text{otherwise.}
\end{aligned} \tag{4.4}
$$

We now give the probabilistic interpretation of Theorem 4.1. Introduce

$$k_s = P^s(\exists\ n \geq 0 : X_n = r) \quad \text{for} \quad s \geq r.$$

Recall that if $k_x \equiv 1$, the chain is recurrent, otherwise it is transient. Since all a_s are positive and finite, recurrence of the chain is equivalent to recurrence of the process. In the transient case, k_s is clearly decreasing, and by a simple contradiction argument, $k_\infty = 0$. The next result is easily obtained from Theorem 4.1 and standard probabilistic arguments (e.g. Karlin and Taylor, 1975, 146–147).

THEOREM 4.2. *$\{X_t\}$ is recurrent iff* $\min_{\mathscr{U}_r} \| u \|^2 = 0$.

If $\{X_t\}$ is transient, k_s minimizes $\| u \|^2$ over $\mathscr{U}_r$, solves the boundary value problem (1.6) and has the representation (4.1).

COROLLARY. *If $d(x)$ is admissible then the corresponding birth and death chain $\{X_n\}$ is recurrent.*

REMARK 4.1. The simplicity of the proofs in this section depends heavily on discreteness of the sample space and the availability of explicit representations using $1/a_x$ in the one dimensional case. The results of this section (except those involving $\sum_{r+1}^{\infty} 1/a_x$) remain valid for arbitrary dimension $k \geq 1$ and are proved in II. The main tool there is the maximum principle for the self ajoint difference operator A.

Suppose now that $\{X_n\}$ is recurrent. We seek conditions on the prior $P(d\lambda)$ and the estimator $d_P(x)$ under which $\{X_n\}$ is null recurrent. The distinction between null and positive recurrence in the Gaussian case is important in discussing "immunity" to the Stein effect. (Gutmann, 1982, 1983; Johnstone and Lalley, 1984.) Recall that $\{X_n\}$ possesses an invariant measure μ which is unique up to a multiplicative constant (Feller, 1967, Chapter 14), and that $\{X_n\}$ is null recurrent if and only if μ is infinite. Harris (1952) gives a necessary and sufficient condition for a recurrent irreducible birth and death chain on $\mathbb{Z}_+$ to be null-recurrent and calculates the invariant distribution in the positive recurrent case. In our situation (cf. (4.4)), this reduces to

LEMMA 4.3. *$\{X_n\}$ is null recurrent iff both*

$$\sum_{r+1}^{\infty} a_k = \infty \quad \textit{and} \quad \sum_{r+1}^{\infty} 1/a_k = \infty.$$

In the positive recurrent case, the invariant distribution $\{\mu_k\}_{k\geq r}$ is

$$\mu_k = c(a_k + a_{k+1}), \quad k \geq r; \quad c^{-1} = \sum_{k=r}^{\infty} (a_k + a_{k+1}).$$

EXAMPLE. For the loss function $L(d, \lambda) = \lambda^{-1}(d - \lambda)^2$, the conjugate priors $\pi_\gamma(d\lambda) = \lambda^\gamma d\lambda$ lead to generalized Bayes estimates $d_\gamma(x) = (x + \gamma)^+$, with $a_x = (x + \gamma)\Gamma(x + \gamma + 1)/\Gamma(x + 1)$. It is easy to check that the corresponding chains are recurrent for $\gamma \leq 0$, and null recurrent for $-2 \leq \gamma \leq 0$. Also, d_γ is admissible iff $\gamma \leq 0$ (use Theorem 1 in Karlin, 1958, or Theorem 1.2).

The priors $\pi_\alpha(d\lambda) = e^{-\alpha\lambda}\, d\lambda$ for $\alpha \geq 0$ yield Bayes estimates

$$d_\alpha(x) = x/(1 + \alpha),$$

which are admissible for all $\alpha \geq 0$, with $a_x = x/(1 + \alpha)^{x+2}$. In this case the corresponding chains are recurrent for all $\alpha \geq 0$, but null recurrent only if $\alpha = 0$. Combining the two cases, one sees that only estimates that are "very close" to the MLE $d(x) = x$ lead to null-recurrent chains.

The condition for positive recurrence can be expressed in terms of the prior $P(d\lambda)$, at least under a mild technical side restriction. On $(0, \infty)$, let $\pi(d\lambda) = \lambda^{-r}e^\lambda P(d\lambda)$.

LEMMA 4.4. *If* $\int^\infty \lambda\, d\pi(\lambda) < \infty$ *then* $\sum a_k < \infty$. *Conversely, if* $\int^\infty \lambda\, d\pi(\lambda) = \infty$, *and either*

$$d_P(k + 1) \in O(d_P(k)) \tag{4.5}$$

or

$$\liminf_{k\to\infty} d_P(k)/k > 0 \tag{4.6}$$

then $\sum a_k = \infty$.

PROOF. By Fubini's Theorem,

$$\begin{aligned}\int_0^\infty & \lambda\, P_\lambda(X \geq r + 1)\, d\pi(\lambda) \\ &= \sum_{k=r+1}^{\infty} p_{k+1}/k! \\ &= \sum_{k=r+1}^{\infty} ((k + 1)/d_P(k + 1))a_{k+1} = \sum_{k=r+1}^{\infty} (d_P(k + 1)/d_P(k))a_k.\end{aligned}$$

Since $d_P(x)$ is monotone in x, this suffices for the proof.

REMARK 4.2. The analogous result for the normal distribution asserts that positive recurrence is equivalent to the prior having finite total mass—i.e., a proper prior. This effect would also occur in the Poisson problem if the loss function were changed to $\lambda^{-2}(d - \lambda)^2$. Admissibility can be trivially established in the positive recurrent case—see Section 6.2.

REMARK 4.3. The hierarchy of possible cases can be summarized as follows: (with an example of the corresponding $\pi(d\lambda)$ in parentheses)

Recurrence $\sum 1/\alpha_x = \infty$	positive	$\sum a_x < \infty$	$(\lambda^{-3}\, d\lambda)$
	null	$\sum \alpha_x = \infty$	$(d\lambda)$
Transience $\sum 1/\alpha_x < \infty$	non-explosion	$\sum x/\alpha_x = \infty$	$(\lambda d\lambda)$
	explosion	$\sum x/\alpha_x < \infty$	$(\lambda^2\, d\lambda)$.

