Solutions to Review Problems for Exam 3

1. Let X have mgf given by
$$\psi_X(t) = \frac{1}{3}e^t + \frac{2}{3}e^{2t}, \qquad \text{for } t \in \mathbb{R}. \tag{1}$$

   (a) Give the distribution of X

   ***Solution**:* The mgf in (1) corresponds to a discrete random variable with pmf
$$p_X(k) = \begin{cases} \dfrac{1}{3}, & \text{if } k = 1; \\[2ex] \dfrac{2}{3}, & \text{if } k = 2; \\[2ex] 0, & \text{elsewhere.} \end{cases}$$
□

   (b) Compute the expected value and variance of X.

   ***Solution**:* Compute the derivatives of the mgf in (1) to get
$$\psi_X'(t) = \frac{1}{3}e^t + \frac{4}{3}e^{2t}, \qquad \text{for } t \in \mathbb{R}, \tag{2}$$
and
$$\psi_X''(t) = \frac{1}{3}e^t + \frac{8}{3}e^{2t}, \qquad \text{for } t \in \mathbb{R}. \tag{3}$$
Using (2) and (3) we then obtain
$$E(X) = \psi_X'(0) = \frac{5}{3},$$
$$E(X^2) = \psi_X''(0) = 3.$$
Thus, the variance of X is
$$\text{Var}(X) = E(X^2) - (E(X))^2 = 3 - \frac{25}{9} = \frac{2}{9}.$$
□

2. Let X have mgf given by

$$\psi_X(t) = \begin{cases} \dfrac{e^t - e^{-t}}{2t}, & \text{if } t \neq 0; \\ \\ 1, & \text{if } t = 0. \end{cases} \tag{4}$$

   (a) Give the distribution of X

   ***Solution**:* Looking at the handout on special distributions we see that the mgf given in (4) corresponds to that of a uniform$(-1, 1)$ random variable. Thus, by the mgf uniqueness theorem, $X \sim \text{Uniform}(-1, 1)$, Consequently, the pdf of X is given by

$$f_X(x) = \begin{cases} \dfrac{1}{2}, & \text{if } -1 < x < 1; \\ \\ 0, & \text{elsewhere.} \end{cases}$$

$\square$

   (b) Compute the expected value and variance of X.

   ***Solution**:* The expected value and variance of X can also be obtained by reading the special distributions handout:

$$E(X) = \frac{-1+1}{2} = 0$$

   and

$$\text{Var}(X) = \frac{(1-(-1))^2}{12} = \frac{4}{12} = \frac{1}{3}.$$

$\square$

3. A random point (X, Y) is distributed uniformly on the square with vertices $(-1, -1)$, $(1, -1)$, $(1, 1)$ and $(-1, 1)$.

   (a) Give the joint pdf for X and Y.

   (b) Compute the following probabilities:

      (i) $\Pr(X^2 + Y^2 < 1)$,

      (ii) $\Pr(2X - Y > 0)$,

      (iii) $\Pr(|X + Y| < 2)$.

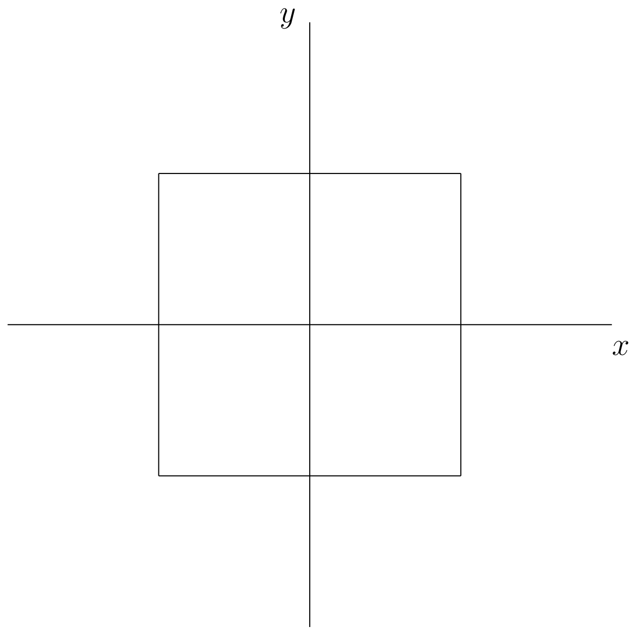

Figure 1: Sketch of square in Problem 3

Solution: The square is pictured in Figure 1 and has area 4.

