# Chapter 4

# Continuous Random Variables

- Learning objectives:
  - understand the definition of a continuous random variable
  - be familiar with the definition of the CDF and pdf for a continuous random variable, as well as their relationship to one another
  - be able to compute probabilities, expected value, and variance for continuous random variables and basic transformations of continuous random variables
  - be familiar with common types of continuous random variables and their distributions, including the uniform, exponential, standard normal, and normal distributions
  - be able to compute probabilities, expected value, and variance for the uniform, exponential, standard normal, and normal distributions
  - be able to approximate a binomial probability using an appropriate normal distribution

In the last chapter, we studied discrete random variables. Discrete random variables can only take on, or produce, a finite or countable number of possible values. That is, the support is a finite or countable set.

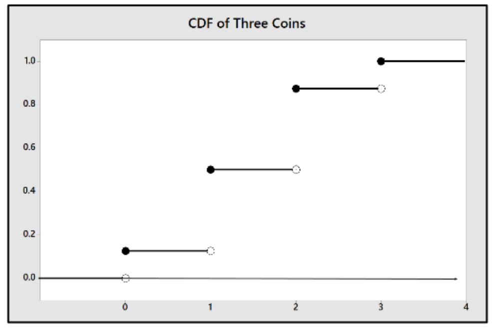
In this chapter, we study a totally different class of random variables called **continuous random variables**. Like discrete random variables, continuous random variables provide a mechanism for mapping a sample space into the real numbers. As was the case with discrete random variables, we are very interested in the distribution of a continuous random variable. The distribution of a continuous random variable is provided by its CDF. We will see that the CDF and support are what distinguishes a continuous random variable from a discrete random variable.

Suppose that the support of a random variable is an interval or a union of intervals. Such a random variable would not be discrete since the support is uncountable. Common examples of random variables for which this is the case include time, height, weight, or any measurement that can take on any value in some interval.

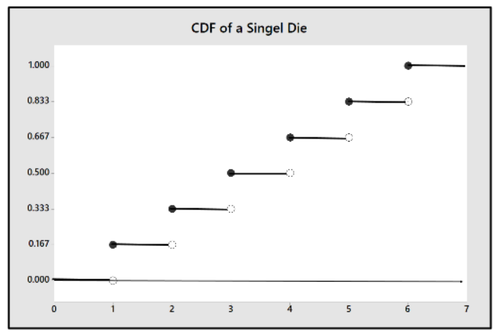
## 4.1 The CDF of a Continuous Random Variable

Recall that for a discrete random variable $X$, we defined the pmf (probability mass function) as $f(x) = P(X = x)$ and the CDF (cumulative distribution function) as $F(x) = P(X \leq x)$. In this case, the CDF was a non-decreasing step function, as illustrated in the examples below. These examples consider two of the experiments studied in the previous chapter: (i) three coins are tossed and the number of heads is counted and (ii) the result of a single die roll is observed.

| $x$ | $F(x)$ |
|---|---|
| 0 | $\frac{1}{8}$ |
| 1 | $\frac{4}{8}$ |
| 2 | $\frac{7}{8}$ |
| 3 | 1 |




| $x$ | $F(x)$ |
|---|---|
| 1 | $\frac{1}{6}$ |
| 2 | $\frac{2}{6}$ |
| 3 | $\frac{3}{6}$ |
| 4 | $\frac{4}{6}$ |
| 5 | $\frac{5}{6}$ |
| 6 | 1 |




Notice that neither of the CDFs above are continuous. Each is a discontinuous (but continuous from the right) step function. In the case of a discrete random variable, it turns out that the CDF can never be continuous. We now introduce a new class of random variables whose CDFs are continuous functions.

**Definition 4.1.1**
Let $X$ be a random variable with CDF $F(x) = P(X \leq x)$. If $F(x)$ is a continuous function, then $X$ is said to be a **continuous random variable**.

Note that the definition of the CDF has not changed. It is still defined as $F(x) = P(X \leq x)$. The CDF still satisfies the same properties discussed previously:

(a) $\lim_{x\to\infty} F(x) = 1$ (the probability goes to 1 in limit, but may never actually equal 1)

(b) $\lim_{x\to-\infty} F(x) = 0$ (as $x$ becomes large negative, the probability to the left goes to 0)

(c) $A \leq B \implies F(A) \leq F(B)$ (this means that the CDF is non-decreasing)

So the only difference between the CDFs we will consider here (for continuous random variables) and the CDFs we have considered in the past (for discrete random variables) is that the CDF is now required to be a continuous function. In such a case, we refer to the associated random variable as a continuous random variable. We now consider several examples to get a feel for this new type of random variable.

**Example 4.1.1**

Suppose we were interested in the random variable $X$ that measures how long an individual can hold their breath. The set of possible outcomes would clearly be uncountable. Therefore, $X$ cannot be a discrete random variable since the support is uncountable. Intuitively, it seems clear that the CDF should have no discontinuities. Since the CDF is continuous, we refer to $X$ as a **continuous random variable**.

Since time, weight, height, and other types of measurements can often take on an uncountably infinite number of possible values, random variables related to these measurements would all be examples of continuous random variables.

**Example 4.1.2**

Below are some examples of functions that could serve as the CDF for some continuous random variable. Each of these functions is a valid CDF since the functions are non-decreasing continuous functions that satisfy the two limit statements provided above.

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ x - 1 & \text{if } 1 \le x \le 2 \\ 1 & \text{if } 2 < x \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } 1 < x \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{if } 0 \le x \end{cases}$$

$$F(x) = \frac{e^x}{1 + e^x} \quad \text{for} \quad -\infty < x < \infty$$

**Example 4.1.3**

A random variable $X$ has the CDF shown below. Determine the value of $a$.

$$F(x) = \begin{cases} 0 & \text{if } x \le a \\ \dfrac{x^2 - 3}{x^2 + 7} & \text{if } a < x \end{cases}$$

**Example 4.1.4**

Suppose a random variable $X$ has CDF

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ x - 1 & \text{if } 1 \leq x \leq 2 \\ 1 & \text{if } 2 < x \end{cases}$$

Determine $P(1.2 < X \leq 1.3)$.