We note that if $0 < \liminf_{x\to\infty} d(x)/x \le \limsup_{x\to\infty} d(x)/x < \infty$, then explosion occurs if and only if the marginal density π of P satisfies $\sum^{\infty} 1/\pi_x < \infty$.

REMARK 4.4 As all results in this section apply for arbitrary sequences $\{a_s\}$ of positive constants, they hold for the discrete exponential families of Remark 3.4.

5. The main admissibility theorem. This section is devoted to the proof of Theorem 1.2. The heuristics for the proof of this result are similar to those for the normal case, but the details of the argument reflect the lack of any invariant structure in the Poisson problem. Suppose the inequality in (1.2) were an equality; it would then suffice for admissibility to take a minimizing sequence $\{u^{(n)}\}$ in $\mathscr{U}_r$ and construct priors $\{Q_n\} \in \mathscr{Q}$ such that the resulting functions $\hat{u}_x^{(n)} = (q_x^{(n)}/p_x)^{1/2} \in \mathscr{U}_{\mathscr{Q}}$ also form a minimizing sequence. The priors Q_n are defined by

$$Q_n(d\lambda) = [u^{(n)}([\lambda])]^2 P(d\lambda),$$

where $[\lambda]$ denotes the integer part of λ.

It follows that $\hat{u}^2(x) = q_x/p_x = E(u^2([\lambda]) \mid x)$, where we have omitted reference to n. Assume here a strengthened form of (1.4): that $d(x) - x \in O(\sqrt{x})$. Suppose also that $P(\{0\}) = 0$, so that d_P is generalized Bayes for the prior measure $\pi(d\lambda) = \lambda^{-r}e^{\lambda}P(d\lambda)$ in the original problem $\mathscr{P}$. It is easy to check from this and (1.5) that the posterior distribution of λ given x becomes asymptotically "concentrated" at x; as $x \to \infty$:

$$E(\lambda \mid x) = d_P(x+1) \sim x \quad \text{and}$$

$$\operatorname{Var}(\lambda \mid x) = d_P(x+1)(d_P(x+2) - d_P(x+1)) \in O(x).$$

This suggests that if u is fairly smooth, then

$$\hat{u}_x \approx u_x \quad \text{and} \quad (D\hat{u}_x)^2 \approx (Du_x)^2,$$

so that $\{\hat{u}^{(n)}\}$ should be a minimizing sequence of the desired type.

These arguments will be legitimized "in mean" in steps $1^\circ - 3^\circ$ below. Throughout, M denotes a constant depending on P but not on λ or x, but not necessarily the same at each occurrence.

Suppose that $d = d_P \in \mathscr{C}_r$ and that (1.3) holds. We shall show that d_P is admissible in $\mathscr{P}_r$ and then appeal to Lemma 2.1. The sufficiency part of the

Stein-LeCam theorem, originally due to Blyth, calls for a sequence of finite measures $\{Q_i\}$ on $[0, \infty)$ such that

(i) $B_r(Q_i) < \infty$,
(ii) $Q_i([p, p+1]) \geq \eta$ for some positive constants p and η and
(iii) $B_r(d_p, Q_i) - B_r(Q_i) \to 0$ as $i \to \infty$.

Given (i)–(iii), the proof is standard (cf. Berger, 1980; Brown and Hwang, 1982).

1°. Let v: $\mathbb{Z}_{r+1} \to \mathbb{R}$, and by convention set $v_i = v_{r+1}$ for $0 \leq i \leq r+1$. Let $\overline{\lambda}$ denote the least integer greater than or equal to λ. Following the heuristics of Section 1, we let $Q(d\lambda) = (v_{\overline{\lambda}})^2 P(d\lambda)$. If v is such that $B_r(Q) < \infty$, then the basic equation (2.2) holds with d replaced by d_P. This paragraph expresses the right side of (2.2) in terms of global differences of v at (5.4), which are converted to one-step differences in 2°, leading to the final result in 3°.

Let m_x be an as yet unspecified function and note for $x \geq r+1$ that

$$d_Q(x) - d_P(x) = \frac{1}{q_{x-1}} \int (\lambda - d_P(x))(v_{\overline{\lambda}}^2 - m_x^2) \quad \lambda^{x-r-1}\, dP(\lambda).$$

Factoring the difference in squares and using the Cauchy-Schwartz inequality yields

$$[d_Q(x) - d_P(x)]^2 q_{x-1} \leq \frac{2}{q_{x-1}} \int [v_{\overline{\lambda}}^2 + m_x^2]\lambda^{x-r-1}\, dP(\lambda)$$

$$\times \int (\lambda - d_P(x))^2 (v_{\overline{\lambda}} - m_x)^2 \lambda^{x-r-1}\, dP(\lambda).$$

Now let $m_x = \int v_{\overline{\lambda}} W_{x-r-1}(d\lambda)$, where $W_x(d\lambda)$ is the probability measure with $dW_{x-r-1}/dP \propto \lambda^{x-r-1}$ for $x \geq r+1$. Thus $m_x^2 \leq q_{x-1}/p_{x-1}$, and

$$[d_Q(x) - d_P(x)]^2 q_{x-1} \leq 4 \int [\lambda - d_P(x)]^2 (v_{\overline{\lambda}} - m_x)^2 \lambda^{x-r-1}\, dP(\lambda).$$

Using the inequality

$$(v_{\overline{\lambda}} - m_x)^2 \leq 2(v_{\overline{\lambda}} - v_x)^2 + 2 \int (v_x - v_{\overline{\lambda}})^2 W_{x-r-1}(d\lambda), \tag{5.1}$$

the identity (2.1) and the bounds (1.4), (1.5)

$$B(d_P, Q) - B(d_Q, Q)$$
$$\leq M \int dP(\lambda) \sum_{x>r} [(\lambda - d_P(x))^2 + x](v_x - v_{\overline{\lambda}})^2 \lambda^{x-r-1}/x!. \tag{5.2}$$

It is now convenient to extend the definition of d from $\mathbb{Z}_+$ to $\mathbb{R}_+$ by linear interpolation. Condition (1.5) implies a Lipschitz bound (with constant M) for the (interpolated) rule d, and hence

$$(\lambda - d_P(x))^2 \leq 2(\lambda - d_P(\lambda))^2 + 2M(\lambda - x)^2, \tag{5.3}$$

so that (5.2) is bounded above by

$$M' \int dP(\lambda)\{(\lambda - d_P(\lambda))^2 \textstyle\sum_{x>r} (v_x - v_{\bar\lambda})^2 \lambda^{x-r-1}/x! + \sum_{x>r} [x + (x-\lambda)^2](v_x - v_{\bar\lambda})^2 \lambda^{x-r-1}/x!\}. \tag{5.4}$$

2°. To bound the sums on x in (5.4), we need some Poincaré inequalities for Poisson measures. These are proved in the appendix.

