The Annals of Statistics
1992, Vol. 20, No. 2, 807–831

MINIMAX ESTIMATION OF A CONSTRAINED POISSON VECTOR

BY IAIN M. JOHNSTONE AND K. BRENDA MACGIBBON

Stanford University and Université du Québec à Montréal

Suppose that the mean τ of a vector of Poisson variates is known to lie in a bounded domain T in $[0, \infty)^p$. How much does this a priori information increase precision of estimation of τ? Using error measure $\sum_i(\hat{\tau}_i - \tau_i)^2/\tau_i$ and minimax risk $\rho(T)$, we give analytical and numerical results for small intervals when $p = 1$. Usually, however, approximations are needed. If T is "rectangulary convex" at 0, there exist *linear* estimators with risk at most $1.26\rho(T)$. For general T, $\rho(T) \geq p^2/(p + \lambda(\Omega))$, where $\lambda(\Omega)$ is the principal eigenvalue of the Laplace operator on the *polydisc transform* $\Omega = \Omega(T)$, a domain in *twice-p*-dimensional space. The bound is asymptotically sharp: $\rho(mT) = p - \lambda(\Omega)/m + o(m^{-1})$. Explicit forms are given for T a simplex or a hyperrectangle. We explore the curious parallel of the results for T with those for a Gaussian vector of *double* the dimension lying in Ω.

1. Introduction. In many estimation settings, there is definite prior information concerning the values of a parameter vector τ. There may be bounds on the individual components τ_i—"all τ_i lie between 0 and 1"—or on particular functionals of the whole vector—"the sum of τ_i is at most c" or "most τ_i are zero." Many estimation methods have been developed to capitalize on such information, positivity-constrained least squares and maximum entropy being just two examples.

How does one compare the performance of various possible estimators when such prior information is present? One common, admittedly conservative, approach is the worst-case analysis: Given some error measure, compute the maximum expected error over the restricted parameter space, and then seek the estimator that minimizes this maximum risk. The resulting best or minimax risk provides (i) a benchmark against which to measure other estimators and (ii) a measure of the value of the prior information (by comparison with the minimax risk computed ignoring the prior information).

For Gaussian data, a considerable literature has recently arisen relating such minimax risks to the size and shape of the constraints and to the structure of the loss function. Some references are given in Section 6. This paper and its companion, Johnstone and MacGibbon (1990) (hereafter called II) consider some corresponding questions for Poisson models. As prototypical discrete data settings, these deserve study in their own right. In fact, Aldous

Received September 1989; revised January 1991.
AMS 1980 *subject classifications*. Primary 62F10; secondary 62F11, 62C20.
Key words and phrases. Polydisc transform, Bayes risk lower bound, second-order minimax, Laplace operator, principal eigenvalue, loss estimation, Fisher information, minimax linear risk, minimax risk, hardest rectangular subproblem, isoperimetric inequalities.

[(1989), page 17] even argues that Poisson approximations arise in more contexts than Gaussian approximations.

Here is an evocative, but hardly exhaustive, list of count data settings in which constraint information is available.

1. Mixture problems in which the proportions of various components are sought and bounds on the total quantity of material are available; for example, in various modes of spectroscopy.
2. Spatial process settings in which a bound on the total possible intensity of the process over a set A may be known, but the distribution of the intensities amongst members A_j of a partition of A may be unknown; for example, metabolite distribution in emission tomography.
3. Thinning problems, in which Poisson processes of known rates are subjected to an unknown degree of thinning before observation, and it is desired to estimate the thinning fractions; for example, in shielding of radioactive sources.
4. Sparse signal settings in which it is known that only a small fraction of the components of the parameter vector will be nonzero, as, for example, in star maps.

To be specific, suppose that X is a random vector whose p components X_i independently follow Poisson (τ_i) distributions. We consider estimation of $\tau = (\tau_1, \dots, \tau_p)$ with the prior information that $\tau \in T$, a subset of $R_+^p = \{\tau \in R^p\colon \tau_i \geq 0\ \forall\ i\}$.

We use a quadratic loss function which has been normalized by Fisher information $I(\tau) = \operatorname{diag}(\tau_i^{-1})$; namely,

$$L(\delta, \tau) = \sum_i (\delta_i - \tau_i)^2/\tau_i = (\delta(X) - \tau)^T I(\tau)(\delta(X) - \tau).$$

This loss function is a commonly used compromise between mathematical tractability and statistical relevance; it is discussed further in Section 6. The expected error or (frequentist) risk is $R(\delta, \tau) = E_\tau L(\delta(X), \tau)$. As a measure of the information in the experiment, we study the minimax risk

$$\rho(T) = \inf_\delta \sup_{\tau \in T} R(\delta, \tau) \tag{1}$$

in both finite and asymptotic settings. If $T = R_+^p$, then the minimax risk equals p, the number of variables observed. So our main question becomes: How much does restriction to a *subset* $T \subset R_+^p$ reduce the minimax risk and what is an (approximately) minimax estimator?

Unfortunately, exact analytic description of the minimax rule and risk is generally intractable, since the minimax rule is a Bayes estimator for a prior F^*, the *least favorable distribution*, which is necessarily concentrated on a meager set of complicated form. Some analytic work is possible in the simplest cases, where the support of F^* consists of a small number of points (Section 2). However, the bulk of this paper is concerned with deriving approximate information of various kinds. These may be grouped under the following headings:

1. Numerical calculation (chiefly for $p = 1$) (Section 2).
2. Simpler classes of estimators (Section 3).
3. (Lower) bounds (Section 4).
4. Asymptotic approximations (II).

When T is a small or moderate sized interval $[0, m)$ in R_+, or more generally a hyperrectangle in R_+^p, it is possible to generate the least favourable distribution, minimax rule and minimax value numerically; explicit details are presented in Section 2.

The simplicity of estimators that are linear (strictly speaking, *affine* linear) in the data makes it natural to ask how much is lost by restricting attention to this class. Explicit calculations are possible for such estimators, and often there is not much loss in efficiency; details are given in Section 3.

For larger rectangles and more general sets, approximations appear necessary, either via lower bounds or asymptotics. In each case, a fundamental role is played by a many-to-one mapping $\tau\colon R^{2p} \to R_+^p$,

$$(2) \qquad \tau\colon (\omega_1, \omega_2, \dots, \omega_{2p-1}, \omega_{2p}) \to \left(\omega_1^2 + \omega_2^2, \dots, \omega_{2p-1}^2 + \omega_{2p}^2\right).$$

Note that τ denotes both the function $\tau(\omega)$ and a point in its image, T, the Poisson mean parameter space. We call the set $\Omega = \tau^{-1}(T)$ the *polydisc transform* of T. The name reflects the fact that the transform of a rectangle $[0, a] \subset R_+^p$, namely

$$\left\{\omega\colon \omega_{2i-1}^2 + \omega_{2i}^2 \le a_i,\, i = 1, \dots, p\right\},$$

is termed a polydisc in function theory.

The inverse mapping τ^{-1} is a "dimension-doubling" version of the traditional square-root variance stabilising transformation for Poisson data. It is explained further in Section 4. The virtue of the polydisc transform is that it converts relatively unpleasant optimization problems for T into the well understood Dirichlet problem for the Laplace equation on Ω. For example, Section 4 gives lower bounds for $\rho(T)$ in terms of the minimal eigenvalue of the Laplacian on Ω.

An asymptotic theory is obtained by approximating $\rho(mT)$ as $m \to \infty$. If the variables X_i in the original setting are obtained from observing a Poisson process for a certain time, the asymptotic formulation corresponds to long observation times on the process. The polydisc lower bound derived in Section 4 is sharp to second order asymptotically. This result is stated informally in Section 5, with formal statement and proof given in II.

Section 5 also evaluates the polydisc transforms and lower bounds in a number of important examples, and discusses the connection with isoperimetric inequalities. Finally Section 6 contains references to the literature and further remarks and Section 7 contains proofs. We conclude this section by collecting notation and definitions for later use.

NOTATION. $\mathbf{Z}_+^p = \{n \in \mathbf{Z}^p: n_i \geq 0 \text{ for all } i\}$. Derivatives are denoted by D_i: $D_i u = (\partial/\partial x_i)u(x)$, or D_{x_i} when the variable of differentiation is shown explicitly. If $\psi = (\psi_1, \dots, \psi_p)$ is a vector field, $D \cdot \psi = \Sigma_i D_i \psi_i$.

Let $X \subset R^p$. Then $C_0^k(X, R^p)$ denotes the space of k-times continuously differentiable functions defined on and having compact support in X (in the relative topology of X) and taking values in R^p. Often this is written simply as $C_0^k(X)$, or as C_0^k when $X = R^p$.

DEFINITIONS. (i) We assume throughout that T is relatively open in R_+^p: T equals the intersection with R_+^p of some open set in R^p. We call T a domain if it is R_+^p open and connected. Since the continuity of risk functions ensures that $\rho(T) = \rho(\overline{T})$, we may and shall by convention choose T so that $T = \operatorname{int} \overline{T}$.

(ii) The ith face of R_+^p, $\mathscr{F}_i = \{\tau \in R_+^p: \tau_i = 0\}$, is a *critical face* for T if $\overline{T}$ intersects $\mathscr{F}_i$. Denote by $\mathbf{I} = \mathbf{I}(T) \subset \{1, \dots, p\}$ the set of indices of critical faces. Throughout the paper, we restrict attention to the class of estimators

$$\mathbf{D} = \mathbf{D}(T) = \{\delta(x): x_i = 0 \text{ and } i \in \mathbf{I}(T) \text{ implies } \delta_i(x) = 0\}, \tag{3}$$

since estimators not in $\mathbf{D}$ are easily seen to have infinite maximum risk.

(iii) For a (prior) probability distribution $F(d\tau)$, define the integrated risk $r(\delta, F)$ and the Bayes risk $r(F)$ by

$$r(\delta, F) = \int R(\delta, \tau) F(d\tau), \qquad r(F) = \inf_{\delta \in D} r(\delta, F). \tag{4}$$

(iv) Let $\mathbf{F}^*(X)$ denote the collection of probability measures supported in X. According to the minimax theorem,

$$\rho(T) = \inf_{\mathbf{D}} \sup_{T} R(\delta, \tau) = \sup_{\mathbf{F}^*(\overline{T})} \inf_{\mathbf{D}} r(\delta, F) = \sup_{\mathbf{F}^*(\overline{T})} r(F). \tag{5}$$

A prior distribution attaining the supremum is called least favorable for T. When $\overline{T}$ is compact, least favorable distributions exist.

