Statistics I, Fall 2012
Homework 08

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

1. (15 points; 5 points each) Let X_1, X_2, and X_3 be the outcome of rolling three independent dices. Let $\overline{X} = \frac{1}{3}(X_1 + X_2 + X_3)$ be the sample mean.

   (a) Find $\mathbb{E}[\overline{X}]$.

   (b) Find $\mathrm{Var}(\overline{X})$.

   (c) Characterize the distribution of $\overline{X}$.

2. (15 points; 5 points each) Let X_1, X_2, and X_3 be three independent and identical random variables with the following pmf

$$\Pr(X_i = j) = \frac{1}{4} \quad \forall i \in \{1, 2, 3\}, j \in \{1, 2, 3, 4\}.$$

   Let $Y = \frac{1}{2}(X_1 + X_2 + X_3)$. Note that Y is NOT the sample mean.

   (a) (5 points) Find $\mathbb{E}[Y]$.

   (b) (5 points) Find $\mathrm{Var}(Y)$.

   (c) (5 points) Characterize the distribution of Y.

3. (10 points) Consider two binomial random variables $X_1 \sim \mathrm{Bi}(10, 0.4)$ and $X_2 \sim \mathrm{Bi}(8, 0.4)$.

   (a) (5 points) Characterize the distribution of $X_1 + X_2$. Feel free to use any proposition we introduced in Chapter 5.

   (b) (5 points) Your answer in Part (a) may or may not tell you that $X_1 + X_2$ follows a binomial distribution. Regardless of that, consider the following question: Does $\frac{1}{2}(X_1 + X_2)$ follow a binomial distribution? If so, what are the parameters? If not, why?

      Hint. You need to do almost no mathematics for this problem if you are aware of a basic property of binomial distributions. If your idea is similar to that in my mind, it should be good enough to answer the some question for the sample means of binomial random variables under any sample size.

4. (10 points; 5 points each) Consider two populations as described below. The first population follows a Poisson distribution with rate 10. The second population follow a binomial distribution with size 25 and probability 0.6. Suppose we sample from both populations. For populations 1 and 2, the sample sizes are 20 and 5, respectively. Let $\overline{X}_i$ be the sample mean of the sample from population i, $i = 1, 2$.

   (a) Can you determine whether $\mathrm{Var}(\overline{X}_1)$ is larger than, equal to, or smaller than $\mathrm{Var}(\overline{X}_2)$? If yes, determine it and explain why. If no, explain why.

   (b) Can you determine whether $\overline{X}_1$ is larger than, equal to, or smaller than $\overline{X}_2$? If yes, determine it and explain why. If no, explain why.

5. (25 points; 5 points each) In the MS Excel file "Stat_hw08.xls", consider the sheet "Income". In the sheet, the population data of 1000 workers working in IM City are recorded.

   (a) Use simple random sampling to sample 50 workers' monthly income. List the numbers of the workers in your sample. Calculate the sample mean and sample standard deviation.

   (b) Use systematic random sampling to sample 50 workers' monthly income. List the numbers of the workers in your sample. Calculate the sample mean and sample standard deviation.

(c) Use "Sex" to split the population into strata and then apply proportionate stratified random sampling to sample 50 workers' monthly income. What are the proportions of men and women? List the numbers of the workers in your sample. Calculate the sample mean and sample standard deviation.

(d) Use "Age Range" to split the population into clusters and then apply cluster random sampling to sample two cluster for the monthly income of workers in those clusters. Write down the two clusters in your sample. Calculate the sample mean and sample standard deviation.

(e) What kind of sampling error may occur when you do Part (d)? Why?

Note. You are suggested to use random number generators instead of the table of random numbers.

6. (15 points; 3 points each) Suppose we are going to sample n values from a population with the following distribution:

Outcome	10	20	30	40
Probability	0.2	0.3	0.4	0.1

Answer the following problems.

(a) If $n = 2$, what is the probability that the sample mean is greater than 25?

(b) If $n = 10$, what is the variance of the sample mean?

(c) If $n = 100$, what is (roughly) the probability that the sample mean is greater than 25?

(d) Suppose $n = 50$, what is (roughly) the probability for the sample mean to deviate from the population mean by 1?

(e) Generate a random sample of 100 values from the population and calculate the sample mean and sample variance. List all the values in your sample. Compare your values of the two statistics with their theoretical values.

7. (10 points) Let $\{X_i\}_{i=1,\dots,n}$ be an independent sample from a uniform population Uni$(0,2)$, i.e., $X_i \sim$ Uni$(0,2)$ and X_i and X_j are independent for all $i \neq j$. In this problem, we will derive the sampling distribution of the *sample maximum.* Let

$$X_{\max} \equiv \max_{i=1,\dots,n} \{X_i\}$$

be the sample maximum, i.e., the smallest value in the sample. As you may imagine, $X_{\max}$ is a continuous random variable in this example. To describe its distribution, first note that the set of all possible values is $[0,2]$, the interval between 0 and 2. Then we should derive its pdf.

(a) (3 points) The cdf of $X_{\max}$ is, by definition,

$$F(x) = \Pr(X_{\max} < x) \quad \forall x \in [0,2].$$

Our task is to explicitly characterize $F(x)$. To do this, note that

$$\Pr(X_{\max} < x) = \Pr(X_1 < x)\Pr(X_2 < x)\cdots\Pr(X_n < x).$$

Briefly explain why.

(b) (4 points) Characterize $\Pr(X_i < x)$ for each i and then characterize $F(x)$.

(c) (3 points) Characterize $f(x)$ over $(0,2)$ by taking a first order derivative of $F(x)$ with respect to x.