Minimax lower bounds

Olivia Gradenwitz
David Bernhard
Seminar about Statistics SS07

May 1, 2007

1 Minimax lower bounds

1.1 Abstract

For a functional T and a density function g, we want to find a way of describing the difficulty in estimating Tg. This we do through comparing the expected value of the loss we would have, when we only know the estimation of Tg at n fixed points. By finding a lower bound of the *minimax risk*, we can measure in some sense how hard this problem is.

1.2 Definitions and Notations

- $\mathcal{G}$ density class on $(\mathcal{X}, \mathcal{B})$ measure space
- λ σ-finite measure on $(\mathcal{X}, \mathcal{B})$
- T functional defined on $\mathcal{G}$
- $t_n, (n \geqslant 1)$ sequence of measurable function where $t : \mathcal{X}^n \longrightarrow \mathbb{R}$ is an *estimation procedure*
- $T_n := t_n(X_1 ... X_n)$ random variable with X_i iid $\sim g \in \mathcal{G}$ is an *estimator* for T $\forall$ $(n \geqslant 1)$
- $R_l(n, t_n, g; T) := E_{g^{\otimes n}}[l(|t_n(X) - Tg|)]$ where
  - l is an *increasing loss function* on $[0, \infty)$
  - $g^{\otimes n}$ is the n-fold product of density associated with $g \in \mathcal{G}$
- We say $T_n^{(1)}$ *is better* than $T_n^{(2)}$ at a point $g :\Longleftrightarrow$
  $R_l(n, t_n^{(1)}, g; T) < R_l(n, t_n^{(2)}, g; T)$
- We define by
  $$MR_l(n, t_n, g; T) := \sup_{g \in \mathcal{G}} R_l(n, t_n, g; T)$$
  the maximal risk of t_n over $\mathcal{G}$

- We say $T_n^{(1)}$ *is better* than $T_n^{(2)}$ in general $:\Longleftrightarrow$ $MR_l(n, t_n^{(1)}, g; T) < MR_l(n, t_n^{(2)}, g; T)$

- We define by

$$MMR_l(n, t_n, g; T) := \inf_{t_n} MR_l(n, t_n, g; T) = \inf_{t_n} \sup_{g \in \mathcal{G}} R_l(n, t_n, g; T)$$

the global minimax risk for estimation Tg based on n samples.

  - Note that if $MMR_l(n, t_n, g; T)$ is finite for some $n_0 \geqslant 1$ then $(MMR_l(n, t_n, g; T))_{n=n_0}^{\infty}$ is a decreasing sequence of positive numbers usually $\downarrow 0$ for $n \to \infty$
    If for $n \to \infty$ $MMR_l(n, t_n, g; T) = \Theta(\delta_n)$ we say that Tg is δ_n-estimable with loss function l. δ_n is called the *rate of convergence*, and is a measure for the ill-posedness of the *estimation problem* (EP).
  - Note that for $\mathcal{G}_n \subset \mathcal{G}$, $MMR_l(n; T, \mathcal{G}_n) \leqslant MMR_l(n; T, \mathcal{G})$. Where $MMR_l(n; T, \mathcal{G})$ is called the *global minimax risk*

- For f, h probability densities on $(\Omega, \mathcal{A})$ and σ-finite measure μ. The *Hellinger distance* $\mathcal{H}(f, h)$ between f and h is defined as the square root of

$$\mathcal{H}^2(f, h) = \frac{1}{2} \int_{\mathcal{X}} (\sqrt{f(x)} - \sqrt{h(x)})^2 d\mu(x) = 1 - \int_{\mathcal{X}} \sqrt{(f(x)h(x)} d\mu(x)$$

1.3 A minimax lower bound Theorem

Lemma 1.1. *Let f, h probability densities on $(\Omega, \mathcal{A})$ with respect to a dominating measure μ. Then*

$$(1 - \mathcal{H}^2(f, h))^2 \leqslant 1 - (1 - (\int f \wedge h d\mu)^2 \leqslant 2 \int f \wedge h d\mu. \qquad (1)$$

Theorem 1.2. *Let $\mathcal{G}, \mathcal{X}, T$ and $g \in \mathcal{G}$ be defined as above. Let (g_n) be a sequence of densities on $\mathcal{G}$ such that*

$$\limsup_{n \to \infty} \sqrt{n} \mathcal{H}(g_n, g) \leqslant \tau. \qquad (2)$$

Then

$$\liminf_{n \to \infty} |Tg_n - Tg|^{-1} MMR_1(n; T, \{g, g_n\}) \geqslant \frac{1}{4} e^{-2\tau^2}. \qquad (3)$$

Where the loss function $l(x) = |x|$ is denoted in the corresponding risk as MMR_1.

Considering the modulus of continuity of T over $\mathcal{G}$ locally at g, with respect to the Hellinger metric:

$$m(\epsilon) = \sup\{|Th - Tg| : h \in \mathcal{G} \text{ and } \mathcal{H}(h, g) \leqslant \epsilon\} \qquad (4)$$

we get the following two corollaries:

Corollary 1.3. *Let $\mathcal{G}$ be a class of densities on $\mathcal{X}$ and T a functional on $\mathcal{G}$. Fix $g \in \mathcal{G}$, and let the function m be defined as in (4). Then for each subset $\mathcal{G}_g$ of $\mathcal{G}$ containing some Hellinger ball around g'*

$$\liminf_{n\to\infty} m(\tau/\sqrt{n})^{-1} MMR_1(n; T, \mathcal{G}_g) \geqslant \frac{1}{4} e^{-2\tau^2} \tag{5}$$

for each positive τ.