2. Analytical and numerical results for $p = 1$. When $p = 1$, it is feasible to determine the minimax risk and rules by explicitly constructing the least favorable distribution. Consider an interval $\overline{T} = [m_1, m_2]$ with $0 \leq m_1 < m_2 < \infty$. Concavity arguments given in II show that the least favorable distribution F_{m_1, m_2} is unique, and analyticity considerations imply that it is supported on a finite number of points in $[m_1, m_2]$. Denote by ε_b the probability measure concentrated at $\tau = b$. Let $\mathscr{F}_k[m_1, m_2]$ be the class of distributions of the form $F = F(a, b) = \Sigma_{i=1}^k a_i \varepsilon_{b_i}$, where $\{b_i\} \subset [m_1, m_2]$ and $\{a_i\}$ are probability masses summing to 1. The Bayes rule δ_F associated to such a prior is given by

$$\delta_F(x) = \frac{1}{E(\tau^{-1} | X = x)} = \frac{\Sigma_{i=1}^k a_i b_i^x e^{-b_i}}{\Sigma_{i=1}^k a_i b_i^{x-1} e^{-b_i}} \tag{6}$$

with the convention (3) requiring that $\delta_F(0) = 0$ if $m_1 = 0$. The least favorable distribution F_{m_1, m_2} belongs to $\mathscr{F}_k[m_1, m_2]$ for k large enough. The optimization problem is then to choose $a_1, \dots, a_k, b_1, \dots, b_k$, constrained as above, so as to maximize

$$r_k(a, b) = r(F(a, b)) = \sum_{i=1}^{k} a_i R(\delta_F, b_i).$$

For identifying the maximum, we recall an important and familiar criterion:

LEMMA 1. *If the support of a prior $F(d\tau)$ is contained in the set at which $R(\delta_F, \tau)$ achieves its maximum on $[m_1, m_2]$, then F is least favorable and δ_F is minimax.*

Analytical descriptions for $[0, m]$, m *small.* When m is sufficiently small, it is plausible that the least favorable distribution $F_m := F_{0,m}$ on $[0, m]$ would be given by a point mass at m. For the one point prior $F = \varepsilon_m$, from (3) and (6), the corresponding Bayes rule $\delta_F(x) = mI\{x > 0\}$. The risk function

$$R(\delta_F, \tau) = \tau e^{-\tau} + \tau^{-1}(\tau - m)^2(1 - e^{-\tau})$$

takes values m^2 and me^{-m} at 0 and m, respectively. Let $m_0 \approx 0.57$ be the solution to $m = e^{-m}$: For $0 < m \le m_0$, $R(\delta_F, \tau)$ takes its maximum at the endpoints of $[0, m]$, and so Lemma 1 shows that δ_F is minimax.

For slightly larger m, one expects a second support point to appear at 0. For the two point prior $a\varepsilon_0 + (1 - a)\varepsilon_m$, the Bayes rule has

$$\delta_F(x) = \begin{cases} 0, & \text{if } x = 0, \\ m(1 - a)e^{-m}/[a + (1 - a)e^{-m}], & \text{if } x = 1, \\ m, & \text{if } x \ge 2. \end{cases} \tag{7}$$

The risk function of δ_F is again available explicitly:

$$R = R(\delta_F, \tau) = e^{-\tau}\left[\tau + (\delta_1 - \tau)^2 + (m - \tau)^2\tau^{-1}(e^{\tau} - 1 - \tau)\right].$$

If δ_F is to be minimax, then necessarily $R(\delta_F, 0) = R(\delta_F, m)$ and this forces

$$\delta_{F_m}(1) = (e^m - 1)^{-1}\left[-m + (m^2e^m + me^m - m)^{1/2}\right]. \tag{8}$$

Along with (7), this determines the value of a. In the Appendix, we verify that this prior is least favorable for $m_0 < m \le m_1$, where $m_1 \approx 1.27$ is the first *positive* zero of the equation

$$(1 + \delta_F(m))^2 = 2 + m^2/2. \tag{9}$$

In summary, we have:

THEOREM 2. *If $0 < m \le m_0 \approx 0.57$, the least favorable distribution $F_m = \varepsilon_m$, $\delta_{F_m}(x) = mI\{x > 0\}$ and $\rho([0, m]) = me^{-m}$. If $m_0 \le m \le m_1 \approx 1.27$, the least favorable distribution $F_m = a_m\varepsilon_0 + (1 - a_m)\varepsilon_m$, δ_{F_m} is given by (7) and $\rho([0, m]) = \delta^2_{F_m}(1)$ [cf. (8)].*

TABLE 1
Least favorable distributions and minimax risk on $[0, m]$ *as a function of* m*

m	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4	$\rho_N(m)$ ($\geq$)	$\rho_L(m)$	ρ_L/ρ ($\leq$)
0.100	1.000				1				0.090	0.091	1.005
0.200	1.000				1				0.164	0.167	1.018
0.300	1.000				1				0.222	0.231	1.038
0.600	0.976	0.024			1	0.0			0.330	0.375	1.138
0.900	0.851	0.149			1	0.0			0.396	0.474	1.197
1.200	0.807	0.193			1	0.0			0.447	0.545	1.221
1.500	0.738	0.262			1	0.080			0.483	0.600	1.241
1.800	0.673	0.327			1	0.133			0.515	0.643	1.249
1.900	0.655	0.345			1	0.143			0.524	0.655	1.250
2.000	0.640	0.360			1	0.152			0.533	0.667	1.251
2.100	0.626	0.374			1	0.159			0.542	0.677	1.251
2.200	0.613	0.387			1	0.164			0.550	0.688	1.250
2.300	0.602	0.398			1	0.169			0.558	0.697	1.250
2.400	0.592	0.408			1	0.173			0.565	0.706	1.249
2.700	0.561	0.421	0.018		1	0.195	0.0		0.585	0.730	1.248
3.000	0.526	0.422	0.052		1	0.231	0.0		0.603	0.750	1.245
3.500	0.480	0.441	0.078		1	0.264	0.0		0.628	0.778	1.238
4.000	0.448	0.461	0.091		1	0.281	0.0		0.650	0.800	1.230
4.500	0.413	0.435	0.152		1	0.321	0.040		0.669	0.818	1.223
5.000	0.383	0.425	0.192		1	0.349	0.056		0.685	0.833	1.216
6.000	0.339	0.420	0.237	0.005	1	0.384	0.073	0.0	0.713	0.857	1.202
7.000	0.301	0.392	0.265	0.043	1	0.423	0.113	0.0	0.735	0.875	1.190
8.000	0.273	0.380	0.289	0.058	1	0.447	0.131	0.003	0.754	0.889	1.179
10.000	0.226	0.336	0.295	0.143	1	0.497	0.190	0.035	0.783	0.909	1.160
11.500	0.202	0.322	0.310	0.165	1	0.518	0.208	0.039	0.801	0.920	1.149

*Least favorable distribution has the form $F_m = \sum_{i=1}^{k} a_i \varepsilon_{\beta_i m}$, where for $m \leq 11.5$, $k \leq 4$. Note that $1 = \beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$ and that the corresponding probability masses a_i sum to 1. $\rho(m)$ gives a (lower) bound to the minimax risk, $\rho_L(m) = m/(1 + m)$ gives the linear minimax risk.

Numerical descriptions for $[0, m]$, *m moderate.* As m increases further, extra support points are necessary between 0 and m. As mentioned above, the least favourable distribution F_m and hence δ_{F_m} and $\rho(m) = \rho([0, m])$ can be determined numerically since an optimization over a fixed number of support locations and masses is required. Since F_m and $\rho(m)$ depend continuously on m (Lemma 10 in the Appendix), the optimization can be done incrementally. Since a rule is minimax if and only if the atoms of the prior are contained in the set where the resulting Bayes rule achieves its maximum risk, we have a check (up to numerical accuracy) that a candidate prior F is indeed least favourable.

Selected results of the numerical optimization are displayed in Table 1 and in Figure 1. A fuller version of Table 1 may be found in Johnstone and MacGibbon (1990). The figure shows the variation with m in the support set of the least favourable distribution and in the prior masses attached to each such support point. Note the generally continuous dependence of the least favorable distribution on the interval $[0, m]$. Exceptions occur at certain

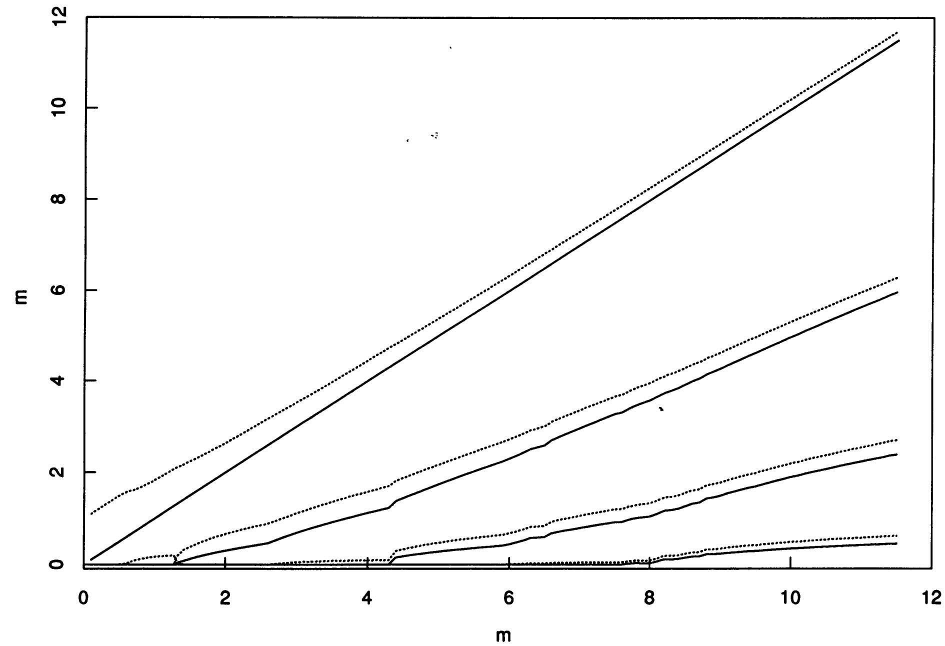

FIG. 1. *Evolution of least favorable distributions on* $[0, m]$ *as* m *increases. Solid lines indicate (as a function of* m*) the location of the support points. Vertical distance from solid line to adjacent dotted line equals prior probability mass at that support point (as a function of* m*).*

critical values at which new support points appear in order to pin down the risk of the minimax rule. The support points become more widely separated for larger values of τ: The variance stabilising transformation would produce a more regular spacing. The asymptotic results for large m of II, summarised in Section 5, show that this spacing would *not* be uniform.

