Math 368 / 621 Fall 2019
Midterm Examination One

Solutions

Professor Adam Kapelner

Septmeber 25, 2019

Full Name ____________________________________ Circle Section and Class: A B C   368 621

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 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 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 the question reads "compute," this means the solution will be a number otherwise you can leave the answer in choose, permutation, exponent, factorial or any other notation which could be resolved to a number with a computer. Questions marked "[MA]" are required for those enrolled in 621 and extra credit for those enrolled in 368. If you are enrolled in 368, I advise you to finish the other questions on the exam and only then attempt the extra credit. 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 Below are some theoretical exercises.

(a) [6 pt / 6 pts] Compute as best as possible: $\sum_{x_1 \in \mathbb{R}} \sum_{x_2 \in \mathbb{R}} \mathbb{1}_{x_1 \in \{3,4,5\}} \mathbb{1}_{x_2 \in \{3,4,5\}}$.

9

(b) [6 pt / 12 pts] Let $X \sim \text{Binomial}(n, p) := \binom{n}{x} p^x (1-p)^{n-x}$ where $n \in \mathbb{N}$ and $p \in (0, 1)$. In class we said that this PMF was valid $\forall x \in \mathbb{R}$. If so, how did we define $\binom{n}{x}$ under the hood?

$$\binom{n}{x} := \frac{n!}{x!(n-x)!} \mathbb{1}_{x \in \{0,1,\ldots,n\}}$$

(c) [10 pt / 22 pts] Let r.v.'s $X_1, X_2 \overset{iid}{\sim}$ with expectation μ and variance σ^2. Let $T = X_1 + X_2$ and $\boldsymbol{Y} = [X_1 \; X_2 \; T]^\top$. Find $\mathbb{V}\text{ar}[\boldsymbol{Y}]$. Simplify to be as clean as possible.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1T} \\ \sigma_{12} & \sigma_2^2 & \sigma_{2T} \\ \sigma_{1T} & \sigma_{2T} & \sigma_T^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \sigma^2 \\ 0 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & 2\sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$\sigma_{12} = 0$

$\sigma_{1T} = \text{Cov}[X_1, X_1 + X_2] = \text{Cov}[X_1, X_1] + \text{Cov}[X_1, X_2] = \sigma^2 + 0$

$\sigma_{2T} = \text{Cov}[X_2, X_1 + X_2] = \text{Cov}[X_2, X_1] + \text{Cov}[X_2, X_2] = 0 + \sigma^2$

$\sigma_T^2 = \sigma^2 + \sigma^2 = 2\sigma^2$

(d) [6 pt / 28 pts] [MA] Consider $X, Y \overset{iid}{\sim} \text{Geometric}(p)$ and find $\mathbb{P}(X \geq Y + c)$ as a function of p and c where $c \in \mathbb{N}_0$, a constant.

$$= \sum_{y \in \mathbb{R}} \sum_{x \in \mathbb{R}} p_{X,Y}(x,y) \mathbb{1}_{x \geq y+c} = \sum_{y \in \mathbb{R}} \sum_{x \in \mathbb{R}} p(x) p(y) \mathbb{1}_{x \geq y+c} = \sum_{y=0}^{\infty} \sum_{x=0}^{\infty} ((1-p)^x p)((1-p)^y p) \mathbb{1}_{x \geq y+c}$$

$$= p^2 \sum_{y=0}^{\infty} (1-p)^y \sum_{x=0}^{\infty} (1-p)^x \mathbb{1}_{x \geq y+c} = p^2 \sum_{y=0}^{\infty} (1-p)^y \sum_{x=y+c}^{\infty} (1-p)^x = p^2 \sum_{y=0}^{\infty} (1-p)^y \sum_{x'=0}^{\infty} \overbrace{(1-p)^{x'+y+c}}^{(1-p)^{x'}(1-p)^c(1-p)^y}$$

let $x' = x - (y+c) \Rightarrow x = x' + y + c$

$$= p^2 (1-p)^c \underbrace{\sum_{y=0}^{\infty} (1-p)^{2y}}_{\frac{1}{1-(1-p)^2} = \frac{1}{p(2-p)}} \underbrace{\sum_{x'=0}^{\infty} (1-p)^{x'}}_{\frac{1}{p}} = \frac{p(1-p)^c}{p(2-p)} = \frac{(1-p)^c}{(2-p)}$$

$-(1-p)^2$

$-(1-2p+p^2)$

$2p - p^2 = p(2-p)$

X and Y are independent.

