LECTURE: GEOMETRIC DISTRIBUTION

1. Geometric Distribution

Recall:

A **Bernoulli trial** is a sequence of independent experiments with probability p of success and $1-p$ of failure

Question: How many trials does it take to get your first success?

This is answered precisely by the geometric distribution.

Example: If you play a slot machine repeatedly until you win, the number of plays it takes to get your first win follows the geometric distribution.

Consider a sequence of Bernoulli trials with probability of success p.

We will perform a as many trials until we get a success.

Let $Y =$ the trial on which the first success occurs

This is a random variable with values $1, 2, 3, \cdots$

We can describe all possible outputs of our experiment by strings consisting of the letters `S` (success) and `F` (failure), like `FFFS`. The event $(Y = 1)$ corresponds `S`, $(Y = 2)$ corresponds to `FS`, etc.

We have the following table for the pmf of Y

y	Event	$p(y)$
1	S	p
2	FS	$(1-p)p$
3	FFS	$(1-p)^2p$
4	FFFS	$(1-p)^3p$
$\vdots$	$\vdots$	$\vdots$
k	$\underbrace{\texttt{FFF}\ldots\texttt{F}}_{k-1 \text{ times}}\texttt{S}$	$(1-p)^{k-1}p$

We can use this to define a geometric random variable.

Definition:

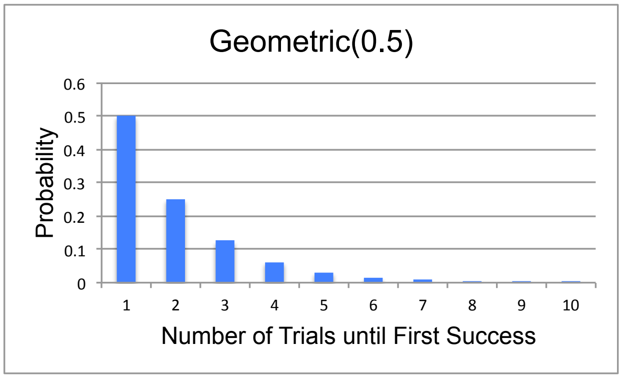
A discrete random variable Y has a **geometric distribution** with probability of success p if:

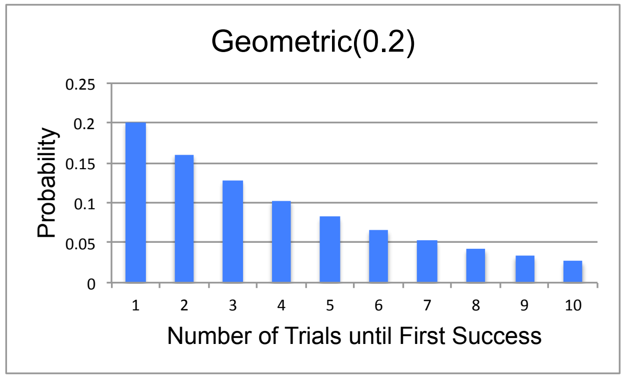
$$p(y) = (1-p)^{y-1}p \qquad\qquad y = 1, 2, 3, \ldots$$

We write $Y \sim \text{Geom}(p)$

Interpretation: This models the number of trials it takes to get the first success, where the probability of a single success is p.

Histograms: Here are histograms for the geometric distribution for parameters 0.5 and 0.2.





In all cases, the probability $p(y)$ decreases as y increases. The $p = 0.5$ case is more steep than the $p = 0.2$ case. This is because for $p = 0.5$ the chance of success is higher, so it takes less trials to win.

2. Properties

Once again, let's show that the sum of probabilities is 1.

Geometric Series:

$$\sum_{k=1}^{\infty} r^{k-1} = 1 + r + r^2 + \cdots = \frac{1}{1-r} \qquad \text{if } |r| < 1$$

Suppose $Y \sim$ Geom (p) then

$$\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} p(1-p)^{k-1} = p\left(\frac{1}{1-(1-p)}\right) = p\left(\frac{1}{p}\right) = 1 \checkmark$$

Example 1:

Suppose $Y \sim$ Geom (p) then what is $E(Y)$?

This involves a nice calculus trick. Assume first that $0 < p < 1$

$$
\begin{aligned}
E(Y) &\overset{\text{DEF}}{=} \sum_{y=1}^{\infty} y p(y) \\
&= \sum_{y=1}^{\infty} y(1-p)^{y-1} p = p \sum_{y=1}^{\infty} y q^{y-1} \text{ where } q = 1-p \\
&= p \sum_{y=1}^{\infty} \frac{d}{dq} q^y = p \left(\frac{d}{dq} \sum_{y=1}^{\infty} q^y \right) = p \frac{d}{dq} \left(\frac{q}{1-q} \right) \text{ Geometric Series} \\
&= p \left(\frac{(1-q)(1) - (q)(-1)}{(1-q)^2} \right) = \frac{p}{p^2} = \frac{1}{p}
\end{aligned}
$$

Finally, if $p = 1$ then you're guaranteed to get your first success in the first trial, so $E(Y) = 1 = \frac{1}{p}$ and if $p = 0$ then you'll never succeed, so $E(Y) = \infty = \frac{1}{p}$.

The proof of the variance is a bit more tedious and will be omitted.

Facts:

If $Y \sim$ Geom (p) then $E(Y) = \frac{1}{p}$ and $\mathrm{Var}(Y) = \frac{1-p}{p^2}$

Application: The geometric distribution is often used the model the distribution of how long you need to wait for a particular event to happen. In this case, the waiting time is given in discrete "chunks" of time such as hours or days.

