IMS Collections
Borrowing Strength: Theory Powering Applications –
A Festschrift for Lawrence D. Brown
Vol. 6 (2010) 32–42

DOI: 10.1214/10-IMSCOLL603

Minimax estimation over hyperrectangles with implications in the Poisson case

Brenda MacGibbon[1]

Université du Québec à Montréal

Abstract: The purpose of this research is to extend the results of Johnstone and MacGibbon [18, 19] to study the asymptotic behavior of the ratio of linear minimax risk to nonlinear minimax risk for the estimation of a Poisson mean that lies in a rescaled compact l^1-ellipsoid using the information-normalized quadratic loss function. This would be an analogue of Pinsker's [24] result for l^2-ellipsoids with quadratic loss in the Gaussian case. In view of the work of Brown et al. [7] which demonstrated the asymptotic equivalence of the problems of estimation under the same loss function of the intensity of a non-homogeneous Poisson process and estimation in the Gaussian white noise model with drift, some results concerning Poisson mean estimation with respect to quadratic loss are also included.

1. Introduction

This work is inspired by the results of Brown et al. in [7] where they establish the asymptotic equivalence of the problems of estimation under the same loss function of the intensity of a non-homogeneous Poisson process and estimation in the Gaussian white noise model with drift. In the light of this equivalence, Brown's [5] contention that minimaxity has played an essential role in many statistical areas including nonparametric function estimation has also been influential. For these reasons, it seems relevant to take another look at the exact results on minimax estimation of an infinite dimensional Poisson mean known to lie in a certain type of compact convex domain under information-normalized loss in [18] and the asymptotic results relating this Poisson minimax estimation problem in finite dimensions to the normal mean vector one with squared error loss in [19]. In [18] and [19], particular attention is paid to the relationship between linear minimax risk and nonlinear minimax risk. It is natural to study this relationship because in many statistical applications linear estimates are often used.

In particular, under information-normalized quadratic loss, Johnstone and MacGibbon in [18] considered minimax estimation of the mean of a Poisson experiment when that mean is known to be in a compact convex domain, which is rectangularly convex at the origin. Sets such as compact ellipsoids or l_1-bodies (with $p \geq 1$) and hyperrectangles satisfy this definition. Their method of proof uses the connection between the estimation problem and solutions to elliptic partial differ-

[1]GERAD and Départment de mathématiques, C.P.8888, Succursale centre-ville, Université du Québec à Montréal, Montréal (Québec) Canada, H3C 3P8, e-mail: macgibbon.brenda@uqam.ca

Keywords and phrases: Estimating a bounded Poisson mean, Pinsker's result, polydisc transform, hardest rectangular subproblems, Ibragimov-Hasminskii constant.

AMS 2000 subject classifications: Primary 62F12; secondary 62F10.

ential equations (see also [2, 21, 22]). An exact bound on the ratio of linear minimax risk to the minimax risk was obtained.

The methods of proof are analogous to those used for the Gaussian bounded mean estimation problem by Donoho et al. in [10], who considered the problem of estimating the (infinite dimensional) mean of a standard Gaussian shift when the mean is known to lie in a compact, orthosymmetric, convex, quadratically convex set in ℓ_2. They used the heuristic that the difficulty of the hardest rectangular subproblem is equal to the difficulty of the full problem and consequently reduced the study to one dimension. They found a bound of 1.25 on the ratio of the linear minimax risk to the minimax risk. Such a bound was called the Ibragimov-Hasminskii [17] constant since its behavior was first studied in [17]. The numerical specification of this bound was achieved using the optimization system NPSOL [15]. They also showed that if the set is not quadratically convex, as in the case of ℓ_1-bodies, then minimax linear estimators may be outperformed arbitrarily by nonlinear estimates.

Because the results in [10] depend on the one dimensional problem, we mention that Casella and Strawderman [8] were the first to study the one dimensional problem of estimating a bounded normal mean with known variance and to provide analytic and numerical results for the minimax risk for this problem. Zinzius [27], Feldman and Brown [11] and Vidakovic and das Gupta [25] also studied these questions. In order to find the minimax solution, the dual problem of finding the least favorable prior distribution for the corresponding Bayes problem is often considered. The duality theory ensures that, if there exist solutions to both problems, then the Bayes procedure with respect to the least favorable prior distribution will be minimax (cf. [1, 4, 12, 20, 26]). For many of these problems, (see, e.g., [13]) the least favorable prior distribution is a discrete measure with finite support. Thus, the finding of a minimax solution becomes a global optimization of a nonlinear, non-convex function and numerical techniques are necessary.