3. Linear versus nonlinear estimators.

3.1. *Rectangularly convex sets containing* 0. We begin with intervals $[0, m] \subset R^1_+$. The minimax risk amongst *linear* rules $\delta(x) = a + bx$ is easily found to be $\rho_L(m) = m/(1 + m)$ and is attained by $\delta(x) = mx/(m + 1)$. The subscript N added to $\rho(T)$ emphasizes the distinction between linear and nonlinear rules. The function $\mu(m) = \rho_L(m)/\rho_N(m)$ is continuous on $(0, \infty)$ and approaches 1 as $m \to 0$ and ∞, so that $\mu^* = \sup_m \mu(m) < \infty$. For the Gaussian case with squared error loss, this phenomenon was noted by Ibragimov and Has'minskii (1984).

Table 1 shows the ratio $\mu(m) = \rho_L(m)/\rho_N(m)$, which may be thought of as the cost due to restricting attention to the linear subclass (cf. also Figure 2). The table suggests that the ratio never exceeds $\mu^* \doteq 1.251$. [This sort of numerically derived upper bound could be verified rigorously along the lines of

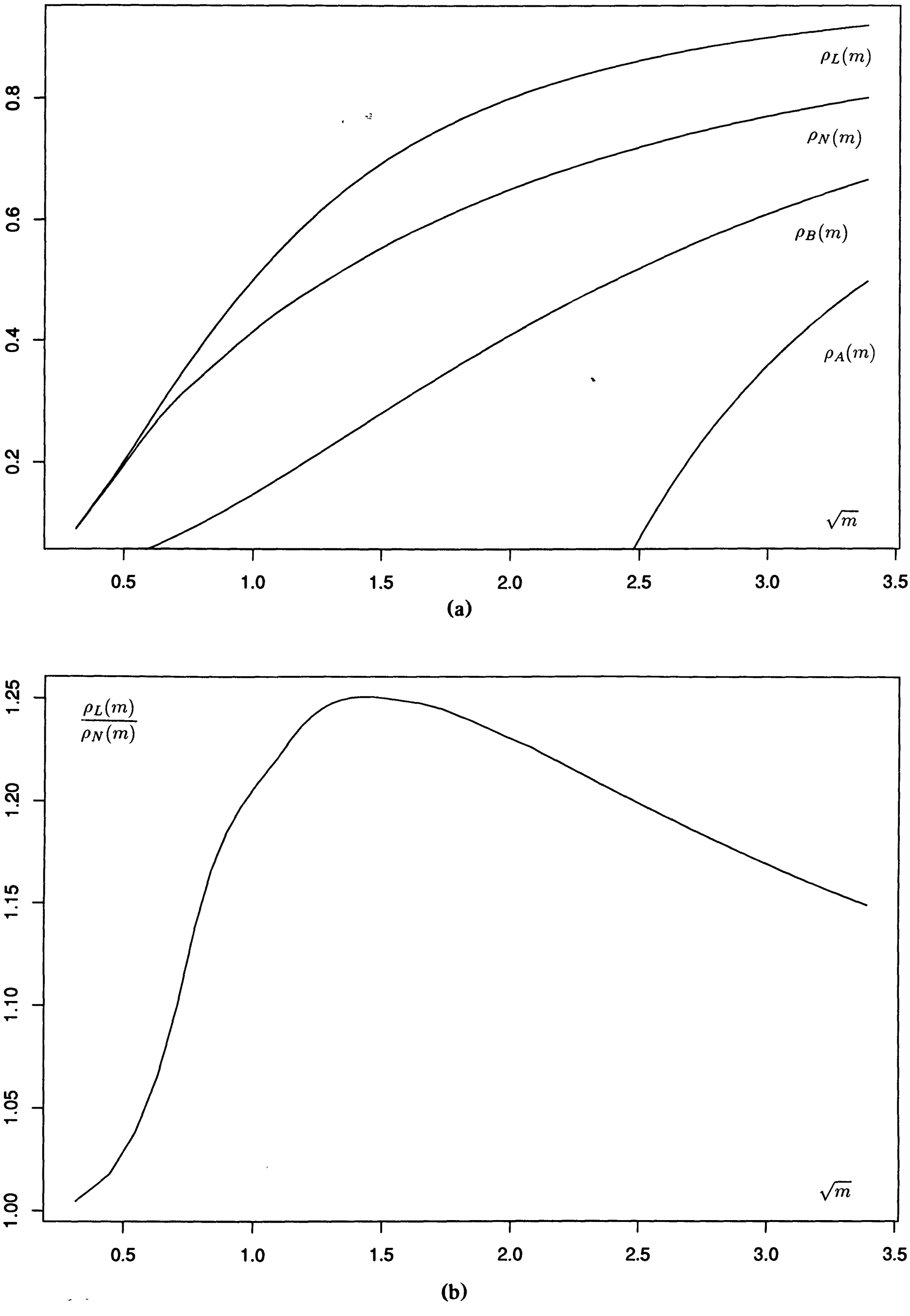

FIG. 2. (a) *Linear* $\rho_L(m)$, *nonlinear* $\rho_N(m)$, *lower bound* $\rho_B(m) = 1/(1 + 5.783/m)$ *and asymptotic* $[\rho_A(m) \sim 1 - 5.783/m]$ *approximations to minimax risk for* $T = [0, m]$ *as a function of* $\sqrt{m}$. (b) *Ratio of linear to nonlinear minimax risks* $\rho_L(m)/\rho_N(m)$.

the extensive appendix to Donoho, Liu and MacGibbon (1988). In the interests of conserving space, we do not do this here.] It is remarkable that the bound on the ratio of the linear minimax risk to the minimax risk for such intervals appears to be almost (but *not* exactly) equal to the numerical value of 1.247 of the analogous bound for Gaussian intervals found by Donoho, Liu and MacGibbon (1990). In fact, using global decomposition optimization techniques, Gourdin, Jaumard and MacGibbon (1990) showed that μ^* in the Poisson problem is contained in the interval [1.250726, 1.250926], while the Gaussian constant μ^* is contained in [1.246408, 1.246805].

Since both minimax linear and minimax risks for hyperrectangles $[0, \tau]$ equal the sum of the corresponding coordinatewise risks, it follows immediately for such hyperrectangles that the ratio of minimax linear to minimax risks is bounded by μ^*. Clearly this holds also for *infinite*-dimensional hyperrectangles of the same form for which the minimax linear risk is finite.

The bound μ^* actually applies to wider classes of sets. Call a set $T \in R^p$ *rectangularly convex at* 0 if whenever $\tau \in T$, the hyperrectangle $[0, \tau] \subset T$ also. Examples of such rectangularly convex sets include l^q bodies

$$T_a = \left\{ \tau \geq 0 \colon \sum_{i=1}^{p} (a_i \tau_i)^q \leq 1 \right\} \quad \text{for } q \geq 1.$$

Thus simplexes, ellipsoids and, of course, hyperrectangles are included. For rectangularly convex sets, the difficulty of T (in the sense of minimax risk) for linear estimates equals the difficulty of its hardest rectangular subproblem:

LEMMA 3. *If T is compact, convex and rectangularly convex at* 0,

$$\rho_L(T) = \sup_{\tau \in T} \rho_L([0, \tau]).$$

PROOF. This is similar to, but somewhat simpler than, the corresponding Theorem 7 in Donoho, Liu and MacGibbon (1990). Indeed, let $J(\tau) = \rho_L([0, \tau]) = \sum_i \tau_i/(1 + \tau_i)$ denote the linear difficulty of $[0, \tau]$. Let m denote the (unique) maximum of J over T and $\delta_m(x) = m_i x_i/(1 + m_i)$ denote the corresponding minimax rule. It suffices to show that $R(\delta_m, \tau)$ attains its maximum over T at m, for then

$$\sup_{\tau} \rho_L([0, \tau]) \leq \rho_L(T) \leq \sup_{\tau} R(\delta_m, \tau) = R(\delta_m, m) = \rho_L([0, m]).$$

But convexity of T and the maximum property of m guarantee that

$$0 \geq \frac{\partial}{\partial \varepsilon} J[(1 - \varepsilon)m + \varepsilon \tau] \bigg|_{\varepsilon = 0} = \sum_i \frac{\tau_i - m_i}{(m_i + 1)^2} = R(\delta_m, \tau) - J(m). \quad \square$$

Since $\rho_N(T) \geq \rho_N([0, \tau])$ for any $\tau \in T$, Lemma 3 and our Ibragimov–Has'minskii bound yield:

PROPOSITION 4. *If T is compact, convex and rectangularly convex at* 0, *then*

$$\rho_L(T)/\rho_N(T) \le \mu^*.$$

Thus linear and nonlinear minimax risks are essentially equivalent for l_1 balls, in sharp contrast to the situation in the corresponding Gaussian setting. Theory developed by Donoho, Liu and MacGibbon [(1990), Section 7ff] and Donoho and Johnstone (1989) indicates that near equivalence of linear and nonlinear minimax risks depends on the relative convexity of the parameter space and the loss function. Since the polydisc transform of the l_1 ball is quadratically convex and the transformed loss function is quadratic, the near equivalence found in Proposition 4 is not surprising. On the other hand, we conjecture that the minimax nonlinear risk will be significantly smaller for sets T whose polydisc transform is not quadratically convex.

3.2. *General hyperrectangles.* We make a start on extending this theory to more general rectangles $[m_1, m_2]$. When $[m_1, m_2] \subset R$, denote the *affine* minimax risk obtained by restricting attention to estimators $\delta_{a,b}(x) = ax + b$ by $\rho_L(m_1, m_2)$. The risk function of an affine linear rule is convex:

$$R(\delta_{a,b}, \tau) = a^2 - 2b(1-a) + (1-a)^2\tau + b^2/\tau \tag{10}$$

and the optimal affine linear rule is quite easily determined (see Section 7 for the proof).

