Statistics I, Fall 2012
Suggested Solution to Homework 09

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

1. As we have proved in the lecture, for each $i = 1, ..., n$, we have

$$Z_i^2 \equiv \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \text{Chi}(1),$$

which means the moment generating function of Z_i^2 is $m_i(t) = (1-2t)^{-\frac{1}{2}}$. Now, as

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^n Z_i^2,$$

the moment generating function of $\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2}$ is the moment generating function of $\sum_{i=1}^n Z_i^2$, which is

$$\prod_{i=1}^n m_i(t) = \left[(1-2t)^{-\frac{1}{2}}\right]^n = (1-2t)^{-\frac{n}{2}}.$$

Because this is exactly the moment generating function of a chi-square random variable with degree of freedom n, the proof is complete.

2. We have

$$Y_n = \frac{1}{\sqrt{(n-1)n}} Z_1 + \frac{1}{\sqrt{(n-1)n}} Z_2 + \cdots + \frac{1}{\sqrt{(n-1)n}} Z_{n-1} - \frac{n-1}{\sqrt{(n-1)n}} Z_n.$$

As Z_is are all normal random variables, Y_n is also a normal random variable. It remains to find the mean and variance of Y_n. Because $\mathbb{E}[Z_i] = 0$ for all $i = 1, ..., n$, the mean is

$$\mathbb{E}[Y_n] = \frac{1}{\sqrt{(n-1)n}} \mathbb{E}[Z_1] + \cdots + \frac{1}{\sqrt{(n-1)n}} \mathbb{E}[Z_{n-1}] - \frac{n-1}{\sqrt{(n-1)n}} \mathbb{E}[Z_n] = 0.$$

Because $\text{Var}(Z_i) = 1$ for all $i = 1, ..., n$, the variance is

$$\text{Var}(Y_n) = \frac{1}{(n-1)n} + \frac{1}{(n-1)n} + \cdots + \frac{1}{(n-1)n} + \frac{(n-1)^2}{(n-1)n} = \frac{1}{n} + \frac{n-1}{n} = 1.$$

It then follows that $Y_n \sim \text{ND}(0, 1)$.

3. We have

$$\begin{aligned}\sum_{i=1}^n (Z_i - \overline{Z})^2 &= \sum_{i=1}^n (Z_i^2 - 2Z_i\overline{Z} + \overline{Z}^2) = \sum_{i=1}^n Z_i^2 - 2\overline{Z}\sum_{i=1}^n Z_i + n\overline{Z}^2 \\ &= \sum_{i=1}^n Z_i^2 - 2\overline{Z}(n\overline{Z}) + n\overline{Z}^2 = \sum_{i=1}^n Z_i^2 - n\overline{Z}^2.\end{aligned}$$

4. (a) The moment generating function is

$$m_n(t) = \frac{1 - \frac{1}{n}}{1 - \frac{1}{n} - t}$$

and the cumulative distribution function is

$$F_n(x) = 1 - e^{-(1-\frac{1}{n})x}.$$

(b) We have

$$\lim_{n\to\infty} m_n(t) = \frac{1}{1-t}.$$

(c) We have

$$\lim_{n\to\infty} F_n(x) = 1 - e^{-x}.$$

(d) The limiting distribution of X_ns is the exponential distribution with rate 1.

5. (a) X_1 follows a Bernoulli distribution with probability 0.3. The probability of having more 1s than 0s when the sample size is 1 is $\Pr(X_1 = 1) = 0.3$.

(b) The distribution of $\hat{p}$ can be derived by finding the probabilities of the three possible outcomes: 0, $\frac{1}{2}$, and 1:

$$\begin{aligned}
\Pr(\hat{p} = 0) &= \Pr(X_1 + X_2 = 0) = 0.7^2 = 0.49, \\
\Pr\left(\hat{p} = \frac{1}{2}\right) &= \Pr(X_1 + X_2 = 1) = 2(0.3)(0.7) = 0.42, \text{ and} \\
\Pr(\hat{p} = 1) &= \Pr(X_1 + X_2 = 2) = 0.09.
\end{aligned}$$

The probability of having more 1s than 0s when the sample size is 2 is thus $\Pr(X_1 + X_2 = 2) = 0.09$.

