Final Exam

Mathematical Statistics, Fall 2014
J. Schmidt-Hieber

January 19, 2015

In total there are five problems. For each of them you can get three points.

1 General understanding

1. The mean square error (MSE) can be decomposed into squared bias + variance. Answer the following questions by giving a short explanation (one or two sentences).

   (a) [1 point] The bias is sometimes called deterministic error and the variance stochastic error. Why?

   (b) [1 point] Nonparametric problems often assume that the observations are independent. Is this assumption helpful to bound the bias or the variance?

   (c) [1 point] In nonparametric function estimation the parameter space is a class of functions with some regularity, for instance a Hölder class. Is the regularity assumption important for the bound of the bias or for the bound of the variance?

2. [3 points] A fundamental concept in nonparametric statistics are lower bounds for the minimax convergence rate. Describe briefly

the reduction of the minimax lower bound to a testing problem. Using this strategy, what type of conditions does one eventually need to prove in order to establish a lower bound?

3. Suppose we observe $\mathbf{Y}^{(n)} = (Y_1, \ldots, Y_n)$ with

$$Y_k = \theta_k + \epsilon_k, \quad k = 1, \ldots, n, \quad \epsilon_k \sim \mathcal{N}(0,1), \text{ i.i.d.} \tag{1.1}$$

Let $\Theta = \mathbb{R}^n$ and consider squared ℓ^2-loss $\ell(\theta, \theta') = \sum_{k=1}^n (\theta_k - \theta'_k)^2 = \|\theta - \theta'\|_2^2$. The risk of the James-Stein estimator is bounded from above by

$$n - \frac{(n-2)^2}{(n-2) + \|\theta\|_2^2}$$

and the risk for the hard thresholding estimator with universal threshold $\tau = \sqrt{2 \log n}$ is bounded from above by

$$8(s_\theta + 1) \log n$$

where s_θ denotes the number of non-zero components of the signal. Given these risk bounds, discuss which of the estimators leads to the better worst case control of the risk for large sample sizes if we additionally know that

(a) [1 point] θ has many small entries

(b) [1 point] θ is sparse (= few non-zero entries)

(c) [1 point] θ has at most $\sqrt{n}$ non-zero entries and all of them are bounded in absolute value by 1.

2 Mathematical problems

1. [3 points] Suppose, we observe n independent and identically distributed random variables $X_1, \ldots, X_n$, generated from a distribution with Lebesgue density f. For a kernel K of order 1, i.e.

$\int K(u)du = 1$ and $\int uK(u)du = 0$, consider the kernel density estimator

$$\widehat{f}_{nh}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{X_i - x}{h}\Big).$$

Show that

$$\int_{-\infty}^{\infty} x\widehat{f}_{nh}(x)dx = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Give an interpretation of the identity.

2. Let $(\phi_k)_{k=1,2,\ldots}$ be an $L^2[0,1]$ basis such that

$$\max_{k=1,2,\ldots}\ \sup_{x\in[0,1]} |\phi_k(x)| \leq \sqrt{2}$$

(this is for instance true for the trigonometric basis). Any function $f \in L^2[0,1]$ has an expansion $f = \sum_{k=1}^{\infty} f_k\phi_k$ with f_k the k-th series coefficients. Consider the sequence model, where we observe $\mathbf{Z}^{(n)} = (Z_1, Z_2, \ldots)$ with

$$Z_k = f_k + n^{-1/2}\epsilon_k, \quad k = 1, 2, \ldots$$

and ϵ_k an i.i.d. sequence of $\mathcal{N}(0,1)$ random variables. The parameter space is $\Theta = \{f = \sum_{k=1}^{\infty} f_k\phi_k \in L^2[0,1] : \sum_{k=1}^{\infty} k|f_k| \leq 1\}$. For a sequence of positive integers $(N_n)_n$ and a fixed $x_0 \in (0,1)$, we define the estimator

$$\widehat{f}_{N_n}(x_0) = \sum_{k=1}^{N_n} Z_k\phi_k(x_0).$$

(a) [1 point] Show that

$$\mathrm{Bias}^2\left(\widehat{f}_{N_n}(x_0)\right) \leq 2N_n^{-2}.$$

(b) [1 point] Show that

$$\operatorname{Var}\left(\widehat{f}_{N_n}(x_0)\right) \leq \frac{2N_n}{n}.$$

(c) [1 point] Show that the MSE of the estimator $\widehat{f}_{N_n}(x_0)$ with the optimal choice of N_n converges to zero with rate $n^{-2/3}$.

3. Recall that a prior is a probability distribution on the parameter space Θ.

(a) [2 points] Show that

$$\inf_{\widehat{\theta}} \sup_{\theta \in \Theta} E_\theta\left[\ell(\widehat{\theta}, \theta)\right] = \inf_{\widehat{\theta}} \sup_{\Pi \text{ prior}} \int_\Theta E_\theta\left[\ell(\widehat{\theta}, \theta)\right] d\Pi(\theta).$$

(b) [1 point] Show that this implies that the minimax risk is bounded from below by the Bayes risk.

4. [3 points] Suppose we observe $Z \sim \mathcal{N}(\theta, 1)$ with $\theta \in \Theta = \mathbb{R}$. Let Π be a prior on the parameter space Θ and consider squared loss $\ell(\theta, \theta') = (\theta - \theta')^2$. Write $\Pi(\cdot|Z)$ for the posterior density. Show that the Bayes estimator $\widehat{\theta}_B$ is given by the posterior mean

$$\widehat{\theta}_B = \int_{-\infty}^{\infty} \theta \ d\Pi(\theta|Z).$$

5. [3 points] Consider model (1.1) and let $\tau > 0$. Show that the soft thresholding estimator

$$\widehat{\theta}_k = \left(1 - \tfrac{\tau}{|Y_k|}\right)_+ Y_k, \quad k = 1, \ldots, n$$

is a solution to the minimization problem

$$\min_{\theta \in \mathbb{R}^n} \sum_{k=1}^n (Y_k - \theta_k)^2 + 2\tau|\theta_k|.$$