Exercise Set 4-1

1) The properties of complements say that $A \cup A^C = \Omega$ and that $A \cap A^C = \emptyset$. The second axiom of probability tells us that $\mathrm{P}(\Omega) = 1$, so we know that $\mathrm{P}(A \cup A^C) = 1$. In more familiar terms, A will either happen, or it will not. The fact that $A \cap A^C = \emptyset$ tells us that A and A^C are mutually exclusive—they share no events in common, so they cannot both happen. Because $A \cap A^C = \emptyset$, we can invoke the third axiom to write $\mathrm{P}(A \cup A^C) = 1 = \mathrm{P}(A) + \mathrm{P}(A^C)$. Thus, $\mathrm{P}(A^C) = 1 - \mathrm{P}(A)$. This is a useful and intuitive fact that we have arrived at through a rather abstract route: the probability of an event occuring is one minus the probability that the event does not happen. The interesting thing is that we have shown that this intuitive result is a logical consequences of the axioms.

2) We have to define the probability function for the roulette problem. We start with two facts, $\mathrm{P}(B) = \beta$ and $\mathrm{P}(R) = \rho$, plus the three axioms of probability. We need to find the probability of every event in $\mathcal{F}$:

$$\mathcal{F} = \big\{\emptyset, \{B\}, \{R\}, \{G\}, \{B, R\}, \{B, G\}, \{R, G\}, \{B, R, G\}\big\}.$$

Let's start with what we can get just from the axioms: Axiom (ii) tells us that $\mathrm{P}(\{B, R, G\}) = 1$ because $\{B, R, G\} = \Omega$. Note that the empty set is the complement of $\{B, R, G\}$, so the result from exercise 1 above tells us that $\mathrm{P}(\emptyset) = 1 - \mathrm{P}(\Omega) = 1 - 1 = 0$.

Now we can use the facts that $\mathrm{P}(B) = \beta$ and $\mathrm{P}(R) = \rho$. First notice that $\{B\} \cap \{R\} = \emptyset$, so we can use axiom (iii) to learn that $\mathrm{P}(\{B, R\}) = \beta + \rho$. Exercise 1 then gives us that $\mathrm{P}(G) = 1 - \beta - \rho$. Similarly, exercise 1 tells us that $\mathrm{P}(\{B, G\}) = 1 - \rho$ and $\mathrm{P}(\{R, G\}) = 1 - \beta$.

3) If $\mathrm{P}(E_1 \cap E_2) = 0$, then the claim to be proven, $\mathrm{P}(E_1 \cup E_2) = \mathrm{P}(E_1) + \mathrm{P}(E_2) - \mathrm{P}(E_1 \cap E_2)$, is equivalent to the third axiom of probability. We need to find a solution that works when E_1 and E_2 share outcomes in common – when it is possible for both events to occur. Define the set difference $E_2 \backslash E_1$ as the set of elements that are members of E_2 but not members of E_1.

Notice that $E_1 \cup (E_2 \backslash E_1) = E_1 \cup E_2$. This implies that $\mathrm{P}(E_1 \cup E_2) = \mathrm{P}(E_1 \cup (E_2 \backslash E_1))$. Also notice that we have removed the elements of E_2 that are present in E_1, so $E_1 \cap (E_2 \backslash E_1) = \emptyset$. We can invoke the third axiom to see that $\mathrm{P}(E_1 \cup (E_2 \backslash E_1)) = \mathrm{P}(E_1) + \mathrm{P}(E_2 \backslash E_1)$.

Putting these two statements together gives us $\mathrm{P}(E_1 \cup E_2) = \mathrm{P}(E_1) + \mathrm{P}(E_2 \backslash E_1)$. Thus, to show that $\mathrm{P}(E_1 \cup E_2) = \mathrm{P}(E_1) + \mathrm{P}(E_2) - \mathrm{P}(E_1 \cap E_2)$, we only have to show that $\mathrm{P}(E_2 \backslash E_1) = \mathrm{P}(E_2) - \mathrm{P}(E_1 \cap E_2)$.

We can apply the third axiom again. Notice that $(E_2 \backslash E_1) \cup (E_1 \cap E_2) = E_2$ and $(E_2 \backslash E_1) \cap (E_1 \cap E_2) = \emptyset$. Therefore, by the third axiom, $\mathrm{P}(E_2) = \mathrm{P}(E_2 \backslash E_1) + \mathrm{P}(E_1 \cap E_2)$. Solving for $\mathrm{P}(E_2 \backslash E_1)$ gives that $\mathrm{P}(E_2 \backslash E_1) = \mathrm{P}(E_2) - \mathrm{P}(E_1 \cap E_2)$. This implies that $\mathrm{P}(E_1 \cup E_2) = \mathrm{P}(E_1) + \mathrm{P}(E_2) - \mathrm{P}(E_1 \cap E_2)$: the probability that one of two events happens is the sum of the probability of each of them minus the probability that they both happen.

