Homework 4
Due Friday April 20 3:00 pm
Submit a pdf file on Canvas

1. Let $d \geq 2$, and let $X_1, \ldots, X_n \sim P$ where $X_i = (X_i(1), \ldots, X_i(d)) \in \mathbb{R}^d$. Assume that $P = N(\mu, I)$ where $\mu = (\mu(1), \ldots, \mu(d))$ and $I$ is the $d \times d$ identity matrix. Let

$$P = \{N(\mu, I) : \mu \in \mathbb{R}^d\}.$$ 

Let

$$R_n = \inf_{\hat{\mu}} \sup_{P \in P} \mathbb{E}_P ||\hat{\mu} - \mu||_\infty.$$ 

Show that

$$\sqrt{\frac{c \log d}{n}} \leq R_n \leq \sqrt{\frac{C \log d}{n}}$$

for some constants $0 < c \leq C < \infty$.

2. Suppose $d \leq n$, and let $\{p_\theta : \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^d$ be a parametric model. Suppose there exists $\theta_0 \in \Theta$ and $r > 0$ such that $\Theta_0 := \{\theta \in \mathbb{R}^d : ||\theta - \theta_0|| < r\}$ is contained in $\Theta$. Assume that the Fisher information matrix $I(\theta)$ is well-defined. Further, assume that there exist a fixed $\lambda_{\text{min}} > 0$ such that $\forall \theta \in \Theta_0$, $\lambda_{\text{max}} I_{d \times d} - I(\theta)$ is always positive definite, where $I_{d \times d}$ is an identity matrix. Further assume there exists $\lambda_{\text{max}} > 0$ such that $\forall \theta \in \Theta_0$, $\lambda_{\text{max}} I_{d \times d} - I(\theta)$ is always positive definite, and assume there exists $M > 0$ such that $\forall \theta \in \Theta_0$, $1 \leq i, j, k \leq d$, $\forall x$, $\left|\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \log p_\theta(x)\right| \leq M$. Let

$$R_n = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta ||\hat{\theta} - \theta||^2.$$ 

(a) Show that

$$R_n \geq \frac{Cd}{n}$$

for some $C > 0$.

(b) Recall that, under appropriate regularity conditions, the mle $\hat{\theta}$ is asymptotically Normal. Thus $\sqrt{n}(\hat{\theta} - \theta) \sim N(0, I_\theta^{-1})$ where $I_\theta$ is the Fisher information matrix. Assume, for simplicity, that $\sqrt{n}(\hat{\theta} - \theta) \sim N(0, I_\theta^{-1})$. Show that the mle achieves the minimax rate of convergence.

3. Let $Y = (Y_1, \ldots, Y_d) \sim N(\theta, I)$ where $\theta = (\theta_1, \ldots, \theta_d)$. Assume that $\theta \in \Theta = \{\theta \in \mathbb{R}^d : ||\theta||_0 \leq 1\}$. Let

$$R_d = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta ||\hat{\theta} - \theta||^2.$$ 

Show that $c \log d \leq R_d \leq C \log d$ for some constants $c$ and $C$. 

1
4. Let $X_1, \ldots, X_n \sim F$ where $F$ is some distribution on $\mathbb{R}$. Suppose we put a Dirichlet process prior on $F$:

$$F \sim DP(\alpha, F_0).$$

(a) Recall the stick-breaking construction. Show that $E(\sum_{j=1}^{\infty} W_j) = 1$.

(b) For any set $A$ write $F(A) \equiv \int_A dF(x)$. What is the prior for $F(A)$? What is the posterior for $F(A)$? Show that the posterior is consistent in the following sense: let $F_*$ be the true distribution (i.e. the data are generated from $F_*$). Fix any $\epsilon > 0$. Let

$$\pi_n = P\left(F_*(A) - \epsilon \leq F(A) \leq F_*(A) + \epsilon \bigg| X_1, \ldots, X_n\right).$$

From the frequentist point of view, the posterior probability $\pi_n$ is a random variable. Show that $\pi_n \overset{P}{\to} 1$.

5. In this question we consider a nonparametric Bayesian estimator and compare to the minimax estimator. For $i = 1, \ldots, n$ and $j = 1, 2, \ldots$ let

$$X_{ij} = \theta_j + \epsilon_{ij}$$

where all the $\epsilon'_{ij}$s are independent $N(0,1)$. The parameter is $\theta = (\theta_1, \theta_2, \ldots)$. Assume that $\sum_j \theta_j^2 < \infty$. Due to sufficiency, we can reduce the problem to the sample means. Thus let $Y_j = n^{-1} \sum_{i=1}^{n} X_{ij}$. So the model is $Y_j \sim N(\theta_j, 1/n)$ for $j = 1, 2, 3 \ldots$ We will put a prior $\pi$ on $\theta$ as follows. We take each $\theta_j$ to be independent and we take $\theta_j \sim N(0, \tau_j^2)$.

(a) Find the posterior for $\theta$. Find the posterior mean $\hat{\theta}$.

(b) Suppose that $\sum_j \tau_j^2 < \infty$. Show that $\hat{\theta}$ is consistent, that is, $||\hat{\theta} - \theta||^2 \overset{P}{\to} 0$.

(c) Now suppose that $\theta$ is in the Sobolev ball

$$\Theta = \left\{ \theta = (\theta_1, \theta_2, \ldots) : \sum_j j^{2p} \theta_j^2 \leq C^2 \right\}$$

where $p > 1/2$. The minimax (for squared error loss) for this problem is $R_n \asymp n^{-2p/(2p+1)}$. Let $\tau_j^2 = (1/j)^{2r}$ where $r = p + (1/2)$. Show that the posterior mean achieves the minimax rate.

This shows that the prior has to be chosen carefully to get minimax estimators.

6. Let $P$ be the uniform distribution on $[0, 1]$. Let $Q$ be the distribution that is uniform on the set $\{1/m, 2/m, \ldots, 1\}$.

(a) Find the total variation distance between $P$ and $Q$.

(b) Find $W_p(P, Q)$.

(c) Let $X_1, \ldots, X_n \sim P$. Let $P_n$ be the empirical distribution. Show that $W_1(P, P_n) \overset{P}{\to} 0$. 

2