Starting Off:

What is a mathematical function?

Probability Distributions

Data Science Immersive

FLATIRON SCHOOL

After this lesson, you will be able to...

- Distinguish between discrete and continuous random variables.
- Compute probabilities, cumulative probabilities, means and variances for discrete random variables.
- Identify binomial random variables and their characteristics.
- Calculate probabilities of binomial random variables.

Terminology

Random Variable

Is a variable whose value results from the process of a random experiment

Types of Random Variables:

Qualitative Random Variables

The possible values vary in kind but not in numerical degree. They are also called categorical variables. (Favorite TV show, Relationship status, College Major)

Quantitative Random Variables

There are two types of quantitative random variables.

- Discrete Random Variable: When the random variable can assume only a countable, sometimes infinite, number of values. (# of cousins, years in school)
- Continuous Random Variable: When the random variable can assume an uncountable number of values in a line interval. (Age, Height, Weight)

Probability Functions

Probability Function

A probability function is a mathematical function that provides probabilities for the possible outcomes of the random variable, X. It is typically denoted as $f(x)$.

Probability Mass Function (PMF)

If the random variable is a discrete random variable, the probability function is usually called the probability mass function (PMF). If X is discrete, then $f(x)=P(X=x)$. In other words, the PMF for a constant, x, if the probability that the random variable X is equal to x.

Probability Density Function (PDF)

If the random variable is a continuous random variable, the probability function is usually called the probability density function (PDF). Contrary to the discrete case, $f(x)\neq P(X=x)$.

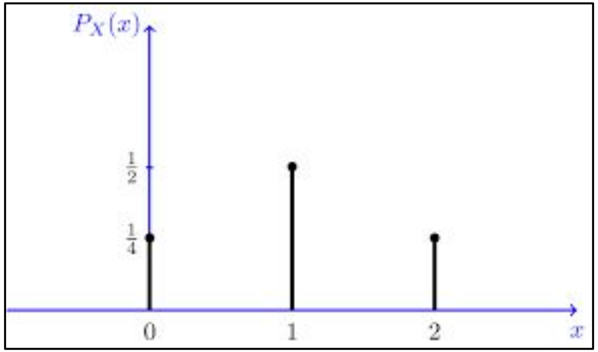
Probability Mass Function

- Probability Mass Function is a function that maps the frequency of a **discrete** set of data to distribution

$$\boldsymbol{f}\mathbf{X(x) = P(X = x)}$$

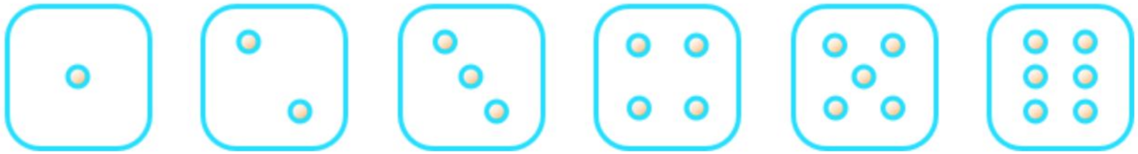
- The values of pmf must sum to 1
- f(x) can only take on values [0,1]

