ST 502 MIDTERM 1

October 7, 2019

NAME:

STUDENT ID:

- You have **75 minutes** to complete this exam.
- This is a **closed book, closed notes** exam. The use of a calculator or computer is NOT permitted.
- Please show all of your work. For this exam, the steps taken to arrive at a particular solution are more important than the final answer.

1. (3 points) Suppose that a sequence of five independent coin tosses are observed, and that the probability of observing tails on any given toss (as opposed to heads) is p. What is the probability of observing at least three consecutive heads in the sequence of 5 independent coin tosses?

   Solution:

$$3 \cdot (1-p)^3 p^2 + 4 \cdot (1-p)^4 p + (1-p)^5$$

2. The random variable $X \sim \text{Weibull}(\alpha, \beta)$ has the cumulative distribution function given by

$$F_X(x) = 1 - e^{-(x/\alpha)^\beta},$$

   for $x \geq 0$, $\alpha > 0$, and $\beta > 0$.

   (a) (3 points) Derive the density function of X.

   Solution: The density function of X is

$$f_X(x) = \frac{dF_X}{dx}(x) = \frac{\beta}{\alpha}(x/\alpha)^{\beta-1} e^{-(x/\alpha)^\beta}.$$

   (b) (3 points) Derive the density function of the random variable $W := (X/\alpha)^\beta$.

Solution:

$$F_W(w) = P(W \le w) = P(X \le \alpha w^{1/\beta}) = F_X(\alpha w^{1/\beta}) = 1 - e^{-w},$$

and so the density function of W is

$$f_W(w) = \frac{dF_W}{dw}(w) = e^{-w}.$$

(c) (3 points) Derive the moment generating function of W.

Solution:

$$M_W(t) = \mathrm{E}(e^{tW}) = \int_0^\infty e^{tw} e^{-w}\ dw = \int_0^\infty e^{(t-1)w}\ dw = \frac{-1}{1-t} \cdot e^{-(1-t)w}\bigg|_0^\infty = \frac{1}{1-t},$$

for $t < 1$.

3. Denote by X_n a sequence of random variables, and c a fixed real number. Suppose that X_n converges in probability to c as $n \to \infty$.

(a) (3 points) Precisely express what it means to say that $X_n \xrightarrow{\mathrm{P}} c$.

Solution: For every $\varepsilon > 0$,

$$P(|X_n - c| > \varepsilon) \longrightarrow 0,$$

as $n \to \infty$.

(b) (3 points) Recall from calculus that a function $g : \mathbb{R} \to \mathbb{R}$ is said to be *continuous* at the point $a \in \mathbb{R}$ if and only if for every $\varepsilon > 0$ there exits a $\delta > 0$ such that if $|x - a| < \delta$ then $|g(x) - g(a)| < \varepsilon$. Show that if g is a continuous function and $X_n \xrightarrow{\mathrm{P}} c$, then $g(X_n) \xrightarrow{\mathrm{P}} g(c)$. Hint: use the expression

$$\{|g(X_n) - g(c)| > \varepsilon\} = \{|g(X_n) - g(c)| > \varepsilon\} \cap \Big(\{|X_n - c| > \delta\} \cup \{|X_n - c| \le \delta\}\Big).$$

Solution: Fix an arbitrary $\varepsilon > 0$, and by the continuity of g, choose $\delta > 0$ such that if $|x - c| \le \delta$ then $|g(x) - g(c)| < \varepsilon$. Next, observe that

$$\begin{aligned}\{|g(X_n) - g(c)| > \varepsilon\} &= \{|g(X_n) - g(c)| > \varepsilon\} \cap \Big(\{|X_n - c| > \delta\} \cup \{|X_n - c| \le \delta\}\Big) \\ &= \Big(\{|g(X_n) - g(c)| > \varepsilon\} \cap \{|X_n - c| > \delta\}\Big) \cup \emptyset \\ &\subseteq \{|X_n - c| > \delta\}.\end{aligned}$$

Then,

$$P(|g(X_n) - g(c)| > \varepsilon) \le P(|X_n - c| > \delta) \longrightarrow 0$$

as $n \to \infty$ since $X_n \xrightarrow{\mathrm{P}} c$.

4. Consider a finite *population* of N hospitals in January 1968, with the proportion of hospitals having fewer than 1000 discharges is equal to some true unknown proportion p. That is, the *population* has the form $\{x_1, \dots, x_N\}$ with $x_i \in \{0, 1\}$ for $i \in \{1, \dots, N\}$. For a simple random sample (without replacement), $\{X_1, \dots, X_n\}$ for some $n < N$, an estimator of p is the sample proportion $\widehat{p} := \frac{1}{n}\sum_{i=1}^n X_i$.

(a) (3 points) Show that $\widehat{p}$ is an unbiased estimator of p.

Solution:

$$\mathrm{E}(\widehat{p}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\big[1\cdot p + 0\cdot(1-p)\big] = p.$$

(b) (3 points) Using a corollary from lecture, it follows that an unbiased estimate of $\mathrm{Var}(\widehat{p})$ is

$$\frac{\widehat{p}(1-\widehat{p})}{n-1}\left(1-\frac{n}{N}\right).$$

Use $\widehat{p}$, this estimate of its variance, and the normal approximation to derive a 95 percent confidence interval for the true proportion p. Recall that $\Phi^{-1}(.975) = 1.96$, where Φ refers to the standard normal CDF. Your final answer should be expressed as an interval.

Solution: Using the normal approximation,

$$Z := \frac{\widehat{p}-p}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}\left(1-\frac{n}{N}\right)}} \sim \mathrm{N}(0,1),$$

and so

$$\begin{aligned}
.95 &= .975 - .025\\
&= .975 - (1-.975)\\
&= \Phi(1.96) - (1-\Phi(1.96))\\
&= \Phi(1.96) - \Phi(-1.96)\\
&= P(-1.96 \le Z \le 1.96)\\
&= P\left(\widehat{p} - 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}\left(1-\frac{n}{N}\right)} \le p \le \widehat{p} + 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}\left(1-\frac{n}{N}\right)}\right).
\end{aligned}$$

Thus, a 95 percent confidence interval for the true proportion p is

$$\left[\widehat{p} - 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}\left(1-\frac{n}{N}\right)}, \widehat{p} + 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}\left(1-\frac{n}{N}\right)}\right].$$

5. Select the correct statement(s).

[**True**](1 point) A confidence interval is a random variable.

[**True**](1 point) Let (X_n, Y_n) be a 95 percent confidence interval for a parameter θ. Then the probability that $(X_n, Y_n) \ni \theta$ is .95.

[**False**](1 point) Let $(1.16, 4.76)$ be an observed 95 percent confidence interval for a parameter θ. Then the probability that $(1.16, 4.76) \ni \theta$ is .95.

[**False**](1 point) Let $(1.16, 4.76)$ be an observed 95 percent confidence interval for a parameter θ. If we sample 100 data sets of the same size, then we would expect that for approximately 95 of the data sets, $(1.16, 4.76) \ni \theta$.