



浙江大学伊利诺伊大学厄巴纳香槟校区联合学院
Zhejiang University-University of Illinois at Urbana Champaign Institute

ECE 448: AI theory and Application

Lecture 11: Random Variables

Dr. Hongwei Wang hongweiwang@intl.zju.edu.cn

Prof. Mark Hasegawa-Johnson jhasegaw@illinois.edu

Spring 2023

1. Random Variables

- RV = function from outcomes to numbers
- Notation
- Probability Mass Function (pmf)
- Expected Value

2. Domain of a Random Variable

- Domain Type: Categorical vs. Numerical
- Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite

3. Joint, Marginal, and Conditional Random Variables

- Marginalization and Conditioning
- Law of Total Probability
- Random Vectors
- Jointly Random Class and Measurement Variables

4. Functions of Random Variables

- Probability Mass Function
- Expectation

1. Random Variables

- RV = function from outcomes to numbers
- Notation
- Probability Mass Function (pmf)
- Expected Value

2. Domain of a Random Variable

- Domain Type: Categorical vs. Numerical
- Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite

3. Joint, Marginal, and Conditional Random Variables

- Marginalization and Conditioning
- Law of Total Probability
- Random Vectors
- Jointly Random Class and Measurement Variables

4. Functions of Random Variables

- Probability Mass Function
- Expectation

- 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]

- A random variable can be viewed as a function, $f: \text{outcomes} \rightarrow \text{numbers}$ (a function that maps outcomes to numbers).
- For example: 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 towtruck is traveling 45mph
 - My car is falling off a cliff, and has reached a terminal velocity of 45mph...

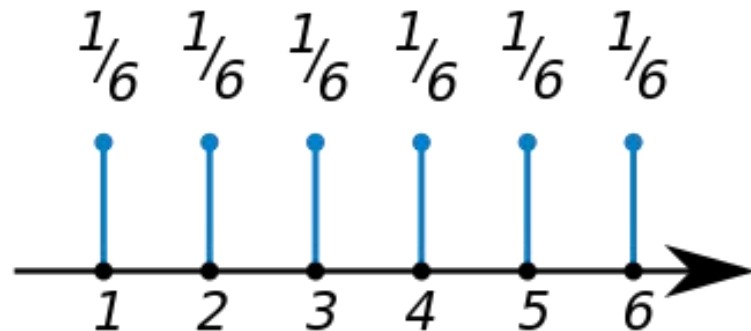
We use an **UPPERCASE** letter for a random variable, and a **lowercase** letter for the actual value that it takes after any particular experiment.

- $X_1 = x_1$ is the event “random variable X_1 takes the value x_1 ”
- X_1 is an RV, which is a **function**, X_1 :outcomes \rightarrow 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).

- $X_1 = x_1$ is the event “random variable X_1 takes the value x_1 ”
- $p(X_1 = x_1)$ is a **number**: 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: $p_{X_1}(x_1)$
- $p(X_1)$ using a capital letter X_1 is a **function**, specifically, this is the **probability mass function (pmf)**: the entire table of the probabilities $X_1 = x_1$ for every possible x_1

- The “Probability Mass Function” (pmf) of a random variable X is defined to be the function $P(X=\text{value})$, 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.”

Axioms of Probability

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

Requirements for a pmf

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.

1. Random Variables

- RV = function from outcomes to numbers
- Notation
- Probability Mass Function (pmf)
- Expected Value

2. Domain of a Random Variable

- Domain Type: Categorical vs. Numerical
- Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite

3. Joint, Marginal, and Conditional Random Variables

- Marginalization and Conditioning
- Law of Total Probability
- Random Vectors
- Jointly Random Class and Measurement Variables

4. Functions of Random Variables

- Probability Mass Function
- Expectation

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



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

$$\begin{aligned} E[D] &= \lim_{n \rightarrow \infty} \frac{1}{n} \left(1 \times (\# \text{ times } D = 1) + \dots + 6 \times (\# \text{ times } D = 6) \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \left(1 \times (n/6) + \dots + 6 \times (n/6) \right) = 3.5 \end{aligned}$$

