Statistics 654 Comprehensive Written Exam

August, 2021

Instructions: Each question is weighted equally. Partial credit will be given for each part of a problem. In some cases, different parts of a problem can be done independently, so if you are unsure how to handle one part of a problem, don't hesitate to try the others. No answer should require a great deal of computation. Please show your work, and briefly explain your reasoning: correct answers without supporting work will not receive full credit. Please write clearly.

Note that this exam is internet free, closed notes, and closed book. You only need a pencil or pen, and paper.

1. Let $X \in \mathbb{R}$ be a random variable with $\mathbb{E}X = 0$ and $\mathrm{Var}(X) = 1$. Define $h(t) = \mathbb{E}[X\,\mathbb{I}(X \geq t)]$.

   a. Show that $h(t) = -\mathbb{E}[X\,\mathbb{I}(X < t)]$. Is $h(0) = 0$?

   b. Show that $h(t)$ is a probability density function.

2. Let P and Q be distributions on $\mathbb{R}$ with densities f and g, respectively.

   a. Define the total variation distance $\mathrm{TV}(P, Q)$ in terms of f and g.

   b. Define the Hellinger distance $\mathrm{H}(P, Q)$ in terms of f and g.

   c. Show that for $a, b \geq 0$ one has $|a - b| = a + b - 2\min(a, b)$ and $\sqrt{ab} \geq \min(a, b)$.

   d. Show that $\mathrm{TV}(P, Q)$ can be lower bounded by a multiple of $\mathrm{H}^2(P, Q)$.

3. Let M_n be the maximum of n iid $\mathcal{N}(0, 1)$ random variables.

   a. As carefully as you can, state the Gaussian extreme value theorem in terms of M_n. You need not prove anything.

   b. What can you say about the limiting value of $\mathbb{P}(M_n \geq \sqrt{2\log n})$ as $n \to \infty$? Carefully justify your answer.

4. Let $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$ be a family of densities on a set $\mathcal{X}$, and suppose that we are interested in estimating θ from an observation $X \in \mathcal{X}$ with $X \sim f(x|\theta) \in \mathcal{P}$.

   a. Define what is meant by an estimator and a loss function in this setting.

b. Define the risk function of an estimator.

c. Define the Bayes risk of an estimator under a prior distribution π on Θ.

d. Let $\mathcal{D}$ be a family of estimators. Define the minimax risk and optimal Bayes risk under π for the family $\mathcal{D}$.

e. Establish an inequality between the minimax risk and the Bayes risk.

5. Let $X = (X_1, \ldots, X_d)^t \in \mathbb{R}^d$ be a random vector with $\mathbb{E}X = 0$ and $\text{Var}(X) = \Sigma$, and let $Y \in \mathbb{R}$ be a random variable defined on the same probability space as X with $\mathbb{E}Y^2$ finite. Let

$$A = \{j : X_j \text{ is independent of } Y\},$$

and let $B = \{1, \ldots, d\} \setminus A$ be the complement of A. Assume that A and B are non-empty. Let $X_A = (X_j)_{j \in A}$ and $X_B = (X_j)_{j \in B}$ be the restriction of X to the components indexed by A and B, respectively. Now consider a solution

$$\beta^* \in \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \mathbb{E}(Y - \beta^t X)^2$$

of the standard linear regression problem. (Note that there may be more than one minimizing vector.) It is reasonable to expect that $\beta_A^* = (\beta_j^*)_{j \in A} = 0$, as the the variables X_A are independent of Y. Show that $\beta_A^* = 0$ if the entries of $\text{Cov}(X_A, X_B)$ are zero and an additional condition is satisfied; clearly state the additional condition you need.