Corollary 1.4. *Let $\mathcal{G}$ be a class of densities on $\mathcal{X}$ and T a functional on $\mathcal{G}$. Fix $g \in \mathcal{G}$, and let the function m be defined as in (4), allowing an asymptotic expansion of m*

$$m(\epsilon) = (c\epsilon)^r(1 - o(1)) \text{ as } \epsilon \downarrow 0$$

for some positive parameters c and r. Then for each subset $\mathcal{G}_g$ in $\mathcal{G}$ containing some Hellinger ball around g,

$$\liminf_{n\to\infty} n^{\frac{r}{2}} MMR_1(n; T, \mathcal{G}_g) \geqslant \frac{1}{4}(\frac{1}{2}c\sqrt{r})^r e^{-\frac{r}{2}}. \tag{6}$$

2 The Van Trees Inequality

2.1 Abstract

Another way to measure the minimax risk of estimating a quantity is to consider the information content of a random variable, that is how much information about the unknown θ is contained in an X distributed according to g_θ. This leads to the concept of *Fisher information*. With the intuition that no estimator can get more about θ out of a sample than it contains, the *Van Trees inequality* gives a lower bound.

2.2 Problem

Let $\mathcal{G}$ be a *convex* set of densities on a sample space Ω and $g \in \mathcal{G}, g_n \in \mathcal{G}$ with the conditions

$$(i) \limsup_{n\to\infty} \sqrt{n}\mathcal{H}(g_n, g) \leq \tau < \infty$$

$$(ii) \{x : g_n(x) > 0\} \subseteq \{x : g(x) > 0\} \text{ for large } n$$

$$(iii) \sup_{g(x)>0} \left| \frac{g_n(x) - g(x)}{g(x)} \right| \to 0 \text{ as } n \to \infty$$

Define

$$g_\theta := g + \theta(g_n - g), \mathcal{G}_\theta := \{g_\theta : \theta \in [0, 1]\}$$

I will only consider the case where T is a *linear* functional on $\mathcal{G}$ and the loss function is the L^2 loss $l(x) = x^2$. Further assume θ is distributed according to π which has density function $\lambda(\theta)$ absolutely continuous to the Lebesgue measure on $[0, 1]$ and $\lambda(0) = \lambda(1) = 0$. t_n is an estimator for T.

2.3 Fisher Information

We define the Fisher information $I(\theta)$ as

$$I(\theta) := E_\theta\left[\left(\frac{\partial}{\partial\theta}\log g_\theta(x)\right)^2\right] = \int_{\mathbb{R}} \frac{\left(\frac{\partial}{\partial\theta}g_\theta(x)\right)^2}{g_\theta(x)}dx$$

and

$$\tilde{I}(\lambda) := E\left[\left(\frac{\partial}{\partial\theta}\log\lambda(\theta)\right)^2\right] = \int_0^1 \frac{\left(\frac{\partial}{\partial\theta}\lambda(\theta)\right)^2}{\lambda(\theta)}d\theta$$

2.4 Theorem

(Van Trees) [2] For any estimator $\hat{t}$ absolutely continuous w.r.t. θ we have

$$E\left[\left(\hat{t}(X) - t(\theta)\right)^2\right] \geq \frac{\left(E\left[\frac{\partial}{\partial\theta}t(\theta)\right]\right)^2}{E[I(\theta)] + \tilde{I}(\lambda)}$$

Note that this holds under more general conditions than I will cover here, in particular t need not be linear.

2.5 Van Trees Minimax bound

Inserting the definition of MMR_2 and the Fisher information we get

$$MMR_2(n; T_n, \mathcal{G}_n) \geq \frac{\left(\int_0^1 \frac{\partial}{\partial\theta}Tg_\theta\lambda(\theta)d\theta\right)^2}{n\int_0^1 I(\theta)\lambda(\theta)d\theta + \tilde{I}(\lambda)}$$

Some calculation gives us the following inequality, where (without proof) $\tilde{I}(\lambda)$ takes a minimum of $4\pi^2$ (this is $\pi = 3.14...$, not the probability associated with λ):

$$\liminf_{n\to\infty}\left(T(g_n - g)\right)^2\left(8\tau^2 + 4\pi^2\right)MMR_2 \geq 1$$

Compare this with the constant $16e^{4\tau^2}$ of Theorem 1.2.

2.6 Example: Exponential distribution

Let $\mathcal{G} := \{g_\theta : g_\theta(x) = \frac{e^{-\frac{x}{\theta}}}{\theta}\}$ be a parametrized family of exponential distributions on $[0,\infty)$ and $\theta > 0$. Our aim is to estimate $\theta = \int x \cdot g_\theta(x)dx$.
We can calculate the Hellinger distance

$$\mathcal{H}(g_\theta, g_\nu) = \frac{\left|\sqrt{\theta} - \sqrt{\nu}\right|}{\sqrt{\theta + \nu}}$$

The modulus of continuity evaluates to

$$m(\epsilon) = sup\{|\delta| : \mathcal{H}(\theta, \theta+\delta) \leq \epsilon\} \approx 2\sqrt{2}\theta\epsilon$$

Using the Corollary 1.4 , we get (e is Euler's constant)

$$\liminf_{n\to\infty} \sqrt{n} MMR_1 \geq \frac{\theta}{4}\sqrt{\frac{2}{e}}$$

For the standard estimator $\hat{\theta}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ the constant becomes $\theta\sqrt{\frac{2}{\pi}}$. So this estimator is not "optimal" for L_1 loss.

References

[1] *Inverse Problems in Statistics* AIO-course 1999 Geurt Jongbloed, Vrije Universiteit

[2] *Applications of the Van Trees inequality: A Bayesian Cramer-Rao bound* Gill and Levit March 2001