IMS Collections
Borrowing Strength: Theory Powering Applications –
A Festschrift for Lawrence D. Brown
Vol. 6 (2010) 22–31

DOI: 10.1214/10-IMSCOLL602

Chi-square lower bounds

Mark G. Low

University of Pennsylvania

Abstract: The information inequality has been shown to be an effective tool for providing lower bounds for the minimax risk. Bounds based on the chi-square distance can sometimes offer a considerable improvement especially when applied iteratively. This paper compares these two methods in four examples including the bounded normal mean problem as well as obervations from a Poisson distribution.

1. Introduction

Lower bounds for the minimax risk under squared error loss are often used as a benchmark for the evaluation of statistical procedures. One useful technique for providing such bounds is based on the information inequality. The basic idea as given in [2] can be explained as follows. Fix a measure μ and suppose that X has a density f_θ where θ is assumed to lie in an interval $[a, b]$. Suppose that M is the minimax risk under squared error loss for estimating θ and that $b(\theta)$ is the bias function of a minimax estimator. Let $J(\theta)$ be the Fisher information. The information inequality then implies that

$$M \geq \frac{(1 + b^{'}(\theta))^2}{J(\theta)} + b^2(\theta), \tag{1}$$

where b is the bias function. Now suppose that the minimax value M in (1) is replaced by a parameter r and that there are no differentiable functions β on $[a, b]$ which satisfy the differential inequality

$$\beta^{'}(\theta) \leq (J(\theta)(r - \beta^2(\theta)))^{1/2} - 1.$$

It then follows that r is a lower bound for the minimax risk.

This technique of finding lower bounds was used in [2] to find lower bounds for the risk in estimating the value of a probability density. In [4] this method was contrasted to Bayes risks bounds as developed in [3] and shown in many cases to yield much stronger lower bounds.

Another approach to finding lower bounds is contained in [5]. This constrained risk inequality focuses just on two points in a parameter space. Once again fix a measure a measure μ and suppose that X has either density f_{θ_1} or f_{θ_2} with respect to μ. Let $R(\phi, \delta)$ be the risk under square error loss, namely

$$R(\phi, \delta) = \int (\phi - \delta(x))^2 f_\phi(x) \mu(dx)$$

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104

Keywords and phrases: minimax lower bound, information inequality, bounded normal mean, Chi-square distance.

AMS 2000 subject classifications: Primary 62G07; secondary 62G20.

and let $I(\theta_1, \theta_2) = (\int \frac{f_{\theta_2}^2(x)}{f_{\theta_1}(x)} \mu(dx))^{1/2}$. Set $\theta = \theta_2 - \theta_1$. If $0 < \epsilon < \frac{|\theta|}{I}$ and $R(\theta_1, \delta) \leq \epsilon^2$ then

$$R(\theta_2, \delta) \geq (|\theta| - \epsilon I)^2.$$

This two point inequality was originally developed as a way to provide lower bounds for adaptive estimation problems. Given the mean squared error at one point it bounds the mean squared error at another point. In applications it was used to show that a minimax rate optimal procedure over one function class may sometimes need to pay a penalty over a another parameter space. See for example [6].

It can however also be effectively used to provide bounds for the minimax risk. In Section 2 we give two examples where sharp minimax results can be attained in this way. In Section 3 we show that the constrained risk inequality follows from some bounds originally due to Hammersley [12] and Chapman and Robbins [9]. The Hammersley-Chapman-Robbins bounds were originally focused on unbiased estimators but they are redeveloped here in terms of trading bias and variance. An upper bound on the variance at one point leads to a bound on how quickly the bias must change from that point to another point. Finally in Section 4 we show these bias-variance tradeoff bounds can be applied iteratively to yield effective bounds on the minimax risk. The technique is analagous to bounds obtained from the information inequality. Moreover it is at least as strong and often yields significantly improved bounds. In this section we compare these new bounds to those from the information inequality as well as to exact minimax results when known.

2. A simple minimax result using the constrained risk inequality

In this section we present two examples where the inequality is sharp. Both of these examples were also considered in [14] but in that paper constraints were not put on the parameter space and the attention was on unbiased estimators.

