Probability Theory

August 6, 2021

Acknowledgment of Sources

Slides based on content from:

Introduction To Probability, 2nd edition, by Dimitri P. Bertsekas and John N. Tsitsiklis

Prof. Mitesh’s course on LARP

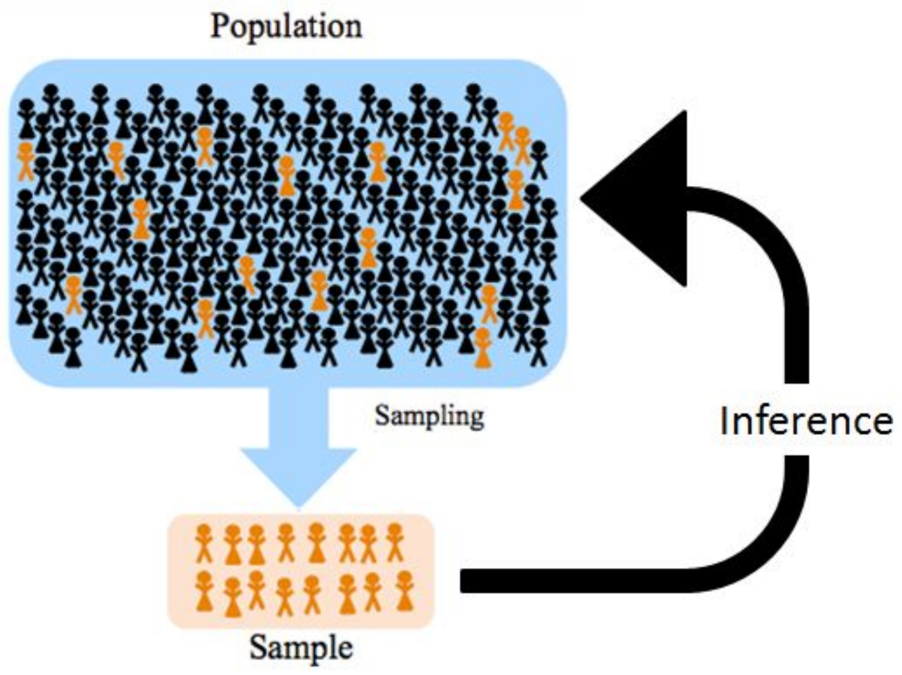
Google Images

Wikipedia

Learning objectives

- Why do we need probability theory in PRML class?
- Probability basics:
  - Counting
  - Probability axioms
  - Conditional probability
  - Multiplication rule
  - Bayes’ theorem
  - Random variable
  - Probability Distributions
- Pointers to the topics not covered today

Why do we need Probability Theory?



Statistics (from sample)

- Mean sugar level
- Variance in fertility rate
- Mean height or weight

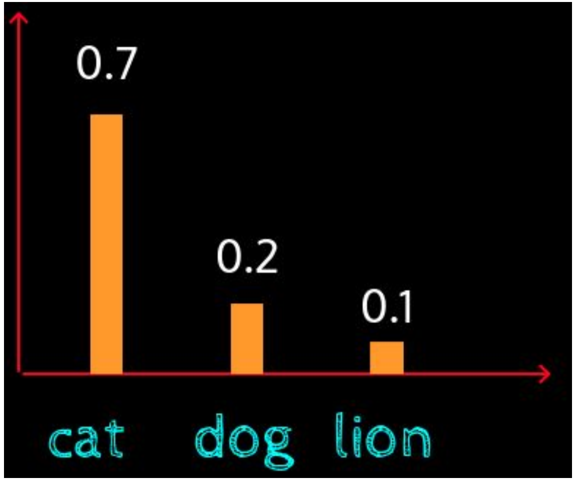
Q: What is the probability that a statistic computed from a sample is close to that computed from a population?

Why do we need Probability Theory?

Cat? Dog? Owl? Lion?

Machine Learning

$$P(label = cat|image)$$



Predict a distribution over class

A simple example

- What is the probability of getting a heads?

    ½ or 0.5 or 50%

- How did you compute this?