LEMMA 5.1. *For any function* $u\colon \mathbb{Z}_+ \to \mathbb{R}$, *such that* $Eu^2(x) < \infty$, *and for all* $\lambda \geq 0$,

$$E[u(x) - u(\bar\lambda)]^2 \leq 2(\lambda + 1)E[D^+u(x)]^2$$

$$E\frac{x}{\lambda}[u(x) - u(\bar\lambda)]^2 \leq 2(\lambda + 1)E\frac{x}{\lambda}[D^+u(x)]^2$$

$$E\left(\frac{(x-\lambda)^2}{\lambda}\right)[u(x) - u(\bar\lambda)]^2 \leq 5(\lambda + 1)E\left(\frac{(x-\lambda)^2}{\lambda} + \frac{x}{\lambda}\right)[D^+u(x)]^2.$$

To apply Lemma 5.1 to (5.4), note that the function $v_{\bar\lambda}$ is constant for $\lambda \leq r + 1$, and hence there is no harm in inserting $I_r = I\{x > r\}$ in all integrands in the statement of Lemma 5.1. It follows that (5.4) is bounded by

$$M \textstyle\sum_{x>r} (D^+v_x)^2 \int dP(\lambda)(\lambda + 1)\{[\lambda - d_P(\lambda)]^2 + x + (x-\lambda)^2\}\dfrac{\lambda^{x-r-1}}{x!}. \tag{5.5}$$

Use the Lipschitz bound on d_P to bound the term in curly brackets in terms of

$$(d_P(x) - x)^2 + x + (\lambda - d_P(x))^2.$$

Now evaluate the integral on λ, using the definition of p_x, (2.1), (1.4) and (1.5) to bound (5.5) by

$$M \textstyle\sum_{x>r} (D^+v_x)^2[(d_P(x) - x)^2 + x + d_P(x)][\pi_x + \pi_{x-1}/x]. \tag{5.6}$$

3°. Since $d_P(x) \geq d_P(r+1) = M_0 > 0$, it follows that

$$\pi_{x-1}/x = \pi_x/d_P(x) \leq M\pi_x \quad \text{for} \quad x \geq r + 1.$$

From (1.4) and (5.6), this implies

$$B_r(d_P, Q) - B_r(Q) \leq M \textstyle\sum_{x>r} (D^+v_x)^2 x[1 + (d_P(x) - x)^2/x]\pi_x.$$

Let $\alpha_x = x^{1/2}((d_P(x)/x) - 1) \leq B < \infty$ by (1.4). Then, writing d_x for $d_P(x)$,

$$[1 + (d_x - x)^2/x]\pi(x)/\pi(x - [x^{1/2}] + 1) \leq (1 + \alpha_x^2) \textstyle\prod_{i=0}^{[x^{1/2}]-2} (1 + \alpha_{x-i}/(x-i)^{1/2}) = L_x.$$

From (1.5), it is easy to show that $D^-\alpha_x \in O(x^{-1/2})$, so for $0 \leq i \leq [x^{1/2}]$, there

exists a constant M such that $\alpha_{x-i} \le \alpha_x + M$. Thus the product in L_x is bounded by

$$\exp\{(\alpha_x + M) \sum_0^{[x^{1/2}]-2} (x - i)^{-1/2}\}.$$

If $x - [x^{1/2}] > x/4$, then the sum above is bounded between ½ and 2. Letting $\varepsilon_x = \pm 1$ according as $\alpha_x + M$ is greater or less than zero, we obtain

$$L_x \le (1 + \alpha_x^2)\exp\, 2^{\varepsilon_x}(\alpha_x + M) \le M',$$

since $\alpha_x \le B$. Noting that $(x + 1)\pi_{x+1} = \pi_x d_x(d_{x+1}/d_x) \le Ma_x$ by (1.5) we find

$$x[1 + (d_P(x) - x)^2/x]\pi_x \le M(x - [x^{1/2}] + 1)\pi_{x-[x^{1/2}]+1} \le M' a_{x-[x^{1/2}]}.$$

Now it is clear that $\sum^\infty 1/a(x - [x^{1/2}]) = \infty$ if $\sum^\infty 1/a(x) = \infty$, so it follows from Lemma 3.1 that d_P is admissible. □

REMARK. The Poincaré inequalities of Lemma 5.1 are no longer valid without extra hypotheses on the rate of growth of v in dimensions $d > 2$. Typically such hypotheses are satisfied by solutions of elliptic difference equations or by suitably smoothed versions of v. See II or Johnstone and Shahshahani (1983) for examples and further details.

Theorem (1.2) can be applied to yield a more specific criterion for admissibility.

COROLLARY 5.2. *Suppose that $d(x) = d_P(x)$ for x large. If for some $\alpha > 1$, $d(x) \ge x + \alpha/\log x$ for large x, then $d(x)$ is inadmissible. Conversely, if $d(x)$ satisfies* (1.4), (1.5), *and for some $\alpha \le 1$, $d(x) \le x + \alpha/\log x$ for large x, then $d(x)$ is admissible.*

PROOF. Note first that for any large r_0

$$\pi_x = p_x/x! = c(r_0) \prod_{i=r_0}^x d_i/i.$$

In the first case, suppose that the inequality holds for $x \ge r_0$. Choose $\beta < 1$ so that $\alpha\beta > 1$. Then for r_1 sufficiently large,

$$\log(\prod_{r_0}^x d_i/i) \ge \log \prod_{r_0}^x (1 + (\alpha/i \log i))$$

$$\ge \alpha\beta \sum_{r_1}^x (1/i \log i) \sim \alpha\beta \log\log x, \quad \text{as} \quad x \to 0.$$

Consequently, for large x, $\pi_x > c(\log x)^\gamma$, where $\gamma > 1$, so that

$$\sum^\infty 1/a_x \le \sum^\infty (1/x(\log x)^\gamma) < \infty.$$

In the second case, $\sum^\infty 1/a_x = \infty$ follows from a similar argument, this time using $\log(1 + x) \le x$. Now apply Theorems 1.1, 1.2 and Lemma 3.1. □

REMARK. A more general sharp comparison criterion can be given using Dini functions (cf. Srinivasan, 1981; or Section 6 of II for a statement in the Poisson context).

