Introduction to Probability

Dr. Ganesh Ghalme

Nov 14-18

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Introduction

What is the probability of getting a head?

Introduction

What is the probability of getting a head?

Answer: we don't know

Introduction

What is the probability of getting a head?

Answer: we don't know

We assume it is 1/2 (for a fair coin)

Throwing Dice

- Dice has 6 faces
- We assume each face is equally likely to show at top; we ignore the role played by uneven edges etc.

Throwing Dice

- Dice has 6 faces
- We assume each face is equally likely to show at top; we ignore the role played by uneven edges etc.

Playing cards

- Total cards 52

Throwing Dice

- Dice has 6 faces
- We assume each face is equally likely to show at top; we ignore the role played by uneven edges etc.

Playing cards

- Total cards 52
- 4 suites; clubs, hearts, spades, diamonds (13 cards each)

Throwing Dice

- Dice has 6 faces
- We assume each face is equally likely to show at top; we ignore the role played by uneven edges etc.

Playing cards

- Total cards 52
- 4 suites; clubs, hearts, spades, diamonds (13 cards each)
- Face cards; 4 in each suit

Throwing Dice

- Dice has 6 faces
- We assume each face is equally likely to show at top; we ignore the role played by uneven edges etc.

Playing cards

- Total cards 52
- 4 suites; clubs, hearts, spades, diamonds (13 cards each)
- Face cards; 4 in each suit
- Draw two cards from the pack. What is the probability of getting

Throwing Dice

- Dice has 6 faces
- We assume each face is equally likely to show at top; we ignore the role played by uneven edges etc.

Playing cards

- Total cards 52
- 4 suites; clubs, hearts, spades, diamonds (13 cards each)
- Face cards; 4 in each suit
- Draw two cards from the pack. What is the probability of getting
  - both cards of same suit

Throwing Dice

- Dice has 6 faces
- We assume each face is equally likely to show at top; we ignore the role played by uneven edges etc.

Playing cards

- Total cards 52
- 4 suites; clubs, hearts, spades, diamonds (13 cards each)
- Face cards; 4 in each suit
- Draw two cards from the pack. What is the probability of getting
  - both cards of same suit
  - both face cards.

Preliminaries: Sets I

A set is a collection of objects called elements.

Examples

- $S = \{H, T\}$
- Power set $2^S = \{\emptyset, H, T, \{H, T\}\}$
- Is this a set $\{H, H\}$?
- Cardinality of a set $|S|$; number of elements in a set
- What is the cardinality of the set of all natural number? What about rational numbers and irrational numbers?
- Suppose a coin is tossed twice. What is the set of all possible outcomes?

Preliminaries: Sets II

Set Operations

- $A \cup B$ (union), $A \cap B$ (intersection), $\overline{A}$ or A^c (complement)
- Union operation is commutative and associative; $A \cup B = B \cup A$ and $(A \cup B) \cup C = A \cup (B \cup C)$
- If $A \subseteq B$ then $A \cup B = B$
- Intresection is commutative, associative and also distributive $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- If $A \subseteq B$ then $A \cap B = A$
- DeMorgan's law
  - $\overline{A \cup B} = \overline{A} \cap \overline{B}$
  - $\overline{A \cap B} = \overline{A} \cup \overline{B}$
  - If $A \cap B = \emptyset$ we say A and B are mutually exclusive sets

Probability space I

- A random experiment: A repeatable procedure such as rolling a dice, tossing a coin, drawing a card
- Set of outcomes: collection of all possible outcomes of a random experiments $\{H, T\}$ for coin toss, denoted by Ω
- Events: collection of subset of outcome set (for infinite outcome sets this definition is slightly involved)
- Probability function P: assigns a number to each event such that
  - $P(\Omega) = 1$
  - $P(A) \geq 0$ for all $A \subseteq \Omega$
  - If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

Probability space II

Properties

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  Proof: Write $A \cup B = A \cup (\overline{A} \cap B)$ and $B = (A \cap B) \cup (\overline{A} \cap B)$. By the third property we have

$$P(A \cup B) = P(A) + P(\overline{A} \cap B) \tag{1}$$

$$P(B) = P(A \cap B) + P(\overline{A} \cap B)$$

$$\implies P(\overline{A} \cap B) = P(B) - P(A \cap B) \tag{2}$$

  replacing this value in (1) we get desired result

Probability space III

Examples

- Tossing a fair coin twice
  - What is Ω, $\mathcal{F}$ and P?
  - What is the probability of event: the outcomes of both tosses is same
- Rolling a dice
  - What is $(\Omega, \mathcal{F}, P)$?
  - What is the probability of getting an even number ?
- Drawing cards from well shuffled pack
  - Write the probability of getting a diamond face card.

