

Math 241 Fall 2014-2015
Midterm Examination Two *Solutions*

Professor Adam Kapelner

November 13, 2014

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 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. Please read the questions carefully. If I say "compute," this means the solution will be a number. 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 You are opening a new luxurious fitness center in Key Garden Hills on Main St. since the closest gym is in Forest Hills. This problem will ask questions about membership that your business partner is interested in.



Before even opening the gym, you have 20 members who pre-signed up for a ~~special discount price~~. 13 signed up at the base price (the "base clients") and 7 signed up at the premium price (the "premium clients").

- (a) [3 pt / 3 pts] The gym doors open and on the first day, 8 of the 20 pre-signed members show up. Create a r.v. X that models the number of premium clients who show on that first day. Write $X \sim$ something and give the values of the parameter(s) if they are known.

$$X \sim \text{hypergeometric} \left(\underset{\substack{\uparrow \\ n}}{8}, \underset{\substack{\uparrow \\ K}}{7}, \underset{\substack{\uparrow \\ N}}{20} \right)$$

- (b) [3 pt / 6 pts] What are all the possible numbers of premium clients that can show on the first day out of those 8 members present? That is, what is the $\text{Supp}[X]$?

$$\text{Supp}[X] = \{0, 1, \dots, 7\}$$

- (c) [5 pt / 11 pts] What is the probability there are 4 premium clients among those 8 members present? You can leave as an expression and not compute.

$$f(4) = \frac{\binom{7}{4} \binom{13}{4}}{\binom{20}{8}}$$

- (d) [2 pt / 13 pts] You project that 2 months down the line there will be $n = 100$ members in addition to the 20 who presigned. You also project the ratio of the number premium clients to number of base clients will remain the same. What is the probability that the next member who signs up will sign up as a premium client? Denote this probability " p ". So write " $p =$ something" below and round to the nearest two digits.

$$p = \frac{7}{20} = 0.35$$

- (e) [3 pt / 16 pts] With every new client of the 100 who signs up, we do not know beforehand whether they will be a base or premium client. John walks in the door. Let X model 1 if John is a premium client and 0 if he's a basic client. What is the distribution of X ? Write " $X \sim$ something" below and give the values of the parameter(s) if they are known.

$$X \sim \text{Bernoulli}(p = 0.35)$$

- (f) [3 pt / 19 pts] Assume every potential client doesn't talk to each other and they all have the same propensity to be a premium client. This propensity is the answer to part (d). Let X model the number of premium clients out of the 100 members who are projected to sign up (do not concern yourself with the 20 who pre-signed). Write $X \sim$ something and give the values of the parameter(s) if they are known.

$$X \sim \text{Binomial}(n = 100, p = 0.35)$$

- (g) [6 pt / 25 pts] Explain in English why your answer to (f) may not be a valid model. That is, explain why the assumptions written in the question text of part (f) may not hold.

Clients do talk to each other and sign up in groups. (violation of independence)
 Not everyone has the same propensity to sign as a premium client since some clients are rich, etc. (violation of identical dist.)

- (h) [1 pt / 29 pts] Find $\mathbb{P}(X = 50)$, the probability that 50 of the 100 clients will be premium. You can leave your answer as an expression; you do not have to compute it.

$$f(x) = \binom{100}{x} p^x (1-p)^{100-x}, \quad f(50) = \binom{100}{50} 0.35^{50} 0.65^{50}$$

- (i) [1 pt / 33 pts] Find the probability that between 40 and 75 of the 100 clients are premium. You can leave your answer as an expression; you do not have to compute it. Do not leave in terms of $f(x)$.

$$\mathbb{P}(X \in [40, 75]) = \sum_{x=40}^{75} \binom{100}{x} 0.35^x 0.65^{100-x}$$

- (j) [3 pt / 36 pts] You have torn out page 9 which displays Table 1. This table provides the values of the PMF and CDF of the r.v. you created in (f). The rows represent the values of the PMF and CDF for different values of $x \in \text{Supp}[X]$. Some rows are left out (they are denoted with a vertical ellipses, ...). Which values of x are left out? Write the explicit set of left out x values below.

$$\{0, \dots, 19\} \cup \{52, 53, \dots, 100\}$$

- (k) [3 pt / 39 pts] Write one sentence *in English* about why you think those rows are left out.