**Example 4.1.5**

Using the same CDF as in the previous example, determine $P(1.2 \leq X \leq 1.3)$.

The above examples demonstrate an important property of continuous distributions that is NOT true in general for discrete distributions:

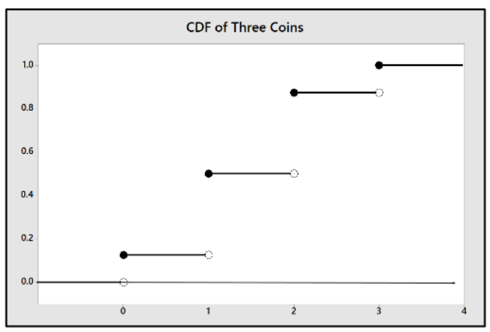
$$P(A \leq X \leq B) = P(A \leq X < B) = P(A < X \leq B) = P(A < X < B)$$

**Example 4.1.6**

Using the same CDF as in the previous example, determine $P(0 \le X \le 1.4)$ and $P(1.5 \le X \le 3)$.

Recall that with discrete random variables, such as $X$ = number of heads in three coin tosses, we were able to use the CDF to determine probabilities via $P(X = x) = F(x) - F(x^-)$ (recall that $x^-$ indicates the value in the support immediately before $x$). For instance, $P(X = 1) = F(1) - F(1^-) = F(1) - F(0) = \frac{4}{8} - \frac{1}{8} = \frac{3}{8}$ and $P(X = 1.5) = F(1.5) - F(1.5^-) = F(1.5) - F(1) = \frac{4}{8} - \frac{4}{8} = 0$.

| $x$ | $F(x)$ |
|---|---|
| 0 | $\frac{1}{8}$ |
| 1 | $\frac{4}{8}$ |
| 2 | $\frac{7}{8}$ |
| 3 | 1 |




For a continuous random variable $X$ with CDF $F(x)$, we could attempt to do the same as above, but the continuity of $F(x)$ means that $P(X = x) = F(x) - F(x^-) = F(x) - \lim_{x^- \to x} F(x^-) = F(x) - F(x) = 0$. In other words, if $X$ is a continuous random variable, then $P(X = x) = 0$ for all $x$!! This may be a surprising feature of continuous random variables that deserves further investigation.

**Example 4.1.7**

Again consider a random variable $X$ with CDF

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ x-1 & \text{if } 1 \le x \le 2 \\ 1 & \text{if } 2 < x \end{cases}$$

As already noted above, $P(1.2 \le X \le 1.3) = 0.1$. This means that while the probability of $X$ taking on any exact value between 1.2 and 1.3 is zero, we still accumulate some probability over the interval $[1.2, 1.3]$. More generally, for $x$ and $x + \Delta x \in [1, 2]$:

$$\begin{aligned} P(x \le X \le x + \Delta x) &= F(x + \Delta x) - F(x^-) \\ &= F(x + \Delta x) - F(x) \\ &= (x + \Delta x - 1) - (x - 1) \\ &= \Delta x \end{aligned}$$

We see that the probability is "evenly spread" amongst the $x$-values between 1 and 2. If we look at the ratio of the probability of being in an interval to the length of the interval we get:

$$\frac{P(x \le X \le x + \Delta x)}{(x + \Delta x) - x} = \frac{\Delta x}{\Delta x} = 1.$$

Note that this is the same as $F'(x)$ over the interval $[1, 2]$.

**Example 4.1.8**

Repeat the investigation done in the previous example for a random variable $X$ with CDF

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } 1 < x \end{cases}$$

## 4.2 The pdf of a Continuous Random Variable

**Definition 4.2.1**

Let $X$ be a continuous random variable with CDF $F(x)$. The **probability density function (pdf)** of $X$ is defined as $f(x) = F'(x)$ for all $x$ where $F'(x)$ exists.

**Definition 4.2.2**

The **support** of a continuous random variable $X$ is defined to be the set of all $x$ such that $f(x) > 0$.

**Example 4.2.1**

The CDF of a random variable $X$ is given below. Use the CDF to determine the pdf of $X$ and the support of $X$. Then graph the pdf.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } 1 < x \end{cases}$$

**Example 4.2.2**

The CDF of a random variable $X$ is given below. Use the CDF to determine the pdf of $X$ and the support of $X$. Then graph the pdf.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{if } 0 \le x \end{cases}$$

**Example 4.2.3**

The CDF of a random variable $X$ is given below. Use the CDF to determine the pdf of $X$ and the support of $X$. Then graph the pdf.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.5x & \text{if } 0 \leq x \leq 1 \\ 0.5x^2 & \text{if } 1 < x \leq \sqrt{2} \\ 1 & \text{if } \sqrt{2} < x \end{cases}$$

**Theorem 4.2.1**

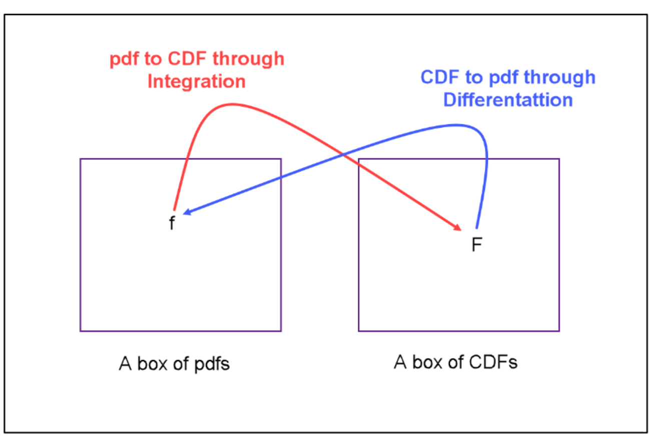
Given a random variable $X$ with CDF $F(x)$ and pmf $f(x)$ (determined from $F(x)$),

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

Note that this implies that

$$P(A \leq X \leq B) = \int_{A}^{B} f(x)\,dx = F(B) - F(A).$$

***Proof:*** Note that this is a direct result of the Fundamental Theorem of Calculus. The connection between the pdf and CDF is illustrated in the figure below. □



**Note 4.2.1**

For a continuous random variable $X$, the CDF $F(x)$ is always a continuous functions. The pdf $f(x)$ may not be continuous. This is why we start with the CDF when defining a continuous random variable, since the continuity of the CDF is what characterizes a continuous random variable.