    1/n

2 possible outcomes: each equally likely

Another simple example

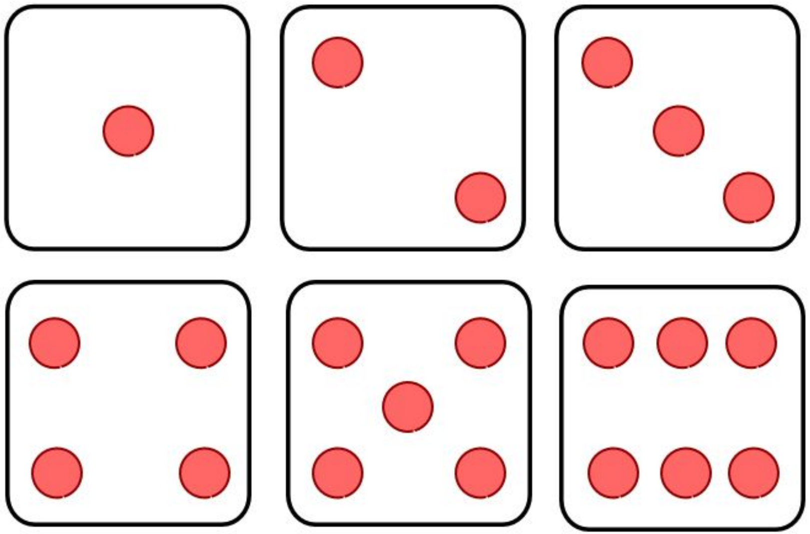
- What is the probability of getting a five?

  $\frac{1}{6}$ or 0.1667 or 16.67%

- How did you compute this?

  1/n

6 possible outcomes: each equally likely

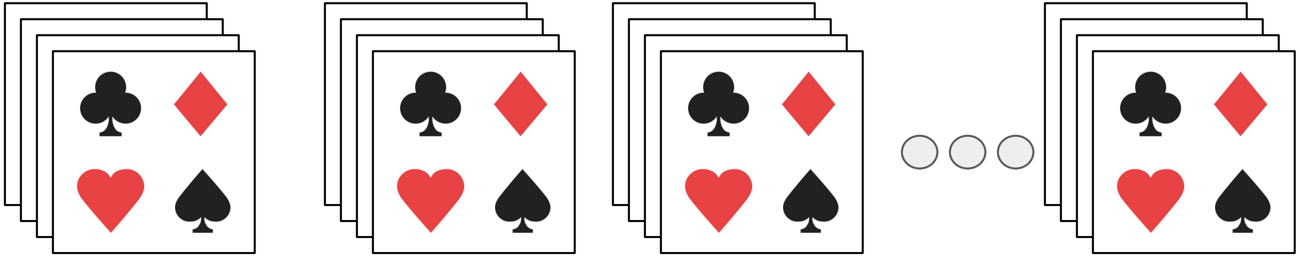


Another example

- What is the probability of getting 4 aces?

  1/n

- But what is n?



n is the number of possible outcomes, i.e., all possible combinations of 4 cards

How do you count n?

[1.6] Introduction To Probability, 2nd edition, by Dimitri P. Bertsekas and John N. Tsitsiklis

Definitions alert!

Ω: Sample space (all outcomes)

Event: subsets of Ω

What is the chance of an event?

Goal: Assign a number to each event such that this number reflects the chance of the experiment resulting that event

Probability

$$P(A) = ?$$

Probability Function

Event

What are the conditions that such a probability function must satisfy?

(Axioms of Probability)

- Axiom 1 (non-negativity)

$$P(A) \geq 0 \forall A$$

- Axiom 2 (normalisation)

$$P(\Omega) = 1$$

- Axiom 3 (finite additivity)

$$P(A_1 \cup A_2 \cup \ldots A_n) = \Sigma_{i=1}^{n} P(A_i)$$

Properties of a Probability Function: [1.2] Introduction To Probability Book

Change in belief

Before start of play: What is the chance of India winning?

0.5

India scores 395 batting first: What is the chance of India winning?

> 0.5

What exactly happened?

