STA261 (Summer 2022) - Assignment 1

These problems are meant to test your understanding of the concepts in Module 1. They are *not* to be handed in. Some of these have been modified (or in some cases taken directly) from questions in the *Additional Resources* listed in the course syllabus, and no claims of originality are made.

---

1. Suppose $\mathcal{X} = \{1, 2, 3, 4\}$ and $\Theta = \{a, b\}$. Two mass functions on $\mathcal{X}$ – one for each value of $\theta \in \Theta$ – are specified in the following table:

	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$p_a(x)$	1/2	1/6	1/6	1/6
$p_b(x)$	1/4	1/4	1/4	1/4

   Suppose $X \sim p_\theta$, and let $T(x) = \begin{cases} 0, & x = 1 \\ 1, & x \in \{2, 3, 4\} \end{cases}$.

   Going from the definition, show that $T(X)$ is a sufficient statistic for θ. You can do this by working out a table of $\mathbb{P}_a(X = x \mid T(X) = t)$ for each $x \in \mathcal{X}$ and $t \in \{0, 1\}$, and then doing the same for $\mathbb{P}_b(X = x \mid T(X) = t)$.

2. Let $X \sim \mathcal{N}(0, \sigma^2)$, where $\sigma^2 > 0$. Prove that $T(X) = |X|$ is sufficient for σ^2.

3. Suppose $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$, where $\theta \in \Theta$ is unknown. Show that $T(\mathbf{X}) = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is sufficient for θ.

4. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}[0, \theta]$, where $\theta > 0$. Find a minimal sufficient statistic for θ.

5. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}[\theta, \theta + 1]$, where $\theta \in \mathbb{R}$. Find a minimal sufficient statistic for θ.

6. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}[\theta_1, \theta_2]$, where $\theta_1, \theta_2 \in \mathbb{R}$ and $\theta_1 < \theta_2$. Find a minimal sufficient statistic for $\theta = (\theta_1, \theta_2)$.

7. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Geo}(\theta)$, where $\theta \in (0, 1)$. Find a complete sufficient statistic for θ.

8. Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with density

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

   Show that $T(\mathbf{X}) = \prod_{i=1}^n X_i$ is a complete sufficient statistic for θ.

9. This will give you some practice dealing with multi-parameter exponential families (which is a hint!). Let $X_1, X_2, \ldots, X_n$ be a random sample from an inverse Gaussian distribution, which has density

$$f_{\mu,\lambda}(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x - \mu)^2}{2\mu^2 x}\right), \quad x \in \mathbb{R}, \quad \lambda > 0, \quad \mu \in \mathbb{R}.$$

   Show that

$$T(\mathbf{X}) = \left(\bar{X}_n, \sum_{i=1}^n X_i \cdot \sum_{i=1}^n \frac{1}{X_i}\right)$$

   is a complete sufficient statistic for $\theta := (\lambda, \mu)$.

10. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Beta$(\alpha, k\alpha)$ distribution with $k \in \mathbb{N}$ known, which has density

$$f_\alpha(x) = \frac{\Gamma((k+1)\alpha)}{\Gamma(\alpha)\Gamma(k\alpha)} x^{\alpha-1}(1-x)^{k\alpha-1}, \quad x \in (0,1), \quad \alpha > 0.$$

Find a complete sufficient statistic for α.

11. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Pareto(θ) distribution, which has density

$$f_\theta(x) = \frac{\theta x_0^\theta}{x^{\theta+1}}, \quad x \geq x_0, \quad \theta > 1.$$

Here $x_0 > 0$ is *known*. Find a complete sufficient statistic for θ.

12. Using our notation, the Evans/Rosenthal textbook initially defines a sufficient statistic like this:

> A function $T(\cdot)$ defined on the sample space $\mathcal{X}^n$ is a *sufficient statistic* for θ if the following holds: $T(\mathbf{x}) = T(\mathbf{y})$ implies $f_\theta(\mathbf{x}) = c(\mathbf{x}, \mathbf{y}) \cdot f_\theta(\mathbf{y})$, for some constant $c(\mathbf{x}, \mathbf{y}) > 0$.

It turns out that their definition is equivalent to ours, but proving that fact is fairly difficult.

