Negative Binomial Distribution

A discrete random variable Y is said to have a negative binomial distribution if

- There are independent and identical trials. Each trial can be thought of as a draw from a population, where the draw 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 r^{th} success occurs.
- The random variable Y is defined as the number of the trial on which the r^{th} success occurs.
- The parameters for the negative binomial random variable Y are the probability of success on each trial p and the number of the r^{th} success.
- The probability distribution function of the negative binomial random variable Y is

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

- The theoretical mean of the negative binomial random variable Y is

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

- The theoretical variance of the negative binomial random variable Y is

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

Working with negative binomial random variables in R.

The built-in function in R for the negative binomial distribution is different from the distribution described in your textbook. To eliminate any confusion, I have written a function in R called **dnbin** that matches the distribution in your textbook. Before you calculate any probabilities for the negative binomial distribution, you will need to type in this function in R.

```
dnbin<- function(y,r,p){choose(y-1,r-1)*p^r*(1-p)^(y-r)}
```

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

```
dnbin(y,r,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 r and y.

```
sum(dnbin(r:y,r,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(dnbin(y1:y2,r,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(dnbin(r:y-1,r,p))
```

Problems.

1. How is the probability distribution function $p(y)$ derived?
2. 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 fifth hole drilled yields the second productive well?
   (b) If the prospector can only afford to drill at most ten wells, what is the probability that the prospector will find two productive wells?
   (c) Find the mean and variance of the number of wells that must be drilled if the prospector wants to establish three wells.
3. 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 second seven will occur on the 6th roll?
   (b) What is the probability that the third seven will not occur in the first ten rolls?
   (c) How many rolls would you expect to make to obtain the second seven?