Change in belief

A: event that India will win

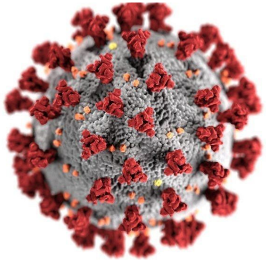
B: India scores 395 runs

$P(A)$ changes once we know that event B has occurred

$$P(A|B) \neq P(A)$$

10% population is infected

Question: What is the probability that a randomly selected person is healthy (not infected)?



A: event that person is healthy $P(A) = 0.9$

B: event that the person has COVID-19 symptoms

The definition of $P(A|B)$

(1 , 1)	**(1 , 2)**	**(1 , 3)**	**(1 , 4)**	**(1 , 5)**	**(1 , 6)**
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

What is the probability that the sum is 8?

$P(A) = 5/36$

(1 , 1)	**(1 , 2)**	**(1 , 3)**	**(1 , 4)**	**(1 , 5)**	**(1 , 6)**
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

The definition of $P(A|B)$

What is the probability that the sum is 8 given that the first dice shows a 4?

A: sum is 8

B: first dice shows a 4

$P(A|B) = 1/6$

(1 , 1)	**(1 , 2)**	**(1 , 3)**	**(1 , 4)**	**(1 , 5)**	**(1 , 6)**
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

(1 , 1)	**(1 , 2)**	**(1 , 3)**	**(1 , 4)**	**(1 , 5)**	**(1 , 6)**
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Conditional Probability