**Example 4.2.4**

Let $X$ be the phase angle of a voltage signal. The pdf is $f(x) = \dfrac{1}{2\pi}$ for $0 \leq x \leq 2\pi$. Determine $P\left(0 \leq X \leq \frac{\pi}{2}\right)$.

**Example 4.2.5**

Determine the value of $k$ so that $f(x) = \begin{cases} kx & \text{if } 1 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$ is a valid pdf for some random variable $X$.

**Example 4.2.6**

For the random variable $X$ with pdf given in the previous example, determine $F(x)$.

**Example 4.2.7**

Consider a random variable $X$ with pdf

$$f(x) = \begin{cases} \dfrac{2x}{15} & \text{if } 1 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

Determine $F(x)$ and $P(X \leq 2)$. The sketch the graph of the CDF.

## 4.3 The Expected Value of a Continuous Random Variable

As was the case with discrete random variables, we are often interested in the theoretical mean (average) of a continuous random variable. That is, if our random variable were to produce endless amounts of data, what should we expect the mean of the data to be?

**Definition 4.3.1**

The **mean (or expected value)** of a continuous random variable $X$ is defined as

$$\mu_X = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx,$$

provided the integral converges.

**Example 4.3.1**

Consider a random variable $X$ with pdf

$$f(x) = \begin{cases} \dfrac{x^2}{3} & \text{if } -1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Determine the mean of $X$.

**Example 4.3.2**

Consider a random variable $X$ with pdf

$$f(x) = \begin{cases} \dfrac{x}{8} & \text{if } 0 < x < 4 \\ 0 & \text{otherwise} \end{cases}$$

Determine the mean of $X$.

**Example 4.3.3**

Consider a random variable $X$ with pdf $f(x) = \dfrac{1}{\pi(1+x^2)}$ and support $-\infty < x < \infty$. Determine the mean of $X$.

**Note 4.3.1**

The previous example demonstrates that a continuous random variable need not have a finite mean. However, we won't consider such examples in this course.

## 4.4 The Expected Value of a Function of a Continuous Random Variable

**Definition 4.4.1**

The variance of a random variable $X$ is defined by $\mathrm{Var}[X] = \mathrm{E}\left[(X-\mu)^2\right]$.

**Theorem 4.4.1**

Suppose that $X$ is a random variable with pdf $f(x)$, and let $Y = g(X)$ be some function of $X$. The mean (or expected value) of $Y$ can be calculated using

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx.$$

Using the above theorem, we see that the variance of $X$ can be thought of as the expected value of $Y = (X-\mu)^2$, which can be calculated as

$$\mathrm{Var}[X] = \mathrm{E}\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx.$$

This leads to the usual computational formula for variance, $\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$, which we verify below:

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}\left[(X-\mu)^2\right] \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx \\
&= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\, dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)\, dx - 2\mu \int_{-\infty}^{\infty} x f(x)\, dx + \mu^2 \int_{-\infty}^{\infty} f(x)\, dx \\
&= \mathrm{E}\left[X^2\right] - 2\mu(\mu) + \mu^2(1) \\
&= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2
\end{aligned}$$

**Example 4.4.1**

Consider a random variable $X$ with pdf

$$f(x) = \begin{cases} \dfrac{x^2}{3} & \text{if } -1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Determine the variance of $X$.

**Example 4.4.2**

Consider a random variable $X$ with pdf

$$f(x) = \begin{cases} \dfrac{x}{8} & \text{if } 0 < x < 4 \\ 0 & \text{otherwise} \end{cases}$$

Determine the variance of $X$.

## 4.5 The (Continuous) Uniform Distribution

**Definition 4.5.1**

A continuous random variable $X$ with pdf

$$f(x) = \begin{cases} \dfrac{1}{B-A} & \text{if } A < x < B \\ 0 & \text{otherwise} \end{cases}$$

is said to be a **uniformly distributed** random variable. This distribution is known as the **uniform distribution**, and we often use the shortcut notation $X \sim \text{Unif}(A, B)$ to indicate that $X$ is a uniform random variable on the interval $(A, B)$.

**Theorem 4.5.1**

If $X \sim \text{Unif}(A, B)$, then $\mu_X = \text{E}[X] = \dfrac{A+B}{2}$.

***Proof:*** It's clear that the mean should be the midpoint of the interval $(A, B)$. More formally, we can use the definition of expected value as follows:

$$\begin{aligned} \text{E}[X] &= \int_A^B x \cdot \frac{1}{B-A}\, dx \\ &= \frac{x^2}{2(B-A)}\Big|_A^B \\ &= \frac{B^2 - A^2}{2(B-A)} \\ &= \frac{A+B}{2} \end{aligned}$$

□

**Theorem 4.5.2**

If $X \sim \text{Unif}(A, B)$, then $\text{Var}[X] = \dfrac{(B-A)^2}{12}$.

