ANNALES DE L'I. H. P., SECTION B

IAIN M. JOHNSTONE
K. BRENDA MACGIBBON

Asymptotically minimax estimation of a constrained Poisson vector via polydisc transforms

Annales de l'I. H. P., section B, tome 29, n° 2 (1993), p. 289-319
<http://www.numdam.org/item?id=AIHPB_1993__29_2_289_0>

Ann. Inst. Henri Poincaré,

Vol. 29, n° 2, 1993, p. 289-319. *Probabilités et Statistiques*

Asymptotically minimax estimation of a constrained Poisson vector via polydisc transforms

by

Iain M. JOHNSTONE

Department of Statistics,
Stanford University, California 94305 U.S.A.

and

K. Brenda MACGIBBON

Département de Mathématiques et d'Informatique,
Université du Québec à Montréal,
C.P. 8888, Succursale « A », Montréal, Canada H3C 3P8

ABSTRACT. – Suppose that the mean σ of a vector of independent Poisson variates $(X_1, \ldots, X_p)$ lies in a subset $m\mathrm{T}$ of $\mathbb{R}^p$, where T is a bounded domain and $m>0$. We study the asymptotic behavior of the minimax risk $\rho(m\mathrm{T})$ and the construction of asymptotic minimax estimators as $m \nearrow \infty$, using the information normalized loss $\sum_{i=1}^{p} \sigma_i^{-1}(d_i-\sigma_i)^2$. With the use of the polydisc transform, a many-to-one mapping from $\mathbb{R}^{2p}$ to $\mathbb{R}_+^p$, we show that $\rho(m\mathrm{T})=p-m^{-1}\lambda(\Omega)+o(m^{-1})$, where $\lambda(\Omega)$ is the principal eigenvalue for the Laplace operator on the pre-image Ω of T under this transform. The proofs exploit the connection between p-dimensional Poisson estimation in T and $2p$-dimensional Gaussian estimation in Ω.

Key words : Polydisc transform, second-order minimax, Laplace operator, principal eigenvalue, Fisher information, minimax risk.

Classification A.M.S. : 62 F 10, 62 F 12.

RÉSUMÉ. — On suppose que la moyenne σ d'un vecteur $X=(X_1, \ldots, X_p)$ de p variables de Poisson indépendantes se trouve dans un sous-ensemble $m\mathrm{T}$ de $\mathbb{R}_+^p$, où T est (relativement) ouvert et borné et $m>0$. On étudie le comportement asymptotique du risque minimax et la construction des estimateurs asymptotiquement minimax quand $m \nearrow \infty$ avec la fonction de perte normalisée $\sum_{i=1}^{p} \sigma_i^{-1}(d_i-\sigma_i)^2$. Avec la transformation polydisque, une transformation de $\mathbb{R}^{2p}$ à $\mathbb{R}_+^p$, on démontre que $\rho(m\mathrm{T})=p-m^{-1}\lambda(\Omega)+o(m^{-1})$, où $\lambda(\Omega)$ est la plus petite valeur propre positive pour le problème de Dirichlet sur l'image inverse Ω de T sous la transformation polydisque. La démonstration utilise la relation entre l'estimation poissonnienne sur $\mathrm{T} \subset \mathbb{R}_+^p$ et l'estimation gaussienne sur $\Omega \subset \mathbb{R}^{2p}$.

1. INTRODUCTION

Let $X=(X_1, \ldots, X_p)$ be a vector of independent Poisson variates, having means $\sigma=(\sigma_1, \ldots, \sigma_p)$. This paper is concerned with minimax estimation of σ given the prior information that σ lies in a set $m\mathrm{T}$ and using the information normalised loss function

$$\mathrm{L}(d, \sigma)=\sum_{i=1}^{p} \sigma_i^{-1}(d_i-\sigma_i)^2.$$

We consider the asymptotic behavior of the minimax risk $\rho(m\mathrm{T})=\inf_{\delta} \sup_{\sigma \in m\mathrm{T}} \mathrm{E}_\sigma \mathrm{L}(\delta(X), \sigma)$ and the construction of asymptotically minimax estimators as $m \nearrow \infty$. This paper is a companion to Johnstone and MacGibbon (1992), henceforth called I, in which background motivation for the problem and a variety of non-asymptotic results were given.

The connection between asymptotic minimax estimation and the principal eigenvalue of elliptic equations was first elaborated in a series of papers by Levit (1980, 1982, 1985*a*) and Berkin and Levit (1980). They studied asymptotic second-order minimax estimators under a general class of loss functions in Gaussian and locally asymptotic Gaussian settings, and connected with the principal eigenvalue of the Laplace (or more generally, second order uniformly elliptic) equation in the domain in which the parameter lies. Bickel (1981) independently derived the results for intervals and spheres in the Gaussian setting for squared error loss.

Melkman and Ritov (1987) extended Bickel's univariate results to a class of location problems. Levit (1986) considers (amongst other things) the information normalised loss function for exponential families including Poisson and establishes a variety of second order admissibility results. Our approach to second-order asymptotic estimation in the Poisson case is inspired by Bickel's (1981) method for Gaussian data.

A fundamental role in our study is played by a many-to-one mapping $\tau : \mathbb{R}^{2p} \to \mathbb{R}_+^p$, called the polydisc transform, where

$$\tau : \quad (\omega_1, \omega_2, \ldots, \omega_{2p-1}, \omega_{2p}) \to (\omega_1^2 + \omega_2^2, \ldots, \omega_{2p-1}^2 + \omega_{2p}^2). \tag{1}$$

For each set T in $\mathbb{R}_+^p$, the Poisson mean parameter space, $\Omega = \tau^{-1}(\mathrm{T})$ will denote the pre-image of T. The name reflects the fact that the pre-image of a rectangle $[0, a] \subset \mathbb{R}_+^p$, namely

$$\{ \omega : \omega_{2i-1}^2 + \omega_{2i}^2 \leqq a_i, \ i = 1, \ldots, p \},$$

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. The virtue of the polydisc transform is that its inverse converts relatively unpleasant optimization problems for T into the well understood Dirichlet problem for the Laplace equation on Ω.

An asymptotic theory is obtained by approximating $\rho(m\mathrm{T})$ 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 chief purposes of the paper are

(1) To present conditions under which the asymptotic expansion

$$\rho(m\mathrm{T}) = p - m^{-1} \lambda(\Omega) + o(m^{-1}) \tag{2}$$

is valid. Here Ω is the pre-image of T under the *polydisc transform* τ defined by (1) and λ denotes the principal eigenvalue of the Laplace operator on Ω: *i.e.*, the smallest constant λ for which there exists a non-trivial solution to the equation

$$\left.\begin{aligned} \Delta u &= -\lambda u, \qquad \omega \in \Omega \\ u &= 0, \qquad \omega \in \partial\Omega. \end{aligned}\right\} \tag{3}$$

In the situations for which we establish (2), we also exhibit an asymptotically minimax sequence of estimators built from the principal eigenfunction of (3) corresponding to $\lambda(\Omega)$.

(2) To study the information-like functionals that arise in studying Bayes risks in Poisson estimation. We explore analogies with the role of Fisher information in Bayes estimation of a Gaussian mean vector. In the latter case, when $\mathrm{X} \sim \mathrm{N}_{2p}(\theta, \mathrm{I})$, Brown's identity connects the Bayes risk

$r(\mathrm{H}) = \inf_{\delta} \int \mathrm{E}_\theta |\delta(\mathrm{X}) - \theta|^2 \mathrm{H}(d\theta)$ for estimation of θ with absolutely continuous prior density $\mathrm{H}(d\theta) = h(\theta)\, d\theta$ with Fisher information $\mathrm{I}(\mathrm{H}) = \int |\mathrm{D}h|^2/h$ via the identity

$$r(\mathrm{H}) = 2p - \mathrm{I}(\mathrm{H} * \Phi), \tag{4}$$

where Φ denotes the standard Gaussian distribution in $\mathbb{R}^{2p}$. If $\theta = m\tau$ and the prior $\mathrm{H} = \sigma_m \mathrm{F}$ are transforms of $\mathrm{F}(d\tau)$ under the scaling $\sigma_m : \tau \to m\tau$, then

$$r(\sigma_m \mathrm{F}) = 2p - m^{-2} \mathrm{I}(\mathrm{F} * \Phi_{1/m}), \tag{5}$$

where $\Phi_{1/m}$ denotes the Gaussian distribution $\mathrm{N}_{2p}(0, m^{-1}\mathrm{I})$.

The corresponding quantity in estimation of a Poisson mean $\sigma = \sigma_m(\tau) = m\tau$ given prior $\mathrm{F}(d\tau) = f(\tau)\, d\tau$ is

$$r(\sigma_m \mathrm{F}) = p - m^{-1} \mathrm{J}_m(\mathrm{F}),$$

where for the present we take this as the definition of J_m.

As $m \to \infty$, we show that J_m approaches a limit

$$\mathrm{J}(f) = \int \sum f^{-1} (\mathrm{D}_i f)^2 (\tau)\, \tau_i\, d\tau. \tag{6}$$

The properties of J_m and J are essential to our method of establishing (2).

(3) To study the connection of the p-dimensional Poisson estimation problem with the $2p$-dimensional Gaussian location estimation problem induced by the transform (1). For example, the functional (6) is related to Fisher information via the identity

$$\mathrm{J}(f) = 4^{-1} \pi^{-p} \mathrm{I}(g), \qquad g(w) = f(\tau(\omega)). \tag{7}$$

The paper is structured as follows. Section 2 collects preliminary technical material on existence and uniqueness of solutions to the equation (3), smoothness conditions on the boundary $\partial\Omega$ and regularity properties of the solutions. Section 3 outlines the main results on asymptotic minimaxity and sketches the proof in order to bring out the roles of the functionals J_m and J. Section 4 focuses on the properties of the limiting functional J; and along the way we obtain a multivariate extension of Huber's (1964, 1981) operator norm characterisation of Fisher information for location. Section 5 studies the discrete functionals J_m and establishes the limiting continuity and semi-continuity relations connecting J_m with J. Finally, Section 6 collects details of the proofs.

HEURISTICS. — Here is an informal explanation of the connection with the Laplace equation. Assume first that the Poisson mean parameter σ

lies in S (later we set $S=m\,T$). The unbiased estimate of risk of any estimator $\delta_g(x)=x+g(x)$ has the form

$$R(\delta_g, \sigma)=E_\sigma\{p+\mathbf{D}(g, X)\} \tag{8}$$

$$=E_\sigma\{p+\sum_i 2[g_i(X+e_i)-g_i(X)]+(X_i+1)^{-1} g_i^2(X+e_i)\}, \tag{9}$$

where e_i denotes a unit vector in the i-th co-ordinate direction and equality (9) defines the difference operator $\mathbf{D}(g, X)$. When S is large, the standard deviation of X will be small relative to S, and we seek the smallest κ for which there exists a solution to

$$\mathbf{D}(g, x)\leqq -\kappa \qquad \text{for} \quad x\in S \tag{10}$$

(or, strictly speaking, for x in a neighborhood of S that contains X with high probability whenever $\sigma\in S$). Further, our heuristic purposes will be served by replacing inequality with equality.

The increments x to $x+e_i$ are small relative to the standard deviation $\sigma_i^{1/2}$ of typical parameter points in $S=mT$ for m large, so consider a differential equation approximation to (10):

$$\mathbf{D}(g, x)\approx\sum_i 2\,D_i g_i(x)+x_i^{-1} g_i^2(x). \tag{11}$$

Complete class theorems imply that the search for solutions g_i can be restricted to the class of Bayes rules. For large x, the Bayes rule corresponding to a prior density $p(\sigma)\,d\sigma$ has the form (Corollary 18)

$$\delta_i^{(p)}(x)\approx x_i+x_i p^{-1} D_i p(x). \tag{12}$$

Thus the vector of functions $g_i^{(p)}\approx x_i p^{-1} D_i p$ and is thus determined by the single function p. Although this could be substituted into (11), the variational discussion in Section 4 of I suggests that we write $p=q^2$ and substitute $g_i^{(q)}\approx 2\,x_i\,q^{-1} D_i\,q$ into (11). This yields

$$\mathbf{D}(2\,x_i q^{-1} D_i q, x)\approx 4\,q^{-1}\sum_i [x_i D_i^2\,q(x)+D_i\,q(x)]=4\,q^{-1}\,L\,q,$$

where L is the indicated second order differential operator. Since the prior density p and hence q is supported on S, we are again led to seek the smallest value of κ for which a solution exists to

$$4\,L\,q+\kappa\,q=0 \quad \text{on S}. \tag{13}$$

(Note that this heuristic argument does not seem to yield the boundary condition $q=0$ on ∂S.)

In studying admissibility in Poisson estimation for the information-normalized loss, Brown (1979, p. 983) noted the resemblance of the differential inequality for p-dimensional estimation to the differential inequality occurring in $2p$-dimensional Gaussian estimation for squared-error loss.

The polydisc transformation $\tau(\omega)$ of (1) is an elaboration of this observation. Defining $u(\omega)=q(\tau(\omega))$ on $\Omega=\tau^{-1}(S)$, computation shows that $(D^2_{2i-1}+D^2_{2i})\,u(\omega)=4x_i D^2_i q(x)+4D_i q(x)$ and hence that equation (13) takes the Laplacian form (3).

