

# Statistics 100A

## Homework 5 Solutions

Ryan Rosario

### Chapter 5

1. Let  $X$  be a random variable with probability density function

$$f(x) = \begin{cases} c(1-x^2) & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- (a) What is the value of  $c$ ?

We know that for  $f(x)$  to be a probability distribution  $\int_{-\infty}^{\infty} f(x)dx = 1$ . We integrate  $f(x)$  with respect to  $x$ , set the result equal to 1 and solve for  $c$ .

$$\begin{aligned} 1 &= \int_{-1}^1 c(1-x^2)dx \\ &= \left[ cx - c\frac{x^3}{3} \right]_{-1}^1 \\ &= \left( c - \frac{c}{3} \right) - \left( -c + \frac{c}{3} \right) \\ &= \frac{2c}{3} - \frac{-2c}{3} \\ &= \frac{4c}{3} \\ c &= \frac{3}{4} \end{aligned}$$

Thus,  $\boxed{c = \frac{3}{4}}$ .

- (b) What is the cumulative distribution function of  $X$ ?

We want to find  $F(x)$ . To do that, integrate  $f(x)$  from the lower bound of the domain on which  $f(x) \neq 0$  to  $x$  so we will get an expression in terms of  $x$ .

$$\begin{aligned} F(x) &= \int_{-1}^x c(1-x^2)dx \\ &= \left[ cx - \frac{cx^3}{3} \right]_{-1}^x \\ &\quad \text{But recall that } c = \frac{3}{4}. \\ &= \frac{3}{4}x - \frac{1}{4}x^3 + \frac{1}{2} \\ &= \boxed{\begin{cases} \frac{3}{4} \left( x - \frac{x^3}{3} + \frac{2}{3} \right) & -1 < x < 1 \\ 0 & \text{elsewhere} \end{cases}} \end{aligned}$$

4. The probability density function of  $X$ , the lifetime of a certain type of electronic device (measured in hours), is given by,

$$f(x) = \begin{cases} \frac{10}{x^2} & x > 10 \\ 0 & x \leq 10 \end{cases}$$

- (a) Find  $P(X > 20)$ .

There are two ways to solve this problem, and other problems like it. We note that the area we are interested in is bounded below by 20 and unbounded above. Thus,

$$P(X > c) = \int_c^{\infty} f(x)dx$$

Unlike in the discrete case, there is not really an advantage to using the complement, but you can of course do so. We could consider  $P(X > c) = 1 - P(X < c)$ ,

$$P(X > c) = 1 - P(X < c) = 1 - \int_{-\infty}^c f(x)dx$$

$$\begin{aligned} P(X > 20) &= \int_{20}^{\infty} \frac{10}{x^2} dx \\ &= \left[ -\frac{10}{x} \right]_{20}^{\infty} \\ &= \lim_{x \rightarrow \infty} \left( -\frac{10}{x} \right) - \left( -\frac{1}{2} \right) \\ &= \boxed{\frac{1}{2}} \end{aligned}$$

- (b) What is the cumulative distribution function of  $X$ ?

We want to find  $F(x)$ . To do that, integrate  $f(x)$  from the lower bound of the domain on which  $f(x) \neq 0$  to  $x$  so we will get an expression in terms of  $x$ .

$$\begin{aligned} F(x) &= \int_{10}^x \frac{10}{x^2} dx \\ &= \left[ -\frac{10}{x} \right]_{10}^x \\ &= -\frac{10}{x} - (-1) \\ &= \boxed{\begin{cases} 1 - \frac{10}{x} & x > 10 \\ 0 & x \leq 10 \end{cases}} \end{aligned}$$

- (c) *What is the probability that of 6 such types of devices at least 3 will function for at least 15 hours? What assumptions are you making?*

This is another one of those problems that must be considered hierarchically in steps. Break the problem up into two steps: finding the probability that 1 device functions for at least 15 hours, and finding the probability that 3 of the 6 devices function for at least 15 hours.

### Step 1: One Device

First we must find the probability that one randomly selected device will function for at least 15 hours. It is given that each device follows a probability distribution  $f(x)$  given earlier. Let  $X$  be the number of hours that the device functions. Then,

$$\begin{aligned} P(X \geq 15) &= \int_{15}^{\infty} \frac{10}{x^2} dx \\ &= \left[ -\frac{10}{x} \right]_{15}^{\infty} \\ &= \lim_{x \rightarrow \infty} -\frac{10}{x} - \left( -\frac{10}{15} \right) \\ &= -\lim_{x \rightarrow \infty} \frac{10}{x} + \frac{2}{3} \\ &= \boxed{\frac{2}{3}} \end{aligned}$$

### Step 2: 3 of the 6 Devices

We are interested in  $n = 6$  devices. Note that each device constitutes a trial with one of two outcomes: device functions at least 15 hours or it doesn't. Each of the 6 trials are Bernoulli trials **and we assume the trials are independent**. We also have a constant probability of success  $p = \frac{2}{3}$ . **We use the binomial model.**