***Proof:*** We start by calculating the second moment, and then use the computational formula for variance along with the result of the previous theorem:

$$\begin{aligned}
\mathrm{E}\left[X^2\right] &= \int_A^B x^2 \cdot \frac{1}{B-A}\,dx \\
&= \frac{x^3}{3(B-A)}\bigg|_A^B \\
&= \frac{B^3-A^3}{3(B-A)} \\
&= \frac{(B-A)(A^2+AB+B^2)}{3(B-A)} \\
&= \frac{A^2+AB+B^2}{3} \\
\mathrm{Var}\left[X\right] &= \mathrm{E}\left[X^2\right] - \left(\mathrm{E}\left[X\right]\right)^2 \\
&= \frac{A^2+AB+B^2}{3} - \left[\frac{A+B}{2}\right]^2 \\
&= \frac{A^2+AB+B^2}{3} - \frac{A^2+2AB+B^2}{4} \\
&= \frac{4A^2+4AB+4B^2}{12} - \frac{3A^2+6AB+3B^2}{12} \\
&= \frac{A^2-2AB+B^2}{12} \\
&= \frac{(B-A)^2}{12}
\end{aligned}$$

□

**Theorem 4.5.3**

If $X \sim \mathrm{Unif}(A, B)$, then the CDF of $X$ is given by

$$F(x) = \begin{cases} 0 & \text{if } x \leq A \\ \dfrac{x-A}{B-A} & \text{if } A < x < B \\ 1 & \text{if } B \leq x \end{cases}$$

***Proof:*** It should be clear that $F(x) = 0$ for $x \leq A$ and $F(x) = 1$ for $x \geq B$. For $A < x < B$, use the definition of CDF to obtain the following:

$$F(x) = P(X \leq x) = \int_A^x \frac{1}{B-A}\,dt = \frac{t}{B-A}\bigg|_A^x = \frac{x-A}{B-A}.$$

□

**Example 4.5.1**

If $X \sim \text{Unif}(-1, 5)$, determine the mean, variance, and CDF of $X$.

**Example 4.5.2**

If $X \sim \text{Unif}(-1, 5)$, determine $P(-0.5 < X < 2.5)$.

## 4.6 The Exponential Distribution

**Definition 4.6.1**

A continuous random variable $X$ with pdf

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ be^{-bx} & \text{if } 0 < x \end{cases}$$

is said to be an **exponentially distributed** random variable. This distribution is known as the **exponential distribution**, and we often use the shortcut notation $X \sim \text{Exp}(b)$ to indicate that $X$ is an exponential random variable with parameter $b$.

**Theorem 4.6.1**

If $X \sim \text{Exp}(b)$, then $\mu_X = \text{E}[X] = \dfrac{1}{b}$.

***Proof:*** Use integration by parts to show that $\mu_X = \text{E}[X] = \displaystyle\int_0^\infty bxe^{-bx}\,dx = \frac{1}{b}$. □

**Theorem 4.6.2**

If $X \sim \text{Exp}(b)$, then $\text{Var}[X] = \dfrac{1}{b^2}$.

***Proof:*** Use integration by parts twice to show that $\text{E}\left[X^2\right] = \displaystyle\int_0^\infty bx^2e^{-bx}\,dx = \frac{2}{b^2}$. Then use the computational formula for variance to obtain $\text{Var}[X] = \text{E}\left[X^2\right] - (\text{E}[X])^2 = \dfrac{2}{b^2} - \left(\dfrac{1}{b}\right)^2 = \dfrac{1}{b^2}$. □

**Theorem 4.6.3**

If $X \sim \text{Exp}(b)$, then the CDF of $X$ is given by

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-bx} & \text{if } 0 < x \end{cases}$$

***Proof:*** It should be clear that $F(x) = 0$ for $x \leq 0$. For $x > 0$, use the definition of CDF to obtain the following:

$$F(x) = P(X \leq x) = \int_0^x be^{-bt}\,dt = -e^{-bt}\Big|_0^x = 1 - e^{-bx}.$$

□

**Example 4.6.1**

Suppose the random variable $X$ is known to follow an exponential distribution with parameter $b = 3$. Determine the mean, variance, standard deviation, pdf, and CDF of $X$.

**Example 4.6.2**

Suppose the random variable $X$ is known to follow an exponential distribution with mean $\mu = 2$. Determine the variance, standard deviation, pdf, and CDF of $X$.

**Example 4.6.3**

Suppose the random variable $X$ is known to follow an exponential distribution with mean $\mu = 2$. Determine $P(3 < X)$ and $P(-2 < X < 2)$.

Exponential random variables are closely related to the Poisson random variables discussed in Chapter 3. Recall that a Poisson random variable counts the number of arrivals in a Poisson process during a certain period. In such a process, an exponential random variable measures the **inter-arrival time**, i.e., the time between two consecutive Poisson events. See pages 37–41 in `https://drive.google.com/file/d/1dAKWJYmlhTzbEE2Oy02oUXqwmlwV4-NV/view` for some applications, including some additional details regarding the example below.

**Example 4.6.4**

Energy efficiency is an important aspect of designing electrical systems. In some modern buildings (e.g., airports), traditional escalators are being replaced by a new type of "smart" escalator which can automatically switch between a normal operating mode and a standby mode depending on the flow of pedestrians.

(a) The arrival of pedestrians can be modeled using a Poisson random variable. Let $\lambda$ be the arrival rate (people per minute), and let $N$ be the number of arrivals in a period of length $t$ minutes. Note that $N$ is a Poisson random variable with rate parameter $\lambda t$ (using the scaling property). Therefore, the pmf of the number of pedestrians in a period of length $t$ minutes is

$$\pi(n) = P(N = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

for $n = 0, 1, 2, 3, \ldots$.

(b) Let $T$ be a random variable denoting the inter-arrival time (i.e., the time between two consecutive arrivals). Show that $P(T > t) = e^{-\lambda t}$, and use this to determine $F_T(t)$ and $f_T(t)$.

