

Math 241 Fall 2015
Midterm Examination Two

Solutions

Professor Adam Kapelner

November 12, 2015

Full Name _____ Section (A, B or C) _____

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 seventy five minutes and closed-book. You are allowed one page (front and back) of a "cheat sheet." You may use a graphing calculator of your choice but *no cell phones*. Please read the questions carefully. If the question reads "compute," this means the solution will be a number otherwise you can leave the answer in choose, permutation, factorial, summation or any other notation which could be resolved to a number with a computer. 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

We will be investigating insurance on tornado damage across America. Note that tornadoes are most likely to hit Texas, Kansas, Oklahoma, South Dakota and others and tornadoes are least likely to hit Vermont, Utah, New Hampshire and others (see <http://www.homefacts.com/tornadoes.html> for more information).



A large insurance company underwrites policies for tornado damage. This question uses mock data and mock assumptions.

- (a) [2 pt / 2 pts] The nationwide ^{mean} average proportion for a house having a claim is 0.5% per year. Create a random variable model B for the number of claims for one house. Write " $B \sim$ [something]" below.

$$B \sim \text{Bernoulli}(0.5\%) \quad \text{or} \quad B \sim \text{Bernoulli}(1, 0.005)$$

- (b) [3 pt / 5 pts] Assume each home is modeled by an iid random variable of the model you built in part (a). If there are 300,000 homes covered under tornado insurance, create a random variable model X for the total number of claims per year. Write " $X \sim$ [something]" below.

$$X \sim \text{Binom}(300000, 0.5\%)$$

- (c) [4 pt / 9 pts] What is the probability that exactly 54,345 claims are filed in one year? No need to compute explicitly.

$$P(X = 54,345) = \binom{300,000}{54,345} (0.5\%)^{54,345} (99.5\%)^{300,000 - 54,345}$$

- (d) [5 pt / 14 pts] What is the probability that one or more claims are filed in a year? No summation is allowed in your answer. No need to compute explicitly.

$$P(X \geq 1) = 1 - P(X=0) = 1 - (99.5\%)^{300,000}$$

- (e) [3 pt / 17 pts] Claims pay \$80,000. Create a r.v. Y as a function of X from part (b) to model the total amount the insurance company must pay per year.

$$Y = \$80,000 X$$

- (f) [4 pt / 21 pts] What is the expected amount of money the insurance company pays out in a year? Compute explicitly. Include the unit in your answer.

$$E(Y) = E(\$80,000 X) = \$80,000 E(X) = \$80,000 \cdot 300,000 \cdot 0.5\% = \$120 M$$

- (g) [5 pt / 26 pts] What is the standard error of the amount of money the insurance company pays out in a year? Compute explicitly.

$$SE(Y) = SE[\$80,000 X] = \$80,000 \sigma = \$80,000 \sqrt{300,000 \cdot 0.5\% \cdot 99.5\%} = \$3.09 M$$

- (h) [8 pt / 34 pts] Write two separate concerns or problems you have with using this $\overset{iid}{\sim}$ assumption for r.v. X for the 300,000 houses.

1) Not independent: If there's a tornado that strikes my house and you live next door, there's a high chance it will strike your house too

2) Not identically distributed: As the Problem 1 Leader states: houses in TX, KS, OK, SD etc have a greater chance of being hit by a tornado than houses in VT, VT, NH, etc. 0.5% is only the nationwide average; it is not the risk for all houses.

- (i) [3 pt / 37 pts] Given the concerns from part (h), we no longer believe the $\tilde{\sim}$ model. Now, what can you say about the probability calculations you made in parts (c) and (d)? Write "correct" if the calculations in (c) and (d) are still correct or "not correct" if no longer correct.

not correct (the binomial model is wrong)

- (j) [3 pt / 40 pts] Given the concerns from part (h), we no longer believe the $\tilde{\sim}$ model. Now, what can you say about the expected value calculation you made in part (f)? Write "correct" if the calculation in (f) is still correct or "not correct" if no longer correct.

correct (it is still the sum of r.v.'s)

- (k) [3 pt / 43 pts] Given the concerns from part (h), we no longer believe the $\tilde{\sim}$ model. Now, what can you say about the standard error calculation you made in part (g)? Write "correct" if the calculation in (g) is still correct or "not correct" if no longer correct.

not correct (the SE calc. assumes additive variance which is only true if the r.v.'s are independent)

