Math 241 Spring 2015
Final Examination

Solutions

Professor Adam Kapelner

May 21, 2015

Full Name _____ Section (A or B) _____

Code of Academic Integrity

Since the college is an academic community, its fundamental purpose is the pursuit of knowledge. Essential to the success of this educational mission is a commitment to the principles of academic integrity. Every member of the college community is responsible for upholding the highest standards of honesty at all times. Students, as members of the community, are also responsible for adhering to the principles and spirit of the following Code of Academic Integrity.

Activities that have the effect or intention of interfering with education, pursuit of knowledge, or fair evaluation of a student's performance are prohibited. Examples of such activities include but are not limited to the following definitions:

Cheating Using or attempting to use unauthorized assistance, material, or study aids in examinations or other academic work or preventing, or attempting to prevent, another from using authorized assistance, material, or study aids. Example: using a cheat sheet in a quiz or exam, altering a graded exam and resubmitting it for a better grade, etc.

I acknowledge and agree to uphold this Code of Academic Integrity.

_____   _____
            signature   date

Instructions

This exam is 120 minutes and closed-book. You are allowed three pages (front and back) of a "cheat sheet." You may use a graphing calculator of your choice. Please read the questions carefully. If I say "compute," this means the solution will be a number. **Please round to two decimal places.** I advise you to skip problems marked "[Extra Credit]" until you have finished the other questions on the exam, then loop back and plug in all the holes. I also advise you to use pencil.

The exam is 100 points total plus extra credit. Partial credit will be granted for incomplete answers on most of the questions. Box in your final answers. Good luck!

Problem 1

Some theoretical questions are below. The subparts are all independent unless otherwise indicated.

(a) [4 pt / 4 pts] Show that the moment generating function (MGF) for the r.v. X where $X \sim U(a, b)$ is $M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$.

$$M_X(t) := E[e^{tX}] = \int_{Supp[X]} e^{tX} f(x) \, dx = \int_a^b e^{tx} \frac{1}{b-a} \, dx = \frac{1}{b-a} \left[\frac{e^{tx}}{t} \right]_a^b = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

(b) [4 pt / 8 pts] Using the result from the last problem, find the MGF for $W \sim U(0, 1)$.

$$X \sim U(a,b) \Leftrightarrow M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)} \Rightarrow W \sim U(0,1) \Leftrightarrow M_W(t) = \frac{e^{t(1)} - e^{t(0)}}{t((1)-(0))} = \boxed{\frac{e^t - 1}{t}}$$

(c) [4 pt / 12 pts] Using the result from the two previous problems, show that if $X = kW + m$ where k and m are constants $\in \mathbb{R}$, then $X \sim U(m, m+k)$.

$$X = kW + m \Rightarrow M_X(t) = e^{tm} M_W(kt) = e^{tm} \left(\frac{e^{(kt)} - 1}{kt} \right)$$

$$= \frac{e^{tm} e^{kt} - e^{tm}}{k(t)} = \frac{e^{t(m+k)} - e^{t(m)}}{t((m+k) - (m))} \Rightarrow X \sim U(m, m+k)$$

$$\text{since } X \sim U(a,b) \Leftrightarrow M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

(d) [4 pt / 16 pts] If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(m, m+k)$ and n is large, what is the approximate distribution of T where $T := X_1 + X_2 + \ldots + X_n$?

$$\to \text{CLT kicks in}$$

$$\text{From the CLT, } T \sim N\left(n\mu, \left(\sqrt{n} \, \sigma \right)^2 \right)$$

$$X_i \sim U(m, m+k) \Rightarrow \mu = E[X] = \frac{a+b}{2} = \frac{(m) + (m+k)}{2} = m + \frac{k}{2}$$

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{(b-a)^2}{12}} = \frac{b-a}{\sqrt{12}} = \frac{(m+k) - (m)}{\sqrt{12}} = \frac{k}{\sqrt{12}}$$

$$\Rightarrow \boxed{T \sim N\left(n\left(m + \frac{k}{2}\right), \left(\sqrt{n} \, \frac{k}{\sqrt{12}} \right)^2 \right)}$$

(e) [4 pt / 20 pts]   [Extra Credit] Evaluate $\int_{\mathbb{R}} e^{-(x-3,142.67)^2} dx$ below.