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

$$= \text{sum}\{ \text{value} * P(\text{variable}=\text{value}) \}$$

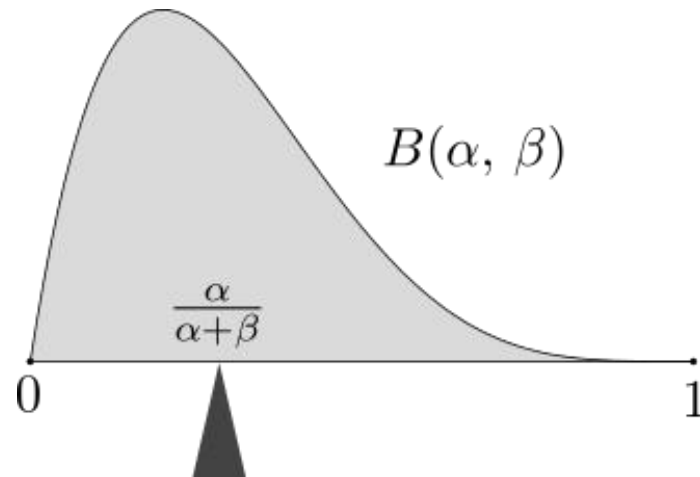
Center of Mass

$$= \text{sum}\{ \text{position} * \text{Mass}(\text{position}) \}$$



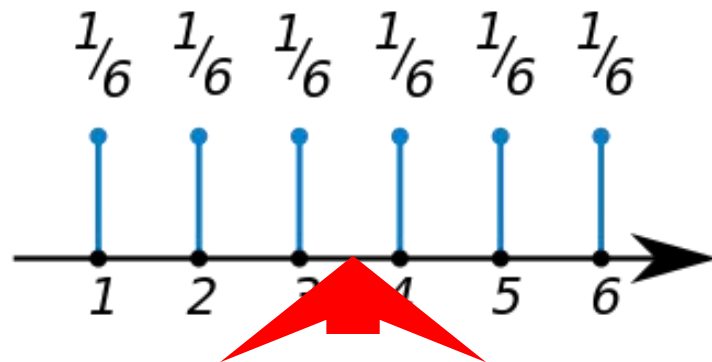
Expected Value of a random variable = the average value, averaged over an infinite number of independent trials

$$= \text{sum}\{ \text{value} * P(\text{variable}=\text{value}) \}$$



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

- The “Probability Mass Function” (pmf) of a random variable X is defined to be the function $P(X=\text{value})$, 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.

1. Random Variables

- RV = function from outcomes to numbers
- Notation
- Probability Mass Function (pmf)
- Expected Value

2. Domain of a Random Variable

- Domain Type: Categorical vs. Numerical
- Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite

3. Joint, Marginal, and Conditional Random Variables

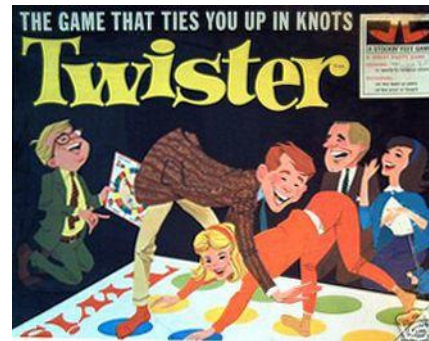
- Marginalization and Conditioning
- Law of Total Probability
- Random Vectors
- Jointly Random Class and Measurement Variables

4. Functions of Random Variables

- Probability Mass Function
- Expectation

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 }



- Expected Value is only well defined for numerical domains.

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

- pmf is well defined even for categorical domains.