Exercise Set 4-2

1) If events A and B are independent, then $\mathrm{P}(A|B) = \mathrm{P}(A)$. Replacing the conditional probability with its definition gives

$$\frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \mathrm{P}(A).$$

Multiplying both sides by $\mathrm{P}(B)$ gives $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$. To prove the second part, divide both sides of $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$ by $\mathrm{P}(A)$. The left side becomes $\mathrm{P}(A \cap B)/\mathrm{P}(A)$, which, by the definition in equation 4.1, is $\mathrm{P}(B|A)$. Thus, $\mathrm{P}(B|A) = \mathrm{P}(B)$.

2) Start by examining the definitions.

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)},$$

and

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)}.$$

If we can get $\mathrm{P}(A \cap B)$ on its own, then to get $\mathrm{P}(B|A)$, we would only need to divide by $\mathrm{P}(A)$. We can do this:

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} = \frac{\frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}\mathrm{P}(B)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A|B)\mathrm{P}(B)}{\mathrm{P}(A)}.$$

The last step, $\mathrm{P}(A \cap B)/\mathrm{P}(B) = \mathrm{P}(A|B)$, follows from equation 4.1. The statement

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(A|B)\mathrm{P}(B)}{\mathrm{P}(A)}$$

is called Bayes' Theorem, and we will have more to say about it in the next section.

Exercise Set 4-3

1) The eight possible sequences and the number of heads each contains are

HHH 3
THH 2
HTH 2
HHT 2
HTT 1
THT 1
TTH 1

TTT 0

Thus, the probability mass function is:

$f_X(0) = f_X(3) = \frac{1}{8}$, $f_X(1) = f_X(2) = \frac{3}{8}$, and $f_X(x) = 0$ for all other x.

2) The sum would be 1. Axiom (ii) of probability (in optional section 4.1) tells us that the probability of the event that includes all possible outcomes is 1.

3) $F_X(b) - F_X(a)$. Remember that $F_X(b) = \mathrm{P}(X \le b)$ and $F_X(a) = \mathrm{P}(X \le a)$. Thus, $F_X(b) - F_X(a) = \mathrm{P}(X \le b) - \mathrm{P}(X \le a)$, or the probability that X is less than or equal to b but not less than or equal to a, $\mathrm{P}(a < X \le b)$.

Exercise Set 4-4

1) There are three parts to consider. For any $x < 0$, $\mathrm{P}(X \le x) = 0$, so the c.d.f. starts with a flat line at height 0 extending from negative infinity to $x = 0$. Similarly, if $x > 1$, then $\mathrm{P}(X \le x) = 1$, so the c.d.f. has another flat line, this one at height 1 and extending from $x = 1$ to positive infinity. The remaining interval is between 0 and 1, the values that X can take. The requirement that all equally-sized intervals in $[0,1]$ are equally likely to contain X results in a line of uniform slope connecting (0,0) and (1,1). In this case, the slope required is 1. You can simply draw the function, but here is code for plotting it in `R`:

```
x <- c(0,1)
Fx <- x
plot(x, Fx, type = "l", xlim = c(-1,2))
lines(c(-1, 0), c(0, 0))
lines(c(1,2), c(1, 1))
```

2) Just as was the case for discrete random variables, $\mathrm{P}(a \le X < b) = F_X(b) - F_X(a)$. This means that if the probability of landing in two intervals of equal sizes differs, then the average slope of F_X in those two intervals must differ. Specifically, the interval with the higher probability must have the higher average slope. Here is `R` code for drawing one possible c.d.f. that meets the description in the problem. Specifically, $\mathrm{P}(0.4 \le X < 0.6) = 0.4$:

```
x1 <- c(0,0.4)
x2 <- c(0.4, 0.6)
x3 <- c(0.6, 1)
Fx1 <- c(0,0.3)
Fx2 <- c(0.3, 0.7)
Fx3 <- c(0.7, 1)
plot(x1, Fx1, type = "l", xlim = c(-1,2), ylim = c(0,1), xlab =
"x", ylab = "Fx")
lines(x2, Fx2)
lines(x3, Fx3)
lines(c(-1, 0), c(0, 0))
```

```
lines(c(1,2), c(1, 1))
```

Exercise Set 4-5

1) $\int_{-\infty}^{\infty} f_X(x)dx = 1$. Recall that for any cumulative distribution function, $\lim_{x\to\infty} F_X(x) = 1$, and recall that $\lim_{x\to\infty} F_X(x) = \int_{-\infty}^{\infty} f_X(x)dx$. Probability density functions always have an area under the curve of exactly 1.