6. Discussion.

1. *Risk restrictions.* The growth conditions on $d_P(x)$ imposed in Theorem 1.2 may be reexpressed in terms of boundedness of risk and posterior risk. In addition, generalized Bayes rules with bounded risk necessarily have bounded increments, thus condition (6.1) is then redundant.

Thus, if one regards rules with unbounded risk as undesirable, then (6.2) is a reasonable restriction on $d_P(x)$ itself. This is the analog of Brown's result characterizing bounded risk generalized Bayes rules as bounded perturbations of $d(x) = x$ in the normal case.

PROPOSITION 6.1.

(6.1) $d_P(x+1) - d_P(x) \in 0(1) \Leftrightarrow E[L(d_P(x), \lambda) \mid x]$ *is bounded as* $x \to \infty$.

(6.2) $d_P(x) - x \in O(\sqrt{x})$ *as* $x \to \infty \Leftrightarrow R(\lambda, d_P)$ *is bounded as* $\lambda \to \infty$, *and either condition implies* (6.1).

(6.3) $d_P(x) - x \le M\sqrt{x} \Leftrightarrow R(\lambda, d_P(x) \vee x)$ *is bounded as* $\lambda \to \infty$.

PROOF. Strictly speaking, $E[L(d_P(x), \lambda) \mid x]$ is well defined only when $r \ge 0$ or $P(\{0\}) = 0$, so that the prior measure π on $[0, \infty)$ in problem $\mathscr{P}$ can be defined by $\pi(d\lambda) = \lambda^{-r}e^{\lambda}P(d\lambda)$. In general, the posterior risk is defined as $\int (\lambda - d_P(x))^2\lambda^{x-r-1}\, dP(\lambda)/p_x$. This is of course consistent with the usual definition in the special (but usual) case described above. Now (6.1) is just a trivial consequence of (2.1).

If (6.2) holds, then for some constant c,

$$\begin{aligned} E_\lambda[d(x) - \lambda]^2 &\le E_\lambda(x - \lambda + c\sqrt{x})^2 + (x - \lambda - c\sqrt{x})^2 \\ &= 2E(x - \lambda)^2 + 2c^2Ex = 2(1 + c^2)\lambda, \end{aligned}$$

from which the conclusion follows.

For the converse, we use the monotonicity of the Bayes rule $d_P(x)$. If (6.2) fails, there exists an increasing sequence $\{x_n\}$ such that $|d_P(x_n) - x_n| > n\sqrt{x_n}$. It follows that $|d_P(x) - x_n| \ge n\sqrt{x_n}/2$ for all x in at least one of the intervals $A_{1n} = (x_n, x_n + n\sqrt{x_n}/2)$ or $A_{2n} = (x_n - n\sqrt{x_n}/2, x_n)$. Then

$$\begin{aligned} (1/x_n)E_{\lambda=x_n}(d_P(x) - \lambda)^2 &\ge (n^2/4)\ \min_{i=1,2}P_{\lambda=x_n}(x \in A_{ni}) \\ &\sim (n^2/4)[\Phi(n/2) - \Phi(0)] \sim n^2/8 \quad \text{as} \quad n \to \infty, \end{aligned}$$

since $\mathscr{L}\{(x - \lambda)/\sqrt{x}\} \to N(0, 1)$ as $\lambda \to \infty$. Thus $R(x_n, d_P)$ is unbounded. (6.3) is proved in a similar fashion.

For the second part of (6.2), note from (2.1) that

$$\begin{aligned} D^+d_P(x) &= \operatorname{Var}(\lambda \mid x - 1)/E(\lambda \mid x - 1) \\ &\le (2/d_P(x))\{E[(\lambda - x)^2 \mid x - 1] + (d_P(x) - x)^2\}. \end{aligned}$$

(Note that for x large, the posterior distribution of $\lambda \mid x$ for the "prior" $\pi(d\lambda) = \lambda^{-r}e^{\lambda}P(d\lambda)$ in problem $\mathscr{P}$ *is* well-defined.) From (6.2), the second term is

bounded, and the first equals $(2/\pi(x)) \int ((\lambda - x)^2/\lambda)p_\lambda(x)dP(\lambda)$. Borrowing from II, Proposition 5.5, the fact that for a suitable constant c, $\lambda^{-1}(\lambda - x)^2 p_\lambda(x) \leq M[p_\lambda(x + c\sqrt{x}) + p_\lambda(x - c\sqrt{x})]$, where $x - c\sqrt{x}$ is taken as zero for $x < c\sqrt{x}$, we obtain as in 3° of the proof of Theorem 1.2 that

$$D^+ d_P(x) \leq M_1 + M_2[\pi(x + c\sqrt{x}) + \pi(x - c\sqrt{x})]/\pi(x) \leq M. \square$$

2. *Bounded support and positive recurrent priors.* In both these cases admissibility can be easily established without recourse to Theorem 1.2. Suppose that $d \in \mathscr{L}_r$, so that d is (unique) Bayes with respect to a prior P on $[0, \infty)$. Admissibility of d in $\mathscr{P}_r$ and $\mathscr{P}$ will follow trivially from finiteness of $B_r(d_P, P)$. If supp P is bounded, finiteness of $B_r(P)$ is a consequence of the boundedness of $R_r(d, \lambda)$ at $\lambda = 0$. Suppose now that the birth-death process corresponding to P is positive recurrent and that one of the mild conditions (4.6) or (4.6)′ hold. A short calculation shows that $B_r(d_P, P) = \sum_{r+1}^{\infty} (p_{x+1}/x! - a_x)$. Now $\sum a_x < \infty$ by assumption and finiteness of the first term follows as in the proof of Lemma 4.4. Thus the force of Theorem 1.3 results in showing admissibility of "null recurrent" estimators. Incidentally, finiteness of integrated risk does not imply positive recurrence, as consideration of the conjugate prior $\pi(d\lambda) = \lambda^{-2}\, d\lambda$ shows.

3. *Other approaches to proving admissibility.* Although the primary purpose of this theory (at least in one-dimension) is the characterization of admissibility, it can of course be used to establish admissibility of estimates not covered by existing methods. Brown and Hwang (1982) have developed a simple and unified approach (also using Blyth's method) for proving admissibility of generalized Bayes estimates of the mean vector of a multiparameter exponential family. The simplicity is achieved by using a *single* sequence of priors for all estimators. There is a cost in terms of regularity conditions and we give here an example of an admissible estimate which fails the Brown-Hwang sufficient conditions.