The most important aspect of the work in [10] is the application of the results to minimax nonparametric function estimation under quadratic loss. As argued in [18] and [19], the Poisson distribution models prototypical discrete data settings and deserves independent study. It is to be hoped that the results in this paper would lead to applications in minimax estimation with respect to normalized loss of an intensity function of an inhomogeneous Poisson process or a signal which may be contaminated by Poisson "white noise" or a Poisson regression function which satisfies certain smoothness conditions.

This, however, is not feasible at the moment without a deeper understanding of how to match this problem of estimating the intensity of a Poisson process under information-normalized loss with the appropriate Gaussian analogue. Our goal for the moment must be more limited in scope and we have chosen to find the Poisson analogue of one of the more elegant results in minimax estimation of constrained normal means. This is Pinsker's [24] result, which shows that, for a Gaussian mean vector (with standard covariance matrix) that lies in an ellipsoid of the form: $\{\sum_{i=1}^{n} a_i\omega_i^2 \leq \frac{c^2}{\epsilon^2(n)}\}$, the ratio of the minimax linear risk to the minimax risk approaches 1 as $n \uparrow \infty$ and $n\epsilon^2(n)$ is constant. This problem is studied in Section 1.

In Section 2 some results about Poisson mean estimation with quadratic loss (cf. [23]) are explored in view of the asymptotic equivalence given by [7].

2. The Poisson analogue of Pinsker's result using information-normalized loss

In order to examine Pinsker's result for Poisson estimation using information-normalized loss, we will first recall some of the results of Johnstone and MacGibbon ([18],[19]). They studied the problem of minimax estimation of a bounded Poisson vector in $\mathbb{R}_+^n$ using the normalized quadratic loss function given by $\sum_i (\delta_i - \tau_i)^2 \tau_i^{-1}$.

The problem studied in [18] in one dimension can be described as follows. Let X denote a Poisson random variable with mean τ. The information-normalized quadratic loss function can be written as, $L(\delta, \tau) = \tau^{-1}(\delta - \tau)^2$, with risk, $R(\delta, \tau) = \sum_{x=0}^{\infty} \tau^{-1}(\delta(x) - \tau)^2 \frac{\tau^x}{x!} e^{-\tau}$.

Under the additional assumption that τ lies in an interval of the form $[0, m], m > 0$, an estimator δ_m is minimax for the above problem if, for all $\delta \in \mathcal{D}$, the space of decision procedures,

$$\sup_{0 \le \tau \le m} R(\delta_m, \tau) = \inf_{\delta \in \mathcal{D}} \sup_{0 \le \tau \le m} R(\delta, \tau). \tag{2.1}$$

In this Poisson model, in order to determine the minimax risk $\rho_P(m)$, the corresponding Bayes problem was considered. A distribution or prior probability measure π is specified on the parameter space $[0, m]$ and a measure of the performance of a procedure δ is given by its Bayes risk:

$$r(\delta, \pi) = \int_0^m R(\delta, \theta)\pi(d\theta). \tag{2.2}$$

The estimator δ_π is called the Bayes procedure with respect to the prior probability measure π if δ_π minimizes the Bayes risk.

The Bayes risk $r(\pi)$ of a prior probability measure π on $[0, m]$ is defined as $r(\pi) = r(\delta_\pi, \pi)$. A prior probability measure π^* is "least favorable" if its Bayes risk is greater than or equal to that of any other prior. Subject to the decision problem satisfying sufficient regularity conditions, a least favorable prior distribution exists (and has finite support in the bounded case considered here) and the corresponding Bayes procedure is minimax (cf. Wald [26], Brown [4], Kempthorne [20]). If the least favorable prior π^* has finite support on a set of k points denoted by $\{b_i m\}$ with corresponding non-negative weights $\{a_i\}$ (such that the sum of the weights is one), then the Bayes rule δ_{π^*} under information-normalized quadratic loss with respect to this prior is given by

$$\delta_{\pi^*}(x) = \frac{\sum_{i=1}^k a_i (b_i m)^x e^{-b_i m}}{\sum_{i=1}^k a_i (b_i m)^{x-1} e^{-b_i m}}. \tag{2.3}$$

The minimax estimator δ_m is equal to this Bayes rule with the convention that $\delta_m(0) = 0$. For this Poisson model, the minimax risk $\rho_P(m)$ is defined as the Bayes risk of this least favorable prior.

For $n = 1$ and small m, they found the exact minimax estimator. For larger m, they used a modification of a convergent iterative procedure for the numerical specification of least favorable priors described by Kempthorne in [20].