Let  $X$  be the number of devices that function at least 15 hours. Then,

$$\begin{aligned} P(X \geq 3) &= \binom{6}{3} \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^3 + \binom{6}{4} \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^2 + \binom{6}{5} \left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right)^1 \\ &\quad + \binom{6}{6} \left(\frac{2}{3}\right)^6 \left(\frac{1}{3}\right)^0 \\ &= \frac{656}{729} \\ &\approx \boxed{0.9} \end{aligned}$$

6. Compute  $E(X)$  if  $X$  has a density function given by

$$(a) \quad f(x) = \begin{cases} \frac{1}{4}xe^{-\frac{x}{2}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Recall that by definition,

$$E(X) = \int_a^b xf(x)dx$$

Then, for the PDF given in this problem,

$$E(X) = \int_0^\infty x \left( \frac{1}{4}xe^{-\frac{x}{2}} \right) dx = \frac{1}{4} \int_0^\infty x^2 e^{-\frac{x}{2}} dx$$

We use what my Calculus II professor called a “tricky  $u$ -substitution.” First, let  $y = \frac{x}{2}$ . Then  $x = 2y$  and  $dx = 2dy$ . The 4 from  $x^2 = (2y)^2 = 4y^2$  cancels the  $\frac{1}{4}$  in front of the integral, and we must carry the 2 from  $dx$  outside the integral to get,

$$E(X) = \frac{1}{4} \int_0^\infty x^2 e^{-\frac{x}{2}} dx = \frac{1}{4} \int_0^\infty (2y)^2 e^{-y} (2dy) = 2 \int_0^\infty y^2 e^{-y} dy$$

Recall that when we have an integral of the form  $h(x) = \int f(x)g(x)dx$  we use integration by parts. We choose values for  $u$  and  $dv$ . Then we compute  $du$  and  $v$  and substitute into  $h(x) = uv - \int vdu$ .

Now use integration by parts, with  $u = y^2, dv = e^{-y}$  so  $du = 2ydy, v = -e^{-y}$ . Then,

$$\begin{aligned} E(X) &= 2 \int_0^\infty y^2 e^{-y} dy \\ &= 2 \left[ -y^2 e^{-y} - \int -e^{-y} (2y) dy \right] \\ &= 2 \left[ -y^2 e^{-y} + 2 \int y e^{-y} dy \right] \\ &\quad \text{integrate by parts again. Use: } u = y, dv = e^{-y} \text{ thus } du = dy, v = -e^{-y} \\ &= 2 \left[ -y^2 e^{-y} + 2 \left\{ -y e^{-y} - \int -e^{-y} dy \right\} \right] \\ &= 2 \left[ -y^2 e^{-y} + 2 \left\{ -y e^{-y} - e^{-y} \right\} \right] \\ &= -2y^2 e^{-y} - 4y e^{-y} - 4e^{-y} \\ &= [-2e^{-y} (y^2 + 2y + 2)]_0^\infty \\ &= \lim_{x \rightarrow \infty} [-2e^{-y} (y^2 + 2y + 2)] - [-2e^{-y} (y^2 + 2y + 2)]_{x=0} \\ &\quad \text{The limit evaluates to an indeterminate form } 0 \cdot \infty. \\ &= -2 \lim_{x \rightarrow \infty} \left[ \frac{1}{e^y} (y^2 + 2y + 2) \right] - [-2e^{-y} (y^2 + 2y + 2)]_{x=0} \\ &\quad \text{The limit evaluates to } \frac{\infty}{\infty} \text{ so use l'Hôpital's Rule.} \\ &\stackrel{L'H}{=} -2 \lim_{x \rightarrow \infty} \frac{2y + 2}{e^y} + 2 [e^{-y} (y^2 + 2y + 2)]_{x=0} \\ &\quad \text{The limit evaluates to } \frac{\infty}{\infty} \text{ still so use l'Hôpital's Rule again.} \\ &\stackrel{L'H}{=} -2 \lim_{x \rightarrow \infty} \underbrace{\frac{2}{e^y}}_{=0} + 2 [e^{-y} (y^2 + 2y + 2)]_{x=0} \\ &= \boxed{4} \end{aligned}$$

$$(b) \ f(x) = \begin{cases} c(1-x^2) & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(X) &= \int_{-1}^1 x [c(1-x^2)] dx \\ &= c \int_{-1}^1 x(1-x^2) dx \\ &= c \int_{-1}^1 x - x^3 dx \\ &= c \left[ \frac{x^2}{2} - \frac{x^4}{4} \right]_{-1}^1 \\ &= c \left\{ \left[ \frac{1}{2} - \frac{1}{4} \right] - \left[ \frac{1}{2} - \frac{1}{4} \right] \right\} \\ &= \boxed{0} \end{aligned}$$

We know the value of  $c$  from a previous problem, but it does not matter here!

$$(c) \ f(x) = \begin{cases} \frac{5}{x^2} & x > 5 \\ 0 & x \leq 5 \end{cases}$$

$$\begin{aligned} E(X) &= \int_5^\infty x \frac{5}{x^2} dx \\ &= \int_5^\infty \frac{5}{x} dx \\ &= 5 \int_5^\infty \frac{1}{x} dx \\ &= 5 [\ln x]_5^\infty \\ &= 5 \left\{ \left( \lim_{x \rightarrow \infty} \ln x \right) - \ln 5 \right\} \\ &\rightarrow \boxed{\infty} \end{aligned}$$

7. The density function of  $X$  is given by

$$f(x) = \begin{cases} a + bx^2 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

if  $E(X) = \frac{3}{5}$ , find  $a$  and  $b$ .

