

Math 241 Fall 2014-2015  
Final Examination *Solutions*

Professor Adam Kapelner

December 18, 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 definition:

**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 an unauthorized 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

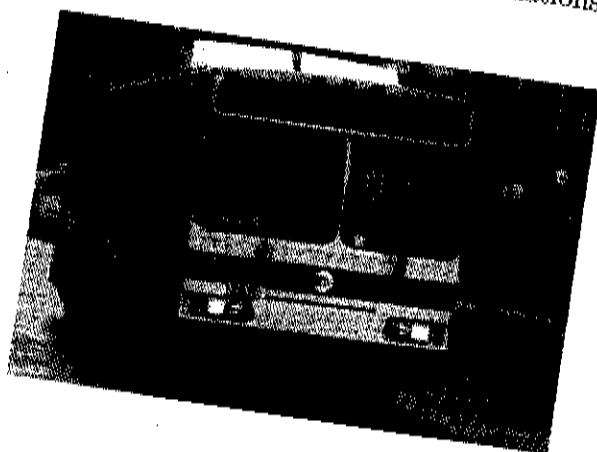
\_\_\_\_\_  
date

## 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. I also advise you to use pencil.

The exam is 100 points total. Partial credit will be granted for incomplete answers on most of the questions. I advise you to skip problems marked "[Extra Credit]" until you have finished the other questions on the exam Box in your final answers. Good luck!

Problem 1 In this problem, we will investigate mock situations involving waiting for the Q64 bus.



- (a) [4 pt / 4 pts] Let's say the Q64 is scheduled during rush hour to come every 8 min. Assume the bus really does keep to this schedule. You show up at the bus stop during rush hour at a random time. What is the probability you wait more than 2min for the next Q64 bus?

$$T \sim U(0, 8) \Rightarrow F_T(t) = \frac{t-a}{b-a} = \frac{t}{8} \quad \begin{array}{l} \swarrow \text{for } U(a,b) \\ \searrow \text{for exp } U(a,b) \end{array}$$

$$P(T \geq 2) = 1 - F_T(2) = 1 - \frac{2}{8} = \frac{6}{8} = \boxed{0.75}$$

- (b) [4 pt / 8 pts] Let's say the time when the next Q64 comes can be modeled as  $T \sim \text{Exp}(\lambda = 0.125)$  where the support of the r.v. is measured in minutes. What is the probability you wait more than 2min for the next Q64 bus? *Compute explicitly to 3 decimals.*

$$F_T(t) = 1 - e^{-\lambda t} = 1 - e^{-0.125t}$$

$$P(T \geq 2) = 1 - F_T(2) = 1 - (1 - e^{-0.125(2)}) = e^{-0.25} = \boxed{.779}$$

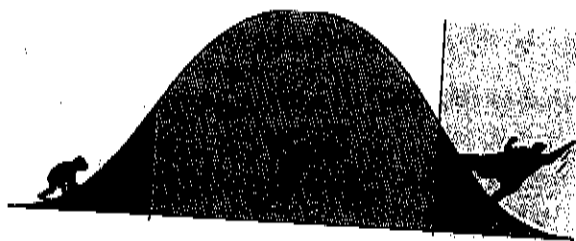
- (c) [3 pt / 11 pts] Let's say the time when the next Q64 comes can be modeled as  $T \sim \text{Exp}(\lambda = 0.125)$  where the support is measured in minutes. You've been sitting at the bus stop and have already waited two minutes. What is the probability given that you've been waiting that you wait more than *another* 2min for the next Q64 bus?

$$P(T \geq 2+2 \mid T \geq 2) = P(T \geq 2) = \boxed{0.779}$$

↑  
by memoryless  
property of exponential  
r.v.'s

↑  
computed in  
part (b)

**Problem 2** In this problem, we will investigate moment generating functions (mgf's) and the Central Limit Theorem.



- (a) [4 pt / 15 pts] Let  $X \sim \text{Bernoulli}(\frac{1}{2})$ . Show that  $M_X(t)$ , the mgf of  $X$  is  $\frac{1}{2}(1 + e^t)$ .