LEMMA 5. *Let $\nu = \sqrt{m_2} - \sqrt{m_1}$. Then*

$$\rho_L(m_1, m_2) = \frac{\nu^2}{1+\nu^2} \quad \textit{and} \quad \delta^*(x) = \frac{\nu^2 x + \sqrt{m_1 m_2}}{1+\nu^2}. \tag{11}$$

It is easily checked that $m_1 \le \delta^*(m_1)$ and $\delta^*(m_2) \le m_2$, so that δ^* maps $[m_1, m_2]$ into itself. Recalling the square-root variance stabilizing transformation, one sees that (11) closely matches the corresponding results for the one-dimensional Gaussian shift experiment which it approaches as m_1, $m_2 \to \infty$. Indeed $Z = 2\sqrt{X}$ has an approximate $N(\theta, 1)$ distribution, with $\theta = 2\sqrt{\tau}$. If $\sqrt{m_2} = \sqrt{m_1} + \nu$, then the τ interval $[m_1, m_2]$ corresponds to a θ interval $I = [z_0 - \nu, z_0 + \nu]$ with midpoint $z_0 = 2\sqrt{m_1} + \nu$. The linear minimax rule (11) transforms to $\hat{\theta}^*(z) = 2\sqrt{\delta^*(z^2/4)}$, and Taylor expansion about z_0 yields

$$\hat{\theta}^*(z) = \left[z_0 + \nu^2(1+\nu^2)^{-1}(z - z_0)\right]\left[1 + O(\nu^2 m^{-1})\right],$$

where the first term is precisely the linear minimax estimator for estimating the $N(\theta, 1)$ parameter constrained to lie in I [cf. Donoho, Liu and MacGibbon (1990)].

Comparing linear and nonlinear minimax risks over intervals $[m_1, m_2]$ leads to a bivariate analogue of the Ibragimov–Has'minskii constant. Although stated for intervals $[m_1, m_2] \subset R^+$, it extends immediately to finite- and

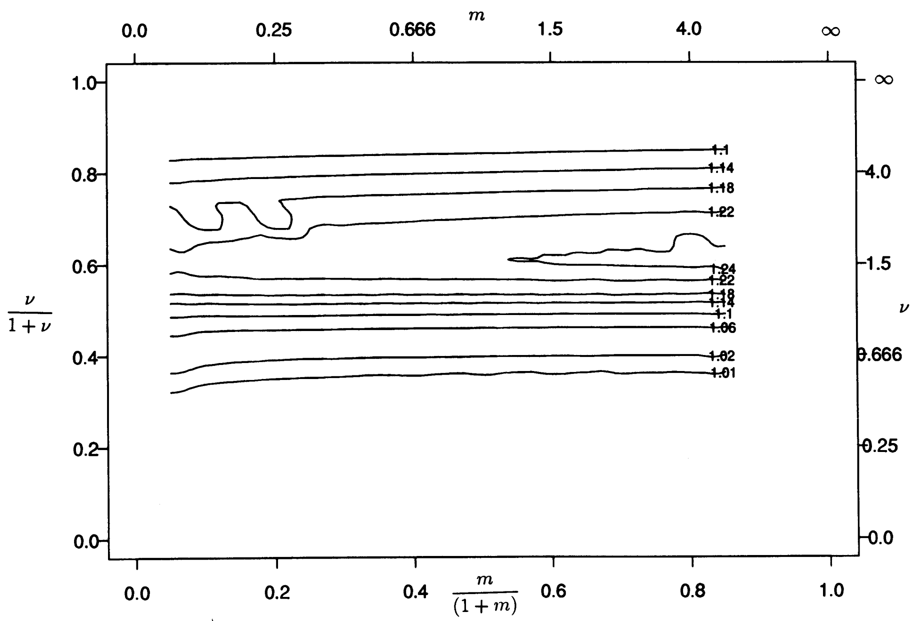

FIG. 3. *Contours of the risk ratio $\mu(m_1, m_2)$. The (x, y) axes are defined, respectively, by $m_1 = x/(1-x)$ and $\nu = y/(1-y)$ where $\nu = \sqrt{m_2} - \sqrt{m_1}$. Contours obtained in S after interpolation of values from raw ratios evaluated at $x = 0.05(0.05)0.85$ and $y = 0.05(0.05)0.85$. Limiting behavior near boundaries described in the proof of Theorem 7.*

infinite-dimensional hyperrectangles. We do not know if there is an extension of Lemma 3 to a class of sets not containing 0.

PROPOSITION 6. *The function $\mu(m_1, m_2) = \rho_L(m_1, m_2)/\rho_N(m_1, m_2)$ is continuous and bounded on $D = \{(m_1, m_2): 0 \le m_1 \le m_2 \le \infty\}$.*

PROOF. We have already seen that ρ_N is continuous on int D, so to complete the proof, we show that μ has bounded limits as $(m_1, m_2) \to \partial D$. For this purpose it is helpful to set $m = m_1$ and $\nu = \sqrt{m_2} - \sqrt{m_1}$, so that D transforms to R_+^2 in (m, ν) space. We also set $\mu'(m, \nu) = \mu(m, (\sqrt{m} + \nu)^2)$. We outline below the behaviour at the four boundaries in turn. Numerically determined contours of the function $\mu(m_1, m_2)$ in (a transformation of) these coordinates are shown in Figure 3.

Large ν. Use the lower bound (16) derived from the polydisc transform, with domain Ω equal to the plane annulus between radii $r_1 = \sqrt{m_1}$ and $r_2 = \sqrt{m_2}$. By further restricting attention to only radially symmetric functions, we obtain

$$\rho_N(m_1, m_2) \ge 1 - \inf\left\{ \int_{r_1}^{r_2} [v'(r)]^2 r\,dr \Big/ \int_{r_1}^{r_2} v^2(r) r\,dr,\ v \in C_0^1[r_1, r_2] \right\}.$$

The change of variables $s = (r - r_1)/(r_2 - r_1)$ transforms the ratio of integrals into

$$\frac{1}{(r_2 - r_1)^2} \frac{r_1 \int w'^2 + (r_2 - r_1) \int s w'^2}{r_1 \int w^2 + (r_2 - r_1) \int s w^2} \le \frac{C(w)}{(r_2 - r_1)^2},$$

where $w(s) = v(r) \in C_0^1[0,1]$ and we have used the inequality

$$(\alpha x + \beta y)/(x + y) \le \alpha \vee \beta$$

in order to define $C(w)$. It follows that $\rho_N(m_1, m_2) \ge 1 - C\nu^{-2}$ and that

$$\mu'(m,\nu) \le \frac{1 - (1+\nu^2)^{-1}}{1 - C\nu^{-2}}, \tag{12}$$

which is *independent of m* and converges to 1 as $\nu \to \infty$.

Small ν. Consider the two point prior $F_0 = (\varepsilon_{m_1} + \varepsilon_{m_2})/2$ and write L_{12} for the likelihood ratio $p_{m_1}(x)/p_{m_2}(x)$. As in (6) the Bayes rule $\delta_{F_0}(x) = p_1(x)m_1 + p_2(x)m_2$, where

$$p_1(x) = \frac{m_2 L_{12}}{m_1 + m_2 L_{12}}, \qquad p_2(x) = 1 - p_1(x).$$

The error $\delta_{F_0}(x) - m_i = (m_2 - m_1)(1 - p_i(x))$, and writing $E_{m_1} Z$ as $E_{m_2} L_{12} Z$, we find

$$\rho_N(m_1, m_2) \ge (1/2) R(\delta_{F_0}, m_1) + (1/2) R(\delta_{F_0}, m_2) \tag{13}$$

$$= (1/2)(m_2 - m_1)^2 m_2^{-1} E_{m_2} p_1(X). \tag{14}$$

Comparing the linear and nonlinear expressions (11) and (13) yields

$$\mu'(m,\nu) \le \frac{2m_2}{(1+\nu^2)\left(\sqrt{m_2} + \sqrt{m_1}\right)^2 E_{m_2} p_1(X)}. \tag{15}$$

The dominated convergence theorem assures that $E_{m_2} p_1(X) \to 1/2$ as $(m,\nu) \to (m_0, 0)$ $(m_0 > 0)$, which entails convergence of $\mu'(m,\nu)$ to 1.

Small m. The boundary $(m, 0)$ is covered by previous continuity arguments except at $(0,0)$. It may be shown that $\mu(m_1, m_2) \to 1$ as $(m_1, m_2) \to (0,0)$ through values $0 \le m_1 < m_2$ by using the prior ε_{m_2} to obtain the lower bound $\rho_N(m_1, m_2) \ge m_2 e^{-m_2}$.

The Gaussian limit (large m). As (m,ν) converges to (∞, ν_0), the Poisson problem approaches a Gaussian one: $\rho'_N(m,\nu) \to \rho_N^G(\nu_0)$, where the latter denotes the minimax risk for squared error loss in estimating $\zeta \in [-\nu_0, \nu_0]$ on the basis of $Y \sim N(\zeta, 1)$.

More specifically, for fixed (m,ν), define $Y = 2\sqrt{X} - 2\sqrt{m} - \nu$ and $\zeta = 2\sqrt{\tau} - 2\sqrt{m} - \nu$ and denote the distribution of Y by $P_\zeta^{m,\nu}$. Let Φ_ζ denote the distribution of an $N(\zeta, 1)$ variate. Then, for fixed ν, the experiment [in the sense of Le Cam (1986)] $\mathbf{E}^{m,\nu} = \{P_\zeta^{m,\nu}\colon |\zeta| \le \nu\}$ converges to $\mathbf{E}^{\infty,\nu} = \{\Phi_\zeta\colon |\zeta| \le \nu\}$ as $m \to \infty$.

Given an estimator $\hat{\tau}(X)$, define a corresponding estimator

$$\hat{\zeta}(Y) = 2\sqrt{\hat{\tau}(X)} - 2\sqrt{m} - \nu.$$

Since we may write

$$\frac{(\hat{\tau} - \tau)^2}{\tau} = \frac{1}{4}\left(1 + \frac{2\sqrt{m} + \nu + \hat{\zeta}}{2\sqrt{m} + \nu + \zeta}\right)^2 (\hat{\zeta} - \zeta)^2 = L^{m,\nu}(\hat{\zeta}, \zeta),$$

it follows that $\rho_N(m_1, m_2)$ equals the minimax risk for estimating ζ in experiment $\mathbf{E}^{m,\nu}$ using loss $L^{m,\nu}$. Since $L^{m,\nu}(\hat{\zeta}, \zeta) \to L(\hat{\zeta}, \zeta) = (\hat{\zeta} - \zeta)^2$ as $m \to \infty$, we conclude that $\rho'_N(m, \nu) \to \rho_N^{\mathrm{G}}(\nu)$. Since $\nu \to \rho'_N(m, \nu)$ and $\nu \to \rho^{\mathrm{G}}(\nu)$ are monotone increasing and continuous, the apparently stronger result that $\rho'_N(m, \nu) \to \rho_N^{\mathrm{G}}(\nu_0)$ as $(m, \nu) \to (\infty, \nu_0)$ follows from a "sandwich argument." The missing steps in these arguments were kindly provided by Nussbaum and are sketched in the Appendix: They form a simple and concrete illustration of Le Cam's theory of convergence of experiments. □

4. A lower bound. The polydisc transformation established a connection between minimax Poisson estimation in R_p^+ and the Laplace equation in R^{2p}. This section is devoted to establishing the following lower bound.