There are two things we have to worry about here. First, we must find  $a$  and  $b$  such that  $E(X) = \frac{3}{5}$ . Additionally, we must find values for  $a$  and  $b$  such that  $f(x)$  is a probability distribution. That is,  $\int_0^1 f(x)dx = 1$ .

First, find the expected value. Recall that,

$$E(X) = \int_a^b xf(x)dx$$

where  $a$  and  $b$  are the lower and upper boundary points of the domain on which  $f(x)$  is non-zero, and the values may be infinite.

$$\begin{aligned} E(X) &= \int_0^1 x(a + bx^2)dx \\ &= \int_0^1 ax + bx^3dx \\ &= \left[ \frac{ax^2}{2} + \frac{bx^4}{4} \right]_0^1 \\ &= \frac{a}{2} + \frac{b}{4} \end{aligned}$$

So we must solve  $\frac{a}{2} + \frac{b}{4} = \frac{3}{5}$ .

Now we impose the restriction that  $\int_0^1 f(x)dx = 1$ .

$$\begin{aligned} \int_0^1 a + bx^2dx &= \left[ ax + b\frac{x^3}{3} \right]_0^1 \\ &= a + \frac{b}{3} \end{aligned}$$

We must solve  $a + \frac{b}{3} = 1$ . Since we have two equations and we must solve for  $a$  and  $b$ , we use a system of linear equations. That is,

$$\begin{cases} \frac{1}{2}a + \frac{1}{4}b &= \frac{3}{5} \\ a + \frac{1}{3}b &= 1 \end{cases}$$

First, let  $R_2 = R_2 - 2R_1$ , to get that  $-\frac{1}{6}b = -\frac{1}{5}$ . Thus,  $b = \frac{6}{5}$ .

Plug in  $b = \frac{6}{5}$  into  $R_2$  to solve  $a + \frac{1}{3}\frac{6}{5} = 1$ . Thus,  $a = \frac{3}{5}$ . You can verify on your own that these values for  $a$  and  $b$  are valid for both  $R_1$  and  $R_2$ .

We see that  $\boxed{a = \frac{3}{5}, \quad b = \frac{6}{5}}$ .

8. The lifetime in hours of an electronic tube is a random variable having a probability density function given by

$$f(x) = xe^{-x} \quad x \geq 0$$

Compute the expected value lifetime of such a tube.

We must solve,

$$E(x) = \int_0^{\infty} x^2 e^{-x} dx$$

Let  $u = x^2$ ,  $dv = e^{-x}$ , then  $du = 2xdx$ ,  $v = -e^{-x}$ .

Then,

$$E(x) = \left[ \int_0^{\infty} x^2 e^{-x} dx = -x^2 e^{-x} - \int -2xe^{-x} dx = -x^2 e^{-x} + 2 \underbrace{\int xe^{-x} dx}_{\text{integrate by parts}} \right]_0^{\infty}$$

which we integrate using integration by parts again. Our substitutions must remain consistent with the first integration, so  $u = x$ ,  $dv = e^{-x}$  and  $du = dx$ ,  $v = -e^{-x}$  giving us,

$$-x^2 e^{-x} + 2 \left\{ -xe^{-x} - \int -e^{-x} dx \right\} = -x^2 e^{-x} + 2 \left\{ -xe^{-x} + \underbrace{\int e^{-x} dx}_{=-e^{-x}} \right\}$$

so,

$$\begin{aligned} E(x) &= \left[ -x^2 e^{-x} + 2 \{ -xe^{-x} + -e^{-x} \} \right]_0^{\infty} \\ &= \left[ -e^{-x} (x^2 + 2x + 2) \right]_0^{\infty} \\ &= \lim_{x \rightarrow \infty} \left[ -e^{-x} (x^2 + 2x + 2) \right] - \left[ -e^{-x} (x^2 + 2x + 2) \right]_{x=0} \\ &\quad \text{the limit evaluates to } 0 \cdot \infty \text{ so we must invert of the terms and try again.} \\ &= \lim_{x \rightarrow \infty} \left[ -\frac{1}{e^x} (x^2 + 2x + 2) \right] - \left[ -e^{-x} (x^2 + 2x + 2) \right]_{x=0} \\ &\quad \text{now the limit evaluates to } \frac{\infty}{\infty} \text{ so we use l'Hôpital's Rule.} \\ &\stackrel{L'H}{=} - \lim_{x \rightarrow \infty} \left[ \frac{2x + 2}{e^x} \right] - \left[ -e^{-x} (x^2 + 2x + 2) \right]_{x=0} \\ &\quad \text{the limit still evaluates to } \frac{\infty}{\infty} \text{ so we use l'Hôpital's Rule again.} \\ &\stackrel{L'H}{=} - \lim_{x \rightarrow \infty} \left[ \frac{2}{e^x} \right] - \left[ -e^{-x} (x^2 + 2x + 2) \right]_{x=0} \\ &= \left[ e^{-x} (x^2 + 2x + 2) \right]_{x=0} \\ &= \boxed{2} \end{aligned}$$

