UNIVERSITY OF TORONTO
Faculty of Arts and Science

STA261H1: Probability and Statistics II
Final Examination
August 17, 2022

Instructor: Robert Zimmerman

- This test is worth 50% of the final grade.
- Do not open this test until you are told to begin.
- You are may use a one-sided handwritten (not printed) cheat sheet. No other aids are permitted.
- There are six questions (worth a total of 60 points) on the exam, plus one bonus question (worth 10 additional points). Take a quick scan through the questions first and prioritize your time accordingly. Do not attempt the bonus question until you are completely satisfied with your work on the remaining questions.
- Show all of your work for full marks, and ensure your notation is legible, correct, and consistent with that used in the course. Be sure to clearly distinguish between random variables and constants, and between vectors and scalars (you can write $\mathbf{X}_n$ as $\vec{X}_n$ and $\mathbf{x}_n$ as $\vec{x}_n$).
- 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. (10 points) Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with density

$$f_\theta(x) = \frac{1}{x \cdot \log(\theta)} \cdot \mathbb{1}_{1 \le x \le \theta}, \quad \theta > 1.$$

(a) (3 points) Show that $T(\mathbf{X}) = X_{(n)}$ is sufficient for θ.

(b) (3 points) Find the MLE of θ. Differentiation will fail you here.

(c) (3 points) Find the MOM estimator of θ using the information in the footnote.[1]

(d) (1 point) Explain in one sentence why the conditions of the Cramér-Rao Lower Bound fail for this model.

[1] When $x > r \ge 1$, the function $g(x) = (x - r)/\log(x)$ has a unique inverse given by $g^{-1}(y) = -y \cdot W\left(-\frac{e^{-r/y}}{y}\right)$, where $W(\cdot)$ is some nice function you don't need to know anything about (it's called the *Lambert W function*, if you want to look it up later).

2. (10 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Bernoulli (θ), where $\theta \in (0,1)$. Recall that the MLE of θ is given by $\hat{\theta}_n(\mathbf{X}_n) = \bar{X}_n$. Let $\tau(\theta) = 1/\theta$ and let $T_n = 1/\bar{X}_n$.

(a) (1 point) Explain in one sentence why T_n is the MLE of $\tau(\theta)$.

(b) (2 points) Show that $T_n \xrightarrow{p} 1/\theta$.

(c) (2 points) Does this mean that the asymptotic distribution of T_n is degenerate at $1/\theta$? Answer YES or NO and explain in one sentence.

(d) (4 points) By checking sufficient regularity conditions, confirm that T_n is a consistent and asymptotically efficient estimator of $1/\theta$. I'll give you $I_1(\theta) = \frac{1}{\theta(1-\theta)}$ for free.

(e) (1 point) Determine the asymptotic distribution of T_n.

3. (a) (5 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$, and suppose that regularity conditions hold which guarantee asymptotic normality of the MLE $\hat{\theta}_n$. Prove that if we test $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$ using the Wald statistic $W_n(\mathbf{X}_n) = (\hat{\theta}_n - \theta_0)^2 I_n(\hat{\theta}_n)$, then $W_n(\mathbf{X}_n) \xrightarrow{d} \chi^2_{(1)}$ under H_0. You may assume that $I_n(\theta)$ is a continuous function of θ.

(b) (5 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, where $\theta \in (0, 1)$. Construct an approximate $(1 - \alpha)$-confidence interval for θ based on the Wald statistic.

4. (10 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, where $\theta \in (0,1)$. Suppose we elicit a $\text{Unif}(a,b)$ prior to θ, where $0 < a < b < 1$, and suppose we observe $\sum_{i=1}^{n} x_i = n$.

   (a) (4 points) Calculate $\pi(\theta \mid \mathbf{x})$, including the normalizing constant.

   (b) (4 points) What effect has the observed data had on our original beliefs concerning the true value of θ? Would a frequentist point estimator of θ based on the MLE be compatible with the Bayesian viewpoint in this situation? Explain.

   (c) (2 points) Explain what is meant by **prior elicitation**.

5. (a) (5 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ and let $\pi(\theta)$ be a prior on θ. Prove that if $T(\mathbf{X})$ is sufficient for θ (in the frequentist sense), then the posterior densities of $\theta \mid \mathbf{x}$ and $\theta \mid T(\mathbf{x})$ are identical.

(b) (5 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ where $\mathbb{E}_\theta [X_i] = \theta$ and $\mathrm{Var}_\theta (X_i) = \sigma^2$ is known. Show that

$$\left(\bar{X}_n - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}} \right)$$

is an approximate $(1-\alpha)$-confidence interval for θ.

6. (a) (5 points) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ where f_θ is in an exponential family:

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

Prove that if we choose an exponential family prior of the form

$$\pi(\theta) \propto g(\theta)^\nu \cdot \exp\left(\sum_{j=1}^{k} w_j(\theta) \cdot \eta_j\right)$$

where ν and $\eta_1, \ldots, \eta_k$ are hyperparameters, then $\pi(\theta)$ is a conjugate prior for f_θ.

(b) (5 points) Let $\Theta_0 \subset \Theta$ and suppose that $H_0 : \theta \in \Theta_0$ and $H_A : \theta \in \Theta_0^c$ are two competing hypotheses about plausible values of θ. Show that $BF_{H_0} = 1/BF_{H_A}$, assuming $\Pi(\Theta_0) \in (0, 1)$.

7. (a) (BONUS: 5 points) Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with pdf f_θ and cdf F_θ, where n is odd. Suppose that $\mathbb{P}_\theta (X_i \leq \theta) = 1/2$, so that θ is the (unique) **median** of the distribution. Find the asymptotic distribution of the **sample median** $M_n(\mathbf{X}_n) = X_{\left(\frac{n+1}{2}\right)}$.

Hint: the probability integral transform might help.

(b) (BONUS: 5 points) Let $g(\cdot)$ and $h(\cdot)$ be smooth (i.e., infinitely-differentiable) functions on $\mathbb{R}$, and suppose that h is maximized at θ^*. The **Laplace approximation** says that

$$\int_{-\infty}^{\infty} g(\theta)e^{nh(\theta)}\,\mathrm{d}\theta \approx \sqrt{-\frac{2\pi}{nh''(\theta^*)}}g(\theta^*)\cdot e^{nh(\theta^*)},$$

and moreover, the approximation becomes exact as $n \to \infty$. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ and write $\hat{\theta}_n := \hat{\theta}_{\mathrm{MLE}}(\mathbf{x}_n)$. Under whatever regularity conditions you need, use the Laplace approximation to derive a version of the Bernstein-von Mises theorem: if $\theta_n \sim \pi(\theta \mid \mathbf{x}_n)$ and n is large, then θ_n is approximately distributed as

$$\mathcal{N}\left(\hat{\theta}_n, \frac{1}{J_n(\mathbf{x}_n)}\right).$$

Hint: if θ is close to $\hat{\theta}_n$, you can assume that a second-order Taylor expansion gives $e^{\ell(\theta|\mathbf{x}_n)-\ell(\hat{\theta}_n|\mathbf{x}_n)} \approx e^{-\frac{1}{2}(\theta-\hat{\theta}_n)^2\cdot J_n(\mathbf{x}_n)}$.