To do: **Do conditional probabilities satisfy the axioms of probability?**

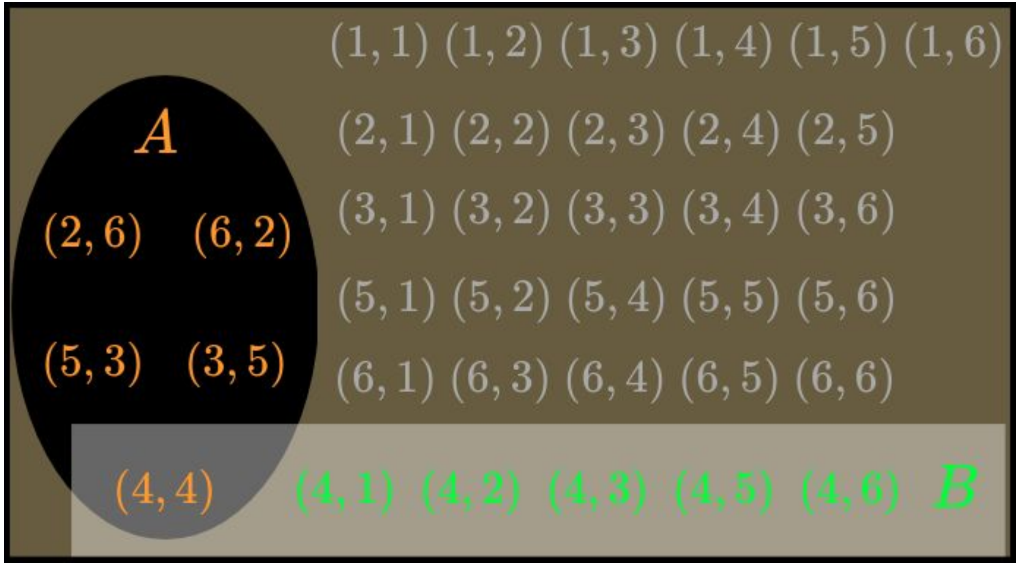
A: sum is 8

B: first dice shows a 4

$P(B) = 6/36$

$P(A \cap B) = 1/36$

$P(A|B) = P(A \cap B) \,/\, P(B)$



Conditional Probability

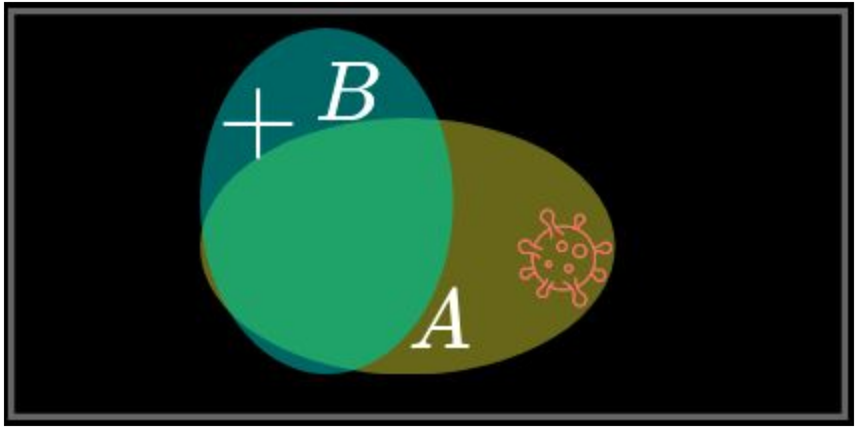
Regular Probabilities

The multiplication principle

The chain rule of probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$\therefore P(A \cap B) = P(A|B).P(B)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$
$$\therefore P(B \cap A) = P(B|A).P(A)$$



A: event that person is infected

B: event that the test result is +ve

$$\therefore P(A \cap B) = P(A|B).P(B) = P(B|A).P(A)$$

The chain rule of probability (for n events)

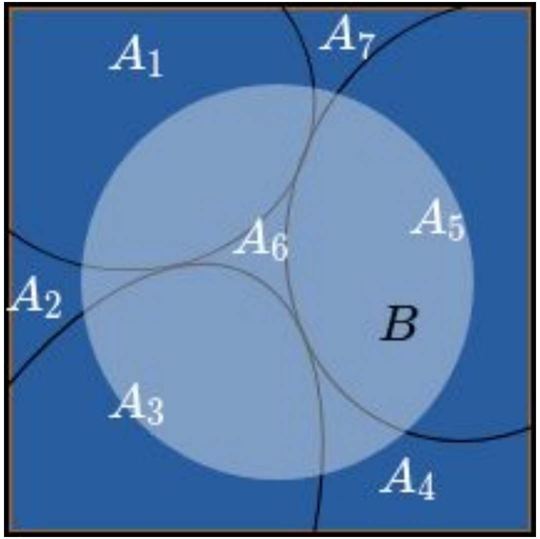
$$P(A \cap B \cap C) = P((A \cap B) \cap C)$$
$$Let (A \cap B) = X$$
$$\therefore P(A \cap B \cap C) = P(X \cap C)$$
$$\therefore P(A \cap B \cap C) = P(X).P(C|X)$$
$$\therefore P(A \cap B \cap C) = P(A \cap B).P(C|A \cap B)$$
$$\therefore P(A \cap B \cap C) = P(A).P(B|A).P(C|A \cap B)$$

for n events

$$P(A \cap B \cap C \cap D) = P(A).P(B|A).P(C|A \cap B).P(D|A \cap B \cap C)$$

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \prod_{i=2}^{n} P(A_i|A_1, \ldots, A_{i-1})$$

Total Probability Theorem



$$A_1, A_2, \ldots A_n \text{ partition } \Omega$$
$$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega \text{ and } A_i \cap A_j = \phi \forall i \neq j$$

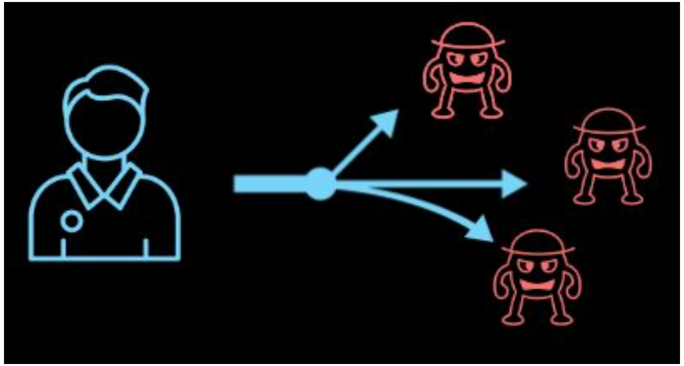
$$B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n)$$
$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n)$$
$$P(B) = P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + \cdots + P(A_n).P(B|A_n)$$

$$P(B) = \sum_{i=1}^{n} P(A_i).P(B|A_i)$$

Example

$P(B^c) = ?$

$$= P(A_1)P(B^c|A_1) + P(A_2)P(B^c|A_2) + P(A_3)P(B^c|A_3)$$
$$= \frac{1}{3} \times 0.7 + \frac{1}{3} \times 0.4 + \frac{1}{3} \times 0.25$$