(a) Consequently, the joint pdf of (X, Y) is given by

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{1}{4}, & \text{for } -1 < x < 1, -1 < y < 1; \\ \\ 0 & \text{elsewhere.} \end{cases} \tag{5}$$

(b) Denoting the square in Figure 1 by R, it follows from (5) that, for any subset A of $\mathbb{R}^2$,

$$\Pr[(x,y) \in A] = \iint_A f_{(X,Y)}(x,y)\, dxdy = \frac{1}{4} \cdot \text{area}(A \cap R); \tag{6}$$

that is, $\Pr[(x,y) \in A]$ is one–fourth the area of the portion of A in R.

We will use the formula in (6) to compute each of the probabilities in (i), (ii) and (iii).

(i) In this case, A is the circle of radius 1 around the origin in $\mathbb{R}^2$ and pictured in Figure 2.

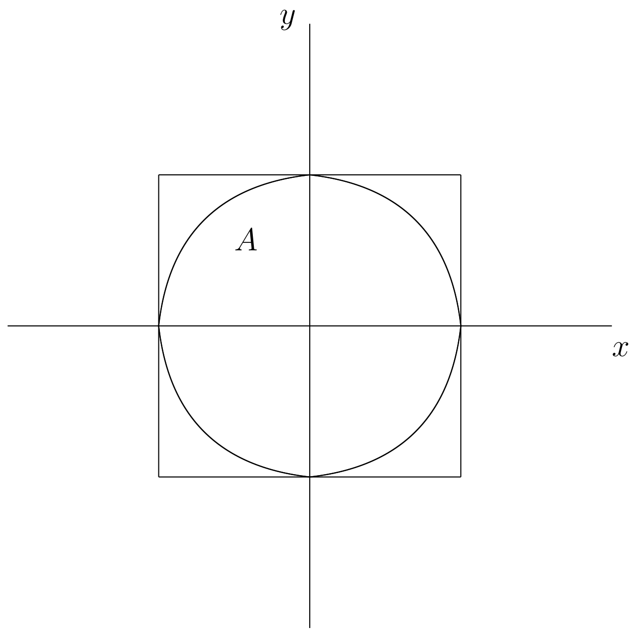

Figure 2: Sketch of A in Problem 3(b)(i)

Note that the circle A in Figure 2 is entirely contained in the square R so that, by the formula in (6),

$$\Pr(X^2 + Y^2 < 1) = \frac{\text{area}(A)}{4} = \frac{\pi}{4}.$$

(ii) The set A in this case is pictured in Figure 3 on page 5. Thus, in this case, $A \cap R$ is a trapezoid of area $2 \cdot \dfrac{\frac{1}{2} + \frac{3}{2}}{2} = 2$, so that, by the formula in (6),

$$\Pr(2X - Y > 0) = \frac{1}{4} \cdot \text{area}(A \cap R) = \frac{1}{2}.$$

(iii) In this case, A is the region in the xy–plane between the lines $x+y = 2$ and $x + y = -2$ (see Figure 4 on page 6). Thus, $A \cap R$ is R; so that, by the formula in (6),

$$\Pr(|X + Y| < 2) = \frac{\text{area}(R)}{4} = 1.$$

$\square$

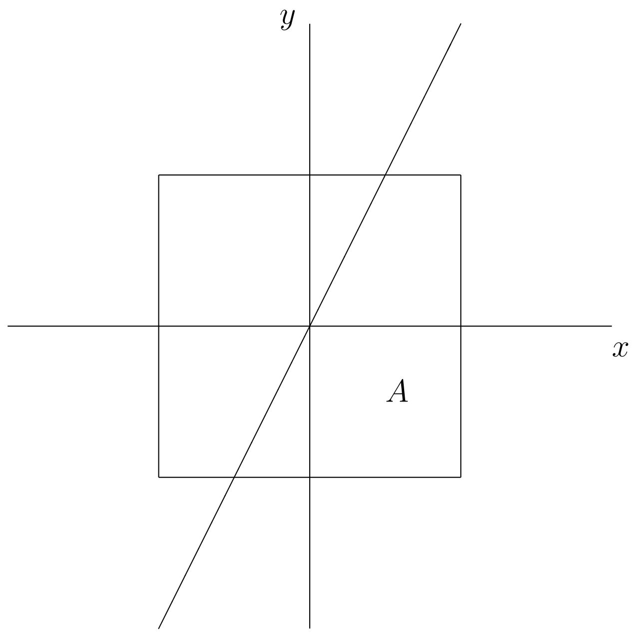

Figure 3: Sketch of A in Problem 3(b)(ii)

