XIII. Geometric Distribution

Geometric Distribution

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

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

Formula for the Geometric Distribution

- Fixed parameters:

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

- Random variable:

$$Y := \text{number of trials (for one success)}$$

- Probability distribution:

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

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 Binomial Distribution

- **Mean**:

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

- **Variance**:

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

- **Standard deviation**:

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

Geometric Series

- **Recall from Calculus II**:

$$\begin{aligned} a + ar + ar^2 + \cdots &= \frac{a}{1-r} \\ &= \boxed{\frac{\text{first term}}{1 - \text{common ratio}}} \end{aligned}$$

- **Application to geometric distribution**:

$$\begin{aligned} P(Y \geq y) &= p(y) + p(y+1) + p(y+2) + \cdots \\ &= q^{y-1}p + q^y p + q^{y+1}p + \cdots \\ &= \frac{q^{y-1}p}{1-q} = \frac{q^{y-1}p}{p} = \boxed{q^{y-1}} \end{aligned}$$

Example I

You draw cards from a deck (with replacement) until you get an ace. What is the chance that you will draw exactly 3 times?

$p = \frac{1}{13}; p(3) = q^2 p = \boxed{\frac{12^2}{13^3}}$

Example II

Each year the Akron Aardvarks have a 10% chance of winning the North-midwestern championship in Pin The Tail On The Donkey. Let Y be the number of years until they next win. Find the mean and standard deviation of Y.

$$\begin{aligned}
p &= \frac{1}{10} \\
q &= \frac{9}{10} \\
\mu &= \frac{1}{p} = \boxed{10 \text{ years}} \\
\sigma^2 &= \frac{q}{p^2} = \frac{\frac{9}{10}}{\frac{1}{10^2}} = 90 \\
\sigma &= \sqrt{90} = 3\sqrt{10} \approx \boxed{9.487 \text{ years}}
\end{aligned}$$

Example III

You and a friend take turns rolling a die. You roll first, and the first person to roll a six wins.

A. What is the chance that you will win on your third roll?

B. What is the chance that your friend will get to roll three times or more?

C. What is the chance that you will win?

This is the geometric distribution with $p = \frac{1}{6}, q = \frac{5}{6}$.

A. $p(5) = q^4 p = \boxed{\frac{5^4}{6^5}}$.

B. The first five rolls must fail to be six, so the chance is $\boxed{\left(\frac{5}{6}\right)^5}$.

C.

$$\begin{aligned} p(1) + p(3) + p(5) + \cdots &= p + q^2 p + q^4 p + \cdots \\ &= \frac{p}{1 - q^2} \\ &= \frac{\frac{1}{6}}{1 - \frac{25}{36}} \\ &= \boxed{\frac{6}{11}} \end{aligned}$$

This is a geometric series with $r = q^2$.

Example III

A. You will win on your third roll?

B. Your friend will roll three times or more?

C. You will win?

This is the geometric distribution with $p = \frac{1}{6}, q = \frac{5}{6}$.

A. $p(5) = q^4 p = \boxed{\frac{5^4}{6^5}}$.

B. The first five rolls must fail to be six, so the chance is $\boxed{\left(\frac{5}{6}\right)^5}$.

C.

$$\begin{aligned} p(1) + p(3) + p(5) + \cdots &= p + q^2 p + q^4 p + \cdots \\ &= \frac{p}{1 - q^2} \\ &= \frac{\frac{1}{6}}{1 - \frac{25}{36}} \\ &= \boxed{\frac{6}{11}} \end{aligned}$$

This is a geometric series with $r = q^2$.

Example IV

10% of applicants for a job possess the right skills. A company interviews applicants one at a time until they find a qualified applicant.

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?

This is the geometric distribution with $p = \frac{1}{10}, q = \frac{9}{10}$.

A. $p(10) = q^9 p = \boxed{\frac{9^9}{10^{10}}}$.

B. $P(Y \geq 10) = q^{y-1} = \boxed{\frac{9^9}{10^9}}$.

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 $T = 3(Y-1)+5 = 3Y+2$. The mean is $E(T) = 3E(Y) + 2 = 3\left(\frac{1}{p}\right) + 2 = \boxed{32 \text{ hours}}$. The variance $(V(aY+b) = a^2V(Y))$ is $V(T) = 9V(Y) = 9\frac{q}{p^2} = 810 \text{ hours}^2$, so the standard deviation is $\sqrt{810} = 9\sqrt{10} \approx \boxed{28.46 \text{ hours}}$.