A_i : i-th path taken
B : monster encountered

Can we find $P(A_1|B)$?

$$P(B|A_1) = 0.3$$
$$P(B|A_2) = 0.6$$
$$P(B|A_3) = 0.75$$

Breaking it down

$$P(A_1|B) = ?$$

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)}$$

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + P(A_3).P(B|A_3)}$$

Total Probability Theorem

$$P(A_1|B) = \frac{P(A_1).P(B|A_1)}{P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + P(A_3).P(B|A_3)}$$

Multiplication rule

$$P(A_1|B) = \frac{P(A_1).P(B|A_1)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Bayes' Theorem

Independence and more..

Probability of Compound Events

Independent Events

$$P(A \text{ and } B) = P(A) \times P(B)$$

Dependent Events

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$

Mutually Exclusive

$$P(A \text{ or } B) = P(A) + P(B)$$

Mutually Inclusive

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Conditional Independence: $P(A|B,C) = P(A|C)$

https://www.onlinemathlearning.com/compound-probability-7sp8a.html

Random Variable

Mapping outcomes to R

Focus on numerical quantities associated with the outcomes of experiments.

In board games, we care about the sums and not the numbers that led to the sum.

$(1,1)$

$(1,2)\ (2,1)$

$(1,3)\ (2,2)\ (3,1)$

$(1,4)\ (2,3)\ (3,2)\ (4,1)$

$(1,5)\ (2,4)\ (3,3)\ (4,2)\ (5,1)$

$(1,6)\ (2,5)\ (3,4)\ (4,3)\ (5,2)\ (6,1)$

$(2,6)\ (3,5)\ (4,4)\ (5,3)\ (6,2)$

$(3,6)\ (4,5)\ (5,4)\ (6,3)$

