CS440/ECE448 Lecture 11: Random Variables

edited by Julia Hockenmaier, February 2019

Random Variables

- Random Variables
  - RV = function from outcomes to numbers
  - Notation
  - Probability Mass Function (pmf)
  - Expected Value
- Domain of a Random Variable
  - Domain Type: Categorical vs. Numerical
  - Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite
- Joint, Marginal, and Conditional Random Variables
  - Marginalization and Conditioning
  - Law of Total Probability
  - Random Vectors
  - Jointly Random Class and Measurement Variables
- Functions of Random Variables
  - Probability Mass Function
  - Expectation

Sample space, Events, Probabilities

An **experiment/trial** is a procedure with a well-defined set of possible outcomes: flipping a coin, flipping a coin twice in a row,

The **sample space Ω** is the **set of all possible outcomes**

Single coin flips: {Head, Tail}

Sequence of two coin flips: { (Head, Head), (Head, Tail),…}

An **event** is a subset of the sample space

The empty subset has probability 1

The sample space itself (the set of all outcomes) has probability 1

If A and B are disjoint events, $P(A \cup B) = P(A) + P(B)$

Random variables

- **We describe the (uncertain) state of the world using *random variables***
  - Denoted by capital letters
  - **R**: *Is it raining?*
  - **W**: *What's the weather?*
  - **D**: *What is the outcome of rolling two dice?*
  - **S**: *What is the speed of my car (in MPH)?*
- Just like variables in CSPs, **random variables take on values in a *domain***
  - Domain values must be ***mutually exclusive* and *exhaustive***
  - **R** in {True, False}
  - **W** in {Sunny, Cloudy, Rainy, Snow}
  - **D** in {(1,1), (1,2), … (6,6)}
  - **S** in [0, 200]
- **Because domain values are mutually exclusive and exhaustive, each random variable defines a partition of the sample space**

Random variables

- **A random variable can be seen as a function that maps outcomes (elements of the sample space) to numbers**
  f:outcomes→numbers
- In the **partition of the sample space** defined by the random variable, each **number** corresponds to **one equivalence class of outcomes**
- The event "Speed=45mph" is the set of all outcomes for which the speed of my car is 45mph:
  - I have my foot on the accelerator pedal, and I'm traveling 45mph
  - My car is being towed, and the tow truck is traveling 45mph
  - My car is falling off a cliff, and has reached a terminal velocity of 45mph…

Random Variables are Uppercase, Instances are Lowercase

We use **an UPPERCASE** letter for a random variable,
and a **lowercase** letter for the actual **value** that it takes after any particular experiment.

- $X_1 = x_1$ is the **event** "random variable X_1 takes the value x_1"
- X_1 is a **RV**, which is a **function**, X_1:outcomes→numbers
- x_1 is just a number.

So, for example, the statement $X_1 = 3$ is a particular outcome of the experiment (the outcome in which the variable X_1 took the value of 3).

Probability Mass Function (pmf) is a lowercase p.

- $X_1 = x_1$ is the **event** "random variable X_1 takes the value x_1"

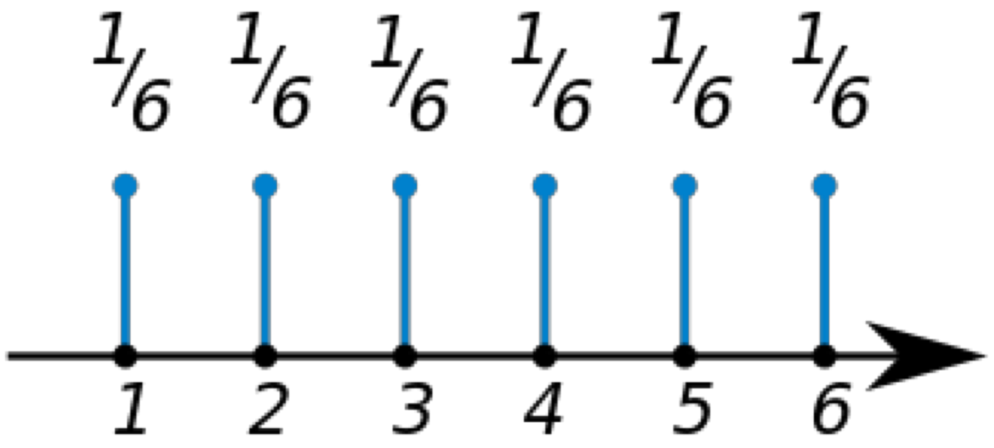
- $p(X_1 = x_1)$ is **the probability** that this event occurs.
  - We call this number the "probability mass" of the event $X_1 = x_1$
  - Shorthand: $p(x_1)$ using a small letter x_1, implies X_1
  - Subscript notation, which we won't use in this class: $\boldsymbol{p_{X_1}(x_1)}$