If $X \sim N(\mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ $\Rightarrow$ $\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = 1$
and $\text{supp}(X) = \mathbb{R}$ for all $\mu \in \mathbb{R}, \sigma \in (0, \infty)$

$e^{-(x-3,142.67)^2} = \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-3,142.67)^2}$ where $\sigma^2 = \frac{1}{2}$

$\Rightarrow$ $\sqrt{\pi}\left(\frac{1}{\sqrt{2\pi(\frac{1}{2})}} e^{-\frac{1}{2(\frac{1}{2})}(x-3,142.67)^2}\right) dx = \sqrt{\pi}(1) = \boxed{\sqrt{\pi}}$

(f) [5 pt / 25 pts]   Let $g(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x} dx$ where $\lambda \in (0, \infty)$. What is $g'(0)$?

If $X \sim Exp(\lambda) := \lambda e^{-\lambda x}$ $\Rightarrow$ $g(t) = M_X(t)$ $\Rightarrow$ $g'(0) = E[X] = \boxed{\frac{1}{\lambda}}$

(g) [5 pt / 30 pts]   State the central limit theorem. Include all assumptions. Your answer should read like "If ... then ..."

$X_1, \ldots, X_n$ are iid from any distribution with finite μ and σ. If n is large, then $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.

(h) [4 pt / 34 pts]   [Extra Credit] Solve for the MGF of a discrete r.v. X using a sum over the discrete sample space Ω.

$M_X(t) := E[e^{tX}] = \sum_{x \in \text{supp}(X)} e^{tx} p(x) = \sum_{x \in \text{supp}(x)} e^{tx} P(X=x) = \boxed{\sum_{\omega \in \Omega} e^{tX(\omega)} p(\omega)}$

Problem 2

In this problem, we will investigate mock situations involving waiting for the Professor Kapelner's "favorite" bus, the Q64 (service between Forest Hills Continental Ave and Elechester via Jewel Ave).

3

(a) [4 pt / 38 pts]   If we consider the waiting time (in minutes) for the bus to be $T \sim \text{Exp}(\lambda)$, find the probability I wait exactly 4 minutes as a function of λ.

Since T is continuous, $P(T=c)=0 \ \forall c \Rightarrow \boxed{P(T=4)=0}$

(b) [4 pt / 42 pts]   If we consider the waiting time (in minutes) for the bus to be $T \sim \text{Exp}(\lambda)$, find the probability I wait between 4 and 10 minutes as a function of λ.

$$P\left(T \in [4,10]\right) = F(10) - F(4) = \left(1 - e^{-10\lambda}\right) - \left(1 - e^{-4\lambda}\right) = \boxed{e^{-4\lambda} - e^{-10\lambda}}$$

(c) [4 pt / 46 pts]   I've already stood at the bus stop 10 minutes waiting for my favorite bus. Find the probability I wait more than 14 minutes *in total* as a function of λ.

$$P\left(T>14 \mid T>10\right) = P(T>4) = 1 - P(T\leq 4) = 1 - \left(1 - e^{-4\lambda}\right) = \boxed{e^{-4\lambda}}$$

by the
memoryless
property
of the exponential
distribution

4

(d) [4 pt / 50 pts]   Assuming the waiting time is a realization from that same data generating process $T \sim \text{Exp}(\lambda)$ every morning and each morning is independent of other mornings, find the probability I wait between 4 and 10 minutes for the bus three days out of a five-day workweek as a function of λ.

$$X \sim Bin\left(n=5, \ p = e^{-4\lambda} - e^{-10\lambda}\right) := \binom{n}{x} p^x (1-p)^{n-x}$$

$$P(X=3) = \boxed{\binom{5}{3}\left(e^{-4\lambda} - e^{-10\lambda}\right)^3 \left(1-\left(e^{-4\lambda} - e^{-10\lambda}\right)\right)^2}$$

(e) [4 pt / 54 pts]   Why is the model you created in (d) a bad model that can potentially yield wrong answers?