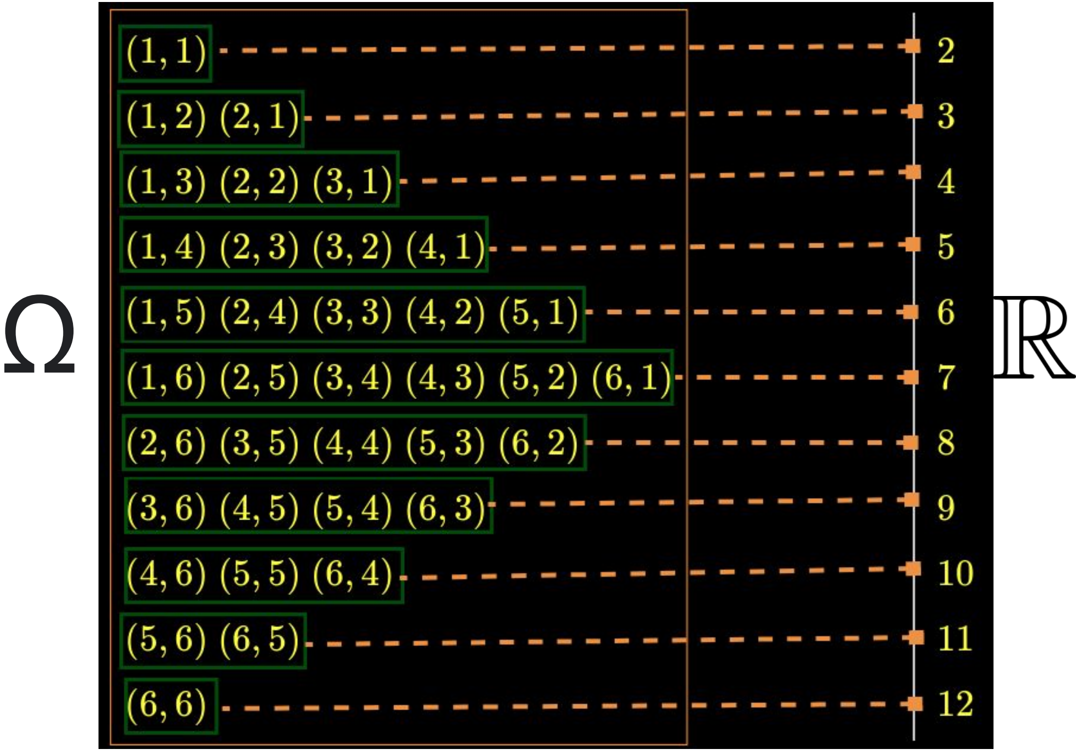
$(4,6)\ (5,5)\ (6,4)$

$(5,6)\ (6,5)$

$(6,6)$

Mapping outcomes to R

Question of interest: What is the probability that the sum will be 10?



Mapping outcomes to R

Experiment: Randomly select an employee

Ω : All employees of the organisation

ℝ : Number of years of experience, number of projects, salary, income tax, num. children

Qs of Interest:

What is the probability that an employee has 2 children?

What is the probability that an employee's monthly salary is greater than 50K

What are the values a random variable can take?

- Discrete
- Continuous

What are the probabilities of the values that a discrete random variable can take?

Probability Mass Function (PMF)

Mapping outcomes to R

What is the probability that the value of the random variable will be x?

$$X : \Omega \to \mathbb{R}$$

$P(X = x) = [0,1]$

x	$P(X = x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

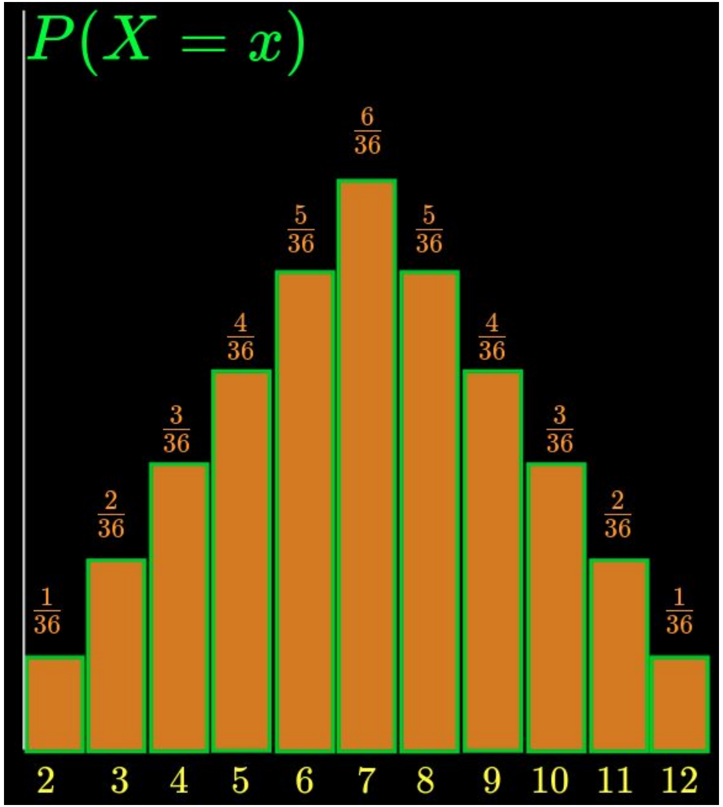
$Event : X = x$	x	$P(X = x)$
$(1,1)$	2	$\frac{1}{36}$
$(1,2)\ (2,1)$	3	$\frac{2}{36}$
$(1,3)\ (2,2)\ (3,1)$	4	$\frac{3}{36}$
$(1,4)\ (2,3)\ (3,2)\ (4,1)$	5	$\frac{4}{36}$
$(1,5)\ (2,4)\ (3,3)\ (4,2)\ (5,1)$	6	$\frac{5}{36}$
$(1,6)\ (2,5)\ (3,4)\ (4,3)\ (5,2)\ (6,1)$	7	$\frac{6}{36}$
$(2,6)\ (3,5)\ (4,4)\ (5,3)\ (6,2)$	8	$\frac{5}{36}$
$(3,6)\ (4,5)\ (5,4)\ (6,3)$	9	$\frac{4}{36}$
$(4,6)\ (5,5)\ (6,4)$	10	$\frac{3}{36}$
$(5,6)\ (6,5)$	11	$\frac{2}{36}$
$(6,6)$	12	$\frac{1}{36}$