As many statistical procedures are linear, it was also of interest in [18] to consider linear estimation. For this Poisson problem, the linear minimax estimator is defined to be the linear estimator δ_m^L satisfying (2.1) with the infimum over δ referring only to linear procedures. Let $\rho_{L,P}(m)$ denote the linear minimax risk for this Poisson problem. The linear minimax estimator is given by $\delta_m^L(x) = \frac{m}{m+1}x$ and the linear minimax risk $\rho_{L,P}(m) = \frac{m}{m+1}$. The analogous minimax and linear minimax estimation problems on a general bounded interval $[m_1, m_2]$ were also resolved in [18], provided that affine estimators were included in the definition of the minimax risk. They also found an upper bound of 1.251 for the Ibragimov-Hasminskii constant for this problem of estimating a bounded Poisson mean which lies in an interval of the form $[0, m], m > 0$.

Johnstone and MacGibbon [18] also considered minimax estimation on hyperrectangles and more general convex domains using the information-normalized loss. For each $\tau = (\tau_i)_{i=1}^{\infty}$, $((\tau_1 \geq 0$ for each $i)$, let $[0, \tau]$ denote the hyperrectangle $= \Pi_i[0, \tau_i]$. A set $\mathcal{T}$ is said to be rectangularly convex at 0, if for each $\tau \in \mathcal{T}$, the hyperrectangle $[0, \tau]$ is in $\mathcal{T}$. In [18], it was shown that if the mean is known to lie in $\mathcal{T}$, a compact convex domain, rectangularly convex at 0, then the minimax risk among linear estimates denoted by $\rho_{L,P}(\mathcal{T})$ is within a factor 1.251 of the minimax risk. It was also shown that $\rho_{L,P}(\mathcal{T}) = \sup_{\tau_i}\{\sum_i \frac{\tau_i}{1+\tau_i} : [0, \tau] \in \mathcal{T}\}$. A similar problem for infinite dimensional hyperrectangles which are centered at a non-zero point was also considered in [18]. For this problem affine estimators were used in the calculation of the linear minimax risk.

In [19], using the special mathematical properties of this information-normalized loss function for the Poisson model, information inequalities as found in Brown and Gajek [6], and the polydisc transform, the asymptotic minimax estimator for the Poisson model was obtained in a fashion analogous to Bickel's work [3] for the normal estimation problem. These results will be used here to prove the analogue of Pinsker's result [24] concerning Gaussian random variables for the Poisson case.

Now let us consider this result of Pinsker's [24], one of the most interesting ones in the study of the estimation of constrained Gaussian means. A special case of it can be described in detail as follows.

Let us consider a Gaussian sequence model:

$$y_i = \omega_i + z_i, \quad \text{with } z_i \overset{\text{i.i.d}}{\sim} N(0, 1), i = 1, 2, \ldots,$$

where the problem is to estimate (ω_i) belonging to an ellipsoid

$$\Omega_{\epsilon(n)} = \left\{\omega : \sum_{i=1}^{n} a_i^2 \omega_i^2 \leq \frac{c^2}{\epsilon^2(n)}\right\}, \tag{2.4}$$

where $\epsilon(n) > 0$ and $n \uparrow \infty$ and $n\epsilon^2(n) \sim$ constant.

Henceforth, in order to distinguish between the Poisson and Gaussian cases, the subscripts P and G will be used. (The subscript L will always refer to linear estimation. Let us denote by $\rho_G(\Omega_{\epsilon(n)})$ the minimax risk (in the Gaussian problem) on $\Omega_{\epsilon(n)}$. Pinsker [24] gave an exact evaluation of the asymptotic minimax risk for this problem as $n \uparrow \infty$ and $n\epsilon^2(n) \sim$ constant:

$$\rho_G\left(\Omega_{\epsilon(n)}\right) \sim \sup\left\{\sum_i \frac{\omega_i^2}{1 + \omega_i^2} : \sum_i a_i^2 \omega_i^2 \leq \frac{c^2}{\epsilon^2(n)}\right\}, \tag{2.5}$$

and thus proving that $\frac{\rho_{L,G}(\Omega_\epsilon(n))}{\rho_G(\Omega_{\epsilon(n)})} \to 1$ as $n \uparrow \infty$ and $n\epsilon^2(n) \sim$ constant where $\rho_G(\Omega_{\epsilon(n)})$ denotes the minimax risk and $\rho_{L,G}(\Omega_{\epsilon(n)})$ denotes the linear minimax risk on $\Omega_{\epsilon(n)}$. The linear minimax risk has been shown to equal $\sup_{\omega_i}\{\sum_i \frac{\omega_i^2}{1+\omega^2} : \sum_i a_i^2 \omega_i^2 \leq \frac{c^2}{\epsilon^2(n)}\}$ for the Gaussian problem. A generalization of this result over ℓ_p-balls for ℓ_q-error was given by Donoho and Johnstone [9].

