March 5, 2020

Math 318 Assignment 7: Due Wednesday, March 11 at start of class

I. Problems to be handed in:

1. A certain coin comes up Heads with an unknown probability p each time it is tossed. It is tossed 6 times, giving 5 Heads.

   (a) Find an unbiased estimate for p, giving your reasoning.

   (b) Assuming the coin is fair ($p = 1/2$), what is the probability that the number of Heads is not in the range $\{2, 3, 4\}$?

   (c) Do the data provide grounds to reject the hypothesis that the coin is fair with 90% confidence?

2. Ten measurements of the percentage of water in a methanol solution yielded the sample mean $\bar{X} = 0.552$ and the sample variance $S^2 = (0.037)^2$. Assuming a normal distribution for the measurements, find a 90% confidence interval for the true percentage of water in the methanol solution.

3. A local company called Gastown Gas produces gas pumps. Recently, increased gas prices and rumours of connections between their CEO and Big Oil have led to increased scrutiny. Some claim that their pumps are intentionally cheating customers by shortchanging them at the pump. (That is, if the pump says it has pumped one litre of gas, the actual amount dispensed is less than that.)

   The CEO denies this and states that while the pumps' output is subject to random fluctuations, every litre purchased from one of their pumps produces a quantity of gas that is normally distributed with mean 1; she does not say what the variance is. Each such quantity is independent and identically distributed.

   To test this hypothesis, an independent investigator goes to a number of Gastown Gas pumps and purchases a litre of gas 10000 times. The investigator measures each dispensed quantity and publishes the results (which are completely accurate) in the file `gasquantities.mat`. A second investigator, independently of the first, does the same and publishes the results in the file `gasquantities2.mat`.

   (a) Using Octave, compute the sample mean $\overline{X}$ and sample variance S^2 of each data set. Download the data files `gasquantities.mat` and `gasquantities2.mat` from the course webpage and put them in your working directory. Then you can use the commands

   ```
   import scipy.io

   mat = scipy.io.loadmat('gasquantities.mat')

   a = mat["gasquantities"][0]
   ```

   to load the data from the first investigator and save it as an array called `a`. Similarly, you can load the second dataset and save it as a different array. Submit your code and your computed sample means and sample variances.

   (b) For each data set, construct a 95% confidence interval for the true mean μ of the volume of gas (where volume is measured in litres). Here you may find the command `scipy.stats.t.interval` useful. If T has a Student t-distribution with k degrees of freedom, then `t.interval(alpha,k)` will give you the interval containing alpha fraction of the probability mass. Submit your code and your computed confidence intervals. Note that you are treating the two data sets independently, so you are finding a different confidence interval for each data set; do not combine the two data sets.

   (c) Based on the first data set, can you reject the hypothesis that $\mu = 1$ at the 5% level? Based on the second data set, can you reject the same hypothesis?

(d) It turns out that the CEO was telling the truth, and that $\mu = 1$ exactly. Find the probability of the outcomes of your hypothesis tests; that is, the probability that someone doing two separate tests with independently collected new data would see your first result for the first test, and see your second result for the second.

(e) Suppose that $\mu = 1$ exactly, and that 100 investigators perform the same test. What is the expected number of tests that will reject the hypothesis that $\mu = 1$ at the 5% level? What does this tell you about the trustworthiness of such hypothesis tests?

4. (a) For the gambler's ruin problem, let M_i denote the expected number of games that will be played when Gambler initially has \$$i$ ($i = 0, 1, \ldots, N$). Let $q = 1 - p$. Show that

$$M_0 = M_N = 0, \quad M_i = 1 + pM_{i+1} + qM_{i-1} \quad (i = 1, \ldots, N-1).$$

Hint: Compute the expectation of the number of games X by conditioning on the outcome of the first game:
$EX = E[X|\text{win first game}]P(\text{win first game}) + E[X|\text{lose first game}]P(\text{lose first game})$.

(b) Solve the equations in (a) to obtain

$$M_i = i(N-i) \quad \text{if } p = \frac{1}{2},$$
$$M_i = \frac{i}{q-p} - \frac{N}{q-p}\frac{1-(q/p)^i}{1-(q/p)^N} \quad \text{if } p \neq \frac{1}{2},$$

by proceeding as follows. First, find the general solution to the homogeneous equation $M_i = pM_{i+1} + qM_{i-1}$ (already done in class). Next, find a particular solution to the inhomogeneous equation $M_i = 1 + pM_{i+1} + qM_{i-1}$ (try $M_i = ci$ for $p \neq \frac{1}{2}$ and $M_i = ci^2$ for $p = \frac{1}{2}$; find c that produces a solution). Add the general solution of the homogeneous equation to the particular solution of the inhomogeneous equation. Finally, solve for the two unknown constants in the general solution by using the boundary conditions.