- $p(X_1)$ (with a capital letter X_1) is the **probability mass function (pmf):** a function from values of X_i to probabilities

  This is the entire table of the probabilities $X_1 = x_1$
  for every possible value x_1

Probability Mass Function

- The “Probability Mass Function” (pmf) of a random variable X is defined to be the function P(X=value), as a function of the different possible values.



Wikipedia: “The probability mass function of a fair die. All the numbers on the die have an equal chance of appearing on top when the die stops rolling.”

Requirements for a Probability Mass Function

<u>Axioms of Probability</u>

1. $P(A) \geq 0$ for every event A
2. $1 = P(True)$
3. $P(A \vee B) = P(A) +$
   $P(B) - P(A \wedge B)$

<u>Requirements for a pmf</u>

1. $P(X = x) \geq 0$ for every x
2. $1 = \sum_x P(X = x)$
3. $P((X = x_1) \vee (X = x_2)) =$
   $P(X = x_1) + P(X = x_2)$

Notice: the last one assumes that $X = x_1$ and $X = x_2$ are mutually exclusive events.

Expected Value

Expected Value of a random variable
= the average value of the random variable,
averaged over an infinite number of independent trials

= the **weighted average of the values** of the random variable,
where each value is **weighted by its probability**

$$E[\mathrm{D}] = \sum_{d \in D} \mathrm{P}(d) \times d$$

NB: The expected value might not be an actual outcome

With P(D = 1) = 0.5 and P(D = 0) = 0.5: E[D] = 0.5

Expected Value

Example: D = number of pips showing on a die

Expected Value of a random variable = the average value, averaged over an infinite number of independent trials

$$E[D] = \lim_{n\to\infty} \frac{1}{n}\big(1\times(\#\ times\ D = 1) + \cdots + 6\times(\#\ times\ D = 6)\big)$$

$$= \lim_{n\to\infty} \frac{1}{n}\big(1\times(n/6) + \cdots + 6\times(n/6)\big) = 3.5$$

Center of Mass (from physics)

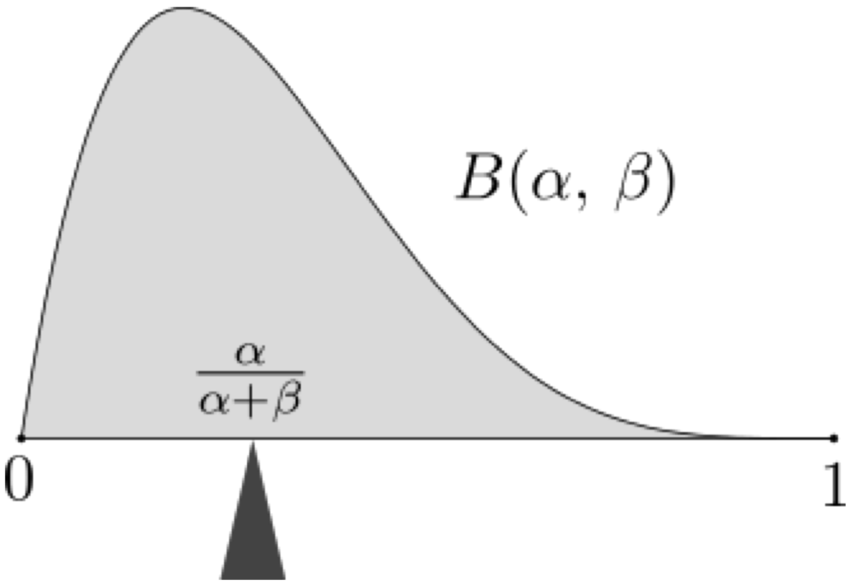
Center of Mass

= sum{ position * Mass(position) }

Expected Value = Center of Probability "Mass"

Expected Value of a random variable = the average value, averaged over an infinite number of independent trials