We focus here on minimax statements for the mean squared error. Note that by balancing the risk at the two parameter points in the constrained risk inequality the following minimax result immeadiately follows.

$$\inf_{\hat{\theta}} \sup_{\theta=\theta_0,\theta_1} E(\hat{\theta} - \theta)^2 \geq \frac{\theta^2}{(1+I)^2}. \tag{2}$$

As a first example where this inequality is sharp let f_θ be the density of a uniform $[0, 1-\theta]$ for $0 \leq \theta \leq a$. It follows that if $\theta_1 > \theta_0$ then

$$I^2(\theta_0, \theta_1) = \frac{(1-\theta_0)}{(1-\theta_1)}$$

whereas if $\theta_0 > \theta_1$ then $I(\theta_0, \theta_1)$ is infinite. Let us consider the case when $\theta_0 = 0$ and $\theta_1 = a$. Then (2) shows that

$$\inf_{\hat{\theta}} \sup_{\theta=\theta_0,\theta_1} E(\hat{\theta} - \theta)^2 \geq \frac{\theta^2}{(1+I)^2} \tag{3}$$

$$= a^2 \frac{(1-a)}{(\sqrt{1-a}+1)^2}. \tag{4}$$

Now let

$$\delta(x) = 0 \quad if \quad 1-a \le x \le 1, \tag{5}$$

$$\delta(x) = \frac{a}{1+\sqrt{1-a}} \quad if \quad 0 \le x \le 1-a. \tag{6}$$

Then it is simple to check that

$$R(\theta,\delta) = \theta^2 \frac{a-\theta}{1-\theta} + \left(\frac{a}{1+\sqrt{1-a}} - \theta\right)^2 \frac{1-a}{1-\theta}. \tag{7}$$

This risk function is maximized when a is close to 0.75 but it is not always maximized at the end points. It is however easy to check numerically that the risk function is maximized at the endpoints when $a \le 0.5$. In particular if $0 \le \theta \le a \le 0.5$ the two point bound gives a sharp lower bound for the minimax risk. It should be mentioned that the information inequality method mentioned in the introduction is not applicable here as the Fisher Information is infinite.

2.1. Shifted exponential

Let X have density f_θ where $f_\theta = \exp(x-\theta)1(x \ge \theta)$. X is a shifted expontial. Assume that $0 \le \theta \le 1$. For $\theta_2 > \theta_1$, $I^2(\theta_1,\theta_2) = \exp(\theta_2 - \theta_1)$. It then follows that

$$\inf_{\hat\theta} \sup_{\theta=0,a} E(\hat\theta - \theta)^2 \ge a^2 \left(\frac{1}{1+\exp(a/2)}\right)^2.$$

In this problem it is easy to calculate the minimax procedure when the parameter space if restricted to the two values 0 and a. Note that a sufficient statistic for this problem is $1(x \ge \theta)$. Hence a minimax procedure is given by

$$\delta(x) = 0 \quad if \quad 0 \le x < a, \tag{8}$$

$$\delta(x) = ca \quad if \quad x \ge \theta. \tag{9}$$

Then it is simple to check that

$$R(0,\delta) = c^2a^2\exp(-a), \tag{10}$$

$$R(a,\delta) = (c-1)^2a^2 \tag{11}$$

and that the minimax procedure corresponds to a choice of $c = \frac{1-\exp(-a/2)}{1-\exp(-a)}$ and that the maximum risk at the points 0 and a are equal to the lower bound.

It is also easy to calculate the risk of the procedure δ for other values of θ. With the choice of $c = \frac{1-\exp(-a/2)}{1-\exp(-a)}$

$$R(\theta,\delta) = \theta^2 - 2ca\theta\exp(-(a-\theta)) + c^2a^2\exp(-(a-\theta)).$$

It is easy to check numerically that the maximum is attained at the endpoints when $a \le 0.9$ but not when $a = 1$. It then follows that for the parameter space $0 \le \theta \le 0.9$ the procedure given above will be minimax over $0 \le \theta \le 0.9$ and not just minimax over the two points of 0 and 0.9. Note also that since the Fisher information is also infinite in this example the information inequality method is also not applicable here.