For $t \geq 0$, we have the following:

$$\begin{aligned} P(T > t) &= P(\text{inter-arrival time} > t) \\ &= P(\text{no arrivals in a period of length } t) \\ &= P(N = 0) \\ &= \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \\ &= e^{-\lambda t} \end{aligned}$$

Using this, we see that $F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t}$ for $t \geq 0$. Using the fact that $F_T'(t) = f_T(t)$, we see that $f_T(t) = \lambda e^{-\lambda t}$ for $t \geq 0$. In other words, $T$ is an exponential random variable with parameter $\lambda$.

(c) Suppose that the escalator will go into standby mode if there are no pedestrians for $t_0 = 30$ seconds. Let $Y$ be a random variable denoting the amount of time that the escalator is in standby mode. That is, let

$$Y = \begin{cases} 0 & \text{if } T \leq 0.5 \\ T - 0.5 & \text{if } T > 0.5 \end{cases}$$

Determine $\mathrm{E}[Y]$, i.e., the expected amount of time the escalator will be in standby mode.

Note that both $T$ and $Y$ are continuous random variables, and we can view $Y$ as a transformation of $T$, i.e., $Y = g(T)$. Using the usual formula for expected value, we have:

$$\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}[g(T)] \\
&= \int_0^\infty g(t) f_T(t)\, dt \\
&= \int_0^{0.5} 0 \cdot f_T(t)\, dt + \int_{0.5}^\infty (t - 0.5) f_T(t)\, dt \\
&= \int_{0.5}^\infty t f_T(t)\, dt - 0.5 \int_{0.5}^\infty f_T(t)\, dt \\
&= \int_{0.5}^\infty \lambda t e^{-\lambda t}\, dt - 0.5 \int_{0.5}^\infty f_T(t)\, dt \\
&= -te^{-\lambda t}\Big|_{0.5}^\infty + \int_{0.5}^\infty e^{-\lambda t}\, dt - 0.5P(T > 0.5) \\
&= \left(0 + 0.5e^{-0.5\lambda}\right) - \frac{e^{-\lambda t}}{\lambda}\Big|_{0.5}^\infty - 0.5P(T > 0.5) \\
&= 0.5e^{-0.5\lambda} - \left(0 - \frac{e^{-0.5\lambda}}{\lambda}\right) - 0.5\left[1 - F_T(0.5)\right] \\
&= 0.5e^{-0.5\lambda} + \frac{e^{-0.5\lambda}}{\lambda} - 0.5\left[1 - (1 - e^{-0.5\lambda})\right] \\
&= 0.5e^{-0.5\lambda} + \frac{e^{-0.5\lambda}}{\lambda} - 0.5e^{-0.5\lambda} \\
&= \frac{e^{-0.5\lambda}}{\lambda}
\end{aligned}$$

Note that as $\mathrm{E}[T]$ is the expected time between arrivals, $\mathrm{E}[Y]$ would be interpreted as the expected amount of time between arrivals during which the escalator is in standby mode.

## 4.7 The Standard Normal Distribution

**Definition 4.7.1**

A continuous random variable $Z$ with pdf

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

and support $-\infty < z < \infty$ is said to be a **standard normal random variable**. This distribution is called the **standard normal distribution** and is often denoted using the shortcut notation $Z \sim \mathrm{N}(0, 1)$.

We will always use the letter $Z$ to refer to a standard normal random variable. Showing that the function provided above is a valid pdf cannot be done "by hand", as there is no way to integrate this function in terms of elementary functions to show that the definite integral over the support is 1. A computer can evaluate this via numerical methods, however, to show that the function is, indeed, a valid pdf.

**Theorem 4.7.1**

The mean of a standard normal random variable is 0, i.e., if $Z \sim \mathrm{N}(0,1)$, then $\mathrm{E}[Z] = 0$.

***Proof:*** Using the definition of expected value, we have the following:

$$\begin{aligned}
\mathrm{E}[Z] &= \int_{-\infty}^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} z e^{-z^2/2}\, dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-z^2/2}\, dz
\end{aligned}$$

It should be clear that the two integrals are equal in magnitude and opposite in sign. So the expected value of $Z$ will be 0 as long as each integral converges. We show below that the second integral converges by making the substitution $u = \frac{-z^2}{2}$:

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-z^2/2}\, dz &= \frac{-1}{\sqrt{2\pi}} \int_{0}^{-\infty} e^{u}\, du \\
&= \frac{-1}{\sqrt{2\pi}} e^{u} \Big|_{0}^{-\infty} \\
&= \frac{-1}{\sqrt{2\pi}} (0 - 1) \\
&= \frac{1}{\sqrt{2\pi}}
\end{aligned}$$

Thus, using the comments above regarding the relationship between the two integrals, we have the desired result:

$$\begin{aligned}
\mathrm{E}[Z] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} z e^{-z^2/2}\, dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-z^2/2}\, dz \\
&= \frac{-1}{\sqrt{2\pi}} + \frac{1}{\sqrt{2\pi}} \\
&= 0
\end{aligned}$$

$\square$

**Theorem 4.7.2**

The variance of a standard normal random variable is 1, i.e., if $Z \sim \mathrm{N}(0,1)$, then $\mathrm{Var}[Z] = 1$.

***Proof:*** We compute the second moment using the integration by parts formula $\int u\, dv = uv - \int v\, du$ with $u = z$, $du = dz$, $dv = \frac{1}{\sqrt{2\pi}} z e^{-z^2/2}\, dz$, and $v = \frac{-1}{\sqrt{2\pi}} e^{-z^2/2}$ as shown below:

$$\begin{aligned}
\mathrm{E}\left[Z^2\right] &= \int_{-\infty}^{\infty} z^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \frac{-1}{\sqrt{2\pi}} z e^{-z^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= 0 + 1 \\
&= 1.
\end{aligned}$$

Note that the second integral is the integral of the pdf of $Z$ over its support, so must evaluate to 1. Using $\mathrm{E}\left[Z^2\right] = 1$ along with result of the previous theorem (that $\mathrm{E}[Z] = 0$), we have $\mathrm{Var}[Z] = \mathrm{E}\left[Z^2\right] - (\mathrm{E}[Z])^2 = 1 - 0^2 = 1$. $\square$

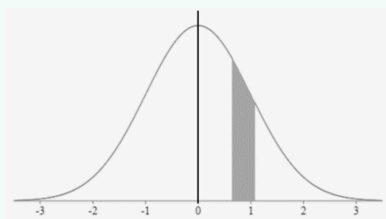
The CDF of $Z$ is defined in the usual manner using the special notation $\Phi(z)$ rather than $F(z)$:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} f(t)\,dt,$$

where $f$ denotes the pdf of $Z$. The antiderivative of $f$ does not exist in closed form, so for any value $z$, $\Phi(z)$ must be computed numerically. Since we obviously cannot do this by hand, we will rely on a CDF chart as demonstrated in the next example.