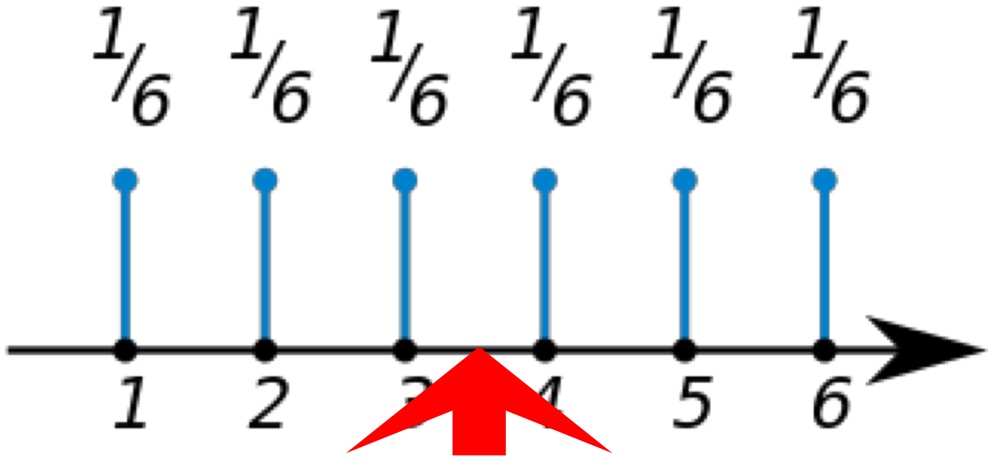
= sum{ value * P(variable=value) }



Wikipedia: "The mass of probability distribution is balanced at the expected value."

Probability Mass Function

- The "Probability Mass Function" (pmf) of a random variable X is defined to be the function P(X=value), as a function of the different possible values.
- Why it's useful: expected value = center of mass.



Wikipedia: "The probability mass function of a fair die. All the numbers on the die have an equal chance of appearing on top when the die stops rolling." The **expected value** is 3.5.

Random Variables

- Random Variables
  - RV = function from outcomes to numbers
  - Notation
  - Probability Mass Function (pmf)
  - Expected Value
- Domain of a Random Variable
  - Domain Type: Categorical vs. Numerical
  - Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite
- Joint, Marginal, and Conditional Random Variables
  - Marginalization and Conditioning
  - Law of Total Probability
  - Random Vectors
  - Jointly Random Class and Measurement Variables
- Functions of Random Variables
  - Probability Mass Function
  - Expectation

Domain of a Random Variable

- The "Domain" of a Random Variable
  is the set of its possible values.
- The domain can be **numerical**. For example:
  - The number of pips showing on a die
  - The age, in years, of a person that you choose at random off the street
  - The number of days of sunshine in the month of March
  - The minimum temperature tonight, in degrees Celsius
- The domain can also be **categorical**. For example:
  - The color chosen by a spinner in the game of Twister
  - The color of the shirt worn by a person chosen at random
  - The type of weather tomorrow: { sunny, cloudy with no precipitation, raining, snowing, sleet }

Domain of a Random Variable

- The domain can also be **categorical**… e.g. colors, etc.
- Hang on, didn't you just say random variable map outcomes to numbers?
- How does this work for categorical RVs?
- Solution: Thinking of the outcomes of RVs as numbers is a mathematical convenience
- We can map each category label to an integer:
  - Red = 1, Blue = 2, …

Expectation and PMF

- Expected Value is only well defined for numerical domains.

$$E[X] = \text{sum value} * P(X=\text{value})$$

- pmf is well defined even for categorical domains.

Example: X = color shown on the spinner

P(X=red) = (1/4)

P(X=blue) = (1/4)

P(X=green) = (1/4)

P(X=yellow) = (1/4)

Size of the Domain = # Different Possible Values

- <u>The domain of a random variable can be finite</u>.

Example: D = value, in dollars, of the next coin you find. Domain = {1.00, 0.50, 0.25, 0.10, 0.05, 0.01}, Size of the domain=6.

- <u>The domain of a random variable can be "countably infinite</u>."

Example: X = number of words in the next Game of Thrones novel. No matter how large you guess, it's possible it might be even longer, so we say the domain is infinite.

Requirement: $1 = \text{sum } P(X=x)$

- <u>The domain of a random variable can be "uncountably infinite."</u>

Example: a variable whose value can be ANY REAL NUMBER.

How we deal with this: $P(X=x)$ is ill-defined, but $P(a \le X < b)$ is well-defined.

Expectation and PMF

- Expected value can be calculated from PMF only if the domain is finite, or countably infinite.

$$E[X] = \text{sum value} * P(X=\text{value})$$

Example: X = number of words in the next GoT novel.

$$E[X] = P(X=1) + 2*P(X=2) + 3*P(X=3) + \ldots$$

If you know P(X=x) for all x (even if “all x” is an infinite set), then you can compute this expectation by solving the infinite series.