3. Some improved bounds

There are two basic strategies. Both use Cauchy-Schwartz. Let δ be an estimator of the parameter θ. Assume that $\theta_1 > \theta_0$

$$E_{f_{\theta_0}}\left\{(\delta - a)\left(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}\right)\right\} = \theta_1 + B(\theta_1) - a$$
$$\leq (E_{f_{\theta_0}}(\delta - a)^2)^{1/2}\left(E_{f_{\theta_0}}\left(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}\right)^2\right)^{1/2}.$$

This then yields

$$B(\theta_1) \leq a - \theta_1 + (E_{f_{\theta_0}}(\delta - a)^2)^{1/2} I(\theta_0, \theta_1)$$

giving an upper bound on the bias at the point θ_1 as a function of a.

Now assume that $\theta_1 < \theta_0$. Then

$$E_{f_{\theta_0}}\left\{(a - \delta)\left(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}\right)\right\} = a - (\theta_1 + B(\theta_1))$$
$$\leq (E_{f_{\theta_0}}(\delta - a)^2)^{1/2}\left(E_{f_{\theta_0}}\left(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}\right)^2\right)^{1/2}.$$

This then yields

$$B(\theta_1) \geq a - \theta_1 - (E_{f_{\theta_0}}(\delta - a)^2)^{1/2} I(\theta_0, \theta_1).$$

By taking $a = \theta_0$ and assuming as in [5] that $(E_{f_{\theta_0}}(\delta - a)^2)^{1/2} I(\theta_0, \theta_1) \leq \theta_1 - \theta_0$ it immediately follows that

$$B^2(\theta_1) \geq (|\theta| - \epsilon I(\theta_0, \theta_1))^2.$$

This bound thus improves on the original constrained risk bound in the sense that it now shows that the bound for mean squared error at the point θ_1 is really a bound for the squared bias at that point. Moreover it shows that although the original bound in [5] can be sharp, it can only be sharp in cases where the variance at one of the parameter points is in fact equal to zero.

Now the choice of $a = \theta_0$ is not the only possible choice. Let $a = \theta_0 + B(\theta_0) + y$. Then

$$B(\theta_1) \leq \theta_0 + B(\theta_0) + y - \theta_1 + (E_{f_{\theta_0}}(\delta - E_{f_{\theta_0}}\delta - y)^2)^{1/2} I(\theta_0, \theta_1)$$

and hence

$$B(\theta_1) - B(\theta_0) \leq \theta_0 - \theta_1 + y + (Var_{f_{\theta_0}}(\delta) + y^2)^{1/2} I(\theta_0, \theta_1).$$

Now the right hand side of this inequality is clearly minimized for some $y \leq 0$ and it is easy to check that the minimum is attained when $y = -\sqrt{\frac{Var_{f_{\theta_0}}(\delta)}{I(\theta_0,\theta_1)^2 - 1}}$ which then yields

$$B(\theta_1) - B(\theta_0) \leq \theta_0 - \theta_1 + \sqrt{Var_{f_{\theta_0}}(\delta)}(I^2(\theta_0, \theta_1) - 1)^{1/2}.$$

Now suppose that we focus one again on a bound on the mean squared error at zero with $B^2(\theta_0) + Var_{f_{\theta_0}}(\delta) \le \epsilon^2$. It follows that

$$B(\theta_1) \le \theta_0 - \theta_1 + B(\theta_0) + \sqrt{\epsilon^2 - B^2(\theta_0)}(I^2(\theta_0,\theta_1) - 1)^{1/2}.$$

The right hand side of this inequality is maximized when $B(\theta_0) = \frac{\epsilon}{I(\theta_0,\theta_1)}$ yielding the original inequality of Brown and Low [5] in terms of the bound on bias. Putting all these results together in the case when $\theta_1 > \theta_0$ it is clear that whenever the inequality (7) sharp it follows that $Var_{f_{\theta_0}}(\delta) = \epsilon^2 \frac{I(\theta_0,\theta_1)-1}{I(\theta_0,\theta_1)}$, $B(\theta_0) = \frac{\epsilon}{I(\theta_0,\theta_1)}$, $Var_{f_{\theta_1}}(\delta) = 0$ and $B(\theta_1) = \epsilon - (\theta_1 - \theta_0)$. In the case when $\theta_1 < \theta_0$ the variances remain unchanged but $B(\theta_0) = -\frac{\epsilon}{I(\theta_0,\theta_1)}$ and $B(\theta_1) = (\theta_0 - \theta_1) - \epsilon$. As a consequence the bounds given in above improve on (7) whenever these conditions do not all hold.

