XIV. Negative Binomial Distribution

Negative Binomial Distribution

- The negative binomial distribution describes a sequence of trials, each of which can have two outcomes (success or failure).
- We continue the trials indefinitely until we get r successes.
- The prototypical example is flipping a coin until we get r heads.
- Unlike the binomial distribution, we don't know the number of trials in advance.
- The geometric distribution is the case $r = 1$.

Could be rolling a die, or the Yankees winning the World Series, or whatever.

Formula for the Negative Binomial Distribution

- Fixed parameters:

$$\begin{aligned} p &:= \text{probability of success on each trial} \\ q &:= \text{probability of failure } = 1 - p \\ r &:= \text{number of successes desired} \end{aligned}$$

- Random variable:

$$Y := \text{number of trials (for } r \text{ successes)}$$

- Probability distribution:

$$\boxed{p(y) = \binom{y-1}{r-1} p^r q^{y-r}, r \leq y < \infty}$$

Warning: These are different p's!
p is the probability of success on any given trial.
$p(y)$ is the probability of y trials overall.

Key Properties of the Negative Binomial Distribution

- **Mean**:

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

- **Variance**:

$$\boxed{\sigma^2 = V(Y) = \frac{rq}{p^2}}$$

- **Standard deviation**:

$$\boxed{\sigma = \sqrt{V(Y)} = \frac{\sqrt{rq}}{p}}$$

Example I

You draw cards from a deck (with replacement) until you get four aces. What is the chance that you will draw exactly 20 times?

$$\begin{aligned}
p &= \frac{1}{13} \\
p(20) &= \binom{y-1}{r-1} p^r q^{y-r} \\
&= \binom{19}{3} \left(\frac{1}{13}\right)^4 \left(\frac{12}{13}\right)^{16} = \boxed{\binom{19}{3} \left(\frac{12^{16}}{13^{20}}\right)}
\end{aligned}$$

Example II

Each year the Akron Aardvarks have a 10% chance of winning the trophy in chinchilla grooming. Their trophy case has space for five trophies. Let Y be the number of years until their case is full. Find the mean and standard deviation of Y.

$$\begin{aligned}
p &= \frac{1}{10} \\
q &= \frac{9}{10} \\
r &= 5 \\
\mu &= \frac{r}{p} = \boxed{50 \text{ years}} \\
\sigma^2 &= \frac{rq}{p^2} = \frac{5\frac{9}{10}}{\frac{1}{10^2}} = 450 \\
\sigma &= \sqrt{450} = 15\sqrt{2} \approx \boxed{21.21 \text{ years}}
\end{aligned}$$

Example III

You roll a die until you get four sixes (not necessarily consecutive). What is the mean and standard deviation of the number of rolls you will make?

This is the negative binomial distribution with $p = \frac{1}{6}, r = 4$.

$$\begin{aligned}
\mu &= \frac{r}{p} \\
&= \boxed{24 \text{ rolls}} \\
\sigma^2 &= \frac{rq}{p^2} \\
&= \frac{4 \cdot \frac{5}{6}}{\frac{1}{36}} \\
&= 120 \\
\sigma &= \sqrt{120} = \boxed{2\sqrt{30} \approx 10.95 \text{ rolls}}
\end{aligned}$$

Example IV

10% of applicants for a job possess the right skills. A company has three positions to fill, and they interview applicants one at a time until they fill all three positions.

A. What is the probability that they will interview exactly ten applicants?

B. What is the probability that they will interview at least ten applicants?

Example IV

A. Exactly ten applicants?

B. At least ten applicants?

This is the negative binomial distribution with $p = \frac{1}{10}, q = \frac{9}{10}, r = 3$.

A.

$$\begin{aligned} p(10) &= \binom{y-1}{r-1} p^r q^{y-r} \\ &= \binom{9}{2} \left(\frac{1}{10}\right)^3 \left(\frac{9}{10}\right)^7 = \boxed{36 \left(\frac{9^7}{10^{10}}\right)} \end{aligned}$$

B. What is the probability that they will find two or fewer out of the first nine? Use the binomial distribution:

$$\begin{aligned} p(y) &= \binom{n}{y} p^y q^{n-y} \\ p(0) + p(1) + p(2) &= \left(\frac{9}{10}\right)^9 + 9 \cdot \frac{9^8}{10^9} + 36 \cdot \frac{9^7}{10^9} \\ &= \frac{9^7(81 + 81 + 45)}{10^9} \\ &= \boxed{\frac{473513931}{5 \cdot 10^8} \approx 94.7\%} \end{aligned}$$

Example V

The company from Example IV takes three hours to interview an unqualified applicant and five hours to interview a qualified applicant. Calculate the mean and standard deviation of the time to conduct all the interviews.

The time is $T = 3(Y-3)+15 = 3Y+6$. The mean is $E(T) = 3E(Y)+6 = 3\left(\frac{r}{p}\right)+6 = 3\left(\frac{3}{\frac{1}{10}}\right)+6$
$\boxed{96 \text{ hours}}$.
The variance ($V(aY+b) = a^2V(Y)$) is $V(T) = 9V(Y) = 9\frac{rq}{p^2} = 9\left(\frac{\frac{27}{10}}{\frac{1}{100}}\right)$ 2430 hours^2.
The standard deviation is $\sqrt{2430} = 9\sqrt{30} \approx$ $\boxed{49.295 \text{ hours}}$.