Math 241 Fall 2015
Final Examination

Solutions

Professor Adam Kapelner

December 17, 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.

______________________________ ______________

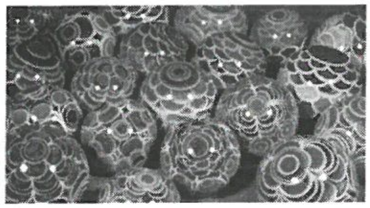
signature date

Instructions

This exam is 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 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. $\boxed{\text{Box}}$ in your final answers. Good luck!

Problem 1

Imagine a bag with 171 marbles each unique.



(a) [3 pt / 3 pts] Imagine you select 19 balls with replacement. How many ways are there to select balls if their order matters and selection order matters?

$$171^{19}$$

(b) [3 pt / 6 pts] Imagine you select 19 balls with replacement. How many ways are there to select balls if their order does not matter?

$$\frac{171^{19}}{19!}$$

(c) [3 pt / 9 pts] Imagine you select 19 balls without replacement. How many ways are there to select balls if their order matters?

$${}_{171}P_{19}$$

(d) [3 pt / 12 pts] Imagine you select 19 balls without replacement. How many ways are there to select balls if their order does not matter?

$$\binom{171}{19}$$

Problem 2

Some theoretical exercises are below.

(a) [2 pt / 14 pts] Consider the following r.v.,

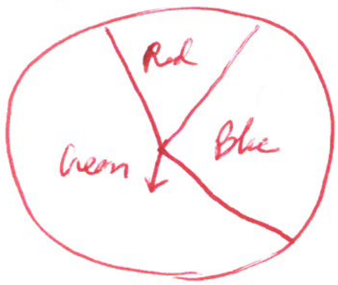
$$X \sim \begin{cases} 2 & \text{w.p. } 1/7 \\ 4 & \text{w.p. } 4/7 \\ -3 & \text{w.p. } 2/7 \end{cases}$$

Provide a set Ω for the domain of this r.v.

$\Omega = \{\text{Red}, \text{Green}, \text{Blue}\}$ or any set s.t. $|\Omega| \geq 3$

(b) [3 pt / 17 pts] For the r.v. above, describe or illustrate (any way you can) of a physical device that will create the r.v. Be sure to use Ω in your illustration from (a).

A spinner:



Red $\to 2$

Green $\to 4$

Blue $\to -3$

(c) [2 pt / 19 pts] Could it be possible that $|\text{Supp}[X]| < |\Omega|$? Write "yes" or "no" only.

Yes

Problem 3

Some more theoretical exercises are below.

(a) [4 pt / 23 pts] Consider $X \sim \text{Uniform}(\{a_1, a_2, a_3, a_4\})$ i.e. $|\text{Supp}[X]| = 4$. Find the MGF of X denoted by $M_X(t)$.

$$M_X(t) = E[e^{tX}] = \sum_{x \in Supp[X]} e^{tx} p(x) = e^{ta_1}\frac{1}{4} + e^{ta_2}\frac{1}{4} + e^{ta_3}\frac{1}{4} + e^{ta_4}\frac{1}{4} = \frac{1}{4}\left(e^{a_1 t} + e^{a_2 t} + e^{a_3 t} + e^{a_4 t}\right)$$

(b) [7 pt / 30 pts] Consider a large number of $X_1, \ldots, X_n \overset{iid}{\sim} \text{Uniform}(\{a_1, a_2, a_3, a_4\})$. Consider $X_1 + \ldots + X_n$. What is its approximate distribution? Answer in as much detail as you can provide (partial credit will be given). Use sum notation for compactness.

$$T = X_1 + \ldots + X_n \sim N(n\mu, (\sqrt{n}\,\sigma)^2)$$

where $\mu = E[X] = \sum_{x \in Supp[X]} x\, p(x) = \frac{1}{4} \sum_{i=1}^{4} a_i$

and $\sigma = \sqrt{Var[X]} = \sqrt{E[(X-\mu)^2]} = \sqrt{\sum_{x \in Supp[X]} (x-\mu)^2 p(x)} = \sqrt{\frac{1}{4} \sum_{i=1}^{4} (a_i - \mu)^2} = \frac{1}{2}\sqrt{\sum_{i=1}^{4} (a_i-\mu)^2}$

