Math/Stat 425
Solutions to Problem Set 5

Instructor: Michael Woodroofe
GSIs: Olugbenga Olumoladeand and Runlong Tang

Statistics Department, UM

Note: For each problem, just one possible solution is provided. You solutions can be different.

1 Solution to Question 2:

Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. from $f(x)$. Then by Central Limit Theorem, we have

$$\frac{\sum_{i=1}^n X_i - E[\sum_{i=1}^n X_i]}{\sqrt{Var[\sum_{i=1}^n X_i]}} \Rightarrow Z, \tag{1.0.1}$$

where Z is a standard normal random variable.

Since

$$E[\sum_{i=1}^n X_i] = \sum_{i=1}^n E[X_i] = nE[X_1] = n \times 0 = 0$$

and

$$Var[\sum_{i=1}^n X_i] = \sum_{i=1}^n Var[X_i] = nVar[X_1] = n \times 1/5 = n/5,$$

from (1.0.1) we have

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n/5}} \Rightarrow Z. \tag{1.0.2}$$

Now $n = 80$, from the definition of S and (1.0.2) we have

$$\frac{S}{4} \overset{d}{\approx} Z,$$

which is

$$S \overset{d}{\approx} 4Z. \tag{1.0.3}$$

Then

$$\begin{aligned} P\{S > 3\} &= P\{4Z > 3\} \\ &= P\{Z > 3/4\} \\ &\approx 0.227. \end{aligned}$$

On the other hand, since

$$\begin{aligned} P\{S \le c\} &= P\{4Z \le c\} \\ &= P\{Z \le c/4\} \end{aligned}$$

and the lower 0.95 quantile of the standard normal distribution is approximately 1.645, we have

$$c/4 \approx 1.645,$$

which is

$$c \approx 6.579.$$

2 Solution to Question 4:

For $1 \le i \le 1000$ and $1 \le j \le 100$, define

$$X_{ij} = \begin{cases} 1 & \text{if the } i\text{th people wins in the } j\text{th day,} \\ 0 & \text{if the } i\text{th people loses in the } j\text{th day,} \end{cases}$$

and suppose they are independent. Then

$$S = \sum_{j=1}^{100} \left(1000 - 500 \sum_{i=1}^{1000} X_{ij} \right). \tag{2.0.4}$$

For $1 \le j \le 100$, define

$$X_j = \sum_{i=1}^{1000} X_{ij}.$$

Then X_j approximately follows $Poisson(\lambda = 1)$ (Same to the reasoning in Problem 3).

Thus

$$S = \sum_{j=1}^{100} (1000 - 500X_j). \tag{2.0.5}$$

Since

$$\nu = E[S] = E[\sum_{j=1}^{100} (1000 - 500X_j)] = \sum_{j=1}^{100} (1000 - 500E[X_j]) = \sum_{j=1}^{100} (1000 - 500) = 50000$$

and

$$Var[S] = Var[\sum_{j=1}^{100} (1000 - 500X_j)] = Var[\sum_{j=1}^{100} (500X_j)] = 100 \times 500^2 Var[X_j] = 5000^2,$$

Still by the Central Limit Theorem (1.0.1), we have

$$\frac{S - 50000}{5000} \Rightarrow Z,$$

where $Z \sim N(0, 1)$.

Since

$$P\{\nu - c < S \le \nu + c\} = P\{|S - \nu|/5000 \le c/5000\}$$

and the upper 0.025 quantile of $N(0, 1)$ is about 1.96, we have

$$c/5000 \approx 1.96,$$

which is

$$c \approx 9800.$$

3 Solution to Question 6:

The distribution function of X_1 or X_2 is

$$F(x) = \int_0^x \frac{1}{2\sqrt{u}} du = x^{1/2},$$

where $x \in (0,1]$; $F(x) = 0$, where $x \in (-\infty, 0]$; and $F(x) = 1$, where $x \in [1, \infty)$. Thus for $y \in [0,1]$, $F^{-1}(y) = y^2$. Then is $U \sim U[0,1]$, $F^{-1}(U) \sim F(x)$.

Thus we can generate $\{(U_{i1}, U_{i2})\}_{i=1}^N$ i.i.d. from $U[0,1]$ and have the corresponding $X_{i1} = F^{-1}(U_{i1})$ and $X_{i2} = F^{-1}(U_{i2})$, where $1 \le i \le N$.

Define $Y_i = 1(X_{i1} + X_{i2} \le 1)$, where $1 \le i \le N$. Then $p = P\{X_1 + X_2 \le 1\}$ can be estimated by $\hat{p} = \sum_{i=1}^N Y_i / N$.

Since a simple bound of error $\hat{p} - p$ is $1/\sqrt{N}$ (see the notes), let $1/\sqrt{N} < 0.01$, which is $N > 10,000$. That is, we need to generate al least 10000 pairs of sample points.

4 Solution to Question 8:

$$\begin{aligned}
\bar{X}_n &= \frac{1}{n} \sum_{k=1}^n \frac{Y_{k-1} + 2Y_k + Y_{k+1}}{4} \\
&= \frac{1}{4} [\frac{1}{n} \sum_{k=1}^n Y_{k-1} + 2\frac{1}{n} \sum_{k=1}^n Y_k + \frac{1}{n} \sum_{k=1}^n Y_{k+1}] \\
&\xrightarrow{P} \frac{1}{4} [0 + 2 \times 0 + 0] \\
&= 0,
\end{aligned}$$

since

$$\frac{1}{n} \sum_{k=1}^n Y_{k-1} \xrightarrow{P} E[Y_k] = 0,$$

$$\frac{1}{n} \sum_{k=1}^n Y_k \xrightarrow{P} E[Y_k] = 0,$$

and

$$\frac{1}{n} \sum_{k=1}^n Y_{k+1} \xrightarrow{P} E[Y_k] = 0.$$

5 Solution to Question 10:

For $1 \le i \le 10,000$, define X_i as the yearly claim of the ith policyholder. Then by the Central Limit Theorem (1.0.1), we know

$$\frac{\sum_{i=1}^{10,000} X_i - 240 \times 10,000}{\sqrt{10,000} \times 800} \overset{d}{\approx} N(0,1).$$

Thus

$$\begin{aligned}
P\{\sum_{i=1}^{10{,}000} X_i > 2.7\times 10^6\} &= P\{\frac{\sum_{i=1}^{10{,}000} X_i - 240\times 10,000}{\sqrt{10,000}\times 800} > \frac{2.7\times 10^6 - 240\times 10,000}{\sqrt{10,000}\times 800}\} \\
&\approx P\{N(0,1) > 15/4\} \\
&= 8.8\times 10^{-5},
\end{aligned}$$

which is almost 0.

6 Solution to Question 12:

From the Problem 4 of Homework 4, we know the density function of Y_n is

$$f(y) = ne^{-ny}1(y \geq 0),$$

for $y \in \mathbb{R}$.

Thus we have

$$\begin{aligned}
E[(Y_n - 0)^2] &= E[Y_n^2] \\
&= \int_{-\infty}^{+\infty} y^2 f(y)dy \\
&= \int_{0}^{+\infty} y^2 ne^{-ny}dy \\
&\overset{z=ny}{=} n^{-2}\int_{0}^{+\infty} z^2 e^{-z}dz \\
&= n^{-2}\times 2 \\
&= 2n^{-2} \\
&\to 0,
\end{aligned}$$

as $n \to \infty$. Therefore Y_n converges to 0 in mean square.