Homework_04

October 1, 2019

1 Homework 4

1.0.1 Instructions

Your homeworks have two components: a written portion and a portion that also involves code. Written work should be completed on paper, and coding questions should be done in the notebook. You are welcome to LaTeX your answers to the written portions, but staff will not be able to assist you with LaTeX related issues. It is your responsibility to ensure that both components of the homework are submitted completely and properly to Gradescope. Refer to the bottom of the notebook for submission instructions.

```
In [1]: # Run this cell to set up your notebook

        import numpy as np
        from scipy import stats
        from scipy import special
        from datascience import *
        from prob140 import *

        # These lines do some fancy plotting magic
        import matplotlib
        %matplotlib inline
        import matplotlib.pyplot as plt
        plt.style.use('fivethirtyeight')

        # These lines make warnings look nicer
        import warnings
        warnings.simplefilter('ignore', FutureWarning)
```

1.1 1. Aces and Face Cards

A standard deck consists of 52 cards of which 4 are aces, 4 are kings, and 12 (including the four kings) are "face cards" (Jacks, Queens, and Kings).

Cards are dealt at random without replacement from a standard deck till all the cards have been dealt.

Find the expectation of the following. Each can be done with almost no calculation if you use symmetry.

a) The number of aces among the first 5 cards

b) The number of face cards that *do not* appear among the first 13 cards

c) The number of aces among the first 5 cards minus the number of kings among the last 5 cards

d) The number of cards before the first ace

e) The number of cards strictly in between the first ace and the last ace

f) The number of face cards before the first ace

1.1.1 [Solution] Aces and Face Cards

Key observations, all by symmetry (though you can calculate by counting permutations if you want):

(1) Specify a type of card and specify a position. The chance that the specified type appears in the specified position is $\frac{\text{num type}}{52}$ where num type is the number of that type of card in the deck.

(2) Specify a set of n positions. The chance that a specified card appears somewhere in those n positions is $\frac{n}{52}$.

(3) The number of cards of Type T in a hand of n cards is a sum of n indicators; $I_j = 1$ if Card j in the hand is of Type T, and 0 otherwise.

Throughout, you will be using additivity of expectation, which does not depend on the nature of the dependence between the variables being added.

a) $5 \times \frac{4}{52}$. Apply (4) with T = ace, and then (1).

b) $12 \times \frac{39}{52}$. The variable is the number of face cards that are among the 39 cards that are not selected. Apply (4) with T = face card and the "hand" being the 39 cards that are not selected.

Alternatively, let X be the variable and $Y = 12 - X$ the number of face cards that are in the hand. Then $E(Y) = 13 \times \frac{12}{52}$ analogously to (a); now subtract from 12.

c) 0. "First five" versus "last five" makes no difference, by symmetry. Apply (a) twice to see that the difference is 0.

d) $\frac{48}{5}$. There are 48 non-ace cards, and if we imagine the aces distributed evenly among the 52 cards, there are 5 "spacings" or "gaps" in which these non-ace cards can fall. By symmetry, $\frac{48}{5}$ cards are expected to fall into each gap.

e) $2 + 3 \times \frac{48}{5}$. There are 3 spacings between the first and last ace, as well as 2 aces.

f) $\frac{12}{5}$. Same as **d** for the deck consisting of just the face cards and the aces.

1.2 2. Unbiased Estimators

a) A population of known size N contains an unknown number G of good elements. Let X be the number of good elements in a simple random sample of size n drawn from this population. Use X to construct an unbiased estimator of G.

See the example in Section 8.2 for a refresher: http://prob140.org/textbook/chapters/Chapter_08/02_Addit

b) Would your answer to Part **a** have been different if X had been the number of good elements in a random sample drawn with replacement from the population? Why or why not?

c) A flattened die lands 1 and 6 with chance $p/2$ each, and the other faces 2, 3, 4, and 5 with chance $(1-p)/4$ each. Here $p \in (0,1)$ is an unknown number. Let $X_1, X_2, \ldots, X_n$ be the results of n rolls of this die. First find $E(|X_1 - 3.5|)$, and use the answer to construct an unbiased estimator of p based on all of $X_1, X_2, \ldots, X_n$.

1.2.1 [Solution] Unbiased Estimators

a) The goal is to construct a random variable Y so that $E(Y) = G$. Start with what you know: $E(X) = n\frac{G}{N}$. Now rearrange the constants to just have G on the right hand side: $\frac{NE(X)}{n} = G$. So take $Y = \frac{NX}{n}$. Then $E(Y) = \frac{NE(X)}{n} = G$.

b) No because $E(X) = n\frac{G}{N}$ for both sampling schemes.

c) $E(|X_1 - 3.5|) = \sum_{k=1}^{6} |k - 3.5| P(X_1 = k) = 2(2.5 \cdot \frac{p}{2}) + 2(1.5 \cdot \frac{1-p}{4}) + 2(0.5 \cdot \frac{1-p}{4}) = 1.5p + 1$

Unbiased estimator of p**:** The goal is to construct a random variable Y that satisfies $E(Y) = p$. As always, start with what you know.

Let $W_i = |X_i - 3.5|$ and $S = \sum_{i=1}^{n} W_i$. This incorporates the entire sample.

$E(W_i) = 1.5p + 1$ for each i, so $E(S) = n(1.5p + 1)$.

