Geometric Distribution

A discrete random variable is said to have a geometric distribution if

- There are independent and identical trials. Each trial can be thought of as a draw from a population, where the drawing is done with replacement.
- Each trial has two possible outcomes, success and failure.
- The probability of success on each trial is the same, p. Therefore, the probability of failure on each trial is $1 - p = q$.
- The experiment is repeated until the first success occurs.
- The random variable Y is defined as the number of the trial on which the first success occurs.
- The parameter for the geometric random variable Y is the probability of success on each trial p.
- The probability distribution function of the geometric random variable Y is

$$p(y) = p(1-p)^{y-1} \qquad \text{for} \qquad y = 1, 2, \ldots$$

- The theoretical mean of the geometric random variable Y is

$$\mu = E(Y) = \frac{1}{p}$$

- The variance of the geometric random variable Y is

$$\sigma^2 = V(Y) = \frac{1-p}{p^2}$$

Working with geometric random variables in R.

The R built-in function for the geometric distribution defines the random variable Y differently than your textbook. To avoid any confusion, I have written functions in R called **geo** which use the R built-in functions for the geometric distribution to calculate probabilities and generate random values from the geometric random variable Y as defined in your textbook. Before you use R for the geometric distribution, you will need to copy and paste the following functions into the command window.

```
dgeo<- function(y,p){dgeom(y-1,p)}
rgeo<- function(n,p){rgeom(n,p) + 1}
```

To find a probability $P(Y = y) = p(y)$ for a single value y, the command in R is

```
dgeo(y,p)
```

To find the probability $P(Y \leq y)$, use the sum command to add up all $p(y)$ values for y between and including 1 and y.

```
sum(dgeo(1:y,p))
```

To find the probability $P(y_1 \leq Y \leq y_2)$, use the sum command to add up all $p(y)$ values for y between and including y_1 and y_2.

```
sum(dgeo(y1:y2,p))
```

To find the probability $P(Y \geq y) = 1 - P(Y < y) = 1 - P(Y \leq y - 1)$, use the sum command to find $P(Y \leq y - 1)$ and subtract this value from 1.

```
1 - sum(dgeo(1:y-1,p))
```

Problems.

1. An oil prospector will drill a succession of holes in a given area to find a productive well. The probability that he is successful on a given trial is 0.2.

   (a) What is the probability that the third hole drilled is the first that yields a productive well?

   (b) If the prospector can only afford to drill at most ten wells, what is the probability that he fails to find a productive well?

   (c) How many wells would the prospector expect to drill before he finds a productive well?

2. In the game of craps, two different dice are rolled and the sum of the 2 dice is determined.

   (a) What is the probability that the first seven will occur on the 5th roll?

   (b) What is the probability that the first seven will occur somewhere in the first 5 rolls?

   (c) How many rolls would you expect to make to obtain the first seven?