Chapter 8
Random Variables

R is designed to provide the probabilities of the more common random variables. This chapter will demonstrate how to use the binomial, uniform, and normal distributions. The needed commands will consist of a prefix (d, q, p, or r) and a suffix (binom, unif, norm) to identify the desired output and also the distribution of interest. Examples are the best way to explain the commands, and thus will be used in this chapter to help you understand how to use R with distributions.

Section 8.4 Binomial Random Variables

Commands dealing with binomial random variables use the “binom” suffix and are dbinom, pbinom, qbinom, and rbinom. The prefixes stand for density (d), probability (p), quantile (q) and random (r). Suppose you would like to find the probabilities \( P(X = 0) \), \( P(X = 1), \ldots, P(X = 10) \) for the binomial distribution with \( n = 10 \) and \( p = 0.25 \). These exact probabilities are calculated by the dbinom() function. You need to provide the values for which you would like the exact probabilities (x=0 through 10), the number of trials (size=10), and the probability of success for each trial (prob=0.25).

So to find the probability of 4 successes from 10 trials, each with probability 0.25 of success, the simple command can be used to find out it is 0.146.

\[
> \text{dbinom}(4, \text{size}=10, \text{prob}=0.25)
\]
\[
[1] 0.145998
\]

Often you would like to quickly find the probabilities for many possible values, such as 0 through 10. It is simply done by providing the dbinom() x argument a vector of values.

\[
> \text{dbinom}(0:10, \text{size}=10, \text{prob}=0.25)
\]
\[
[1] 5.631351e-02 1.877117e-01 2.815676e-01 2.502823e-01 1.459980e-01
\]

Note that \( 5.63e-02 \) is equivalent to \( 5.63 \times 10^{-2} = 0.0563 \). If you desire to avoid this scientific notation, use the round() function to round the output out to the desired number of significant figures. The following example will round them to 8 significant figures.

\[
> \text{round(dbinom(0:10, size=10, prob=0.25),8)}
\]
\[
[1] 0.05631351 0.18771172 0.28156757 0.25028229 0.14599800 0.05839920
[7] 0.01622200 0.00308990 0.00038624 0.00002861 0.00000095
\]

To find the cumulative probabilities, use the pbinom() function. This calculates \( P(X \leq k) \). In the previous example, suppose we wanted to know \( P(X \leq 4) \). Then we would type the command pbinom(4, size=10, prob=0.25) as shown below giving us the probability of 0.92.

\[
> \text{pbinom}(4, \text{size}=10, \text{prob}=0.25)
\]
\[
[1] 0.9218731
\]
Since $P(X \leq 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$, we could have also used the following approach of summing the individual probabilities of 0 through 4 to get the same answer.

```r
> sum( dbinom( 0:4, size=10, prob=0.25 ) )
[1] 0.9218731
```

We will now use Example 8.16 for more practice. Using 10 births ($n = 10$), and probability of a girl being 0.488 ($p = 0.488$) calculate the probability of having exactly 7 girls, $P(X = 7)$, the probability of at most 7 girls, $P(X \leq 7)$, and the probability of at least 7 girls, $P(X \geq 7) = 1 - P(X \leq 6)$. The first question is answered by the `dbinom()` function and the second by the `pbinom()` function. For the first two questions, we only need to specify to R that there are 10 girls (size=10) and the probability of each child being a girl is 0.488 (prob=0.488). The respective answers to $P(X = 7)$ and $P(X \leq 7)$ are probabilities 0.106 and 0.953.

```r
> dbinom( 7, size=10, prob=0.488)
[1] 0.1061524
> pbinom( 7, size=10, prob=0.488)
[1] 0.9532567
```

Calculating $P(X \geq 7) = 1 - P(X \leq 6)$ requires a simple extra step of subtraction giving the probability 0.153 of getting at least 7 girls.

```r
> 1-pbinom( 6, size=10, prob=0.488 )
[1] 0.1528957
```

Section 8.5 Continuous Random Variables.

With continuous random variables we are sometimes interested in the height of the density curve at a certain value, but most often with the cumulative probability for a certain value.