The bus waiting time each morning are not identically distributed (i.d.). Bad weather days exhibit lower λ than good weather days. If each day is not i.d., you cannot use the binomial model in (d) to compute probabilities.

Problem 3

In this problem, we will investigating 5th graders and playing video games. The photograph below was part of a larger image of the entire auditorium at a boys' private school in Brooklyn.

Altogether, of 156 fifth grade students in this auditorium, 109 said they play video games before they do their homework.

(a) [2 pt / 56 pts]   Let p denote the true proportion of fifth graders who play video games before they do their homework. What is your best guess for p? Round to the nearest two decimal places.

best guess of p is $\hat{p} = \frac{109}{156} = \boxed{0.70}$

(b) [4 pt / 60 pts]   Construct a 99% confidence interval (CI) for the true proportion of fifth graders who play video games before they do their homework. Use the correct notation for $CI_{a,b}$ and round to the nearest two decimal places.

$$CI_{p, 99\%} := \left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] = \left[0.70 \pm 2.56 \sqrt{\frac{0.70(1-0.70)}{156}} \right] = \boxed{[0.61, 0.79]}$$

(c) [4 pt / 64 pts]   Assuming that the CLT has kicked in by $n = 156$ and that the approximation $p(1-p) \approx \hat{p}(1-\hat{p})$ is not an issue, why could the CI you produced in (b) be not valid?

p: proportion of <u>all</u> fifth graders who play video games before homework. But $\hat{p}$ is derived from a sample of only boys in a private school in brooklyn and hence the sample could be <u>biased</u>.

(d) [3 pt / 67 pts]   Regardless of whatever you wrote in (c), give an interpretation of what the interval in (b) means.

There are three questions:

① If we repeated many samples, 99% of them would cover p. (classic A)

② Before we did our survey $P\left(p \in CI_{p, 99\%} \right) = 99\%$.   (classic B)

③ Our interval in (b) has a 95% probability that p is inside. (subjective)

6

(e) [3 pt / 70 pts]   Provide two ways to make the interval in (b) have smaller width.

① Increase α above 1%,

② Increase n above 156.

(f) [8 pt / 78 pts]   The parents association of the school are concerned that more than 50% of their children play video games before doing their homework. Use a hypothesis test to evaluate their concern. Pick your own α. Report your findings in English.

$H_0: p \leq 50\%$.
$H_a: p > 50\%$.
$\alpha = 2.5\%$ (since I know the z_α!)

Acc Region $:= \left(-\infty, \; p + z_\alpha \sqrt{\frac{p(1-p)}{n}} \right]$

$= \left(-\infty, \; 0.5 + 2\sqrt{\frac{0.5(1-0.5)}{156}} \right]$

$= \left(-\infty, \; 0.58 \right]$

$\hat{p} = 0.7 \notin$ Acc Region $\Rightarrow$ Reject H_0 / Accept H_a $\Rightarrow$ More than 50% of the students in this school are playing video games before doing their homework.

(g) [4 pt / 82 pts]   Assume $p = 0.5$, and that each students' response is $\overset{iid}{\sim}$ what is the *exact* probability of getting this $\hat{p}$ on this survey? No need to compute explicitly.

$X \sim Bin\left(n = 156, \; p = 0.5 \right)$ $\hat{p} = 0.7 \Leftrightarrow X = 109$

$P(X = 109) = \binom{156}{109} 0.5^{109} (1-0.5)^{156-109} = \boxed{\binom{156}{109} 0.5^{156}}$

→ a normal approx. is not acceptable here!

