Solutions to Assignment #21

1. Forty–eight measurements are recorded to several decimal places. Each of these 48 numbers is rounded off to the nearest integer. The sum of the original 48 numbers is approximated by the sum of those integers. Assume that the errors made in rounding off are independent, identically distributed random variables with a uniform distribution over the interval $(-0.5, 0.5)$. Compute approximately the probability that the sum of the integers is within two units of the true sum.

   ***Solution**:* Let $X_1, X_2, \ldots, X_n$, where $n = 48$ denote the 48 measurements, and $Y_1, Y_2, \ldots, Y_n$ be the corresponding nearest integers after rounding off. Then

   $$X_i = Y_i + U_i \quad \text{for} \quad i = 1, 2, \ldots, n,$$

   where $U_1, U_2, \ldots, U_n$ are independent identically distributed uniform random variables on the interval $(-0.5, 0.5)$. We then have that $E(U_i) = 0$ for all i, and

   $$\text{var}(U_i) = \frac{1}{12},$$

   for all i. Thus, $\sigma = 1/\sqrt{12}$.
   The sum, S, of the original measurements is

   $$S = \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} U_i,$$

   where $\sum_{i=1}^{n} Y_i$ is the sum of the integer approximations. Put

   $$W = S - \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} U_i.$$

   We would like to estimate $\Pr(|W| \leqslant 2)$. We will do this by applying the Central Limit Theorem to $U_1, U_2, U_3, \ldots$
   Observe that $W = n\overline{U}_n$, where $\overline{U}_n$ is the sample mean of the rounding off values $U_1, U_2, \ldots, U_n$. By the Central Limit Theorem

   $$\Pr\left(\frac{\overline{U}_n}{\sigma/\sqrt{n}} \leqslant z\right) \approx \Pr(Z \leqslant z),$$

for all $z \in \mathbb{R}$, where $Z \sim \text{Normal}(0,1)$. We therefore get that

$$\Pr\left(\frac{|\overline{U}_n|}{\sigma/\sqrt{n}} \leqslant z\right) \approx \Pr(|Z| \leqslant z),$$

for $z > 0$.

It then follows that

$$\begin{aligned}
\Pr(|W| \leqslant 2) &= \Pr(|\overline{U}_n| \leqslant 2/n) \\
&= \Pr\left(\frac{|\overline{U}_n|}{\sigma/\sqrt{n}} \leqslant \frac{2}{\sigma\sqrt{n}}\right) \\
&\approx \Pr\left(|Z| \leqslant \frac{2}{\sigma\sqrt{n}}\right) \\
&= 2F_Z\left(\frac{2}{\sigma\sqrt{n}}\right) - 1 \\
&= 2F_Z(1) - 1 \\
&\approx 0.6826,
\end{aligned}$$

or about 68.3%. □

2. Let X denote a random variable with pdf

$$f_X(x) = \begin{cases} \dfrac{1}{x^2} & \text{if } 1 < x < \infty, \\ \\ 0 & \text{otherwise.} \end{cases}$$

Consider a random sample of size 72 from this distribution. Compute approximately the probability that 50 or more observations of the random sample are less than 3.

***Solution**:* The probability that a random observation from the distribution is less than 3 is given by

$$p = \int_{-\infty}^{3} f_X(x)\, \mathrm{d}x = \int_{1}^{3} \frac{1}{x^2}\, \mathrm{d}x = \frac{2}{3}.$$

Let Y denote the number of observation out of the 72 which are less than 3. Then, $Y \sim \text{Binomial}(p, n)$, where $n = 72$. It then follows that the mean of Y is $\mu = np = 48$ and the standard deviation of Y is $\sigma = \sqrt{np(1-p)} = 4$. By the Central Limit Theorem we then get that

$$\Pr\left(\frac{Y - np}{\sqrt{np(1-p)}} \leqslant z\right) \approx \Pr(Z \leqslant z),$$

for all $z \in \mathbb{R}$, where $Z \sim \text{Normal}(0, 1)$, or

$$\Pr\left(\frac{Y - 48}{4} \leqslant z\right) \approx \Pr(Z \leqslant z).$$

We want to estimate $\Pr(Y > 49) = 1 - \Pr(Y \leqslant 49)$, where

$$\begin{aligned}
\Pr(Y \leqslant 49) &= \Pr\left(\frac{Y - 48}{4} \leqslant \frac{49 - 48}{4}\right) \\
&\approx \Pr\left(Z \leqslant 0.25\right) \\
&\approx 0.5987.
\end{aligned}$$

Thus, $\Pr(Y > 49) \approx 1 - 0.5987 = 0.4013$ or about 40.13%. □

3. *[Exercise 5 on page 235 in the text]*

   How large a random sample must be taken from a given distribution in order for the probability to be at least 0.99 that the sample mean will be within 2 standard deviations of the mean of the distribution?