$X\backslash Y$	2	3	4
1	$\frac{1}{12}$	$\frac{1}{6}$	0
2	$\frac{1}{6}$	0	$\frac{1}{3}$
3	$\frac{1}{12}$	$\frac{1}{6}$	0

Table 1: Joint Probability Distribution for X and Y, $p_{(X,Y)}$

4. A random vector (X, Y) has the joint distribution shown in Table 1.

   (a) Show that X and Y are not independent.

   ***Solution**:* Table 2 shows the marginal distributions of X and Y on the margins.

   Observe from Table 2 that

$$p_{(X,Y)}(1, 4) = 0,$$

   while

$$p_X(1) = \frac{1}{4} \quad \text{and} \quad p_Y(4) = \frac{1}{3}.$$

   Thus,

$$p_X(1) \cdot p_Y(4) = \frac{1}{12};$$

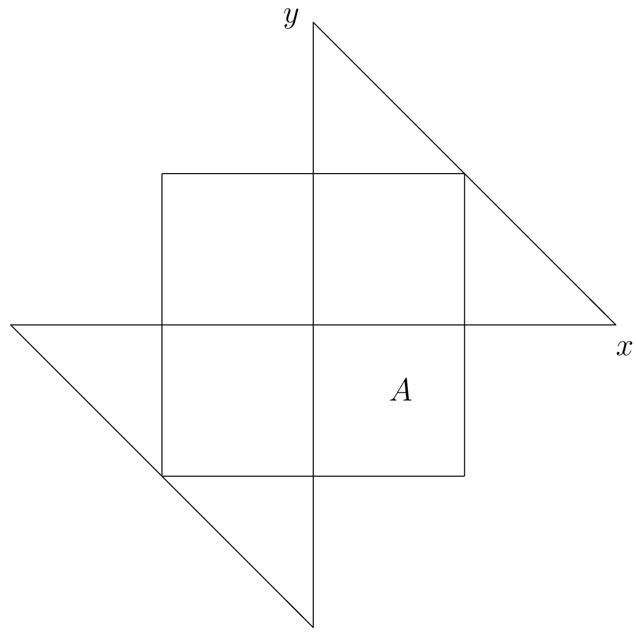

Figure 4: Sketch of A in Problem 3(b)(iii)

$X\backslash Y$	2	3	4	p_X
1	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
2	$\frac{1}{6}$	0	$\frac{1}{3}$	$\frac{1}{2}$
3	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
p_Y	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Table 2: Joint pdf for X and Y and marginal distributions p_X and p_Y

so that

$$p_{(X,Y)}(1,4) \neq p_X(1) \cdot p_Y(4),$$

and, therefore, X and Y are not independent. □

(b) Give a probability table for random variables U and V that have the same marginal distributions as X and Y, respectively, but are independent.

Solution: Table 3 on page 7 shows the joint pmf of (U,V) and the marginal distributions, p_U and p_V. □

5. An experiment consists of independent tosses of a fair coin. Let X denote the number of trials needed to obtain the first head, and let Y be the number of

$U\backslash V$	2	3	4	p_U
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$
p_V	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Table 3: Joint pdf for U and V and their marginal distributions.

trials needed to get two heads in repeated tosses. Are X and Y independent random variables?

***Solution**:* X has a geometric distribution with parameter $p = \dfrac{1}{2}$, so that

$$p_X(k) = \frac{1}{2^k}, \qquad \text{for } k = 1, 2, 3, \ldots \text{ and 0 elsewhere.} \tag{7}$$

On the other hand,

$$\Pr[Y = 2] = \frac{1}{4}, \tag{8}$$

since, in two repeated tosses of a coin, the events are HH, HT, TH and TT, and these events are equally likely.