$$\begin{aligned} M_X(t) &:= E[e^{tx}] = \sum_{x \in \text{supp}(X)} e^{tx} f(x) = \sum_{x=0}^1 e^{tx} f(x) = \sum_{x=0}^1 e^{tx} p^x (1-p)^{1-x} = e^{t(0)} p^0 (1-p)^{1-0} + e^{t(1)} p^1 (1-p)^{1-1} \\ &= 1-p + pe^t \\ &\text{Here } p = \frac{1}{2} \end{aligned}$$

$$\Rightarrow M_X(t) = 1 - \frac{1}{2} + \frac{1}{2}e^t = \frac{1}{2} + \frac{1}{2}e^t = \boxed{\frac{1}{2}(1 + e^t)}$$

- (b) [4 pt / 19 pts] Set up an infinite sum that if solved would provide the mgf for  $X$ , a Poisson r.v. with parameter  $\lambda$ . Do not attempt to solve or simplify!

$$X \sim \text{Poisson}(\lambda) := \frac{e^{-\lambda} \lambda^x}{x!}$$

$$M_X(t) = E[e^{tx}] = \sum_{x \in \text{supp}(X)} e^{tx} f(x) = \boxed{\sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda} \lambda^x}{x!}}$$

- (c) [4 pt / 23 pts] If  $X \sim \text{Erlang}(k, \lambda)$ , which is the continuous analogue of the negative binomial distribution, then the PDF of  $X$  is  $f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$  and its mgf is  $M_X(t) = (1 - t/\lambda)^{-k}$ . Show that  $\mathbb{E}[X] = k/\lambda$  using any method you wish. If you cannot solve all the way, describe the method you would use to solve for partial credit. Hint: be careful with the negative sign.

$$M_X'(0) = \mathbb{E}(X)$$

$$M_X(u) = \left(1 - \frac{u}{\lambda}\right)^{-k} \quad \text{let } u = -\frac{t}{\lambda} \quad \frac{du}{dt} = -\frac{1}{\lambda} \quad \Rightarrow M_X(u) = (1+u)^{-k} \Rightarrow M_X'(u) = -k(1+u)^{-k-1}$$

$$M_X'(t) = \frac{dM_X(u)}{du} \cdot \frac{du}{dt} = -k(1+u)^{-k-1} \cdot -\frac{1}{\lambda} = \frac{k}{\lambda} \left(1 - \frac{t}{\lambda}\right)^{-k-1}$$

$$M_X'(0) = \frac{k}{\lambda} \left(1 - \frac{0}{\lambda}\right)^{-k-1} = \frac{k}{\lambda} 1^{-k-1} = \frac{k}{\lambda} \quad \checkmark$$

- (d) [4 pt / 27 pts] If  $Z_1, Z_2 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$  what is the mgf of  $\frac{1}{\sqrt{2}}Z_1 + \frac{1}{\sqrt{2}}Z_2$ ? Hint: the mgf for the standard normal is  $M_Z(t) = e^{t^2/2}$ .

$$M_Z(t) = e^{t^2/2} \quad M_{\frac{Z}{\sqrt{2}}}(t) = M_Z\left(\frac{t}{\sqrt{2}}\right) = e^{\left(\frac{t}{\sqrt{2}}\right)^2/2} = e^{\frac{t^2}{2} \cdot \frac{1}{2}} = e^{t^2/4}$$