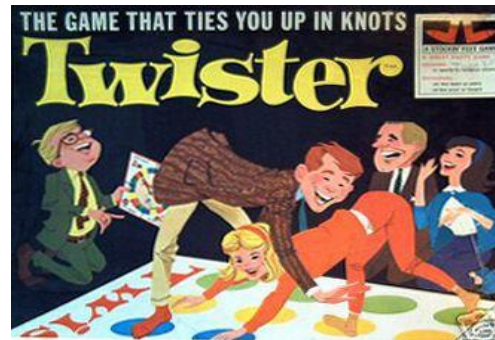
Example: X = color shown on the spinner

$$P(X=\text{red}) = (1/4)$$

$$P(X=\text{blue}) = (1/4)$$

$$P(X=\text{green}) = (1/4)$$

$$P(X=\text{yellow}) = (1/4)$$



- Domain of a random variable can be finite.

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.

- Domain of a random variable can be “countably infinite.”

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: $1 = \sum P(X=x)$

- Domain of a random variable can be “uncountably infinite.”

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

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

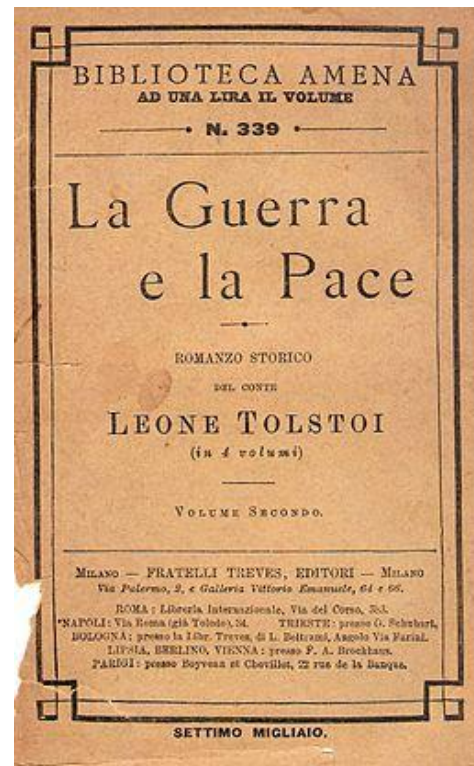
- Expected value can be calculated from PMF only if the domain is finite, or countably infinite.

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

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

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

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.



1. Random Variables

- RV = function from outcomes to numbers
- Notation
- Probability Mass Function (pmf)
- Expected Value

2. Domain of a Random Variable

- Domain Type: Categorical vs. Numerical
- Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite

3. Joint, Marginal, and Conditional Random Variables

- Marginalization and Conditioning
- Law of Total Probability
- Random Vectors
- Jointly Random Class and Measurement Variables

4. Functions of Random Variables

- Probability Mass Function
- Expectation

- $p(X_1=x_1, X_2=x_2, \dots, X_N=x_N)$ refers to the probability of a particular outcome (the outcome $X_1=x_1, \dots, X_N=x_N$).
 - Shorthand: $p(x_1, x_2, \dots, x_N)$
 - Subscript notation, which we won't use in this class:
 $p_{X_1, \dots, X_N}(x_1, \dots, x_N)$
- $p(X_1, X_2, \dots, X_N)$ 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

- 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

$$P(X = x) = \sum_w \sum_y \sum_z P(W = w, X = x, Y = y, Z = z)$$

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

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

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

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

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

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:

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

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:

$$P(X = x) = \sum_y P(X = x, Y = y)$$

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:

$$P(X = x) = \sum_y P(X = x|Y = y)P(Y = y)$$

This is called the “Law of Total Probability:”

$$P(X = x) = \sum_y P(X = x|Y = y)P(Y = y)$$

Example:

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

1. Bones lied.

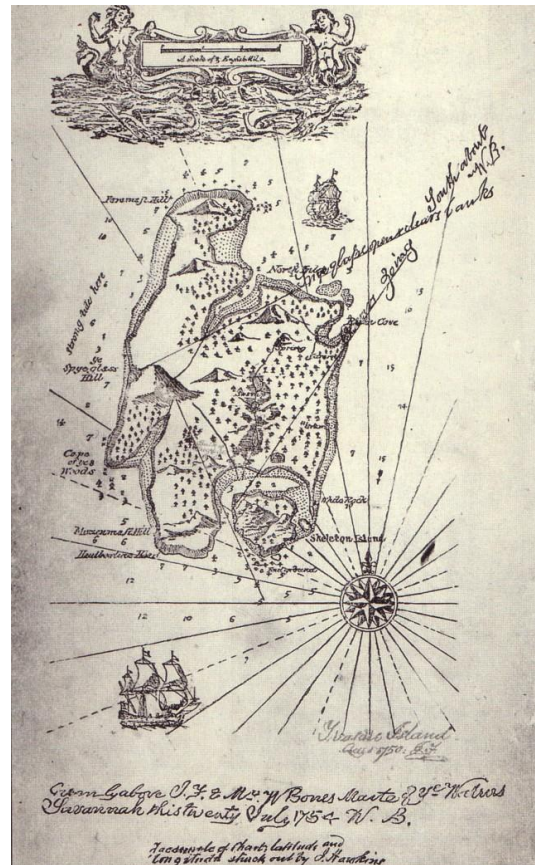
$$P(\text{TreasureChest} = \text{full} | \text{Bones lied}) = 0.0$$

2. Bones told the truth.

$$P(\text{TreasureChest} = \text{full} | \text{Bones told truth}) = 0.7$$

- Law of Total Probability:

$$\begin{aligned} P(\text{TreasureChest} = \text{full}) \\ = 0.0 \times P(\text{Bones lied}) + 0.7 \times P(\text{Bones true}) \end{aligned}$$



1. Random Variables

- RV = function from outcomes to numbers
- Notation
- Probability Mass Function (pmf)
- Expected Value

2. Domain of a Random Variable

- Domain Type: Categorical vs. Numerical
- Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite

3. Joint, Marginal, and Conditional Random Variables

- Marginalization and Conditioning
- Law of Total Probability
- Random Vectors
- Jointly Random Class and Measurement Variables

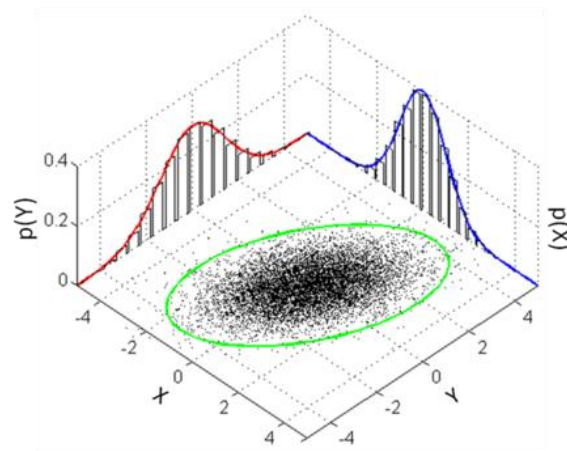
4. Functions of Random Variables

- Probability Mass Function
- Expectation

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

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

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



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).

1. Random Variables

- RV = function from outcomes to numbers
- Notation
- Probability Mass Function (pmf)
- Expected Value

2. Domain of a Random Variable

- Domain Type: Categorical vs. Numerical
- Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite

3. Joint, Marginal, and Conditional Random Variables

- Marginalization and Conditioning
- Law of Total Probability
- Random Vectors
- Jointly Random Class and Measurement Variables

4. Functions of Random Variables

- Probability Mass Function
- Expectation

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
...

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

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

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

$$E[g(X)] = \sum_g g * P(g(X) = g)$$

$$g(E[X]) = g\left(\sum_x x * P(X = x)\right)$$

Those are not the same thing!!

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

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

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

Those are not the same thing!!

1. Random Variables

- RV = function from outcomes to numbers
- Notation
- Probability Mass Function (pmf)
- Expected Value

2. Domain of a Random Variable

- Domain Type: Categorical vs. Numerical
- Domain Size: Finite vs. Countably Infinite vs. Uncountably Infinite

3. Joint, Marginal, and Conditional Random Variables

- Marginalization and Conditioning
- Law of Total Probability
- Random Vectors
- Jointly Random Class and Measurement Variables

4. Functions of Random Variables

- Probability Mass Function
- Expectation



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