5. This problem concerns 2-dimensional random walk on the square lattice $\mathbb{Z}^2$ (consisting of points $(x^{(1)}, x^{(2)})$ with $x^{(1)}$ and $x^{(2)}$ both integers).

Let $\vec{X} = (X^{(1)}, X^{(2)})$ be a random vector which takes the values $(1,0)$, $(-1,0)$, $(0,1)$, $(0,-1)$ with equal probabilities $\frac{1}{4}$. For $j = 1, 2$, the marginal p.m.f. of $X^{(j)}$ takes value 1 with probability $\frac{1}{4}$, -1 with probability $\frac{1}{4}$, and 0 with probability $\frac{1}{2}$.

Let $\vec{X}_i$ be i.i.d. with the same distribution as $\vec{X}$, and let $\vec{S}_n = \vec{X}_1 + \cdots + \vec{X}_n$. Then $\vec{S}_n$ represents the position after n steps of a random walker on $\mathbb{Z}^2$, starting from $(0,0)$, whose random steps are equally likely to be any of the four unit vectors.

(a) The expectation of a random vector $\vec{Y} = (Y^{(1)}, Y^{(2)})$ is given by $E\vec{Y} = (EY^{(1)}, EY^{(2)})$ and the variance is given by $\mathrm{Var}(\vec{Y}) = E[(\vec{Y} - E\vec{Y}) \cdot (\vec{Y} - E\vec{Y})]$ with the dot indicating the dot product. Show that, for the random walk, $E\vec{S}_n = (0,0)$ and $\mathrm{Var}(\vec{S}_n) = n$.

(b) If the random walk is instead 1-dimensional (probability $\frac{1}{2}$ of steps left or right) or 3-dimensional (probability $\frac{1}{6}$ of steps north, south, east, west, up, down), what is the expected position and the variance of the walk after n steps?

II. Recommended problems: These provide additional practice but are not to be handed in.

A. Let $\Gamma(n) = \int_0^\infty e^{-t} t^{n-1}\, dt$. Prove that $\Gamma(n) = (n-1)!$ if $n = 1, 2, \ldots$ by two methods: (i) use integration by parts to show that $\Gamma(n) = (n-1)\Gamma(n-1)$, (ii) consider $\frac{d^m}{d\lambda^m} \int_0^\infty e^{-\lambda t}\, dt$ and set $\lambda = 1$.

B. 25 measurements are made of the splitting tensile stress (lb/in^2) of concrete cylinders. The following table shows the frequency of each measured value, with the strength on the first line and the frequency on the second line. Assuming a normal distribution, determine a 99% confidence interval for the mean splitting tensile stress μ of the population from which the sample was drawn. $[350.3, 369.7]$.

320	330	340	350	360	370	380	390
1	1	3	3	8	3	5	1

C. Chapter 4 #57 $[p(1 - q/p)/(1 - (q/p)^n) + q(1 - p/q)/(1 - (p/q)^n)]$, 58

D. Consider the gambler's ruin scenario discussed in class, in which Smith has initial fortune $\$n$ and the bank has initial fortune $\$(1000 - n)$, so the total is \$1000. Smith plays roulette and bets on red. Determine the value of n so that the probability is 99% that Smith goes broke. [956]

Simple random walks on $\mathbb{Z}^2$ taking $n = 1,000$, 10,000 and 60,000 steps. The circles have radius $\sqrt{n}$, in units of the step size of the random walk. Since $E\vec{S}_n = \vec{0}$ (see #5), $\mathrm{Var}(\vec{S}_n) = E(|\vec{S}_n|^2)$ is the *mean-square displacement*. Since the variance is n, the standard deviation is $\sqrt{n}$, so the typical distance of a random walk from its starting point, after n steps, is $\sqrt{n}$. This is illustrated here.

Quote of the week: *"I think you're begging the question," said Haydock, "and I can see looming ahead one of those terrible exercises in probability where six men have white hats and six men have black hats and you have to work it out by mathematics how likely it is that the hats will get mixed up and in what proportion. If you start thinking about things like that, you would go round the bend. Let me assure you of that!"*

Agatha Christie in *The Mirror Crack'd*