The growth condition used by Brown and Hwang is in this problem

$$\int_0^{1-\varepsilon} + \int_{1+\varepsilon}^{\infty} \frac{p(\lambda)\, d\lambda}{\lambda(\log \lambda)^2 \log^2(|\log \lambda| \vee 2)} < \infty,$$

where $p(\lambda)$ is the prior density. This condition fails if

$$p(\lambda) = (\log \lambda)\log(|\log \lambda| \vee 2).$$

However one can show fairly easily that $\pi(x) = \log x(\log\log x) + O(x^{-1/2})$ as $x \to \infty$, and this implies condition (5.1), that $a_x \sim x(\log x)(\log\log x)$ for large x and hence the admissibility of $d_P(x)$.

7. A counterexample: a recurrent, inadmissible estimator. In view of Proposition 6.1, the characterization theorems of previous sections assert that generalized Bayes rules with bounded risk are admissible iff the tail of the prior is light, in the sense that $\sum^{\infty} 1/a_x = \infty$. Unfortunately, the bounded risk (of $d_P(x) \vee x$) requirement in Theorem 1.3 cannot be completely dispensed with. We use a "large gap" construction to build a prior with light tail such that the correspond-

ing Bayes estimator has risk > 1 for all λ. An earlier construction of this sort was given in the normal case by Brown (1979).

Let $b_0 = 0$. Place mass $\nu_1 = 1$ at $a_1 = 1$, choose an integer $b_1 \geq 160$ and let $c_0 = b_1 \nu_1 p_{a_1}(b_1)$. Inductively, for $i \geq 1$, choose $a_{i+1} \geq 2b_i + 5$ such that

$$\tfrac{1}{4}(a_{i+1} - b_i)^2 P_{b_{i-1}}(b_i + \sqrt{b_i} < x \leq a_{i+1}) > b_i. \tag{7.1}$$

Now choose b_{i+1} as the smallest integer $\geq a_{i+1}$ such that $p_{a_{i+1}}(b_{i+1} - 1) < p_{a_{i+1}}(b_i)$. By increasing a_{i+1} slightly (which does not affect (7.1)), we can ensure

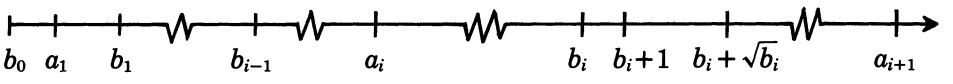
$$p_{a_{i+1}}(b_i) = p_{a_{i+1}}(b_{i+1} - 1). \tag{7.2}$$

Note from (7.2) and Jensen's inequality in turn that

$$\log\left[\frac{a_{i+1}}{b_{i+1} - 1}\right] \leq (b_{i+1} - b_i - 1)^{-1} \int_{b_i+1}^{b_{i+1}} \log\left[\frac{s}{b_{i+1} - 1}\right] ds \leq \log \frac{3}{4}$$

so that $a_{i+1} \leq (3/4)(b_{i+1} - 1)$. Finally place mass ν_{i+1} at a_{i+1}, where ν_{i+1} satisfies

$$\nu_{i+1} p_{a_{i+1}}(b_i + 1) = \nu_i p_{a_i}(b_i) b_i/(b_i + 1). \tag{7.3}$$



In outline, (7.2) and (7.3) ensure that the marginal density $\pi(x)$ of the prior is small at $b_i + 1$ (less than $3c_0/(b_i + 1)$), and approximately equal to $\pi(b_i)$ (so that $d(b_i + 1) \leq 3(b_i + 1)$), so that $\sum_{x=1}^{\infty} 1/d(x)\pi(x) = \infty$ and the tail of the prior is light. However, the gaps between a_i and a_{i+1} are so large that if $x \in [b_i + \sqrt{b_i}, a_{i+1}]$, then $d_P(x) \geq a_{i+1} - 80$. Condition (7.1) says that for $\lambda \in [b_{i-1}, b_i]$, this event has sufficiently large probability that d_P has risk > 1. This argument applies for all $i \geq 1$, so d_P is inadmissible, being dominated by the MLE.

In completing the verification we need a useful bound (proved in the Appendix) on the tails of the Poisson distribution, which will also be extensively used in II.

PROPOSITION 7.1. *If $x, y \in \mathbb{R}^+$, $y > x$ or $y \leq x - 1$, then $p_\lambda(y)/p_\lambda(x) \leq (\lambda/x)^{y-x}$. If $0 \leq x - 1 < y \leq x$, then $p_\lambda(y)/p_\lambda(x) \leq e^{|\psi(1)|}(\lambda/x)^{y-x}$, where $\psi(z) = (d/dz)\log \Gamma(z)$ is the digamma function.*

Returning to the counterexample, to see first that $\pi(b_i + 1) \leq 3c_0/(b_i + 1)$, note that (7.3) and (7.2) imply

$$\nu_i b_i p_{a_i}(b_i) \equiv c_0, \quad \text{for all } i. \tag{7.4}$$

Thus it suffices to show that $\sum_{j=1}^{\infty} \nu_j p_{a_j}(b_i + 1) \leq 3\nu_{i+1} p_{a_{i+1}}(b_i + 1)$. Using (7.3) and (7.4) together, then (7.2), Proposition 7.1 and $b_{j-1}/a_j \leq \frac{1}{2}$, we find

$$\begin{aligned} &[\nu_{i+1} p_{a_{i+1}}(b_i + 1)]^{-1} \textstyle\sum_{j>i+1} \nu_j p_{a_j}(b_i + 1) \\ &\quad = \textstyle\sum_{j>i+1} p_{a_j}(b_i)/p_{a_j}(b_{j-1}) \leq \sum_{j>i+1} (\tfrac{1}{2})^{b_{j-1} - b_i} \leq \tfrac{1}{2}. \end{aligned} \tag{7.5}$$

A similar argument is used for $j \le i$, but bounds $p_{a_j}(b_i)/p_{a_j}(b_j - 1)$ and employs the relation $a_j/(b_j - 1) \le \frac{3}{4}$. For the bound on $d(b_i + 1)$, use (7.4):

$$\frac{d(b_i + 1)}{b_i + 1} = \frac{\pi(b_i + 1)}{\pi(b_i)} \le \frac{3c_0/(b_i + 1)}{\nu_i p_{a_i}(b_i)} \le 3.$$