More Examples

Your friend has tossed a coin twice and you know it is not a fair coin. You are asked to choose between following two predictions. 1) The outcomes of both tosses is same and 2) the outcome is different. Which will you choose and why?

More Examples

Your friend has tossed a coin twice and you know it is not a fair coin. You are asked to choose between following two predictions. 1) The outcomes of both tosses is same and 2) the outcome is different. Which will you choose and why?

A box contains three marbles(red, green and blue). COnsider an experiment of drawing a marble from a box, replacing it and drawing one marble again. What is the sample space of this experiment? If each marble is equally likely to to be selected what is the probability of each point in sample space. Repeat the same when the marble is not replaced after the first draw.

Conditional Probability I

the probability of A given B

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) \neq 0 \tag{3}$$

Examples

- Toss a fair coin 4 times. Let $A =$ atleast 3 heads and $B =$ first toss is tails. Find $P(A|B)$.
- Answer: Clearly $P(B) = 1/2$ and
  $A = \{\{HHHH\}, \{THHH\}, \{HTHH\}, \{HHTH\}, \{HHHT\}\}$
  $B = \{HXXX\}$.
  $A \cap B = \{\{HHHH\}, \{HTHH\}, \{HHTH\}, \{HHHT\}\}$
  $P(A|B) = \frac{4/16}{1/2} = 1/2$

Conditional Probability II

Example 2

A box contains three coins: two fair coins and one two-headed coin $(P(H) = 1)$,
Pick a coin at random and toss it. You pick a coin at random and toss it, and get heads. What is the probability that it is the two-headed coin?

Answer: Let A = coin lands heads and B = chosen coin is two-headed.
We want to find $P(B|A) = \frac{P(A \cap B)}{P(A)}$

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \overline{B}) \\ &= P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) \\ &= 1 \cdot 1/3 + 1/2 \cdot 2/3 = 2/3 \end{aligned}$$

$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{1 \cdot 1/3}{2/3} = 1/2.$

Law of Total Probability

$P(A) = \sum_n P(A \cap B_n)$, here B_n are events whose union is the entire sample space; $\cup_n B_n = \Omega$.

Law of Total Probability

$P(A) = \sum_n P(A \cap B_n)$, here B_n are events whose union is the entire sample space; $\cup_n B_n = \Omega$.
Alternatively,

$$P(A) = \sum_n P(A|B_n)P(B_n)$$

Example

Two factories supply light bulbs. Factory X's bulbs work for over 5000 hours in 99% of cases, whereas factory Y's bulbs work for over 5000 hours in 95% of cases. It is known that factory X supplies 60% of the total bulbs available and Y supplies 40% of the total bulbs available. What is the chance that a purchased bulb will work for longer than 5000 hours?

Law of Total Probability

$P(A) = \sum_n P(A \cap B_n)$, here B_n are events whose union is the entire sample space; $\cup_n B_n = \Omega$.
Alternatively,

$$P(A) = \sum_n P(A|B_n)P(B_n)$$

Example

Two factories supply light bulbs. Factory X's bulbs work for over 5000 hours in 99% of cases, whereas factory Y's bulbs work for over 5000 hours in 95% of cases. It is known that factory X supplies 60% of the total bulbs available and Y supplies 40% of the total bulbs available. What is the chance that a purchased bulb will work for longer than 5000 hours?

Answer: $0.99 \cdot 0.6 + 0.95 \cdot 0.4$

Independent Events I

Let A and B be two events. We call them (pairwise) independent if $P(A \cap B) = P(A) \cdot P(B)$.

Independent Events II

Example 2

A biased coin is tossed till we get a head. What is the probability that the number of required tosses is odd?