Next, consider the joint event $(X = 2, Y = 2)$. Note that

$$(X = 2, Y = 2) = [X = 2] \cap [Y = 2] = \emptyset,$$

since $[X = 2]$ corresponds to the event TH, while $[Y = 2]$ to the event HH. Thus,

$$\Pr(X = 2, Y = 2) = 0,$$

while

$$p_X(2) \cdot p_Y(2) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16},$$

by (7) and (8). Thus,

$$p_{(X,Y)}(2, 2) \neq p_X(2) \cdot p_X(2).$$

Hence, X and Y are not independent. $\square$

6. Let $g(t)$ denote a non–negative, integrable function of a single variable with the property that

$$\int_0^\infty g(t) \;\; \mathrm{d}t = 1.$$

Define

$$f(x,y) = \begin{cases} \dfrac{2g(\sqrt{x^2+y^2})}{\pi\sqrt{x^2+y^2}}, & \text{for } 0 < x < \infty,\ 0 < y < \infty, \\ \\ 0, & \text{otherwise.} \end{cases}$$

Show that $f(x,y)$ is a joint pdf for two random variables X and Y.

***Solution**:* First observe that f is non–negative since g is non–negative. Next, compute

$$\iint_{\mathbb{R}^2} f(x,y)\ \mathrm{d}x\ \mathrm{d}y = \int_0^\infty \int_0^\infty \frac{2g(\sqrt{x^2+y^2})}{\pi\sqrt{x^2+y^2}}\ \mathrm{d}x\ \mathrm{d}y.$$

Switching to polar coordinates we then get that

$$\begin{aligned} \iint_{\mathbb{R}^2} f(x,y)\ \mathrm{d}x\ \mathrm{d}y &= \int_0^{\pi/2} \int_0^\infty \frac{2g(r)}{\pi r}\ r\ \mathrm{d}r\ \mathrm{d}\theta \\ &= \frac{\pi}{2} \int_0^\infty \frac{2}{\pi} g(r)\ \mathrm{d}r \\ &= \int_0^\infty g(r)\ \mathrm{d}r \\ &= 1; \end{aligned}$$

therefore, $f(x,y)$ is indeed a joint pdf for two random variables X and Y. $\square$

7. Suppose that two persons make an appointment to meet between 5 PM and 6 PM at a certain location and they agree that neither person will wait more than 10 minutes for each person. If they arrive independently at random times between 5 PM and 6 PM, what is the probability that they will meet?

   ***Solution**:* Let X denote the arrival time of the first person and Y that of the second person. Then, X and Y are independent and uniformly distributed on the interval (5 PM, 6 PM), in hours. It then follows that the joint pdf of X and Y is

$$f_{(X,Y)}(x,y) = \begin{cases} 1, & \text{if } 5 \text{ PM} < x < 6 \text{ PM}, 5 \text{ PM} < x < 6 \text{ PM}, \\ 0, & \text{elsewhere.} \end{cases}$$

Define $W = |X - Y|$; this is the time that one person would have to wait for the other one. Then, W takes on values, w, between 0 and 1 (in hours). The probability that that a person would have to wait more than 10 minutes is

$$\Pr(W > 1/6),$$

since the time is being measured in hours. It then follows that the probability that the two persons will meet is

$$1 - \Pr(W > 1/6) = \Pr(W \leqslant 1/6) = F_{_W}(1/6).$$

We will therefore need to find the cdf of W. To do this, we compute

$$\begin{aligned} \Pr(W \leqslant w) &= \Pr(|X - Y| \leqslant w), \quad \text{for } 0 < w < 1, \\ &= \iint_A f_{_{(X,Y)}}(x, y) \; \mathrm{d}x \; \mathrm{d}y, \end{aligned}$$

where A is the event

$$A = \{(x, y) \in \mathbb{R}^2 \mid 5 \text{ PM} < x < 6 \text{ PM}, 5 \text{ PM} < y < 6 \text{ PM}, |x - y| \leqslant w\}.$$

This event is pictured in Figure 5.

We then have that

$$\begin{aligned} \Pr(W \leqslant w) &= \iint_A \; \mathrm{d}x \; \mathrm{d}y \\ &= \text{area}(A), \end{aligned}$$

where the area of A can be computed by subtracting from 1 the area of the two corner triangles shown in Figure 5:

$$\begin{aligned} \Pr(W \leqslant w) &= 1 - (1 - w)^2 \\ &= 2w - w^2. \end{aligned}$$