10. *Trains headed for destination A arrive at the train station at 15-minute intervals starting at 7 AM, whereas trains headed to destination B arrive at 15-minute intervals starting at 7:05 AM.*

- (a) *If a certain passenger arrives at the station at a time uniformly distributed between 7 and 8 AM and then gets on the first train that arrives, what proportion of time does he or she go to destination A?*

For this random person to randomly wander onto the train going to destination A, he/she must arrive right after the train to B leaves, but before the next train to A leaves. This means the person must arrive some time between 7:05-7:15, or between 7:20-7:30 or between 7:35-7:45 or between 7:50-8:00. The person must arrive during one of these 10 minute windows, and there are 4 of these non-overlapping windows. Thus, the probability that this random person will arrive at location A is,

$$P(A) = \frac{40}{60} = \frac{2}{3}$$

- (b) *What if the passenger arrives at a time uniformly distributed between 7:10 and 8:10 AM?*

All of the 10 minute windows in the previous part still apply, except for the first. To randomly jump onto a train headed to location A, the person must arrive between 7:10-7:15, 7:20-7:30, 7:35-7:45, 7:50-8:00 **or** between 8:10-8:15. So we have a total of 3 windows of 10 minutes each, and 2 windows of 5 minutes each. There are a total of 40 minutes (still) during which a person could arrive and end up at location A. Therefore,

$$P(A) = \frac{40}{60} = \frac{2}{3}$$



13. You arrive at a bus stop at 10 o'clock, knowing that the bus will arrive at some time uniformly distributed between 10 and 10:30.

(a) What is the probability that you will have to wait longer than 10 minutes?

This person arrives at the bus stop at 10am. We want to find the probability that the person has to wait longer than 10 minutes. That is, the bus arrives after 10:10 but before or at 10:30. Let  $X$  be the waiting time for the bus and let  $X \sim U(0, 30)$ . Then to wait 10 minutes means that the bus arrives in the last 20 minutes of this 30 minute domain. For the Uniform distribution, we know that  $P(X \leq x) = \frac{x}{b-a}$ . Therefore,

$$P(X > 10) = 1 - P(X \leq 10) = 1 - \frac{10}{30 - 0} = \frac{20}{30} = \boxed{\frac{2}{3}}$$

(b) If at 10:15 the bus has not yet arrived, what is the probability that you will have to wait at least an additional 10 minutes?

Let  $X$  again be the number of minutes the person waits for the bus. Since it is 10:15, we know the person has been waiting at least 15 minutes. That is,  $X \geq 15$ . We want to find the probability that the person waits at least 10 more minutes until the bus arrives. In total, the person will have waited at least 25 minutes in total. That is, we want to find  $P(X \geq 25 | X \geq 15)$ . Using Bayes' rule,

$$P(X \geq 25 | X \geq 15) = \frac{P(X \geq 25 \cap X \geq 15)}{P(X \geq 15)} = \frac{P(X \geq 25)}{P(X \geq 15)}$$

The bus arrives uniformly within a 30 minute period so for a person to wait at least 25 minutes means that the bus still has 5 more minutes to arrive. If a person waits at least 15 minutes, then the bus still has 15 minutes to arrive. That is,

$$P(X \geq 25 | X \geq 15) = \frac{P(X \geq 25)}{P(X \geq 15)} = \frac{\frac{5}{30}}{\frac{15}{30}} = \frac{5}{15} = \boxed{\frac{1}{3}}$$

14. Let  $X$  be a uniform  $(0, 1)$  random variable. Compute  $E(X^n)$  by using Proposition 2.1 and then check the result by using the definition of expectation.