Here, we prove the Poisson analogue of Pinsker's result which can be stated as follows. The risk functions have the subscript P in order to denote the Poisson estimation problem with respect to information-normalized loss.

Theorem 2.1. *Let us consider the $\epsilon^2(n)-$scaled ℓ_1-ellipsoid $\mathcal{T} = \mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}$ in $\mathbb{R}^n_+$ where*

$$\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}} = \left\{\tau : \sum_{i=1}^{n} a_i^2 \tau_i \leq \frac{c^2}{\epsilon^2(n)}\right\}. \tag{2.6}$$

If $n \uparrow \infty$ and $\epsilon(n) \to 0$ such that $n\epsilon^2(n) \sim$ constant, and with $\rho_P(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}})$ denoting the minimax risk for estimating the Poisson mean vector under the information-normalized loss $= \sum_{i=1}^n \tau_i^{-1}(\delta_i - \tau_i)^2$, then the limit is given by:

$$\left(\rho_P\left(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}\right) \Big/ \sup\left\{\sum_i \frac{\tau_i}{1+\tau_i} : \sum_i a_i^2\tau_i \leq \frac{c^2}{\epsilon^2(n)}\right\}\right) = 1 + o(1) \tag{2.7}$$

and this limit is equal to the linear minimax risk on $\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}$. In other words, the limit as $n \uparrow \infty$, such that $n\epsilon^2(n) \sim$ constant, of the ratio of the linear minimax risk $\rho_{L,P}$ to the minimax risk,

$$\frac{\rho_{L,P}(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}})}{\rho_P(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}})} \to 1, \tag{2.8}$$

where

$$\left(\sup\left\{\sum_i \frac{\tau_i}{1+\tau_i} : \sum_i a_i^2\tau_i \leq \frac{c^2}{\epsilon^2(n)}\right\} \Big/ \rho_P\left(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}\right)\right) = 1 + o(1). \tag{2.9}$$

Proof. The polydisc transform plays a fundamental role in proving this result as it did in the asymptotic minimax results for Poisson mean estimation under information–normalized loss obtained by Johnstone and MacGibbon [19]. Recall that the polydisc transform is defined as follows: the polydisc transform is the many–to–one mapping $\tau_* : \mathbb{R}^{2n} \to \mathbb{R}^n_+$ defined by

$$\tau_* : (\omega_1, \omega_2, \ldots, \omega_{2n-1}, \omega_{2n}) \to (\omega_1^2 + \omega_2^2, \ldots, \omega_{2p-1}^2 + \omega_{2p}^2).$$

For each set $\mathcal{T}$ in $\mathbb{R}^n_+$, the Poisson mean parameter space, $\Omega = \tau_*^{-1}(\mathcal{T})$ will denote the pre-image of $\mathcal{T}$. Note that

$$\Omega_{2,2n,a^*,\frac{c^2}{\epsilon^2(n)}} = \tau_*^{-1}\left(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}\right), \tag{2.10}$$

where $a = a_1, a_2, a_3, \ldots$ and $a^* = a_1, a_1, a_2, a_2, a_3, a_3, \ldots$.

Let us recall the following necessary terminology and results from [19]. In [19] one of the deepest results from the theory of elliptic partial differential equations [14] was used to show that under suitable regularity conditions on the boundary of $\mathcal{T}$, which $\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}$ obviously satisfies, then, for any positive m, the minimax risk with respect to the information-normalized loss function satisfies

$$\rho_P(mT) = \inf_{\delta\in\Delta} \sup_{\tau\in mT} E_\sigma \sum_{i=1}^{n} \tau_i^{-1}[\delta_i(X) - \tau_i]^2 = n - m^{-1}\lambda(\Omega) + o(m^{-1}). \tag{2.11}$$

where $\lambda(\Omega)$ denotes the minimum eigenvalue of the Laplace operator $= \sum_1^{2n} \partial^2/\partial\omega_j^2$ on $\Omega = \tau_*^{-1}(T)$, that is, the smallest λ for which the equation

$$\begin{cases} \Delta u(\omega) = -\lambda u(\omega), & \omega \in \operatorname{int}(\Omega), \\ u(\omega) = 0, & \omega \in \partial(\Omega), \end{cases} \tag{2.12}$$

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 $-u(\omega)$] is strictly positive on Ω (cf. [14]). Assume that u_Ω is normalized so that $\int_\Omega u_\Omega^2 = 1$.