Consequently, $F_{_W}(w) = 2w - w^2$ for $0 < w < 1$. Thus the probability that the two persons will meet is

$$F_{_W}(1/6) = 2 \cdot \frac{1}{6} - \left(\frac{1}{6}\right)^2 = \frac{11}{36},$$

or about 30.56%. $\square$

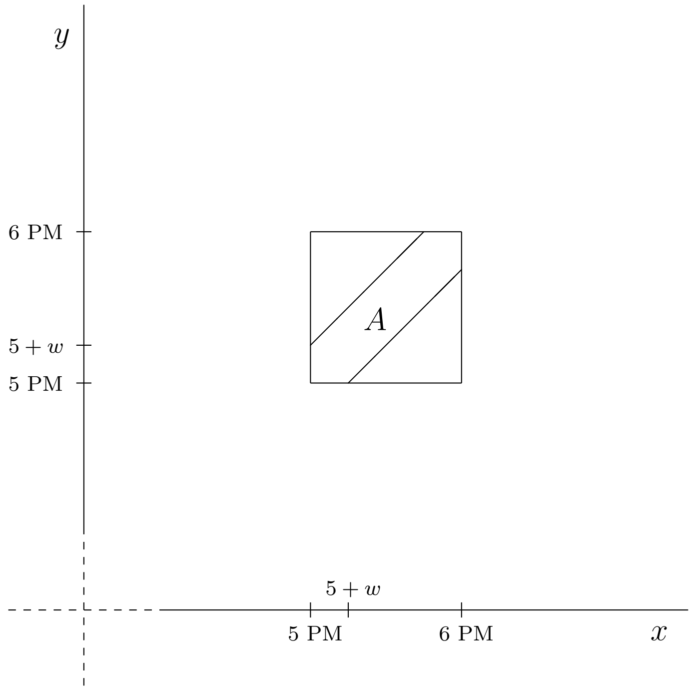

Figure 5: Event A in the xy–plane

8. Assume that the number of calls coming per minute into a hotel's reservation center follows a Poisson distribution with mean 3.

   (a) Find the probability that no calls come in a given 1 minute period.

   Solution: Let Y denote the number of calls that come to the hotel's reservation center in one minute. Then, $Y \sim \text{Poisson}(3)$; so that,

   $$p_{_Y}(k) = \frac{3^k}{k!}\, e^{-3}, \quad \text{for } k = 0, 1, 2, \ldots$$

   Then, the probability that no calls will come in the given minute is

   $$\Pr(Y = 0) = p_{_Y}(0) = e^{-3} \approx 0.05,$$

   or about 5%. □

   (b) Assume that the number of calls arriving in two different minutes are independent. Find the probability that at least two calls will arrive in a given two minute period.

***Solution**:* Let Y_1 denote the number of calls that arrive in one minute and Y_2 denote the number of calls that arrive in another minute. We then have that

$$Y_i \sim \text{Poisson}(3), \quad \text{for } i = 1, 2,$$

and Y_i and Y_2 are independent. We want to compute

$$\Pr(Y_1 + Y_2 \geqslant 2).$$

To do this, we determine the distribution of $W = Y_1 + Y_2$.
Since Y_1 and Y_2 are independent,

$$\psi_{_W}(t) = \psi_{_{Y_1+Y_2}}(t) = \psi_{_{Y_1}}(t) \cdot \psi_{_{Y_2}}(t);$$

so that,

$$\psi_{_W}(t) = e^{3(e^t-1)} \cdot e^{3(e^t-1)} = e^{6(e^t-1)},$$

which is the mgf of a Poisson(6) distribution. Thus, by the mgf Uniqueness Theorem, $W \sim \text{Poisson}(6)$. We then have that

$$p_{_W}(k) = \frac{6^k}{k!}\, e^{-6}, \quad \text{for } k = 0, 1, 2, \ldots$$

Therefore,

$$\begin{aligned}
\Pr(Y_1 + Y_2 \geqslant 2) &= \Pr(W \geqslant 2) \\
&= 1 - \Pr(W < 2) \\
&= 1 - \Pr(W = 0) - \Pr(W = 1) \\
&= 1 - e^{-6} - 6e^{-6} \\
&= 1 - \frac{7}{e^6} \\
&\approx 0.9826.
\end{aligned}$$

Hence, the probability that at least two calls will arrive in a given two minute period is about 98.3%. □

9. Let $Y \sim \text{binomial}(100, 1/2)$. Use the central limit theorem to estimate the value of $\Pr(Y = 50)$.

   Suggestion: Observe that $\Pr(Y = 50) = \Pr(49.5 < Y \leq 50.5)$, since Y is discrete.