THEOREM 7. *Let T be an R_+^p-open and connected set. The minimax risk is*

$$\rho(T) = \inf_{\delta \in \mathbf{D}} \sup_{\tau \in T} R(\delta, \tau) \geq \frac{p^2}{p + \lambda(\Omega)} \geq p - \lambda(\Omega), \tag{16}$$

where $\Omega = \tau^{-1}(T)$ is the polydisc transform (2) *of T and $\lambda(\Omega)$ is the minimum eigenvalue of the Laplace equation on Ω.*

That is, $\lambda(\Omega)$ is the smallest value of λ for which the equation

$$\begin{aligned} \Delta u(\omega) &= -\lambda u(\omega), \qquad \omega \in \Omega, \\ u(\omega) &= 0, \qquad \omega \in \partial\Omega, \end{aligned} \tag{17}$$

has a nonzero solution. Here D_j denotes $\partial/\partial\omega_j$ and $\Delta = \sum_{j=1}^{2p} D_j^2$ denotes the Laplace operator. This solution is unique up to a constant multiple and we denote by u_Ω the positive solution normalized to satisfy $\int_\Omega u_\Omega^2 = 1$. As a shorthand notation, we sometimes write $\lambda_+(T)$ for $\lambda(\tau^{-1}(T))$.

Lower bound. A widely applicable lower bound follows from an extension of the Cramér–Rao information inequality due to Borovkov and Sakhanienko (1980) and further studied in Brown and Gajek (1990). Let the prior distribution $F(d\tau)$ have a density $f(\tau)$ w.r.t. Lebsegue measure which is C^1 and has compact support in $(0, \infty)^p$. Then from the multivariate analogs of Theorem 2.1 and Corollary 2.3 of Brown and Gajek (1990), applied to Poisson estimation,

$$r(F) \geq \frac{p^2}{p + J(f)} \geq p - J(f),$$

where

$$(18)\qquad J(f) = \int \sum_i f^{-1}(D_i f)^2(\tau)\tau_i \, d\tau.$$

REMARK. Given a specific candidate for a prior, better lower bounds for the Bayes risk can be obtained from Theorems 2.7 and 2.9 of Brown and Gajek (1990).

In a worst case analysis, however, one wishes to maximise (18) over a *class* of priors. Specifically, let T be an R_+^p-open set and maximize $r(F)$ over prior densities with compact support contained in $T \cap (0,\infty)^p$. This yields

$$(19)\qquad \rho(T) \geq \frac{p^2}{p + \lambda(T)} \geq p - \lambda(T),$$

where

$$(20)\quad \lambda(T) = \inf\left\{J(f)\colon f \in C_0^1(T \cap (0,\infty))^p,\ f \geq 0 \text{ and } \int_T f(\tau)\,d\tau = 1\right\}.$$

Scaling properties of the lower bound. Let $\sigma = m\tau$: A prior $P(d\sigma) = p(\sigma)\,d\sigma$ on mT induces a prior $F(d\tau) = f(\tau)\,d\tau$ on T and vice versa. Since $D_i p(\sigma) = m^{-1}D_i f(\tau)$, one verifies easily that

$$(21)\qquad \rho(mT) \geq \frac{p^2}{p + \sigma(T)/m} \geq p - \frac{\lambda(T)}{m}.$$

Alternative versions. We now derive equivalent forms of this bound that are both theoretically and computationally more convenient. Since $f \geq 0$, set $f = v^2$ so that (20) is equivalent to

$$(22)\qquad \lambda(T) = 4\inf\left\{\int_T \sum_i |D_i v(\tau)|^2 \tau_i \, d\tau\colon v \in C_0^1(T),\ \int_T v^2 = 1\right\}.$$

[In passing from (20) to (22), we have increased the class of test functions from $C_0^1(T \cap (0,\infty)^p)$ to $C_0^1(T)$. This does not change the minimization problem since any $v \in C_0^1(T)$ can be approximated by a sequence $v_n \in C_0^1(T \cap (0,\infty)^p)$ with $\int_T v_n^2 = \int_T v^2$ and $J(v_n^2) \to J(v^2)$. For example, if $T \subset [0, a_1]^p$ and $h_n(\tau) = (\log n)^{-1}\log^+(na_1^{-p}\prod_i \tau_i)$, then $v_n = c_n h_n v$ suffices, for suitable constants c_n, and $\log^+(x) = \max(0, \log(x))$.]

The Euler–Lagrange equation associated with (22) is derived from the lagrangian $H_\lambda(v) = J(v^2) - (\lambda/4)\int_T v^2$ by evaluating $\partial/\partial\varepsilon\, H_\lambda(v + \varepsilon\psi)|_{\varepsilon=0}$ for test functions $\psi \in C_0^\infty(T \cap (0,\infty)^p)$. Hence a minimizer v of (22), if it exists, satisfies

$$(23)\quad \sum_i \left[\tau_i D_i^2 v(\tau) + D_i v(\tau)\right] + (\lambda/4)v(\tau) = 0, \qquad \tau \in T \cap (0,\infty)^p.$$

This differential equation is also associated with the unbiased risk estimate corresponding to a Bayes rule for the prior $v^2(\tau)\,d\tau$; see II for further discussion. Equation (23) is a nonuniformly elliptic equation, as the coefficients τ_i of $D_i^2 v$ can become arbitrarily small if $\mathscr{F}_i$ is a critical face of T. Rather than pursue directly the theory of (22) and (23), we use the polydisc transform, defined in (2), to recover the classical (and *uniformly* elliptic) Laplace equation.

Role of polydisc transform. We first note that if T is R_+^p-open and connected, then $\Omega = \tau^{-1}(T)$ is open and connected in R^{2p} (Lemma 11). Our convention that $T = \operatorname{int} \overline{T}$ makes $\Omega = \tau^{-1}(T)$ as large as possible: $\tau^{-1}((0, m))$ is a punctured disc, but $\operatorname{int}\overline{(0, m)} = [0, m)$, so $\tau^{-1}(\operatorname{int} \overline{T})$ has the puncture removed.

A function $v(\tau)$ defined on T induces a function $u(\omega) = v(\tau(\omega))$ on Ω. It is useful to associate polar coordinates (r_i, θ_i), $i = 1, \dots, p$, with ω: We then have $\bar{u}(r)$ such that $u(\omega) = \bar{u}(r) = v(\tau)$. Since $d\omega_{2i-1}\, d\omega_{2i} = r_i\, dr_i\, d\theta_i = 2^{-1}\, d\tau_i\, d\theta_i$,

$$\int_\Omega u(\omega)\, d\omega = (2\pi)^p \int \bar{u}(r) \prod_i r_i\, dr_i = \pi^p \int_T v(\tau)\, d\tau. \tag{24}$$

In particular, the polydisc transform is *volume-preserving*: If T and T' have the same p-dimensional measure, then the $2p$-dimensional Lebesgue measures of $\tau^{-1}(T)$ and $\tau^{-1}(T')$ are equal.

Derivatives are easily calculated:

$$D_{2i-1} u(\omega) = 2\, D_i v(\tau) \omega_{2i-1}, \qquad D_{2i} u(\omega) = 2\, D_i v(\tau) \omega_{2i}, \tag{25}$$

and hence

$$\int_\Omega |Du|^2\, d\omega = 4 \int_T \sum_i (D_i v)^2 \tau_i\, d\tau. \tag{26}$$

REMARK. Define $g(\omega) = f(\tau(\omega))$. In a similar vein we find that

$$J(f) = 4^{-1} \pi^{-p} \int_\Omega g^{-1} |Dg|^2 = 4^{-1} \pi^{-p} I(g), \tag{27}$$

where $I(g)$ is a multivariate extension of Fisher information for location of density g (see II).

The Dirichlet problem. We define the class of *polyradial* functions as

$$\mathbf{P} = \{u \in L^2(\Omega) : u(\omega) = v(\tau(\omega)) \text{ for } \omega \in \Omega \text{ and some function } v : \mathbf{R}_+^p \to R\}.$$

The polydisc transform expresses $\lambda(T)$ in (22) [via (24) and (26)] as

$$\lambda(T) = \inf\left\{ \int_\Omega |Du|^2 : u \in C_0^1(\Omega) \cap \mathbf{P},\ \int_\Omega u^2 = 1 \right\}.$$

Let $W^{1,2}(\Omega)$ denote the Sobolev space consisting of once-weakly differentiable functions having norm

$$\|u\|_{W^{1,2}(\Omega)}^2 = \int_\Omega |Du|^2 + u^2 < \infty.$$

The energy functional $I^*(u) = \int_\Omega |Du|^2$ is naturally defined on $W_0^{1,2}(\Omega)$, the closure of $C_0^1(\Omega)$ in $W^{1,2}(\Omega)$. It is related to Fisher information by the equation $I^*(u) = I(u^2)/4$. Define

$$\lambda(\Omega) = \inf\left\{I^*(u): u \in W_0^{1,2}(\Omega), \int_\Omega u^2 = 1\right\}. \tag{28}$$

The *important fact* $\lambda(T) = \lambda(\Omega)$ is easily verified. By construction, Ω and hence $W_0^{1,2}(\Omega)$, $\int_\Omega |Du|^2$ and $\int_\Omega u^2$ are invariant under the obvious action of transformations $R_\theta = (R_{\theta_1}, \dots, R_{\theta_p})$, where R_{θ_i} denotes a rotation of R^2 about 0 through angle θ_i. Since the minimizer u_Ω of (28) is unique (Theorem 8) and since $R_\theta u_\Omega$ is an equally valid minimizer, it follows that $R_\theta u_\Omega = u_\Omega$ for all θ, and hence that $u_\Omega \in \mathbf{P}$.

Finally, we recall here some basic facts about the Dirichlet problem [see, for example, Gilbarg and Trudinger (1983), page 214]. Combining these with the preceding discussion completes the proof of Theorem 7.

