Statistics I, Fall 2012
Suggested Solution to Homework 08

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

1. (a) We have

$$\begin{aligned}\mathbb{E}[\overline{X}] &= \mathbb{E}\left[\frac{1}{3}(X_1+X_2+X_3)\right] \\ &= \frac{1}{3}\Big(\mathbb{E}[X_1]+\mathbb{E}[X_2]+\mathbb{E}[X_3]\Big) = \frac{1}{3}\left(\frac{7}{2}+\frac{7}{2}+\frac{7}{2}\right) = \frac{7}{2}.\end{aligned}$$

   (b) We have

$$\begin{aligned}\text{Var}(\overline{X}) &= \text{Var}\left(\frac{1}{3}(X_1+X_2+X_3)\right) \\ &= \frac{1}{9}\Big[\text{Var}(X_1)+\text{Var}(X_2)+\text{Var}(X_3)\Big] = \frac{1}{9}\left(\frac{35}{12}+\frac{35}{12}+\frac{35}{12}\right) = \frac{35}{12}.\end{aligned}$$

   (c) To characterize the distribution of $\overline{X}$, we need to first list all its possible outcomes and then assign probabilities to these outcomes. The set of possible values of $\overline{X}$ is $\{1, \frac{4}{3}, \frac{5}{3}, 2, ..., \frac{17}{3}, 6\}$. The probability of each possible outcome need to be derived through detailed counting:

   - $\Pr(\overline{X}=1)$: All three dices must result in 1 and thus the probability is $\frac{1}{216}$.
   - $\Pr(\overline{X}=\frac{4}{3})$: We must get two 1s and one 2. The probability is thus $\frac{\binom{3}{1}}{216} = \frac{3}{216}$.
   - $\Pr(\overline{X}=\frac{5}{3})$: We must get either two 1s and one 3 or one 1 and two 2s. The probability is thus $\frac{\binom{3}{1}+\binom{3}{1}}{216} = \frac{6}{216}$.
   - $\Pr(\overline{X}=2)$: There are three possibilities: (1) Two 1s and one 4, (2) one 1, one 2, and one 3, and (3) three 2s. The probability is thus $\frac{\binom{3}{1}+\binom{3}{1}\binom{2}{1}+\binom{3}{3}}{216} = \frac{10}{216}$.
   - And so on...

   The distribution of $\overline{X}$ is summarized below:

Outcome	1	$\frac{4}{3}$	$\frac{5}{3}$	2	$\frac{7}{3}$	$\frac{8}{3}$	3	$\frac{10}{3}$
Probability	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$
Outcome	$\frac{11}{3}$	4	$\frac{13}{3}$	$\frac{14}{3}$	5	$\frac{16}{3}$	$\frac{17}{3}$	6
Probability	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$

2. (a) We have

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}\left[\frac{1}{2}(X_1+X_2+X_3)\right] \\ &= \frac{1}{2}\Big(\mathbb{E}[X_1]+\mathbb{E}[X_2]+\mathbb{E}[X_3]\Big) = \frac{1}{2}\left(\frac{5}{2}+\frac{5}{2}+\frac{5}{2}\right) = \frac{15}{4}.\end{aligned}$$

   (b) We have

$$\begin{aligned}\text{Var}(Y) &= \text{Var}\left(\frac{1}{2}(X_1+X_2+X_3)\right) \\ &= \frac{1}{4}\Big[\text{Var}(X_1)+\text{Var}(X_2)+\text{Var}(X_3)\Big] = \frac{1}{4}\left(\frac{5}{4}+\frac{5}{4}+\frac{5}{4}\right) = \frac{15}{16}.\end{aligned}$$

(c) To characterize the distribution of Y, we need to first list all its possible outcomes and then assign probabilities to these outcomes. The set of possible values of $\overline{X}$ is $\{\frac{3}{2}, 2, \frac{5}{2}, 3, ..., 6\}$. The probability of each possible outcome need to be derived through detailed counting:

- $\Pr(Y = \frac{3}{2})$: All three dices must result in 1 and thus the probability is $\frac{1}{64}$.
- $\Pr(\overline{X} = 2)$: We must get two 1s and one 2. The probability is thus $\frac{\binom{3}{1}}{64} = \frac{3}{64}$.
- $\Pr(\overline{X} = \frac{3}{2})$: We must get either two 1s and one 3 or one 1 and two 2s. The probability is thus $\frac{\binom{3}{1}+\binom{3}{1}}{64} = \frac{6}{64}$.
- $\Pr(\overline{X} = 3)$: There are three possibilities: (1) Two 1s and one 4, (2) one 1, one 2, and one 3, and (3) three 2s. The probability is thus $\frac{\binom{3}{1}+\binom{3}{1}\binom{2}{1}+\binom{3}{3}}{216} = \frac{10}{64}$.
- And so on...

The distribution of $\overline{X}$ is summarized below:

Outcome	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{7}{2}$	4	$\frac{9}{2}$	5	$\frac{11}{2}$	6
Probability	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{12}{216}$	$\frac{12}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$

3. (a) The random variable $X_1 + X_2$ follows a binomial distribution with $10 + 8 = 18$ trials and probability 0.4.

   (b) $\frac{1}{2}(X_1 + X_2)$ does not follow a binomial distribution because one of its possible outcome is $\frac{1}{2}$, which cannot occur for a binomial distribution.

4. (a) $\mathrm{Var}(\overline{X}_1) = \frac{10}{20} = \frac{1}{2}$ and $\mathrm{Var}(\overline{X}_2) = \frac{25\times0.6\times0.4}{5} = \frac{6}{5}$? It is then clear that $\mathrm{Var}(\overline{X}_1) < \mathrm{Var}(\overline{X}_2)$.

   (b) As $\overline{X}_1$ and $\overline{X}_2$ are random variables with overlapping sets of possible outcomes, we cannot determine whether one is larger than the other one.

5. (a) The 50 workers sampled by me with simple random sampling are

528	868	989	696	114	35	957	879	401	819
377	166	592	514	940	862	287	209	842	557
200	121	547	469	895	190	1000	633	348	774
528	242	668	301	112	537	96	885	807	440
155	713	502	928	850	276	908	719	278	67

   The sample mean is 30501 and the sample standard deviation is 2140.26.

   (b) The 50 workers sampled by me with simple random sampling are

6	26	46	66	86	106	126	146	166	186
206	226	246	266	286	306	326	346	366	386
406	426	446	466	486	506	526	546	566	586
606	626	646	666	686	706	726	746	766	786
806	826	846	866	886	906	926	946	966	986

   The sample mean is 29809.04 and the sample standard deviation is 2235.14.

   (c) In the sample, there are 594 men and 406 women. With proportionate stratified random sampling, we should sample $50 \times \frac{594}{1000} \approx 29.7$ men and 20.3 women. As no rule specifies how to deal with the fractions, we may choose to sample 30 men and 20 women (sampling 29 men and 21 women is also fine). The 50 workers sampled by me with proportionate stratified random sampling are

938	418	842	123	540	461	16	816	540	257
170	92	521	940	371	929	713	352	60	485
296	348	59	687	403	967	751	766	401	120
544	468	881	810	739	164	86	10	430	350
51	608	46	967	391	27	741	369	87	11

where the first 20 are women and the last 30 are men. The sample mean is 30023.46 and the sample standard deviation is 1565.85.

(d) The two clusters sampled by me are "31–35" and "41–45". There are 382 entities in these two clusters. The sample mean is 30005.11 and the sample standard deviation is 2054.12.

(e) As income has some correlation with ages, entities in each cluster may not be heterogeneous enough. Moreover, the sizes of different clusters are quite different. These may be the sources of sampling error when we use "Age Range" as clusters and perform cluster random sampling.