$f(x)$ in those rows is near zero, therefore those values of x are essentially $\notin \text{Supp}(X)$

- (l) [1 pt / 40 pts] The median of X is 35.
- (m) [1 pt / 41 pts] The 10%ile of X is ~~28~~ 29.
- (n) [1 pt / 42 pts] The mode of X is 35.
- (o) [4 pt / 46 pts] Compute $\mathbb{E}[X]$ to two decimals.

$$\mathbb{E}(X) = 4p = 100 - 0.35 = \boxed{35}$$

- (p) [3 pt / 49 pts] Let $Y = 100 - X$ which are the number of base clients. Find $\mathbb{E}[Y]$ to two decimals using your answer to part (o). If you did not solve part (o), pretend the answer to (o) is 30.12.

$$\mathbb{E}(Y) = \mathbb{E}(100 - X) = 100 - \mathbb{E}(X) = 100 - 35 = \boxed{65}$$

- (q) [3 pt / 52 pts] Your profit on a base client is \$17.78/mo and your profit on a premium client is \$26.43/mo. Let C be the r.v. that represents profit based on clients. Write C as a function of X , Y and constants below.

$$C = 26.43X + 17.78Y$$

- (r) [2 pt / 54 pts] You have to pay \$2000/mo on expenses. Let P be the r.v. representing total profit which is your answer in (q) combined with your expenses. Write P below as a function of X , Y and constants below.

$$P = -2000 + 26.43X + 17.78Y$$

- (s) [5 pt / 59 pts] Find $E[P]$. Compute explicitly to the nearest cent.

$$E[P] = -2000 + 26.43 \underbrace{E[X]}_{35} + 17.78 \underbrace{E[Y]}_{65} = \boxed{81.25}$$

- (t) [5 pt / 64 pts] Your friend calculates $\text{Var}[P] = (\$26.43/\text{mo})^2 \sigma_X^2 + (\$17.78/\text{mo})^2 \sigma_Y^2$. Why is his answer wrong? Write one sentence in English below.

X, Y are not independent. Thus, he must account for the $\text{Cov}(X, Y)$ in the calculation.

- (u) [5 pt / 69 pts] New clients start signing up and everytime a new premium client signs up we give a bonus to the sales rep. What is the probability we have to wait for 3 new signups to give a bonus? You can leave as an expression and not compute.

$$X \sim \text{Geometric}(0.35) \Rightarrow f(x) = (1-p)^{x-1} p \Rightarrow f(3) = \boxed{0.65^2 0.35}$$

- (v) [3 pt / 72 pts] What is the expected number of clients that sign up before we give a bonus to the sales rep? Round to the nearest two decimals.

$$E[X] = \frac{1}{p} = \frac{1}{0.35} = \boxed{2.86}$$

- (w) [5 pt / 77 pts] What is the probability we wait for 12 new signups to give 5 bonuses?
You can leave as an expression and not compute.

$$X \sim \text{Neg Bin}(5, 0.35)$$

$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

$$f(12) = \binom{11}{4} p^5 (1-p)^7 = \boxed{\binom{11}{4} (0.35)^5 (0.65)^7}$$

- (x) [3 pt / 80 pts] Using your answer to part (v), what is the expected number of clients that need to sign up before we give out 5 bonuses? Round to the nearest two decimals. If you did not get an answer to (v), use 3.03 as the answer to (v).

$$E(X) = r \left(\frac{1}{p} \right) = 5(2.86) = \boxed{14.29}$$

- (y) [2 pt / 82 pts] Fast forward to the future. The gym is successful. There are hundreds and hundreds of clients. You are now worried about capacity on Monday evening which is the busiest gym day and time period. For five weeks you count the maximum number of people in the gym during this time period and record the data: {13, 17, 10, 19, 13}. Calculate the sample average. Use the notation for the sample average we used in class.

$$\bar{X} = 14.8$$

- (z) [2 pt / 81 pts] If you collected data for many more weeks, what would the sample average become closer to? One word or one piece of notation is fine.

$$\mu \text{ or } E[X]$$

- (aa) [2 pt / 86 pts] The law that explains the phenomenon in part (z) is the law of ...

large numbers

- (bb) [3 pt / 89 pts] There are hundreds and hundreds of clients and a small chance each of them will show up on Monday evening. Model the number of people who show up one Monday evening using the r.v. X . Approximate $\mathbb{E}[X]$ using the sample average \bar{x} you found in part (y) and use this result to come up with the parameter(s) of X . Write " $X \sim$ something" below.

$$X \sim \text{Poisson}(\lambda = 14.4)$$

- (cc) [6 pt / 95 pts] Once again, you are concerned with capacity. The fire department says maximum occupancy is 23 people. If the capacity surpasses this amount, the gym can face heavy fines from the fire department. What is the probability that on a future Monday night, the gym goes over capacity? Leave as a computable expression. That is, an expression that can be put into a calculator (you can not put ∞ or $-\infty$ into a calculator because they are not real numbers). DO NOT compute explicitly.