   ***Solution**:* We use the central limit theorem to estimate

$$\Pr(49.5 < Y \leqslant 50.5).$$

   By the central limit theorem,

$$\Pr(49.5 < Y \leqslant 50.5) \approx \Pr\left(\frac{49.5 - n\mu}{\sqrt{n}\sigma} < Z \leqslant \frac{50.5 - n\mu}{\sqrt{n}\sigma}\right), \tag{9}$$

   where $Z \sim \text{normal}(0, 1)$, $n = 100$, and $n\mu = 50$ and

$$\sigma = \sqrt{\frac{1}{2}\left(1 - \frac{1}{2}\right)} = \frac{1}{2}.$$

   We then obtain from (9) that

$$\begin{aligned} \Pr(49.5 < Y \leqslant 50.5) &\approx \Pr(-0.1 < Z \leqslant 0.1) \\ &\approx F_{Z}(0.1) - F_{Z}(-0.1) \\ &\approx 2F_{Z}(0.1) - 1 \\ &\approx 2(0.5398) - 1 \\ &\approx 0.0796. \end{aligned}$$

   Thus,

$$\Pr(Y = 50) \approx 0.08,$$

   or about 8%. □

10. Roll a balanced die 36 times. Let Y denote the sum of the outcomes in each of the 36 rolls. Estimate the probability that $108 \leq Y \leq 144$.

    Suggestion: Since the event of interest is $(Y \in \{108, 109, \ldots, 144\})$, rewrite $\Pr(108 \leq Y \leq 144)$ as

$$\Pr(107.5 < Y \leqslant 144.5).$$

Solution: Let $X_1, X_2, \ldots, X_n$, where $n = 36$, denote the outcomes of the 36 rolls. Since we are assuming that the die is balanced, the random variables $X_1, X_2, \ldots, X_n$ are identically uniformly distributed over the digits $\{1, 2, \ldots, 6\}$; in other words, $X_1, X_2, \ldots, X_n$ is a random sample from the discrete uniform(6) distribution. Consequently, the mean of the distribution is

$$\mu = \frac{6+1}{2} = 3.5, \tag{10}$$

and the variance is

$$\sigma^2 = \frac{(6+1)(6-1)}{12} = \frac{35}{12}. \tag{11}$$

We also have that

$$Y = \sum_{k=1}^{n} X_k,$$

where $n = 36$.

By the central limit theorem,

$$\Pr(107.5 < Y \leqslant 144.5) \approx \Pr\left(\frac{107.5 - n\mu}{\sqrt{n}\sigma} < Z \leqslant \frac{144.5 - n\mu}{\sqrt{n}\sigma}\right), \tag{12}$$

where $Z \sim \text{normal}(0,1)$, $n = 36$, and μ and σ are given in (10) and (11), respectively. We then have from (12) that

$$\begin{aligned}
\Pr(107.5 < Y \leqslant 144.5) &\approx \Pr(-1.81 < Z \leqslant 1.81) \\
&\approx F_Z(1.81) - F_Z(-1.81) \\
&\approx 2F_Z(1.81) - 1 \\
&\approx 2(0.9649) - 1 \\
&\approx 0.9298;
\end{aligned}$$

so that the probability that $108 \leqslant Y \leqslant 144$ is about 93%. □

11. The standard voltage in residences in the United States of America is 120 volts. Assume that this voltage can be modeled by a random variable with mean 120 and variance 25. Suppose that some sensitive electrical appliances can be damaged if the voltage is not between 110 and 130. Use Chebyshev's inequality

to find an upper bound for the probability that damage will occur to a sensitive electrical appliance.

***Solution**:* Let X denote the voltage at a residence. We are assuming that X is a random variable with mean $\mu = 120$ and variance $\sigma^2 = 25$. We find an upper bound for the probability

$$\Pr(X < 110 \text{ or } X > 130). \tag{13}$$

The probability in (13) is the same as the probability

$$\Pr(X - \mu < -10 \text{ or } X - \mu > 10);$$

so that,

$$\Pr(X < 110 \text{ or } X > 130) = \Pr(|X - \mu| > 10).$$

Thus, bu the monotonicity property of probability,

$$\Pr(X < 110 \text{ or } X > 130) \leqslant \Pr(|X - \mu| \geqslant 10). \tag{14}$$