$$f_X(x) = \Pr(X = x) = P(\{s \in S : X(s) = x\})$$

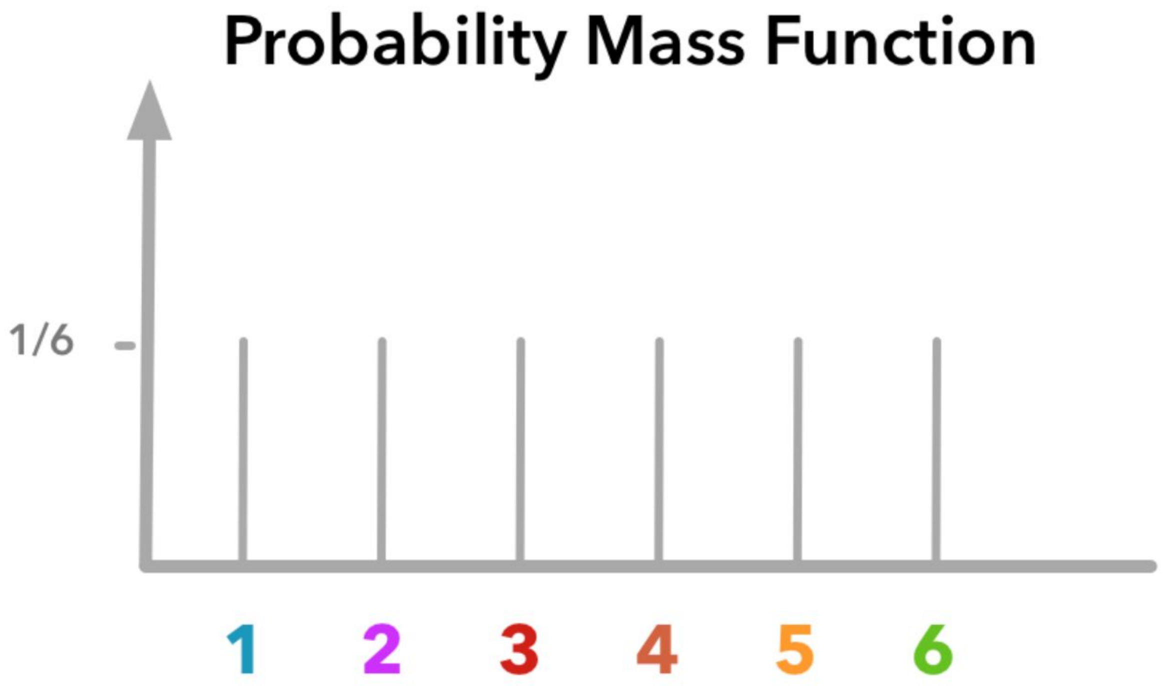


Probability Mass Function - Example

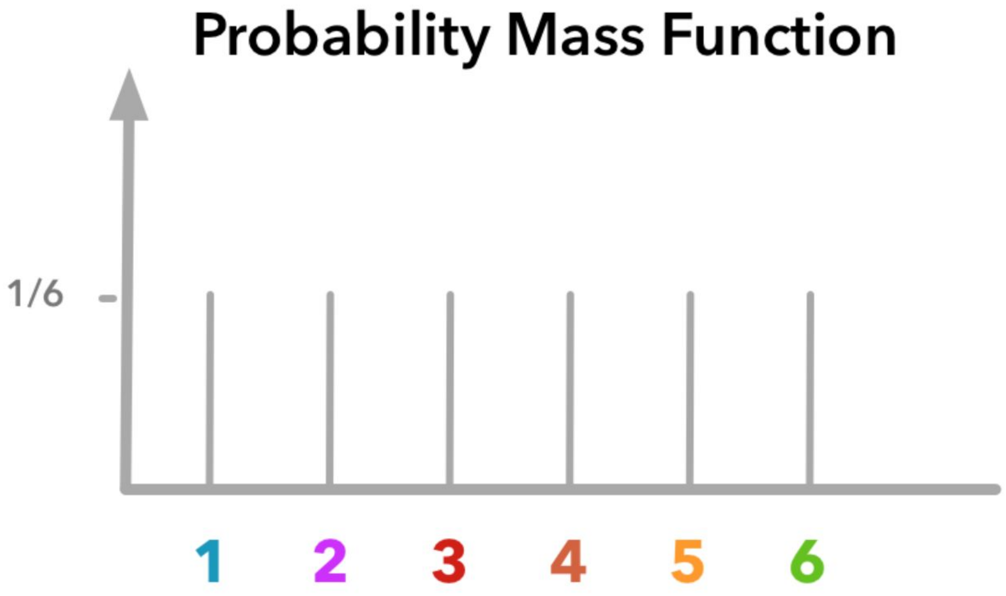
- Draw the probability distribution of a fair, standard dice roll
  - What are the values the random variable could possibly take on?
  - What is the probability of rolling each value?