- (l) [2 pt / 45 pts] Tornadoes are relatively common across America. It is rare in an entire year that there are no tornadoes and hence no claims. Per day, there is about a 4% chance of a tornado happening somewhere in America. Assume each new day is independent of the previous days. This question is no longer concerned with the number of houses or number of claims. Create a random variable model T for the time it takes to see one tornado somewhere across America. Write " $T \sim$ [something]" below.

$$T \sim \text{Geom}(4\%)$$

- (m) [5 pt / 50 pts] What is the probability it takes exactly two weeks (i.e. 14 days) until you see one tornado? (The tornado occurs on the 14th day). Compute explicitly.

$$P(T=14) = (1-4\%)^{13} 4\% = \boxed{0.0235}$$

- (n) [4 pt / 54 pts] What is the expected number of days that pass before you see one tornado? Compute explicitly.

$$E(T) = \frac{1}{4\%} = \boxed{25}$$

- (o) [2 pt / 56 pts] The first time there is ten tornadoes in a year (beginning January 1st), there is a company board meeting the night of that tenth tornado. Create a random variable T for the time (number of days) it takes to have this board meeting. Write " $T \sim$ [something]" below.

$$T \sim \text{NegBm}(10, 4\%)$$

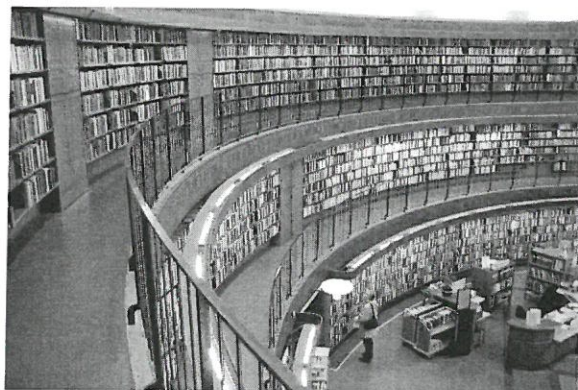
- (p) [5 pt / 61 pts] What is the probability of having a board meeting in 90 days or less? No need to compute but your answer must be computable exactly (no approximations).

$$P(T \leq 90) = \sum_{t=10}^{90} \binom{t-1}{9} (1-4\%)^{t-10} (4\%)^{10} \quad \leftarrow \text{CDF of NegBm}$$

$$(\text{or}) = 1 - \sum_{i=0}^9 \binom{90}{i} (4\%)^i (1-4\%)^{90-i} \quad \leftarrow \text{using the CDF of the binomial equivalence computation}$$

Problem 2

A library has 100,000 books of which 5,000 are academic, 10,000 are science fiction, 10,000 are romance novels and 75,000 are of other genres.



- (a) [3 pt / 64 pts] If I randomly take 10 books off the shelves, build a random variable model X for the number of science fiction books I get. Write " $X \sim$ [something]" below.

$$X \sim \text{Hyper}(10, 10000, 100000)$$

- (b) [4 pt / 68 pts] What is the probability I get 3 science fiction books of the 10 I randomly select? No need to compute.

$$P(X=3) = \frac{\binom{10000}{3} \binom{90000}{7}}{\binom{100000}{10}}$$

- (c) [4 pt / 72 pts] Approximate your answer to (b) by assuming the number of books in the library is effectively infinite. No need to compute.

$$X \approx \text{Bin}(10, 0.1)$$

$$P(X=3) = \binom{10}{3} 0.1^3 0.9^7$$

Problem 3

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

- (a) [5 pt / 77 pts] Give an example of a brand-name r.v. X where $\sigma^2 \notin \text{Supp}[X]$. You must give the parameter values and demonstrate that its variance is not in the support. Make sure you write " $X \sim$ [something]" below.

$$X \sim \text{Bern}(0.5)$$

$$\text{Supp}[X] = \{0, 1\}$$

$$\sigma^2 = 0.5(1-0.5) = 0.25$$

$$0.25 \notin \{0, 1\} \checkmark$$

- (b) [5 pt / 82 pts] Give an example of a brand-name r.v. X where $\sigma^2 \in \text{Supp}[X]$. You must give the parameter values and demonstrate that its variance is in the support. Make sure you write " $X \sim$ [something]" below.

$$X \sim \text{Bin}(4, \frac{1}{2})$$

$$\text{Supp}[X] = \{0, 1, 2, 3, 4\}$$

$$\sigma^2 = 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = 1$$

$$1 \in \{0, 1, 2, 3, 4\} \checkmark$$

(c) [4 pt / 86 pts] Compute $\sum_{x=0}^{\infty} 0.9^x$ explicitly.