Johnstone and MacGibbon [19] considered the connection between the n-dimensional Poisson estimation problem and the $2n$-dimensional Gaussian location estimation problem induced by the polydisc transform. They studied the information-like functionals that arise in studying Bayes risks in Poisson estimation and found analogies with the role of Fisher information in Bayes estimation of a Gaussian mean vector. In the latter case (see [3]), when $X \sim N_{2n}(\theta, I)$, Brown's identity connects the Bayes risk $r_{2n}(H) = \inf_\delta \int E_\theta[\delta(X) - \theta]^2 H(d\theta)$ for estimation of θ with respect to an absolutely continuous prior density $H(d\theta) = h(\theta)\, d\theta$ with Fisher information $I(H) = \int |Dh|^2/h$ via the identity,

$$r_{2n}(H) = 2n - I_{2n}(H \star \Phi),$$

where Φ denotes the standard Gaussian distribution function in $\mathbb{R}^{2n}$, and I_{2n} represents the multivariate Fisher information of dimension $2n$. If $\theta = m\tau$ and the prior $H = \sigma_m F$ are transforms of $F(d\tau)$ under the scaling $\sigma_m : \tau \to m\tau$, then

$$r_{2n}(\sigma_m F) = 2n - m^{-2} I_{2n}(F \star \Phi_{1/m}),$$

where $\Phi_{1/m}$ denotes the Gaussian distribution $N_{2n}(0, m^{-1}I)$.

The corresponding quantity in the estimation of an $n-$dimensional Poisson mean $\sigma = \sigma_m(\mathcal{T}) = m\tau$ given the prior $F(d\tau) = f(\tau)d\tau$ is defined in [19] and its limit asymptotically as $m \to \infty$ is shown to be given by

$$J_n(F) = \int \sum_{i=1}^n f^{-1}(D_i f)^2(\tau)\tau_i \, d\tau.$$

$J_n(F)$ is related to Fisher information via the identity

$$J_n(F) = 4^{-1}\pi^{-p} I_{2n}(g), \qquad g(\omega) = f(\tau_*(\omega)),$$

where τ_* represents the polydisc transform.

Such identities were used in [19] to prove (2.11). Now (2.11) can be combined with the well known result (see, e.g., the solution for the normal problem on a sphere [3]) that if the convex domain Ω satisfies certain regularity conditions, which are obviously satisfied by $\Omega_{2,2n,a^*,c^2}$, then asymptotically as $m \to \infty$,

$$\rho_G(m\Omega) = 2n - \frac{4}{m^2}\lambda(\Omega) + o(m^{-2}). \tag{2.13}$$

Letting $\Omega_2 = \Omega_{2,2n,a^*,c^2}$ and letting $m = \epsilon^{-1}(n)$ we obtain:

$$\rho_G\left(\frac{\Omega_2}{\epsilon(n)}\right) = 2n - 4\epsilon^2(n)\lambda(\Omega_2) + o(\epsilon^2(n)).$$

Now scaling by $(\sqrt{2})^{-1}$,

$$\rho_G\left(\frac{\sqrt{2}\Omega_2}{\epsilon(n)}\right) = 2n - 2\epsilon^2(n)\lambda(\Omega_2) + o(\epsilon^2(n)).$$

In an analogous fashion in the Poisson case, with $\mathcal{T}_1 = \mathcal{T}_{1,n,a,c^2}$, using (2.11) and letting $m = \epsilon^{-2}(n)$, and multiplying by $\epsilon^2(n)$,

$$\rho_P\left(\frac{\mathcal{T}_1}{\epsilon^2(n)}\right) = n - \epsilon^2(n)\lambda(\Omega_2) + o(\epsilon^2(n)). \tag{2.14}$$

Thus if $n\epsilon^2(n) \sim$ constant and $\epsilon(n) \to 0$ as $n \uparrow \infty$, then it follows from (2.14) and (2.14) that

$$2\,\rho_P\left(\frac{\mathcal{T}_1}{\epsilon^2(n)}\right) \sim \rho_G\left(\frac{\sqrt{2}\Omega_2}{\epsilon(n)}\right). \tag{2.15}$$

Using Pinsker's result [24]and (2.15), with $\epsilon^2(n)$ satisfying the above conditions, we have the following asymptotic equivalence:

$$\begin{aligned}\rho_P\left(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}\right) &\sim \left(\frac{1}{2}\right)\sup\left\{\sum_i \frac{\omega_i^2}{1+\omega_i^2} : \sum_i (a_i^*)^2\omega_i^2 \le \frac{2c^2}{\epsilon^2(n)}\right\} \\ &= \sup\left\{\sum_i \frac{\tau_i}{1+\tau_i} : \sum_i a_i^2\tau_i \le \frac{c^2}{\epsilon^2(n)}\right\}.\end{aligned} \tag{2.16}$$

The above equality follows from the properties of the polydisc transform, of $a = (a_i)$ and $a^* = (a_i^*)$ and of the function $f(x) = x/(1+x)$.