(c) [7 pt / 37 pts] Consider Y_1 being distributed as 5 or 10 with equal probability and $Y_2 \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ where Y_1 and Y_2 are independent. Show that $Y_1 + Y_2$ is a uniform discrete r.v. with $|\text{Supp}[X]| = 4$ using MGF's.

$$M_{Y_1}(t) = \frac{1}{2}e^{t5} + \frac{1}{2}e^{t10} = \frac{1}{2}\left(e^{5t} + e^{10t}\right)$$

$$M_{Y_2}(t) = 1 - p + pe^t = \frac{1}{2} + \frac{1}{2}e^t = \frac{1}{2}\left(1 + e^t\right)$$

$$M_{Y_1+Y_2}(t) = \underbrace{M_{Y_1}(t)\, M_{Y_2}(t)}_{\text{by ind}} = \frac{1}{2}\left(e^{5t} + e^{10t}\right)\frac{1}{2}\left(1 + e^t\right) = \frac{1}{4}\left(e^{5t} + e^{10t} + e^{6t} + e^{11t}\right)$$

which is the same form as the mgf of (a) $\Rightarrow$ it is a unif. discrete with $|\text{Support}(X)| = 4$ with $a_1 = 5$, $a_2 = 10$, $a_3 = 6$, $a_4 = 11$

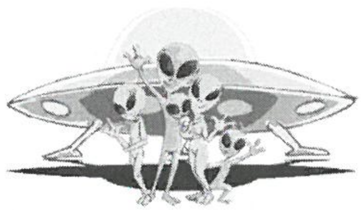
(d) [3 pt / 40 pts] [E.C.] Give examples of a_1, a_2, a_3, a_4 above which would most likely break the approximation from part (d) for a modest size n.

$a_1, a_2 \approx -\infty$ and $a_3, a_4 \approx \infty$

(e) [3 pt / 43 pts] [E.C.] Find $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^{13}\right]$ where μ and σ are the mean and standard error of X respectively.

Problem 4

We will investigate more about aliens and UFOs by looking at data from the National UFO Reportion Center Online Database.



monthly

Below is the number of events in each month in 2014 and represents the number of "UFO sightings" across America, a country with 319 million people. Assume that each sighting is unique — that is each person either makes a sighting for a given month or does not and there are no duplicates. That is, a person cannot make more than one sighting per 12-month period.

Month	Number
12/2014	521
11/2014	543
10/2014	786
09/2014	829
08/2014	922
07/2014	1096
06/2014	776
05/2014	652
04/2014	666
03/2014	517
02/2014	554
01/2014	715

significant

(a) [2 pt / 45 pts] We are interested in the true proportion of people who sight a UFO per year (not per month). Provide an estimate of this proportion below. Round to the nearest two significant digits.

$$\hat{p} = \frac{521+543+\ldots+715}{3.19e8} = 0.000027$$

(b) [6 pt / 51 pts] Provide a 99.7% confidence interval for the true proportion of people worldwide who see a UFO yearly. Round to two digits.

$z_{\frac{\alpha}{2}} = 3$ for 99.7% confidence

$$CI_{p,99.7\%} = \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = \left[.0027 \pm 3\sqrt{\frac{.0027(1-.0027)}{3.19e6}}\right] = [.000026, .000028]$$

(c) [3 pt / 54 pts] Give two possible concerns as to why the confidence interval you produced in (b) may not be valid.

1. Since $\hat{p} \simeq 0$ then $\hat{p}(1-\hat{p})$ may not be $\simeq p(1-p)$ so the MOE may be wrong

2. Main issue: the sample was American which not be representative of the whole world

+ others...

3. Selection bias for people who like to call in reports of UFO's.

(d) [4 pt / 58 pts] Regardless of the issues in (c), provide four possibe interpretations of the CI from (b).