**Example 4.7.1**

Suppose that we wish to determine $P(0.65 < Z < 1.07)$. In other words, we need to determine the probability that a randomly selected value from the standard normal distribution will be between 0.65 and 1.07. Note that this probability is equal to the area under the graph of the pdf of $Z$ between $z = 0.65$ and $z = 1.07$ as shown in the figure below:




Using CDF notation, this probability/area is equal to $P(0.65 < Z < 1.07) = \Phi(1.07) - \Phi(0.65)$. We can obtain these values using the CDF chart for $Z$ (which we will refer to as the $z$-chart). Note that the chart consists of two pages, one for positive $z$-values and one for negative $z$-values. Start by finding the row that corresponds to the desired ones and tenths places, then find the column that corresponds to the desired hundredths place. The intersection of the row and column contains the desired CDF probability. In other words, the $z$-values are decomposed into ones and tenths (appear at the side of chart) and hundredths (appear at the top of chart), and the "center" of the chart contains cumulative (CDF) probabilities.

Hundredths

Whole Plus Tenths

| | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8079 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |

Returning to the original problem, we have

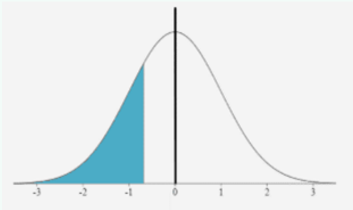
$$P(0.65 < Z < 1.07) = \Phi(1.07) - \Phi(0.65) = 0.8577 - 0.7422 = 0.1155.$$

**Example 4.7.2**

Determine each standard normal probability:

(a) $P(-1.24 < Z < 1.78) = \Phi(1.78) - \Phi(-1.24) = 0.9625 - 0.1075 = 0.8550$


(b) $P(Z < -0.68) = \Phi(-0.68) = 0.2483$




(c) $P(1.88 < Z) = 1 - \Phi(1.88) = 1 - 0.9699 = 0.0301$


**Example 4.7.3**

Determine each standard normal probability:

(a) $P(0.22 < Z < 1.28)$

(b) $P(-0.85 < Z)$

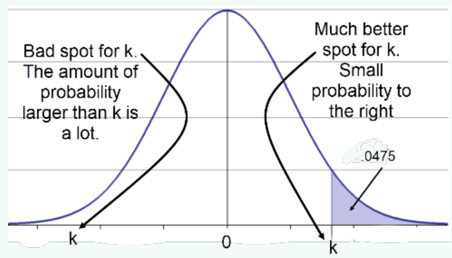
(c) $P(0.56 < Z)$

Note that in each of the problems above, we were given the $z$-value and asked to determine the associated probability. In many cases, we need to do the "opposite" of this, i.e., for a given probability, determine the corresponding $z$-value.

**Example 4.7.4**

Determine the value of $k$ so that $P(k < Z) = 0.0475$.

Since there is probability/area 0.0475 to the right of $k$, there must be probability/area 0.9525 to the left of $k$.




We now look for probability 0.9525 in the CDF chart (search in the "center" part of the chart where the probabilities are located). We find this probability in the chart and see that it corresponds to the $z$-value 1.67, so that $k = 1.67$.

**Note 4.7.1**

If the probability we are looking for does not appear in the CDF chart, we often go with the probability that's closest to the one we're looking for. If the one we're looking for is exactly halfway between two probabilities in the chart, a common approach is to use the average/midpoint of the two corresponding $z$-values.

**Note 4.7.2**

There is a special notation for the $z$-value with probability $\alpha$ to its right: $z_\alpha$. So in the previous example, rather than saying "determine the value of $k$ so that $P(k < Z) = 0.0475$", an equivalent problem is to "determine $z_{0.0475}$". Note that there is no special notation for a left-sided probability question of this type. But any left-sided probability question can be re-written as a right-sided probability question for which this special notation can be used.

**Example 4.7.5**

Determine the value of $k$ so that $P(Z < k) = 0.8907$, i.e., determine $z_{0.1093}$.

## 4.8 Other Normal Distributions

**Definition 4.8.1**

Two random variables $X$ and $Y$ are said to be **identically distributed** if (i) they both have the same support $\mathcal{S}$ and (ii) for all $w \in \mathcal{S}$, $F_X(w) = F_Y(w)$. Often, we simply say that $X$ and $Y$ are identically distributed if and only if they have the same probability distribution (which implies the two aforementioned conditions).

**Definition 4.8.2**

A continuous random variable $X$ with pdf

$$f(x) = \frac{1}{\sqrt{2\pi b^2}} e^{-(x-a)^2/2b^2}$$

and support $-\infty < x < \infty$ (where $a, b \in \mathbb{R}$ are constants) is said to be a **normally distributed random variable**. This distribution is called the **normal distribution** and is often denoted using the shortcut notation $X \sim \mathrm{N}(a, b)$.

**Note 4.8.1**

The standard normal distribution, discussed in the previous section, is the special case of a normal distribution for which $a = 0$ and $b = 1$.

**Theorem 4.8.1**

If $X \sim \mathrm{N}(a, b)$, then $\mathrm{E}[X] = a$ and $\mathrm{Var}[X] = b^2$.