$$|q| < 1 \Rightarrow \sum_{x=0}^{\infty} q^x = \frac{1}{1-q} \Rightarrow \sum_{x=0}^{\infty} 0.9^x = \frac{1}{1-0.9} = \frac{1}{0.1} = 10$$

(d) [3 pt / 89 pts] Define the concept of "data" below in one English sentence.

data is realization of random variables

(e) [4 pt / 93 pts] The following numbers come from an $\overset{iid}{\sim}$ data generating process:

-0.93 0.24 -1.24 0.26 -0.20 1.32 0.24
 -0.33 -0.10 -0.50 1.13 0.68 1.44 -0.59
 0.22 -1.82 1.47 -1.22 -0.65 -1.37 0.81
 0.84 -0.41 -0.11 0.67 1.15 -2.39 -0.45
 -0.15 -1.15 0.87 -0.52 -3.37 0.69 -0.26
 0.37 -1.01 0.46 1.37 0.47 1.22 1.89
 0.12 1.69 -1.50 -1.42 0.24 -1.41 1.38

of which the $\bar{x} = -0.03$. What is this \bar{x} close to? One short English sentence only.

The sample average (\bar{x}) is close to the ^{mean (or)} expectation (or) expected value of the data generating process (μ).

(f) [3 pt / 96 pts] What theorem / law did you invoke to answer the previous question (part e)?

Law of Large Numbers (LLN)

(g) [4 pt / 100 pts] I read in the newspaper that America's GDP in 2014 was \$46,405.26. Could this number have been different? Yes/no and explain.

Yes, this is a realization of a r.v.'s model; hence it is random; hence it could've been different.

- (h) [3 pt / 103 pts] [Extra credit] Find $SE[aX + bY]$ where X and Y are r.v.'s and $a, b \in \mathbb{R}$. Use the notation μ_X and μ_Y for the expectations of X and Y and σ_X^2 and σ_Y^2 for the variances of X and Y and σ_{XY} for the $\text{Cov}[X, Y]$.

$$\begin{aligned}
 SE(aX + bY) &= \sqrt{Var(aX + bY)} = \sqrt{Var(aX) + Var(bY) + 2Cov(aX, bY)} \\
 &= \sqrt{a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2(E[aXbY] - E[aX]E[bY])} \\
 &= \sqrt{a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2(ab E[XY] - ab E[X]E[Y])} \\
 &= \sqrt{a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab(E[XY] - \mu_X \mu_Y)} \\
 &= \sqrt{a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab \sigma_{XY}}
 \end{aligned}$$

- (i) [3 pt / 106 pts] [Extra credit] Consider $X_1, \dots, X_n \stackrel{iid}{\sim} \text{NegBin}(r, p)$ Find $SE[\bar{X}_n]$.

$$SE[\bar{X}] = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{r \frac{1-p}{p^2}}}{\sqrt{n}} = \sqrt{\frac{r}{n} \frac{1-p}{p^2}}$$

- (j) [4 pt / 110 pts] [Extra credit] Prove that the negative binomial r.v. is not memoryless.

(See next page)

$$X \sim \text{NegBin}(r, p)$$

Want to show:

$$P(X = x_0 + x_0 \mid X > x_0) \neq P(X = x) \quad \text{i.e. not memoryless}$$

$$\Rightarrow \frac{P(X = x_0 + x_0)}{P(X > x_0)} \neq P(X = x)$$

$= 1 - P(X \leq x_0)$

$$\Rightarrow \frac{\binom{x+x_0-1}{r-1} (1-p)^{x+x_0-r} p^r}{1 - \sum_{t=r}^{x_0} \binom{t-1}{r-1} (1-p)^{t-r} p^r} \neq \binom{x-1}{r-1} (1-p)^{x-r} p^r$$

It suffices to show for any valid r, p, x_0, x

so choose $p = \frac{1}{2}$, $r = 2$, $x_0 = 10$, $x = 10$

$$\frac{\binom{19}{1} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2}{1 - \sum_{t=2}^{10} \binom{t-1}{1} \left(\frac{1}{2}\right)^t} \neq \binom{9}{1} \left(\frac{1}{2}\right)^{10} \Rightarrow \frac{19 \frac{1}{2^{12}}}{1 - \frac{1}{2^2}} \neq 9 \frac{1}{2^{10}} \Rightarrow 0.00618 \neq 0.00879$$