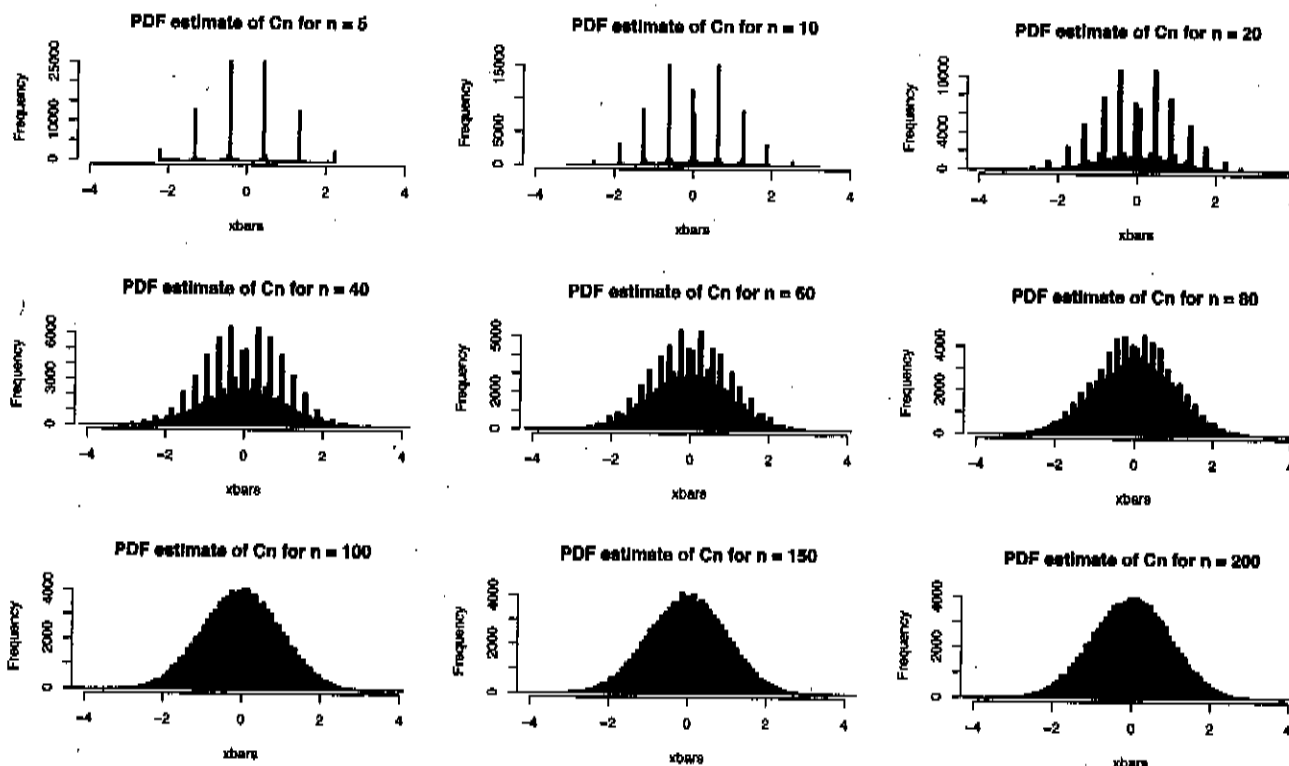
$$M_{\frac{Z_1}{\sqrt{2}} + \frac{Z_2}{\sqrt{2}}}(t) = \left(M_{\frac{Z}{\sqrt{2}}}(t)\right)^2 = \left(e^{t^2/4}\right)^2 = \boxed{e^{t^2/2}}$$

by independence  
and identical distr.  
of  $Z_1$  and  $Z_2$

- (e) [3 pt / 30 pts] If  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim}$  with some unknown distribution that has finite mean and variance, what is the distribution of the following quantity as  $n$  gets large? Write your answer to the right of the "distributed as" symbol. The brand name notation with the correct parameters clearly marked or the PDF is valid. No other answer is valid.

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1) \quad \left(\text{by the central limit theorem}\right)$$

- (f) [3 pt / 33 pts] We are interested in using confidence intervals and hypothesis testing for a sample and we have an idea of the distribution of the underlying r.v. which is sampled  $\overset{iid}{\sim}$  from a well-specified population. The following image is the result of some simulations. Below what *minimum* sample size <sup>would</sup> you not trust the validity of confidence intervals and hypothesis tests? Write  $n < \text{something}$  below the image.