Probability Mass Function

Think of the event corresponding to $X = x$

Once we know this event (subset of sample space) we know how to compute $P(X=x)$

PMF: $p_X(x) = P(X = x)$



Properties of PMF: [2.1 and 2.2] Introduction To Probability Book

How can we describe distributions compactly?

An assignment of probabilities to all possible values that a discrete RV take can be tedious.

Can PMF be specified compactly?

X: random variable indicating the number of tosses after which you observe the first heads

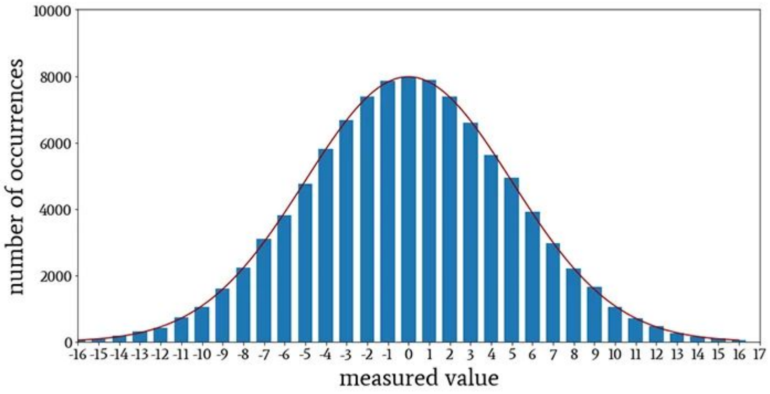
$$p^x * (1 - p)^{1-x}$$

$$p_X(x) = \begin{cases} .. & if\ x = 1 \\ .. & if\ x = 2 \\ .. & if\ x = 3 \\ .. & if\ x = 4 \\ .. & if\ x = 5 \\ .. & if\ x = 6 \\ .. & .. \\ .. & .. \\ .. & if\ x = \infty \end{cases}$$

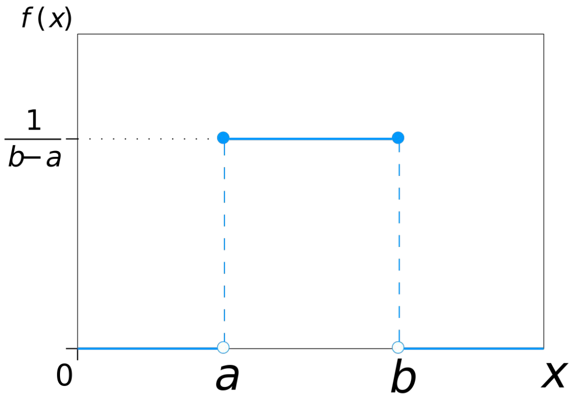
Probability distributions

Bernoulli Distribution (experiments with only two outcomes)

Binomial Distribution (Repeat a Bernoulli trial n times)



Normal distribution



Uniform distribution

Probability Distributions: [2.2] Introduction To Probability Book

More on Random Variables

Expectation of a RV: $$E[X] = \sum_{\text{all possible } x} xP(X = x)$$

Variance of a RV: $$\sigma_X^2 = E[(X - \mu_X)^2]$$

Exercise: Find the expectation and variance of the distributions discussed in the previous slide.

Other topics

Markov inequality

Chebyshev inequality

Law of large numbers

Central Limit Theorem

Cross Entropy

KL Divergence