   Solution: Applying Chebyshev's Inequality to the sample mean, $\overline{X}_n$, we get that

$$\Pr(|\overline{X}_n - \mu| \geqslant \varepsilon) \leqslant \frac{\text{var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2},$$

   for any $\varepsilon > 0$. Thus,

$$\Pr(|\overline{X}_n - \mu| < \varepsilon) \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

   Taking $\varepsilon = 2\sigma$ we get that

$$\Pr(|\overline{X}_n - \mu| < 2\sigma) \geqslant 1 - \frac{\sigma^2}{n(2\sigma)^2} = 1 - \frac{1}{4n}.$$

We want n to be so that

$$1 - \frac{1}{4n} \geqslant 0.99,$$

which yields $4n \geqslant 100$, or $n \geqslant 25$. Thus, we want the sample size to be at least 25. □

4. *[Exercise 6 on page 235 in the text]*

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of size n from a distribution for which the mean is 6.5 and the variance is 4. Determine how large the value of n must be in order for the following relation to be satisfied:

$$\Pr(6 \leqslant \overline{X}_n \leqslant 7) \leqslant 0.8.$$

Solution: Observe that $6 \leqslant \overline{X}_n \leqslant 7$ if and only if

$$-0.5 \leqslant \overline{X}_n - 6.6 \leqslant 0.5$$

or $|\overline{X}_n - 6.5| \leqslant 0.5$.
Using the inequality

$$\Pr(|\overline{X}_n - \mu| < \varepsilon) \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2},$$

with $\mu = 6.5$, $\sigma^2 = 4$, and $\varepsilon = 0.5$, we obtain that

$$\Pr(|\overline{X}_n - 6.5| < 0.5) \geqslant 1 - \frac{4}{n(0.5)^2},$$

so that

$$\Pr(|\overline{X}_n - 6.5| \leqslant 0.5) \geqslant \Pr(|\overline{X}_n - 6.5| < 0.5) \geqslant 1 - \frac{8}{n}.$$

Thus, in order to make $\Pr(|\overline{X}_n - 6.5| \leqslant 0.5)$ bigger than or equal to 0.8, we need to choose n so that

$$1 - \frac{8}{n} \geqslant 0.8.$$

This yields $n \geqslant 40$. □

5. *[Exercise 8(a) on page 235 in the text]*

   Suppose that 30% of the items in a large manufactured lot are of poor quality. Suppose also that a random sample of n items is to be taken from the lot, and let Q_n denote the proportion of the items in the sample that are or poor quality. Use the Chebyshev inequality to find the value of n such that

$$\Pr(0.2 \leqslant Q_n \leqslant 0.4) \geqslant 0.75.$$

***Solution**:* From the Chebyshev inequality we obtain that

$$\Pr(|\overline{X}_n - \mu| < \varepsilon) \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Applying this inequality to the case in which $\overline{X}_n = Q_n$, $\mu = p = 0.3$, $\sigma^2 = p(1-p) = 0.21$ and $\varepsilon = 0.1$, we obtain that

$$\Pr(|Q_n - 0.3| < 0.1) \geqslant 1 - \frac{0.21}{n(0.1)^2}.$$

We then have that

$$\Pr(0.2 \leqslant Q_n \leqslant 0.4) \geqslant \Pr(|Q_n - 0.3| < 0.1) \geqslant 1 - \frac{21}{n}.$$

Thus, to make $\Pr(0.2 \leqslant Q_n \leqslant 0.4) \geqslant 0.75$, we need to choose n so that

$$1 - \frac{21}{n} \geqslant 0.75.$$

This yields $n \geqslant 84$. □