There is a second approach that also yields the above improved bound.

$$\begin{aligned} E_g\left\{(\delta - E_g\delta)\left(\frac{f_{\theta_1}(X) - f_{\theta_0}(X)}{g(X)}\right)\right\} &= \theta_1 + B(\theta_1) - (\theta_0 + B(\theta_0)) \\ &\le (Var_g\delta)^{1/2}\left(E_g\left(\frac{f_{\theta_1} - f_{\theta_0}}{g}\right)^2\right)^{1/2}. \end{aligned} \tag{12}$$

By varying g we can create a collection of bounds for the change in bias and it is often not immediately clear how best to select a g. In particular in parametric problems where we are estimating θ we usually have all $\theta \in [\theta_0, \theta_1]$ and we are free to select for g any f_θ. In such a case $(E_g(\frac{f_{\theta_1} - f_{\theta_0}}{g})^2)^{1/2}$ is often minimized for a g other than f_{θ_0} or f_{θ_1}. This does not however imply that it is best to use such a choice of g because in applications we must also bound the value of $Var_g\delta$. For minimax estimation this is usually done by first assuming a bound on the minimax risk equal to say r and then bounding $Var_g(\delta)$ by $r - B^2(g)$. This bound is often minimized at one of the f_{θ_1} or f_{θ_0} and in the applications considered here it appears best to use one of these values. Thus for the examples in the present paper we focus on taking g to be equal to f_{θ_1} or f_{θ_0} which yield

$$B(\theta_1) - B(\theta_0) \le (Var_{f_{\theta_0}}\delta)^{1/2}(I^2(\theta_0,\theta_1) - 1)^{1/2} - (\theta_1 - \theta_0) \tag{13}$$

and

$$B(\theta_1) - B(\theta_0) \le (Var_{f_{\theta_1}}\delta)^{1/2}(I^2(\theta_1,\theta_0) - 1)^{1/2} - (\theta_1 - \theta_0). \tag{14}$$

These bounds are actually a consequence of a bound due to Hammersley [12] and Chapman and Robbins [9]. The focus in those papers was on unbiased estimators. However every estimator is clearly an unbiased estimator of its expectation and taking this point of view these bounds are identical to the bounds in these papers. See also [7] where a generalization of the Hammersley-Chapman-Robbins inequality is given for other loss functions.

4. Iterative bounds

In many applications the most effective use of the inequalities (13) and (14) for bounding minimax risk is through an iterative scheme. In this section we consider several examples some of which are concerned with weighted squared error loss.

Suppose that

$$r \ge g(\theta)(Var(\theta) + B^2(\theta))$$

for $0 \le \theta \le 1$. It then follows that for every pair θ_1 and θ_2 we must have

$$B(\theta_2) - B(\theta_1) \le (Var_{\theta_1}(\delta))^{1/2}(I^2(\theta_1, \theta_2) - 1)^{1/2} - (\theta_2 - \theta_1) \tag{15}$$

$$\le \left(\frac{r}{g(\theta_1)} - B^2(\theta_1)\right)^{1/2}(I^2(\theta_1, \theta_2) - 1)^{1/2} - (\theta_2 - \theta_1) \tag{16}$$

and also have

$$B(\theta_2) - B(\theta_1) \le (Var_{\theta_2}(\delta))^{1/2}(I^2\theta_2, \theta_1) - 1)^{1/2} - (\theta_2 - \theta_1) \tag{17}$$

$$\le \left(\frac{r}{g(\theta_2)} - B^2(\theta_2)\right)^{1/2}(I^2(\theta_2, \theta_1) - 1)^{1/2} - (\theta_2 - \theta_1). \tag{18}$$

It follows that if there does not exist a function B which satisfies this inequality for all θ then r must be a lower bound for the minimax risk. These bounds will always be at least as strong as theose given in [2].

In this section we consider three examples which illustate the use of these iterative bounds. The first example concerns the well known bounded normal mean problem. Let $X \sim N(\theta, 1)$ with $|\theta| \le L$. This model is important in the development of the theory of estimating linear functionals in nonparametric function estimation problems. See for example and Donoho [10]. It was initially studied by Casella and Strawderman [8] as well as Bickel [1]. In particular Casella and Strawderman calculated the minimax risk for a number of values of $L \le 2$ and these results were later extended by Kuks [15], Donoho, Liu and Macgibbon [11] as well as Feldman and Brown (1989). These results allow for a direct comparison of our lower bounds as well as a comparison with the Information inequality bounds obtained in [4].