By definition of expectation of a  $U(a, b)$  random variable, we know that

$$E(X) = \int_a^b x \frac{1}{b-a} dx$$

So in our case,

$$E(X) = \int_0^1 x \frac{1}{1-0} dx = \int_0^1 x dx$$

Using the definition of expectation, we find  $E(X^n)$  by simply replacing the the stray  $x$  in the integral by  $x^n$ .

$$E(X^n) = \int_0^1 x^n dx = \left[ \frac{x^{n+1}}{n+1} \right]_0^1 = \boxed{\frac{1}{n+1}}$$

Proposition 2.1 applied to the uniform says the following. We have that  $X$  is a continuous random variable with the uniform probability density function  $f(x)$ . Then for any  $g(x)$  (real),  $E(g(X)) = \int_a^b g(x) \frac{1}{b-a} dx$ . Simply let  $g(x) = x^n$  and voilà we are done.

$$E(X^n) = \int_0^1 x^n dx = \boxed{\frac{1}{n+1}}$$

15. If  $X$  is a normal random variable with parameters  $\mu = 10$  and  $\sigma^2 = 36$ , compute

Whenever  $\mu$  and  $\sigma$  are given, we know we are working with the normal distribution. The PDF for the normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The trick to this problem is that we are going to use  $\Phi$  to represent the CDF of the normal distribution so no integration is necessary. That is,

$$P(Z \leq z) = \Phi(z)$$

The value at which we compute  $\Phi$  is called a  $z$ -score and,

$$z = \frac{x - \mu}{\sigma}$$

The value  $\Phi(z)$  can be found by using a standard normal distribution  $z$ -table, using a calculator with the Normal CDF or by integration.

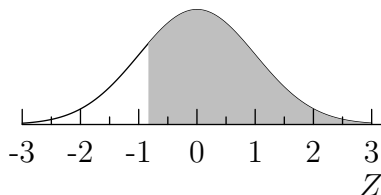
Also, remember that for continuous distributions,  $P(Z \leq z) = P(Z < z)$  because  $P(Z = z) = 0$ . I may use these interchangeably.

- (a)  $P(X > 5)$

Recall that  $P(X > 5) = 1 - P(X \leq 5)$  and  $P(Z \leq z) = \Phi(z)$ .

Then,

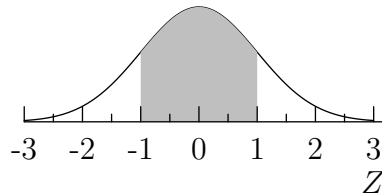
$$\begin{aligned} P(X > 5) &= 1 - P(X \leq 5) \\ &= 1 - \Phi\left(\frac{5 - 10}{6}\right) \\ &= 1 - \Phi\left(-\frac{5}{6}\right) \\ &\approx 1 - 0.202 \\ &= \boxed{0.798} \end{aligned}$$



(b)  $P(4 < X < 16)$

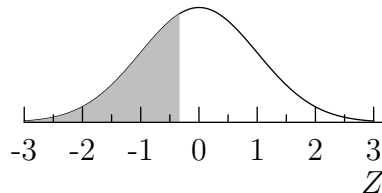
$$\begin{aligned}P(4 < X < 16) &= P(X < 16) - P(X < 4) \\&= \Phi\left(\frac{16 - 10}{6}\right) - \Phi\left(\frac{4 - 10}{6}\right) \\&= \Phi(1) - \Phi(-1) \\&\approx 0.841 - 0.159 \\&= \boxed{0.683}\end{aligned}$$

If you have taken introductory Statistics, you may know that this particular problem can also be solved using the Empirical Rule. By the Empirical Rule, we know that 68% of the data lies within one standard deviation of the mean, 95% lies within two standard deviations and 99.7% lies within three standard deviations.



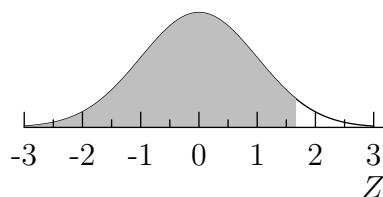
(c)  $P(X < 8)$

$$\begin{aligned}P(X < 8) &= \Phi\left(\frac{8 - 10}{6}\right) \\&= \Phi\left(-\frac{1}{3}\right) \\&= \boxed{0.369}\end{aligned}$$



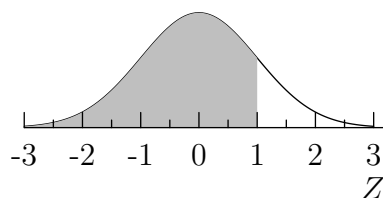
(d)  $P(X < 20)$

$$\begin{aligned}P(X < 20) &= \Phi\left(\frac{20 - 10}{6}\right) \\&= \Phi\left(\frac{5}{3}\right) \\&\approx \boxed{0.952}\end{aligned}$$



(e)  $P(X > 16)$

$$\begin{aligned}P(X > 16) &= 1 - \Phi\left(\frac{16 - 10}{6}\right) \\&= 1 - \Phi(1) \\&\approx \boxed{0.159}\end{aligned}$$



20. If 65 percent of the population of a large community is in favor of a proposed rise in school taxes, approximate the probability that a random sample of 100 people will contain

(Um, when was this book published??)

A random sample of 100 people is drawn from a population. Random sampling implies that the  $n = 100$  people drawn are independent. Each person (trial) is a Bernoulli random variable: either the person approves of the proposal or does not. The probability of success (given in the problem) is constant,  $p = 0.65$ . We use the **binomial** model. Let  $X$  be the number of people that approve of the proposal. Then  $X \sim \text{Bin}(65, 4.77)$ .