It follows from arguments used in Lemma 3 of [19] that the minimax linear risk on $\mathcal{T}_1$, $\rho_{L,P}(\mathcal{T}_1) = \sup\left(\sum_{i=1}^n \frac{\tau_i}{1+\tau_i} : \sum_i a_i^2\tau_i \le c^2\right)$. Thus,

$$\rho_{L,P}\left(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}\right) = \sup\left\{\sum_i \frac{\tau_i}{1+\tau_i} : \sum_i a_i^2\tau_i \le \frac{c^2}{\epsilon^2(n)}\right\}.$$

Now combining this with the previous three equations, we have as $n \uparrow \infty$ with $n\epsilon^2(n) \sim$ constant

$$\left(\sup\left\{\sum_i \frac{\tau_i}{1+\tau_i} : \sum_i a_i^2\tau_i \le \frac{c^2}{\epsilon^2(n)}\right\} \Big/ \rho_P\left(\mathcal{T}_{1,n,a,\frac{c^2}{\epsilon^2(n)}}\right)\right) = 1 + o(1),$$

thus proving the desired result. For the full argument of this proof see an updated version of this paper on arXiv.org. □

3. Some results concerning estimation under quadratic loss

Since the application of the asymptotic equivalence results in [7] for estimating the intensity function of a Poisson process would use unweighted quadratic loss as is usual with Gaussian estimation, it would be fruitful to look at the estimation of constrained Poisson mean vectors under quadratic loss.

We recall some of the results from MacGibbon et al. [23] who studied the problem of Poisson mean estimation under quadratic loss with the additional assumption that Θ lies in an interval on a bounded interval $[0, m], m > 0$. In [23] the minimax risk, $\rho(m)$, the linear minimax risk, $\rho_L(m)$, and the affine minimax risk, $\rho_A(x)$ and their associated minimax estimators were determined for this problem.

The linear minimax estimator was given by $\delta_m^L(x) = \frac{m}{m+1}x$ with the linear minimax risk $\rho_L(m) = \frac{m^2}{m+1}x$, and the affine minimax estimator was given by

$$\delta_m^A(x) = \sqrt{m+1} - 1 + \left(1 - \frac{1}{\sqrt{m+1}}\right) x$$

with the affine minimax risk $\rho_A(m) = \left[\sqrt{m+1} - 1\right]^2$.

For small $m (m \leq 0.91)$, they found that the least favorable prior π_m is a two point prior given by $a_m \epsilon_{\{0\}}(n) + (1 - a_m)\epsilon_{\{m\}}(n)$ where a_m satisfies $R(0, \delta_{\pi_m}) = R(m, \delta_{\pi_m})$. The minimax estimator is then given by

$$\delta_m(x) = m \quad \text{if} \quad x \geq 1$$
$$\text{and} \quad \delta_m(0) = \frac{(1 - a_m)e^{-m}m}{a_m + (1 - a_m)e^{-m}}, \text{ with } \ a_m = \frac{1}{1 + e^{\frac{m}{2}}}.$$

This yields a minimax risk of $\rho(m) = r(\pi_m) = \frac{m^2}{(1+e^{\frac{m}{2}})^2}$, for $0 < m \leq m_0$.

A global optimization technique (cf. [16]) for finding the maximum of a Lipschitz continuous function is used to find the minimax solution in the case of a three point prior. Since, in order to determine the Ibragimov-Hasminskii constant for this problem, a modification of Kempthorne's [20] iterative procedure for numerical specification of discrete least favorable priors was used as m and the number of points in the prior increases to obtain the values of the Ibragimov-Hasminskii constants, $\mu_A(m) = \frac{\rho_A(m)}{\rho(m)}$ and $\mu_L(m) = \frac{\rho_L(m)}{\rho(m)}$, the bound m was increased until $\mu_A(m)$ was clearly decreasing. It was shown numerically that $\mu_L(m)$ decreased from its value of 4 at $m = 0$. By a judicious choice of prior in Brown's information inequality for the Bayes risk (cf. [6]), $\lim_{m\to\infty} \mu_A(m)$ and $\lim_{m\to\infty} \mu_L(m)$ can be bounded.

The Ibragimov-Hasminskii constant, $\mu_A = \sup_m \frac{\rho_A(m)}{\rho(m)}$, for the bounded Poisson mean problem under quadratic loss was shown numerically to be approximately 1.56 in [23]. When only linear estimators were considered then obviously it was shown that numerically μ_L was equal to 4.