I wouldn't trust

$$n < 100$$

**Problem 3** In this problem, we will be investigating a call center. Callers call in and are helped by a customer service representative and then answer a survey later if they're satisfied or not: "yes" or "no." Note that the vast majority of callers do not bother with the survey and just hang up, thus we don't know if they were satisfied or not. Assume the satisfaction outcome of all calls are  $\overset{iid}{\sim}$  Bernoulli r.v.'s.



- (a) [2 pt / 35 pts] The company is interested in investigating George, one of the customer service representatives, on the basis of surveyed customer satisfaction. They look at a random sample of 200 of George's calls and find that 168 are satisfied. Calculate  $\hat{p}$  to two decimal places.

$$\hat{p} = \frac{168}{200} = \boxed{0.84}$$

- (b) [5 pt / 40 pts] Provide a 95% confidence interval for the true proportion of customers satisfied by George's phone call on the post-call survey. Use correct notation for the CI with both subscripts clearly notated. Compute explicitly and round to 3 decimal places.

$$\begin{aligned} CI_{p, 1-\alpha} &= \left[ \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})} \right] \\ CI_{p, 95\%} &= \left[ 0.84 \pm 2 \sqrt{\frac{0.84(1-0.84)}{200}} \right] \\ &= \left[ 0.84 \pm 2 \cdot 0.0257 \right] \\ &= \boxed{[0.788, 0.892]} \end{aligned}$$

- (c) [3 pt / 43 pts] Employ one of the two correct interpretations of confidence intervals we discussed in class to interpret this interval appropriately in English. Do not write two interpretations, only one of the two interpretations.

① If we run ~~to~~ run many sampling experiments and compute the CI each time, 95% of these intervals would contain  $p$ , the true binomial proportion

—OR— ② Before running this sampling experiment, the probability of this interval containing  $p$  is 95%.

- (d) [2 pt / 45 pts] What can you NOT say about this interval that you may really want to say? Answer in English.

The probability of  $p$  being within  $[.778, .892]$  is 95%.

- (e) [2 pt / 47 pts] Does this also provide 95% coverage for the true proportion of all callers who satisfied with George's phone assistance? Explain why or why not. Yes or no without explanation will not receive credit.

No. Callers who do not wish to take the survey hang up.  
Thus, the estimate  $\hat{p}$  and its CI are biased.

- (f) [2 pt / 49 pts] The company is aiming for 90% customer satisfaction (as measured by the survey). If the customers are satisfied less than that proportion of the time, they will be unhappy. If they are satisfied more than 90% of the time, then representatives are probably spending too much time with each customer which will mean increased expenditure for the company. So having customer satisfaction *too low* is bad and having customer satisfaction *too high* is also bad albeit for different reasons. Write the null hypothesis and alternative hypothesis clearly below.

$$H_0: p = 0.9$$

$$H_a: p \neq 0.9$$

- (g) [5 pt / 54 pts] For a sample size of  $n = 200$ , what is the acceptance region at  $\alpha = 5\%$ ? Your answer must be in set notation. Compute explicitly and round to three decimals.

$$\text{Acc Region} = \left[ p \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right] = \left[ 0.9 \pm 2 \sqrt{\frac{0.9 \cdot 0.1}{200}} \right] = \left[ 0.9 \pm 0.0424 \right] = \boxed{[0.858, 0.942]}$$

$\alpha = 5\% \Rightarrow 1 - \alpha = 95\% \Rightarrow z_{\frac{\alpha}{2}} = 2$

- (h) [3 pt / 57 pts] Set your probability of Type I error to be 5%. Then, using the  $\hat{p}$  (the result of the experiment) you computed in (a), run a hypothesis test using the hypotheses you provided in (f) and the acceptance region you computed in (g) and finally interpret the result in English.

$$\hat{p} = 0.84 \notin [0.858, 0.942] \Rightarrow \text{Reject } H_0 \text{ and conclude}$$

that George's Customer Satisfaction rate is not 0.9.