However, consider how difficult it would be to use the binomial. We would have to compute

$$P(X \leq c) = \sum_{i=0}^c \binom{100}{i} p^i (1-p)^{100-i}$$

The factorial function on a standard calculator would throw an exception, and by hand the computation would take forever.

Instead, we use the **normal approximation to the binomial** since  $n$  is large, and since  $np > 10$  and since  $n(1-p) > 10$ . Then, let  $Y \sim N(65, 4.77)$ .

There is one small complication. We are using a continuous distribution to model a *discrete* distribution so we must correct for this discrepancy using the *continuity correction* which is bolded in the equation below.

- (a) *at least 50 who are in favor of the proposition.*

We want to compute the probability that at least 50 people are in support of the proposal.

$$\begin{aligned} P(Y \geq 50) &= P\left(Z \geq \frac{50 - \mathbf{0.5} - 65}{4.77}\right) \\ &= P(Z \geq -3.25) \\ &= 1 - P(Z \leq -3.25) \\ &= 1 - \Phi(-3.25) \\ &= 1 - 0.0006 \\ &\approx \boxed{0.9994} \end{aligned}$$

- (b) *between 60 and 70 inclusive who are in favor.*

$$\begin{aligned} P(60 \leq Y \leq 70) &= P(Y \leq 70) - P(Y \leq 60) \\ &= P\left(Z \leq \frac{70 + \mathbf{0.5} - 65}{4.77}\right) - P\left(Z \leq \frac{60 - \mathbf{0.5} - 65}{4.77}\right) \\ &= P(Z \leq 1.15) - P(Z \leq -1.15) \\ &= \Phi(1.15) - \Phi(-1.15) \\ &\approx \boxed{0.75} \end{aligned}$$

(c) fewer than 75 in favor.

$$\begin{aligned}
 P(Y < 75) &= 1 - P(Y \geq 75) \\
 &= 1 - P\left(Z \geq \frac{75 - 0.5 - 65}{4.77}\right) \\
 &= 1 - P(Z \geq 2.08) \\
 &= 1 - [1 - P(Z < 2.08)] \\
 &= 1 - [1 - \Phi(2.08)] \\
 &= \Phi(2.08) \\
 &= \boxed{0.9812}
 \end{aligned}$$

23. One thousand independent rolls of a fair die will be made. COmpute an approximation to the probability that number 6 will appear between 150 and 200 times inclusively. If number 6 appears exactly 200 times, find the probability that number 5 will appear less than 150 times.

(a) First, let's compute an approximation to the probability that number 6 will appear between 150 and 200 times. The die is rolled 1,000 times. These are independent trials which are Bernoulli trials since a die shows either a 6 or not. Thus,  $n = 1000$ . The probability of success is constant,  $p = \frac{1}{6}$ . We can use the **binomial** model. Let  $X$  be the number of times the die shows a 6.

Again, consider how difficult it would be to use the binomial. We would have to compute the following:

$$\begin{aligned}
 P(150 \leq X \leq 200) &= P(X \leq 200) - P(X \leq 150) \\
 &= \sum_{i=0}^{200} \binom{200}{i} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{200-i} - \sum_{i=0}^{150} \binom{150}{i} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{150-i}
 \end{aligned}$$

No way. Instead, we can use the **normal approximation to the binomial model** because  $np = 1000 \cdot \frac{1}{6} > 10$  and  $n(1-p) = 1000 \cdot \frac{5}{6} > 10$ .

$X \sim \text{Bin}(166.7, 11.79)$ . But, using the normal approximation, let  $Y$  be normally distributed using the parameters from the binomial model. That is, in the normal model,  $\mu = np$  and  $\sigma = \sqrt{np(1-p)}$ . So, let  $Y \sim N(166.7, 11.79)$ .

There is one small complication. We are using a continuous distribution to model a *discrete* distribution. We must correct for this discrepancy using the *continuity correction* which is bolded in the equation below.

$$\begin{aligned}
 P(150 \leq Y \leq 200) &= P(Y \leq 200) - P(Y \leq 150) \\
 &= P\left(Z \leq \frac{200 + \mathbf{0.5} - 166.7}{11.79}\right) - P\left(Z \leq \frac{150 - \mathbf{0.5} - 166.7}{11.79}\right) \\
 &= \Phi(2.87) - \Phi(-1.46) \\
 &= 0.998 - 0.072 \\
 &= \boxed{0.9258}
 \end{aligned}$$

- (b) Now we consider the probability that we obtain less than 150 6s if exactly 200 5s have been shown.