Probability Mass Function - Example



Probability Mass Function - Example



Practice: Using the PMF to the left, calculate the probability of each of the follow questions:

A) $P(X = 1)$

B) $P(X \leq 2)$

C) $P(X > 3)$

D) $P(1 \leq X \leq 3)$

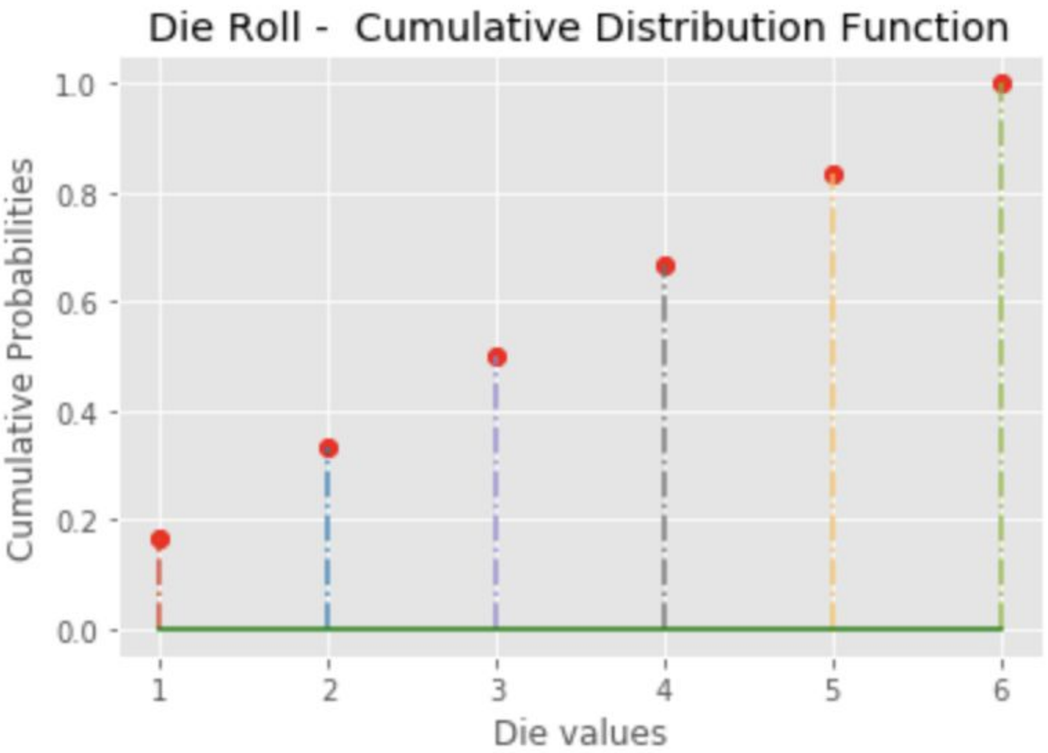
Cumulative Mass Function

Cumulative frequency sums the frequencies <= a particular value.

$$F(x) = P(X \leq x)$$

Outcome	Frequency	Cumulative Frequency	Cumulative frequency probability
1	1	1	.167
2	1	2	.33
3	1	3	.5
4	1	4	.67
5	1	5	.833
6	1	6	1

Cumulative Mass Function



Cumulative Mass Function

The following is a table showing how 11 students rated the movie “US” on a scale of 1-5. Complete the table and then draw the CMF.

Outcome	Frequency	Cumulative Frequency	Cumulative frequency probability
1	1	1	
2	2		
3	3		
4	3		
5	2		

Discrete Random Variable

By continuing the dice rolling example, if we were to roll the die 10,000 times what value should we expect to get? What would be the average value?

Expected Value (or mean) of a Discrete Random Variable:

$$\mu = E(X) = \sum x_i f(x_i)$$

The formula means that we multiply each value, *x*, in the support by its respective probability, *f(x)*, and then add them all together. It can be seen as an average value but weighted by the likelihood of the value.