THEOREM 8. (i) *There is a unique* (*up to sign*) *function* $u_\Omega \in W_0^{1,2}(\Omega)$ *achieving the minimum of* $I^*(u)$ *in* (28). *It satisfies the equation*

$$\Delta u + \lambda(\Omega) u = 0 \quad \text{in } \Omega.$$

(ii) *The minimum eigenvalue* $\lambda(\Omega) > 0$ *and is simple*; *the corresponding eigenfunction* u_Ω (*or* $-u_\Omega$) *is positive throughout* Ω.

(iii) *The minimum eigenvalue* $\lambda(\Omega)$ *is monotone in* Ω: $\Omega \subset \Omega'$ *implies* $\lambda(\Omega) \geq \lambda(\Omega')$.

5. Asymptotics and examples.

Asymptotics. A simple scaling argument shows that the bound (16) transforms to $\rho(mT) \geq p - m^{-1}\lambda(\Omega)$. Our main asymptotic result (stated precisely in II) asserts that this bound is sharp:

THEOREM 9. *Let* T *be* R_+^p*-open with compact closure and sufficiently smooth boundary. Let* $\Omega = \tau^{-1}(T)$. *Then*

$$\rho(mT) = p - \lambda(\Omega)/m + o(m^{-1}). \tag{29}$$

Further, the least favourable prior distribution on mT, *when rescaled back to* T, *converges weakly as* $m \to \infty$ *to a distribution with density* $f_T(\tau) = u_\Omega^2(\omega)$.

Figure 2 plots the asymptotic approximation (29), from which it is apparent that the asymptotics are appropriate for moderately large m. This phenomenon is equally present in the Gaussian case.

The theorem also suggests that f_T could be used to construct a second order asymptotically minimax estimator of the form

$$\delta_{m,i}(x) = x_i + m^{-1}x_i(D_i f/f)(m^{-1}x)$$

(see the heuristic discussion appearing in II). This does not quite work due to singularity of Df/f near the boundary of T, but under smoothness conditions on T, an asymptotically second order minimax version of δ_m is constructed in II.

Examples. Separation of variables in (17) can be used to evaluate $\lambda(\Omega)$ and u_Ω explicitly in a number of interesting cases.

EXAMPLE A. Intervals in R^1_+. (i) $[0, b]$. The polydisc transform yields a disc of radius $\sqrt{b}$ in the plane:

$$\Omega = \tau^{-1}([0, b]) = \{\omega: \omega_1^2 + \omega_2^2 \le b\}.$$

The eigenvalues and eigenfunctions of the disc are given in terms of the Bessel functions of the first kind of index n, solution to $x^2 J_n''(x) + xJ_n'(x) + (x^2 - n^2)J_n(x) = 0$ and regular at $x = 0$. Let v_n denote the smallest positive zero of $J_n(x)$ [cf., e.g., Abramowitz and Stegun (1972), Chapter 9]. Then the minimum eigenvalue $\lambda_+([0, b]) = b^{-1}v_0^2$ and the corresponding asymptotically least favourable density is $f_T(\tau) = cJ_0^2(v_0\tau^{1/2}b^{-1/2})$. A close approximation to $f_T(\tau)$ is given by $c' \cos^2(\pi\tau^{1/2}b^{-1/2}/2)$. Figure 2 shows that the asymptotic approximation $\rho_A(m) = 1 - 5.782m^{-1}$ provided by Theorem 9 is poor for small values of $\sqrt{m}$, but the lower bound provided by (16) and (21) of Section 4 is considerably better. Work of Levit (1987) for the Gaussian case suggests that the error in $\rho_A(m)$ here would be $O(m^{-3/2})$.

(ii) $[b_1, b_2]$ (with $b_1 > 0$). The polydisc transform now gives an annulus between radii $\sqrt{b_1}$ and $\sqrt{b_2}$ in the plane:

$$\Omega = \tau^{-1}([b_1, b_2]) = \{\omega: b_1 \le \omega_1^2 + \omega_2^2 \le b_2\}.$$

Separation of variables produces zero boundary conditions at $\tau = b_1$ *and* b_2, and the second linearly independent solution to Bessel's equation, $Y_n(x)$, is now also required. The smallest eigenvalue is found to be $\lambda_+([b_1, b_2]) = x_{0,1}^2(k)/b_1$ in terms of the smallest positive zero $x_{0,1}$ of the cross-product equation

$$J_0(x)Y_0(kx) - J_0(kx)Y_0(x) = 0, \qquad k = (b_2/b_1)^{1/2}.$$

The roots $x_{0,1}(k)$ are tabulated in Jahnke, Emde and Lošch [(1960), Table 34] and for these values $(k - 1)x_{0,1}(k)$ decreases from π to v_0 for $k \in [1, \infty)$. We therefore write $\lambda_+([b_1, b_2]) = c^2/(\sqrt{b_2} - \sqrt{b_1})^2$, with the tables supporting the conjecture that c decreases from π to v_0 as the annulus fattens from a ring into a disc. Thus the previous expression for [0, b] appears as a limiting case. The asymptotically least favourable distribution $f_T(\tau) = cu_\Omega^2(\sqrt{\tau})$, where

$$u_\Omega(r) = Y_0(x_{0,1})J_0(x_{0,1}rb_1^{-1/2}) - J(x_{0,1})Y_0(x_{0,1}rb_1^{-1/2})$$

is the principal eigenfunction of the annulus.

EXAMPLE B. Solid simplex in R^p_+. $T_S(b) = \{\tau \ge 0: \Sigma_1^p\tau_i < b\}$. T_S expresses an upper bound on the total of the cell means. As noted earlier, this is appropriate when the distribution of intensity amongst p cells of a partition is sought, when a bound on the total intensity is given. The polydisc transform leads to a ball in R^{2p}:

$$\Omega_S = \tau^{-1}(T_S) = \left\{\omega: \omega_1^2 + \cdots + \omega_{2p}^2 < b\right\}.$$

We now have $\lambda_+(T_S(b)) = b^{-1}v_{p-1}^2$ and the corresponding asymptotically least favourable density is

$$f_T(\tau) = c\left(|\tau| b^{-1}v_{p-1}^2\right)^{1-p} J_{p-1}^2\left(|\tau|^{1/2} b^{-1/2} v_{p-1}\right) \quad \text{for } 0 \le |\tau| = \sum_1^p \tau_i \le b.$$

This is obtained from the principal eigenfunction

$$u_\Omega(\omega) = \left(|\omega|^2 b^{-1} v_{p-1}^2\right)^{(1-p)/2} J_{p-1}\left(|\omega| b^{-1/2} v_{p-1}\right)$$

of the Laplace operator on the ball of radius b in R^{2p}. This eigenvalue and $u_\Omega(\omega)$ occur in Berkhin and Levit [(1980), Example 1] and Bickel (1981) in the corresponding *Gaussian* minimax estimation problem on the ball Ω_S.

EXAMPLE C. Hyperrectangle. $T_R(b) = \{\tau: b_{1i} \le \tau_i \le b_{2i},\ i = 1, \dots, p\}$. T_R corresponds to independent prior constraints on each of the components. The polydisc transform produces a polydisc:

$$\Omega_R = \tau^{-1}(T_R) = \left\{\omega: b_{1i} < \omega_{2i-1}^2 + \omega_{2i}^2 < b_{2i},\ i = 1, \dots, p\right\}.$$

However, when T is a product $\prod_{i=1}^p T_i$, it is easily seen from the independence of the components that the minimax problem decomposes into separate univariate problems. Hence $\lambda_+(T) = \sum_1^p \lambda_+(T_i)$ and $f_T(\tau) = \prod_1^p f_{T_i}(\tau_i)$. Thus if $T = \prod_i [b_{1i}, b_{2i})$, everything follows from the case $p = 1$ discussed above. Thus $\lambda_+(T_R(b)) = \sum_1^p x_{0,1}^2(\sqrt{b_2}/\sqrt{b_1})/b_{1i}$ and $f_T(\tau)$ is proportional to the product of the univariate least favourable densities. For a hyperrectangle $[0, b]$ containing 0, the eigenvalue has the simpler form $\lambda_+([0, b]) = v_0^2 \sum_1^p b_i^{-1}$.

Isoperimetric inequalities. Classical inequalities derived for $\lambda(\Omega)$ as a function of Ω yield statistical information via (29). For example, the Faber–Krahn inequality states that amongst domains in R^n with a given volume, $\lambda(\Omega)$ is minimised when Ω is a ball. For further discussion and references, see Payne (1967) and, in the planar case, Kuttler and Sigillito (1984). Since the polydisc transform is volume-preserving [cf. (24)], it follows that among all intervals T in $[0, \infty)$ of a given length b say, $\rho(mT)$ is maximised (to second order in m) by setting $T = [0, b]$. It is easy to see also that $T = [0, b]$ is most difficult even among sets T that are a finite union of disjoint intervals of total length at most b. More generally, the simplex $T_S(b)$ is most difficult asymptotically among all sets in R_+^p of equal volume composed of a finite union of connected open sets.

6. Discussion. The dependence of the minimax risk $\rho_N(T)$ on the constraint set T has been intensively studied in the finite-dimensional Gaussian translation model in recent years. Numerical and analytical results for small intervals $T \subset R^1$ were given by Casella and Strawderman (1981). Asymptotic approximations for large T (or small noise level) are given by Levit (1980, 1982, 1985), Berkhin and Levit (1980) and Bickel (1981). The works of Levit and Melkman and Ritov (1987) consider non-Gaussian situations also. The Gaussian minimax risk $\rho_N(T)$ was later shown to have an unexpected and important role in nonparametric estimation problems over compact infinite-dimensional parameter spaces in comparing the behaviour of linear to nonlinear estimators [Ibragimov and Has'minskii (1984); Donoho and Liu (1988); Donoho, Liu and MacGibbon (1990)].

REMARKS 1. This paper considers only the error measure

$$L(d,\tau) = \sum_i (\delta_i - \tau_i)^2/\tau_i. \tag{30}$$

For this loss function, the trivial estimator $\delta(X) = X$ has constant risk throughout the parameter space, which partially compensates for the lack of any group invariant structure [see Clevenson and Zidek (1975) for further discussion]. It seems plausible, however, from the work of Brown (1979) and Levit that the role of the polydisc transform, principal eigenvalue of Laplacian and form of asymptotically second order minimax estimator will be insensitive to choice of loss function within a class of the form $\sum_i w[\tau_i^{-1/2}(\delta_i - \tau_i)]$ for suitable weight function w: $R \to R_+$. For losses of the form $\sum_i w(\delta_i - \tau_i)$, results will be qualitatively different: For asymptotic results for compact $\overline{T} \subset (0,\infty)^p$, see Levit (1982).