Answer: Let A_i = head appears in ith toss for the first time.
Assume each trial is independent of the rest and let $P(H) := p$ and $P(T) := q = 1 - p$. We have $P(A_i) = q^{i-1}p$.

$$\begin{aligned} P(\text{Head apeears on odd toss}) &= P(A_1 \cup A_3 \cdots) \\ &= \sum_{i=0}^{\infty} P(A_{2i+1}) = \sum_{i=0}^{\infty} q^{2i} p = p \sum_{i=0}^{\infty} q^{2i} \\ &= \frac{p}{1-q^2} = \frac{p}{p(1+q)} = \frac{1}{2-p} \end{aligned}$$

Independent Events III

Question

Two events are mutually exclusive. Can they be independent?

Question

If A_1 and A_2 are independent events, are $\overline{A}_1$ and $\overline{A}_2$ independent? If so, why?

Baye's Formula

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(B_j|A)P(B_j)}$$

- We have seen this earlier
- Denominator is $P(A)$; use law of total probability
- Numerator is $P(A \cap B_i)$; use conditional probability

Examples I

Monty Hall Problem

- Consider a game where there are three closed doors.
- Behind one door there is a car and behind other two doors there is nothing
- Contestant is asked to choose a door and the host opens the door which has nothing
- contestant is then given a choice to switch the choice or remain with original choice.
- what will you do and why?

Monty Hall Problem

- Consider a game where there are three closed doors.
- Behind one door there is a car and behind other two doors there is nothing
- Contestant is asked to choose a door and the host opens the door which has nothing
- contestant is then given a choice to switch the choice or remain with original choice.
- what will you do and why?

Answer: Always switch. Read wikipedia article for more details of the problem and its solution.

Random Variables

A mapping X from outcome space to $\mathbb{R}$.

Random Variables

A mapping X from outcome space to $\mathbb{R}$.

Bernoulli Random Variable

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability} 1 - p \end{cases}$$

Random Variables

A mapping X from outcome space to $\mathbb{R}$.

Bernoulli Random Variable

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability} 1-p \end{cases}$$

Rademacher RV

$$X = \begin{cases} -1 & \text{with probability } 1/2 \\ 1 & \text{with probability} 1/2 \end{cases}$$

Definition

- PMF $f_X(k) := P(X = k)$
- CDF $F_X(k) = P(X \leq k)$

Continuous RVs

- The set of values a random variable takes is uncountably infinite
- definition: X is continuous RV if there exists a non-decreasing function f on $\mathbb{R}$ such that for any set B of real numbers

$$P(X \in B) = \int_B f(x)dx \tag{4}$$

Continuous RVs

- The set of values a random variable takes is uncountably infinite
- definition: X is continuous RV if there exists a non-decreasing function f on $\mathbb{R}$ such that for any set B of real numbers

$$P(X \in B) = \int_B f(x)dx \tag{4}$$

- PDF $f(x)$
- CDF $F_X(x) := \int_{-\infty}^{x} f(x)dx$

Properties of CDF,PDF

- $F_X(\infty) = 1$ and $F_X(-\infty) = 0$
- For any $k_1 < k_2$, $F_X(k_1) \leq F_X(k_2)$ (monotone increasing)
- For any $k_1 < k_2$ $P(k_1 < X \leq k_2) \leq F_X(k_2) - F_X(k_1)$

Expectation and variance

- $\mathbb{E}[X] = \sum_x x \cdot P(X = x)$
- Also seen as average or mean value of RV
- Conditional Expectation:

$$\mathbb{E}[X|A] = \sum_x xP(x|A)$$

- Variance: $\mathbb{E}(X - \mathbb{E}[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
- kth Moment: $\mathbb{E}[X^k]$

Properties

- Linearity of Expectation: Let X and Y be two random variables. Then $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Properties

- Linearity of Expectation: Let X and Y be two random variables. Then $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ if X and Y are independent.