Variance

Variance of a Discrete Random Variable

The variance of a discrete random variable is given by:

$$\sigma^2 = \mathrm{Var}(X) = \sum (x_i - \mu)^2 f(x_i)$$

$$\sigma^2 = \mathrm{Var}(X) = \sum x_i^2 f(x_i) - E(X)^2 = \sum x_i^2 f(x_i) - \mu^2$$

Standard Deviation of a Discrete Random Variable

The standard deviation of a random variable, *X*, is the square root of the variance.

$$\sigma = \mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{\sigma^2}$$

Deal or No Deal

You have 5 suitcases in front of you that have a value of \$1, \$1,000, \$10,000, \$50,000 and \$100,000. You can choose 1 suitcase and win that amount of money or take a offer of \$30,000.

- Calculate the expected value of picking a suitcase.
- If you were to base your decision solely on the expected value of a suitcase, what would you do?

Deal or No Deal - Bonus question

Let’s rethink the game a little. This time, you get to pick a suitcase and keep the amount in the suitcase or you can put the suitcase back reshuffle and pick again. The amount in the second pick will be your final amount.

What is the expected value of this this proposition?

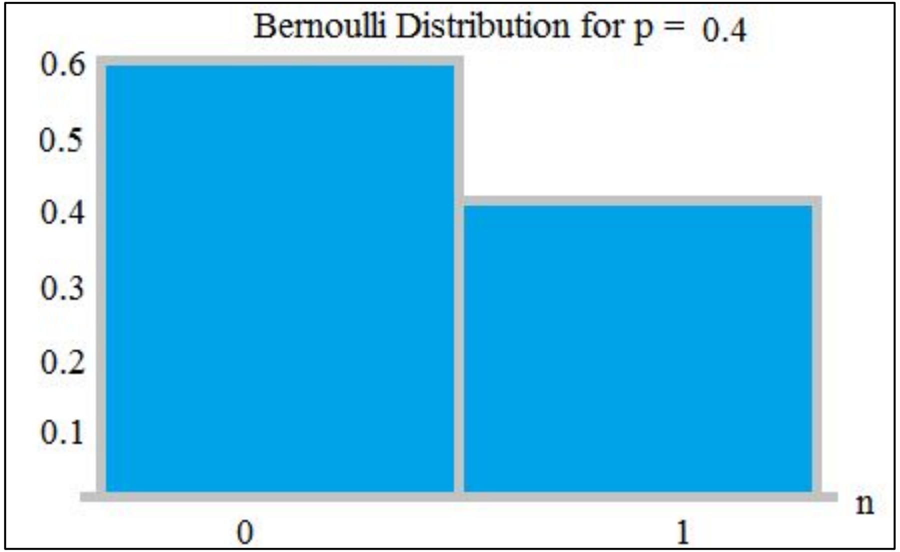
Probability Distributions

In probability theory and statistics, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment.It is a mathematical description of a random phenomenon in terms of its sample space and the probabilities of events (subsets of the sample space).

The Bernoulli Trial

Bernoulli Trials

Bernoulli trials are experiments with two outcomes, such as coin flip, win/lose etc



How can a dice roll experiment be thought of as Bernoulli Trials?

The Binomial Distribution

Binomial distribution

The binomial distribution is a special discrete distribution where there are two distinct complementary outcomes, a “success” and a “failure”.

A binomial random variable is random variable that represents the number of successes in *n* successive independent trials of a Bernoulli experiment.

Some example uses include the number of heads in n coin flips, the number of disk drives that crashed in a cluster of 1000 computers, and the number of advertisements that are clicked when 40,000 are served.

The Binomial Distribution

The Binomial Distribution

We have a binomial experiment if ALL of the following four conditions are satisfied:

1. The experiment consists of *n* identical trials.
2. Each trial results in one of the two outcomes, called success and failure.
3. The probability of success, denoted p, remains the same from trial to trial.
4. The n trials are independent. That is, the outcome of any trial does not affect the outcome of the others.

If the four conditions are satisfied, then the random variable *X*=number of successes in *n* trials, is a binomial random variable with

$\mu = E(X) = np$ (Mean)

$\mathrm{Var}(X) = np(1-p)$ (Variance)

$\mathrm{SD}(X) = \sqrt{np(1-p)}$, where p is the probability of the "success." (Standard Deviation)

The Binomial Distribution

Imagine you flipped a fair coin 3 times. What is the probability that you will get heads exactly 1 time in 3 flips?