- (i) [2 pt / 59 pts] [Extra Credit] Regardless of your answer to (g), the company has a policy of firing those people who do not have 90% customer satisfaction over their phone calls. What would a type II error be from the company's perspective? Answer in English.

Type II error is to retain the  $H_0$  when  $H_a$  is true. Here, it would be to keep George employed even though he should've been fired.

- (j) [4 pt / 63 pts] [Extra Credit] The company now wants to test against  $H_a : p = 0.7$  to find representatives who are not satisfying the customers. How much more likely is George to belong to the  $H_0$  distribution relative to the  $H_a$  distribution? Partial credit given. Do not attempt until you have finished the exam.

$$H_0: p = 0.9 \Rightarrow \hat{p} \sim N(0.9, 0.0212^2)$$

$$H_a: p = 0.7 \Rightarrow \hat{p} \sim N(0.7, 0.0324^2)$$

POF of normal distribution

$$R = \frac{f_{\hat{p}|H_0}(\hat{p})}{f_{\hat{p}|H_a}(\hat{p})} = \frac{\frac{1}{\sqrt{2\pi}} \frac{1}{0.0212} e^{-\frac{1}{2 \cdot 0.0212^2} (0.84 - 0.9)^2}}{\frac{1}{\sqrt{2\pi}} \frac{1}{0.0324} e^{-\frac{1}{2 \cdot 0.0324^2} (0.84 - 0.7)^2}} = \frac{0.015667}{2.4030 \times 10^{-7}} = \boxed{65,199} \text{ times more likely}$$



- (k) [5 pt / 68 pts] [Extra Credit] At  $\alpha = 5\%$ , find the probability of a Type II error given  $H_a : p = 0.7$ . Partial credit given. Do not attempt until you have finished the exam.

Type II error: rejects  $H_0$  when  $H_0$  is true

$$= P(\text{Reject} | H_0) \quad \text{Reject is } \hat{p} \geq .858$$

$$= P(\hat{p} \geq .858 | \hat{p} \sim N(0.7, .0329))$$

$$= P\left(\frac{\hat{p} - 0.7}{\sqrt{.0329}} \geq \frac{.858 - .7}{\sqrt{.0329}}\right)$$

$$= P(Z \geq 4.876) = 5.4 \times 10^{-7} \approx \boxed{0}$$

**Problem 4** In the standard American "pick up sticks" game, there are 30 sticks — 1 black, 7 red, 7 blue, 8 green and 7 yellow. The sticks are shuffled at random and allowed to drop on the floor in a messy pile strewn in random directions with random overlaps such as seen below:



Gameplay proceeds as follows. A player attempts to pick up a stick out of the pile without allowing any other stick to move. If they remove a stick without moving any other sticks, they collect that stick and score points and then they go again. Points are as follows: 25 points for black, 10 points for red, 5 points for blue, 2 points for green and one point for yellow. Number of sticks and points is summed up in Table 1.

| Stick color | Number of sticks | Point value |
|-------------|------------------|-------------|
| Black       | 1                | 25          |
| Red         | 7                | 10          |
| Blue        | 7                | 5           |
| Green       | 8                | 2           |
| Yellow      | 7                | 1           |
| Totals:     | 30               | 153         |

Table 1: Pick up Sticks game information

- (a) [2 pt / 70 pts] Assume a new random shuffle. I successfully remove the top three sticks off the pile. What is the probability of (in those three sticks) getting the black stick, a red stick and a blue stick *in that order*? Do *not* compute explicitly.

$$\frac{1}{30} \cdot \frac{7}{29} \cdot \frac{7}{28}$$

- (b) [5 pt / 75 pts] Assume a new random shuffle. At the end of the game, I have 15 sticks. What is the probability I have 3 red, 3 blue, 5 green and 4 yellow *in any order*? Do *not* compute explicitly.

