Consider again a **binomial experiment** :

- Each trial results in one of two outcomes: success, S, and failure, F.
- $P(S) = p$ and $P(F) = q = (1 - p)$; remain the same from trial to trial.
- The trials are independent.
- Perform the experiment until we obtain the r^{th} success.

Let Y be the *waiting time* until the r^{th} success, where $r \geq 1$ is an integer:

Y = the number of the trial on which the r^{th} success occurs, or

Y = the number of trials up to, and including, the r^{th} success.

Then Y is said to have a **negative binomial distribution with parameters r and p** $(0 < p < 1)$. We write **$Y \sim$ NegBin(r, p)**.

Warning: Some let Y be the # of failures *before* the r^{th} success.

Notice: NegBin(1, p) ~ Geometric(p), so the Negative Binomial is a generalization of the Geometric, and the Geometric is a special case of the Negative Binomial.

$Y \sim$ NegBin(r, p) is the waiting time until the r^{th} success in a binomial experiment with $P(S) = p$ and $P(F) = q = 1 - p$.

The support of Y is $\{r, r+1, r+2, r+3, \ldots\}$.

The probability function of Y: If $Y = y$, there were y trials with:

- $(y - r)$ failures,
- r successes, and
- the last trial was a success.

	___	___		___	S
trial	#1	#2	. . .	#(y-1)	#y

So there were exactly $(r-1)$ successes in the first $(y-1)$ trials, which can happen in $\binom{y-1}{r-1}$ ways: choose $(r-1)$ of the first $(y-1)$ trials for S's.

Each sequence of r S's and $(y - r)$ F's has probability $p^r q^{y-r}$, so

$$p(y) = \begin{cases} \binom{y-1}{r-1} p^r q^{y-r}, & \text{for } y = r, r+1, r+2, \cdots \\ 0, & \text{elsewhere} \end{cases}.$$

Definition. Y is said to have a **negative binomial distribution with parameters r and p**, $0 < p < 1$, $r = 1, 2, 3, \ldots,$ if Y has probability function

$$p(y) = \begin{cases} \binom{y-1}{r-1} p^r q^{y-r}, & \text{for } y = r, r+1, r+2, \ldots \\ 0, & \text{elsewhere} \end{cases}.$$

The reason for the name is that $\binom{y-1}{r-1} = \binom{y-1}{y-r}$ is the coefficient

of x^{y-r} in the expansion of $(1-x)^{-r} = \sum_{k=0}^{\infty} \binom{r+k-1}{k} x^k$

in *Newton's Binomial Theorem*, a generalization of the Binomial Theorem to include negative exponents. [Let $k = y - r$ in the coefficient.]

$Y \sim$ NegBin(r, p), so $p(y) = \begin{cases} \binom{y-1}{r-1} p^r q^{y-r}, & \text{for } y = r, r+1, r+2, \ldots \\ 0, & \text{elsewhere} \end{cases}$

Example. An experiment consists of drawing a card at random from a standard deck and replacing it. If the experiment is done repeatedly:

(a) What is the probability the third heart appears on the tenth draw?

(b) What is the mean number of non-hearts drawn ***before*** the third heart is drawn?

Solution. Let Y be the number of cards drawn up to and including the third heart. Then Y has a negative binomial distribution with $r = 3$, $p = \frac{1}{4}$, and probability function

$$p(y) = \binom{y-1}{3-1} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^{y-3}, \text{ for } y = 3, 4, 5, \cdots$$

$Y \sim \text{NegBin}\left(3, \frac{1}{4}\right)$, so $p(y) = \binom{y-1}{3-1}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^{y-3}$, for $y = 3, 4, 5, \ldots$

(a) What is the probability the third heart appears on the tenth draw?

Restating the problem in terms of the random variable Y, we have:

The third heart (success) occurs on the 10^{th} draw (trial) is equivalent to the event that $Y = 10$. Thus we are asked for

$$P(Y=10) = p(10) = \binom{10-1}{3-1}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^{10-3} = \binom{9}{2}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^7$$

$$= 36\,(.25)^3\,(.75)^7 \approx 0.075$$

6

(b) What is the mean number of non-hearts drawn ***before*** the 3^{rd} heart? In terms of the random variable, Y, this is asking for $E(Y - 3) = E(Y) - 3$.

To motivate the formula for $E(Y)$, consider the following:

If $Y \sim \text{NegBin}(r, p)$, the number of trials up to (and including) the r^{th} success, then $Y = Y_1 + Y_2 + \ldots + Y_r$, where

Y_1 = # trials up to the 1st success;

Y_2 = # trials after the 1st and up to the 2nd success;

Y_3 = # trials after the 2nd and up to the 3rd success; etc.

Then each $Y_i \sim \text{Geometric}(p)$ and has mean $1/p$.

So we might expect that since Y is the sum of the Y_i, the average value of Y is the sum of the average values of the Y_i; that is,

$$E(Y) = E(Y_1) + E(Y_2) + \cdots + E(Y_r) = r(1/p).$$

It is, but this is not a proof.

7

Theorem. If $Y \sim \text{NegBin}(r, p)$, then $E(Y) = \frac{r}{p}$ and $V(Y) = \frac{rq}{p^2}$.

Example. $Y \sim \text{NegBin}\left(3, \frac{1}{4}\right)$; $p(y) = \binom{y-1}{3-1}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^{y-3}$, for $y = 3, 4, \ldots$

(b) What is the mean number of non-hearts drawn before the 3^{rd} heart?

In terms of the random variable, Y, this is asking for $E(Y - 3)$, so

$$E(Y-3) = E(Y) - 3 = \frac{3}{(1/4)} - 3 = 12 - 3 = 9\,.$$

(c) What are the variance and standard deviation of the number of non-hearts drawn before the 3^{rd} heart? In terms of the random variable, Y, this is asking for $V(Y - 3)$ and σ_{Y-3}, so

$$V(Y-3) = V(Y) = \frac{3\cdot(3/4)}{(1/4)^2} = \frac{(9/4)}{(1/16)} = 36 \text{ and } \sigma_{Y-3} = \sqrt{V(Y-3)} = 6.$$

8

Comparison of the two forms of the Negative Binomial Distributions.

If X = # failures before the r^{th} success
and Y = # trials up to and including the r^{th} success, then

- $Y = X + r$ or $X = Y - r$,
- Support $(X) = \{0, 1, 2, 3, \ldots\}$ and Support $(Y) = \{r, r+1, r+2, \ldots\}$
- $p_X(x) = \binom{x+r-1}{r-1} q^x p^r,\ x = 0,1,2,\ldots$ and

  $p_Y(y) = \binom{y-1}{r-1} q^{y-r} p^r,\ y = r, r+1, r+2, \ldots$
- $\left\{\begin{array}{l} E(Y) = \frac{r}{p} \\ V(Y) = \frac{rq}{p^2} \end{array}\right\} \Rightarrow \left\{\begin{array}{l} E(X) = E(Y-r) = E(Y) - r = \frac{r}{p} - r = \frac{r - rp}{p} = \frac{rq}{p} \\ V(X) = V(Y-r) = V(Y) = \frac{rq}{p^2} \end{array}\right\}$

9