***Proof:*** The proof is the same as the proof for the standard normal distribution given in the previous section. $\square$

Note that since the mean of a normal distribution is $\mathrm{E}[X] = a$, and the variance of a normal distribution is $\mathrm{Var}[X] = b^2$, the notation $\mathrm{N}(a, b)$ is equivalent to $\mathrm{N}(\mu, \sigma)$.

**Theorem 4.8.2**

If $X \sim \mathrm{N}(\mu, \sigma)$, then

$$F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

where $F_Z$ and $\Phi$ refer to the CDF of the standard normal distribution.

***Proof:*** We use the definition of CDF below, along with the substitution $y = \dfrac{t-\mu}{\sigma}$ (note that we do not use the usual letter $u$ for our substitution here so as not to cause any confusion between $u$ and $\mu$). Notice that $dy = \frac{1}{\sigma}\, dt$

so that $\sigma\, dy = dt$. This leads to the following:

$$\begin{aligned} F_X(x) &= P(X \le x) \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-\mu)^2/2\sigma^2}\, dt \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \\ &= F_Z\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

$\square$

**Theorem 4.8.3**

If $X \sim \mathrm{N}(\mu, \sigma)$, then $\dfrac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$.

***Proof:*** We have just shown that $F_X(x) = F_Z\left(\dfrac{x-\mu}{\sigma}\right)$ in the previous theorem, and it should be clear that $F_X(x) = F_{\frac{X-\mu}{\sigma}}\left(\dfrac{x-\mu}{\sigma}\right)$. This gives $F_Z\left(\dfrac{x-\mu}{\sigma}\right) = F_{\frac{X-\mu}{\sigma}}\left(\dfrac{x-\mu}{\sigma}\right)$. Thus, $Z$ and $\dfrac{X-\mu}{\sigma}$ are identically distributed as defined above, so that $\dfrac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$. $\square$

**Example 4.8.1**

Suppose that $X \sim \mathrm{N}(\mu = 10, \sigma = 2)$. Determine $P(8.5 < X < 13.2)$.

$$\begin{aligned} P(8.5 < X < 13.2) &= P\left(\frac{8.5-10}{2} < \frac{X-10}{2} < \frac{13.2-10}{2}\right) \\ &= P(-0.75 < Z < 1.6) \\ &= \Phi(1.6) - \Phi(-0.75) \\ &= 0.9452 - 0.2266 \\ &= 0.7186 \end{aligned}$$

The process used in the previous example in which a question about a normally distributed random variable is transformed to a question about a standard normal random variable is often called **standardizing**.

**Example 4.8.2**

Suppose that $X \sim \mathrm{N}(\mu = 10, \sigma = 2)$. Determine $P(11.68 < X)$.

**Example 4.8.3**

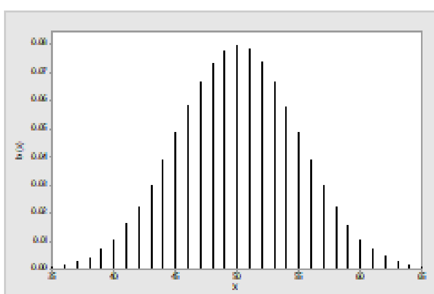
Suppose that $X \sim \mathrm{N}(\mu = 10, \sigma = 2)$. Determine the number $k$ so that $P(k < X) = 0.10$.

The importance of normal random variables can be seen by considering a simple example that involves the sum obtained when $n$ dice are rolled, as described on pages 50–51 in https://drive.google.com/file/d/1dAKWJYmlhTzbEE2Oy02oUXqwmlwV4-NV/view. It turns out that the sum of $n$ independent random variables, regardless of their distributions, converges to a normal distribution as $n \to \infty$. Consequently, it is often reasonable to assume that the mean $\overline{X}$ of a random sample is approximately normally distributed (as the mean involves the sum of the random variables), which is one of the most widely used and well-known results in probability and statistics.

## 4.9 The Normal Approximation to the Binomial Distribution

Suppose $X \sim \mathrm{Bin}(n, p)$. If $n$ is large and $p$ is not too small or too large, then the normal distribution provides a good approximation to the binomial distribution. A general rule of thumb that can be used to decide if the parameters $n$ and $p$ satisfy the size conditions is that both $np \geq 5$ and $n(1-p) \geq 5$ should be satisfied.

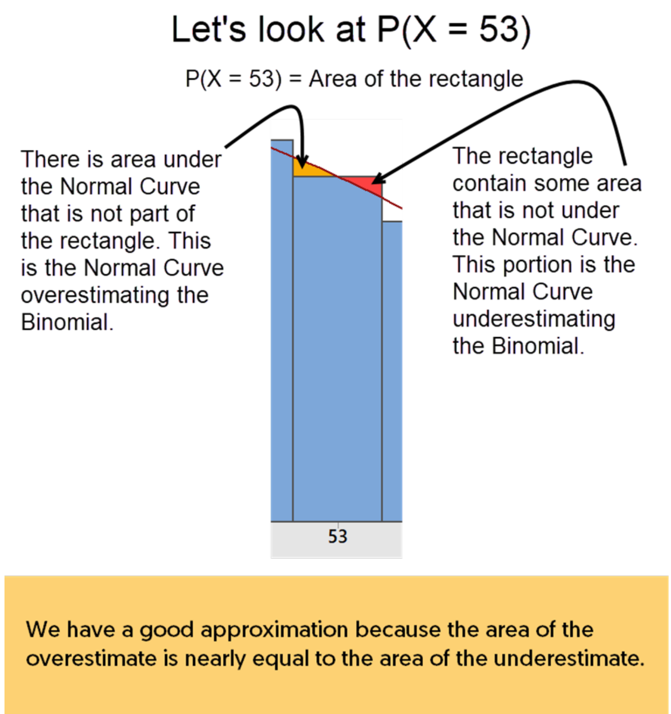
Suppose that $X \sim \mathrm{Bin}(n = 100, p = 0.5)$. We see that the pmf is bell-shaped in the figure below, so using the normal distribution to approximate binomial probabilities seems like a reasonable idea.