(e) [6 pt / 34 pts] [MA] Let $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ and $Y \sim \text{U}\left(\{c, c+2, c+4, \ldots, c+2m\}\right)$ where $c \in \mathbb{R}$ and $m \in \mathbb{N}_0$ are both constants. Find the PMF of $T = X + Y$ using a convolution formula. Is the final answer a brand name r.v. you know? If so, denote it using the notation we learned in class.

$$p_T(t) = \sum_{x \in \text{Supp}[X]} p_X(x)\, p_Y(t-x)\, \mathbb{1}_{t-x \in \text{Supp}[Y]} = \sum_{x \in \{0,1\}} \left(\frac{1}{2}\right)\left(\frac{1}{m+1} \mathbb{1}_{t-x \in \{c, c+2, \ldots, c+2m\}}\right)$$

$$= \frac{1}{2(m+1)} \left(\mathbb{1}_{t \in \{c, c+2, \ldots, c+2m\}} + \mathbb{1}_{t \in \{c+1, c+3, \ldots, c+2m+1\}}\right)$$

$$= \frac{1}{2(m+1)} \mathbb{1}_{t \in \{c, c+1, \ldots, c+2m+1\}} \qquad \text{since } \mathbb{1}_A + \mathbb{1}_B = \mathbb{1}_{A \cup B} \text{ if } A \cap B = \emptyset$$

$$= U\left(\{c, c+1, \ldots, c+2m+1\}\right)$$

Problem 2 Consider the following bag of fruit with 5 apples, 6 bananas, 2 oranges and 1 grape cluster.

$$\vec{p} = \frac{1}{14}\begin{bmatrix} 5 \\ 6 \\ 2 \\ 1 \end{bmatrix}$$

Let the vector r.v. $\boldsymbol{X} = [X_1, X_2, X_3, X_4]^\top$ represent a sample of 20 fruits uniformly at random *with replacement* from a bag where X_1 denotes the number of apples sampled, X_2 denotes the number of bananas sampled, X_3 denotes the number of oranges sampled and X_4 denotes the number of grape clusters sampled. Assume this sampling procedure and this notation for all questions in this problem.

(a) [7 pt / 41 pts] Find the JMF of $\boldsymbol{X}$ in a format that can be computed. Do not use choose notation. No brand name notation allowed in the final answer. Assume $\boldsymbol{x} \in \text{Supp}[\boldsymbol{X}]$ so no need to ensure your answer is valid $\forall \boldsymbol{x} \in \mathbb{R}^K$. Please write all parameter values explicitly.

$$\vec{X} \sim \text{Multi}\left(n = 20, \vec{p} = \frac{1}{14}\begin{bmatrix} 5 \\ 6 \\ 2 \\ 1 \end{bmatrix}\right) = \binom{20}{x_1, x_2, x_3, x_4} \left(\frac{5}{14}\right)^{x_1} \left(\frac{6}{14}\right)^{x_2} \left(\frac{2}{14}\right)^{x_3} \left(\frac{1}{14}\right)^{x_4}$$

$$= \frac{20!}{x_1!\, x_2!\, x_3!\, x_4!} \left(\frac{5}{14}\right)^{x_1} \left(\frac{6}{14}\right)^{x_2} \left(\frac{2}{14}\right)^{x_3} \left(\frac{1}{14}\right)^{x_4}$$

(b) [7 pt / 48 pts] What is the probability you draw 3 apples, 10 bananas, 5 oranges and 2 grape clusters? Do not compute explicitly but leave your answer in a format that a computer can do the calculation. Do not use choose notation.

$$P\left(\vec{X} = \begin{bmatrix} 3 \\ 10 \\ 5 \\ 2 \end{bmatrix}\right) = \frac{20!}{3!\,10!\,5!\,2!} \left(\frac{5}{14}\right)^3 \left(\frac{6}{14}\right)^{10} \left(\frac{2}{14}\right)^5 \left(\frac{1}{14}\right)^2$$

(c) [5 pt / 53 pts] What is the probability you draw 7 apples, 8 bananas, 3 oranges and 3 grape clusters? Do not compute explicitly but leave your answer in a format that a computer can do the calculation. Do not use choose notation.

