

Department of Statistics, The Chinese University of Hong Kong
STAT5010 Advanced Statistical Inference (Term 1, 2022–23)

Assignment 4 · due on 5 December 2022
 Please submit your answers in .pdf format via Blackboard.

1. (Hint for Problem 3 in Asg 5) Based on Theorem 3.5 in TPE, we first show the risk of the Bayes estimator η is given by

$$R[\eta, E(\eta|X)] = R[\eta, -\nabla \log h(X)] - \frac{2p}{\sigma^2 + \tau^2} + \sum_{i=1}^p E\left(\frac{X_i - \mu}{\sigma^2 + \tau^2}\right)^2,$$

since $m(x)$ in the normal hierarchical model is given by

$$m(x) = C(\sigma, \tau) \exp\left(-\frac{1}{2(\sigma^2 + \tau^2)} \sum_{i=1}^p (X_i - \mu)^2\right) \Rightarrow \frac{\partial}{\partial x_i} \log m(x) = -\frac{(x_i - \mu)}{\sigma^2 + \tau^2}.$$

As the normal hierarchical model has the following properties:

$$h(x) = \prod_{i=1}^p \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_i^2}{\sigma^2}\right) \right] \Rightarrow \frac{\partial}{\partial x_i} \log h(x) = -\frac{x_i}{\sigma^2},$$

$$R[\eta, -\nabla \log h(X)] = \dots = \frac{p}{\sigma^2},$$

which leads to

$$\begin{aligned} R[\eta, E(\eta|X)] &= \dots \\ &= \frac{p\tau^4}{\sigma^2(\sigma^2 + \tau^2)^2} + \left(\frac{\sigma^2}{\sigma^2 + \tau^2}\right)^2 \sum a_i^2, \end{aligned}$$

where $a_i = \eta_i - \mu/\sigma^2$. Finally, since the risk is given by

$$R[\eta, E(\eta|X)] = \frac{p\tau^4}{\sigma^2(\sigma^2 + \tau^2)^2} + \left(\frac{\sigma^2}{\sigma^2 + \tau^2}\right)^2 k$$

when $\sum a_i^2 = k$. Try to find the expression of η_i that solves $\sum a_i^2 = k$.

2. (K12.11) Laplace's law of succession gives a distribution for Bernoulli variables X_1, X_2, \dots in which $\Pr(X_1 = 1) = 1/2$, and

$$\Pr(X_{j+1} = 1 \mid X_1 = x_1, \dots, X_j = x_j) = \frac{1 + x_1 + \dots + x_j}{j + 2}, \quad j \geq 1.$$

Consider testing the hypothesis H_1 that X_1, \dots, X_n have this distribution against the null hypothesis H_0 that the variables are iid with $\Pr(X_i = 1) = 1/2$. If $n = 10$, find the best test with size $\alpha = 0.05$. What is the power of this test?

3. (K12.17, p -values) Suppose we have a family of tests φ_α , $\alpha \in (0, 1)$ indexed by level (so φ_α has level α), and that these tests are “nested” in the sense that $\varphi_\alpha(x)$ is nondecreasing as a function of α . We can then define the “ p -value” or “attained significance” for observed data x as $\inf\{\alpha : \varphi_\alpha(x) = 1\}$, thought of as the smallest value for α where test φ_α rejects H_0 . Suppose we are testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ and that the densities for data X have monotone likelihood ratios in T . Further suppose T has continuous distribution.

- (a) Show that the family of uniformly most powerful tests are nested in the sense described.
 (b) Show that if $X = x$ is observed, the p -values $P(x)$ is

$$\Pr_{\theta_0}(T(X) > t),$$

where $t = T(x)$ is the observed value of T .

- (c) Determine the distribution of the p -value $P(X)$ when $\theta = \theta_0$.
4. (K12.19) Suppose X has a Poisson distribution with parameter λ . Determine the uniformly most powerful test of $H_0 : \lambda \leq 1$ versus $H_1 : \lambda > 1$ with level $\alpha = 0.05$.
5. (K12.22) Suppose we observe a single observation X from $N(\theta, \theta^2)$.
- (a) Do the densities for X have monotone likelihood ratios?
 (b) Let ϕ^* be the best level α test of $H_0 : \theta = 1$ versus $H_1 : \theta = 2$. Is ϕ^* also the best level α test of $H_0 : \theta = 1$ versus $H_1 : \theta = 4$?
6. (K12.29) Suppose Y_1 and Y_2 are independent variables, both uniformly distributed on $(0, \theta)$, but our observation is $X = Y_1 + Y_2$.
- (a) Show that the densities for X have monotone likelihood ratios.
 (b) Find the UMP level α test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ based on X .

7. Let the variables X_i , $1 \leq i \leq n$ be independently distributed with distribution $Poisson(\lambda_i)$, $1 \leq i \leq n$ respectively. For testing the hypothesis

$$H_0 : \sum_{i=1}^n \lambda_i \leq a \quad v.s. \quad H_1 : \sum_{i=1}^n \lambda_i > a.$$

(for example, that the combined radioactivity of a number of pieces of radioactive material does not exceed a), show that there exists a UMP test, which rejects when $\sum_{i=1}^n X_i > C$.

8. (Optional) Suppose we wish to test n hypotheses H_1, H_2, \dots, H_n . And we assume that the null p -values are uniformly distributed. In this problem, we are interested in procedures which operate in two steps:

- **Step 1.** Select a set $\mathcal{S} \subset \{1, 2, \dots, n\}$ of 'promising' hypotheses.
- **Step 2.** Apply a multiple testing procedure to test those hypotheses in \mathcal{S} , namely, $\{H_i\}_{i \in \mathcal{S}}$.

Below we shall assume that the selection step is monotone in the following sense: if $\mathcal{S}(p)$ is the set of selected hypotheses on the basis of the n p-values (p_1, \dots, p_n) , then $p_i \leq p'_i$ for all i ($p \leq p'$ for short) implies that $\mathcal{S}(p') \subset \mathcal{S}(p)$.

- Suppose we apply the Benjamini-Hochberg (BH) procedure to the selected set of hypotheses with an FDR target level set to q (this means that the critical thresholds would be equal to $q_i/|\mathcal{S}|$ for $i = 1, 2, \dots, |\mathcal{S}|$). Under independence of all n p-values, would you expect FDR control at level q ? Explain why or why not. Similarly, imagine you were to apply the Bonferroni correction at level $\alpha/|\mathcal{S}|$, would you expect FWER control at level α ?
- Suppose now that you apply the BH procedure to the selected hypotheses with an FDR target set to $q|\mathcal{S}|/n$. Under independence between all the p-values, show that this two-step procedure would control the FDR at level q .
Hint: You may use the following claim: whenever a function $f : (p_1, \dots, p_n) \rightarrow [0, 1]$ is nonincreasing (recall that this means that $p \leq p'$ implies $f(p) \geq f(p')$), we have

$$\mathbb{E} \left[\frac{I_{\{p_i < f(p)\}}}{f(p)} \right] \leq 1,$$

provided the p-values obey the PRDS property.

- Suppose then that the n p-values actually obey the PRDS property, would FDR control at level q continue to hold? Explain why or why not.
- Under independence between the p-values, can I set a nominal threshold higher than $q|\mathcal{S}|/n$ and expect FDR control in general? Explain why or why not.
- Describe an application where it might make sense to use the two-step procedure we have just described.
- Prove the claim from the hint.