In the second example we observe X a binomial (n,p) random variable and the focus is on esimating the canonical binomial parameter under a weighted squared error loss where the weight function is Fisher information. In this model we do not constrain the canonical parameter. This model was also considered in [4]. Finally we shall turn to estimating the mean of a Poisson random variable, once again under the assumption that the mean vector is bounded by a given value. This model was extensively studied in [13] and we shall compare our bounds to the exact minimax results given in that paper.

4.1. Bounded normal mean problem

For $X \sim N(\theta, 1)$ we have $I^2(\theta_1, \theta_2) = \exp((\theta_2 - \theta_1)^2)$. Hence if r is a lower bound for the minimax risk there must be a function B on $[-L, L]$ such that

$$B(\phi) - B(\theta) \le (r - \max(B^2(\phi), B^2(\theta)))^{1/2}(\exp((\phi - \theta)^2) - 1)^{1/2} - (\phi - \theta)$$

for all $\phi > \theta$. It is easy to see such a solution exists if and only there is a function B on $[0, L]$ satisfying $B(0) = 0$ and

$$B(\phi) - B(\theta) \le (r - B^2(\phi))^{1/2}(\exp((\phi - \theta)^2) - 1)^{1/2} - (\phi - \theta) \tag{19}$$

for all $\phi > \theta \ge 0$.

In practice we do not try to discover values of r for which there are solutions but rather values for which we can show that there cannot be a solution. Any such value for which a solution does not exist is thus a lower bound. Moreover we do not in practice consider all possible choices of ϕ and θ but we only consider a grid of values. In this case it is convenient to select an equally spaced grid of points on the interval $[0, L]$. Let N be the number of grid points and for $i = 0, 1, \ldots, N-1$ let $\phi_i = iL/N - 1$. We shall then check for any given value of r whether there exists a solution to (19) on this grid. This can be done as follows. In the first stage starting at the value of $B(0)$ solve for $i > 0$

$$B(\theta_i) - B(0) \leq (r - B^2(\theta_i))^{1/2}(\exp(\theta_i^2) - 1)^{1/2} - \theta_i. \tag{20}$$

The solution gives us an upper bound for each of the values $B(\theta_i)$ for $i > 0$.

At the second stage start from the upper bound on the value $B(\theta_1)$ and solve for $j > 1$

$$B(\theta_j) - B(\theta_1) \leq (r - B^2(\theta_j))^{1/2}(\exp(\theta_j - \theta_1)^2 - 1)^{1/2} - (\theta_j - \theta_1), \tag{21}$$

where $B(\theta_1)$ is actually the upper bound from the first stage. From these first two stages we have two upper bounds for the value of $B(\theta_2)$. Take the smallest of these two values and continue to the next stage solving for each $j > 2$.

$$B(\theta_j) - B(\theta_2) \leq (r - B^2(\theta_j))^{1/2}(\exp(\theta_j - \theta_2)^2 - 1)^{1/2} - (\theta_j - \theta_2). \tag{22}$$

This process is continued. At the *ith* step we start at the point θ_{i-1} where we take for $B(\theta_i)$ the smallest of the upper bounds from the previous $i-1$ stages. At each of these stages we have a quadratic inequality which may or may not have a solution. If it does not have a solution there cannot then be a solution to (19) and r is then a lower bound to the minimax risk.

For this scheme there are really only two parameters, the value r and the number of grid points. For a very fine grid the above scheme is very closely connected to the information inequality method. In fact in terms of numerical solutions the application of the inequality method also involves taking a grid of values. However when finding a solution to that grid of values we only look from the point θ_i to the next point θ_{i+1} rather than to all point θ_j with $j > i$. This is the primary advantage of the method. In fact for small values of L it suffices to take $N = 2$ and consider just $\theta = 0$ and $\theta = L$. Taking a finer grid does not improve the lower bound at least when $L \leq 0.5$. In this case an exact answer can be given since we need only find the smallest r for which the quadratic equation

$$(B(L) + L)^2 = (r - B^2(L))(\exp(L^2) - 1)$$

has a solution. It is easy to check that the smallest r is $r = L^2 \exp(-L^2)$ which in the case of $L = 0.5$ yields the bound 0.1947. For $L = 1$ the bound from this approach no longer gives the best lower bound. For the two endpoint it only yields 0.368 and as can be seen in the table it is possible to improve this bound by considering a finer grid of values. As L grows it is important to also increase the number of points. It should be clear from the following table that the major improvement over the incormation inequality is when L is small. In this example once $L = 10$ the methods essentially yield the same results.

