30100: Mathematical Statistics I **Winter 2022**

Homework 4

Lecturer: Chao Gao

1. In the class, I analyzed the Gaussian sequence model, which is an approximation to the original density estimation model. In this problem, please directly analyze the density estimation problem. Consider independent observations $X_1, ..., X_n \sim f$, where f is a density function on $[0, 1]$. The density function belongs to the following class

$$\mathcal{F}_\alpha(R) = \left\{ f(\cdot) = a_0 + \sum_{j\geq 1}(a_j \cos(2\pi j\cdot) + b_j \sin(2\pi j\cdot)) \in L^2[0,1] : \sum_j (2\pi j)^{2\alpha}(a_j^2 + b_j^2) \leq R^2, f \geq 0, \int f = 1 \right\}.$$

   Consider the following estimator

$$\hat{f}(x) = \hat{a}_0 + \sum_{j=1}^{k}(\hat{a}_j \cos(2\pi jx) + \hat{b}_j \sin(2\pi jx)),$$

   where $\hat{a}_j$'s and $\hat{b}_j$'s are empirical Fourier coefficients. Choose an appropriate k and show that

$$\sup_{f\in\mathcal{F}_\alpha(R)} \mathbb{E}_f \int_0^1 (\hat{f}(x) - f(x))^2 \leq Cn^{-\frac{2\alpha}{2\alpha+1}},$$

   for some constant C that only depends on α and R.

2. Consider the Gaussian sequence model $X_j \sim N(\theta_j^*, n^{-1})$ for $j \in \mathbb{N}$, and θ^* belongs to

$$\Theta_\alpha(R) = \left\{ \theta : \sum_j j^{2\alpha}\theta_j^2 \leq R^2 \right\}.$$

   We study a Bayesian procedure with the prior $\theta \sim \Pi = \bigotimes_j N(0, j^{-2\alpha-1})$.

   (a) Find the posterior mean $\hat{\theta} = \mathbb{E}(\theta|X)$.

   (b) Show $\sup_{\theta^*\in\Theta_\alpha(R)} \mathbb{E}_{\theta^*}\|\hat{\theta} - \theta^*\|^2 \leq Cn^{-\frac{2\alpha}{2\alpha+1}}$, for some constant C that only depends on α and R.

   (c) Find the prior probability of the parameter space $\Pi(\Theta_\alpha(R))$.

   (d) Find the posterior probability of the parameter space $\Pi(\Theta_\alpha(R)|X)$.

   (e) Are you surprised? What is going on? Discuss your understandings.

3. The blockwise James-Stein estimator not only achieves the optimal rate, but also leads to the sharp Pinsker constant. If you are interested, read Chapter 6 of Johnstone. Also try to finish reading Chapters 2, 4 and 5 whenever you have time.

4. If you are interested in asymptotic equivalence, read (`https://www.stat.uchicago.edu/~chaogao/stat366/lecture04.pdf`).

5. The optimal error of a two-point testing problem with i.i.d. data is $\inf_\phi(P^n\phi + Q^n(1-\phi))$, where P^n denotes the product distribution. Though the testing error can be related to the total variation between P^n and Q^n, it is not clear how fast the error converges to zero as n tends to infinity.

   (a) For any $\alpha \in (0,1)\cup(1,\infty)$, the Rényi divergence of order α is defined as $D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\int p^\alpha q^{1-\alpha}$, another information distance between two probability distributions. Show $D_\alpha(P\|Q) \geq 0$ via Jensen's inequality.

   (b) Define $C(P,Q) = -\min_{t\in(0,1)}\log\int p^{1-t}q^t$, known as the Chernoff information between P and Q. Prove $\inf_\phi(P^n\phi + Q^n(1-\phi)) \leq 2\exp\left(-nC(P,Q)\right)$. This implies the testing error goes to zero at least exponentially fast. Hint: apply Markov inequality after some rearrangement of the likelihood ratio test.

   (c) The Hellinger affinity is defined by $\int\sqrt{pq}$. Show $\int\min(p,q) \geq \frac{1}{2}\left(\int\sqrt{pq}\right)^2$. Hint: write $|p-q|$ as $|\sqrt{p}-\sqrt{q}||\sqrt{p}+\sqrt{q}|$ and apply Cauchy-Schwarz to $\int|p-q|$.

   (d) Show $\inf_\phi(P^n\phi + Q^n(1-\phi)) \geq \frac{1}{2}\exp\left(-nD_{1/2}(P\|Q)\right)$. In other words, the testing error cannot go to zero faster than the exponential rate.

   (e) Consider $P = N(\theta, I_p)$ and $Q = N(\eta, I_p)$. Find $D_{1/2}(P\|Q)$ and $C(P,Q)$.