Statistics (Ph. D. Version)

Instructions to the Student

a. Answer all six questions. Each will be graded from 0 to 10.

b. Use a different booklet for each question. Write the problem number and your code number (NOT YOUR NAME) on the outside cover.

c. Keep scratch work on separate pages in the same booklet.

d. If you use a “well known” theorem in your solution to any problem, it is your responsibility to make clear which theorem you are using and to justify its use.

1. Consider data $X = (X_1, \ldots, X_{20})$, where $X_i \sim \mathcal{N}(\theta, 1)$ are iid, and the parameter $\theta \in \Theta = \mathbb{R}$ follows the standard normal prior density $\pi(\theta) = \phi(\theta)$. Find a statistical procedure $a = a(X) \in A \equiv \{0, 1\}$ (the ‘Bayes action’) for which the expected loss $L(\theta, a(X))$ is minimized, with expectations taken jointly over $(\theta, X)$ and where the loss function has the form

$$L(\theta, 1) \equiv 2 \text{ if } \theta \in \Theta_0 = (-\infty, 0] \text{, } L(\theta, 0) \equiv 1 \text{ if } \theta \in \Theta_1 = (0, \infty)$$

and $L(\theta, a) = 0$ otherwise. (That is, losses occur only when $\theta \in \Theta_0$ and you choose the action $a = 1$ to “Reject”, or when $\theta \in \Theta_1$ and your action is $a = 0$ to “Accept”.) Your formula for the Bayes action should be closed-form except that it may involve the standard normal distribution function or its quantiles.

2. Let $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ be independent samples from $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ respectively, with $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ as unknown parameters.

(a). Develop the size $\alpha$ Likelihood Ratio test of $H_0 : \sigma_1^2 = \sigma_2^2$, and express the constant rejection threshold in terms of (the quantiles of) a well known distribution.
(b). State the definition of the consistency of a test and show that the Likelihood Ratio test in (a) is consistent.

3. Suppose that an $n$-dimensional Gaussian random vector $X \sim \mathcal{N}(\mu, \sigma^2 A)$ is observed, where $\mu, \sigma^2$ are unknown, the vector $\mathbf{1}$ has all entries equal to 1, and the $n \times n$ matrix $A$ has diagonal entries $A_{ii} = 1$ and all off-diagonal entries $A_{ij}$ for $i \neq j$ equal to 0.1.

(a). Let $U$ be a matrix with first column $n^{-1/2} \mathbf{1}$ whose columns form an orthonormal basis of $\mathbb{R}^n$. Find the distribution of $W = U'X$.

(b). Find a pivotal quantity from the transformed vector $W$ in (a) from which you can define explicitly a 2-sided 95% confidence interval for $\mu$ that makes use of all the data $W$.

4. Two independent samples are drawn from Exponential populations, respectively $X = (X_1, \ldots, X_n) \sim \text{Expon}(\lambda)$ with density $\lambda e^{-\lambda x}$ for $x > 0$ and $Y = (Y_1, \ldots, Y_n) \sim \text{Expon}(1/\lambda)$ with density $\lambda^{-1} e^{-y/\lambda}$ for $y > 0$.

(a). Find the Maximum Likelihood Estimator $\hat{\lambda}$ of $\lambda$, and give its (centered and scaled) nondegenerate limiting distribution for large $n$.

(b). Compare the large-sample asymptotic variance of $\hat{\lambda}$ in (a) with the large-sample asymptotic variance of the minimum mean-squared error estimator of $\lambda$ of the form $a/\bar{X} + (1-a)\bar{Y}$ for constant $a \in [0, 1]$.

5. Let $Y = (Y_1, \ldots, Y_n)$ with $n \geq 2$ denote a vector of nonnegative random variables with unknown parameter $\vartheta = (\beta, j) \in (0, \infty) \times \{0, 1\}$ such that $Y_i$ are iid Gamma$(2, \beta)$ (with density $\beta^2 ye^{-\beta y}$) when $j = 0$ and are iid Gamma$(3, \beta)$ when $j = 1$.

(a). Show that $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ is not sufficient for $\vartheta$.

(b). Find a two-dimensional minimal sufficient statistic for $\vartheta$ based on $\bar{Y}$.

(c). Show that the parameter $\vartheta$ of the data $Y$ is identifiable, in the sense that the mapping from parameter $\vartheta$ to the space of densities $f_Y(\cdot; \vartheta)$ is one-to-one.

6. A sample of 150 Multinomial$(n; p_1, p_2, p_3)$ observations $(M_1, M_2, M_3)$ yields values $M_1 = 40, M_2 = 60, M_3 = 50$.

(a). Find an approximate numerical expression involving the tail probabilities of a standard tabulated distribution for the p-value of a significance
test of the null hypothesis that the multinomial parameters \((p_1, p_2, p_3)\) are of the form \((\theta^2, (1 - \theta)\theta, 1 - \theta)\) for some \(\theta \in (0, 1)\).

(b). Explain how you could evaluate through a statistical simulation study the quality of the large-sample approximation you used in (a).