0 since $\vec{x} \notin Supp(\vec{X})$ $\quad 7+8+3+3 \neq 20$

(d) [5 pt / 58 pts] What is the probability you draw 5 apples? Do not compute explicitly but leave your answer in a format that a computer can do the calculation. Do not use choose notation.

$$X_1 \sim Bin\left(20, \frac{5}{14}\right), \quad P(X_1 = 5) = \frac{20!}{5!\,15!} \left(\frac{5}{14}\right)^5 \left(\frac{9}{14}\right)^{15}$$

(e) [8 pt / 66 pts] Knowing that you drew 5 apples, what is the probability you drew 3 bananas? Do not compute explicitly but leave your answer in a format that a computer can do the calculation without using choose notation.

$$\vec{X}_{-1} \sim Multi\left(n - x_1, \vec{p}/1-p_1\right) = Multi\left(15, \frac{1}{1-\frac{5}{14}} \begin{bmatrix} 6/14 \\ 3/14 \\ 1/14 \end{bmatrix}\right)$$

$$X_{-1,2} \sim Bin\left(n - x_1, \frac{p_2}{1-p_1}\right) = Bin\left(15, \frac{\frac{6}{14}}{1-\frac{5}{14}} = \frac{2}{3}\right)$$

$$P(X_{-1,2} = 3) = \frac{15!}{3!\,12!} \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^{10}$$

(f) [5 pt / 71 pts] You cannot sell apples and bananas, but you can sell each orange for \$0.50 and each grape cluster for \$2.00. Let R be the r.v. the models sale revenue on the sample of 20 fruits from part (a). Compute $\mathbb{E}[R]$ to the nearest cent.

$$R = 0 \cdot X_1 + 0 \cdot X_2 + \$0.5 \cdot X_3 + \$2 \cdot X_4$$

$$\Rightarrow E(R) = \$0.50\, \underbrace{E[X_3]}_{np_3} + \$2 \cdot \underbrace{E(X_4)}_{np_4} = \$0.50 \cdot 20 \cdot \frac{2}{14} + \$2 \cdot 20 \cdot \frac{1}{14} = \$4.26$$

(g) [7 pt / 78 pts] Write an expression to compute $\mathbb{V}\text{ar}[R]$. Factor out constants but leave in terms of matrix and vector multiplication. Do not compute.

$$\sigma_R^2 = c^T \Sigma c = \begin{bmatrix} 0 & 0 & 0.5 & 2 \end{bmatrix} \Sigma \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix} = \begin{bmatrix} np_1(1-p_1) & -np_1p_2 & -np_1p_3 & -np_1p_4 \\ -np_1p_2 & np_2(1-p_2) & -np_2p_3 & -np_2p_4 \\ -np_1p_3 & -np_2p_3 & np_3(1-p_3) & -np_3p_4 \\ -np_1p_4 & -np_2p_4 & -np_3p_4 & np_4(1-p_4) \end{bmatrix} = \frac{20}{14^2} \begin{bmatrix} 5.9 & -5.6 & -5.2 & -5.1 \\ -5.6 & 6.8 & -6.2 & -6.1 \\ -5.2 & -6.2 & 2.12 & -2.1 \\ -5.1 & -6.1 & -2.1 & 1.11 \end{bmatrix}$$

$$= \frac{20}{14^2} \begin{bmatrix} 45 & -30 & -10 & -5 \\ -30 & 48 & -12 & -6 \\ -10 & -12 & 24 & -2 \\ -5 & -6 & -2 & 11 \end{bmatrix}$$

Problem 3 In soccer, each team scores a number of "goals" and whoever has the most goals wins. We will consider modeling soccer game point spreads. This is the number of goals one time wins (or loses) by. We will be using the Skellam distribution:

$$D \sim \text{Skellam}(\lambda_1, \lambda_2) := e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{d/2} I_{|d|}(2\sqrt{\lambda_1\lambda_2}) \mathbb{1}_{d\in\mathbb{Z}}$$

where $I_c(a)$ denotes the modified Bessel function of the first kind.