The standard arguments that the problem on hyperrectangles can be reduced to considering the bounded interval problem used in [10] show that the Ibragimov-Hasminskii constants remain the same over hyperrectangles in this Poisson problem.

In [10] the Gâteaux derivative was used to establish the result that when the mean of an infinite dimensional Gaussian vector is known to lie in a compact convex orthosymmetric domain, quadratically convex in ℓ_2, then the minimax risk among linear estimates is within a factor 1.25 of the minimax risk. In order to compare the linear minimax risk with the minimax risk when estimating the mean of an infinite dimensional Poisson vector known to lie in Θ, a compact convex domain, rectangularly convex at 0 and quadratically convex in ℓ_2, Gâteaux derivatives were also used in [23]. They showed that the Ibragimov-Hasminskii constant in this case was equal to 4. Since the arguments in [23] are more subtle than those in [10], we give a brief summary of the proof of the result from [23] here.

For each $\theta \in \Theta$, let $J(\theta) = \rho_L([0,\theta]) = \sum_i \frac{\theta_i^2}{\theta_i+1}$ denote the linear minimax risk on the hyperrectangle $[0,\theta]$. Arguing as is in the proof of Theorem 7 [10], J has a maximum τ^{**} on Θ. In order to complete the proof, it suffices to show that $R(\delta,\theta) \leq J(\tau^{**})$ for all $\theta \in \Theta$.

The same change of variable as in [10] was used in [23]; that is, let $t = (t_i)_i$ be defined by $t_i = \theta_i^2$ for each i. Clearly, $T = \{t : t_i = \theta_i^2,\ \theta_i \in \Theta\}$ is convex, because Θ is quadratically convex. Let us define $\tilde{J}$ on T by $\tilde{J}(t) = J(\theta)$. Clearly, since $\tau^* = \tau^{**2}$ is a maximum of $\tilde{J}$, the Gâteaux derivative of $\tilde{J}$ at $\tau^* = D_{\tau^*}\tilde{J}$ is negative at τ^*. To complete the proof, it must be shown that $R(\delta,\theta) - J(\tau^*) \leq D_{\tau^*}\tilde{J}$ for all $\theta \in \Theta$.

Now, $R(\delta_L^*,\theta) - J(\tau^*) = \sum \frac{(\sqrt{t_i}-\sqrt{\tau_i^*})(\sqrt{t_i}+\sqrt{\tau_i^*}+\tau_i^*)}{(\sqrt{\tau_i^*}+1)^2}$. Let $\tilde{J}(t) = \sum_i \frac{t_i}{\sqrt{t_i}+1}$. Then the Gâteaux derivative of $\tilde{J}(t)$ at τ^*, $D_{\tau^*}\tilde{J}$, which is negative, is given by

$$D_{\tau^*}\tilde{J} = \sum \left(\sqrt{t_i} - \sqrt{\tau_i^*}\right)\left(\frac{1}{2}\sqrt{t_i}\sqrt{\tau_i^*} + \frac{1}{2}\tau_i^* + \sqrt{t_i} + \sqrt{\tau_i^*}\right).$$

It was shown in [23] that by considering for all $\theta \in \Theta^*$ (with $t = \theta^2$) the two cases, $\sqrt{\tau_i} - \sqrt{\tau_i^*}$, positive or negative, term by term, that $D_{\tau^*}\tilde{J} \geq R(\delta_L^*,\theta) - J((\tau^*)$.

Remark 3.1. We conjecture that for the Poisson estimation problem, the form of infinite dimensional subsets Θ for which the ratio of $\rho_L(\Theta)$ to $\rho(\Theta)$ is bounded by the univariate Ibragimov-Hasminskii constant can be different for each of the loss functions studied here. In particular, from [18] it is known that the bound holds for l_1-rectangles with respect to information-normalized loss. An argument was given in [10] that linear estimators can be arbitrarily outperformed by nonlinear ones in the sense of minimax risk for ℓ_1-cubes in the Gaussian estimation problem using quadratic loss. We conjecture that a similar argument can be used for Poisson mean estimation on ℓ_1-cubes under quadratic loss to prove an analogous result.

4. Conclusion

A first step has been taken to extend the results of Johnstone and MacGibbon [19]. Using the methods of proof developed here it has been possible to prove the Poisson analogue of Pinsker's result [24] with respect to information-normalized loss. Although the methods of [19] are not applicable to the problem of proving an analogous result to that of Pinsker [24] for the case of Poisson minimax estimation with respect to quadratic loss, the asymptotic equivalence results of [7] can be used to obtain minimax estimators of the intensity of a non-homogeneous Poisson process with respect to quadratic loss. It would be interesting to develop a suitable asymptotic equivalence in order to find minimax estimators of the intensity function of a Poisson process with respect to the information-normalized loss function.