In order to determine this, let's first calculate the outcome space and the event spaces.

How many different outcomes can you come up with for 3 flips of a fair coin?

How many of those have exactly 1 heads?

Can use combinations and permutations to calculate this more quickly.

Create a PMF for the number of heads you can get from this experiment.

The Binomial Distribution

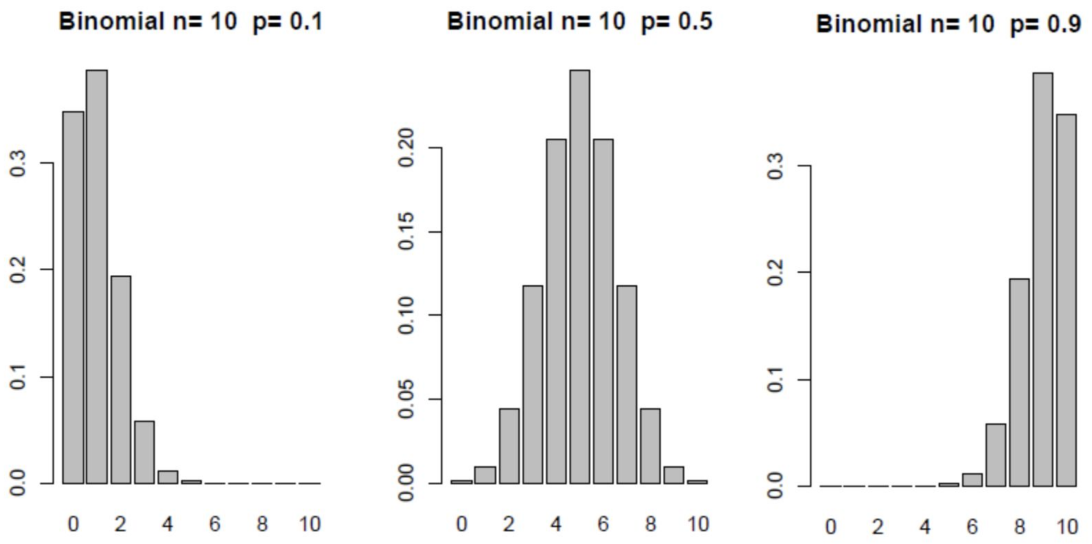
The Binomial Formula

For a binomial random variable with probability of success, p, and n trials...

$$f(x) = P(X = x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

Binomial Distributions

The formula defined above is the probability mass function, pmf, for the Binomial. We can graph the probabilities for any given *n* and *p*. The following distributions show how the graphs change with a given n and varying probabilities.



The Binomial Distribution

An FBI survey shows that about 80% of all property crimes go unsolved. Suppose that in your town 3 such crimes are committed and they are each deemed independent of each other. What is the probability that 1 of 3 of these crimes will be solved?

First, we must determine if this situation satisfies ALL four conditions of a binomial experiment:

1. Does it satisfy a fixed number of trials? YES the number of trials is fixed at 3 (n = 3.)
2. Does it have only 2 outcomes? YES (Solved and unsolved)
3. Do all the trials have the same probability of success? YES (p = 0.2)
4. Are all crimes independent? YES (Stated in the description.)

To find the probability that only 1 of the 3 crimes will be solved we first find the probability that one of the crimes would be solved. With three such events (crimes) there are three sequences in which only one is solved:

The Binomial Formula

For a binomial random variable with probability of success, p, and n trials...

$$f(x) = P(X = x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

The Binomial Distribution

If a fair coin is flipped 10 times, what is the probability that it will come up heads at least eight out of those 10 times?

The Binomial Formula

For a binomial random variable with probability of success, p, and n trials...

$$f(x) = P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

Brain Teaser

If 1024 fair coins are each tossed 10 times, what are the chances that at least one coin will come up heads 10 times in a row?