Finally, suppose that $b_i + \sqrt{b_i} < x \le a_{i+1}$. From the monotonicity of $d_P(x)$, (7.4), (7.2) and Proposition (7.1), we have

$$\begin{aligned} a_{i+1} - d(x) &\le a_{i+1} - d(b + \sqrt{b_i} + 1) \le \textstyle\sum_1^i (a_{i+1} - a_j) P(a_j \mid x = b_i + \sqrt{b_i}) \\ &\le \textstyle\sum_1^i \dfrac{a_{i+1}^2}{a_j} \dfrac{p_{a_j}(b_i + \sqrt{b_i})}{p_{a_j}(b_j - 1)} \dfrac{p_{a_{i+1}}(b_i)}{p_{a_{i+1}}(b_i + \sqrt{b_i})} \\ &\le \textstyle\sum_1^i (b_i^2/a_j)\ (\frac{3}{4})^{b_i + \sqrt{b_i} - b_j} (\frac{1}{2})^{\sqrt{b_i} - 2} \\ &\le 4b_i^2 (\tfrac{3}{8})^{\sqrt{b_i}} + 2b_i^2 (\tfrac{3}{4})^{b_i/2} \le 80. \end{aligned}$$

8. Relation to difference inequality methods for inadmissibility. The present method for establishing inadmissibility is nonconstructive: convergence of $\sum 1/a_x$ and the consequent availability of a solution to ($\mathscr{P}$) do not seem to yield an improved estimator. In view of the extensive recent work on constructing improvements of inadmissible estimates by solving differential (or difference) inequalities (see e.g. Berger, 1980a; Hwang, 1982a; Ghosh, Hwang, Tsui, 1983), it is of interest to connect the two approaches.

For a large class of continuous exponential families with density of the form $e^{\theta x - \psi(\theta)} t(x)\, dx$ and squared error loss for estimating θ, the Euler-Lagrange equation for the Stein-LeCam minimization coincides with the differential equation obtained via Stein's (1973) unbiased estimator of risk. Specifically, if F and G are priors, with $f(x) = \int e^{\theta x - \psi(\theta)}\, dF(\theta)$, $g(x) = \int e^{\theta x - \psi(\theta)} dG(\theta)$, $j^2 = g/f$ and corresponding generalized Bayes estimates $d_F = \nabla f/f$, $d_g = \nabla g/g$, then under appropriate conditions (cf. Brown, 1971, for the normal case):

$$B(d_F, G) - B(d_g, G) = -4 \int j\, \nabla \cdot (ft\, \nabla j)\, dx,$$

while using Hudson (1978) or Lemma 1 of Berger (1980a), one finds that

$$R(d_F, \theta) - R(d_G, \theta) = -4\, E_\theta(1/ftj)\, \nabla \cdot (ft\, \nabla j).$$

Consequently, in the transient case, the solution to the minimization can in principle be used to construct a better estimator.

For discrete exponential families of power series type, the situation is less clear cut. Stein's unbiased estimate of the difference in risk between two estimators was extended to estimation of the parameter of discrete power series families by Hudson (1978), Hwang (1982a) and others. They obtain a difference operator $\mathscr{L}$ such that solutions to $\mathscr{L}\phi_x \le 0$ will allow construction of improved estimators. We show below that the equation $\mathscr{L}\phi_x = 0$ is the Euler-Lagrange equation for the original minimization problem related to the expression (3.3), whereas (1.6) is the Euler-Lagrange equation of the "linearized" expression (1.3).

Thus the simpler equation suffices to resolve the issue of admissibility for a given estimator d_P; but to proceed further to construction of a better estimate, the more complex operator (8.3) (or equivalently) (8.2) is required.

Using the notation of Sections 3 and 5, with the discrete exponential family density $f(x \mid \theta) = \theta^x \phi(\theta) t_x$, we recall that

$$
\begin{aligned}
B_r(d_P, Q) - B_r(Q) &= \sum_{r+1}^{\infty} \left(\frac{q_x}{p_x} - \frac{q_{x-1}}{p_{x-1}} \right)^2 \frac{p_x^2}{q_{x-1}^2} q_{x-1} t_x \\
&= \sum_{r+1}^{\infty} \frac{(Dh_x)^2}{h_{x-1}} a_x = T(h), \quad \text{say.}
\end{aligned} \tag{8.1}
$$

Here $h_x = q_x/p_x$. Note that we do not write $h = j^2$ or use the inequality $1 + j_x/j_{x-1} \geq 1$. Suppose that d_P is inadmissible (so that $\inf_{h \in \mathscr{H}_r} T(h) > 0$), and that the infimum is attained by some function $h^\circ \in \mathscr{H}_r$. Calculating the Euler-Lagrange equation by considering $\lim_{\varepsilon \to 0} [T(h + \varepsilon k) - T(h)]/\varepsilon$ as in the argument before (4.1), we find that h° satisfies the quasilinear elliptic equation

$$
\mathscr{L}_1 h_x = 2D\left(\frac{a_{x+1}}{h_x} Dh_{x+1} \right) + a_{x+1} \frac{(Dh_{x+1})^2}{h_x^2} = 0, \quad x \geq r + 1. \tag{8.2}
$$

Hwang (1982) considers the difference in risk between two estimators $\bar{d}$ and d (which we assume to lie in $\mathscr{D}_r$) by writing $d_x = \bar{d}_x + r_x \phi_x$, where r_x depends on $\bar{d}$, and ϕ_x is to be determined. For the one-dimensional case with loss $\theta^{-1}(d - \theta)^2$, his results specialize to give

$$
\begin{aligned}
&R(\theta, d) - R(\theta, \bar{d}) \\
&\quad = E_\theta[2(\bar{d}_x - d_x) + (d_{x+1}^2 - \bar{d}_{x+1}^2)(t_{x+1}/t_x)] + (d_0^2 - \bar{d}_0^2) P_\theta(X = 0) \\
&\quad = E_\theta(\varepsilon_x + \mathscr{L}_x) + (d_0^2 - \bar{d}_0^2) P_\theta(X = 0),
\end{aligned}
$$

where $\varepsilon_x = 2\{\bar{d}_{x+1} r_{x+1}(t_{x+1}/t_x) - r_x\} \phi_{x+1}$, and

$$
\mathscr{L}_x = 2 r_x D\phi_{x+1} + r_{x+1}^2 (t_{x+1}/t_x) \phi_{x+1}^2. \tag{8.3}
$$

One chooses r_x recursively to make ε_x zero and then seeks a solution ϕ_x to the now first order difference inequality $\mathscr{L}\phi \leq 0$.