Example 2:

Suppose the probability of your computer crashing in any given 1-hour period is 0.05

(a) What is the probability that the computer crashes within the first two hours after start-up?

Let X be the number of one-hour intervals until your computer crashes.

We will model $X \sim$ Geom (0.05) where "success" is defined as your computer crashing.

$P(X \leq 2) = P(X = 1) + P(X = 2) = p + (1-p)p = (0.05) + (0.95)(0.05) = 0.0975$

(b) What is the probability that your computer will still be running two hours after start-up?

$$P(X > 2) = 1 - P(X \leq 2) = 1 - 0.0975 = 0.9025$$

Other Method: If your computer is still running after 2 hours, this means it did not crash in Hour 1 and in Hour 2

$$\begin{aligned} P(X > 2) &= P(\text{no crash in Hour 1} \cap \text{no crash in Hour 2}) \\ &= P(\text{no crash in Hour 1})\, P(\text{no crash in Hour 2}) \\ &= (1-p)^2 \\ &= 0.95^2 \\ &= 0.9025 \end{aligned}$$

So there is a rougly 90% chance your computer is still running after 2h

(c) What is the expected number of hours your computer will run until it crashes?

$$E(X) = \frac{1}{p} = \frac{1}{0.05} = 20$$

3. Memoriless Property

(d) Given that your computer is still running after 10h, what is the probability that it is still running for two more hours?

We want the probability that the computer does not crash in Hour 11 or Hour 12 given that's it's been running after 10h

In other words, we know that $X > 10$ and we want to know when $X > 12$

So we are interested in $P(X > 12 | X > 10)$

By definition of conditional probability

$$P(X > 12 | X > 10) = \frac{P(X > 12 \cap X > 10)}{P(X > 10)} = \frac{P(X > 12)}{P(X > 10)}$$

This is because if $X > 12$ is more specific than $X > 10$, that is $(X > 12) \subseteq (X > 10)$ (subsets)

$$\begin{aligned}
& P(X > 12) \\
=& P(\text{ No crash in Hour 1 } \cap \text{ No crash in Hour 2 } \cdots \cap \text{ No crash in Hour 12 }) \\
=& (1-p)(1-p)\cdots(1-p) \\
=& (1-p)^{12}
\end{aligned}$$

Similarly, $P(X > 10) = (1-p)^{10}$ and therefore

$$P(X > 12|X > 10) = \frac{P(X > 12)}{P(X > 10)} = \frac{(1-p)^{12}}{(1-p)^{10}} = (1-p)^2 = 0.9025$$

Important Observation: This is the same as $P(X > 2)$!!

$$P(X > 12|X > 10) = P(X > 2)$$

In other words, the probability that the computer is still running after two more hours given that it has already been running for 10h is the same as the probability that the computer is still running two hours after startup.

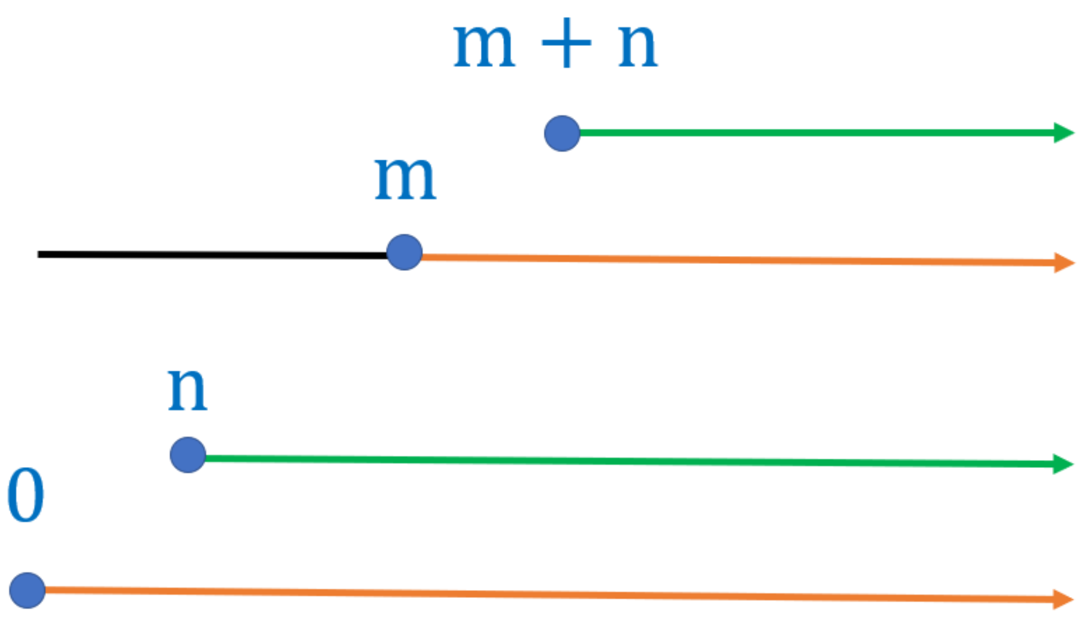
So it's as if the first 10 hours didn't matter at all!

This property of is called the *memoriless property* of the geometric distribution:

Memoriless Property:

If $Y \sim$ Geom (p) then for all m and n

$$P(Y > m + n|Y > m) = P(Y > n)$$



In other words, the number of trials for the first success after you've done m trials is the same as if you started from scratch, you're somehow forgetting about the first m trials.

This makes sense in practice. For example, if you already flipped 50 coins, then the outcome of the 51st flip doesn't depend on the previous coin flips, so no betting strategy based on past outcomes should work.