MATH 340/640 Fall 2023 Homework #3

Professor Adam Kapelner

Due by email 11:59PM October 7, 2023

(this document last updated Friday 22nd September, 2023 at 5:04pm)

Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out." Feel free to "work out" with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, review Math 241 concerning random variables, support, parameter space, PMF's, the central limit theorem (CLT), variance, covariance.

The problems below are color coded: green problems are considered *easy* and marked "[easy]"; yellow problems are considered *intermediate* and marked "[harder]", red problems are considered *difficult* and marked "[difficult]" and purple problems are extra credit. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems. "[MA]" are for those registered for 621 and extra credit otherwise.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 5 points are given as a bonus if the homework is typed using LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `overleaf.com`. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: ____________________

Problem 1

These exercises will introduce implications of the CLT.

(a) [easy] State the CLT with all of its assumptions.

(b) [easy] State the asymptotic distribution of $\bar{X}_n$, the average of the rv's.

(c) [easy] State the asymptotic distribution of T_n, the total of the rv's.

(d) [harder] According to this site, the S&P500 delivers an average of 7.7% per year with a standard deviation of 19.1%. What is the probability the average yearly return over 20yr is positive? Note: this is a different question from "do you make money over 20yr given an initial investment"? We will learn how to answer that latter question later in the semester when we learn about the LogNormal rv.

Problem 2

These exercises will introduce the Multinomial distribution.

(a) [easy] If $\boldsymbol{X} \sim \text{Multinomial}\,(n, \boldsymbol{p})$ where $\dim\,[\boldsymbol{X}] = k$, what is the parameter space for both n and $\boldsymbol{p}$?

(b) [easy] If $\boldsymbol{X} \sim \text{Multinomial}\left(n,\, \boldsymbol{p}\right)$ where $\dim\left[\boldsymbol{X}\right] = k$, what is $\mathbb{S}_{\boldsymbol{X}}$?

(c) [easy] If $\boldsymbol{X} \sim \text{Multinomial}\left(n,\, \boldsymbol{p}\right)$ where $\dim\left[\boldsymbol{X}\right] = k$, what is $\dim\left[\boldsymbol{p}\right]$?

(d) [easy] If $\boldsymbol{X} \sim \text{Multinomial}\left(n,\, \boldsymbol{p}\right)$ where $\dim\left[\boldsymbol{X}\right] = 2$, express p_2 as a function of p_1.

(e) [easy] If $\boldsymbol{X} \sim \text{Multinomial}\left(n,\, \boldsymbol{p}\right)$ where $\dim\left[\boldsymbol{X}\right] = 2$, prove that $X_2 \sim \text{Binomial}\left(n,\, p_2\right)$.

(f) [easy] If $\boldsymbol{X} \sim \text{Multinomial}\left(n,\, \boldsymbol{p}\right)$ and $n = 1$ and $\dim\left[\boldsymbol{X}\right] = 7$ as a column vector, give an example value of $\boldsymbol{x}$, a realization of the r.v. $\boldsymbol{X}$. Use the notation $[\ldots]^\top$ to write it as a row vector transposed.

(g) [easy] If $\boldsymbol{X} \sim \text{Multinomial}\left(n,\, \boldsymbol{p}\right)$ and $n = 10$ and $\dim\left[\boldsymbol{X}\right] = 7$ as a column vector, give an example value of $\boldsymbol{x}$, a realization of the r.v. $\boldsymbol{X}$. Use the notation $[\ldots]^\top$ to write it as a row vector transposed.

(h) [harder] Is a binomial rv a multinomial rv? Yes or no and explain. This is subtle.

(i) [harder] If $\boldsymbol{X} \sim \text{Multinomial}\left(n,\, \boldsymbol{p}\right)$ where $\dim\left[\boldsymbol{X}\right] = k$, prove that the sum over the JMF is 1. To do this, use the multinomial theorem.