1. Before 2014, there is a 99.7% chance that this survey will capture the true proportion of people who sight a UFO in a given year.
2. If the survey is repeated many years, 99.7% of the intervals will capture the true proportion (in expectation).
3. The interval captured the true proportion or it did not.
4. The interval has a 99.7% chance the true proportion is inside (if certain prior beliefs are assumed).

American

(e) [3 pt / 61 pts] Let's assume the point estimate of (a) is equal to the real population proportion (assume $p = \hat{p}$ for the remainder of the problem). in America Assume 2015 is the same as 2014 with regards to UFO sightings (and the population of America remained the same to the nearest million people). What is the probability of less than 100,000 sightings in 2015? No need to compute exactly. Sum notation allowed.

Let N be the r.v for # of sightings of UFO's in America in 2015

$N \sim \text{Binom}(n=3.19e6, p=.0027)$ $P(N<100{,}000) = \sum_{x=0}^{99{,}999} \binom{3.19e6}{x}(.00027)^x(1-.00027)^{3.19e6-x}$

(f) [2 pt / 63 pts] It is possible some states have more UFO sightings than other states. Hawaii had 67 sightings in 2014 for a population of 1.42 million people. Here are some examples:

Timestamp	City	Shape	Duration	Description
1/15/14 03:30	Paia HI	Sphere	2 min	I Saw a golden sphere shaped object hovering over the water on the North Shore at 3:30am.
1/15/14	Kealia HI	Fireball	20 min	Fireball w- chinese laturns and parade of triangles.
1/14/14 20:50	Waimanalo HI	Fireball	5 sec	I saw an orange fireball with flames trailing as it shot diagonally across the sky and disappeared behind the nearby mountain range.

Find the sample proportion for seeing a UFO sighting in a year in Hawaii. Round to two digits.

$$\hat{p} = \frac{67}{1.42e6} = 0.000047$$

(g) [7 pt / 70 pts] Test whether this proportion is different than the proportion in America overall.

$p := P(\text{UFO sighting} \mid \text{Hawaii})$

H_0: $p = 0.000027$

$:= P(\text{UFO sighting})$

H_a: $p \neq 0.000027$

$\alpha =$ 5% $\Rightarrow z_{\alpha/2} = 2$

$$\text{Retention Region} = \left[p \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right] = \left[.0027 \pm 2\sqrt{\frac{.000027(1-.000027)}{1.42e6}}\right] = [.000026, .000028]$$

Since $\hat{p} = .000047 \notin$ Retention Region $\Rightarrow$ Reject $\Rightarrow$ the proportion of UFO sightings in Hawaii is <u>not</u> the same the proportion of sightings in America

(h) [2 pt / 72 pts] Based on your result in (g), is the event of being in Hawaii and the event of seeing a UFO over the course of a year independent or dependent? Explain using the definition of dependence and independence.

Since $P(\text{UFO sighting} \mid \text{Hawaii}) \neq P(\text{UFO sighting})$

$\Rightarrow$ UFO sighting event and the Hawaii event are $\boxed{\text{dependent.}}$

Problem 5

We return to our discussion of Chevalier de Mere and his compulsive dice gambiling.

(a) [3 pt / 75 pts] Recall the dice game he played: you win if you get at least one 6-6 in 24 rolls of two dice. What is the probability of winning? Compute to four decimal places.

$P_w := P(\geq 1 \text{ 6-6 in 24 rolls of two dice}) = 1 - P(\text{zero 6-6 in 24 rolls})$

$= 1 - P(\text{no 6-6})^{24} = 1 - (1 - P(\text{6-6}))^{24} = 1 - \left(1 - \frac{1}{36}\right)^{24} = \boxed{.4914}$

(b) [6 pt / 81 pts] Chevalier de Mere most likely thought the probability of winning was 0.5 (not your answer from the previous question). Assuming Chevalier de Mere went easy on the wine and cigarettes and kept perfect records of wins and losses of everyone playing this game, how many games (denoted n) did he have to watch to be sure (at the 2.5% level) that the probability of winning was less than 50%. You can leave as an algebraic equation of n; you do not need to solve.