2) Yes, this could be a density for continuous random variable. The area under this function is 1 (axiom ii), and the function is non-negative, which means that the probability of the random variable falling in any interval is non-negative (axiom i). This is different from a probability mass function because mass functions can never take values larger than 1, whereas this density is equal to 10 when $0 \le x \le 1/10$. The difference is that whereas mass functions must *sum* to 1, density functions must *integrate* to 1.

Exercise Set 4-6

1) The probability mass function of the Poisson distribution is $\mathrm{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$. Plugging in the appropriate values for k and λ gives: i) e^{-5}, ii) $5e^{-5}$, iii) $(25/2)e^{-5}$.

2) Use the probability mass function of the geometric distribution with parameter 1/2. If our first "heads" occurs on the 6th flip, then we have five tails before it. We plug $p = 1/2$ and $k = 5$ into $\mathrm{P}(X = k) = (1-p)^k p$ to get $\mathrm{P}(X = 5) = \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right) = 1/64$.

3) a) Use:

```
> x <- seq(-3, 3, length.out = 1000)
> plot(x, dnorm(x, mean = 0, sd = 1), type = "l")
```

b) Use:

```
> x <- seq(-3, 3, length.out = 1000)
> plot(x, pnorm(x, mean = 0, sd = 1), type = "l")
```

c) What value of x is at the 97.5th percentile of the standard normal distribution?

```
> qnorm(0.975, mean = 0, sd = 1)
[1] 1.959964
```

This value, 1.96, is actually a useful number to remember in applied statistics, for reasons that will be discussed in chapter 7.

4) a) Use:

```
> normsims <- rnorm(1000, mean = 0, sd = 1)
> hist(normsims)
```

b) Use:

```
> unifsims <- runif(1000, 0 , 1)
> hist(qnorm(unifsims, mean = 0, sd = 1))
```

We simulated uniform random draws, but we were able to transform them to draws from a normal distribution by feeding them through the normal quantile function, or the inverse of the normal distribution function.

Use this code to see a picture that shows how this works:

```
r <- seq(-3, 3, length.out = 1000)
cdf <- pnorm(r)

#Draw the normal cumulative distribution function.
plot(r, cdf, type = "l", xaxs = "i", yaxs = "i", xlim = c(-3,
3), xlab = expression(italic(x)), ylab =
expression(paste(italic(F[X]), "(", italic(x), ")", sep = "")),
lwd = 2)

#Draw light grey lines representing random samples from the
#standard normal distribution.
x <- rnorm(500)
for(i in x){
  lines(c(i,i), c(min(x), pnorm(i)), col = rgb(190, 190, 190,
     alpha = 60, max = 255))
  lines(c(min(x)-1,i), c(pnorm(i), pnorm(i)), col = rgb(190,
     190, 190, alpha = 60, max = 255))
}
```

The cumulative distribution function of X is drawn in solid black. The light grey lines represent 500 random draws from the distribution of X. Start on the horizontal axis. Each light grey line traces from the value of one of the random samples of X on the x-axis up to the cumulative distribution function. Once it hits the cumulative distribution function, it turns left until it hits the vertical axis. Notice that the positions where the lines hit the x-axis are centered on zero, symmetric, and concentrated near the middle—they look like a normal distribution. Entering `hist(x)` will confirm the suspicion. In contrast, the lines hit the y-axis with roughly uniform density from zero to one. `hist(pnorm(x))` will confirm.

To make the plot, we simulated x-values from the normal distribution and fed them into `pnorm()` to get uniformly distributed data. Effectively, we traced grey lines from the horizontal axis up to the cumulative distribution function and then to the left, ending with a uniform distribution along the vertical axis. But we can also go backwards, starting with uniformly distributed data on the vertical axis, tracing lines to the right until we get to the cumulative distribution function, and then drawing lines straight down to get normally distributed data. This is what happens when we apply `qnorm()` to uniformly distributed data.

This is a powerful idea, not just a curiosity. This approach lets us draw pseudorandom samples from any distribution with a known cumulative distribution function as long as it is possible to generate pseudorandom samples from a continuous uniform distribution.