UNIVERSITY OF TORONTO
Faculty of Arts and Science

STA261H1: Probability and Statistics II
Midterm 1
July 19, 2022

Instructor: Robert Zimmerman

- Do not open this test until you are told to begin.
- Midterm 1 is closed-book; no aids are allowed.
- There are five questions (worth a total of 25 points) on the midterm, plus one bonus question (worth five additional points). Take a quick scan through the questions first and prioritize your time accordingly.
- Show all of your work for full marks, and ensure your notation is legible, correct, and consistent with that used in the course.
- If you need to use a result from lecture, either refer to it by its name (if it is a named theorem), or briefly describe it.

Good luck!

1. (5 points) Answer the following.

   (a) (1 point) State the definition of a **sufficient statistic**.

   (b) (2 points) State the definition of a **UMVUE** (including what the abbreviation stands for).

   (c) (1 point) State the **factorization theorem**.

   (d) (1 point) State the **Rao-Blackwell theorem**.

2. (5 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bin}(k, \theta)$, where $\theta \in (0,1)$ and k is known.[1] Let $\tau(\theta) = \log(\theta)$.

(a) (2.5 points) Find the MLE of $\tau(\theta)$. You can skip the second derivative test.

(b) (2.5 points) Compute the Cramér-Rao Lower Bound for unbiased estimators of $\tau(\theta)$.

[1] Each X_i has pmf given by $f_\theta(x) = \binom{k}{x}\theta^x(1-\theta)^{k-x}$ and satisfies $\mathbb{E}_\theta[X_i] = k\theta$ and $\text{Var}_\theta(X_i) = k\theta(1-\theta)$.

3. (5 points) Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with pdf

$$f_\theta(x) = \frac{\log(\theta)}{\theta - 1}\theta^x, \quad x \in (0,1), \quad \theta > 1.$$

Let $T(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n X_i$.

(a) (2.5 points) Without referring to completeness, show that $T(\mathbf{X})$ is minimal sufficient for θ.

(b) (2.5 points) Show that $T(\mathbf{X})$ is a complete sufficient statistic θ.

4. (5 points) Answer each of the following questions by writing YES or NO (1 point), and justify your answer in *at most* three sentences (1.5 points).

   (a) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$, where f_θ is in a location family with location parameter $\theta \in \mathbb{R}$. Can the range statistic $R(\mathbf{X}) = X_{(n)} - X_{(1)}$ ever be a complete sufficient statistic for θ?

   (b) A non-statistician friend of yours collects a data sample $\mathbf{x}$ generated by some pdf f_θ and calculates $L(261 \mid \mathbf{x}) = 10000$. Thinking this number to be very high, your friend concludes that $\theta = 261$ must a very plausible value of the true parameter. Is this reasoning sound?

5. (5 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ be a random sample from an exponential family, where

$$f_\theta(x) = h(x) \cdot g(\theta) \cdot \exp\left(\sum_{j=1}^{k} w_j(\theta) \cdot T_j(x)\right).$$

Prove that

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i)\right)$$

is sufficient for θ.

6. (5 points) Let $\mathbf{X} \sim f_\theta$, and suppose $T(\mathbf{X})$ is an unbiased point estimator of $\tau(\theta)$ which is uncorrelated with *all* unbiased estimators of 0. Prove that $T(\mathbf{X})$ must be the UMVUE of $\tau(\theta)$.

   Hint: let $V(\mathbf{X})$ *be unbiased for* $\tau(\theta)$ *and write* $V(\mathbf{X}) = T(\mathbf{X}) + (V(\mathbf{X}) - T(\mathbf{X}))$.

7. (BONUS: 5 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}\,(0, \theta)$, where $\theta > 0$.[2] Prove that $T(\mathbf{X}) = X_{(n)}$ is a complete statistic.

[2]Each X_i has pdf given by $f_\theta(x) = \theta^{-1} \cdot \mathbb{1}_{0<x<\theta}$.