4.2. Binomial

We now turn to the problem of estimating the canonical exponential parameter $\theta = \ln(p/(1-p))$ when we observe X a binomial (n,p) random varaiable. As in [4]

TABLE 1
Bounds for the minimax risk when $X \sim N(\theta, 1)$ and $|\theta| \leq L$

L	Information inequality	Chi-square iterative	Minimax risk
0.5	0.126	0.194	0.199
1	0.300	0.403	0.450
2	0.547	0.617	0.645
3	0.688	0.733	0.751
5	0.829	0.848	0.857
10	0.937	0.938	0.945

we focus here on lower bounds for the minimax risk under a weighted squared error loss $L(\theta, a) = J(\theta)(\theta - a)^2$ where $J(\theta) = ne^\theta/(1+e^\theta)^2$ is the Fisher Information. Note that for X, binomial(n,p) the chi-square distance is given by

$$I^2(p_1, p_2) = \left(\frac{(1-p_2)^2}{(1-p_1)} + \frac{p_2^2}{p_1}\right)^n$$

and since $p = \frac{e^\theta}{1+e^\theta}$ it follows that

$$I^2(\theta_1, \theta_2) = \left(\frac{1+e^{\theta_1}}{(1+e^{\theta_2})^2}(1 + \frac{e^{2\theta_2}}{e^{\theta_1}})\right)^n.$$

It then follows that

$$B(\theta_2) - B(\theta_1) \leq \left(\frac{r}{J(\theta_1)} - B^2(\theta_1)\right)^{1/2} (I^2(\theta_1, \theta_2) - 1)^{1/2} - (\theta_2 - \theta_1) \tag{23}$$

for all $\theta_2 > \theta_1$. However since $I(\theta_1, \theta_2) = I(-\theta_1, -\theta_2)$ and $J(\theta) = J(-\theta)$ there is a solution to (23) if and only if there is a solution to if there is also a solution to

$$B(\phi) - B(\theta) \leq \left(\frac{r}{J(\theta)} - B^2(\theta)\right)^{1/2} (I^2(\theta, \phi) - 1)^{1/2} - (\phi - \theta) \tag{24}$$

for all $0 \leq \theta \leq \phi < \infty$ where $B(0) = 0$.

Solutions to such an equation only exist if r is sufficiently large. The smallest such r is thus a lower bound for the minimax risk. The following table gives a comparison of the bounds obtained via this method to the bounds given in [4] obtained using the information inequality method. Also included for comparison are bounds for the minimax risk based on lower bounds for the Bayes risk given in [3].

4.3. Poisson

Johnstone and MacGibbon [13] considered the problem of finding the minimax risk for estimating the mean of a Poisson random variable under a weighted squared

TABLE 2
Bounds for the minimax risk for binomial

n	Brown-Gajek	Information inequality	Chi-square iteration
1	0.1314	0.211	0.2208
3	0.3058	0.400	0.4097
5	0.4186	0.505	0.5156
10	0.5829	0.646	0.6587
25	0.7707	0.800	0.8135
100	0.9279	0.934	0.9403

TABLE 3
Bounds for the minimax risk for Poisson

m	Information inequality	Iterative Chi-square	Minimax value	Ratio
0.1	0.066	0.090	0.090	1.0
0.2	0.115	0.164	0.164	1.0
0.3	0.154	0.222	0.222	1.0
0.6	0.243	0.3295	0.333	1.011
0.9	0.307	0.389	0.396	1.018
1.2	0.358	0.435	0.447	1.0276
1.5	0.400	0.472	0.483	1.0233
1.8	0.435	0.503	0.515	1.0239
1.9	0.445	0.512	0.524	1.0234
2.0	0.455	0.521	0.533	1.0230
2.1	0.465	0.529	0.542	1.0246
2.2	0.474	0.537	0.550	1.0242
2.3	0.483	0.545	0.558	1.0239
2.4	0.491	0.552	0.565	1.0236
2.7	0.514	0.572	0.585	1.0227
3.0	0.535	0.590	0.603	1.0220
3.5	0.565	0.615	0.628	1.0211
4.0	0.591	0.638	0.650	1.0188
4.5	0.613	0.657	0.669	1.0183
5.0	0.633	0.673	0.685	1.0178
6.0	0.666	0.701	0.713	1.0171
7.0	0.693	0.724	0.735	1.0152
8.0	0.715	0.743	0.754	1.0148
10.0	0.751	0.773	0.783	1.013
11.5	0.771	0.790	0.801	1.014