To connect this with the minimization problem, take $\bar{d} = d_P = p_x/p_{x-1}$, $d = d_Q = q_x/q_{x-1}$ for $x \geq r + 1$, with both equal to zero for $x \leq r$. Solving $\varepsilon_x = 0$ for $x \geq r$ then implies that $r_x = c/t_x p_x$ for some constant c, (which is set to 1) and hence that

$$
\phi_x = \frac{1}{r_x} (d_x - \bar{d}_x) = p_x t_x \left(\frac{q_x}{q_{x-1}} - \frac{p_x}{p_{x-1}} \right) = a_x \frac{Dh_x}{h_{x-1}}
$$

for $x \geq r + 1$. Therefore

$$
\begin{aligned}
\mathscr{L}_x &= \frac{1}{p_x t_x} \left[2\, D\left(a_{x+1} \frac{Dh_{x+1}}{h_x} \right) + a_{x+1} \frac{(Dh_{x+1})^2}{h_x^2} \right] \\
&= \frac{1}{p_x t_x} \mathscr{L}_1 h_x \quad \text{for} \quad x \geq r + 1.
\end{aligned}
$$

It is easily checked that $\varepsilon_x + \mathscr{L}_x$ equals 0 for $x < r$ and equals

$$[d_Q^2(r+1) - d_P^2(r+1)][t_{r+1}/t_r] \quad \text{for} \quad x = r.$$

Thus, if h_x solves the Euler-Lagrange equation (or inequality with ≤ 0) for $x \geq r+1$ and $h_{r+1} \leq h_r$, then the estimator $d = q_x/q_{x-1} I_{\{x>r\}}$ is as good as $\bar{d}$. It is better if strict inequality occurs somewhere; alternatively, if $\mathscr{L}_1 h \equiv 0$ and $h \not\equiv$ constant, then the estimate $d' = (d + \bar{d})/2$ will strictly dominate $\bar{d}$.

Let us note that treating (8.2) as a quadratic in Dh_{x+1} leads to the following recursion for a solution:

$$\frac{Dh_{x+1}}{h_x} = -1 + \left(1 + \frac{2a_x}{a_{x+1}} \frac{Dh_x}{h_{x-1}}\right)^{1/2}, \quad x \geq r+1$$

and this solution will be nonnegative and decreasing (if it is well defined).

APPENDIX

Poincaré and other inequalities for Poisson measures on $\mathbb{Z}$

As a preliminary, we record a Poisson version of the simplest bound on Mill's ratio for the normal distribution (Feller, 1968, page 175).

LEMMA A1. *Let $P(s) = P_\lambda(X \leq s)$ and $\tilde{P}(s) = P_\lambda(X \geq s)$.*

If $s \leq \lambda$, then $P(s) \leq (s+1)p_\lambda(s) \leq (\lambda+1)p_\lambda(s)$.

If $s \geq \lambda$, then $\tilde{P}(s) \leq (s+1)p_\lambda(s)$.

PROOF. The first statement follows from the inequality $\lambda^x/x! \leq \lambda^s/s!$ for $x \leq s \leq \lambda$. For the second, note that for $s \geq \lambda + 1$, $\tilde{P}(s) - \tilde{P}(s+1) \leq (s+1)p_\lambda(s) - (s+2)p_\lambda(s+1)$. Finally, if $s = \bar{\lambda}$,

$$\frac{\tilde{P}(\bar{\lambda})}{p_\lambda(\bar{\lambda})} = \sum_{t=0}^{\infty} \frac{\lambda^t}{(\bar{\lambda}+1)\cdots(\bar{\lambda}+t)} \leq \bar{\lambda} + 1. \quad \square$$

PROOF OF LEMMA 5.1. Let $k(x, \lambda) = 1$, x/λ or $(x-\lambda)^2/\lambda$. From the Cauchy-Schwartz inequality, if $x > \bar{\lambda}$, then $[u(x) - u(\bar{\lambda})]^2 \leq (x - \bar{\lambda}) \sum_{\bar{\lambda} \leq s < x} [D^+ u_s]^2$. Thus

$$\text{(A.1)} \quad \sum_{x>\bar{\lambda}} [u_x - u_{\bar{\lambda}}]^2 k(x,\lambda) p_\lambda(x) \leq \sum_{s \geq \bar{\lambda}} [Du_s^+]^2 \sum_{x>s} (x - \bar{\lambda}) k(x,\lambda) p_\lambda(x).$$

The inner sum is evaluated using "summation by parts" identities for the Poisson distribution, based on the relation $x p_\lambda(x) = \lambda p_\lambda(x-1)$. Let $\tilde{P}(s) = \sum_{x \geq s} p_\lambda(x)$; then for all $s \geq 0$, $\lambda > 0$,

$$\begin{aligned} \sum_{s+1}^{\infty} (x-\lambda) p_\lambda(x) &= \lambda p_\lambda(s) \\ \text{(A.2)} \qquad \sum_{s+1}^{\infty} (x-\lambda)^2 p_\lambda(x) &= \lambda p_\lambda(s)[s-\lambda] + \lambda \tilde{P}(s) \\ \sum_{s+1}^{\infty} (x-\lambda)^3 p_\lambda(x) &= \lambda p_\lambda(s)[(s-\lambda)^2 + 2s] + \lambda \tilde{P}(s). \end{aligned}$$

When taken instead over $x < \overline{\lambda}$, the sum on the left side of (A.1) is bounded by

$$\sum_{0 \le s < \overline{\lambda}} [Du_s^+]^2 \sum_{0 \le x \le s} (\overline{\lambda} - x) k(x, \lambda) p_\lambda(x).$$

To evaluate the inner sum here, convert the sums in (A.2) to the range $0 \le x \le s$ by using the relations $E(x - \lambda) = 0$, $E(x - \lambda)^2 = E(x - \lambda)^3 = \lambda$. In the case $k(x, \lambda) \equiv 1$, combining the two bounds leads to

$$E[u(x) - u(\overline{\lambda})]^2 = \lambda E[Du^+(x)]^2 + (\lambda - \overline{\lambda}) \sum_{s \ge \overline{\lambda}} [Du_s^+]^2 \tilde{P}(s + 1)$$
$$+ (\overline{\lambda} - \lambda) \sum_{s < \overline{\lambda}} [Du_s^+]^2 P(s),$$

where $P(s) = \sum_0^s p_\lambda(x)$. The second term on the right side is negative, and the third is bounded by $(1 + \lambda)E[Du^+(x)]^2$ from the first part of Lemma A.1. When $k(x, \lambda) = x/\lambda$, a similar argument applies, using instead the second part of Lemma A.1.