7
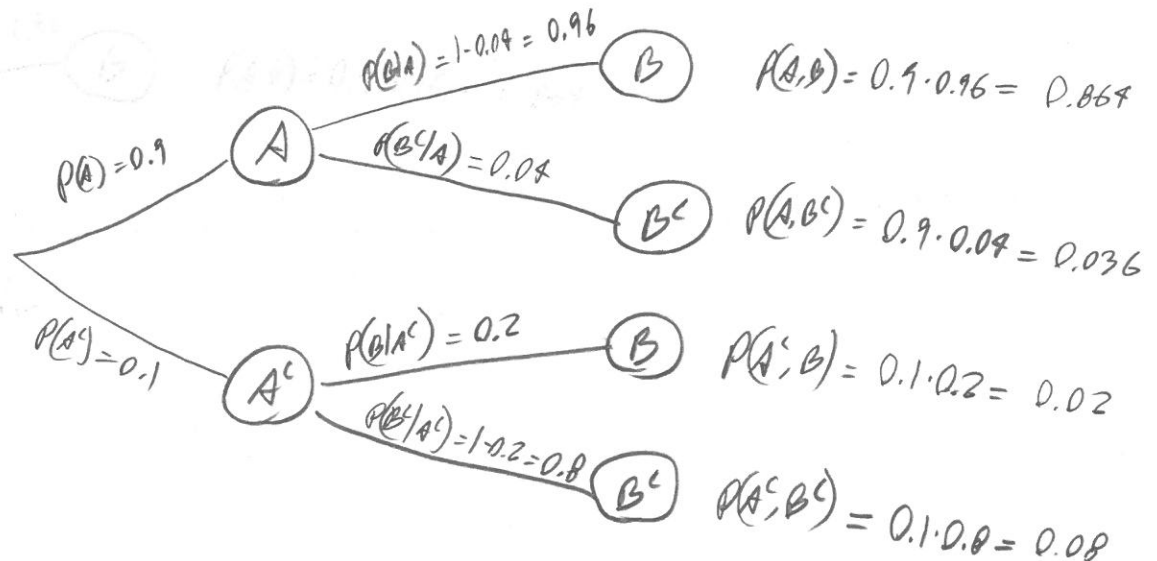
Problem 4

We will be investigating an important topic to society, although grisly: murder and the justice system. It is remarkably difficult to find good data on this, so I will use what I came up with. Of all murder trials, about 90% of defendents are convicted (the justice system says they are guilty of murder). Call the event of a guilty conviction A. $\Rightarrow P(A) = 0.9$

But some people who are guilty are actually innocent! Guardian magazine recently published an article that claims that "... 4% of people sentenced to death row are innocent..." as well as other claims. How they did their study and came up with this estimate doesn't concern us. Further assume that this proportion is the same for all those convicted of murder. That means of those convicted, 4% are really innocent. Call the event of being truly guilty B. $\Rightarrow P(B^c/A) = 0.04$

Now, for the purposes of this problem only, assume that if someone is given the verdict of not guilty, there is a 20% chance they actually committed the murder *and got away with it scott-free*. I can't find any data supporting this but it seems reasonable to me. Assume it is true for this exercise. $\Rightarrow P(B/A^c) = 0.2$

(a) [5 pt / 87 pts]   Draw a probability tree with the first event level being A and the second event level being B. Make sure to mark all branches and provide the joint probabilities on the right of the four joint events. Do not round your answers.

$P(B/A) = 1-0.04 = 0.96$

$P(A) = 0.9$

$P(B^c/A) = 0.04$

$P(A,B) = 0.9 \cdot 0.96 = 0.864$

$P(A,B^c) = 0.9 \cdot 0.04 = 0.036$

$P(A^c) = 0.1$

$P(B/A^c) = 0.2$

$P(B^c/A^c) = 1-0.2 = 0.8$

$P(A^c, B) = 0.1 \cdot 0.2 = 0.02$

$P(A^c, B^c) = 0.1 \cdot 0.8 = 0.08$

(b) [3 pt / 90 pts]   Of 12,000 people who are brought to trial for murder annually, how many in expectation are both guilty and given a guilty verdict? Assume all defendents' trials are independent.