error loss with weight equal to the Fisher information. The minimax risk was evaluated when the mean vector was assumed to be bounded by a given value which we shall write as m. Minimax values were given for m ranging from 0.1 to 11.5. In this case the bounds available from the iterative application of the chi-square bounds can be found as follows.

For X a Poisson random variable with mean λ it is easy to check that

$$I^2(\lambda_1, \lambda_2) = e^{\lambda_1 - 2\lambda_2} e^{\frac{\lambda_2{}^2}{\lambda_1}}$$

and that the Fisher information $J(\lambda) = \frac{1}{\lambda}$. Note that if

$$r \geq J(\lambda)(Var_\lambda(\delta) + B^2(\lambda))$$

for all $0 \leq \lambda \leq m$ it follows, since the Fisher information vanishes at zero that $B(0) = 0$. Bounds for the minimax risk can thus be found by seeking solutions to

$$B(\lambda_2) - B(\lambda_1) \leq (r\lambda_2 - B^2(\lambda_2))^{\frac{1}{2}} (e^{\lambda_2 - 2\lambda_1} e^{\frac{\lambda_1{}^2}{\lambda_2}} - 1)^{\frac{1}{2}} - (\lambda_2 - \lambda_1)$$

for all $0 \leq \lambda_1 < \lambda_2 \leq m$ under the assumption that $B(0) = 0$. In order to find a lower bound for the minimax risk we can as before only look at a grid of points. For small values of m only two points are needed and using just two points one at 0 and the other at m it is easy to check that a solution can exist only if

$$r \geq me^{-m}.$$

It turns out that this inequality is in fact sharp if $m \leq 0.3$. For larger values of m we need to put more points in the grid. Although the resulting lower bounds are not sharp a comparison with the minimax risks given by Johnstone and MacGibbon [13] shows that the ratio of the minimax risk to the lower bound is no larger than about 1.025. Detailed results are given in Table 3 where lower bounds are also given based on the information inequality method.

References

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

[2] Brown, L. D. and Farrell, R. H. (1990). A lower bound for the risk in estimating the value of a probability density. *J. Amer. Statist. Assoc.* **85** 1147–1153.

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

[4] Brown, L. D. and Low, M. G. (1991). Information inequality bounds on the minimax risk (with an application to nonparametric regression. *Ann. Statist.* **19** 329–337.

[5] Brown, L. D. and Low, M. G. (1996). A constrained risk inequality with applications to nonparametric function estimation. *Ann. Statist.* **24** 2524–2535.

[6] Cai, T. and Low, M. G. (2005). On adaptive estimation of linear functionals. *Ann. Statist.* **33** 2311–2343.

[7] Cai, L., Low, M. G. and Zhao, L. (2007). Tradeoffs between global and local risks in nonparametric function estimation. *Bernoulli* **13** 1–19.

[8] Casella, G. and Strawderman, W. E. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** 870–878.

[9] Chapman, D. G. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions. *Ann. Math. Statist.* **22** 581–586.

[10] Donoho, D. L. Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270.

[11] Donoho, D. L., Liu, R. C. and Macgibbon, B. (1990). Minimax risks over hyperrectangles and implications. *Ann. Statist.* **18** 1416–1437.

[12] Hammersley, J. M. (1950). On estimating restricted parameters. *J. Roy. Statist. Soc., Ser. B* **12** 192–229.

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

[14] Kiefer, J. (1952). On minimum variance estimators. *Ann. Math. Statist.* **23** 627–629.

[15] Kuks, J. (1983). An approximately minimax estimator of the mean of a one-dimensional normal distribution. *Izv. Akad. Nauk Estonskoj SSR* **32** 286–293. (In Russian).