STA 332/MATH 343: Homework 2

Duke University, Fall 2026

Due: September 15th, 11:59 PM on Gradescope

Probability and Data Reduction

1. Let $X \sim \mathsf{Bin}(m, p)$. Starting from the probability mass function, compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

2. Let $n \geq 4$, and suppose
$$T(\underline{X}) = \big(T_1(\underline{X}), T_2(\underline{X})\big)$$
is sufficient for a parameter θ. Determine which of the following statistics must be sufficient for θ. Justify each answer.
$$S_1(\underline{X}) = \big(T_1(\underline{X}), e^{T_2(\underline{X})}\big), \qquad S_2(\underline{X}) = \big(-T_1(\underline{X}), T_2(\underline{X}), X_4\big), \qquad S_3(\underline{X}) = e^{T_2(\underline{X})}.$$

3. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Poisson}(\theta)$, where $\theta > 0$, and let
$$T = \sum_{i=1}^{n} X_i.$$
   (a) Find the distribution of T. *Hint: use the result of problem 13(a) on HW1.*
   (b) Without using the factorization theorem, prove directly from the conditional definition that T is sufficient for θ.

Exponential Families, Ancillarity, and Minimal Sufficiency

4. Let $Z_1, \ldots, Z_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau)$, where $\tau > 0$ is the variance and $n \geq 2$, and suppose that we observe
$$X_i = Z_i^2, \qquad i = 1, \ldots, n.$$
   (a) Show that $X_i \sim \mathsf{Gamma}\big(\frac{1}{2}, \frac{1}{2\tau}\big)$.
   (b) Let f_τ be the pdf of the X_i. Show that $\{f_\tau(x) : \tau > 0\}$ is a scale family with scale parameter τ.
   (c) Use the likelihood-ratio criterion to prove that $T = \sum_{i=1}^n X_i$ is minimal sufficient for τ.
   (d) Show that
$$A = \frac{X_1}{\sum_{i=1}^n X_i}$$
is ancillary for τ.

5. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Gamma}(\alpha, \beta)$, where $\alpha > 0$ is known and $\beta > 0$ is the unknown rate. Thus,
$$f_\beta(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\{x>0\}}.$$

(a) Write f_β in the form
$$f_\beta(x) = g(x)\exp\{b(\beta)R(x) - c(\beta)\},$$
and identify g, R, b, and c.

(b) Identify the minimal sufficient statistic $R_n(\underline{x})$ and explain why it is minimal sufficient for β.

(c) Use the exponential-family moment identity to find $\mathbb{E}_\beta[X_1]$.

6. Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}(\theta, \theta+1)$, where $\theta \in \mathbb{R}$.

   (a) Show that this is a location family.

   (b) Show that the sample range
   $$A = X_{(n)} - X_{(1)}$$
   is ancillary for θ.

   (c) Find a sufficient statistic for θ whose dimension does not depend on n.

   (d) Prove that
   $$T(\underline{X}) = \big(X_{(1)}, X_{(n)}\big)$$
   is minimal sufficient for θ. Because likelihoods may be zero, compare the likelihood functions directly rather than dividing them.

7. Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Poisson}(\theta)$, where $\theta > 0$ and $n \geq 2$. Define
$$S(\underline{X}) = \sum_{i=1}^n X_i \qquad \text{and} \qquad V(\underline{X}) = \left(\sum_{i=1}^n X_i, X_1\right).$$

   (a) Show that V is sufficient for θ.

   (b) Show that V is not minimal sufficient by using the definition of minimal sufficiency and the fact that S is sufficient.

   (c) Confirm the conclusion in part (b) using the likelihood-ratio criterion.