Error terms can be given (at least in order of magnitude) in the above heuristics if $S=mT$ and $m\to\infty$. Suppose that $f(\tau)$ is a probability density supported on T and that the sequence of priors $p_m(\sigma)$ is defined via $p_m(\sigma)\,d\sigma=f(\tau)\,d\tau$, with $\sigma=m\tau$. Then, if $x=mz$

$$\delta_i^{(p_m)}(mz)=mz_i+z_i f^{-1} D_i f(z)+o(1) \quad \text{as } m\to\infty \tag{14}$$

uniformly in z belonging to compact subsets of int T (Corollary 18). Writing $f=v^2$ as before, so that $q^2(\sigma)\,d\sigma=v^2(\tau)\,d\tau$ it is easily found that

$$q^{-1}\,L\,q(x)=m^{-1}\,v^{-1}\,L\,v(z), \qquad z\in T$$

and hence that the eigenvalues κ_+ for $S=mT$ in (13) are related to the eigenvalues λ_+ for T in (3) by $\kappa_+=\lambda_+/m$.

In particular, formula (14) also suggests the form of an asymptotically second-order minimax estimator in terms of the prior density $f(\tau)=u^2_\Omega(\omega)$, where $\Omega=\tau^{-1}(T)$ (*cf.* Theorem 6).

We conclude this section by collecting notation and definitions for later use.

NOTATION. – $\mathbf{Z}^p_+=\{n\in\mathbf{Z}^p: n_i\geqq 0 \text{ for all } i\}$. 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 $\psi=(\psi_1,\ldots,\psi_p)$ is a vector field, $D.\psi=\sum_i D_i\psi_i$.

Let $X\subset R^p$. Then $C^k_0(X, R^p)$ 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 $C^k_0(X)$, or as C^k_0 when $X=R^p$. We make the convention throughout that $0^0=1$ and $0/0=0$. The indicator function $I\{A\}$ of a set A will sometimes be denoted simply $\{A\}$: for example $g(x)I\{x\in B\}$ will be written $g(x)\{x\in B\}$.

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 $\rho(T)=\rho(\bar{T})$. we may and shall by convention choose T so that $T=\operatorname{int}\bar{T}$.

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

$$\Delta=\Delta(T)=\{\delta(x): x_i=0 \text{ and } i\in I(T) \text{ implies } \delta_i(x)=0\}, \tag{15}$$

since estimators not in Δ 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

$$r(\delta, F) = \int R(\delta, \tau) F(d\tau), \qquad r(F) = \inf_{\delta \in \Delta} r(\delta, F). \tag{16}$$

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

$$\rho(T) = \inf_{\Delta} \sup_{T} R(\delta, \tau) = \sup_{F^*(\bar{T})} \inf_{\Delta} r(\delta, F) = \sup_{F^*(\bar{T})} r(F). \tag{17}$$

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

2. HÖLDER CONTINUITY OF SOLUTIONS AND BOUNDARIES

This section contains preliminary technical information on properties of solutions to (3) and the essentially equivalent classical Dirichlet problem

$$\lambda(\Omega) = \inf \left\{ \int_{\Omega} |Du|^2 : u \in W_0^{1,2}(\Omega), \int_{\Omega} u^2 = 1 \right\}. \tag{18}$$

Here $W_0^{1,2}$ denotes the closure of $C_0^1(\Omega)$ in $W^{1,2}(\Omega)$, the Sobolev space consisting of once-weakly differentiable functions having norm $\|u\|^2_{W^{1,2}(\Omega)} = \int_{\Omega} |Du|^2 + u^2 < \infty$. We use Gilbarg and Trudinger (1983, GT) as a convenient reference for certain standard definitions, notation and theorems. *See* also Levit (1982, Theorem 5) for an overview. The following is a standard result (*see* GT, p. 214, for example).

THEOREM 1. — (1) *There is a unique (up to sign) function* $u_\Omega \in W_0^{1,2}(\Omega)$ *achieving the minimum in* (18). *It satisfies the equation*

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

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

(3) *The minimum eigenvalue* $\lambda(\Omega)$ *is monotone in* $\Omega : \Omega \subset \Omega'$ *implies* $\lambda(\Omega) \geqq \lambda(\Omega')$.

HÖLDER CONTINUITY. — We shall need Hölder-continuity properties of u_Ω to establish weak convergence of the least favorable distributions and to verify asymptotic minimaxity of estimator (25). Let $\mathbf{j} = (j_1, \ldots, j_p)$ be a multi-index, $|\mathbf{j}| = \sum j_i$ and $\alpha \in (0, 1]$. Following the conventions of GT

(Section 4.1), define

$$\left.\begin{aligned} |u|_{\Omega} &= \sup_{x\in\Omega} |u(x)| \\ [u]_{\alpha;\,\Omega} &= \sup\left\{\frac{|u(x)-u(y)|}{|x-y|^{\alpha}},\ x,\ y\in\Omega,\ x\neq y\right\} \\ \|u\|_{C^{k,\alpha}(\bar{\Omega})} &= \sum_{j\leq k}\sup_{|\mathbf{j}|=j}|D^{\mathbf{j}}u|_{\Omega}+\sup_{|\mathbf{j}|=k}[D^{\mathbf{j}}u]_{\alpha;\,\Omega}. \end{aligned}\right\} \tag{19}$$

We need the cases $k=0$ and 2. In particular, when $k=0$, $\|u\|_{C^{\alpha}(\bar{\Omega})}=|u|_{\Omega}+[u]_{\alpha;\,\Omega}$, and the following easily checked properties will be used later.

$$\left.\begin{aligned} &\|uv\|_{C^{\alpha}(\bar{\Omega})}\leqq\|u\|_{C^{\alpha}(\bar{\Omega})}\|v\|_{C^{\alpha}(\bar{\Omega})} \\ &\text{If } \inf_{\Omega} u=m>0, \qquad \|u^{-1}\|_{C^{\alpha}(\bar{\Omega})}\leqq m^{-2}\|u\|_{C^{\alpha}(\bar{\Omega})}. \end{aligned}\right\} \tag{20}$$

BOUNDARY ASSUMPTIONS. — A domain Ω is said to be of class $C^{2,\alpha}$ if at every point $\omega_0\in\partial\Omega$, there exists a change of co-ordinates $u=u(\omega)$ having Hölder continuous second derivatives with index $\alpha\in(0,1]$ in which Ω is specified near ω_0 by the inequality $u_1>0$. We call Ω $C^{2,\alpha}$-*approximable* if there exists a sequence of $C^{2,\alpha}$ domains $\Omega_{\varepsilon}\uparrow\Omega$ with $\lambda(\Omega_{\varepsilon})\downarrow\lambda(\Omega)$. A domain $T\subset R_{+}^{p}$ is called $C^{2,\alpha}$ approximable if $\tau^{-1}(T)$ is also.

Consider now domains $T\subset R_{+}^{p}$. The relevant part of the boundary of T is $\partial\bar{T}$. Here boundary is computed in the relative topology of R_{+}^{p}, thus for example $\partial[0,m]=\{m\}$, and $\partial\{\tau:\tau_1+\tau_2\leqq 1\}=\{\tau:\tau_1+\tau_2=1\}$. In particular, $\tau^{-1}(\bar{T}\setminus\partial\bar{T})=\tau^{-1}(\operatorname{Int}\bar{T})=\operatorname{Int}\bar{\Omega}$. For an arbitrary $\tau_0\in\partial\bar{T}$, let $n(\tau_0)$ denote the number of zero components of τ_0. Call T of class $C^{2,\alpha}$ if at each τ_0, there is a $C^{2,\alpha}$ change of co-ordinates $\sigma=\sigma(\tau)$ in which T is specified near τ_0 by the inequalities $\sigma_i>0$ for $1\leqq i\leqq n(\tau_0)+1$. This definition is consistent with that for $\Omega=\tau^{-1}(T)$ in the following sense:

LEMMA 2. — *If* T *is of class* $C^{2,\alpha}$ *then* $\Omega=\tau^{-1}(T)$ *is also.*

(The proof is deferred to Section 6. We believe the converse to be true also.)

There are certain sufficient conditions for a domain Ω to be $C^{2,\alpha}$ approximable. Firstly, Berkhin and Levit (1980) call Ω a "$C^{2,\alpha}$ domain with non-zero corners" if at each $\omega_0\in\partial\Omega$ there exists a $C^{2,\alpha}$ co-ordinate change $u(\omega)$ in which Ω is specified near ω_0 by the inequalities $u_i>0$, $1\leqq i\leqq j=j(\omega_0)$. Similarly, call $T\subset R_{+}^{p}$ a $C^{2,\alpha}$ domain with non-zero corners if for each $\tau_0\in\partial\bar{T}$, there is a $C^{2,\alpha}$ change of co-ordinates σ, in which T is specified near τ_0 by the inequalities $\sigma_i>0$, $1\leqq i\leqq j_1(\tau_0)$, with $j_1(\tau_0)>n(\tau_0)$. The obvious extension of Lemma 2 in conjunction with Proposition 3 below ensures that this is a sufficient condition for $C^{2,\alpha}$ approximability of T.

Secondly, Dancer (1988) proves continuous dependence of $\lambda(\Omega)$ on Ω under what we shall call *Dancer's condition* on Ω: let B be a ball containing Ω. If $u \in W^{1,2}(B)$ and $u=0$ on $(B \setminus \Omega)$, then $u \in W_0^{1,2}(\Omega)$. We believe that Dancer's condition holds for $\Omega = \tau^{-1}(T)$ for essentially all sets T arising in applications. In particular, we expect it to hold for sets T which at each $\tau_0 \in \partial \bar{T}$ are specified (after a smooth co-ordinate change σ) by a finite number of linear inequalities on σ.

PROPOSITION 3. — (i) *Let Ω be a piecewise-smooth domain with non-zero corners and let Ω_ε be a sequence of regions of the same type tending uniformly to Ω. Then $\lambda(\Omega_\varepsilon) \to \lambda(\Omega)$. In particular, Ω is $C^{2,\alpha}$-approximable.*

(ii) *Suppose that Ω is a domain satisfying Dancer's condition and such that there exists a sequence of $C^{2,\alpha}$ domains $\Omega_\varepsilon \uparrow \Omega$ such that for any compact $K \subset \Omega$, $\Omega_\varepsilon \supseteq K$ for small ε. Then Ω is $C^{2,\alpha}$-approximable.*

Proof. — An informal argument is given in Courant-Hilbert (1953, Theorem VI.11). Part (i) is proved in Berkhin and Levit (1980, Theorem 7). For part (ii) apply Dancer's (1988) Theorem 1 to $f(u)=\mu u$, $u_0=0$ with $\mu=\lambda(\Omega)\pm\delta$ for sufficiently small δ to conclude that $\lambda(\Omega_\varepsilon) \in (\lambda(\Omega)-\delta, \lambda(\Omega)+\delta)$ for $\varepsilon \leqq \varepsilon(\delta)$. ■

Finally, we collect properties of solutions to the Dirichlet problem (3) that depend on the above smoothness hypotheses.

THEOREM 4. — (i) *If the domain Ω is of class $C^{2,\alpha}$, then a) $u_\Omega \in C^{2,\alpha}(\bar{\Omega})$ and b)*

$$\| u_\Omega \|_{C^{2,\alpha}(\bar{\Omega})} \leqq C_1 (\| \partial\Omega \|^*_{2,\alpha}, \operatorname{vol}(\Omega)).$$

(ii) *If Ω is of class $C^{2,\alpha}$ there exists $C_2>0$ such that $\partial u_\Omega/\partial n \geqq C_2$ on $\partial\Omega$, where n is the inner normal to Ω.*

(iii) *Let $\Omega^\varepsilon = \{\omega : d(\omega, \Omega) < \varepsilon\}$, where $d(.,.)$ denotes Euclidean distance. For sufficiently small ε, Ω^ε is a domain of class $C^{2,\alpha}$ and the constants C_1, C_2 in* (i) *and* (ii) *may be taken independent of* ε.

Proof. — Part (i)*a*) is in GT, p. 214, part (i)*b*) may be found in Ladyzhenskaya and Ural'tseva (1968, 1973, chap. 3) though Levit (1982, Theorem 5) makes the dependence on Ω more explicit. In particular, $\| \partial\Omega \|^*_{2,\alpha}$ is a Hölder norm on $\partial\Omega$ defined by Levit. Part (ii) follows from Giraud's theorem (Miranda, 1970, p. 7) and the positivity of u_Ω. Finally, part (iii) follows from part (i) and the refinement of Giraud's theorem given by Berkhin and Levit (1980). ■

3. MAIN ASYMPTOTIC RESULTS AND OUTLINE OF PROOFS

Let independent $X_i \sim$ Poisson $(m\tau_i)$, $i=1, \ldots, p$ and suppose that $\tau \in T$, a relatively open connected subset of R^p_+ having compact closure. Suppose also that T is $C^{2,\alpha}$ approximable. As in Section 1, let $\Omega=\tau^{-1}(T)\subset R^{2p}$, where $\tau_i(\omega)=\omega_{2i-1}^2+\omega_{2i}^2$, $i=1,\ldots,p$. Finally set $\sigma=m\tau$.

