30100: Mathematical Statistics I **Winter 2022**

Homework 3

Lecturer: Chao Gao

1. Read Chapters 4.1 & 4.2 of Lehmann & Casella, and answer the question: why are Bayes estimators biased?

2. Consider i.i.d. observations $X_1, ..., X_n \sim N(\theta, \sigma^2)$ with parameter space $(\theta, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$, where $\mathbb{R}^+ = (0, \infty)$. Show $\bar{X}$ is the minimax estimator for the loss $(\hat{\theta} - \theta)^2/\sigma^2$. In other words, it minimizes $\sup_{\theta \in \mathbb{R}, \sigma^2 > 0} \mathbb{E}_{\theta,\sigma^2} \frac{(\hat{\theta}-\theta)^2}{\sigma^2}$.

3. Consider $y \sim N(X\beta, \sigma^2 I_n)$ with $\beta \in \mathbb{R}^p$. For the loss $\|X(\hat{\beta} - \beta)\|^2$, show $\hat{\beta} = (X^T X)^- X^T y$ is minimax. Do not assume $(X^T X)$ is invertible. The notation $(X^T X)^-$ means the generalized inverse.

4. For $Z \sim N(0, I_p)$, show $\mathbb{E} Z^T g(Z) = \mathbb{E} \nabla^T g(Z)$.

5. For $Z \sim N(0, I_p)$ with $p \geq 3$, show $\mathbb{E} \frac{1}{\|Z+\mu\|^2}$ only depends on $\|\mu\|$ and is a decreasing function of $\|\mu\|$.

6. Consider $X_1, ..., X_n \sim N(\theta, \sigma^2 I_p)$ for $p \geq 3$. The parameter space is $\theta \in \mathbb{R}^p$ and the loss is $\|\hat{\theta} - \theta\|^2$. Let $\hat{\theta}_{\text{JS}+} = \left(1 - \frac{\sigma^2(p-2)}{n\|\bar{X}\|^2}\right)_+ \bar{X}$, where $x_+ = \max(x, 0)$. Prove $\mathbb{E}_\theta \|\hat{\theta}_{\text{JS}+} - \theta\|^2 < \mathbb{E}_\theta \|\hat{\theta}_{\text{JS}} - \theta\|^2$ for all $\theta \in \mathbb{R}^p$. Is $\hat{\theta}_{\text{JS}}$ admissible?

7. Is $\hat{\theta}_{\text{JS}+}$ admissible? Briefly explain why. Hint: Read Chapter 5.7 of Lehmann & Casella.

8. Does $\hat{\theta}_{\text{JS}}$ improves MLE for each coordinate? Without loss of generality, we can study this problem with $n = 1$, $\sigma^2 = 1$ and $p \geq 3$, and compare $\mathbb{E}(\hat{\theta}_{\text{JS},1} - \theta_1)^2$ and $\mathbb{E}(X_1 - \theta_1)^2$.

   (a) Show $\mathbb{E}(\hat{\theta}_{\text{JS},1} - \theta_1)^2 - \mathbb{E}(X_1 - \theta_1)^2 = (p-2)\mathbb{E}\left(\frac{(p+2)X_1^2 - 2\|X\|^2}{\|X\|^4}\right)$.

   (b) When $\theta_1 = \theta_2 = ... = \theta_p$, show $\mathbb{E}(\hat{\theta}_{\text{JS},1} - \theta_1)^2 - \mathbb{E}(X_1 - \theta_1)^2 < 0$.

   (c) When $\theta_2 = ... = \theta_p$, show as θ_1 moves away from θ_2, eventually $\mathbb{E}(\hat{\theta}_{\text{JS},1} - \theta_1)^2 - \mathbb{E}(X_1 - \theta_1)^2 > 0$.

   (d) Plot $\mathbb{E}(\hat{\theta}_{\text{JS},1} - \theta_1)^2$ and $\mathbb{E}(X_1 - \theta_1)^2$ as functions of θ_1, and illustrate your conclusions in (b) and (c). Discuss your new understanding of $\hat{\theta}_{\text{JS}}$.

9. Read the paper (https://projecteuclid.org/download/pdfview_1/euclid.ss/1331729980) and write a very brief summary. Do not exceed 0.5 pages.

10. If you want more intuition of the construction of the JS estimator, read (or do) Chapter 5, Problem 4.6 and Chapter 5, Problem 5.3 in Lehmann & Casella. You do not need to submit anything for this problem.

11. (***Extra credit***) Consider i.i.d. $X_1, \cdots, X_n \sim N(\theta, 1)$. You proved in the last homework that the Bayes estimator induced by the Gaussian prior $\theta \sim N(0, \tau^2)$ satisfies $\sup_{\theta \in \mathbb{R}} \mathbb{E}_\theta (\hat{\theta} - \theta)^2 = \infty$. This is because of the strong shrinkage effect of the Gaussian prior. In other words, a Gaussian distribution implies a very strong prior belief around its mean. Can we construct a Bayes estimator that satisfies $\sup_{\theta \in \mathbb{R}} \mathbb{E}_\theta (\hat{\theta} - \theta)^2 < \infty$? This can be achieved by using a prior distribution with a heavier tail. For example, consider a Laplace prior $\theta \sim \frac{1}{2\tau} \exp\left(-\frac{|\theta|}{\tau}\right)$. For the posterior mean induced by this prior, show that you can achieve

$$\sup_{\theta \in \mathbb{R}} \mathbb{E}_\theta (\hat{\theta} - \theta)^2 \leq \frac{1.00001}{n},$$

with some sufficiently large τ. An interesting fact is that any prior distribution with a heavier tail would also work (e.g. Cauchy). On the other hand, the Laplace is the lightest tail such that the induced Bayes estimator satisfies $\sup_{\theta \in \mathbb{R}} \mathbb{E}_\theta (\hat{\theta} - \theta)^2 < \infty$.