ISyE 2028, Fall 2015
Homework 3
100 points total.

This homework is due Tuesday Sept 29 in class.

- Please remember to staple if you turn in more than one page.
- Please make sure to **SHOW ALL WORK** in order to receive full credit.

1. **Continuous distributions.** (20 points.)

   (a) (Normal distribution. 4-69 in textbook.) The life of a semiconductor laser at a constant power is normally distributed with a mean of 7000 hours and a standard deviation of 600 hours.

   i. What is the probability that a laser fails before 5000 hours?
   ii. What is the life in hours that 95% of the lasers exceed?
   iii. If three lasers are used in a product and they are assumed to fail independently, what is the probability that all three are still operating after 7000 hours?

   (b) (χ^2 distribution. 8-42 in textbook.) Determine the values of the following percentage points:
   $$\chi^2_{0.05,10}, \chi^2_{0.025,15}, \chi^2_{0.01,12}, \chi^2_{0.95,20}, \chi^2_{0.99,18}, \chi^2_{0.995,16}, \chi^2_{0.005,25}.$$
   And use linear interpolation to determine $\chi^2_{0.30,8}$.

   (c) (t distribution. 8-22 in textbook.) Find the values of the following percentage points:
   $$t_{0.025,15}, t_{0.05,10}, t_{0.10,20}, t_{0.005,25}, t_{0.001,30}.$$
   And use linear interpolation to determine $t_{0.002,3}$.

   (d) (F distribution. 10-50 in textbook.) For an F distribution, find the following
   $$f_{0.25,7,15}, f_{0.10,10,12}, f_{0.01,20,10}, f_{0.75,7,15}, f_{0.90,10,12}, f_{0.99,20,10}$$
   And use linear interpolation to determine $f_{0.15,2,3}$.

2. **Sample distribution for sample mean and sample variance using R.** (30 points.)

   Suppose that an experimenter observes a set of variables that are taken to be normally distributed with an unknown mean and variance. Using simulation methods, for given values of the mean and variance, we can simulate the data values that the experimenter might obtain. More interestingly, we can simulate lots of possible samples of which, in reality, the experimenter would observe only one. Performing this simulation allows us to check on sampling distributions of the parameter estimates.

   Let us assume that $\mu = 100$ and $\sigma^2 = 9$, which, in fact, the experimenter does not know. In our simulation study, we assume that the experimenter will observe 100 observations, which are normally distributed. To simulate a sample of 100 observations from $N(100, 9)$, which the experimenter might observe, the R command is

```
x = rnorm(100,mean=100,sd=3)
```

The vector x will contain 100 values which are observations from a normal distribution $N(100, 9)$.

(a) What is the mean and the variance of this sample? How do the sample mean and sample variance compare to true values of the mean and variance?

Instructions. Use functions `mean` and `var` in R to find the mean and the variance.

```
mean(x)
var(x)
```

(b) Obtain random samples from the sampling distributions for the sample mean and the sample variance.

Instructions. In order to check the sampling distribution of the sample mean $\widehat{\mu}$ and of the sample variance $\widehat{\sigma}^2$, we will simulate 100 samples for several times (say 500 times). To simulate 500 times, we run the `rnorm` command within a for loop and create a matrix X with 500 rows and 100 columns, each row corresponding to one sample of 100 observations:

```
n = 100 #number of observations in one sample
S = 500 #number of simulations
X = matrix(0,nrow=S, ncol=n)
for(i in 1:S){
  X[i,] = rnorm(n,mean=100,sd=3)
}
```

To obtain the sample means and sample variances of the 500 samples, we apply the function `apply` as follows:

```
means = apply(X,1,mean)
variances = apply(X,1,var)
```

The vectors `means` and `variances` will contain the 500 sample means and 500 sample variances of the 500 samples.

(c) Find the 5-numerical summary for the sample means and sample variances from the 500 samples (using the R functions `summary` and `var`). Plot the sample means and sample variances using a histogram.

