Stat 200A, Fall 2018
A. Adhikari and Z. Bartha

WEEK 6 EXERCISES

You are expected to do all these problems, but for **Homework 6** please turn in **only Problems 2, 3, and 5** on **Thursday October 4 at the start of lecture**.

1. Linear Combinations of Independent Normals

Let Z_1, Z_2, Z_3, Z_4 be i.i.d. standard normal variables. Find the following without integration. You can leave answers in terms of the standard normal cdf Φ if necessary.

(a) $P(3Z_1 + 2Z_2 > Z_3 + 4Z_4)$

(b) $P(Z_1 + Z_2 > Z_3 + Z_4 + 1)$

(c) $E(3Z_1 + 2Z_2 - Z_3 - 4Z_4 + 10)$

(d) $SD(3Z_1 + 2Z_2 - Z_3 - 4Z_4 + 10)$

2. Distance Between Normal Points

(a) Show that X has the normal (μ, σ^2) distribution if and only if $X = \sigma Z + \mu$ where Z has the standard normal distribution.

(b) Consider two points thrown independently on the plane, such that the two coordinates of each point are i.i.d. normal (μ, σ^2) random variables. Find the expectation and variance of the distance between the two points.

3. The Cauchy Density

Let X be uniform on the interval $(-\pi/2, \pi/2)$, and let $Y = \tan(X)$.

(a) Find the density of Y. This is called the *Cauchy* density.

(b) Show that the distribution of Y is symmetric about 0 but $E(Y)$ is undefined.

4. Spacings

Let U_1, U_2, U_3, U_4 be i.i.d. uniform $(0,1)$ random variables. Let $U_{(1)}, U_{(2)}, U_{(3)}, U_{(4)}$ denote the four variables arranged in increasing order. As a visualization, if you mark the points U_1, U_2, U_3, U_4 on the unit interval, then from left to right the marks will be at $U_{(1)}, U_{(2)}, U_{(3)}, U_{(4)}$.

(a) Find the density of $U_{(1)}$.

(b) Find the density of $U_{(4)}$.

(c) Find the density of $1 - U_{(4)}$, and compare with the answer to Part **a**.

(d) Review the method of Exercise **6c** of Week 2 Exercises and explain why $U_{(2)}$ and $U_{(3)}$ have the joint density given by

$$f(x,y) = \begin{cases} cx(1-y), & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Here $c > 0$ is a constant. Find it without integration.

(e) Find the density of $U_{(3)} - U_{(2)}$ and compare it with the answer to Part **a**.

5. A Ratio

Let X and Y have the joint density given by

$$f(x,y) \;=\; \begin{cases} \frac{1}{y}e^{-(x+y^2)/y}, & x>0, y>0 \\ 0 & \text{otherwise} \end{cases}$$

Show that X/Y and Y are i.i.d. exponential variables and hence find $E(X)$ and $Var(X)$.

6. The Chi-Squared Distributions

Let n be a positive integer. In statistics, the gamma $(n/2, 1/2)$ distribution is known as the *chi-squared distribution with n degrees of freedom.* We will denote that distribution χ^2_n.

(a) Let Z be a standard normal variable. We showed in class that Z^2 has the gamma $(1/2, 1/2)$ distribution. Now let $Z_1, Z_2, \ldots, Z_n$ be i.i.d. standard normal variables,. Explain why $Z_1^2 + Z_2^2 + \cdots + Z_n^2$ has the χ^2_n distribution.

(b) Find the expectation and variance of the χ^2_n distribution. Use Exercise **5ef** of Week 4 Exercises.

(c) Sketch the graph of the χ^2_n distribution for large n, and explain your choice of shape.

Problems left from last week:

5/4. Correlation

The covariance of random variables X and Y has nasty units: the product of the units of X and the units of Y. Dividing the covariance by the two SDs results in an important pure number.

The *correlation coefficient* of the random variables X and Y is defined as

$$r(X,Y) \;=\; \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

It is called the correlation, for short. The definition explains why X and Y are called *uncorrelated* if $Cov(X,Y) = 0$.

(a) Let X^* be X is standard units and let Y^* be Y in standard units. Check that

$$r(X,Y) = E(X^*Y^*)$$

(b) Use the fact that both $(X^* + Y^*)^2$ and $(X^* - Y^*)^2$ are non-negative random variables to show that $-1 \le r(X,Y) \le 1$.

[First find the numerical values of $E(X^*)$ and $E(X^{*2})$. Then find $E(X^* + Y^*)^2$.]

(c) Show that if $Y = aX + b$ where $a \neq 0$, then $r(X,Y)$ is 1 or -1 depending on whether the sign of a is positive or negative.

(d) Consider a sequence of i.i.d. Bernoulli (p) trials. For any positive integer k let X_k be the number of successes in trials 1 through k. **Use bilinearity** to find $Cov(X_n, X_{n+m})$ and hence find $r(X_n, X_{n+m})$.

(e) Fix n and find the limit of your answer to **(d)** as $m \to \infty$. Explain why the limit is consistent with intuition.

5/5. Relations Between Random Variables

This exercise is about departures from the "independent and identically distributed" (i.i.d.) model, with particular attention to correlation.

(a) Let X_1 and X_2 be the numbers appearing on the first and second rolls of a die. Let $S = X_1 + X_2$ and $D = X_1 - X_2$. Are S and D identically distributed? Are they independent? Are they uncorrelated?

(b) Construct two random variables X and Y such that X and Y are identically distributed and negatively correlated, that is, $Cov(X, Y)$ is negative. You can do this easily on the space of a few tosses of a coin.

(c) Construct two random variables X and Y such that $X \neq Y$, X and Y are identically distributed and positively correlated, that is, $Cov(X, Y)$ is positive. This too can be done on the space of a few tosses of a coin.

5/9. Reliability

Let X_n be the number of successes in n i.i.d. Bernoulli (0.9) trials. About how large does n have to be so that the chance of 100 or more successes is about 99%?

Versions of this calculation are used by airlines to work out by how much they will overbook their flights, or by manufacturers who need to get a minimum number of good items using a process that has some chance of producing duds.