THEOREM 5. — (i) *The minimax risk*

$$\rho(mT)=\inf_{\delta\in\Delta}\sup_{\sigma\in mT}E_\sigma\sum_{i=1}^{p}\sigma_i^{-1}[\delta_i(X)-\sigma_i]^2=p-m^{-1}\lambda(\Omega)+o(m^{-1}). \quad (21)$$

(ii) $\lambda(\Omega)$ *denotes the minimum eigenvalue of the Laplace operator* $\Delta=\sum_1^{2p}\partial^2/\partial\omega_j^2$ *on* Ω, i.e. *the smallest* λ *for which the equation*

$$\left.\begin{aligned}\Delta u(\omega)&=-\lambda u(\omega), &\omega\in\operatorname{int}\Omega\\ u(\omega)&=0, &\omega\in\partial\Omega\end{aligned}\right\} \quad (22)$$

has a non-zero solution. The eigenspace corresponding to $\lambda(\Omega)$ *is one-dimensional, and the corresponding eigenfunction* $u_\Omega(\omega)=u(\omega,\Omega)$ [*or* $-u_\Omega(\omega)$] *is strictly positive on* Ω. *Assume that* u_Ω *is normalised so that* $\int_\Omega u_\Omega^2=1$.

(iii) *Let* $P_m(d\sigma)$ *denote a least favorable prior distribution for the region* $S_m=mT$ *and* $F_m(d\tau)$ *the corresponding prior rescaled to* T. *A probability density may be unambiguously defined on* T *by* $f_0(\tau)=c_p u_\Omega^2(\omega)$, (*with* $c_p=\pi^{-p}$) *and the measure* $F_0(d\tau)=f_0(\tau)\,d\tau$ *is the weak limit of the* (*rescaled*) *least favorable distributions* F_m.

The proof of this theorem is spread over the following sections and is outlined at the end of this section. In the companion paper I, it was shown that $\rho(T)\geqq p^2/(p+\lambda(\Omega))$. Theorem 5 states that this bound is asymptotically sharp.

Now, assume further that T is a domain in R_+^p of class $C^{2,\alpha}$, and let the η-extension of T be defined by $T^\eta=\{\tau\in R_+^p : d(\tau, T)<\eta\}$,where $d(.,.)$ denotes Euclidean distance. If Ω is a domain in R^{2p}, we may similarly define Ω^ε, and it may be shown that $\tau^{-1}(T^\eta)\subset\Omega^{\eta^{1/2}}$ [Lemma 19 (ii)]. Define

$$u_m(\omega)=u(\omega,\Omega^{2\eta_m^{1/2}}),\qquad f_m(\tau)=u_m^2(\omega) \quad (23)$$

$$h_{m,i}(\tau)=\tau_i(f_m^{-1}D_i f_m)(\tau). \quad (24)$$

THEOREM 6. — *For T as above, if* $\eta_m=m^{-\beta}$, *for* $0<\beta<\alpha/2$, *then*

$$\delta_m(x)=x+h_m(xm^{-1})\{x\in mT^{\eta_m}\}. \quad (25)$$

is a second order asymptotically minimax estimator of σ *in* mT:

$$\sup_{mT}R(\delta_m,\sigma)-\inf_\delta\sup_{mT}R(\delta,\sigma)=o(m^{-1}).$$

The proof involves substitution of estimator (25) into the unbiased risk estimate exhibited in (8) and is deferred to the Appendix.

STARSHAPED DOMAINS. – We shall call a domain $T \subset R^p_+$ strictly star-shaped relative to τ_0 if (i) $\tau \in T$ implies that the closed segment $\overline{\tau_0 \tau}$ lies in T, and (ii) for no point $\tau \in \partial_1 T$ does $\tau_0 - \tau$ lie in the tangent plane to $\partial_1 T$ at τ. For strictly star-shaped $C^{2,\alpha}$ domains, a slightly simpler construction of a second order asymptotically minimax estimate can be given in terms of the least favorable prior corresponding to the domain T. For brevity, we describe only the special case in which $\tau_0 = 0$. Then set $f(\tau) = u^2_\Omega(\omega)$ and

$$\tilde{\delta}_{mi}(x) = x_i + \gamma_m m^{-1} x_i (f^{-1} D_i f)(\gamma_m m^{-1} x) \{ x \in \zeta_m m T \}.$$

Here $\gamma_m = 1 - 3\varepsilon_m$, $\zeta_m = 1 + \varepsilon_m$ and $\varepsilon_m = m^{-\beta}$ for $0 < \beta < \alpha/4$.

Examples. – In the companion paper I, the explicit forms of the asymptotically least favorable density $f_m(\tau)$ from (22) are given for different domains T such as rectangles, solid simplexes and hyperrectangles. By use of the polydisc transform, these densities are expessed in terms of Bessel functions.

In the case of a solid simplex in $\mathbb{R}^p_+$ $(p \geqq 2)$, we obtain the following consequence of Theorem 6. An analogous result for Gaussian data is noted by Bickel (1981, p. 1307).

COROLLARY 7. – *For each* $p \geqq 2$, $\delta_m(x) \to \delta_{CZ}(x) = (1 - (p-1)/\sum x_i) x$ *as* $m \to \infty$. *Thus,* $\delta_{CZ}(x)$ *is minimax estimator of* λ *in* $\mathbb{R}^p_+$.

Clevenson and Zidek (1975) introduced δ_{CZ} as a minimax estimator of σ that dominates the maximum likelihood estimator in terms of risk. The corollary is established by noting from paper I that

$$f_m(\tau) = c(|\tau| m^{-1} v^2_{p-1})^{1-p} J^2_{p-1}(|\tau|^{1/2} m^{-1/2} v_{p-1})$$

for $0 \leqq |\tau| = \sum_1^p \tau_i \leqq m$, where $J_p(\tau)$ is the Bessel function of the first kind of index p and v_p is its smallest positive zero. Now take limits as $m \to \infty$ in (23) and (24).

Proof of Theorem 5 *(Outline).* – Suppose that $F(d\tau)$ is a probability measure supported in $\bar{T}$. Let $\sigma_m(\tau) = m\tau$ and denote by $\sigma_m F$ the rescaled measure in $m\bar{T}$. Recall that $r(P)$ denotes the Bayes risk for prior P [*cf.* (16)]. Define

$$J_m(F) = m[p - r(\sigma_m F)], \tag{26}$$

more explicit representations appear in Section 5.

If F has a weakly differentiable density f, we have defined $J(f) = \int \sum_i \tau_i [D_i f(\tau)]^2 / f(\tau) \, d\tau$, and noted at (7) that $J(f) = c I(\tau^{-1} F)$

where $I(.)$ is a scalar form of Fisher information for multivariate distributions. In Section 4, we pursue the analogy with Fisher information and define an extension of J to all probability measures F.

Let P_m^* be a sequence of least favorable distributions for the sets mT, and let $F_m^* = \sigma_m^{-1} P_m^*$ be measures rescaled to have support in $\bar{T}$. Since $\bar{T}$ is compact, the sequence $\{F_m^*\}$ has weak limits that are probability measures supported on $\bar{T}$. Let F^* be any such weak limit: the key to the proof of (iii) lies in showing that $J(F^*) \leqq J(F_0)$.

Consider first the case in which T (and hence Ω, by Lemma 2) is of class $C^{2,\alpha}$. Then the following sequence of inequalities is fundamental:

$$J(F^*) \leqq \liminf J_m(F_m) \leqq \limsup J_m(F_m) \leqq \lim J_m(F_0) = J(F_0) < \infty. \qquad (7)$$

The first inequality follows from the joint lower semi-continuity of $(m, F) \to J_m(F)$ (Theorem 15). The second is trivial, while the third reflects that fact that F_m is least favorable and by definition minimizes J_m. The fourth equality expresses the convergence of J_m to J for sufficiently nice measures (Theorem 16): this step uses the regularity properties of solutions to the boundary problem (22) for smooth domains.

If the domain Ω is merely $C^{2,\alpha}$ approximable, then choose a sequence of $C^{2,\alpha}$ domains $\Omega_\varepsilon \subset \Omega$ for which $\lambda(\Omega_\varepsilon)$ decreases to $\lambda(\Omega)$. Since $\Omega_\varepsilon \subset \Omega$, $J_m(F_m) \leqq J_m(F_\varepsilon)$ and so the argument leading to (27) shows that $J(F^*) \leqq J(F_\varepsilon)$ where $dF_\varepsilon/d\tau = c_p^2 u_{\Omega_\varepsilon}(\omega)$. Since (by Theorem 1) $J(F_\varepsilon) = \lambda(\Omega_\varepsilon)$ and decreases to $\lambda(\Omega) = J(F_0)$, it follows that $J(F^*) \leqq J(F_0)$.

Since $\limsup J_m(F_m^*) < \infty$, it follows from Theorem 15 (ii) that $F^*(\overline{\partial_0 T}) = 0$. On the other hand, finiteness of $J(F^*)$ entails (Theorem 9) that F^* is absolutely continuous on $(0, \infty)^p$ relative to Lebesgue measure and that its density f^* integrates to 1. We know from Theorem 1 (1), however, that $f_0 = dF_0/d\tau$ is the *unique minimizer of* $J(f_0)$ *amongst probability* densities. Consequently $\mathbf{F^* = F_0}$, which is thus the single weak limit of $\{F_m^*\}$. Equality must hold throughout in (27) so that

$$\lim J_m(F_m) = J(F_0) = \lambda(\Omega).$$

This shows that

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

and completes the (outline of the) proof of Theorem 5.

Remark. – The strategy (27) for proving convergence of least favorable distributions is modelled after that of Bickel (1981) in the univariate Gaussian case. There, translation invariance and the continuous sample space allow the analogue of $J_m(F)$ to be represented as $I(F * \Phi_{1/m})$ where $\Phi_{1/m}$ denotes the $N(0, 1/m)$ density, and $I(F)$ is the usual Fisher information for an absolutely continuous measure F [*cf.* (5)]. In the Poisson setting, $J_m(F)$ is not related so directly to $J(F)$. Indeed, each J_m is derived from a discrete sample space, whereas the limit J is continuous. It is helpful to develop representations of J_m and J in terms of test functions (in the Schwartz distribution sense). This is inspired by Huber's (1964, 1981) "test-function" interpretation of Fisher information and some related notions for discrete distributions in Johnstone and MacGibbon (1987). Sections 4 and 5 are largely concerned with developing these representations and further properties of J_m and J required to establish (27).

4. TOTAL FISHER INFORMATION AND ITS POISSON RELATIVES

The purpose of this section is to extend the definition of $J(f)$ in (6) to arbitrary measures F for use in the proof of Theorem 5. The extension is similar in spirit to the extension of Fisher information (for location) from densities to measures on R given by Huber (1964, 1981). We begin by carrying over Huber's construction to "Fisher information" for measures on R^p as this setting is simpler and yet contains most of the ideas needed for the Poisson case also.

4.1. Scalar Fisher information for multivariate location

Let F be a probability measure on R^p and define

$$I(F) = \sup_{\psi \in C_0^1(R^p, R^p)} \frac{\left(\int D.\psi\, dF\right)^2}{\int |\psi|^2\, dF}. \tag{28}$$

THEOREM 8. — *The following two statements are equivalent*

(i) $I(F) < \infty$,

(ii) F *is absolutely continuous with respect to Lebesgue measure with density f that is weakly differentiable and*

$$\int |Df/f|^2 f < \infty.$$

In either case $I(F) = \int |Df/f|^2 f.$

Remark. — Steele (1986) has described a notion of finite Fisher information for densities on R^p. Theorem 8 shows that our definition is consistent with his.

Proof. — (ii) $\Rightarrow$ (i) This is an easy consequence of the Cauchy-Schwartz inequality and the divergence theorem:

$$\left(\int f\mathrm{D}.\psi\right)^2=\left(-\int\psi.\mathrm{D}f\right)^2\leqq\int|\psi|^2 f\int|\mathrm{D}f|^2/f. \tag{29}$$

(i) $\Rightarrow$ (ii) The proof that F is absolutely continuous is by induction on p, the case $p=1$ being Huber's result. Set $y=(x_1, \ldots x_{p-1})$, $z=x_p$, and decompose $\mathrm{F}(dx)=\mathrm{F}_1(dy)\,\mathrm{F}_2(dz\,|\,y)$. F_1 is the marginal distribution of y and F_2 is a regular conditional distribution for z given y. One easily verifies that $\mathrm{I}(\mathrm{F}_1)\leqq\mathrm{I}(\mathrm{F})<\infty$: if $\phi\in\mathrm{C}_0^1(\mathrm{R}^{p-1}, \mathrm{R}^{p-1})$, consider $\psi_i^{(n)}(y, z)=\phi_i(y)\,\gamma^{(n)}(z)$ for $i\leqq p-1$, and 0 for $i=p$, where $\gamma^{(n)}\in\mathrm{C}_0^1(\mathrm{R})\nearrow 1$ in a suitable manner. The induction hypothesis guarantees that F_1 is absolutely continuous with density f_1. Consider the mapping T: $\mathrm{C}_0^1(\mathrm{R}^p, \mathrm{R}^p)\to\mathrm{R}$ given by $\mathrm{T}\psi=-\int\mathrm{D}.\psi\,d\mathrm{F}$. Write $\|\psi\|^2=\int|\psi|^2\,d\mathrm{F}$ and note that

$$\sup_{\mathrm{C}_0^1(\mathrm{R}^p,\mathrm{R}^p)}\frac{|\mathrm{T}\psi|^2}{\|\psi\|^2}=\mathrm{I}(\mathrm{F})<\infty.$$

Thus T may be extended to a linear functional on all of $\mathrm{L}^2(\mathrm{R}^p, d\mathrm{F})$ with the same norm. The Riesz representation theorem provides a vector field $k\in\mathrm{L}^2(\mathrm{F})$ such that $\mathrm{T}\psi=\int\psi.k\,d\mathrm{F}$.