Instructions. The R command for a histogram is `hist`. To divide the figure into two panels each panel with one different plot use the command `par(mfrow=c(2,1))`.

```
par(mfrow=c(2,1))
hist(means)
hist(variances)
```

(d) What is the (theoretical) sampling distribution of $\widehat{\mu}$ if we know that the 500 samples come from a normal distribution $N(100, 9)$? Does the histogram approximate the sampling distribution for the sample mean? Why?

(e) What is the (theoretical) sampling distribution of $\widehat{\sigma}^2$ if we know that the 500 samples come from a normal distribution $N(100, 9)$? Does the distribution of the sample variances from the 500 samples approximate the theoretical sampling distribution?

Instructions. In order to evaluate the sampling distribution for $\widehat{\sigma}^2 = S^2$, we can use the `qqplot`. We will use qqplot to compare the distribution of the sample for $\widehat{\sigma}^2$ and its theoretical distribution. The function used for the sampling distribution of $\widehat{\sigma}^2$ is `rchisq(500, df = (100-1))` and the R command for using `qqplot` function is

```
qqplot(variances, (rchisq(500, df = (100-1))*9/(n-1)))
```

3. **Estimation: general concepts.** (50 points.)

(a) (7-14 in textbook.) A consumer electronics company is comparing the brightness of two different types of picture tubes for use in its television sets. Tube type A has mean brightness of 100 and standard deviation of 16, while type B has unknown mean brightness but the standard deviation of assumed to be identical to type A. A random sample of $n = 25$ tubes of each type is selected and $\bar{X}_B - \bar{X}_A$ is computed. If μ_B equals or exceeds μ_A, the manufacture would like to adopt the type B for use. The observed difference is $\bar{x}_B - \bar{x}_A = 3.5$. What decision would you make and why?

(b) (9-124 in textbook.) A manufacturer of semiconductor devices takes a random sample of size n of chips and tests them, classifying each chip as defective or nondefective. Let $X_i = 0$ if the chip is nondefective and $X_i = 1$ if the chip is defective. The sample fraction defective is

$$\widehat{p} = \frac{X_1 + \cdots X_n}{n}.$$

What are the sampling distribution, the sample mean, and the sample variance estimates of $\widehat{p}$ when

i. The sample size is $n = 50$?
ii. The sample size is $n = 80$?
iii. The sample size is $n = 100$?

(c) Let $X_1, \ldots, X_n$ independent random variables identically distributed with density function

$$f(x) = \begin{cases} (\theta+1)x^\theta & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

i. Find a Method of Moments (MOM) Estimator.
ii. Find the Maximum Likelihood Estimator (MLE).

(d) Let $\bar{X}_1$ and S_1^2 be the sample mean and variance for a sample of size n_1 from a population with mean μ_1 and variance σ_1^2. Similarly, let $\bar{X}_2$ and S_2^2 be the sample mean and variance for a sample of size n_2 from a population with mean μ_2 and variance σ_2^2.

i. Find an unbiased estimator for $\mu_1 - \mu_2$ and find its standard error.
ii. Find the bias of the estimator $\bar{X}_1^2 - \bar{X}_2^2$ for the parameter $\mu_1^2 - \mu_2^2$. What happens to the bias as the sample sizes of n_1 and n_2 increase to ∞?

iii. Assume that both populations have the same variance; that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Show that

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator of σ^2.

(e) Suppose that the expectations of three random variables are equal ($\mathbb{E}(X_1) = \mathbb{E}(X_2) = \mathbb{E}(X_3) = \mu$), and their variances are $Var(X_1) = 7$, $Var(X_2) = 13$, and $Var(X_3) = 20$. Consider the point estimates

$$\widehat{\mu}_1 = \frac{X_1}{3} + \frac{X_2}{3} + \frac{X_3}{3}$$
$$\widehat{\mu}_2 = \frac{X_1}{4} + \frac{X_2}{3} + \frac{X_3}{5}$$
$$\widehat{\mu}_3 = \frac{X_1}{6} + \frac{X_2}{3} + \frac{X_3}{4} + 2$$

i. Calculate the bias of each point estimate. Is any one of them unbiased?

ii. Calculate the variance of each point estimate. Which one has the smallest variance?

iii. Calculates the mean square error of each point estimate. Which point estimate has the smallest mean square error for $\mu = 3$?