This is a conditional probability problem. Let  $Y$  be the number of 6s that appear and let  $X$  be the number of 5s that appear. Then we need to compute

$P(Y < 150 | X = 200)$ . That is, we consider that 200 of the rolls show a 5, therefore, we have 800 trials remaining that can display anything except a 5, so  $n = 800$  now. Since we have already considered those 200 trials where a 5 appears, the probability of success (a 6 appears) is  $p = \frac{1}{5}$ . Again, we use the normal approximation due to the large value for  $n$  and since the success/failure conditions hold ( $np > 10$  and  $n(1 - p) > 10$ ). We know that  $\mu = np = 800 \cdot \frac{1}{5} = 160$  and  $\sigma = \sqrt{800 \cdot \frac{1}{5} \cdot \frac{4}{5}} = 11.31$ .

Let  $Y^* \sim N(160, 11.31)$ . Then,

$$\begin{aligned} P(Y^* < 150) &= P\left(Z < \frac{150 - 0.5 - 160}{11.31}\right) \\ &= \Phi(-0.93) \\ &= \boxed{0.1762} \end{aligned}$$

27. In 10,000 independent tosses of a coin, the coin landed heads 5800 times. Is it reasonable to assume that the coin is not fair? Explain.

We are given that we have  $n = 10000$  independent trials. Each trial is a Bernoulli random variable since the coin lands heads or it does not. Additionally, the probability of success is constant,  $p = \frac{1}{2}$  because we must assume the coin is fair initially. Thus, we can use the **binomial** model, but  $n$  is so large that this is not practical. Since  $np = 10000 \cdot \frac{1}{2} >> 10$  and  $n(1-p) = 10000 \cdot \frac{1}{2} >> 10$  we can use the **normal approximation to the binomial**. Let  $X$  be the number of heads that appear.  $X$  is binomial with mean  $np = 5000$  and standard deviation  $\sigma = \sqrt{np(1-p)} = 50$ . For the normal approximation, let  $Y$  be normally distributed with the same parameters, that is  $Y \sim N(5000, 50)$ .

We want to know how unusual it is to see 5,800 heads when a fair coin is flipped 10,000 times. That is, we want to compute the probability of observing a value of  $Y$  at least as extreme as 5,800 assuming the coin is indeed fair. We want to compute the *p-value*.

$$\begin{aligned} P(X \geq 5800) &= P\left(Z \geq \frac{5799.5 - 5000}{50}\right) \\ &= 1 - P\left(Z \leq \frac{5799.5 - 5000}{50}\right) \\ &= 1 - \Phi(15.99) \\ &\approx \boxed{0} \end{aligned}$$

The probability of observing a number of heads at least 5,800 is practically 0. So, yes, it is very unusual. We are performing a hypothesis test on the proportion of heads that appear. The test statistic is the observed proportion of flips that are heads,  $\hat{p} = 0.58$ . We assume that the coin is fair ( $p = 0.5$ ). We want to see whether or not  $\hat{p}$  is unusual. We are testing the following hypotheses:

$$\begin{cases} H_0 : p = 0.5 \\ H_a : p > 0.5 \end{cases}$$

We reject  $H_0$  if the p-value computed for the test is less than our significance level,  $\alpha$ . Usually the significance level is  $\alpha = 0.05$ , but here we would reject under any  $\alpha > 0$ . We reject the null hypothesis. If the coin is in fact fair ( $H_0$ ), there is about a 0% chance that we would obtain at least 5,800 heads. Most likely, this coin is biased ( $p > 0.5$ ).



## Chapter 5 Theoretical Exercises

9. If  $Z$  is a standard normal random variable, show that for  $x > 0$ ,

The key to solving this problem is to realize that  $-Z$  is also a standard normal random variable. That is,  $-Z \sim N(0, 1)$ . Here is why, if  $Z \sim N(0, 1)$  then,

$$E(-Z) = -E(Z) = -0 = 0 = E(Z)$$

$$\sigma_{-Z} = \sqrt{\text{Var}(-Z)} = \sqrt{(-1)^2 \text{Var}(Z)} = \sqrt{\text{Var}(Z)} = \sigma_Z$$

(a)  $P(Z > x) = P(Z < -x)$

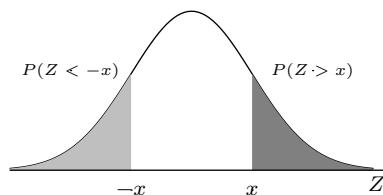
*Proof.* Start with  $P(Z > x)$ . By symmetry, we can negate the inequality. If we negate  $Z$  we also must negate  $x$  and switch the sign in the inequality. So by symmetry,

$$P(Z > x) = P(-Z < -x)$$

but  $-Z \sim N(0, 1)$  and so it is equivalent to  $Z$ . Then,

$$P(-Z < -x) = P(Z < -x)$$

□

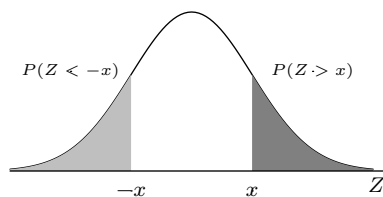


(b)  $P(|Z| > x) = 2P(Z > x)$

*Proof.*

$$\begin{aligned} P(|Z| > x) &= P(Z > x) + P(Z < -x) \\ &\quad \text{by definition of absolute value.} \\ &= P(Z > x) + P(Z > x) \\ &\quad \text{by part (a).} \\ &= 2P(Z > x) \end{aligned}$$