Red Blue Green Yellow

$$\frac{\binom{7}{3} \binom{7}{3} \binom{8}{5} \binom{7}{4}}{\binom{30}{15}}$$

- (c) [2 pt / 77 pts] Assume a new random shuffle. At the end of the game, I have 15 sticks. Create a r.v. that models the number of red sticks I have out of my 15 collected sticks. You should write  $X \sim$  something below and write the name of the r.v. and clearly indicate the parameter(s) in this r.v. You do not have to write the PMF/PDF of this r.v.

$$n=15, K=7, N=30$$

$$X \sim \text{Hypergeometric}(15, 7, 30)$$

- (d) [2 pt / 79 pts] Assume a new random shuffle. Assume the probability of me having a stick at the end of the game is 0.5 regardless of its color. Create a r.v. that models 1 if I have the black stick and 0 if I do not have the black stick. You should write  $X \sim$  something below and write the name of the r.v. and clearly indicate the parameter(s) in this r.v.

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

- (e) [2 pt / 81 pts] Assume a new random shuffle. Assume the probability of me having a stick at the end of the game is 0.5 regardless of its color. Create a r.v. that models the number of sticks I have at the end of the game regardless of their color. You should write  $X \sim$  something below and write the name of the r.v. and clearly indicate the parameter(s) in this r.v.

$$X \sim \text{Binomial}(30, 0.5)$$

- (f) [2 pt / 83 pts] Assume a new random shuffle. Assume the probability of me having a stick at the end of the game is 0.5 regardless of its color. Find the probability I wind up with 15 sticks at the end of the game. Do *not* compute explicitly but the number must be computable (no  $\infty$ 's allowed).

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for a binomial r.v. generally}$$

$$f_X(15) = \binom{30}{15} 0.5^{15} 0.5^{15} = \boxed{\binom{30}{15} 0.5^{30}}$$

- (g) [3 pt / 86 pts] Assume a new random shuffle. Assume the probability of me having a stick at the end of the game is 0.5 regardless of its color. Find the probability I wind up with 20 or more sticks at the end of the game. Do *not* compute explicitly but the number must be computable (no  $\infty$ 's allowed).

$$P(X \geq 20) = \sum_{x=20}^{30} \binom{30}{x} 0.5^x 0.5^{30-x} = \boxed{0.5^{30} \sum_{x=20}^{30} \binom{30}{x}}$$

both  
answers  
acceptable

- (h) [4 pt / 90 pts] Assume a new random shuffle. Assume the probability of me having a stick at the end of the game is 0.5 regardless of its color. Find the expected number of points I have from green sticks only. That is if you add up my score, this is the component of the score from my green sticks.

# of green sticks

$$N_G \sim \text{Binomial}(8, 0.5) \quad P_G = 2 \frac{\text{points}}{\text{stick}} N_G$$

$$E[P_G] = 2 \frac{\text{points}}{\text{stick}} E[N_G] = 2 \frac{\text{points}}{\text{stick}} (8 \text{ sticks})(0.5) = \boxed{8 \text{ points}}$$

- (i) [5 pt / 95 pts] [Extra Credit] Assume a new random shuffle. Assume the probability of me having a stick at the end of the game is 0.5 regardless of its color. Find the expected number of points I have at the end of the game. You can get partial credit on this extra credit.

$$\begin{aligned} N_G &\sim \text{Binomial}(8, 0.5) & P_G &= 2N_G \\ N_{\text{black}} &\sim \text{Binomial}(15, 0.5) & P_{\text{black}} &= 15N_{\text{black}} \\ N_R &\sim \text{Binomial}(2, 0.5) & P_R &= 10N_R \\ N_{\text{blue}} &\sim \text{Binomial}(7, 0.5) & P_{\text{blue}} &= 5N_{\text{blue}} \\ N_Y &\sim \text{Binomial}(7, 0.5) & P_Y &= 1N_Y \end{aligned}$$

$$\begin{aligned} P &= P_G + P_{\text{black}} + P_R + P_{\text{blue}} + P_Y \\ \Rightarrow E[P] &= 2E[N_G] + 15E[N_{\text{black}}] + 10E[N_R] + 5E[N_{\text{blue}}] + 1E[N_Y] \\ &= 2 \cdot 8 \cdot 0.5 + 15 \cdot 0.5 + 10 \cdot 2 \cdot 0.5 + 5 \cdot 7 \cdot 0.5 + 7 \cdot 0.5 \\ &= 0.5(16 + 15 + 20 + 25 + 7) \\ &= 0.5(73) = \boxed{21.5} \end{aligned}$$