Proceeding analogously with $k(x, \lambda) = (x - \lambda)^2/\lambda$, one finds that

$$E(x - \lambda)^2/\lambda [u(x) - u(\overline{\lambda})]^2$$
$$\le \lambda E[(x - \lambda)^2/\lambda + 2x/\lambda][Du_x^+]^2 + \sum_{s \ge \overline{\lambda}} [Du_s^+]^2 \tilde{P}(s)$$
$$+ \sum_{s < \overline{\lambda}} [Du_s^+]^2 (\overline{\lambda} - \lambda)[p_\lambda(s)(\lambda - s) + P(s - 1)].$$

This leads to advertised bound if one replaces $\lambda - s$ by the upper bound $(1 + \lambda)[(s - \lambda)^2/\lambda + s/\lambda]$.

Finally, Proposition 7.1 is a corollary of some simple inequalities for the gamma function $\Gamma(x) = \int_0^\infty e^{-t} t^{z-1}\, dt$. Let $\psi(z) = (d/dt)\log \Gamma(z)$ be the digamma function.

LEMMA A.2.
(i) $z^a \le \Gamma(z + a)/\Gamma(z) \le (z + a - 1)^a$ *for* $a \ge 1$,
(ii) $\Gamma(z + a)/\Gamma(z) \ge e^{\psi(1)} z^a$ *for* $0 \le a \le 1, z \ge 1$.

PROOF. Let $H_z(a) = \log \Gamma(z + a) - a \log z - \log \Gamma(z)$, so that $H_z'(a) = \psi(z + a) - \log z$. Using the integral representations (Lebedev, 1972, page 6–7),

$$\psi(z) = \int_0^\infty \left(\frac{e^t}{t} - \frac{e^{-tz}}{1 - e^{-t}} \right) dt, \quad \log z = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-tz}}{t} \right) dt, \tag{A.3}$$

so that

$$H_z'(a) = \int_0^\infty \left(\frac{1}{t} - \frac{e^{-ta}}{1 - e^{-t}} \right) e^{-tz}\, dt. \tag{A.4}$$

Since $H_z(1) = 0$ and from (A.4) $H_z'(a) \ge 0$ for $a \ge 1$, the left side of (i) follows. For $0 \le a \le 1$ and $z \ge 1$,

$$H_z(a) = \int_0^a H_z'(b)\, db \ge a \int_0^\infty \left(\frac{1}{t} - \frac{1}{1 - e^{-t}} \right) e^{-tz}\, dt \ge \psi(1),$$

which yields (ii). Finally, the right side of (i) is easy to establish in a similar vein, using the identity $\psi(z + 1) = \psi(z) + 1/z$.

Acknowledgements. My profound thanks go to Larry Brown for continued expert and encouraging discussion of this work as it developed. Roger Farrell, Juinn Hwang, David Pollard and the referees made valuable suggestions on both form and content.

REFERENCES

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

BERGER, J. (1980b). *Statistical Decision Theory*. Springer-Verlag, New York.

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

BROWN, L. D. (1976). Lecture notes on statistical decision theory. Unpublished manuscript. Cornell University, Ithaca, N.Y.

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

BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9** 1289–1300.

BROWN, L. D. and FARRELL, R. H. (1985). Complete class theorems for estimation of multivariate Poisson means and related problems. *Ann. Statist.* **13**, to appear.

BROWN, L. D. and HWANG, J. T. (1982). A unified admissibility proof. *Proc. 3rd Purdue Symposium Statistical Decision Theory*, III. S. S. Gupta and J. O. Berger (eds.). Academic, New York.

CLEVENSON, M. and ZIDEK, J. (1975). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705.

DOYLE, P. and SNELL, J. L. (1981). *Random Walk and Electrical Networks*. To appear as a UMAP publication.

DYNKIN, E. B. and YUSHKEVICH, A. A. (1969). *Markov Processes; Theorems and Problems*. Plenum Press, New York.

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

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*. Vol. I 3rd ed. Wiley, New York.

FUKUSHIMA, M. (1980). *Dirichlet Forms and Markov Processes*. North-Holland, Amsterdam.

GHOSH, M., HWANG, J. T. and TSUI, K. W. (1983). Construction of improved estimators in multiparameter estimation for discrete exponential families. *Ann. Statist.* **11** 351–367.

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

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

GUTTMAN, S. (1983). Decisions immune to Stein's effect. *Sankhyā*, to appear.

HARRIS, T. E. (1952). First passage and recurrence distributions. *Trans. Amer. Math. Soc.* **73** 471–486.

HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.

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

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

IGLEHART, D. L. (1974). Weak convergence in applied probability. *Stochastic Process. Appl.* **2** 211–241.

JAMES, W. and STEIN C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379.

JOHNSTONE, I. (1984). Admissible simultaneous estimation of Poisson means, Dirichlet principles and recurrence of birth-death chains in $\mathbb{Z}_+^p$. *Z. Wahrsch. verw. Gebiete*, to appear.

JOHNSTONE, I. and LALLEY, S. (1984). On independent statistical decision problems and products of diffusions. *Z. Wahrsch. verw. Gebiete*, to appear.

JOHNSTONE, I. and SHAHSHAHANI, M. (1983). A Poincaré-type inequality for solutions of linear elliptic equations, with statistical applications. Unpublished.

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

KARLIN, S. and TAYLOR, H. M. (1975). *A First Course in Stochastic Processes.* 2nd ed. Academic, New York.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

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

PENG, J. C. M. (1975). Simultaneous estimation of the parameters of independent Possson distributins. Technical Report 78, Dept. of Statistics, Stanford Univ.

SRINIVASAN, C. (1981). Admissible generalized Bayes estimates and exterior boundary value problems. Purdue University, Dept. of Statistics. Mimeograph series No. 78-2. *Sankhyā A* 1–25.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206.

STEIN, C. (1973). Estimation of the mean of a multivariate normal distribution. In *Proc. Praque Symp. Asymptotic Statist.* 345–381.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305