REMARK 2. The loss function $L(\delta, \lambda) = \sum_i \lambda_i^{-1}(\delta_i - \lambda_i)^2$ is not defined when any $\lambda_i = 0$, so that it is only strictly meaningful to consider parameter spaces $T \subset (0,\infty)^p$. Consider, for example, the case $p = 1$, $T = (0, m)$. The work of Section 2 shows for some values of m that no least favorable distribution exists; rather only a sequence of approximately least favorable priors P_m, whose weak limit has an atom at 0.

Note, however, that since risk functions are continuous in τ, the minimax value of $\rho(T)$ depends only on $\overline{T}$. Thus the difficulty is not serious; but for completeness, we show in the Appendix (Remark A.1) that it can be accommodated by introducing an artificial decision problem, equivalent for the purposes of minimax analysis, in which T may intersect ∂R_+^p.

REMARK 3. This paper elaborates the mathematical connection between a p-variate Poisson estimation problem and a $2p$-variate Gaussian problem, but the statistical connection is still somewhat obscure. Brown's heuristic analysis [(1979), Section 2.3] shows that the variance stabilising transformation $r = \tau^{1/2}$ leads to a location-like estimation setting. Indeed, setting $\gamma(x) = \delta^{1/2}(x)$ and assuming τ large gives

$$\tau^{-1}(\delta(x) - \tau)^2 = r^{-2}(\gamma(x) + r)^2(\gamma(x) - r)^2 \approx 4(\gamma(x) - r)^2.$$

The mathematics here says that r, which is inherently nonnegative, should be thought of as the *radial* components of a bivariate location vector.

REMARK 4. The role of the polydisc transform in this paper is to carry over standard analytical and statistical results from the Gaussian to the Poisson setting. The transform can be expected to play this role in various other Poisson decision theoretic contexts. Examples include admissibility for both point and loss estimation procedures related to the loss function (30); some successful applications are discussed by Lele (1990).

REMARK 5. Do there exist analogues of the polydisc transform appropriate to other parametric families of distributions? Here it has functioned as a sort of "dimension-sensitive" variance stabilising transformation: In the Poisson case, it is derived from the usual square-root transformation, but doubles the dimension of the new parameter space. Here is one result from work with Soren Johansen. For a regular family of densities $p_\tau(x)\nu(dx)$ on R, the analogue, for information-normalised loss $(\delta(x) - \tau)^t I(\tau)(\delta(x) - \tau)$ of the quantity (18) appearing in the Borovkov–Sakhanienko bound is

$$J(f) = \int \sum_i (D_i f/f)^2 I^{-1}(\tau_i) f(\tau)\, d\tau,$$

where $I(\tau)$ is the usual Fisher information $E_\tau[\partial \log p_\tau(X)/\partial\tau]^2$ and I^{-1} denotes $1/I$. One might describe as a "generalised polydisc transform" any mapping $\omega \to (\tau, \sigma)$ which (i) preserves Lebesgue measure: $f(\tau)\, d\tau\, d\sigma = c_1 g(\omega)\, d\omega$ [where $g(\omega) := f(\tau(\omega))$] and (ii) maps $J(f)$ into (scalar) Fisher information for location $J(f) = c_2 \int |\nabla g/g|^2 g$. Some examples of such mappings include:

(a) $\tau = (\omega_1^2 + \cdots + \omega_m^2)^{m/2}$, $\sigma = \omega/|\omega|$ when $I^{-1}(\tau) = \tau^{2-2/m}$ (and $m \geq 1$, $m \in Z$).

(b) $\tau = (\omega_1 + \omega_2)^\alpha$, $\sigma = \omega_1/\alpha(\omega_1 + \omega_2)^{\alpha-1}$ when $I^{-1}(\tau) = \tau^{2-2/\alpha}$ (and $\alpha \geq 2$, $\alpha \in R$).

(c) $\tau = e^{\omega_1+\omega_2}$, $\sigma = \omega_1 e^{-\omega_1-\omega_2}$ when $I^{-1}(\tau) = 2\tau^2$.

Note, for example, that $I^{-1}(\tau) = c\tau^2$ when τ is the mean of a gamma distribution.

REMARK 6. This paper is "parametric" in the asymptotic sense that as the amount of information increases ($m \to \infty$), the number of parameters p remains fixed. Donoho and Johnstone (1989) study an alternative asymptotic model (for Gaussian data) in which the number of parameters grows with the available data, allowing various possible limiting signal to noise behaviors. In this "nonparametric" setting, only first order asymptotic results appear, but Fisher information still plays a critical role, and the shape of T has a profound influence on the relative performance of linear and nonlinear estimators.

APPENDIX

PROOF OF THEOREM 2. It remains to show that the two point prior $a\varepsilon_0 + (1 - a)\varepsilon_1$ specified in the theorem is minimax. The strategy for minimaxity is to show that

$$\tau^2 e^\tau \,\partial R/\partial\tau = \tau^2(m - \delta_F(1))(m + \delta_F(1) - 2\tau) - 2\delta_F(1)\tau^2 + (1+\tau)m^2 + e^\tau(\tau^2 - m^2) \tag{31}$$

changes sign at most once from negative to positive on $[0, m]$. At zero, $\partial R/\partial\tau$ will be nonpositive for $0 < m \le m_1$ (≈ 1.27), where m_1 is the first *positive* zero of the equation $(1 + \delta_F(m))^2 = 2 + m^2/2$. For $m \in [m_0, m_1]$, nonnegativity for $\partial R/\partial\tau$ at $\tau = m$ follows from (8), and careful study of the first four derivatives of the right side of (31) shows that $\partial R/\partial\tau$ has at most one sign change on $[0, m]$.

LEMMA 10. *$\rho_N([m_1, m_2])$ and the corresponding least favorable distribution F_{m_1, m_2} depend continuously on (m_1, m_2).*

PROOF. Since $\rho_N(T)$ is monotone with respect to subsets of T, it follows that for (m'_1, m'_2) in a given neighbourhood N of (m_1, m_2), there is an upper bound k_0 to the cardinality of the support of $F_{m'_1, m'_2}$. In fact, k_0 is derived from $F_{m_1^*, m_2^*}$ where $[m_1^*, m_2^*]$ is the union of all the intervals derived from (m'_1, m'_2) in the closure of N. Continuity of $\rho_N([\cdot, \cdot])$ then flows from continuity for $r_{k_0}(\cdot, \cdot)$. The uniqueness of the least favourable distribution $F_{m'_1, m'_2}$ then ensures that it depends continuously (in the weak topology on probability measures) on (m'_1, m'_2), though extra constraints are needed to eliminate redundancies in the coordinate description using (a, b). □

PROOF OF LEMMA 5. From (10),

$$\sup_{[m_1, m_2]} R(\delta_{a,b}, \tau) = a^2 + (f_1 \vee f_2)(a, b),$$

where $f_i(a, b) = m_i[(1-a) - bm_i^{-1}]^2$. First fix a and optimize over b: The quadratics $b \to f_i(b)$ intersect at $b_\pm^2 = m_1 m_2 (1-a)^2$. One checks that the minima b_i of f_i are positive and lie, respectively, inside (b_1) and outside (b_2) the interval $[b_-, b_+]$. Since $f_1'' > f_2''$, it follows that

$$\min_b (f_1 \vee f_2)(b) = f_1(b_+) = (1-a)^2\left(\sqrt{m_2} - \sqrt{m_1}\right)^2.$$

Optimizing $a^2 + (1-a)^2\nu^2$ over a yields $a^* = \nu^2/(1+\nu^2)$. □

Convergence to Gaussian experiment in Proposition 6. Let P_τ denote a Poisson distribution with mean τ, $Q_{\zeta, m} = P_{(m^{1/2} + (\zeta+\nu)/2)^2}$ and $Q_{\zeta, 0}$ an $N(\zeta, 1)$ distribution.

1. The sequence of experiments $\mathscr{Q}_m = \{Q_{\zeta, m}, |\zeta| \le \nu\}$ converges weakly to $\mathscr{Q}_0 = \{Q_{\zeta, 0}, |\zeta| \le \nu\}$. In view of Le Cam's (1986) Lemma 10.2.1 and Proposition 6.3.7, this would follow if there exist random variables η_m such that for

$|\zeta| \le \nu$, the log likelihood ratio

$$\log \frac{dQ_{\zeta,m}}{dQ_{\zeta,0}} - \zeta\eta_m + \frac{1}{2}\zeta^2 \to 0$$

in $Q_{0,m}$ probability, and $\mathscr{L}(\eta_m | Q_{0,m}) \to N(0,1)$. This follows from calculation and the fact that if $X_m \sim Q_{0,m} = P_{\tau_0}$, $\tau_0 = (m^{1/2} + \nu/2)^2$, then $\eta_m = (X_m - \tau_0)/\sqrt{\tau_o} \to_D N(0,1)$.

2. In fact, $\mathscr{Q}_m$ converges *strongly* to $\mathscr{Q}_0$: This relies on Theorem 6.2 of Le Cam (1986), the continuity of the mapping $\zeta \to N(\zeta, 1)$ and its inverse (with respect to total variation norm), the standard inequality relating Hellinger and total variation distance

$$(1/2)\|P - Q\| \le H(P,Q)\{\|P + Q\| - H^2(P,Q)\}^{1/2}$$

and the identity

$$H^2(Q_{\zeta_1,m}, Q_{\zeta_2,m}) = 2\left\{1 - \exp\{-(1/8)(\zeta_1 - \zeta_2)^2\}\right\}.$$

3. To establish convergence of risks, use the framework and notation of Le Cam [(1986), Chapter 7.4]. The decision space is $\{\zeta\colon |\zeta| \le \nu\}$. The uniform lattice Γ is $C[-\nu,\nu]$ with the supremum norm and the L-space L may be identified with signed measures on R with total variation norm. The loss functions $l_{\zeta,0}(d) = l_0(d - \zeta) = (d - \zeta)^2$ and $l_{\zeta,m}(d) = L^{m,\nu}(d,\zeta)$ belong to Γ and

$$\|l_{\zeta,m} - l_{\zeta,0}\|_\infty \to 0 \quad \text{as } m \to \infty.$$

All decision procedures have a Markov kernel representation and hence are randomized estimators. Thus

$$\rho'_N(m,\nu) = \inf_\sigma \sup_{|\zeta| \le \nu} Q_{\zeta,m}\sigma l_{\zeta,m}$$

and $\rho_N^{\mathrm{G}}(\nu)$ is the corresponding quantity with m replaced by 0. Denote by T transitions $L \to L$. Strong convergence means that

$$\Delta(\mathscr{Q}_m, \mathscr{Q}_0) = \max\left\{\inf_T \sup_{|\zeta|\le\nu} \|Q_{\zeta,m} - Q_{\zeta,0}T\|, \inf_T \sup_{|\zeta|\le\nu} \|Q_{\zeta,0} - Q_{\zeta,m}T\|\right\} = o(1).$$