□

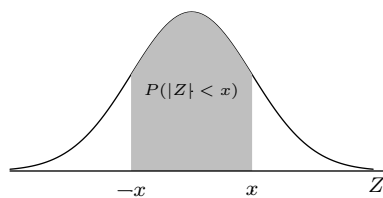


$$(c) \ P(|Z| < x) = 2P(Z < x) - 1$$

*Proof.*

$$\begin{aligned}
 P(|Z| < x) &= 1 - P(|Z| > x) \\
 &= 1 - 2P(Z > x) \\
 &\quad \text{by part (b).} \\
 &= 1 - 2[1 - P(Z < x)] \\
 &= 1 - 2 + 2P(Z < x) \\
 &= -1 + 2P(Z < x) \\
 &= 2P(Z < x) - 1
 \end{aligned}$$

□



10. Let  $f(x)$  denote the probability density function of a normal random variable with mean  $\mu$  and variance  $\sigma^2$ . Show that  $\mu - \sigma$  and  $\mu + \sigma$  are points of inflection of this function. That is, show that  $f''(x) = 0$  when  $x = \mu - \sigma$  or  $x = \mu + \sigma$ .

We must make a substitution to simplify the computation of the derivatives. The PDF for the normal distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Recall that  $z = \frac{x-\mu}{\sigma}$ , so let's replace it with  $z$ . Once we write the function in terms of  $z$ , we then must also remember to use the chain rule since  $z$  is a function of  $x$ ! To further simplify the calculation, we will ignore the constant in front of the exponential. This is typical in Statistics.

$$f(z) \propto C e^{-\frac{z^2}{2}}$$

and  $\frac{dz}{dx} = \frac{1}{\sigma}$ . Then,

$$f'(z) = C e^{-\frac{z^2}{2}} (-z) \frac{1}{\sigma} \propto -C z e^{-\frac{z^2}{2}}$$

$$f''(z) = -z C e^{-\frac{z^2}{2}} (-z) \frac{1}{\sigma} - C e^{-\frac{z^2}{2}} (-z) \propto C e^{-\frac{z^2}{2}} (z^2 - 1)$$

And we must solve  $f''(z) \propto C e^{-\frac{z^2}{2}} (z^2 - 1) = 0$ . The exponential term only goes to 0 if  $z \rightarrow \infty$ , which only happens if  $\sigma = 0$  which is not possible, or if  $x = \infty$  which is irrelevant to this problem since we are asked about two specific values for  $x$ . Thus, we solve  $(z+1)(z-1) = 0$  which yields the result  $z = \pm 1$ .

Now recall that  $z = \frac{x-\mu}{\sigma}$ .

$$\begin{aligned} \pm 1 &= Z \\ &= \frac{x - \mu}{\sigma} \\ \pm \sigma &= x - \mu \\ x &= \boxed{\mu \pm \sigma} \end{aligned}$$

12. The median of a continuous random variable having distribution function  $G$  is that value  $m$  such that  $F(m) = \frac{1}{2}$ . That is, a random variable is just as likely to be larger than its median as it is to be smaller. Find the median of  $X$  if  $X$  is

This problem may look confusing at first. It isn't. The median is the value at which 50% of the data lies beneath it, and 50% lies above it. In other words, we need to find the 50th percentile, that is,  $P(X \leq x) = 0.5$ .

(a) *uniformly distributed over  $(a, b)$ .*

$$\begin{aligned} P(X \leq x) &= \int_a^x \frac{1}{b-a} dx \\ &= \left[ \frac{x}{b-a} \right]_a^x \\ &= \frac{x}{b-a} - \frac{a}{b-a} \\ &= \frac{x-a}{b-a} \end{aligned}$$

Now we set  $\frac{x-a}{b-a} = 0.5$  and solve for  $x$ .

$$\begin{aligned} \frac{1}{2} &= \frac{x-a}{b-a} \\ \frac{1}{2}(b-a) &= x-a \\ x &= \frac{1}{2}(b-a) + a \\ &= \boxed{\frac{a+b}{2}} \end{aligned}$$

(b) *normal with parameters  $\mu, \sigma^2$ .*

Again, we solve  $P(X \leq x) = 0.5$  for  $x$ . **The normal distribution is a pain to work with, but  $z$  scores are not. That is the trick to solving this part of the problem.**

$$\begin{aligned} P(X \leq x) &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= P(Z \leq z) \\ &= \Phi(z) \end{aligned}$$

Then we solve  $\Phi(z) = 0.5$ , which is simply the peak of the normal distribution at  $z = 0$  which is true at  $\boxed{x = \mu}$ .

The normal distribution and the uniform distribution are two examples of symmetric distributions. A common trait of symmetric distributions is that the mean is the same as the median.