MATH 368/621 Fall 2019 Homework #2

Professor Adam Kapelner

Due under the door of KY604 11:59PM Friday, September 20, 2019

(this document last updated Wednesday 11^{th} September, 2019 at 8:54pm)

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 from math 241 about conditional probability, expectation and variance then read on your own about the multinomial distribution, conditional vector expectation, covariances, variance-covariance matrices.

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.

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 10 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 review expectation and variance and introduce covariance as well as expectation and variance of multidimensional (vector) r.v.'s.

(a) [harder] Consider a sequence of independent r.v.'s $X_1, \ldots, X_n$ and prove that

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}\left[X_i\right].$$

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

(c) [easy] Prove that $\mathbb{C}\text{ov}\left[X, X\right] = \mathbb{V}\text{ar}\left[X\right]$.

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

(e) [easy] Prove that $\mathbb{C}\text{ov}\left[a_1 X_1, a_2 X_2\right] = a_1 a_2 \mathbb{C}\text{ov}\left[X_1, X_2\right]$.

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

(g) [harder] [MA] Prove that

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

.

(h) [harder] [MA] Prove that

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

without using the vector formulas.

(i) [easy] Prove $\mathbb{E}[\boldsymbol{X} + \boldsymbol{c}] = \boldsymbol{\mu} + \boldsymbol{c}$ where $\boldsymbol{c} \in \mathbb{R}^K$, a constant.

(j) [easy] Prove $\mathbb{V}\mathrm{ar}\left[\boldsymbol{c}^\top \boldsymbol{X}\right] = \boldsymbol{c}^\top \Sigma \boldsymbol{c}$ where $\boldsymbol{c} \in \mathbb{R}^K$, a constant and $\Sigma := \mathbb{V}\mathrm{ar}[\boldsymbol{X}]$, the variance-covariance matrix of the vector r.v. $\boldsymbol{X}$. This is marked easy since it's in the notes.

(k) [easy] Why is $\boldsymbol{c}^\top \Sigma \boldsymbol{c}$ called a "quadratic form?" Read about it on wikipedia.

Problem 2

These exercises are about the Multinomial distribution.

(a) [easy] Explain in English why $\boldsymbol{B} \sim \text{Multinomial}(1, \boldsymbol{p})$ is the multidimensional generalization of the Bernoulli r.v.

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

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

(c) [harder] Prove the combinatorial identity in (a).

(d) [easy] Consider the following bag of marbles.

Draw from 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.

(e) [difficult] [MA] If $\boldsymbol{X} \sim \text{Multinomial}\,(n,\, \boldsymbol{p})$, prove that its JMF sums to one, i.e. $\sum_{\boldsymbol{x} \in \text{Supp}[\boldsymbol{X}]} p_{\boldsymbol{X}}(\boldsymbol{x}) = 1$.

(f) [difficult] [MA] If $\boldsymbol{X} \sim \text{Multinomial}\,(n,\, \boldsymbol{p})$, prove that any marginal distribution is binomial with n and p_j as parameters i.e.

$$p_{X_j}(x_j) = \sum_{\boldsymbol{x}_{-j} \in \text{Supp}[\boldsymbol{X}_{-j}]} p_{\boldsymbol{X}}(\boldsymbol{x}) = \text{Binomial}\,(n,\, p_j)$$

We only assumed this in class because it makes sense conceptual given balls being sampled from an urn, but it was never explicitly proven.

(g) [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.

(h) [easy] Explain in English why $\boldsymbol{B} \sim \text{Multinomial}\,(1,\, \boldsymbol{p})$ is the multidimensional generalization of the Bernoulli r.v.

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

$$\boldsymbol{B}_1, \ldots, \boldsymbol{B}_n \overset{iid}{\sim} \text{Multinomial}\,(1,\, \boldsymbol{p}) \quad \text{then} \quad \boldsymbol{X} := \sum_{i=1}^{n} \boldsymbol{B}_i \sim \text{Multinomial}\,(n,\, \boldsymbol{p})$$

(j) [harder] Find the answer by reasoning in English. No need to prove mathematically.

$$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r \overset{iid}{\sim} \text{Multinomial}\,(n,\, \boldsymbol{p}) \quad \text{then} \quad \boldsymbol{T} := \sum_{i=1}^{r} \boldsymbol{X}_i \sim \ ?$$

(k) [easy] If $\boldsymbol{X} \sim \text{Multinomial}\,(n,\, \boldsymbol{p})$, find $p_{\boldsymbol{X}_{-j} | X_j}(\boldsymbol{x}_{-j}, x_j)$. This is marked easy since it's in the notes.

(l) [E.C.] [MA] If $\boldsymbol{X} \sim \text{Multinomial}\,(n,\, \boldsymbol{p})$, find a proof for $\mathbb{C}\text{ov}\left[X_i, X_j\right] = -np_ip_j$ that is qualitatively different than the one we did in class.

(m) [harder] If $\boldsymbol{X} \sim \text{Multinomial}(n, \boldsymbol{p})$ where $\dim[\boldsymbol{X}] = K$ and $\boldsymbol{p} = \frac{1}{K}\mathbf{1}_K$. What is the limit of $\mathbb{C}\text{ov}[X_i, X_j]$ as K gets large but n is fixed. Why does this make sense?

(n) [easy] Correlation ρ is a unitless measure bounded between $[-1, 1]$ and is a type of normalized covariance metric. It is defined for two r.v.'s as

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

where $\mathbb{SE}[\cdot]$ denotes the standard error of a r.v., the square root of its variance. Find $\text{Corr}[X_i,\, X_j]$ for two arbitrary elements in the r.v. vector $\boldsymbol{X} \sim \text{Multinomial}(n, \boldsymbol{p})$.