(j) [difficult] [MA] Let $\boldsymbol{X}_1, \boldsymbol{X}_2 \overset{iid}{\sim} \text{Multinomial}\left(n,\, \boldsymbol{p}\right)$ with $\dim\left[\boldsymbol{X}_1\right] = \dim\left[\boldsymbol{X}_2\right] = k$. Find the JMF of $\boldsymbol{T}_2 = \boldsymbol{X}_1 + \boldsymbol{X}_2$ from the definition of convolution. This looks harder than it is! First, use the definition of convolution and factor out the terms that are not a function of $x_1, \ldots, x_K$. Finally, use Theorem 1 in this paper: [click here] for the summation.

(k) [harder] Explain in English why the following should be true. Remember how the sampling from the bag works.

$$\binom{n}{x_1, x_2, \ldots, x_K} = \binom{n}{x_1}\binom{n - x_1}{x_2}\binom{n - (x_1 + x_2)}{x_3} \cdot \ldots \cdot \binom{n - (x_1 + x_2 + \ldots x_{K-1})}{x_K}$$

(l) [easy] Consider the following bag of 4 green, 3 red, 2 blue and 1 yellow marbles:

Draw one marble with replacement 37 times. What is the probability of getting 10 red, 17 green, 6 blue and 4 yellow? Compute explicitly to the nearest two significant digits.

(m) [E.C.] [MA] If $\boldsymbol{X} \sim \text{Multinomial}\,(n, \boldsymbol{p})$, find the JMF of any subset of $X_1, \ldots, X_k$. Is it technically multinomial? This is not much harder than the previous problem if formulated carefully.

Problem 3

These exercises will introduce the concept of covariance (the metric that gauges linear dependence between two rv's).

(a) [easy] Prove that $\mathbb{C}\text{ov}\left[X_1, X_2\right] = \mathbb{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right]$.

(b) [easy] Prove that $\mathbb{C}\text{ov}\left[X_1, X_2\right] = \mathbb{C}\text{ov}\left[X_2, X_1\right]$.

(c) [easy] Prove that $\mathbb{C}\text{ov}\left[X_1 + X_3, X_2\right] = \mathbb{C}\text{ov}\left[X_1, X_2\right] + \mathbb{C}\text{ov}\left[X_3, X_2\right]$.

(d) [difficult] Prove that

$$\mathbb{V}\text{ar}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\left[X_i, X_j\right].$$

Hint: use induction and play with the sum notation.

(e) [harder] [MA] Prove that

$$\mathbb{C}\text{ov}\left[\sum_{i \in A} X_i, \sum_{j \in B} Y_j\right] = \sum_{i \in A}\sum_{j \in B} \mathbb{C}\text{ov}\left[X_i, Y_j\right]$$

.

(f) [easy] Prove the Cauchy-Schwartz Inequality.

(g) [easy] Prove the Covariance Inequality by invoking the Cauchy-Schwartz Inequality.

(h) [harder] Let Q be a non-negative, non-degenerate discrete rv. Prove $\mathbb{E}[Q] > 0$.

(i) [harder] If $\boldsymbol{X} \sim \text{Multinomial}(n, \boldsymbol{p})$ where $\dim[\boldsymbol{X}] \geq 2$, compute an upper bound for the covariance of X_1 and X_2. Hint, to get the best possible upper bound, use the fact that we know this covariance must be negative.

(j) [harder] Correlation ρ is a normalized unitless covariance metric. It is defined for for any two rv's X_1 and X_2 as:

$$\rho_{1,2} := \text{Corr}[X_1, X_2] := \frac{\sigma_{1,2}}{\sigma_1 \sigma_2} = \frac{\mathbb{C}\text{ov}[X_1, X_2]}{\sqrt{\mathbb{V}\text{ar}[X_1]\,\mathbb{V}\text{ar}[X_2]}}.$$

Prove that $\rho \in [-1, 1]$ for any two rv's X_1 and X_2.