$A \cap B$

$\Rightarrow P(A \cap B) = 0.864$

$X \sim Bin\left(n = 12,000, \ p = 0.864\right)$

$E[X] = np = 12,000 \cdot 0.864 = \boxed{10,368}$

8

(c) [2 pt / 92 pts] What is the probability of a person brought to trial for murder being truly guilty?

$$P(B) = P(B, A) + P(B, A^c) = 0.864 + 0.02 = \boxed{0.884}$$

(d) [3 pt / 95 pts] What is the probability of the person being given a not guilty verdict (innocent) if they were in fact innocent of the murder they were charged with? Round to two decimal places.

$$P(A^c \mid B^c) = \frac{P(A^c, B^c)}{P(B^c)} = \frac{P(A^c, B^c)}{1 - P(B)} = \frac{0.08}{1 - 0.884} = \boxed{0.69}$$

(e) [2 pt / 97 pts] What is the probability the justice system makes a mistake for those brought to trial for murder? mistake := say guilty but innocent and/or say innocent but guilty

$$= (A \cap B^c) \cup (A^c \cap B)$$
$$\Rightarrow P(\text{mistake}) = P(A, B^c) + P(A^c, B) = 0.036 + 0.02 = \boxed{0.056}$$

(f) [2 pt / 99 pts] In America, the justice system is based on the *presumption of innocence* which is sometimes summed up in the pithy phrase "guilty until proven innocent." The burden of proof lies on the accuser to "prove beyond a reasonable doubt" the guilt of the accused. This is a long precedence in law dating back to Roman times as well as Islamic Law and Talmudic Law and it has been adopted by most civil law systems.

Thus, in our justice example, the null hypothesis is that the defendent is innocent. What is the alternative hypothesis?

$$\boxed{H_a: \text{the defendent is guilty}}$$

(g) [2 pt / 101 pts]   Assuming the H_0 and H_a of part (e), what does the Type I error rate for our murder convictions case mean in English? Then, create a probability statement with the events A, A^C, B and/or B^C.

Type I err rate := probability of rejecting H_0 when it is true
= probability of convicting when truly innocent
= $P(A|B^C)$

(h) [2 pt / 103 pts]   Assuming the H_0 and H_a of part (e), what does the Type II error rate for our murder convictions mean in English? Then, create a probability statement with the events A, A^C, B and/or B^C.

Type II err rate := probability of retaining H_0 when it is false
= probability of not convicting when truly guilty
= $P(A^C|B)$

(i) [2 pt / 105 pts]   Calculate the Type I error rate.

$$P(A|B^C) = \underbrace{1 - P(A^C|B^C)}_{\text{complement rule for cond. prob.}} = 1 - 0.69 = \boxed{0.31}$$

part (d)

(j) [3 pt / 108 pts]   In an evil and sadistic society, they may have the inverse principle of justice: the accused are considered guilty until they prove their innocence beyond a reasonable doubt. In this case, would the Type I error be greater or lower than the number you just computed in (i)? Why?

H_0: guilty
H_a: innocent

see (g), Type I err rate := prob. of rejecting H_0 when it is true
= prob of saying innocent when guilty
= $P(A^C|B)$

There are three ways to answer this based on how you read the question!

① Type I err. rate is always α and therefore does not change $\Rightarrow$ $\boxed{\text{Same}}$

② $P(A^C|B) = \frac{P(A^C,B)}{P(B)} = \frac{0.02}{0.884} = 0.023 < 0.31 \Rightarrow \boxed{\text{Lower}}$

③ In a sadistic society (H_0/H_a flipped)
saying someone is guilty when they in fact innocent is more likely $\Rightarrow$ $\boxed{\text{greater}}$