It follows from Chebyshev's inequality that

$$\Pr(|X - \mu| \geqslant 10) \leqslant \frac{1}{10^2}\mathrm{Var}(X);$$

so that,

$$\Pr(|X - \mu| \geqslant 10) \leqslant \frac{25}{100} = \frac{1}{4}.$$

Hence, in view of (14),

$$\Pr(X < 110 \text{ or } X > 130) \leqslant \frac{1}{4}.$$

Therefore, the probability that damage will occur to a sensitive electrical appliance is at most 25%. □

12. Many random number generators, like the `RAND()` function in MS Excel, are pseudo–random number generators. These are algorithms that provide a (real) random number in the interval $(0, 1)$. Many of these pseudo–random numbers can be modeled by a uniform$(0, 1)$ random variable.

    Suppose a pseudo–random number generator is used to generate 400 random numbers from the interval $[0, 1]$.

(a) Use Chebyshevs inequality to find a lower bound for the probability that the sum of the numbers lies between 190 and 210.

(b) Use the central limit theorem to estimate the probability that the sum of the numbers lies between 190 and 210.

***Solution**:* Let $U_1, U_2, \ldots, U_n$ be iid uniform$(0,1)$ random variables, where $n = 400$. Then,

$$\mu = E(U_k) = \frac{1}{2}, \quad \text{for all } k,$$

and

$$\sigma^2 = \text{Var}(U_k) = \frac{1}{12}, \quad \text{for all } k.$$

These values were obtained using the special distributions sheet.

Define

$$Y = \sum_{k=1}^{n} U_k, \tag{15}$$

where $n = 400$.

(a) We use Chebyshevs inequality to find a lower bound for the probability

$$\Pr(190 < Y < 210). \tag{16}$$

The probability in (16) can be written as

$$\begin{aligned} \Pr(190 < Y < 210) &= \Pr(190 - 200 < Y - n\mu < 210 - 200) \\ &= \Pr(-10 < Y - n\mu < 10) \\ &= \Pr(|Y - n\mu| < 10); \end{aligned}$$

so that, using the complement rule of probability

$$\Pr(190 < Y < 210) = 1 - \Pr(|Y - n\mu| \geqslant 10). \tag{17}$$

Next, use Chebyshev's inequality to estimate the right–most probability in (17) to get

$$\Pr(|Y - n\mu| \geqslant 10) \leqslant \frac{1}{10^2}\text{Var}(Y),$$

where

$$\text{Var}(Y) = n\sigma^2 = (400)\frac{1}{12} \approx 33.33;$$

so that, approximately,

$$\Pr(|Y - n\mu| \geqslant 10) \leqslant 0.3333. \tag{18}$$

Combining (17) and the estimate in (18), we obtain the estimate

$$\Pr(190 < Y < 210) \geqslant 1 - 0.3333,$$

or

$$\Pr(190 < Y < 210) \geqslant 0.6667. \tag{19}$$

Thus, a lower bound for the probability that the sum of the numbers lies between 190 and 210 is about 67%.

(b) We apply the central limit theorem to estimate

$$\Pr(190 < Y < 210) = \Pr(190 < Y \leqslant 210), \tag{20}$$

where Y is given in (15); observe that Y is a continuous random variable. Rewrite the expression in (20) as follows:

$$\begin{aligned} \Pr(190 < Y < 210) &= \Pr\left(\frac{190-200}{20(0.2887)} < \frac{Y - n\mu}{\sqrt{n}\sigma} \leqslant \frac{210-200}{20(0.2887)}\right) \\ &\approx \Pr\left(-1.73 < \frac{Y - n\mu}{\sqrt{n}\sigma} \leqslant 1.73\right). \end{aligned}$$

Thus, applying the central limit theorem, we get the estimate

$$\Pr(190 < Y < 210) \approx \Pr\left(-1.73 < Z \leqslant 1.73\right),$$

where $Z \sim \text{normal}(0,1)$, or

$$\Pr(190 < Y < 210) \approx F_Z(1.73) - F_Z(-1.73)$$

Using the `NORMDIST` function in MS Excel, we obtain that

$$\Pr(190 < Y < 210) \approx 0.9582 - 0.0418,$$

or

$$\Pr(190 < Y < 2010) \approx 0.9064,$$

or about 91%. We note that this estimate is bigger than the lower bound of 67% obtained in part (a) using Chebyshev's inequality, which to be expected.

$\square$