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

WEEK 11 EXERCISES

You are expected to do all these problems, but for **Homework 10** please turn in **only Problems 2, 5, and 6** on **Thursday November 8 at the start of lecture**.

1. Baseline and Post-Treatment

Weight loss is an undesirable side effect of a new medical treatment. Patients are weighed, in pounds, at the start and end of the treatment. Let X be the baseline (pre-treatment) weight of a randomly picked patient and let Y be that patient's post-treatment weight. Suppose X and Y have the bivariate normal distribution with parameters $(160, 140, 15, 20, 0.6)$.

(a) Find $P((X+Y)/2 > 150)$. Did you use the bivariate normal assumption?

(b) Find $P(Y < X)$.

2. Slices of a Normal Cake

You don't have to integrate in this exercise. In fact, integration is not recommended. Instead, draw some straight lines, think about angles, and remember the shape of the joint density of two i.i.d. standard normal variables.

(a) Let V and W be i.i.d. standard normal variables. Find $P(V > 0, W > \sqrt{3}V)$.

(b) In a population of adult mother-daughter pairs, the heights of the mothers and the daughters have correlation 0.6 and an approximately bivariate normal joint distribution. In roughly what proportion of mother-daughter pairs are both women above average in height?

3. Permutations and Subsets

Let $\mathbf{X} = [X_1 \ X_2 \ \ldots \ X_n]^T$ be multivariate normal with mean vector μ_X and covariance matrix $\mathbf{\Sigma}_X$.

(a) Let $\mathbf{Y}$ be a permutation of $\mathbf{X}$. Show that $\mathbf{Y}$ is multivariate normal.

(b) For $m < n$, let $W_1, W_2, \ldots, W_m$ be any subset of $X_1, X_2, \ldots, X_n$ and let $\mathbf{W} = [W_1 \ W_2 \ \ldots \ W_m]^T$. Show that $\mathbf{W}$ is multivariate normal.

4. Normal Sample Mean and Sample Variance, Part 1

(a) Let R have the chi-squared distribution with n degrees of freedom. What is the mgf of R?

(b) For R as in Part (a), suppose $R = V + W$ where V and W are independent and V has the chi-squared distribution with $m < n$ degrees of freedom. Can you identify the distribution of W? Justify your answer.

(c) Let $X_1, X_2, \ldots, X_n$ be any sequence of random variables and let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Let α be any constant. Prove the **sum of squares decomposition**

$$\sum_{i=1}^n (X_i - \alpha)^2 \ = \ \sum_{i=1}^n (X_i - \bar{X})^2 \ + \ n(\bar{X} - \alpha)^2$$

(d) Now let $X_1, X_2, \ldots, X_n$ be i.i.d. normal with mean μ and variance $\sigma^2 > 0$. Let S^2 be the "sample variance" defined by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Use Parts (b) and (c) with an appropriate value of α to find a constant c such that cS^2 has a chi-squared distribution. Provide the degrees of freedom.

5. Normal Sample Mean and Sample Variance, Part 2

Required reading: Carefully go through Exercise 6 in Homework 5. In Part **(e)** of that exercise you defined the sample variance of an i.i.d. sample. The same definition was used in the exercise above. Note that the Homework 5 exercise made no assumption about the underlying distribution of the elements of the sample, other than the existence of the expectation μ and variance σ^2.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2. Define the sample mean $\bar{X}$ and the sample variance S^2 as in Exercise 4.

(a) For $1 \le i \le n$ let $D_i = X_i - \bar{X}$. Find $Cov(D_i, \bar{X})$.

(b) Now assume in addition that $X_1, X_2, \ldots, X_n$ are i.i.d. normal (μ, σ^2). What is the joint distribution of $\bar{X}, D_1, D_2, \ldots, D_{n-1}$? Explain why D_n isn't on the list.

(c) True or false (justify your answer): The sample mean and sample variance of an i.i.d. normal sample are independent of each other.

6. The F Distribution

The F distribution is named for Sir Ronald Fisher, the creator of much of modern statistical theory. The distribution is on the positive real numbers and has two positive integer parameters which we will call n and d for "numerator" and "denominator".

A random variable X has the $F_{n,d}$ distribution if $X = \frac{N/n}{D/d}$ where N and D are independent random variables such that N has the chi-squared (n) distribution and D has the chi-squared (d) distribution.

In statistics, the distribution arises for example in tests for whether two normal populations have the same variance. The variables N and D both arise as sums of squares of centered normal random variables.

The goal of this exercise is for you to derive the following impressive formula for the $F_{n,d}$ density:

$$f_X(x) = \frac{\Gamma\left(\frac{n}{2}+\frac{d}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{d}{2}\right)}\left(\frac{n}{d}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}\left(1+\frac{n}{d}x\right)^{-\frac{n+d}{2}}, \qquad x > 0$$

That looks horrendous but it really isn't. As always, the most impressive part is the constant of integration. The functional form is in fact rather straightforward. For example, if $n = 50$ and $d = 100$ then the density becomes

$$f_X(x) = Cx^{24}(1+0.5x)^{-75}$$

which doesn't seem so bad after all.

(a) Find $E(N/n)$ and $E(D/d)$. This will help explain why F statistics are often compared to 1.

(b) Since $X = \frac{d}{n}Y$ where $Y = \frac{N}{D}$, you will first find the density of Y and then the density of X. Start by writing Y in terms of $R = \frac{N}{N+D}$.

(c) Go back to Lecture 11 given on 9/27 and review the beta-gamma algebra. Find the density of Y.

(d) Now use Part (b) to find the density of X.

7. The t Distribution

This distribution is much used in tests for the mean of a normal population. It is sometimes called the *Student's* t distribution, and using related constructions is sometimes called Studentizing. That's in honor of the originator of the distribution, William Sealy Gosset. He worked in the Guinness Brewery in Ireland early in the 20th century and also spent time publishing statistical papers under the pseudonym Student.

A random variable T has the t distribution with d degrees of freedom if $T = \frac{Z}{\sqrt{D/d}}$ where Z is standard normal, D has the chi-squared (d) distribution, and Z and D are independent.

The goal of this problem is for you to find the density of T and also to identify a commonly used statistic that has a t distribution.

(a) Show that $T \stackrel{d}{=} -T$.

(b) Show that T^2 has an F distribution and hence use Exercise 6 to find the density of T^2.

(c) Find the density of $\sqrt{T^2}$, the positive square root of T^2.

(d) Find the density of T. At this point this should require almost no calculation.

(e) Let $X_1, X_2, \ldots, X_n$ be i.i.d. normal (μ, σ^2) random variables. As in Exercise 5, let $\bar{X}$ be the sample mean and S^2 the sample variance. Let $S = \sqrt{S^2}$ be the positive square root of S^2. Show that the random variable

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has a t distribution and find the degrees of freedom.