Define $g(z\,|\,y)=-\int_{s>z}k_p(y, s)\,\mathrm{F}_2(ds\,|\,y)$. If $\psi\in\mathrm{C}_0^\infty(\mathrm{R}^p)$, we may apply Fubini's theorem to obtain

$$\begin{aligned}\iint\mathrm{D}_p\psi(y, z)\,g(z\,|\,y)f_1(y)\,dy\,dz&\\ =-\iint f_1(y)\,dy\,\mathrm{F}_2(ds\,|\,y)\,k_p(y, s)&\int_{-\infty}^{s}\mathrm{D}_p\psi(y, z)\,dz\\ =-\int(\psi k_p)(y, z)\,\mathrm{F}(dy, dz)&=\int\mathrm{D}_p\psi(y, z)\,\mathrm{F}(dy, dz).\end{aligned} \tag{30}$$

Let $\bar{\mathrm{F}}(dy, dz)=g(z\,|\,y)\,f_1(y)\,dy\,dz$: we show that $\bar{\mathrm{F}}$ and F are the same measure by verifying that (30) implies $\int\phi\,d\bar{\mathrm{F}}=\int\phi\,d\mathrm{F}$ for all $\phi\in\mathrm{C}_0^\infty(\mathrm{R}^p)$. Fix $\phi\in\mathrm{C}_0^\infty$ and $\varepsilon>0$. Choose R so that $\mathrm{K}=\operatorname{supp}\phi\subset\mathrm{B}_\mathrm{R}(0)$, the ball about 0 of radius R. Define $\psi_n\in\mathrm{C}_0^\infty(\mathrm{R}^p)$ by $\psi_n(y, z)=\int_{-\infty}^{z}\mathrm{D}_p\psi_n(y, z)\,dz$, where we set $D_p \psi_n(y, z) = \phi(y, z) - \phi(y, z - nR)$. Clearly

$$\left| \int D_p \psi_n \, dF - \int \phi \, dF \right| \leqq \| \phi \|_\infty F(K + n R e_p) \to 0$$

as $n \to \infty$. To verify the same result for $\bar{F}$, note that the Cauchy-Schwartz inequality implies

$$g^2(z \mid y) \leqq \int_{s > z} k_p^2(y, s) F_2(ds \mid y)$$

and hence that

$$H(z) = \int | g(z \mid y) | f_1(y) \, dy \leqq \int g^2(z \mid y) f_1(y) \, dy$$

$$\leqq \int_{z > s} k_p^2 \, dF \to 0 \quad \text{as } z \to \infty, \tag{31}$$

since $k_p \in L^2(dF)$. Consequently

$$| \bar{F}(K + n R e_p) | \leqq \int_{(n-1)R}^{(n+1)R} dz \int | g(z \mid y) | f_1(y) \, dy$$
$$\leqq 2R \sup \{ H(z) : | z - nR | < R \} \to 0,$$

as $n \to \infty$. This establishes that $f(y, z) = g(z \mid y) f_1(y)$ is indeed a version of the density of F relative to Lebesgue measure.

To complete the proof, we note that kf is integrable since $\int | k_i f | \leqq \int k_i^2 f \int f < \infty$. Equating the two representations for T shows that

$$\int \psi . kf = - \int f D . \psi \qquad \text{for} \quad \psi \in C_0^1(R^p)$$

It follows from the definitions (e. g. Gilbarg and Trudinger, 1983, p. 149) that f is weakly differentiable with $Df = kf$. Again from the Riesz theorem, we conclude that

$$I(f) = \| T \|^2 = \int | k |^2 f = \int | Df/f |^2 f. \quad \blacksquare$$

4.2. Poisson analogue

The space of test vector fields $\chi = (\chi_i(\tau), i = 1, \ldots, p)$ is given by

$$\mathbf{X} = \{ \chi \in C_0^1(R_+^p, R^p) \text{: for some } \varepsilon > 0 \text{ and all } i, \chi_i(\tau) = 0 \text{ if } \tau_i < \varepsilon \}. \tag{32}$$

The reason for the extra condition appears in the proof below. Let $F(d\tau)$ be a probability measure on R^p_+. Define

$$J(F)=\sup_{\mathbf{X}}\frac{\left(\int\sum_i D_i\chi_i(\tau)\,dF(\tau)\right)^2}{\int\sum_i \chi_i^2\tau_i^{-1}\,dF(\tau)} \tag{33}$$

and for a weakly differentiable probability density on $(0, \infty)^p$, set

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

No ambiguity results if $F(\partial R^p_+)=0$: this is guaranteed by the following analogue of Theorem 8.

THEOREM 9. — *If* $J(F)<\infty$ *then* (i) F *is absolutely continuous with respect to Lebesgue measure on* $(0, \infty)^p$, *with density* f, (ii) f *is weakly differentiable on* $(0, \infty)^p$ *and* (iii) $J(f)<\infty$. *If also* $F(\partial R^p_+)=0$ *then* $J(F)=J(f)$.

Conversely if (i)-(iii) *hold and* $F(\partial R^p_+)=0$, *then* $J(F)=J(f)<\infty$.

Remark. — Note that nothing is said about F on ∂R^p_+. The result could be strengthened by adapting the method of proof to cases in which $F(\partial R^p_+)>0$, but we do not do this here.

Proof. — The converse follows as in the Fisher information case, so long as we note that the equality $\int_T fD.\chi=-\int_T\chi.Df$ (for an R^p_+-open set T containing supp χ) depends on the vanishing of the boundary term $\int_{\partial T} f\chi.n\,d\sigma$. Fortunately, the normal component $\chi.n$ vanishes on ∂R^p_+ by the very choice of $\mathbf{X}$.

The remainder of the theorem is proved by minor modifications of the proof of Theorem 8. Again the proof is by induction, involving the decomposition $F(d\tau)=F_1(dy)\,F_2(dz\mid y)$ for $\tau=(y, z)$. For $\chi\in\mathbf{X}$, set $T\chi=-\int D.\chi\,dF$ and $\|\chi\|^2=\int\sum_i\chi_i^2(\tau)\tau_i^{-1}\,dF(\tau)$. As before, we may assume that $F_1(dy)=f_1(y)\,dy$. The finiteness of $J(F)$ permits us to extend T to the completion H of $\mathbf{X}$ in $L^2(\tau_i^{-1}\,dF)$ with $\|T\|=J^{1/2}(F)$. Consequently, there exists a representer $k\in L^2(\tau_i^{-1}\,dF)$ of T such that

$$T\chi=\int\sum_i\chi_i k_i\tau_i^{-1}\,dF=-\int D.\chi\,dF \qquad \text{for} \quad \chi\in H. \tag{34}$$

Define $g(z \mid y) = -\int_{s>z} k_p(y, s) s^{-1} F_2(ds \mid y)$ and argue as before that $\int D_p \chi \, dF = \int D_p \chi \, d\bar{F}$, for $\bar{F}(dy, dz) = g(z \mid y) f_1(y) \, dy \, dz$. The analogue of (31) is $H(z) \leqq z^{-1} \int_{s>z} k_p^2 \, dF \to 0$ as $z \to \infty$, so it follows that $F = \bar{F}$ on $(0, \infty)^p$, and hence that F has density $f(\tau) = f_1(y) g(z \mid y)$ on $(0, \infty)^p$.

If $\chi \in C_0^1((0, \infty)^p)$, then we may replace dF by $f \, d\tau$ in (34). It follows that $k_i \tau_i^{-1} f$ is integrable on compact subsets of $(0, \infty)^p$, and hence that f is weakly differentiable on $(0, \infty)^p$ with $D_i f = k_i \tau_i^{-1} f$.

Finally, if $F(\partial R_+^p) = 0$, then

$$J(F) = \| T \|^2 = \int \sum_i k_i^2 \tau_i^{-1} \, dF = \int \sum_i (f^{-1} D_i f)^2 \tau_i f \, d\tau. \quad \blacksquare$$

5. INFORMATION FUNCTIONALS FOR PRIORS AND CONTINUITY

The information functionals $J_m(F)$ defined in (26) play a basic role in the proof of Theorem 5. Although only the limiting form $J(F)$ bears an explicit relation to Fisher information $I(F)$, through the polydisc transform (7), many of the standard properties of I (collected, for example, in Chapter 4 of Huber, 1981) have easily established analogues for each J_m which we shall exploit. Proofs are collected in Section 6.

Define the marginal density of P by $\pi(P)(x) = \int p_\tau(x) P(d\tau)$, or just $\pi(x)$ or π_x when there is no ambiguity. Define also

$$\hat{p}(x) = \pi(x) x! = \int \prod_i e^{-\tau_i} \tau_i^{x_i} P(d\tau).$$

We consider now representations of Bayes estimators. If i does not index a critical face [defined at (15)], then $\hat{p}(x) < \infty$ even if $x_i = -1$. Introduce indicators $\kappa_i = 1$ if i is a critical face and $\kappa_i = 0$ otherwise, and let $I = I(T) = \{ i : \kappa_i = 1 \}$. The Bayes rule corresponding to P in $\Delta(T)$, [*cf.* (15)], has the representation

$$\delta_{P,i}(x) \hat{p}(x - e_i) = \hat{p}(x) \qquad \text{if} \quad x_i \geqq \kappa_i. \tag{35}$$

See also Remark 1 in Section 6. When $x_i \geqq 1$, this may be rewritten as

$$\delta_{P,i}(x) = x_i + x_i [\pi(x) - \pi(x - e_i)] / \pi(x - e_i). \tag{36}$$

In the following definitions, in addition to the indicated ranges, the sums are taken only over those x for which $\hat{p}(x-e_i)>0$.

$$\mathrm{K}_\varkappa(\hat{p})=\sum_i \sum_{x_i \geqq \varkappa_i} \frac{[\hat{p}(x)-x_i \hat{p}(x-e_i)]^2}{\hat{p}(x-e_i)\, x!}, \tag{37}$$

$$\mathrm{K}(\pi)=\sum_i \sum_{x:x_i>0} \frac{[\pi(x)-\pi(x-e_i)]^2}{\pi(x-e_i)}, \tag{38}$$

$$\Delta_m(\hat{p})=\sum_{i \notin \mathrm{I}} \sum_{x_i=0} \delta_{\mathrm{P},\, i}(x)\, \pi(x). \tag{39}$$

The following Poisson analogue of Brown's (1971) identity expresses the Bayes risk of a prior distribution in terms of differences of the marginal density of P.

LEMMA 10:

$$p-r(\mathrm{P})=\mathrm{K}_\varkappa(\hat{p})=\mathrm{K}(\pi)+\Delta_m(\hat{p}) \tag{40}$$

Remark. – Given a probability distribution $\pi(x)$ on $\mathbf{Z}_+^p$, the functional $\mathrm{K}(\pi)$ satisfies the inequality

$$\mathrm{K}(\pi) \geqq p^2 \Big/ \sum_i (\mathrm{EX}_i+1), \qquad \Big(\mathrm{EX}_i=\sum_x x_i\, \pi(x)\Big),$$

with equality if and only if π is the product of independent geometric distributions (possibly with unequal success probabilities). This follows from the Cauchy-Schwartz inequality applied to $\sum_i \sum_{x_i>0} [\pi(x)-\pi(x-e_i)]\, x_i$.

Let $\mathbf{P}^*(\mathrm{X})$ denote the collection of probability measures supported in X. The representations (40) and (37) make it easy to show that the least favorable distribution is unique, using

LEMMA 11. – *The function* $\mathrm{P} \to r(\mathrm{P})$ *is strictly concave on* $\mathbf{P}^*(\bar{\mathrm{T}})$.

A "test function" representation for J_m that corresponds to that of (33) for J is useful in studying convergence of J_m to J. We derive this first for the functional $\mathrm{K}_\varkappa$, using a discrete analogue of the function class $\mathbf{X}$ defined in (32), namely $\Phi_\varkappa=\{\phi:\mathbf{Z}_+^p \to \mathbf{R}^p$: each ϕ_i has compact support and $\phi_i(x)=0$ whenever $x_i<\kappa_i\}$.

LEMMA 12:

$$\mathrm{K}_\varkappa(\hat{p})=\sup_{\Phi_\varkappa} \frac{\Big\{\sum_i \sum_x [\phi_i(x+e_i)-\phi_i(x)]\, \hat{p}(x)/x!\Big\}^2}{\sum_i \sum_x \phi_i^2(x)\, \hat{p}(x-e_i)/x!}. \tag{41}$$

We now relate $\mathrm{K}_\varkappa$ to J_m (in Theorem 14 below). Suppose that $\mathrm{F}(d\tau)$ is a probability measure supported on $\bar{\mathrm{T}} \subset \mathbf{R}_+^p$. The rescaling function

$\sigma = \sigma_m(\tau) = m\tau$ induces a measure $\sigma_m \mathrm{F}$ on $\bar{\mathrm{S}} = m\bar{\mathrm{T}}$. Define a sequence of probability measures on $\mathbf{R}_+^p \times \mathbf{Z}_+^p$ by

$$\mathbf{P}_m(d\tau, x) = \mathrm{F}(d\tau)\, p_{m\tau}(x).$$

Expectations computed under $\mathbf{P}_m$ are denoted by $\mathbf{E}_m$, and expectations involving the posterior distribution of τ given x will be written (with slight abuse of notation) as $\mathbf{E}_m[h(\tau)\,|\,x]$. Denote the prior on $\sigma = m\tau$ corresponding to $\mathrm{F}(d\tau)$ by $\mathrm{P}_m(d\sigma)$.