Random Variables

- Random Variables
  - RV = function from outcomes to numbers
  - Notation
  - Probability Mass Function (pmf)
  - Expected Value
- Domain of a Random Variable
  - Domain Type: Categorical vs. Numerical
  - Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite
- Joint, Marginal, and Conditional Random Variables
  - Marginalization and Conditioning
  - Law of Total Probability
  - Random Vectors
  - Jointly Random Class and Measurement Variables
- Functions of Random Variables
  - Probability Mass Function
  - Expectation

Joint probability mass function (joint pmf)

- $p(X_1=x_1,X_2=x_2,\ldots,X_N=x_N)$ refers to the probability of a particular outcome (the outcome $X_1=x_1,\ldots,X_N=x_N$).
  - Shorthand: $p(x_1, x_2, \ldots, x_N)$
  - Subscript notation, which we won't use in this class: $\boldsymbol{p_{X_1,\ldots,X_N}(x_1, \ldots, x_N)}$
- $p(X_1, X_2, \ldots, X_N)$ refers to the entire joint probability mass function, i.e., the entire table, listing all possible outcomes, and the probability of each
- $P(A)$ (capital P) refers to the probability of an event

Joint Random Variables

- For example, suppose W = pips showing on the red die, X = pips on purple die, Y = green, Z = blue.
- The following table shows $p(W,X,Y,Z)$, their joint pmf.

w	x	y	z	P(W=w,X=x,Y=y,Z=z)
1	1	1	1	1/1296
1	1	1	2	1/1296
…		…		…
6	6	6	4	1/1296
6	6	6	5	1/1296
6	6	6	6	1/1296

Marginalization

$$P(X = x) = \sum_{w} \sum_{y} \sum_{z} P(W = w, X = x, Y = y, Z = z)$$

Example: if W, X, Y, Z are four independent dice, then the marginal is just what you would expect:

$$P(X = x) = \sum_{w=1}^{6} \sum_{y=1}^{6} \sum_{z=1}^{6} \left(\frac{1}{1296}\right) = \frac{1}{6}$$

Conditioning

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Example: if W, X, Y, Z are four independent dice, then the marginal is just what you would expect:

$$P(X = 3|Z = 3) = \frac{P(X = 3, Z = 3)}{P(Z = 3)} = \frac{1/36}{1/6} = \frac{1}{6}$$

Conditioning

Here’s a surprise. One of the most useful things you can do with a conditional probability is to turn it around, to calculate the joint pmf:

$$P(X = x, Y = y) = P(X = x | Y = y) P(Y = y)$$

Conditioning+Marginalization

Here's a surprise. One of the most useful things you can do with a conditional probability is to turn it around, to calculate the joint pmf:

$$P(X = x, Y = y) = P(X = x|Y = y)P(Y = y)$$

Remember the law for marginalization:

$$P(X = x) = \sum_{y} P(X = x, Y = y)$$

Conditioning+Marginalization = Law of Total Probability

Here's a surprise. One of the most useful things you can do with a conditional probability is to turn it around, to calculate the joint pmf:

$$P(X = x, Y = y) = P(X = x|Y = y)P(Y = y)$$

Remember the law for marginalization:

$$P(X = x) = \sum_{y} P(X = x, Y = y)$$

Putting those two things together:

$$P(X = x) = \sum_{y} P(X = x|Y = y)P(Y = y)$$

Law of Total Probability

This is called the "Law of Total Probability:"

$$P(X = x) = \sum_{y} P(X = x | Y = y) P(Y = y)$$

Law of Total Probability

Example:

- Billy Bones said that there is treasure in a treasure chest on this island.
- What is $P(TreasureChest = full)$?
- Two possibilities:

  1. Bones lied.

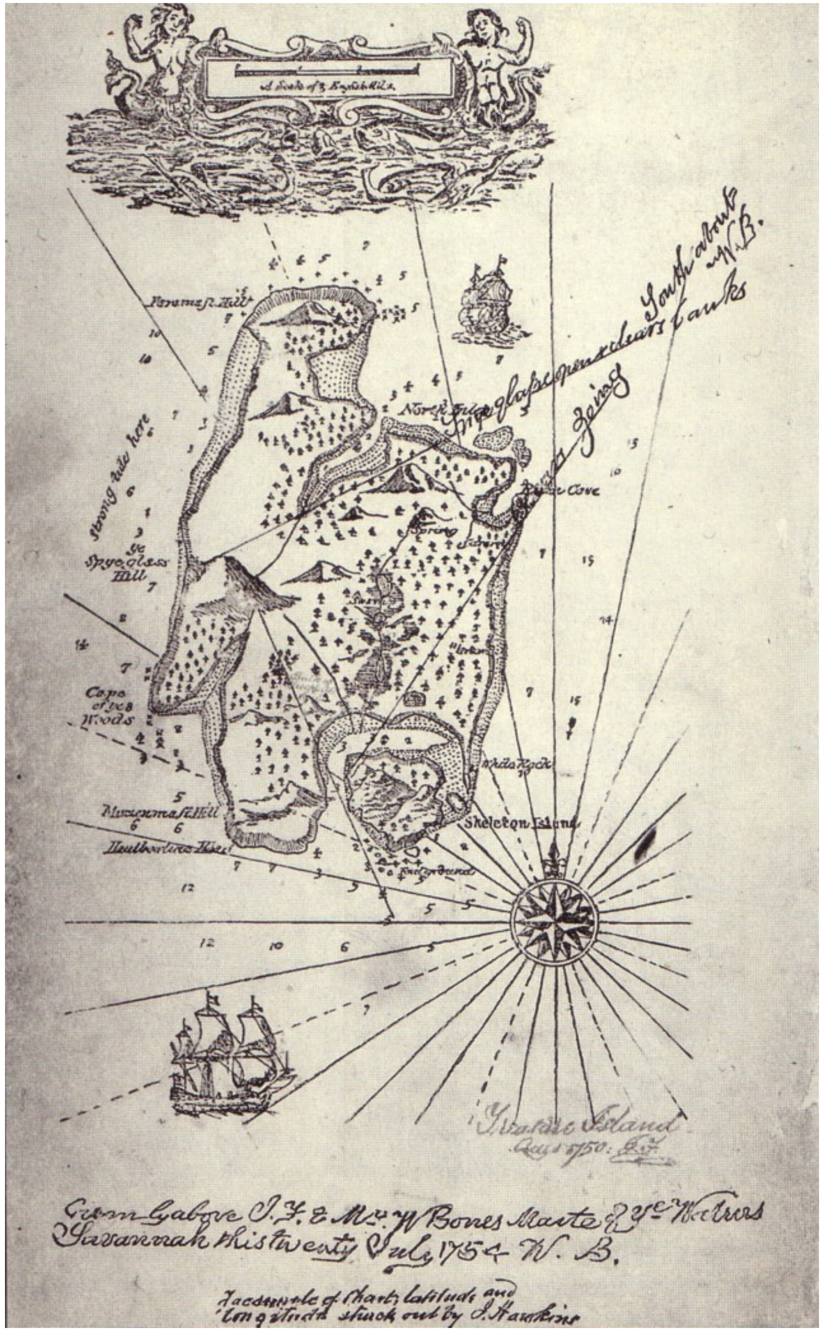
     $$P(TreasureChest = full \mid Bones\ lied) = 0.0$$

  2. Bones told the truth.

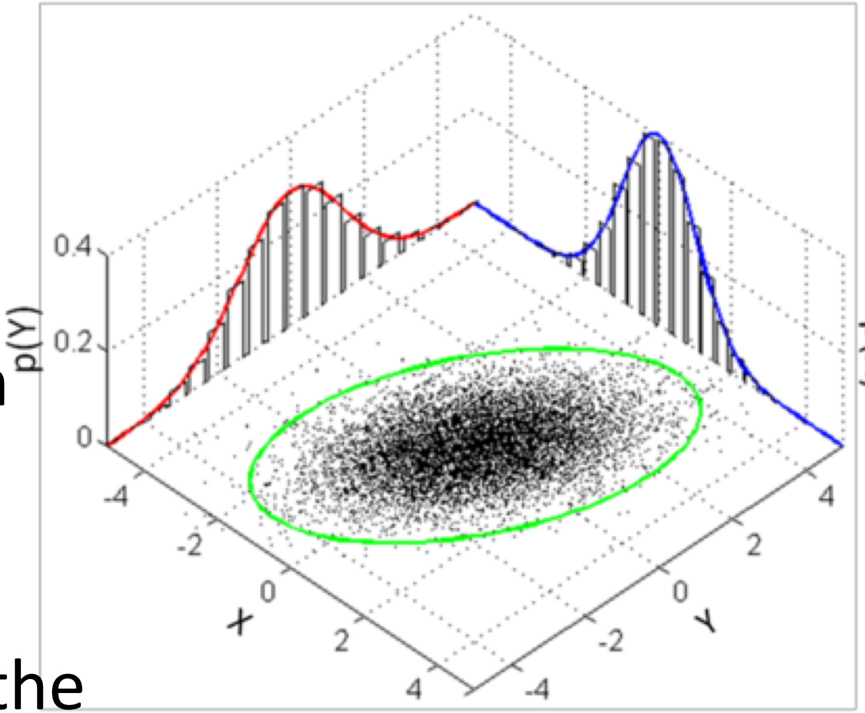
     $$P(TreasureChest = full \mid Bones\ told\ truth) = 0.7$$