(a) Instead, just prove that our definition implies theirs. You can stick to the discrete case. The trick is to observe that the event $\{\mathbf{X} = \mathbf{x}\}$ is a subset of the event $\{T(\mathbf{X}) = T(\mathbf{x})\}$.

(b) If we replace the word "implies" in the textbook definition with "if and only if", what can be said about $T(\cdot)$?

13. Prove that if a statistic $T(\mathbf{X})$ is complete for θ and r is one-to-one, then the statistic $r(T(\mathbf{X}))$ is also complete for θ.

14. Suppose $\mathcal{X} = \{0, 1, 2\}$ and $\Theta = (0, \frac{1}{2})$.

	$x = 0$	$x = 1$	$x = 2$
$p_\theta(x)$	θ	θ^2	$1 - \theta - \theta^2$

Let $X \sim p_\theta$. Prove that $T(X) = X$ is a complete sufficient statistic for θ using the same principle as Example 1.35.

15. Let $X_1, X_2, \ldots, X_n$ be a random sample from a scale family with parameter $\sigma > 0$. Prove that *any* function of the $n - 1$ ratios $X_1/X_n, \ldots, X_{n-1}/X_n$ must be ancillary for σ. Hint: $X_i/\sigma \sim F(x)$.

16. Let $X_1, X_2, \ldots, X_n$ be a random sample from a location family with parameter $\mu \in \mathbb{R}$. Prove that *any* function of the $n - 1$ differences $X_1 - X_n, X_2 - X_n, \ldots, X_{n-1} - X_n$ is ancillary for μ.

17. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Show that $R(\mathbf{X}) = (X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X})$ is ancillary for μ. Is $R(\mathbf{X})$ independent of $\bar{X}$? This example will be very important in Module 4.

18. We've not had any issues checking the "open set" condition of Theorem 1.8, but here's a famous example that shows you what can happen if it's not satisfied. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, \theta^2)$, where $\theta > 0$. One can show that the condition doesn't hold for this parameter space.

(a) Show that in this case, the statistic in the theorem is

$$T(\mathbf{X}) = \left(\sum_{i=1}^n X_i, -\frac{1}{2}\sum_{i=1}^n X_i^2\right).$$

(b) Show that $T(\mathbf{X})$ is a one-to-one function of $(\bar{X}_n, S^2)$.

(c) Show that $\mathbb{E}_\theta \left[\frac{n}{n+1}(\bar{X}_n)^2 - S^2\right] = 0$ for all $\theta > 0$.

(d) Clearly, it's not always true that $\frac{n}{n+1}(\bar{X}_n)^2 \neq S^2$ (try it with a few small numbers if you're skeptical). Explain why this implies that $T(\mathbf{X})$ can't be complete for θ.

19. Show that the following distributions are in exponential families, assuming all parameters are unknown unless otherwise specified. For each one, identify the parameter θ which could be a vector, and the functions $h(x)$, $g(\theta)$, $T_j(x)$, and $w_j(\theta)$ for each j (if there's more than one). Also decide if each one belongs to a location family, a scale family, or a location-scale family (or none).

(a)
$$f_{\mu,\sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right), \quad x > 0, \qquad \mu \in \mathbb{R}, \quad \sigma > 0.$$

(b)
$$f_{k,\lambda}(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}, \quad x > 0, \qquad k \in \mathbb{N}, \quad \lambda > 0.$$

(c)
$$f_\nu(x) = \frac{2^{-\nu/2} x^{-(\nu/2+1)}}{\Gamma(\nu/2)} e^{-\frac{1}{2x}}, \quad x > 0, \qquad \nu > 0.$$

(d)
$$p_\theta(x) = \binom{x+r-1}{x}(1-\theta)^r \theta^x, \quad x \in \{0, 1, 2, \ldots\}, \qquad r \text{ known}, \quad \theta \in [0, 1].$$

(e)
$$p_{\mathbf{p}}(\mathbf{x}) = \binom{n!}{x_1, \ldots, x_k} \prod_{i=1}^k p_i^{x_i}, \quad \mathbf{x} \in \{0, 1, \ldots, n\}^k \text{ s.t. } \sum_{i=1}^k x_i = n, \quad \mathbf{p} \in (0, 1)^k \text{ s.t. } \sum_{i=1}^k p_i = 1,$$

with n known.