This 15 should be a 25, thus the answer should be 76.5

- (j) [2 pt / 97 pts] Me and the other player are both really good at pick up sticks but it is still a difficult game. Each time we go to take a stick off, we have probability 20% that we successfully remove the stick without jiggling another stick. What is the probability we remove three sticks successfully in three turns? Hint: you do not need to consider the fact that there are two players here. Imagine there is just one player playing against himself.

$$(0.2)^3 = \boxed{.008}$$

both players  
acceptable

- (k) [3 pt / 100 pts] Assuming we have probability 20% that we successfully remove a stick without jiggling another stick, what is the probability we take six turns to successfully remove one stick? Compute explicitly and round to 3 decimal places.

$$X \sim \text{Geom}(p) := (1-p)^{x-1} p$$

$$X \sim \text{Geom}(0.2) = (1-0.2)^{x-1} 0.2 = 0.8^{x-1} 0.2$$

$$P(X=6) = 0.8^{6-1} 0.2 = \boxed{.0655}$$

- (l) [3 pt / 103 pts] Assuming we have probability 20% that we successfully remove a stick without jiggling another stick, what is the probability we take ten turns to successfully remove six sticks? You do *not* need to compute explicitly.

$$X \sim \text{Neg Bin}(r, p) := \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

$$X \sim \text{Neg Bin}(6, 0.2) := \binom{x-1}{5} 0.2^6 0.8^{x-6}$$

$$P(X=10) = \boxed{\binom{9}{5} 0.2^6 0.8^4}$$

- (m) [3 pt / 106 pts] Assuming we have probability 20% that we successfully remove a stick without jiggling another stick, what is the expected number of turns to complete the game? Compute explicitly.

$$X \sim \text{Neg Bin}(30, 0.2) \quad E(X) = \frac{r}{p} = \frac{30}{0.2} = \boxed{150}$$

↑  
# of sticks in  
game (total)

- (n) [2 pt / 108 pts] You play  $n$  games where  $n$  is really large. Assume  $n$  is large enough to establish probabilities of certain outcomes. What definition of probability are you using here? You do not need to write a sentence — a few words is fine.

" long run frequency definition "

- (o) [3 pt / 111 pts] The black stick is very helpful to winning the game of pick up sticks. Let  $B$  be the event you have the black stick at some point in the game in your possession and let  $W$  be the event that you win. Having the black stick seems to be 50-50 when looking at game data, so  $P(B) = 0.5$ . But if you have the black stick, you win 80% of the time. What is the probability you both win the game and have the black stick?

$$\begin{aligned}
 P(B) &= 0.5 \\
 P(W|B) &= 0.8 \\
 P(W, B) &= ?
 \end{aligned}
 \quad
 \begin{aligned}
 P(W|B) &= \frac{P(W, B)}{P(B)} \Rightarrow P(W, B) = P(B) P(W|B) \\
 &= 0.5 \cdot 0.8 \\
 &= \boxed{0.4}
 \end{aligned}$$

- (p) [5 pt / 116 pts] Looking at data, you also learn that the probability of both winning and *not* having the black stick is 10%. Given this new information and the information given in the previous problem as well as your answer to the previous problem, what is the probability among games you've won that you had the black stick?

$$P(W, B^c) = 0.1, \quad P(B|W) = ?$$

$$P(B|W) = \frac{P(W, B)}{P(W)} \quad \leftarrow \text{from last prob}$$

$$= \frac{0.4}{0.5} = \boxed{0.8}$$

$$\begin{aligned}
 &\downarrow \\
 &= P(W, B) + P(W, B^c) \quad \text{law of total prob.} \\
 &= 0.4 + 0.1 = 0.5
 \end{aligned}$$