Let $\hat{p}$ be the proportion of wins, $\hat{p} \sim N\left(.4914, \left(\sqrt{\frac{.4914(1-.4914)}{n}}\right)^2\right)$

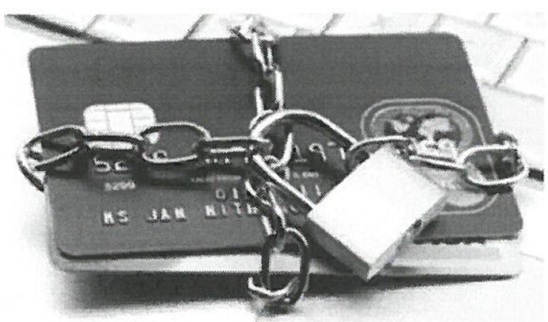
due to CLT

$P(\hat{p} > 0.5) = 2.5\% \Rightarrow P\left(\frac{\hat{p} - .4914}{\sqrt{\frac{.4914(1-.4914)}{n}}} > \underbrace{\frac{0.5 - .4914}{\sqrt{\frac{.4914(1-.4914)}{n}}}}_{f(n)}\right) = 2.5\%$

$\Rightarrow P(Z > f(n)) = 2.5\% \Rightarrow f(n) = 2 \Rightarrow \boxed{\frac{.5 - .4914}{\sqrt{\frac{.4914(1-.4914)}{n}}} = 2} \Rightarrow n = \boxed{13,517}$

Problem 6

We learn about a bank alert system here for detecting fraudulent charges on a credit card.



(a) [2 pt / 83 pts] Assume each charge is legitimate from the start. What are the null and alternative hypotheses for this situation with credit card charges? No credit given for the general definition. Answer in English.

H_0: the charge is legitimate
H_a: the charge is fraudulent

(b) [3 pt / 86 pts] Describe a Type II error for this situation with credit card charges. No credit given for the general definition. Answer in English.

Allowing a fraudulent charge to go unnoticed

(c) [3 pt / 89 pts] Assume the Type I error rate is α with cost $C_I \sim \mathcal{N}(\$100, \$50^2)$ and the Type II error rate is β with cost $C_{II} \sim \mathcal{N}(\$500, \$38^2)$. What is the expected cost to the bank for a single credit card transaction?

$E[\alpha C_1 + \beta C_2] = \alpha E[C_1] + \beta E[C_2] = \boxed{\$100\alpha + \$500\beta}$

Problem 7

Some questions about continuous r.v.'s

(a) [2 pt / 91 pts] The chi-squared distribution (denoted χ^2_k) has a PDF given by

$$\chi^2_k := f(x) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

and its support is all non-negative numbers. Note that the gamma function (denoted Γ) is evaluated to a number using a computer.

What are the parameter(s) of the χ^2_k model?

k

(b) [4 pt / 95 pts] Write an integral expression for $\mathbb{E}[X]$ where $X \sim \chi^2_k$.

$$\mathbb{E}(X) = \int_0^\infty x \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}\, dx$$

(c) [4 pt / 99 pts] Write an integral expression for $F(x)$ where $X \sim \chi^2_k$.

$$F(x) = P(X \leq x) = \int_0^x \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} t^{\frac{k}{2}-1} e^{-\frac{t}{2}}\, dt$$

any other free variable notation is acceptable here

(d) [3 pt / 102 pts] Consider $T \sim \text{Exp}(\lambda)$ which represents the time spent waiting for the Q64 in minutes. What is the probability of waiting more than 10 minutes?

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

(e) [4 pt / 106 pts] A stopwatch beeps every minute. After each beep, which r.v. models the waiting time in minutes for the Q64? Assume of course the bus hasn't come by the beep.

$T_{new} \sim \text{Exp}(\lambda)$ since the exponential has the memorylessness property