LEMMA 13:

$$\delta_{\mathrm{P},\,i}(x) - x_i = x_i\,\pi_{x-e_i}^{-1}(\pi_x - \pi_{x-e_i}), \qquad x_i \geq \kappa_i \tag{42}$$

$$= \mathbf{E}_m[\tau_i f^{-1} \mathrm{D}_i f(\tau)\,|\,x - e_i]. \tag{43}$$

We note also that expectations against marginal densities have the following forms when $\pi = \pi\mathrm{P}$, and secondly when $\pi = \pi(\sigma_m \mathrm{F})$:

$$\sum_x g(x)\,\pi(x) = \int \sum_x g(x)\,p_\sigma(x)\,\mathrm{P}(d\sigma) = \int \mathrm{E}_\sigma g(\mathrm{X})\,\mathrm{P}(d\sigma)$$

$$= \int \mathrm{E}_{m\tau} g(\mathrm{X})\,\mathrm{F}(d\tau) = \mathbf{E}_m g(\mathrm{X}). \tag{44}$$

To state the representations for $\mathrm{J}_m(\mathrm{F})$, let $\mathbf{X}_\kappa = \{\chi \in \mathrm{C}_0^1(\mathrm{R}_+^p, \mathrm{R}^p)$: for each critical face i, there exists $\varepsilon_i > 0$ such that $\chi_i(\tau) = 0$ if $\tau_i < \varepsilon_i\}$.

THEOREM 14. — *Let* F *be a probability measure on* $\bar{\mathrm{T}}$. *If* $\pi = \pi(\sigma_m \mathrm{F})$ *and* $\hat{p}_m(x)$ *correspond to* $\sigma_m \mathrm{F}$ *as described above, then*

(i) $$m^{-1}\mathrm{J}_m(\mathrm{F}) = \mathrm{K}_\kappa(\hat{p}_m) = \mathrm{K}(\pi) + \Delta_m(\hat{p}_m).$$

If $|\mathbf{I}(\mathrm{T})| = p$, $\Delta_m = 0$. *Otherwise, if* $\bar{\mathrm{T}}$ *is compact, there exist positive* c_1, ε_1 *depending only on* T, *such that* $\Delta_m \leq c_1 m e^{-\varepsilon_1 m}$.

(ii) $$\mathrm{J}_m(\mathrm{F}) = \sup_{\chi \in \mathbf{X}_\kappa} \frac{\{\mathbf{E}_m \sum_i m[\chi_i(m^{-1}(x+e_i)) - \chi_i(m^{-1}x)]\}^2}{\mathbf{E}_m \sum_i m\chi_i^2(m^{-1}(x+e_i))/(x_i+1) + m\,\delta_{m\kappa}(\chi)}. \tag{45}$$

where $\delta_{m\kappa}(\chi) \leq c_2 e^{-\varepsilon_1 m}$ *for some positive* $c_2 = c_2(\chi, \mathrm{T})$, $\varepsilon_1 = \varepsilon_1(\mathrm{T})$ *and vanishes if* $|\mathbf{I}| = p$.

(iii) *If* F *is absolutely continuous on* $(0, \infty)^p$ *with weakly differentiable density* f *and* $\mathrm{F}(\partial \mathrm{R}_+^p) = 0$, *then*

$$\mathrm{J}_m(\mathrm{F}) - m\,\Delta_m(\mathrm{F}) = m\,\mathrm{K}(\pi) = \mathbf{E}_m \sum_i \frac{m}{\mathrm{X}_i + 1}\,\mathbf{E}_m^2[\tau_i f^{-1}\mathrm{D}_i f(\tau)\,|\,\mathrm{X}]. \tag{46}$$

Part (i) connects J_m to K through the rescaling $\sigma = m\tau$. The terms Δ_m and $\delta_{m\kappa}$ will be shown to be asymptotically negligible as $m \to \infty$. Formula (45) provides the discrete counterpart for $\mathrm{J}_m(\mathrm{F})$ of formula (33) for J (F) for *arbitrary* priors F, whereas (46) is the analogue of formula (6) for $J(f)$ when F has a differentiable density f.

We now turn to continuity properties of the functionals $J_m(F)$ needed for (27). The first states that $(m, F) \to J_m(F)$ is lower semi-continuous. Since no restrictions are placed on the measures F, the proof necessarily involves the representations (45) and (33).

THEOREM 15. — (i) *If* $\{F_m\}$ *is a sequence of probability measures on* R_+^p *converging weakly to* F_0, *then* $J(F_0) \leqq \liminf J_m(F_m)$.

(ii) *In particular, if* $F_0(\partial R_+^p) > 0$, *then* $\liminf J_m(F_m) = \infty$.

The second property asserts that $J_m(F)$ actually converges to J (F) if F is assumed to have sufficient regularity properties. The proof, although lengthy, amounts to showing that the dominated convergence theorem may be applied to representation (46).

THEOREM 16. — *Suppose that* F *has a continuous density f with compact support. Let* $T = \{\tau : f(\tau) > 0\}$ *and* $\gamma(\tau) = d(\tau, T^c)$. *Assume that* (i) T *is a domain of class* $C^{2,\alpha}$ (GT, p. 94), (ii) f *is* C^1 *on* T *and* $\sup_{i, \bar{T}} |\tau_i D_i f(\tau)| < \infty$, *and* (iii) *there exist* $c, k > 0$ *such that for* $\gamma(\tau)$ *sufficiently small*, $f(\tau) \geqq c\gamma^k(\tau)$. *Then as* $m \to \infty$

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

Remark. — In the proof of Theorem 5, we take $f(\tau) = u_\Omega^2(\omega)$ where u_Ω solves (22). The regularity properties of u_Ω provided by Theorem 4 ensure that Theorem 16 applies to this choice of f. Indeed, since Ω is assumed to be of class $C^{2,\alpha}$, it follows from Theorem 4 (i) that $u_\Omega \in C^{2,\alpha}(\bar{\Omega})$. Thus $|\tau_i^{1/2} D_i f(\tau)| \leqq |Du_\Omega(\omega)| \leqq M$ on $\bar{\Omega}$. Since $\bar{T}$ is bounded, this establishes condition (ii). Let $\gamma_\Omega(\omega) = d(\omega, \Omega^c)$: for ω near $\partial\Omega$, $\gamma(\tau) \leqq c_1 \gamma_\Omega(\omega) \leqq c_2 \gamma^{1/2}(\tau)$. Thus, Giraud's Theorem [Theorem 4 (ii)] guarantees that $f(\tau) = u_\Omega^2(\omega) \geqq c_3 \gamma_\Omega^2(\omega) \geqq c_4 \gamma^2(\tau)$ for $\gamma(\tau)$ sufficiently small, which establishes condition (iii).

Finally, we articulate a simple approximation result that is basic to the proof of Theorem 16, and also to the approximation of Bayes estimates [*cf.* (14) in the introduction].

LEMMA 17. — *If* $g \in C_0(R_+^p)$, *then uniformly on compact subsets of* $(0, \infty)^p$,

$$\int g(\tau) m^p p_{m\tau}(mz) \, d\tau \to g(z). \tag{47}$$

COROLLARY 18. — *Let* $F(d\tau)$ *have weakly differentiable density* $f(\tau)$ *on* R_+^p. *Let* $\sigma = m\tau$ *and define* $P_m(d\sigma)$ *as the rescaled version of* F. *Then if*

$z>0$,

$$\delta_{P_m, i}(mz)=mz_i+z_i\, f^{-1}\, D_i f(z)+o(1) \quad as\ m\to\infty. \tag{48}$$

This is proved by combining the representation (43) with the relation

$$E_m[\tau_i f^{-1} D_i f(\tau)\,|\,X]=\frac{\displaystyle\int \tau_i D_i f(\tau)\, m^p p_{m\tau}(mz)\, d\tau}{\displaystyle\int f(\tau)\, m^p p_{m\tau}(mz)\, d\tau}. \tag{49}$$

6. PROOFS

Remark 1. – If $P\in \mathbf{P}^*(T\cap(0,\infty)^p)$, then (35) may be re-expressed (in the original problem) as

$$\delta_{P, i}(x)=\{E(\lambda_i^{-1}\,|\,x)\}^{-1},$$

but this is false if $\operatorname{supp} P$ intersects ∂R_+^p. For example, if $p=1$ and $P=1/2(\delta_{\{0\}}+\delta_{\{1\}})$, then from (35) $\delta_P(1)=(1+e)^{-1}$, whereas the above representation would give 1.

Proof of Lemma 2. – We prove that if T is $C^{2,\alpha}$ possibly with non-zero corners, then so also is $\Omega=\tau^{-1}(T)$. Some care is needed to allow for points $\tau_0\in\partial\bar{T}\cap\partial R_+^p$. Pick $\omega_0\in\partial\Omega$ and let $\tau_0=\tau(\omega_0)$. Renumbering co-ordinates if necessary, assume that $\tau_{0i}=0$ for $1\leqq i\leqq j_0$. Let $\sigma(\tau)$ be a $C^{2,\alpha}$ change of co-ordinates near τ_0 such that $\det(\partial\sigma/\partial\tau)\neq 0$, and near τ_0, T is described by $\sigma_i>0$, $1\leqq i\leqq j_1$, where $j_1>j_0$. Note that for $i\leqq j_0$, we may take $\sigma_i(\tau)=\tau_i$. Setting $\sigma'=(\sigma_i, i>j_0)$, $\tau'=(\tau_i, i>j_0)$, we conclude that $\partial\sigma'/\partial\tau'$ is non-singular.

For $i>j_0$, $\tau_{0i}>0$, and so we may define polar co-ordinates (r_i, θ_i) on $(\omega_{2i-1}, \omega_{2i})$ such that $\omega\to\theta_i$ is C^∞ near τ_0. Define

$$\begin{aligned} u_i(\omega)&=\omega_i, && 1\leqq i\leqq 2j_0,\\ u_{2j_0+2k}(\omega)&=\sigma_{j_0+k}(\tau(\omega)), && 1\leqq k\leqq p'=p-j_0,\\ u_{2j_0+2k-1}(\omega)&=\theta_{j_0+k}, && 1\leqq k\leqq p'=p-j_0,\end{aligned}$$

which is $C^{2,\alpha}$. Note that Ω is described near ω_0 by the constraints $u_{2i}>0$ for $j_0<i\leqq j_1$. Setting $u'=(u_l, l>2j_0)$ and $\theta'=(\theta_i, i>j_0)$, we find that invertibility of u near ω_0 follows from the invertibility of the $2p'\times 2p'$ Jacobian matrices $\partial u'/\partial(\theta', \tau')$ and $\partial(\theta', \tau')/\partial\omega'$. ∎

For the following lemma, associate (multiple) polar co-ordinates (r_i, θ_i) with components $(\omega_{2i-1}, \omega_{2i})$ of $\omega\in R^{2p}$ in the usual way.

LEMMA 19. – (i) *If* $\omega=(r_i, \theta_i)$ *and* $\bar{\omega}=(\bar{r}_i, \theta_i)$ *have the same angular co-ordinates, then*

$$|\omega-\bar{\omega}|^4\leqq|\tau(\omega)-\tau(\bar{\omega})|^2$$

(ii) $$\tau^{-1}(\mathrm{T}^{\eta}) \subset (\tau^{-1}(\mathrm{T}))^{\eta^{1/2}}.$$

Proof. — For part (i) it suffices to consider $p=1$. Clearly

$$|\omega-\bar{\omega}|^2=(r-\bar{r})^2=\tau(1-\sqrt{\bar{\tau}/\tau})^2 \leqq \tau\,|1-\bar{\tau}/\tau|=|\tau-\bar{\tau}|.$$

For part (ii), if $\bar{\tau}\in\mathrm{T}$ satisfies $|\bar{\tau}-\tau(\omega)|<\eta$, and $\bar{\omega}\leftrightarrow(\bar{\tau}_i^{1/2}, \theta_i)$, where θ_i are the angular co-ordinates of ω, then $\bar{\omega}\in\tau^{-1}(\mathrm{T})$ and $|\omega-\bar{\omega}|\leqq\eta^{1/2}$. ■

Proof of Theorem 6. — The proof simply involves substitution of the estimator (25) into the unbiased risk estimate (8). As seen in Section 1, the choice (25) ensures that the leading term yields the minimum eigenvalue of the Laplacian. The burden of the proof is to show that the error terms are uniformly small—and this depends on regularity properties of u_Ω. As noted by Bickel (1981) and Berkhin and Levit (1980), since u_Ω vanishes on $\partial\Omega$, the behaviour of $\mathrm{D}u_\Omega/u_\Omega$ and hence $\tau_i\,\mathrm{D}_i\,f/f$ is unstable near the boundary; and this accounts for the introduction of the extension sequence T^{η_m}, Ω^{ε_m} and $u_{\Omega^{\varepsilon_m}}$.

