

Homework 1 – STAT 643

Due Friday, September 9 (in class)

1. Consider a probability space $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), P)$ where P is defined in terms of a measurable density $f(x, y) \geq 0$ defined by a product measure $\mu_1 \times \mu_2$ on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ such that

$$P(C) = \int_C f(x, y) d(\mu_1 \times \mu_2)(x, y), \quad C \in \mathcal{B}(\mathbb{R}^2).$$

Define the random vectors X and Y on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), P)$ such that $X(x, y) = x$ and $Y(x, y) = y$ for $(x, y) \in \mathbb{R}^2$ (i.e., X and Y are just first and second coordinate mappings, each taking $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). Let $\sigma(Y) = \{Y^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\} = \{\mathbb{R} \times A : A \in \mathcal{B}(\mathbb{R})\} \subset \mathcal{B}(\mathbb{R}^2)$ denote the σ -algebra generated by Y and let $f_Y(y) = \int_{\mathbb{R}} f(x, y) d\mu_1(x)$ for $y \in \mathbb{R}$. Fix $B \in \mathcal{B}(\mathbb{R})$ and define

$$h(x, y) = \begin{cases} \int_B f(x, y) d\mu_1(x) / f_Y(y) & \text{if } f_Y(y) > 0, \\ \Phi(B) & \text{otherwise,} \end{cases}$$

where $\Phi(\cdot)$ denotes the standard normal distribution.

Show that h is a version of the conditional probability $P(X \in B | Y) \equiv P(X \in B | \sigma(Y))$.

Note: By Fubini's theorem, one has that $\int_B f(x, y) d\mu_1(x)$ is $(\mathcal{B}(\mathbb{R}^2), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$ -measurable function of y (for any given $B \in \mathcal{B}(\mathbb{R})$).

2. Suppose that the distributions $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ are dominated by a σ -finite measure μ and that $\theta_0 \in \Theta$ is such that $f_{\theta_0} = \frac{dP_{\theta_0}}{d\mu} > 0$ a.e. μ . Consider the function of x and θ given by

$$\Delta(\theta, x) = \frac{f_\theta(x)}{f_{\theta_0}(x)}.$$

The random function of θ , $\Delta(\theta, X)$, can be thought of as a “statistic.” Argue that it is sufficient.

3. Suppose that $X = (X_1, X_2, \dots, X_n)$ has independent components, where each X_i is generated as follows. For independent random variables $W_i \sim N(\mu, 1)$ and $Z_i \sim \text{Poisson}(\mu)$, $X_i = W_i$ with probability p and $X_i = Z_i$ with probability $1 - p$. Suppose that $\mu \in [0, \infty)$. Use the factorization theorem and find low-dimensional sufficient statistics in the cases that

- (a) p is known to be $1/2$;
- (b) $p \in [0, 1]$ is unknown.

Note: In the first case, the parameter space is $\Theta = \{1/2\} \times [0, \infty)$, while in the second case it is $\Theta = [0, 1] \times [0, \infty)$.

4. Let X be a sample from $P \in \mathcal{P}$, where \mathcal{P} is a family of distributions on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ (where \mathcal{P} may not necessarily be dominated by a σ -finite measure μ). Show that if $T(X)$ is sufficient for \mathcal{P} and $T = \psi(S)$, where ψ is measurable and $S(X)$ is another statistic, then $S(X)$ is sufficient for \mathcal{P} .

Hint: Consider how conditional expectations work for sub- σ -algebras of different sizes $\mathcal{G}_1 \subset \mathcal{G}_2$

5. Suppose that Z is exponential with mean $1/\lambda > 0$, but that one only observes $X = ZI(Z \geq 1)$ (where $I(\cdot)$ is the indicator function).
- (a) Consider the measure μ on $\mathcal{X} = \{0\} \cup [1, \infty)$ consisting of a point mass of 1 at 0 plus the Lebesgue measure on $[1, \infty)$. Give a formula for the R-N derivative of P_λ^X with respect to μ on \mathcal{X} .
 - (b) Suppose that X_1, X_2, \dots, X_n are iid with the distribution P_λ^X . Find a two-dimensional sufficient statistic for this problem and argue that it is indeed sufficient.

(c) Argue carefully that your statistic from (b) is minimal sufficient.

6. Consider a family of distributions $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which is absolutely continuous with respect to the Lebesgue measure μ , where $f_\theta = \frac{dP_\theta}{d\mu} > 0$ for any $\theta \in \Theta$ a.e. μ . Let F_θ be the cdf of P_θ . For $d \in \mathbb{R}$, let $P_{\theta,d}$ have the density

$$f_{\theta,d}(x) = I(x > d) \frac{f_\theta(x)}{1 - F_\theta(d)}$$

with respect to the Lebesgue measure μ . Let $X = (X_1, \dots, X_n)$ and suppose that $T(X)$ is sufficient for θ in a model where X_1, X_2, \dots, X_n are iid P_θ

- (a) Prove or give a counter-example to that $[T(X), \min_{1 \leq i \leq n} X_i]$ is sufficient for (θ, d) .
 - (b) If $T(X)$ is minimal sufficient for θ , is $[T(X), \min_{1 \leq i \leq n} X_i]$ guaranteed to be minimal sufficient for (θ, d) ?
7. Suppose that X_1, X_2, \dots, X_n are iid P_θ for $\theta = (\theta_1, \theta_2) \in (0, 1) \times \{1, 2\}$, where $P_{(\theta_1, 1)}$ is the Poisson(θ_1) distribution and $P_{(\theta_2, 2)}$ is the Bernoulli(θ_2) distribution. Find a two-dimensional minimal sufficient statistic for θ (argue carefully for minimal sufficiency).
8. Suppose that X_1, X_2, \dots, X_n are iid $N(\gamma, \gamma^2)$ for $\gamma \in \mathbb{R}$. Find a minimal sufficient statistic and show that it is not complete.