However, it is also to be expected that several other interesting results concerning Poisson mean estimation under information-normalized loss can be proved in an analogous fashion to the Gaussian case by a judicious use of the polydisc transform.

Acknowledgements

The author would like to thank NSERC of Canada for supporting this research and to thank Professor Larry Brown and acknowledge her indebtedness to him for introducing her to statistical decision theory. The author would also like to thank the referee for helpful comments that improved this manuscript.

References

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

[2] Berkin, P. E. and Levit, B. Y. (1980). Second order asymptotically minimax estimates for the mean of a normal population. *Problems in Information Transmission* **12** 212–229.

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

[4] Brown, L. D. (1979). Statistical decision theory. Technical report, Cornell Univ. (mimeographed notes).

[5] Brown, L. D. (2000). An essay on statistical decision theory. *J. Amer. Statist. Assoc.* **95** 1277–1281.

[6] Brown, L. and Gajek, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18** 1578–1594.

[7] Brown, L. D., Carter, A. V., Low, M. G. and Zhang, C. H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *Ann. Statist.* **32** 2074–2097.

[8] Casella, G. and Strawderman, W. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** 868–876.

[9] Donoho, D. L. and Johnstone, I. M. (1994). Minimax risk over l_p balls for l_q-error. *Probab. Theory Related Fields* **99** 277–303.

[10] Donoho, D., Liu, R. and MacGibbon, B. (1990). Minimax risk over hyperrectangles and implications. *Ann. Statist.* **18** 1416–1437.

[11] Feldman, I. and Brown, L. D. (1989). The minimax risk for estimating a bounded normal mean. Technical report, Cornell Univ., Cornell Statistical Center Series.

[12] Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic, New York.

[13] Ghosh, M. N. (1964). Uniform approximation of minimax point estimates. *Ann. Math. Statist.* **35** 1031–1047.

[14] Gilbarg, D. and Trudinger, N. S. (1983). *Elliptic Partial Differential Equations of Second Order*, 2nd ed. Springer, Berlin, Heidelberg, New York.

[15] Gill, P., Murray, W., Saunders, M. and Wright, M. (1986). User's guide for NPSOL (Version 4.0): A FORTRAN package for non-linear programming, Systems Optimization Laboratory. Technical report, Dept. Mathematics and Statistics.

[16] Gourdin, E., Jaumard, B. and MacGibbon, B. (1994). Global optimization decomposition methods for bounded parameter minimax risk evaluation. *SIAM J. Sci. Comput.* **15** 16–35.

[17] Ibragimov, I. A. and Hasminskii, R. Z. (1984). Nonparametric estimation of the value of a linear function in Gaussian white noise. *Theory Probab. Appl.* **29** 1–32.

[18] Johnstone, I. M. and MacGibbon, K. B. (1992). Minimax estimation of a constrained Poisson vector. *Ann. Statist.* **20** 807–831.

[19] Johnstone, I. M. and MacGibbon, K. B. (1993). Asymptotically minimax estimation of a constrained Poisson vector via polydisc transforms. *Ann. Inst. H. Poincaré (B)* **29** 289–319.

[20] Kempthorne, P. J. (1987). Numerical specification of discrete least favorable prior distributions. *SIAM J. Sci. Statist. Comput.* **8** 171–184.

[21] Levit, B. Y. (1980). On asymptotic minimax estimates of the second order. *Theory Probab. Appl.* **25** 552–568.

[22] Levit, B. Y. (1982). Minimax estimation and positive solutions of elliptic equations. *Theory Probab. Appl.* **27** 563–586.

[23] MacGibbon, B., Gourdin, E., Jaumard, B. and Kempthorne, P. (2000). Minimax estimation of exponential family means over *lp* bodies under quadratic loss. In *Stochastic Models, A Volume in Honour of Donald A. Dawson* (L. Gorostiza and G. Ivanoff, eds.) **26** 319–330. Canadian Mathematical Society Conference Proceedings.

[24] Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems of Information Transmission* **16** 52–68.

[25] Vidakovic, B. and Dasgupta, A. (1996). Efficiency of linear rules for estimating a bounded normal mean. *Sankhyā* **58** 81–100.

[26] Wald, A. (1950). *Statistical Decision Functions*. Wiley, New York.

[27] Zinzius, E. (1981). Minimaxschätzer für den Mittelwert v einer normalverteilten Zufallgröße mit bekannter Varianz bei vorgegebener oberer und unterer Scranke für v. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 551–557.