6. (a) **Method 1.** We may first derive the distribution of the sample mean. Through the process similar to that for Problems 1 and 2, the distribution of the sample mean with sample size 2 can be characterized as

Outcome	10	15	20	25	30	35	40
Probability	0.04	0.12	0.25	0.28	0.22	0.08	0.01

Now the probability for the sample mean to be greater than 25 is $0.22 + 0.08 + 0.01 = 0.31$.
Method 2. Let X_1 and X_2 be the first and second values sampled, the probability that we are looking for is

$$\begin{aligned}
&\Pr(X_1 + X_2 > 25) \\
=\ &\Pr(X_1 = 20, X_2 = 40 \cap X_1 = 30, X_2 \geq 30 \cap X_1 = 40, X_2 \geq 20) \\
=\ &\Pr(X_1 = 20, X_2 = 40) + \Pr(X_1 = 30, X_2 \geq 30) + \Pr(X_1 = 40, X_2 \geq 20) \\
=\ &\Pr(X_1 = 20)\Pr(X_2 = 40) + \Pr(X_1 = 30)\Pr(X_2 \geq 30) + \Pr(X_1 = 40)\Pr(X_2 \geq 20) \\
=\ &0.03 + 0.2 + 0.08 = 0.31.
\end{aligned}$$

(b) Let X be the outcome of drawing one entity and $\overline{X}$ be the sample mean of drawing ten entities. First, we may calculate $\mathrm{Var}(X) = \sigma^2 = 84$. It then follows that the variance of the sample mean is

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{84}{10} = 8.4.$$

(c) Because $n \geq 30$, we may apply the central limit theorem, which implies that the sample mean follows a normal distribution with mean 24 (the same as the population mean) and standard deviation $\sqrt{\frac{84}{100}} = 0.917$. Let $\overline{X}$ be the sample mean of a sample size 100, it then follows that the probability for the sample mean to be greater than 25 is

$$\Pr(\overline{X} > 25) = \Pr\left(Z > \frac{25 - 24}{0.917}\right) \approx \Pr(Z > 1.091) \approx 0.138,$$

where Z is a standard normal random variable.

(d) Because $n \geq 30$, we may apply the central limit theorem, which implies that the sample mean follows a normal distribution with mean 24 (the same as the population mean) and standard deviation $\sqrt{\frac{84}{50}} \approx 1.296$. Let $\overline{X}$ be the sample mean of a sample size 50, it then follows that the probability for the sample mean to deviate from the population mean by 1 is

$$\begin{aligned}
1 - \Pr(23 < \overline{X} < 25) &\approx 1 - \Pr\left(\frac{23 - 24}{1.296} < Z < \frac{25 - 24}{1.296}\right) \\
&\approx 1 - \Pr(-0.772 < Z < 0.772) \approx 0.44,
\end{aligned}$$

where Z is a standard normal random variable.

(e) The random sample generated by me contains the following 100 values:

10	20	20	20	20	20	30	30	40	30	30	10	20	30	20	30	10	10	20	10
30	30	20	40	20	30	20	20	10	20	30	30	30	20	30	30	10	10	20	20
10	30	10	30	20	30	30	30	30	30	20	30	30	30	30	30	30	40	20	40
30	30	10	20	30	30	30	20	30	10	30	30	20	30	10	30	20	30	20	20
10	20	30	20	30	10	40	10	10	20	10	20	30	20	10	10	40	20	30	20

The sample mean is 20.8 and the sample variance (with denominator 99!) is 82.67. The theoretical values, which are the population mean and variance, are 24 and 84. The sample mean and variances are close to their theoretical values.

7. (a) If the maximum value of a set of values is lower than x, it must be that all these values are lower than x. Moreover, If all the values are lower than x, their maximum must be lower than x. Therefore, the events "$X_{\max} < x$" and "$X_i < x \forall i$" are equivalent. The equality then holds because X_is are independent.

   (b) We have

$$\Pr(X_i < x) = \int_0^x \frac{1}{2} dz = \frac{x}{2}.$$

   Therefore, the cumulative distribution function is

$$F(x) = \left(\frac{x}{2}\right)^n.$$

   (c) The probability density function is

$$f(x) = \frac{d}{dx} F(x) = n\left(\frac{x}{2}\right)^{n-1}\left(\frac{1}{2}\right) = \frac{nx^{n-1}}{2^n}.$$