Let T_m be the transition attaining the first of these infima. Then

$$Q_{\zeta,m}\sigma l_{\zeta,m} = Q_{\zeta,0}T_m\sigma l_{\zeta,0} + Q_{\zeta,m}\sigma(l_{\zeta,m} - l_{\zeta,0}) + (Q_{\zeta,m} - Q_{\zeta,0}T_m)\sigma l_{\zeta,0}$$

and

$$\|Q_{\zeta,m}\sigma(l_{\zeta,m} - l_{\zeta,0})\| \le \|l_{\zeta,m} - l_{\zeta,0}\|_\infty = o(1),$$
$$\|(Q_{\zeta,m} - Q_{\zeta,0}T_m)\sigma l_{\zeta,0}\| \le \sup_\zeta \|Q_{\zeta,m} - Q_{\zeta,0}T_m\| \, \|l_{\zeta,0}\|_\infty = o(1).$$

Thus

$$\inf_\sigma Q_{\zeta,m}\sigma l_{\zeta,m} = \inf_\sigma Q_{\zeta,0}T_m\sigma l_{\zeta,0} + o(1) \ge \inf_\sigma Q_{\zeta,0}\sigma l_{\zeta,0} + o(1).$$

The reverse inequality is proved analogously, and this completes the proof. □

LEMMA 11. (i) *If T is R_+^p-open, then* $\Omega = \tau^{-1}(T)$ *is open in* R^{2p}.
(ii) *If T is connected, then so is* Ω.

PROOF. Part (i) follows from the simple inequality

$$|\omega|, |\bar{\omega}| \le M \quad \Rightarrow \quad |\tau(\omega) - \tau(\bar{\omega})|^2 \le 8M^2|\omega - \bar{\omega}|^2. \tag{32}$$

(ii) If $\tau(t)$ is a path from $\tau(\omega)$ to $\tau(\bar{\omega})$, then, using polar coordinates $(r_i, \theta_i)_{i=1}^p$ on R^{2p}, a path from ω to $\bar{\omega}$ is given by $r_i(t) = \tau_i^{1/2}(t)$, $\theta_i(t) = (1-t)\theta_i + t\bar{\theta}_i$. □

REMARK A.1. As before, let $\kappa_i = 1$ if i indexes a critical face of T and 0 otherwise. In the ith new decision problem, we observe X conditional on $X \ge \kappa_i$ and estimate τ_i with loss function

$$L_{1i}(\delta_i, \tau_i) = \begin{cases} \tau_i^{-1}(\delta_i - \tau_i)^2, & \text{if } i \notin \mathbf{I}, \\ e^{-\tau_i}\tau_i + \phi(\tau_i)(\delta_i - \tau_i)^2, & \text{if } i \in \mathbf{I}, \end{cases}$$

where $\phi(\tau_i) = 1$ if $\tau_i = 0$ and $= \tau_i^{-1}(1 - e^{-\tau_i})$ if $\tau_i > 0$. The point is that the risk function of $\delta_i(X)$, for $\tau_i > 0$, satisfies

$$R_{1i}(\delta_i, \tau) = E_\tau\left[L_{1i}(\delta_i(X), \tau_i) | X_i \ge \kappa_i\right] = E_\tau(\delta_i(X) - \tau_i)^2/\tau_i,$$

since by our convention $\delta_i(x) = 0$ if $x_i < \kappa_i$. Summing the risk functions in the p new decision problems yields a risk function $R_1(\delta, \tau) = \sum_i R_{1i}(\delta_i, \tau)$ equaling $R(\delta, \tau)$ on $(0, \infty)^p$ for $\delta \in \mathbf{D}(T)$. If follows that the minimax value for R_1 equals that for R. On the other hand, $R_1(\delta, \tau)$ is well defined for $\tau \in \bar{T} \cap R_+^p$, and so Bayes rules for priors supported on $\bar{T}$ make sense. By a double application of the minimax theorem, it follows that

$$\sup_{\mathbf{P}^*(\bar{T})} r_1(P) = \inf_{\mathbf{D}} \sup_{\bar{T}} R_1(\delta, \tau) = \inf_{\mathbf{D}} \sup_{T \cap (0,\infty)^p} R(\delta, \tau) = \sup_{\mathbf{P}^*(T \cap (0,\infty)^p)} r(P).$$

It is shown in II that $P \to r(P)$ is strictly concave on $\mathbf{P}^*(\bar{T})$, and hence that a maximising P, the *least favorable distribution*, is unique. For $P \in \mathbf{P}^*(\bar{T})$, define $\hat{p}(x) = \int \prod_i e^{-\tau_i}\tau_i^{x_i} P(d\tau)$. Note that if i does not index a critical face, then $\hat{p}(x) < \infty$ even if $x_i = -1$. It is easily checked that the Bayes rule δ_P in $\mathbf{D}(T)$ for $R_1(\delta, \tau)$ corresponding to P has the representation

$$\delta_{P,i}(x) = \hat{p}(x)/\hat{p}(x - e_i) \quad \text{if } x_i \ge \kappa_i \text{ and } \hat{p}(x - e_i) > 0 \tag{33}$$

and 0 otherwise. Of course, this agrees with the Bayes rule for $R(\delta, \tau)$ if $P \in \mathbf{P}^*(T \cap (0, \infty)^p)$. Furthermore, if P_n converges weakly to P in $\mathbf{P}^*(\bar{T})$, and $P((0, \infty)^p) > 0$, then $\delta_{P_n}(x) \to \delta_P(x)$ for all x.

Acknowledgments. Research support at various times from NSF MCS-84-51750 and MCS-86-00235, NIH PHS-GM-21215-12, SERC, the Sloan Foundation, NSERC, FCAR and GERAD is gratefully acknowledged. For helpful conversations and references, the authors are delighted to thank Larry Brown, David Donoho, Edward Fraenkel, Søren Johansen, Bernard Silverman and John Toland. Michael Nussbaum helped greatly with the proof of Proposition 6. Much of the work on the paper was done during visits by both authors to the University of Bath. The computing of the exact minimax risk was done using some ideas of Peter Kempthorne and using modified grid searches on a CYBER 3D at Concordia University. The authors are grateful to programmer Gabriel Lambert for the implementation of the algorithms. An Associate Editor and referees made stimulating comments leading to reorganization and partition of the original version.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions*. National Bureau of Standards, Washington, D.C.

ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York.

BERKHIN, P. E. and LEVIT, B. YA. (1980). Second-order asymptotically minimax estimates for the mean of a normal population. *Problemy Peredachi Informatsii* **16** 60–79. [Translation: *Problems Inform. Transmission* (1980) **16** 212–229.]

BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309.

BOROVKOV, A. A. and SAKHANIENKO, A. U. (1980). On estimates of the expected quadratic risk. *Probab. Math. Statist.* **1** 185–195 (in Russian).

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

BROWN, L. D. and GAJEK, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18** 1578–1594.

CASELLA, G. and STRAWDERMAN, W. E. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** 870–878.

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

DONOHO, D. L. and JOHNSTONE, I. M. (1989). Minimax risks over l_p-balls. Technical Report 322, Dept. Statistics, Stanford Univ.

DONOHO, D. L. and LIU, R. C. (1988). Hardest one-dimensional subfamilies. Technical Report 178, Dept. Statist., Univ. California, Berkeley.

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1988). Minimax risk over hyperrectangles. Technical Report 123, 3rd rev., Dept. Statistics., Univ. California, Berkeley.

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, with implications. *Ann. Statist.* **18** 1416–1437.

GILBARG, D. and TRUDINGER, N. S. (1983). *Elliptic Partial Differential Equations of Second Order*, 2nd ed. Springer, New York.

GOURDIN, E., JAUMARD, B. and MACGIBBON, B. (1990). Global optimization decomposition methods for bounded parameter minimax estimation. Technical Report G-90-48, GERAD, École des Hautes Études Commerciales, Montréal.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1984). Nonparametric estimation of the value of a linear functional in Gaussian white noise (translation). *Theory Probab. Appl.* **29** 1–32.

JAHNKE, E., EMDE, F. and LOŠCH, F. (1960). *Tables of Functions*, 6th ed. Teubner, Stuttgart.

JOHNSTONE, I. M. and MACGIBBON, K. B. (1990). Minimax estimation of a constrained Poisson vector. Technical Report 364, Dept. Statistics, Stanford Univ.

JOHNSTONE, I. M. and MACGIBBON, K. B. (1991). Asymptotically minimax estimation of a constrained Poisson vector via polydisc transforms. Technical Report 385, Dept. Statistics, Stanford Univ.

KUTTLER, J. R. and SIGILLITO, V. G. (1984). Eigenvalues of the Laplacian in two dimensions. *SIAM Rev.* **26** 163–193.

LE CAM, L. M. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LELE, C. (1990). Admissibility in loss estimation. Ph.D. dissertation, Dept. Statistics, Stanford Univ.

LEVIT, B. YA. (1980). On asymptotic minimax estimates of the second order (translation). *Theory Probab. Appl.* **25** 552–568.

LEVIT, B. YA. (1982). Minimax estimation and positive solutions of elliptic equations (translation). *Theory Probab. Appl.* **27** 563–568.

LEVIT, B. YA. (1985). Second order asymptotic optimality and positive solutions of Schrödinger's equations (translation). *Theory Probab. Appl.* **30** 333–363.

LEVIT, B. YA. (1987). On bounds for the minimax risk. In *Probability Theory and Mathematical Statistics* (Y. V. Prohorov, ed.) **2** 203–216. VNU Science, Utrecht.

MELKMAN, A. A. and RITOV, Y. (1987). Minimax estimation of the mean of a general distribution when the parameter space is restricted. *Ann. Statist.* **15** 432–442.

PAYNE, L. E. (1967). Isoperimetric inequalities and their applications. *SIAM Rev.* **9** 453–488.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DÉPARTMENT DE MATHÉMATIQUES
ET D'INFORMATIQUE
UNIVERSITÉ DU QUÉBEC
À MONTRÉAL
C. P. 8888, SUCC. "A"
MONTREAL
CANADA H3C 3P8