What parameters $\mu$ and $\sigma$ should we use for the normal distribution? Since these represent the mean and standard deviation of the normal distribution, a reasonable choice would be to use the mean and standard deviation of the binomial distribution $\mathrm{Bin}(n = 100, p = 0.5)$. In other words, if using a normal distribution to approximate a binomial distribution, it seems appropriate that both distributions should have the same mean and standard deviation. So in this case, we should use $\mu = np = 100(0.5) = 50$ and $\sigma = \sqrt{np(1-p)} = \sqrt{100(0.5)(1-0.5)} = 5$.

How do we actually go about doing the approximation? First, we must realize that the binomial distribution has a discrete support, while the normal distribution has a continuous (interval) support. Suppose that we draw the pmf of the binomial distribution in such a way that there is a rectangle of width 1 centered at each $x$-value in the support whose height is equal to the probability of that $x$-value. Since the width of each rectangle is 1, this is equivalent to saying that the area of each rectangle is equal to the probability of the corresponding $x$-value, i.e., Area of rectangle $= P(X = x) = b(x; n = 100, p = 0.5)$. This means that approximating a binomial probability is equivalent to approximating the area of a rectangle. If the normal distribution gives a good approximation to the binomial distribution, then the area of the desired rectangle should be approximately equal to the area under the normal curve between $x - 0.5$ and $x + 0.5$ (since the rectangle is centered at $x$ and has width 1).

The figure below provides the rationale for estimating the area of a rectangle by using the area under a normal curve, and also demonstrates why this approximation is often quite good:



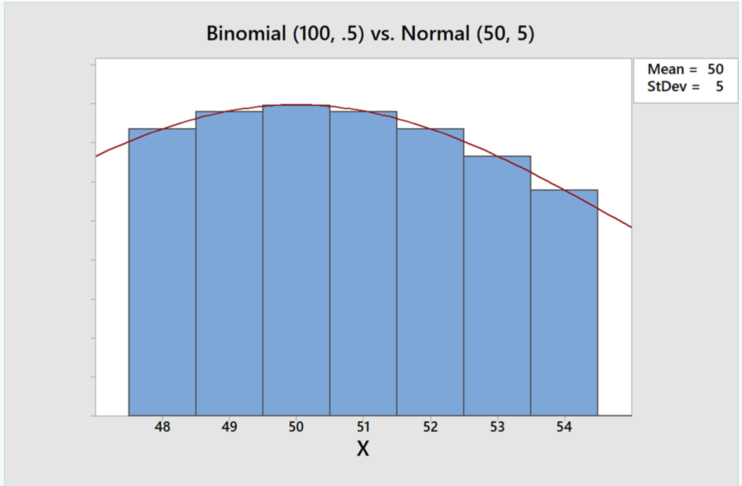

**Example 4.9.1**

Use the normal approximation to the binomial to approximate $P(X = 50)$ in the case where $X \sim \text{Bin}(n = 100, p = 0.5)$. Then compare the approximation obtained from the normal distribution to the actual probability obtained from the binomial distribution.

Using the normal distribution $Y \sim \text{N}(\mu = 50, \sigma = 5)$ as an approximation to $X \sim \text{Bin}(n = 100, p = 0.5)$, we obtain the following:

$$\begin{aligned} P(X = 50) &\approx P(49.5 < Y < 50.5) \\ &= P\left(\frac{49.5 - 50}{5} < \frac{Y - 50}{5} < \frac{50.5 - 50}{5}\right) \\ &= P(-0.1 < Z < 0.1) \\ &= \Phi(0.1) - \Phi(-0.1) \\ &= 0.539828 - 0.460172 \\ &= 0.079656 \end{aligned}$$

This approximation is very close to the actual probability of $P(X = 50) = \binom{100}{50}(0.5)^{50}(1 - 0.5)^{100-50} \approx 0.079589$. The reason the normal approximation does so well in this case is shown in the figure below:




**Example 4.9.2**

Suppose that $X \sim \text{Bin}(100, 0.5)$. Use the normal approximation to the binomial distribution to approximate $P(45 \le X \le 55)$. Then compare the approximation obtained from the normal distribution to the actual probability obtained from the binomial distribution.

Using the normal distribution $Y \sim \text{N}(\mu = 50, \sigma = 5)$ as an approximation to $X \sim \text{Bin}(n = 100, p = 0.5)$, we obtain the following:

$$\begin{aligned}
P(45 \le X \le 55) &\approx P(44.5 < Y < 55.5) \\
&= P\left(\frac{44.5 - 50}{5} < \frac{Y - 50}{5} < \frac{55.5 - 50}{5}\right) \\
&= P(-1.1 < Z < 1.1) \\
&= \Phi(1.1) - \Phi(-1.1) \\
&= 0.864334 - 0.135666 \\
&= 0.728668
\end{aligned}$$

This approximation is very close to the actual probability of $P(45 \le X \le 55) = B(55) - B(44) = 0.864373 - 0.135627 = 0.728747$.

**Example 4.9.3**

Suppose that $X \sim \text{Bin}(400, 0.55)$. If we are going to use the normal approximation to the binomial distribution, what mean and standard deviation would we use? In other words, determine the values of $\mu$ and $\sigma$ such that $\text{N}(\mu, \sigma)$ can be used as an approximation to $X \sim \text{Bin}(400, 0.55)$.

**Example 4.9.4**

Suppose that $X \sim \text{Bin}(400, 0.55)$. Use the normal approximation to the binomial distribution to approximate $P(201 \leq X \leq 300)$. Then compare the approximation obtained from the normal distribution to the actual probability obtained from the binomial distribution.