Performing the indicated substitution and using Taylor's Theorem yields

$$\begin{aligned}
m[\mathrm{R}(\delta_m, \lambda)-p]&\\
&=2m\,\mathrm{E}_{m\tau}\sum_i[h_{mi}(z+m^{-1}e_i)-h_{mi}(z)]\{z,\, z+m^{-1}e_i\in\mathrm{T}^{\eta_m}\}\\
&\quad+m\,\mathrm{E}_{m\tau}\sum_i(mz_i+1)^{-1}h_{mi}^2(z+m^{-1}e_i)\{z+m^{-1}e_i\in\mathrm{T}^{\eta_m}\}\\
&\quad+2m\,\mathrm{E}_{m\tau}\sum_i\begin{bmatrix}h_{mi}(z+m^{-1}e_i)\{z+m^{-1}e_i\in\mathrm{T}^{\eta_m},\, z\notin\mathrm{T}^{\eta_m}\}\\ -h_{mi}(z)\{z+m^{-1}e_i\notin\mathrm{T}^{\eta_m},\, z\in\mathrm{T}^{\eta_m}\}\end{bmatrix}\\
&=\mathrm{E}_{m\tau}\sum_i[2\,\mathrm{D}_i h_{mi}(z)+z_i^{-1}h_{mi}^2(z)]\{z\in\mathrm{T}^{\eta_m}\}+\mathrm{R}_{1m}+\mathrm{R}_{2m}+\mathrm{R}_{3m},
\end{aligned}$$

where $\{\mathrm{A}\}$ denotes the indicator function of the set A. Call the principal term $\mathrm{P}_m(\tau)$ say, and calculate it using (24) and $v_m(\tau)=f_m^{1/2}(\tau)$ exactly as in Section 1 to obtain

$$\mathrm{P}_m(\tau)=4\,\mathrm{E}_{m\tau}\,v_m^{-1}\,\mathrm{L}\,v_m(z)\{z\in\mathrm{T}^{\eta_m}\}=4\,\lambda(\Omega^{2\,\eta_m^{1/2}})\,\mathrm{P}_{m\tau}(\mathrm{T}^{\eta_m}),$$

since $\tau^{-1}(\mathrm{T}^{\eta_m})\subset\Omega^{\eta_m^{1/2}}$ (Lemma 19). Uniform convergence of $\mathrm{P}_m(\tau)$ to $4\lambda(\Omega)$ follows from Proposition 3 and the following uniform large deviations bound, whose proof is indicated below.

PROPOSITION 20. — *Let* $\bar{\mathrm{Y}}_m=m^{-1}\sum_1^m\mathrm{Y}_i$ *where* Y_i *are independent Poisson* (τ) *vectors, with* $\tau\in\mathrm{T}\subset[0, \mathrm{M}]\subset\mathrm{R}_+^p$ *and* $\eta_m\geqq m^{-\alpha}$ *for* $0<\alpha<1/2$. *Then there exists a constant* $\varepsilon_1(\mathrm{M})>0$ *such that*

$$\sup_{\mathrm{T}}\mathrm{P}_\tau(\bar{\mathrm{Y}}_m\notin\mathrm{T}^{\eta_m})\leqq 2pe^{-\varepsilon_1 m^{1-2\alpha}}.$$

It remains to show that the error terms $\mathrm{R}_{im}\to 0$ uniformly in T. First, we express h_{mi} and $\bar{h}_{mi}(z)=z_i^{-1}h_{mi}^2(z)$ and their derivatives in terms of u_m.

For this we need polar co-ordinates (r_i, θ_i) of $\omega \in \mathrm{R}^{2p}$, and (abusing notation) set $\theta_i^{\mathrm{T}} = (0\dots 0, \cos\theta_i, \sin\theta_i, 0\dots 0)$. Let $\mathrm{D}u$ and $\mathrm{D}^2 u$ denote gradient and Hessian matrix of u in the ω co-ordinate system. Calculation gives (dropping the index m)

$$\left.\begin{aligned} h_i &= r_i u^{-1} \theta_i^{\mathrm{T}} \mathrm{D}u, \\ 2\,\mathrm{D}_i h_i &= (r_i u)^{-1} \theta_i^{\mathrm{T}} \mathrm{D}u + u^{-1} \theta_i^{\mathrm{T}} \mathrm{D}^2 u\, \theta_i - (u^{-1} \theta_i^{\mathrm{T}} \mathrm{D}u)^2, \end{aligned}\right\} \tag{50}$$

$$\left.\begin{aligned} \bar{h}_i &= (u^{-1} \theta_i^{\mathrm{T}} \mathrm{D}u)^2, \\ \mathrm{D}_i \bar{h}_i &= (r_i u)^{-1} \theta_i^{\mathrm{T}} \mathrm{D}u\, [u^{-1} \theta_i^{\mathrm{T}} \mathrm{D}^2 u\, \theta_i - (u^{-1} \theta_i^{\mathrm{T}} \mathrm{D}u)^2]. \end{aligned}\right\} \tag{51}$$

To estimate $r_i^{-1} \theta_i^{\mathrm{T}} \mathrm{D}u$, we express it in terms of second derivatives of u by solving for $r^{-1}\, \partial/\partial r$ in the identity

$$\partial^2/\partial x_1^2 + \partial^2/\partial x_2^2 = \partial^2/\partial r^2 + r^{-1}\, \partial/\partial r + \partial^2/\partial \theta^2,$$

and noting that the $\partial^2/\partial\theta^2$ term drops out due to radial symmetry of u. Thus

$$r^{-1} \theta_i^{\mathrm{T}} \mathrm{D}u = \mathrm{D}_{2i-1}^2 u + \mathrm{D}_{2i}^2 u - \theta_i^{\mathrm{T}} \mathrm{D}^2 u\, \theta_i. \tag{52}$$

To estimate R_{1m}, we use the bound

$$|g(\tau+h) - g(\tau) - h^{\mathrm{T}} \mathrm{D}g(\tau)| \leqq |h|^{1+\alpha} [\mathrm{D}g]_{\alpha;\,\mathrm{B}},$$

valid for τ and $\tau + h$ belonging to a given set B. Thus

$$\begin{aligned} |\mathrm{R}_{1m}| \leqq 2\, \mathrm{E}_{m\tau} \sum_i m^{-\alpha} [\mathrm{D}h_{mi}]_{\alpha;\,\mathrm{T}^{\eta_m}} \{\mathrm{T}^{\eta_m}\} \\ + |\mathrm{D}_i h_{mi}|_{\mathrm{T}^{\eta_m}} \{z \in \mathrm{T}^{\eta_m},\ z + m^{-1} e_i \notin \mathrm{T}^{\eta_m}\}. \end{aligned} \tag{53}$$

To simplify notation, we sometimes use $\mathrm{T}(\eta)$ for T^{η} below. Now use the containment $\tau^{-1}(\mathrm{T}(\eta_m)) \subset \Omega(\eta_m^{1/2})$, together with (50), (52) and (20) to bound $[h_{mi}]_{\alpha,\,\mathrm{T}^{\eta_m}}$ in terms of

$$\|u_m\|_{\mathrm{C}^{2,\alpha}\,\overline{(\Omega(2\,\eta_m^{1/2}))}} \qquad \text{and} \qquad b_m = \inf\{u_m(\omega) : \omega \in \Omega(\eta_m^{1/2})\}.$$

Theorem 4 (iii) gives a bound $\|u_m\|_{\mathrm{C}^{2,\alpha}\,\overline{(\Omega^{\varepsilon_m})}} \leqq \mathrm{K}$ uniformly as $\varepsilon_m \to 0$. Since $d(\Omega(\eta_m^{1/2}),\ \partial\Omega(2\,\eta_m^{1/2})) \geqq \eta_m^{1/2}$, it follows from the uniform version of Giraud's theorem [4 (iii)] and the uniform bound on $\mathrm{D}^2 u$ that there exists $c > 0$ for which $b_m \geqq c\, \eta_m^{1/2}$ for large m. For the second term in (53), note that

$$\{z \in \mathrm{T}^{\eta_m},\ z + m^{-1} e_i \notin \mathrm{T}^{\eta_m}\} \subset \mathrm{T}(\eta_m')^c,$$

where we have set $\eta_m' = \eta_m - m^{-1}$. In summary, using c_i for constants not depending on m or τ, we obtain

$$\begin{aligned} |\mathrm{R}_{1m}| \leqq\ & p m^{-\alpha} c_1 (4\, b_m^{-2} \mathrm{K}^2 + b_m^{-4} \mathrm{K}^4)\, \mathrm{P}_{m\tau}(\mathrm{T}^{\eta_m}) \\ & + p\,(4\, b_m^{-1} \mathrm{K} + b_m^{-2} \mathrm{K}^2)\, \mathrm{P}_{m\tau}\{\mathrm{T}(\eta_m')^c\} \\ & \leqq c_2\, m^{-\alpha} \eta_m^{-2} + c_3\, \eta_m^{-1}\, \mathrm{P}_{m\tau}\{\mathrm{T}(\eta_m')^c\}. \end{aligned}$$

Arguing in similar fashion for R_{2m} and $R_{3\,m}$ we obtain

$$|R_{2m}| \leqq E_{m\tau} \sum_i m^{-1} |D_i \bar{h}_{mi}|_{T^{\eta_m}} \{T^{\eta_m}\} + 2 |\bar{h}_{mi}|_{T^{\eta_m}} \{T(\eta'_m)^c\}$$

$$\leqq c_4 m^{-1} \eta_m^{-3}/2 + c_5 \eta_m^{-1} P_{m\tau} \{T(\eta'_m)^c\}$$

$$|R_{3m}| \leqq 2 m E_{m\tau} \sum_i |h_{mi}|_{T^{\eta_m}} \{T(\eta'_m)^c\} \leqq c_6 m \eta_m^{-1/2} P_{m\tau} \{T(\eta'_m)^c\}.$$

It now follows from Proposition 20 that $R_{im} \to 0$ uniformly in $\tau \in T$ if we choose $\eta_m = m^{-\beta}$ for $0 < \beta < \alpha/2$. This completes the proof of Theorem 6. ■

Proof of Proposition 20. — Consider first the one-dimensional case: suppose $X_1, \ldots, X_m$ are i. i. d. distributed as Poisson (τ).

(i) There exists a positive constant ε such that uniformly in τ,

a) if $b \in [2\tau/3, \tau]$, $P_\tau(\bar{X}_m \leqq b) \leqq e^{-m\varepsilon(\tau-b)^2/b}$,

b) if $b \in [\tau, 2\tau]$, $P_\tau(\bar{X}_m \geqq b) \leqq e^{-m\varepsilon(\tau-b)^2/b}$.

The (standard) proof applies Markov's inequality, for example, to $P_\tau(e^{\alpha\bar{X}_m} \geqq e^{\alpha b})$, and optimizes over α to yield

$$\min\{P_\tau(\bar{X}_m \leqq b),\ P_\tau(\bar{X}_m \geqq b)\} \leqq \exp\{-mf(\tau, b)\}$$

where $f(\tau, b) = \tau - b - b \log[1 + (\tau - b)/b]$. Choose $\varepsilon > 0$ so that for $|x| < 1/2$, $\log(1+x) \leqq x - \varepsilon x^2$. Then for

$$|b^{-1}\tau - 1| \leqq 1/2, f(\tau, b) \geqq \varepsilon b^{-1} (\tau - b)^2.$$

(ii) Given $M > 0$, there exists a positive constant ε_1, such that uniformly in $\tau \leqq M$ and b such that $|\tau - b| \geqq m^{-\alpha}$,

$$\min\{P_\tau(\bar{X}_m \leqq b),\ P_\tau(\bar{X}_m \geqq b)\} \leqq e^{-\varepsilon_1 m^{1-2\alpha}}.$$

Suppose first $b \geqq \tau + m^{-\alpha}$, so that

$$P_\tau(\bar{X}_m \geqq b) \leqq P_\tau(\bar{X}_m \geqq \tau + m^{-\alpha}).$$

If $\tau \geqq m^{-\alpha}$, (i) *a)* applies directly, while if $\tau < m^{-\alpha}$, then

$$P_\tau(\bar{X}_m \geqq \tau + m^{-\alpha}) \leqq P_{m^{-\alpha}}(\bar{X}_m \geqq 2 m^{-\alpha}),$$

to which (i) *a)* applies. Secondly, if $\tau \geqq b + m^{-\alpha}$, then

$$P_\tau(\bar{X}_m \leqq b) \leqq P_\tau(\bar{X}_m \leqq \max\{b, 2\tau/3\}),$$

to which (i) *b)* applies.