Examples 8.19 and 8.20 deal with the uniform distribution. The uniform distribution is the simplest of the continuous distributions. In R, you need to specify the lower and upper bounds of the distribution where the probability density function is not zero. Because the waiting time for the bus is 10 minutes, the density curve is 0.1 for values between 0 and 10. Outside of the 0 through 10 range, the height of the density curve is 0. (Note: The area beneath the density curve is 0.1* (10 - 0) = 1 as should be expected.) To calculate the probability of a rider waiting between 5 and 7 minutes would be $P(5 \leq X \leq 7) = P(X \leq 7) - P(X \leq 5) = 0.7 - 0.5 = 0.2$. (Note: Because the uniform distribution is continuous, $P(X \leq 5) = P(X < 5)$.) This is generally not true with discrete random variables.) The `punif()` function can quickly solve these probabilities. Type the following to see that the probability is 0.2. The first two command lines are simply for demonstration and not necessary.

```r
> punif(7, min=0, max=10)
[1] 0.7
> punif(5, min=0, max=10)
[1] 0.5
> punif(7, min=0, max=10) - punif( 5, min=0, max=10)
[1] 0.2
```
Section 8.6 Normal Random Variables

R uses the functions `dnorm()`, `pnorm()`, `qnorm()`, and `rnorm()` to work with normal random variables. As with binomial and uniform distributions, the d, p, q, and r prefixes refer to density, probability, quantile, and random. We will focus on `pnorm()` which calculates $P(X \leq a)$ for a normal random variable $X$ with mean $\mu$ and standard deviation $\sigma$. You need to provide `pnorm()` the value of $a$ and the mean and standard deviation of $X$. Thus if $X$ were a normal random variable with mean 10 and standard deviation 2, $P(X \leq 6)$ would be calculated by `pnorm(6, mean=10, sd=2)` which is 0.023.

```r
> pnorm(6, mean=10, sd=2)
[1] 0.02275013
```

If the mean and sd are not specified within the `pnorm()` command, the default is a mean of 0 and standard deviation of 1. A normal random variable with mean 0 and standard deviation 1 is known as a standard normal variable. Let the variable $Z$ represent a standard normal random variable. Then, for example, $P(Z < 1.31)$ is calculated in R via either of the following `pnorm()` commands giving the probability 0.905.

```r
> pnorm(1.31, mean=0, sd=1)
[1] 0.904902
> pnorm(1.31) # taking advantage of pnorm defaults
[1] 0.904902
```

In Example 8.25 we find the complement of the previous problem, $P(Z > 1.31)$. This is done by simply subtracting 1. That is, $P(Z > 1.31) = 1 - P(Z \leq 1.31)$. This can be done as one step in R as shown in the following command.

```r
> 1-pnorm(1.31)
[1] 0.09509792
```

Use Example 8.25 for more practice with the normal distribution. Assuming the distribution of scores on the math section of the SAT test is normal with mean 515 and standard deviation 100. What is the probability that a randomly selected test-taker had a score less than or equal to 600? That is, calculate $P(X < 600)$ where $\mu=515$ and $\sigma=100$. The following R command quickly finds the answer of 0.802.

```r
> pnorm(600, mean=515, sd=100)
[1] 0.8023375
```

What is the probability that a randomly selected test-taker scored higher than 600? That is, calculate $P(X > 600) = 1 - P(X < 600)$ where $\mu=515$ and $\sigma=100$. The following R command quickly finds the answer of 0.198.

```r
> 1-pnorm(600, mean=515, sd=100)
[1] 0.1976625
```

What is the probability that a randomly selected test-taker scored between 515 and 600. That is, calculate $P(515 < X < 600) = P(X < 600) - P(X \leq 515)$ where $\mu=515$ and $\sigma=100$. The following R command quickly finds the answer of 0.302.

```r
> pnorm(600, mean=515, sd=100)-pnorm(515, mean=515, sd=100)
[1] 0.3023375
```
What is the probability that a randomly selected test-taker’s score was more than 85 points from the mean in either direction? That is, calculate \( P(X < 430) + P(X > 600) = 2 \times P(X < 430) \) where \( \mu = 515 \) and \( \sigma = 100 \). The following R command quickly finds the answer of 0.395.

\[
> 2 \times \text{pnorm}(430, \text{mean}=515, \text{sd}=100) \\
[1] 0.3953251
\]

Example 8.26 deals with finding percentiles also known as quantiles. The example asks to calculate the 75th percentile of systolic blood pressures if we were to assume the distribution of blood pressures is normal with mean 120 and standard deviation 10. This is where the \text{qnorm()} \) command becomes useful. The \text{qnorm()} \ function is essentially the inverse of the \text{pnorm()} \ command. For a given percentile \( p \) it will find \( x \) such that \( P(X \leq x) = p \). In Example 8.26, \( p = 0.75 \) and \( X \) is a normal random variable with \( \mu = 120 \) and \( \sigma = 10 \). The appropriate R command would be \text{qnorm(0.75, mean=120, sd=10)} to find the answer that 75 percent of the men aged 18 to 29 years have systolic blood pressures less than 127.

\[
> \text{qnorm(0.75, mean=120, sd=10)} \\
[1] 126.7449
\]