- Law of Total Probability:

$$P(TreasureChest = full)$$
$$= 0.0 \times P(Bones\ lied) + 0.7 \times P(Bones\ true)$$



Random Vector



A Random Vector, $\vec{X}$, is a vector of joint random variables $\vec{X} = [X_1, X_2, \ldots, X_n]$.

The pmf of the random vector is defined to be the Joint pmf of all of its component variables:

$$P\left(\vec{X} = \vec{x}\right) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Jointly Random Class and Measurement Variables

The most important case of joint random variables for AI:
jointly random categorical (class) and numerical (measurement) variables.

For example, Y= type of fruit, X = weight of the fruit.

x	y	P(X=x,Y=y)
10g	Grape	0.68
10g	Apple	0.06
100g	Grape	0.02
100g	Apple	0.34

We'll talk A LOT more about this in a few lectures (Bayesian inference).

Random Variables

- Random Variables
  - RV = function from outcomes to numbers
  - Notation
  - Probability Mass Function (pmf)
  - Expected Value
- Domain of a Random Variable
  - Domain Type: Categorical vs. Numerical
  - Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite
- Joint, Marginal, and Conditional Random Variables
  - Marginalization and Conditioning
  - Law of Total Probability
  - Random Vectors
  - Jointly Random Class and Measurement Variables
- Functions of Random Variables
  - Probability Mass Function
  - Expectation

Functions of Random Variables: PMF

The PMF for a function of random variables is computed the same way as any other marginal: by adding up the component probabilities.

Example: $S = W + X + Y + Z$

w	x	y	z	s	P(W=w,X=x,Y=y,Z=z,S=s)
1	1	1	1	4	1/1296
1	1	1	2	5	1/1296
1	1	2	1	5	1/1296
…	…	…	…	…	…

Functions of Random Variables: PMF

w	x	y	z	s	P(W=w,X=x,Y=y,Z=z,S=s)
1	1	1	1	4	1/1296
1	1	1	2	5	1/1296
1	1	2	1	5	1/1296
…	…	…	…	…	…

- There is only one outcome for which S=4, so

$$P(S=4)=\frac{1}{1296}$$

- There are four different outcomes for which S=5, so

$$P(S=5)=\frac{1}{1296}+\frac{1}{1296}+\frac{1}{1296}+\frac{1}{1296}=\frac{4}{1296}$$

Functions of Random Variables: Expectation

It's important to know that, for any function g(X), $E[g(X)] \neq g(E[X])$

$$E[g(X)] = \sum_{y} y * P(g(X) = y)$$

$$g(E[X]) = g\left(\sum_{x} x * P(X = x)\right)$$

Those are not the same thing!!

Functions of Random Variables: Expectation

Example: $E[X^2] \neq E[X]^2$

$$E[X^2] = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + \cdots + 6^2\left(\frac{1}{6}\right) = 15.1667$$

$$E[X]^2 = \left(1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \cdots + 6\left(\frac{1}{6}\right)\right)^2 = 12.25$$

Those are not the same thing!!

Random Variables

- Random Variables
  - RV = function from outcomes to numbers
  - Notation
  - Probability Mass Function (pmf)
  - Expected Value
- Domain of a Random Variable
  - Domain Type: Categorical vs. Numerical
  - Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite
- Joint, Marginal, and Conditional Random Variables
  - Marginalization and Conditioning
  - Law of Total Probability
  - Random Vectors
  - Jointly Random Class and Measurement Variables
- Functions of Random Variables
  - Probability Mass Function
  - Expectation

"ABOUT THIS EXPERIMENT FOR GENERATING RANDOM NUMBERS – EACH TIME YOU DO IT, IT COMES OUT DIFFERENT."