Turning to the multivariate case, we note that

$$\{\bar{Y}_m \notin T^{\eta_m}\} \subset \{\max_j |\bar{Y}_{m,j} - \tau_j| > \eta_m\}.$$

Thus

$$P_\tau(Y_m \notin T^{\eta_m}) \leqq \sum_j P_{\tau_j}\{|\bar{Y}_{m,j} - \tau_j| > \eta_m\}$$

to which (ii) applies. ■

Proof of Lemma 10. – This is a minor variation on a standard identity [e. g. Berger (1985), Johnstone, (1986)]. Let $\delta_0(x)=x$ and note that since $\delta_P\in\mathbf{D}$, both $\delta_{0i}(x)$ and $\delta_{P,i}(x)=0$ if $x_i<\kappa_i$. Thus

$$
\begin{aligned}
p-r(\mathrm{P})&=r(\delta_0,\mathrm{P})-r(\delta_P,\mathrm{P})\\
&=\sum_i\int\lambda_i^{-1}\sum_{x_i\geqq\kappa_i}[\delta_{0i}^2(x)-\delta_{Pi}^2(x)-2\lambda_i(\delta_{0i}(x)-\delta_{Pi}(x))]p_\lambda(x)\,\mathrm{P}(d\lambda)\\
&=\sum_i\sum_{x_i\geqq\kappa_i}[\delta_{0i}(x)-\delta_{Pi}(x)]^2\hat{p}(x-e_i)/x!
\end{aligned}
$$

Substitution of (35) and (36) yields (37) and (39) respectively. ■

Proof of Lemma 11. – Choose $\mathrm{P}_0,\mathrm{P}_1\in\mathbf{P}^*(\bar{\mathrm{T}})$ and let $\mathrm{P}_t=(1-t)\,\mathrm{P}_0+t\mathrm{P}_1$, so that $\hat{p}_{tx}=(1-t)\,\hat{p}_{0x}+t_{\hat{p}_{1x}}$. We show that $k:t\to\mathrm{K}_\kappa(\hat{p}_t)$ is strictly convex for $t\in(0,1)$ unless $\hat{p}_{0x}=\hat{p}_{1x}$ for all $x\in\mathbf{Z}_+^p$, in which case $\mathrm{P}_0=\mathrm{P}_1$. Let $w_{itx}=[\hat{p}_{tx}-x_i\hat{p}_{t,x-e_i}]^2/\hat{p}_{t,x-e_i}x!$ and $\mathbf{X}_0=\{x:\max\{\hat{p}_{0x},\hat{p}_{1x}\}>0\}$. Then

$$
\mathrm{K}_\kappa(\hat{p}_t)=\sum_i\sum_x w_{itx}\{x:x_i\geqq\kappa_i,\ x-e_i\in\mathrm{X}_0\}.
$$

If u_t and v_t are linear functions of t such that $v_t>0$ for $0<t<1$, then $w_t=u_t^2/v_t$ is convex for $0<t<1$, indeed $\ddot{w}_t=2v_t^{-3}[\dot{u}_tv_t-\dot{v}_tu_t]^2$. In this application, $u_{itx}=\hat{p}_{tx}-x_i\hat{p}_{t,x-e_i}$ and $v_{itx}=\hat{p}_{t,x-e_i}x!$ Thus $k(t)$ is convex, and $\ddot{k}=0$ implies that $\dot{u}_{itx}v_{itx}=\dot{v}_{itx}u_{itx}$ for all x such that $x_i\geqq\kappa_i$ and $x-e_i\in\mathbf{X}_0$. Calculation shows that this implies

$$
\hat{p}_{1\,x+e_i}\hat{p}_{0\,x}=\hat{p}_{0x+e_i}\hat{p}_{1x},\qquad x\in\mathbf{Z}_+^p\cap\mathbf{X}_0.
$$

Since $\hat{p}_{0x}=\hat{p}_{1x}=1$ at $x=0$, we deduce by induction that $\hat{p}_{0x}$ and $\hat{p}_{1x}$ are either both positive or both zero, and hence equal at all $x\in\mathbf{Z}_+^p$. ■

Proof of Lemma 12. – For $\phi\in\Phi$, summation by parts establishes that

$$
\begin{aligned}
\mathrm{L}(\phi)&=\sum_i\sum_{x_i\geqq\kappa_i}(\hat{p}_x-x_i\hat{p}_{x-e_i})\,\phi_{i,x}/x!\\
&=\sum_i\sum_x\hat{p}_x\phi_{i,x}/x!-\sum_i\sum_{x_i\geqq 1}\hat{p}_{x-e_i}\phi_{i,x}/(x-e_i)!\\
&=\sum_i\sum_x\frac{\hat{p}_x}{x!}(\phi_{i,x}-\phi_{i,x+e_i}).
\end{aligned}
$$

The Cauchy-Schwartz inequality yields

$$
\mathrm{L}^2(\phi)\leqq\sum_i\sum_{x_i\geqq\kappa_i}\frac{(\hat{p}_x-x_i\hat{p}_{x-e_i})^2}{\hat{p}_{x-e_i}x!}\sum_i\sum_x\phi_{i,x}^2\frac{\hat{p}_{x-e_i}}{x!}=\mathrm{K}_\kappa(\hat{p})\,\mathrm{R}(\phi)\quad(54)
$$

say, which establishes inequality in (41) (again since $\phi_{i,x}=0$ if $x_i<\kappa_i$). The condition for equality in (54) suggests the definition $\phi_{i,x}^{(n)}=\hat{p}_{x-e_i}^{-1}(\hat{p}_x-x_i\hat{p}_{x-e_i})$ for x such that $x_i\geqq\kappa_i$ and $|x|=\sum x_i\leqq n$; and zero otherwise. Indeed

$$\mathrm{L}(\phi^n)=\mathrm{R}(\phi^n)=\sum_i \sum_{\{|x|\leq n,\, x_i\geq \varkappa_i\}} (\hat{p}_x - x_i \hat{p}_{x-e_i})^2/\hat{p}_{x-e_i} x!$$

and hence $\mathrm{L}^2(\phi^n)/\mathrm{R}(\phi^n)\uparrow \mathrm{K}_\varkappa(\pi)$ as $n\to\infty$. ■

Proof of Lemma 13. – Equation (42) is just (36). The identity $\mathrm{D}_{\tau_i} p_{m\tau}(x) = -m[p_{m\tau}(x) - p_{m\tau}(x-e_i)]$ permits integration by parts:

$$\pi_x - \pi_{x-e_i} = m^{-1}\int p_{m\tau}(x)\,\mathrm{D}_i f(\tau)\,d\tau = x_i^{-1}\int p_{m\tau}(x-e_i)\,\mathrm{D}_i f(\tau)\,d\tau.$$

Division by π_{x-e_i} yields (43). ■

Proof of Theorem 14. – For part (i), it remains to establish the bound on $\Delta_m(\hat{p})$. Suppose that $\bar{\mathrm{T}}\subset[0, \mathrm{M}]^p$, so that $d_{\mathrm{P},i}(x)\leq m\mathrm{M}$. Relabelling indices if necessary, consider the case when $i=1$ is not a critical index, and set $x'=(x_2, \ldots, x_p)$, $\tau'=(\tau_2, \ldots, \tau_p)$. There exists $\varepsilon_1>0$ such that $\tau\in\bar{\mathrm{T}}$ implies $\tau_i\geq\varepsilon_1$. Hence $p_{m\tau}(x)\leq e^{-m\varepsilon_1}p_{m\tau'}(x')$, and

$$|\Delta_{m1}|\leq m\mathrm{M}\,e^{-m\varepsilon_1}\int\sum_{x'} p_{m\tau'}(x')\,\mathrm{F}(d\tau)\leq m\mathrm{M}\,e^{-m\varepsilon_1}.$$

Arguing similarly for the other indices yields the conclusion (i).

For part (ii), if $\chi\in\mathbf{X}_\varkappa$, we may define $\phi\in\Phi_\varkappa$ by $\phi(x)=\chi(x/m)$. Conversely, given $\phi\in\Phi_\varkappa$, there exists $\chi\in\mathbf{X}_\varkappa$ satisfying this relation. This equivalence and (44) show that the numerators of formulas (45) and (41) agree when $\pi=\pi(\sigma_m\mathrm{F})$. Write the denominator of (41) as

$$\sum_i \sum_{x_i\geq 1} \phi_i^2(x)\frac{\hat{p}(x-e_i)}{x!} + \delta_{m\varkappa}(\phi),$$

where $\delta_{m\varkappa}(\phi)=\sum_{i\notin\mathbf{I}}\sum_{x_i=0}\phi_i^2(x)\hat{p}(x-e_i)/x!$ This leads to the form claimed in (45). To bound $\delta_{m\varkappa}$, consider only the case that $i=1\notin\mathbf{I}$, set $x'=(x_2, \ldots, x_p)$, $\tau'=(\tau_2, \ldots, \tau_p)$ and suppose that $\tau\in\bar{\mathrm{T}}$ implies $\tau_{01}\geq\varepsilon_1>0$. It follows that

$$\delta_{m\varkappa 1}(\phi)\leq\|\phi_1\|_\infty^2\int(m\tau_1)^{-1}e^{-m\tau_1}\sum_{x'}p_{m\tau'}(x')\,\mathrm{F}(d\tau)\leq c_2\,e^{-\varepsilon_1 m}.$$

Finally, for part (iii) write $\mathrm{K}(\pi)$ as $\sum_i\sum_x \pi_x^{-1}(\pi_{x+e_i}-\pi_x)^2(x_i+1)$, and substitute (43). ■

Proof of Theorem 15. – If $\chi\not\equiv 0$ a.e. F, we write

$$\mathrm{J}_m(\mathrm{F}, \chi)=\frac{n_m^2(\mathrm{F}, \chi)}{w_m(\mathrm{F}, \chi)}, \qquad \mathrm{J}(\mathrm{F}, \chi)=\frac{n^2(\mathrm{F}, \chi)}{w(\mathrm{F}, \chi)}, \tag{55}$$

where

$$n_m(\mathrm{F}, \chi)=\int \mathrm{E}_{m\tau} \sum_i m\,[\chi_i(m^{-1}(x+e_i))-\chi_i(m^{-1}x)]\,\mathrm{F}(d\tau),$$

$$n(\mathrm{F}, \chi)=\int \sum_i \mathrm{D}_i \chi_i(\tau)\,\mathrm{F}(d\tau),$$

$$w_m(\mathrm{F}, \chi)=\int \mathrm{E}_{m\tau} \sum_i (\mathrm{X}_i+1)^{-1} m\,\chi_i^2(m^{-1}(x+e_i))\,\mathrm{F}(d\tau)+\delta_{m\varkappa}(\chi),$$

$$w(\mathrm{F}, \chi)=\int \sum_i \chi_i^2(\tau)\,\tau_i^{-1}\,\mathrm{F}(d\tau).$$

Clearly $\mathrm{J}_m(\mathrm{F})=\sup_{\chi \in \mathbf{X}_\varkappa} \mathrm{J}_m(\mathrm{F}, \chi)$ and $\mathrm{J}(\mathrm{F})=\sup_{\chi \in \mathbf{X}} \mathrm{J}(\mathrm{F}, \chi)$.

We first prove part (i) under the assumption that $\mathrm{F}_0(\partial \mathrm{R}_+^p)=0$. The point of the test function representation is that it suffices to show that $\mathrm{J}_m(\mathrm{F}_m, \chi) \to \mathrm{J}(\mathrm{F}, \chi)$ for all $\chi \not\equiv 0$ a.e. F such that $\chi \in \mathbf{X}$ (we may ignore $\chi \in \mathbf{X}_\varkappa \cap \mathbf{X}^c$!). For such χ, $w(\mathrm{F}, \chi)>0$, and so we need only check that $n_m(\mathrm{F}_m, \chi)$ and $w_m(\mathrm{F}_m, \chi)$ converge to $n(\mathrm{F}, \chi)$ and $w(\mathrm{F}, \chi)$ respectively. We consider only n_m as the method for w_m is similar and the bound in Theorem 14 (iii) shows already that we may ignore $\delta_{m\varkappa}(\chi)$. It is enough to consider $i=1$, and to show that

$$\Delta_{1,m}=\int \mathrm{E}_{m\tau}\,\mathrm{L}_m(x, \tau)\,\mathrm{F}_m(d\tau) \to 0,$$

$$\mathrm{L}_m(x, \tau)=m\,[\chi_1(m^{-1}(x+e_1))-\chi_1(m^{-1}x)]-\mathrm{D}_1\chi_1(\tau).$$

We merely sketch the details involved in verifying that $\Delta_{1,m} \to 0$. We have

$$\mathrm{L}_m(x, \tau)=\sum_i (m^{-1}x_i-\tau_i)\,\mathrm{D}_{1i}^2\chi_1(\tau^*)+(2m)^{-1}\mathrm{D}_1^2\chi_1(\tau^+)$$

for appropriate τ^*, τ^+ depending on x, m. Suppose $\operatorname{supp}\chi_1 \subset [0, c]^p$. Decomposing $\mathrm{R}_+^p \times \mathrm{R}_+^p$ into $\mathrm{A}=\{\tau \in [0, 2c]^p\}$, $\mathrm{B}=\{\tau \notin [0, 2c]^p,\ x \notin [0, c]^p\}$ and $\mathrm{C}=\{\tau \notin [0, 2c]^p,\ x \in [0, c]^p\}$, exploiting the smoothness and compact support of χ_1 and a large deviations argument complete the proof.