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

$$P(X > 23) = 1 - P(X \leq 23) = 1 - \sum_{x=0}^{23} e^{-14.4} \frac{14.4^x}{x!}$$

- (dd) [5 pt / 100 pts] Consider collecting n data points just like the five data points you counted in part (y). This will result in a set of n r.v.'s which we will denote X_1, X_2, \dots, X_n which are $\overset{\text{iid}}{\sim}$ with expectation $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}$. Denote the sum of them as the r.v. $T_n = \sum_{i=1}^n X_i$. Now consider the following r.v.

$$\frac{T_n - n\mu}{\sigma\sqrt{n}}$$

and find $\mathbb{E}\left[\frac{T_n - n\mu}{\sigma\sqrt{n}}\right]$, the expectation of this r.v. as a function of μ, σ^2 and constants.

$$E(\bar{X}) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \mu$$

$$E\left[\frac{T_n - n\mu}{\sigma\sqrt{n}}\right] = \frac{1}{\sigma\sqrt{n}} E[T_n - n\mu] = \frac{1}{\sigma\sqrt{n}} (E(T_n) - n\mu) = \frac{1}{\sigma\sqrt{n}} (n\mu - n\mu) = 0$$

- (ee) [3 pt / 103 pts] [E.C] Find $\text{SE} \left[\frac{T_n - n\mu}{\sigma\sqrt{n}} \right]$, the standard error of this r.v. as a function of μ , σ^2 and constants.

$$\text{Var}(T_n) = \text{Var}(X_1 + \dots + X_n) = n\sigma^2$$

$$\text{Var} \left(\frac{T_n - n\mu}{\sigma\sqrt{n}} \right) = \frac{1}{\sigma^2 n} \text{Var}(T_n - n\mu) = \frac{1}{\sigma^2 n} \text{Var}(T_n) = \frac{\sigma^2 n}{\sigma^2 n} = 1$$

$$\text{SE} \left[\frac{T_n - n\mu}{\sigma\sqrt{n}} \right] = \sqrt{1} = 1$$

- (ff) [3 pt / 106 pts] [E.C] Let $Y := X^2$ where X is a r.v. with PMF unknown. Use the notation μ for $\mathbb{E}[X]$, σ^2 for $\text{Var}[X]$, μ_3 for $\mathbb{E}[X^3]$ and μ_4 for $\mathbb{E}[X^4]$. Assume μ , σ^2 , μ_3 , μ_4 all exist in \mathbb{R} . Find $\text{Var}[Y]$ as a function of μ , σ^2 , μ_3 and μ_4 .

$$\text{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$$

$$\mathbb{E}[Y] = \mathbb{E}[X^2] = \sigma^2 + \mu^2$$

$$\mathbb{E}[Y^2] = \mathbb{E}[X^4] = \mu_4$$

$$= \mu_4 - (\sigma^2 + \mu^2)^2$$

- (gg) [3 pt / 109 pts] [E.C] Consider $X \sim \text{Geometric}(p)$. Prove that $\mathbb{P}(X = k) = \mathbb{P}(X = t+k \mid X > t)$ for any $t \in \mathbb{N}$ and $k \in \mathbb{N}$.

$$\mathbb{P}(X=k) = (1-p)^{k-1} p$$

$$\mathbb{P}(X=t+k \mid X > t) = \frac{(1-p)^{t+k-1} p}{\mathbb{P}(X > t)} = \frac{(1-p)^{t+k-1} p}{(1-p)^t} = (1-p)^{k-1} p \quad \checkmark$$

$\mathbb{P}(X > t) = (1-p)^t$ since for all sequences from $\{1, \dots, t\}$ there had to be failure $(1-p) \cdot (1-p) \cdot \dots \cdot (1-p)$
 t times

x	$f(x)$	$F(x)$
\vdots	\vdots	\vdots
20	0.000	0.001
21	0.001	0.002
22	0.002	0.003
23	0.003	0.007
24	0.006	0.012
25	0.009	0.021
26	0.014	0.035
27	0.021	0.056
28	0.029	0.085
29	0.039	0.124
30	0.049	0.173
31	0.060	0.233
32	0.070	0.303
33	0.077	0.380
34	0.082	0.462
35	0.083	0.546
36	0.081	0.627
37	0.076	0.702
38	0.067	0.770
39	0.058	0.828
40	0.047	0.875
41	0.037	0.912
42	0.028	0.941
43	0.021	0.961
44	0.014	0.975
45	0.010	0.985
46	0.006	0.991
47	0.004	0.995
48	0.002	0.997
49	0.001	0.999
50	0.001	0.999
51	0.000	1.000
\vdots	\vdots	\vdots

Table 1: The PMF and CDF for a r.v. described in the problem text.