Properties

- Linearity of Expectation: Let X and Y be two random variables. Then $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ if X and Y are independent.
- $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$

Properties

- Linearity of Expectation: Let X and Y be two random variables. Then $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ if X and Y are independent.
- $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$
- $E[f(X)] = \sum_x f(x) P(X = x)$

Properties

- $\mathsf{Var}(aX + b) = a^2\mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Properties

- $\mathsf{Var}(aX + b) = a^2\mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- $\mathsf{Cov}(X, X) = \mathsf{Var}(X)$

Properties

- $\mathsf{Var}(aX + b) = a^2\mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- $\mathsf{Cov}(X, X) = \mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) = \mathsf{Cov}(Y, X)$

Properties

- $\mathsf{Var}(aX + b) = a^2\mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- $\mathsf{Cov}(X, X) = \mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) = \mathsf{Cov}(Y, X)$
- $\mathsf{Cov}(aX, Y) = a\mathsf{Cov}(X, Y)$

Properties

- $\mathsf{Var}(aX + b) = a^2 \mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- $\mathsf{Cov}(X, X) = \mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) = \mathsf{Cov}(Y, X)$
- $\mathsf{Cov}(aX, Y) = a\mathsf{Cov}(X, Y)$
- $\mathsf{Cov}(X, Y + Z) = \mathsf{Cov}(X, Y) + \mathsf{Cov}(X, Z)$

Properties

- $\mathsf{Var}(aX + b) = a^2\mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- $\mathsf{Cov}(X, X) = \mathsf{Var}(X)$
- $\mathsf{Cov}(X, Y) = \mathsf{Cov}(Y, X)$
- $\mathsf{Cov}(aX, Y) = a\mathsf{Cov}(X, Y)$
- $\mathsf{Cov}(X, Y + Z) = \mathsf{Cov}(X, Y) + \mathsf{Cov}(X, Z)$
- Let X and Y be independent then

$$\mathsf{Var}(XY) = \mathsf{Var}(X)\mathsf{Var}(Y) + \mathsf{Var}(X)(\mathbb{E}[Y])^2 + \mathsf{Var}(Y)(\mathbb{E}[X])^2$$

Properties

- $\mathsf{Var}(aX+b) = a^2\mathsf{Var}(X)$
- $\mathsf{Cov}(X,Y) := \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$
- $\mathsf{Cov}(X,X) = \mathsf{Var}(X)$
- $\mathsf{Cov}(X,Y) = \mathsf{Cov}(Y,X)$
- $\mathsf{Cov}(aX,Y) = a\mathsf{Cov}(X,Y)$
- $\mathsf{Cov}(X,Y+Z) = \mathsf{Cov}(X,Y)+\mathsf{Cov}(X,Z)$
- Let X and Y be independent then

$$\mathsf{Var}(XY) = \mathsf{Var}(X)\mathsf{Var}(Y) + \mathsf{Var}(X)(\mathbb{E}[Y])^2 + \mathsf{Var}(Y)(\mathbb{E}[X])^2$$

- $\mathsf{Var}(X+Y) = \mathsf{Var}(X)+\mathsf{Var}(Y)+2\mathsf{Cov}(X,Y)$

Examples

Bernoulli Random Variable

Bernoulli Random Variable

- $X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability} 1 - p \end{cases}$

Bernoulli Random Variable

- $X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability} 1-p \end{cases}$
- $E[X] = 1 \cdot p + 0 \cdot (1-p)$

Bernoulli Random Variable

- $X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability} 1-p \end{cases}$
- $E[X] = 1 \cdot p + 0 \cdot (1-p)$
- $\mathsf{Var}[X] = p(1-p)$

Bernoulli Random Variable

- $\text{X} = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability} 1-p \end{cases}$
- $E[X] = 1 \cdot p + 0 \cdot (1-p)$
- $\mathsf{Var}[X] = p(1-p)$
- $\mathbb{E}[X^k] = p$

Binomial Random Variable

- Denoted by $Bin(n, p)$
- Interpreted as number of successes (1's) in n independent trials of Bernoulli random variables with probability p i.e. $Bin(n, p) = \sum_{i=1}^{n} X_i$ with $X_i \sim Bern(p)$ independent.
- PMF $f_X(x) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\mathbb{E}[X] = np$
- $\mathsf{Var}[X] = np(1-p)$

Geometric Random variables

Geometric RV

- Probability of getting first success in sequence of independent Bernoulli trials
- $f_X(k) = (1-p)^k p$
- $\mathbb{E}[X] = \frac{1}{p}$
- $\mathsf{Var}[X] = \frac{1-p}{p^2}$

Memoryless property: $P(X > m + n | X \geq m) = P(X > n)$
Geometric distribution is memoryless.