We now assume that $\mathrm{F}_0(\partial \mathrm{R}_+^p)>0$ and proceed to part (ii). Note first that if $\chi_i(\tau) \leqq \sqrt{\tau_i}$, we have from (55) [noting that $\delta_{m\varkappa}(\chi)=0$]

$$p\,\mathrm{J}_m(\mathrm{F}, \chi) \geqq \left\{\int \mathrm{F}(d\tau)\,\mathrm{E}_{m\tau} \sum_i m\,[\chi_i(m^{-1}(x+e_i))-\chi_i(m^{-1}x)]\right\}^2.$$

Write $|\tau|_\infty$ for $\max_i \tau_i$. Fix $c>0$: Let χ_m be a vector field in $\mathbf{X}$ constructed so that $\chi_{i,m}=0$ unless $\max_{j \neq i} \tau_j \leqq c$, in which case $\tau_i \to \chi_{i,m}(\tau)$ is unimodal, with $\chi_{i,m}(\tau)=\sqrt{\tau_i}$ for $\tau_i\in[m^{-1}, c]$ and $=0$ for $\tau\leqq m^{-1}/2$ or $\tau\geqq 2c$. Consideration of $\chi_{m,i}$ near $\tau_i=0$ and for large $|\tau|_\infty$ shows that

$$\begin{aligned}\{p\,\mathrm{J}_m(\mathrm{F}_m, \chi_m)\}^{1/2} &\geqq \int \mathrm{F}_m(d\tau)\,\mathrm{E}_{m\tau}\sum_i m\{((x_i+1)/m)^{1/2}-(x_i/m)^{1/2}\}+o(1)\\ &\geqq 1/2\sum_i\int \mathrm{F}_m(d\tau)\,\mathrm{E}_{m\tau}\,m^{1/2}(x_i+1)^{-1/2}\\ &\geqq 1/2\sum_i\int \mathrm{F}_m(d\tau)(\tau_i+m^{-1})^{-1}+o_m(1),\end{aligned}$$

by Jensen's inequality. Let $\mathrm{N}_{\delta,i}=\{\tau:\tau_i\leqq\delta\}$ and $\mathrm{N}_\delta=\bigcup_i \mathrm{N}_{\delta,i}$. Weak convergence of F_m to F_0 implies the existence of $\delta_m>0$, lim inf $\delta_m=0$ for which $\mathrm{F}_m(\mathrm{N}(\delta_m))\geqq\eta=\mathrm{F}_0(\partial\mathrm{R}^p_+)/2$. Consequently

$$\begin{aligned}\{p\,\mathrm{J}_m(\mathrm{F}_m, \chi_m)\}^{1/2} &\geqq 1/2\sum_i(\delta_m+m^{-1})\,\mathrm{F}(\mathrm{N}_\delta, i)\\ &\geqq 1/2\,\eta\,(\delta_m+m^{-1})^{-1}/2\to\infty \quad \text{as } m\to\infty. \ \blacksquare\end{aligned}$$

Proof of Theorem 16. – We use (46) to write

$$\mathrm{J}_m(\mathrm{F})=\int\sum_i \mathrm{A}_{mi}(\tau)f(\tau)\,d\tau+\Delta_m(\mathrm{F}),$$

where

$$\mathrm{A}_{mi}(\tau)=\mathrm{E}_{m\tau}\,m(\mathrm{X}_i+1)^{-1}\,\mathrm{E}^2_m[\tau_i f^{-1}\mathrm{D}_i f(\tau)\,|\,\mathrm{X}]$$

and the bound of Theorem 14 (i) shows that we may ignore $\Delta_m(\mathrm{F})$. By Cauchy-Schwartz

$$\mathrm{A}_{mi}(\tau)\leqq \mathrm{B}_{mi}(\tau)=\mathrm{E}_{m\tau}\,m(\mathrm{X}_i+1)^{-1}\,\mathrm{E}_m[\tau_i^2(\mathrm{D}_i f/f)^2(\tau)\,|\,\mathrm{X}],$$

and

$$\begin{aligned}\int\sum_i \mathrm{B}_{mi}(\tau)f(\tau)\,d\tau &=\mathrm{E}_m\sum_i m(\mathrm{X}_i+1)^{-1}\tau_i^2(\mathrm{D}_i f/f)^2(\tau)\\ &=\int\sum_i \mathrm{E}_{m\tau}\,m\,\tau_i(\mathrm{X}_i+1)^{-1}\tau_i f^{-1}(\mathrm{D}_i f)^2(\tau)\,d\tau\nearrow \mathrm{J}(f).\end{aligned}$$

Young's extension of the dominated convergence theorem (e.g. Loeve, 1977, pp. 164-165) reduces the task to showing that $\mathrm{A}_{mi}(\tau)\to a_i(\tau)=\tau_i(\mathrm{D}_i f/f)^2(\tau)$ for (almost all) τ in $\mathrm{T}=\{f>0\}$.

Introduce

$$a_{mi}(m^{-1}x)=m(x_i+1)^{-1}\,\mathrm{E}^2_m[\tau_i(\mathrm{D}_i f/f)(\tau)\,|\,\mathrm{X}],$$

so that $\mathrm{A}_{mi}(\tau)=\mathrm{E}_{m\tau}\,a_{mi}(m^{-1}\mathrm{X})$. Let $\gamma(\tau)=d(\tau, \mathrm{T}^c)$ where

$$d(\tau, \mathrm{S})=\inf\{|\tau-s|, s\in\mathrm{S}\},$$

and put $Z_m = X/m$. The following two properties now suffice to prove the theorem:

(1) $a_m(z) \to a(z)$ uniformly on compact subsets of $T \cap (0, \infty)^p$.

(2) For some $\eta > 0$, if $\gamma(\tau_0) \geqq \eta$ then $E_{\tau_0}[a_m(Z_m), \gamma(Z_m) < \eta/2] \to 0$.

Property (1) follows from Lemma 47 and the relation (49). We present only an outline for the proof of step (2). This follows from standard large deviation results if $a_m(z)$ grows at worst polynomially in m and z. Now $a_{mi}(z) = m(mz_i + 1)^{-1} b_{mi}^2(z)$, where $b_{mi}(z) = E_m[\tau_i(D_i f/f)(\tau) \mid mz]$.

Let $T_\beta = \{\tau \in T : \gamma(\tau) \leqq \beta\}$. From the assumed regularity of f, it follows that for β small

$$b_{mi}(z) = \frac{\int_T p_{m\tau}(mz)\,\tau_i D_i f(\tau)\,d\tau}{\int_T p_{m\tau}(mz) f(\tau)\,d\tau} \leqq c_5 \beta^{-k} \left\{ 1 - \frac{\int_{T_\beta} p_{m\tau}(mz)\,d\tau}{\int_T p_{m\tau}(mz)\,d\tau} \right\}^{-1}$$

Call the integral ratio on the right side $R_m(\beta)$. Let

$$\xi(z, \tau) = \sum_i \tau_i - z_i - z_i \log(\tau_i/z_i)$$

and $f_{T, z}(\sigma)$ the density of the measure $\mu_{T, z}(\sigma) = \mathrm{Leb}\{\tau \in T : \xi(z, \tau) \leqq \sigma\}$. Thus

$$R_m(\beta) = \int e^{-m\sigma} f_{T_\beta, z}(\sigma)\,d\sigma \Big/ \int e^{-m\sigma} f_{T, z}(\sigma)\,d\sigma.$$

Now $f_{T, z}$ (and $f_{T_\beta, z}$) have at worst polynomial growth in σ and z near the (common) lower limit of their support. Choosing $\beta = \beta_m(z) = [m(|z| + 1)]^{-\beta}$ for suitable $\beta > 1$, shows that $R_m(\beta_m(z)) \to 0$ uniformly in $z \in T_\eta/2$. This implies a polynomial growth bound on $b_{mi}(z)$ and hence $a_{mi}(z)$. ■

Proof of Lemma 17. – We give the proof for $p = 1$; the multivariate case proceeds similarly. The integral may be written as $E^z g(\Gamma_m)$ where $m\Gamma_m \sim \mathit{Gamma}(mz + 1, 1)$ and the superscript z makes explicit the dependence on the parameters of the Gamma distribution. Let $A_m = \{\phi : |\phi - z - m^{-1}| < \zeta_m (m^{-1} z)^{1/2}\}$, and ω_g be the modulus of continuity of g. Suppose $z \in [c_1, c_2] \subset (0, \infty)$. Then

$$R_m^z = |E^z g(\Gamma_m) - g(z)| \leqq \omega_g(c_3 \zeta_m m^{-1/2}) P^z(A_m) + 2\|g\|_\infty P^z(A_m^c).$$

By Chebychev's inequality, $P^z(A_m) \leqq \zeta_m^{-2} z^{-1} m^{-1}(mz + 1) \leqq c_4 \zeta_m^{-2}$ uniformly in $z \in [c_1, c_2]$. Thus $R_m^z \to 0$ uniformly in such z if $\zeta_m \to \infty$ and $\zeta_m = o(m^{1/2})$. ■

ACKNOWLEDGEMENTS

Research support at various times from N.S.F. (MCS 84-51750, 86-00235), N.I.H. (PHS GM 21215-12), S.E.R.C., the Sloan Foundation, N.S.E.R.C., F.C.A.R. and G.E.R.A.D. is gratefully acknowledged. For helpful conversations and references, the authors are delighted to thank Larry Brown, David Donoho, Edward Fraenkel, Soren Johansen, Bernard Silverman and John Toland. Much of the work on the paper was done during visits by both authors to the University of Bath.

REFERENCES

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

P. E. BERKIN and B. Ya. LEVIT, Second-Order Asymptotically Minimax Estimates for the Mean of a Normal Population, *Problemy Peredachi Informatsii*, Vol. **16**, 3, 1980, pp. 60-79. Translation: *Problems of Information Transmission*, 1981, pp. 212-229.

P. J. BICKEL, Minimax Estimation of the Mean of a Normal Distribution when the Parameter Space is Restricted, *Ann. Statist.*, Vol. **9**, 1981, pp. 1301-1309.

L. D. BROWN, Admissible Estimators, Recurrent Diffusions and Insoluble Boundary Value Problems, *Ann. Math. Statist.*, Vol. **42**, 1971, pp. 855-903. Correction, *Ann. Statist.*, Vol. **1**, pp. 594-596.

L. D. BROWN, An Heuristic Method for Determining Admissibility of Estimators – with Applications, *Ann. Statist.*, Vol. **7**, 1979, pp. 960-994.

M. CLEVENSON and J. ZIDEK, Simultaneous Estimation of the Mean of Independent Poisson Laws, *J. Amer. Statist. Assoc.*, Vol. **70**, 1975, pp. 698-705.

R. COURANT and D. HILBERT, *Methods of Mathematical Physics*, Vol. **1**, Interscience Publishers, New York, 1953.

E. N. DANCER, The Effect of Domain Shape on the Number of Positive Solutions of Certain Nonlinear Equations, *J. Diff. Eq.*, Vol. **74**, 1988, pp. 120-156.

D. GILBARG and N. S. TRUDINGER, *Elliptic Partial Differential Equations of Second Order*, 2nd ed., Springer-Verlag, Berlin-Heidelberg, New York, 1983.

P. J. HUBER, Robust Estimation of a Location Parameter, *Ann. Math. Statist.*, 1964, Vol. **35**, pp. 73-101.

P. J. HUBER, *Robust Statistics*, Wiley, New York, 1981.

I. M. JOHNSTONE, Admissible Estimation, Dirichlet Principles and Recurrence of Birth-Death Chains on $\mathbf{Z}_+^p$, *Prob. Theory and Related Fields* (= *Z. Wahrsch. Th.*), Vol. **71**, 1986, pp. 213-269.

I. M. JOHNSTONE and K. B. MACGIBBON, Une mesure d'information caractérisant la loi de Poisson, *Séminaire de Probabilités*, Vol. **XXI**, 1987, pp. 563-573, *Lecture Notes in Mathematics*, Vol. **1247**, J. AZÉMA, P. A. MEYER and M. YOR Eds., Springer Verlag, New York.

I. M. JOHNSTONE and K. B. MACGIBBON, Minimax Estimation of a Constrained Poisson Vector, *Ann. Statist.*, Vol. **20**, 1992, pp. 807-831.

O. A. LADYZHENSKAYA and N. N. URAL'TSEVA, *Linear and Quasilinear Elliptic Equations*, English translation, New York, Academic Press, 1968 (2nd. Russian edition, 1973).

B. Ya. LEVIT, On Asymptotic Minimax Estimates of the Second Order, Translation: *Theor. Prob. Appl.*, Vol. **25**, 1980, pp. 552-568.

B. Ya. LEVIT, Minimax Estimation and Positive Solutions of Elliptic Equations, Translation: *Theor. Prob. Appl.*, Vol. **27**, 1982, pp. 563-586.

B. Ya. LEVIT, Second Order Asymptotic Optimality and Positive Solutions of Schrödinger's Equation, Translation: *Theor. Prob. Appl.*, Vol. **30**, 1985 *a*, pp. 333-363.

B. Ya. LEVIT, Evaluation of the Minimax Risk, *Fourth International Vilnius Conf. Prob. Theory Math. Stat.*, Vol. **4**, Vilnius, 1985 *b*, pp. 181-183.

B. Ya. LEVIT, On Second Order Admissibility in Simultaneous Estimation, *Proc. 1st World Congress Bernouilli Soc. Tashkent, U.S.S.R.*, 1986, Yu. A. PROHOROV and V. V. SAZONOV Eds., V.N.U. Sc. Press, Utrecht.

M. LOÈVE, *Probability Theory I*, Springer Verlag, New York, 1977.

A. A. MELKMAN and Y. RITOV, Minimax Estimation of the Mean of a General Distribution when the Parameter Space is Restricted, *Ann. Statist.*, Vol. **15**, 1987, pp. 432-442.

C. MIRANDA, *Partial Differential Equations of Elliptic Type*, 2nd. ed., Springer Verlag, New York, Heidelberg, 1970.

J. M. STEELE, Fisher Information and Detection of a Euclidean Perturbation of an Independent Stationary Process, *Ann. Prob.*, Vol. **14**, 1986, pp. 326-335.

(*Manuscript received January* 21, 1992.)