(c) The distribution of $\hat{p}$ can be derived by finding the probabilities of the three possible outcomes: 0, $\frac{1}{3}$, $\frac{2}{3}$, and 1:

$$\begin{aligned}
\Pr(\hat{p} = 0) &= \Pr(X_1 + X_2 + X_3 = 0) = 0.7^3 = 0.343, \\
\Pr\left(\hat{p} = \frac{1}{3}\right) &= \Pr(X_1 + X_2 + X_3 = 1) = 3(0.3)(0.7)^2 = 0.441, \\
\Pr\left(\hat{p} = \frac{2}{3}\right) &= \Pr(X_1 + X_2 + X_3 = 2) = 3(0.3)^2(0.7) = 0.189, \text{ and} \\
\Pr(\hat{p} = 1) &= \Pr(X_1 + X_2 + X_3 = 3) = 0.027.
\end{aligned}$$

The probability of having more 1s than 0s when the sample size is 2 is thus $\Pr(X_1+X_2+X_3 \geq 2) = 0.189 + 0.027 = 0.216$.

(d) As $n \geq 30$, we may apply the central limit theorem and conclude that $\hat{p}$ follows a normal distribution. For the sampling distribution, the mean should be $p = 0.3$ and the standard deviation should be $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3\times0.7}{50}} = 0.065$. Therefore, $\hat{p} \sim \text{ND}(0.3, 0.065)$. The probability of having more 1s than 0s when the sample size is 50 is thus

$$\Pr(\hat{p} > 0.5) = \Pr\left(Z > \frac{0.5 - 0.3}{0.065}\right) \approx \Pr(Z > 3.086) \approx 0.001.$$

(e) No. For example, when the sample size goes from 1 to 2, the probability goes from 0.3 to 0.09. This certainly has something to do with the fact that X_is are discrete. Will increasing the sample size always decrease the probability of making this kind of error when the random variables are continuous?

6. (a) Let X be the number of women in the sample that have volunteering experiences and $\hat{p} = \frac{X}{150}$ be the sample proportion. According to the central limit theorem, we have

$$\hat{p} \sim \text{ND}\left(0.25, \sqrt{\frac{0.25 \times 0.75}{150}}\right) \sim \text{ND}(0.25, 0.035).$$

It then follows that

$$\Pr(X \geq 35) \approx \Pr(\hat{p} \geq 0.233) \approx \Pr\left(Z \geq \frac{0.233 - 0.25}{0.035}\right) \approx \Pr(Z \geq -0.47) \approx 0.681,$$

where Z is a standard normal random variable.

(b) We need to first find the population proportion, which is the probability that a randomly selected person has volunteering experiences. This probability is $0.48 \times 0.25 + 0.52 \times 0.2 = 0.224$. Let X be the number of people in the sample that have volunteering experiences and $\hat{p} = \frac{X}{300}$ be the sample proportion. According to the central limit theorem, we have

$$\hat{p} \sim \text{ND}\left(0.224, \sqrt{\frac{0.224 \times 0.776}{300}}\right) \sim \text{ND}(0.224, 0.024).$$

It then follows that

$$\begin{aligned} \Pr(0.2 \le \hat{p} \le 0.25) &\approx \Pr\left(\frac{0.2 - 0.224}{0.024} \le Z \le \frac{0.25 - 0.224}{0.024}\right) \\ &\approx \Pr(-0.997 \le Z \le 1.08) \approx 0.701. \end{aligned}$$

where Z is a standard normal random variable.

7. (a) We know $1.4S^2 \sim \text{Chi}(n-1) \sim \text{Chi}(14)$. It then implies that $\Pr(1.4S^2 > 10) = 0.762$.

   (b) We know $1.4S^2 \sim \text{Chi}(14)$. It then implies that $\Pr(6 < 1.4S^2 < 14) \approx 0.966 - 0.45 = 0.516$.

   (c) The probability that $6 < S^2 < 14$ cannot be found directly as we do not know the distribution of S^2. However, once we multiply all the terms by 1.4, we may find the desired probability as we know the distribution of $1.4S^2$. The probability is

$$\Pr(6 < S^2 < 14) = \Pr(8.4 < 1.4S^2 < 19.6) \approx 0.867 - 0.143 = 0.724.$$