Rearrange the constants so that you just have p on the right hand side: $\frac{1}{1.5}(\frac{E(S)}{n} - 1) = p$.

So take $Y = \frac{1}{1.5}(\frac{S}{n} - 1)$. Then $E(Y) = p$ by linearity, so Y is unbiased for p.

In terms of the original sample, $Y = \frac{1}{1.5n}\sum_{i=1}^{n} |X_i - 3.5| - \frac{1}{1.5}$.

1.3 3. Poisson Moments

Work out Example 3 of Section 8.3 of the textbook before you start this problem. Notice that we didn't say, "Read Example 3." We said, "Work out Example 3." Do the calculation yourself, don't just read it.

For $k = 1, 2, 3, \ldots$, the *kth moment* of a random variable X is defined as $E(X^k)$. The term comes from physics where the principle of moments is used in finding the center of gravity of a system. Remember that the expectation or first moment $E(X)$ is the center of gravity of the distribution of X.

Let X have the Poisson (μ) distribution. Find the expectations below and prove your answers.

a) $E(X + 1)$

b) $E(1/(X + 1))$

c) $E(X^3)$

1.3.1 [Solution] Poisson Moments

a) $\mu + 1$

b) $\sum_{k=0}^{\infty} \frac{1}{k+1} \cdot e^{-\mu}\frac{\mu^k}{k!} = e^{-\mu}\mu^{-1}\sum_{k=0}^{\infty} \frac{\mu^{k+1}}{(k+1)!} = e^{-\mu}\mu^{-1}(e^{\mu} - 1) = \frac{1-e^{-\mu}}{\mu}$

c) $E(X(X-1)(X-2)) = \sum_{k=3}^{\infty} e^{-\mu}\frac{\mu^k}{(k-3)!} = e^{-\mu}\mu^3\sum_{k=3}^{\infty}\frac{\mu^{k-3}}{(k-3)!} = e^{-\mu}\mu^3 e^{\mu} = \mu^3$

Now $X(X-1)(X-2) = X^3 - 3X^2 + 2X$ and $E(X^2) = \mu^2 + \mu$ by the Example cited in the exercise. So $E(X^3) = \mu^3 + 3(\mu^2 + \mu) - 2\mu = \mu^3 + 3\mu^2 + \mu$

1.4 4. Collecting Distinct Values

This problem is a workout in finding expectations by using all the tools at your disposal. If an answer doesn't appear to fit into a formula that has already been proven, it's a very good idea to try to write the variable as a sum of simpler variables.

a) A fair die is rolled n times. Find the expected number of times the face with six spots appears.

b) A fair die is rolled n times. Find the expected number of faces that *do not* appear, and say what happens to this expectation as n increases.

c) Use your answer to Part **b** to find the expected number of distinct faces that *do* appear in n rolls of a die.

d) Find the expected number of times you have to roll a die till you have seen all of the faces. This is a version of what is known as the *coupon collector's problem*.

1.4.1 [Solution] Collecting Distinct Values

a) Binomial $(n, 1/6)$. So $n/6$.

b) Let X = the number of faces that do not appear in the n rolls. Then X is the sum of six indicators, one per face:

$X = \sum_{j=1}^{6} I_j$ where $I_j = 1$ if Face j does *not* appear in the n rolls and 0 otherwise.

For each j, $P(I_j = 1) = (\frac{5}{6})^n$. So $E(X) = 6(\frac{5}{6})^n$, which goes to 0 as n gets large.

c) $6 - E(\text{number of faces that don't appear}) = 6 - 6(\frac{5}{6})^n$

d) Let X = the number of times I have to roll the die until I see all the faces. Then $X = 1 + W_2 + W_3 + W_4 + W_5 + W_6$ where:

- W_2 is the number of rolls after the first roll till I see the second new face
- W_3 is the number of rolls after I see two distinct faces till I see the third new face
- etc.

For each i, W_i is a geometric random variable with success probability $\frac{6-(i-1)}{6}$, and hence $E(W_i) = \frac{6}{6-i+1}$.

$$E(X) \;=\; 1 + \frac{6}{5} + \frac{6}{4} + \cdots + \frac{6}{1} \;\approx\; 14.7$$

1.4.2 5. Fun with Indicators: The Inclusion-Exclusion Formula

We guessed the general inclusion-exclusion formula (see Section 5.2 of the textbook) but we never proved it. Let's get that done.

a) Let $x_1, x_2, \ldots, x_n$ be numbers. Expand the product $(1 - x_1)(1 - x_2)$ and then expand $(1 - x_1)(1 - x_2)(1 - x_3)$ by using the expansion you got for $(1 - x_1)(1 - x_2)$. Now guess a formula for the expansion of the product

$$\prod_{i=1}^{n}(1 - x_i)$$

and use induction to prove it. The induction shouldn't take many steps. It consists of just two observations, both of which can be expressed in English without complicated notation.

b) Let $A_1, A_2, \ldots, A_n$ be events. For each i in the range 1 through n let I_i be the indicator of A_i. Let I be the indicator of $\cup_{i=1}^{n} A_i$. Explain why

$$I \;=\; 1 - \prod_{i=1}^{n}(1 - I_i)$$

c) Use Parts **a** and **b** to establish the inclusion-exclusion formula.