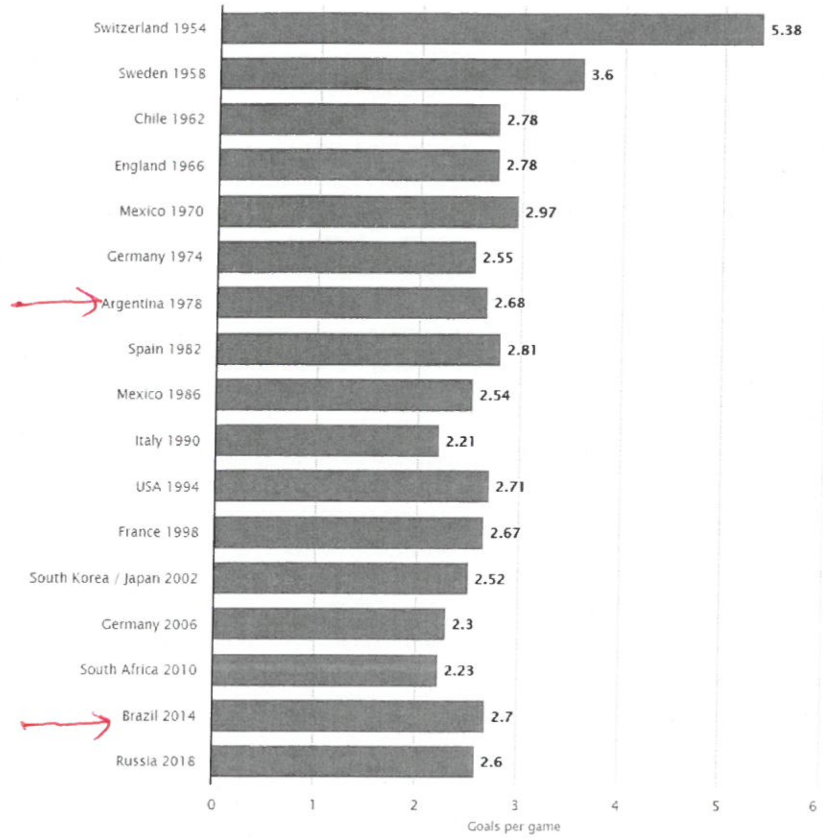
(a) [6 pt / 84 pts] Let $R = -D$ and prove that $R \sim \text{Skellam}(\lambda_2, \lambda_1)$.

$$p_R(r) = p_D(-r) = e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{-r}{2}} I_{|-r|}(2\sqrt{\lambda_1\lambda_2}) \mathbb{1}_{-r\in\mathbb{Z}}$$

$$= e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{r}{2}} I_{|r|}(2\sqrt{\lambda_1\lambda_2}) \mathbb{1}_{r\in\mathbb{Z}} = \text{Skellam}(\lambda_2, \lambda_1)$$

The website linked here compiled statistics on each soccer team for almost 90 years. Below are average number of goals for many countries' teams that won the world cup since 1950:

This is not a class on statistics, so we will not talk about estimation. We will assume (a) the number of goals scored in any game is independent and Poisson-distributed and (b) the average number of goals above is exactly the λ parameter in the Poisson distribution for each team.



(b) [5 pt / 89 pts] If Brazil plays Argentina, who is expected to win and by how much? Brazil by

$D = X_{Brazil} - X_{Argentina}$ $E[D] = \lambda_{Brazil} - \lambda_{Argentina} = 2.7 - 2.68 = 0.02$

(c) [6 pt / 95 pts] Find an expression for the probability that Brazil beats Argentina by 2. You can leave your answer in terms of $I_c(a)$, the modified Bessel function of the first kind. Simplify to be as clean as possible.

$$P(D=2) = e^{-(2.7+2.68)}\left(\frac{2.7}{2.68}\right)^{\frac{2}{2}} I_{|2|}\left(2\sqrt{2.7 \cdot 2.68}\right) = e^{-5.38}(1.0075)\, I_2(5.38)$$
$$= 0.005\, I_2(5.38)$$

(d) [5 pt / 100 pts] [MA] The Skellam model actually has a small problem: you cannot tie in soccer (meaning both teams get the same score), so $d = 0$ *should not be* in the support of our final modeling distribution. With this fact in mind, find an expression for the the probability that Brazil beats Argentina by 2. You can leave your answer in terms of $I_c(a)$, the modified Bessel function of the first kind. Simplify to be as clean as possible.

$$P(D=2 \mid D \neq 0) = \frac{P(D=2 \,\&\, D \neq 0)}{P(D \neq 0)} = \frac{P(D=2)}{1 - P(D=0)} = \frac{0.005\, I_2(5.38)}{1 - e^{-(2.7+2.68)} I_0(5.38)}$$

$$= \frac{0.005\, I_2(5.38)}{1 - 0.005\, I_0(5.38)}$$