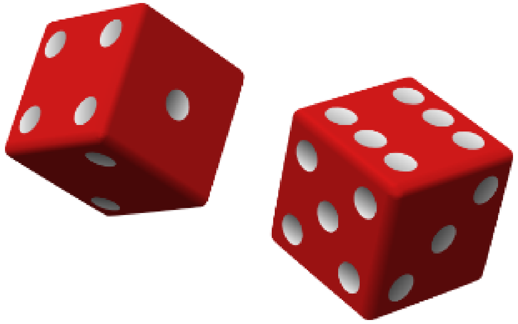
Probability and Combinatorics

The science of uncertainty… and gambling



Yordan Darakchiev

Technical Trainer

iordan93@gmail.com

sli.do

#MathForDevs

Table of Contents

Probability

Definition and principles

Some Definitions

- The scientific method relies on **experiments**
  - Initial conditions → outcome
    - Usually, we control the initial conditions to isolate the outcome
- **Random event**
  - A set of outcomes of an experiment
  - Each outcome happens with a certain **probability**
- **Random variable**
  - An expression whose value is the outcome of the experiment
  - Usually denoted with X, Y, Z... (capital letters)
- **It is not possible to predict the next outcome of a random event!**
  - But we can perform the same experiment many times (trials)
  - The patterns and laws become more apparent with more trials

Frequency

- Let's perform the same experiment many times
  - Under the same conditions
  - … such as throwing a dice
- Assign a frequency to each number $i = \{1, 2, \dots, 6\}$ that the dice shows

$$f_i = \frac{m_i}{n}$$

  - m – number of trials we got i, n – all trials
- As n increases, f_i "stabilizes" around some number
- We cannot perform infinitely many experiments
  - But we can "extend" the trials mathematically

$$p(A) = \lim_{n \to \infty} \frac{m}{n}$$

  - We call this the probability of outcome **A**
    - **Statistical definition** of probability

Examples

- Rolling a dice
  - Possible outcomes: $\{1, 2, 3, 4, 5, 6\}$
  - Fair dice – all outcomes are equally likely

    $$p(1) = p(2) = \cdots = p(6) = 1/6$$
- Tossing a fair coin
  - Possible outcomes: $\{0 = heads, 1 = tails\}$

    $$p(0) = p(1) = 1/2$$
- Tossing an unfair coin

  $$p(0) = 0,3; \; p(1) = 0,7$$
- Note that
  - The probability $p \in [0; 1]$
    - It can also be expressed as percentage: $p \in [0\%; 100\%]$
  - The sum of all probabilities is equal to 1

Countable and Uncountable Outcomes

- In some cases, the number of outcomes is finite
- But some random variables have **infinitely many** outcomes
- Example: intervals
  - What is the probability that a real number $A \in [0; 100]$ chosen at random, is in the interval $[10; 20]$?
  - Answer: it depends only on the lengths of the intervals

$$p = \frac{20 - 10}{100 - 0} = 0.1 = 10\%$$

    - The number of outcomes in infinite, but we are still able to compute probabilities
  - **Probability density** – the probability of the result being in a tiny interval dx

$$dp = \frac{dx}{b - a}$$

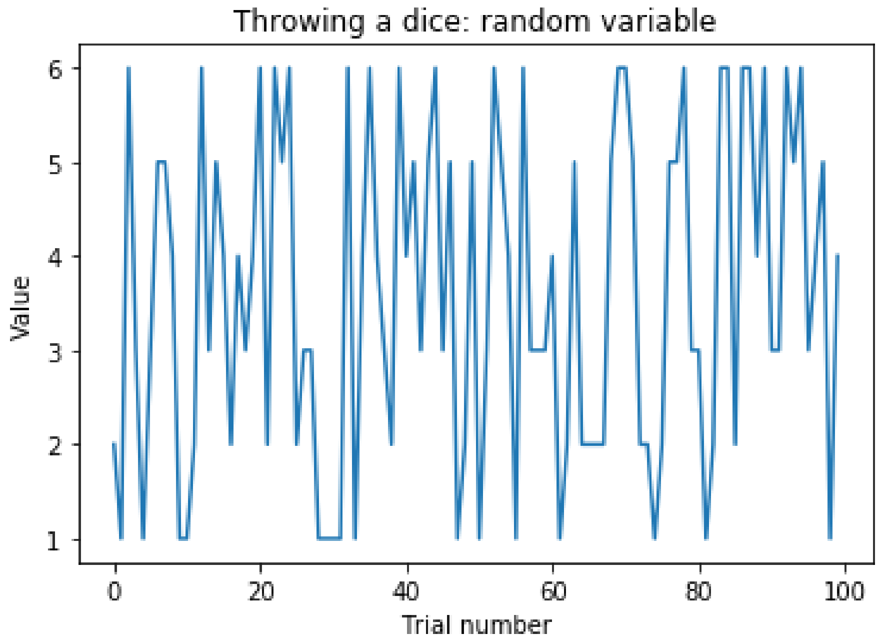
    - a, b – both ends of the interval $[0; 100]$

Visualizing Random Variables

- To visualize a random variable, we plot the value as a function of the trial number
  - We can generate random values using **numpy**
  - Example: throwing a dice

```
def throw_dice():
  return np.random.randint(1, 7) # from 1 to 6

x = [throw_dice() for i in range(100)]
plt.plot(x)
plt.show()
```

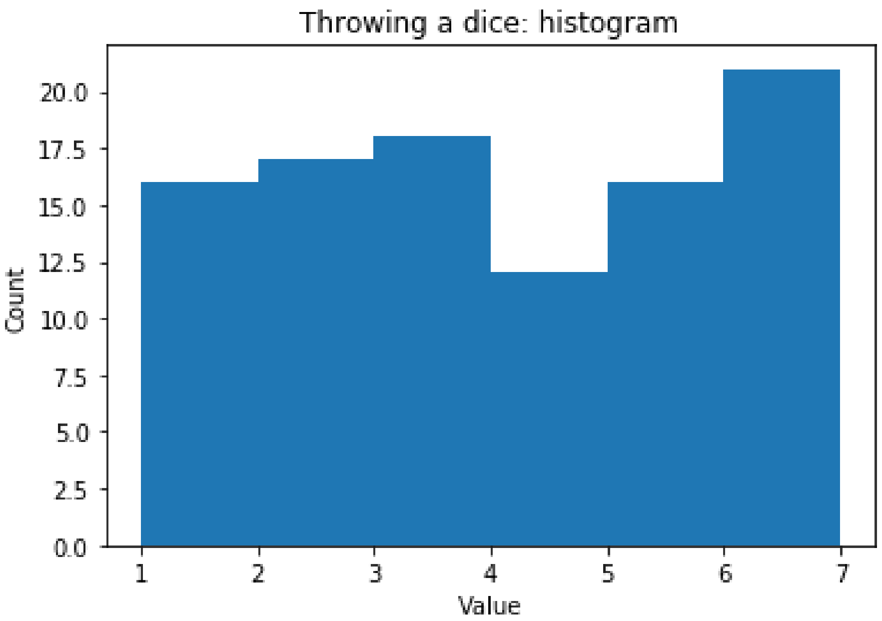


Visualizing Random Variables (2)

- The function we got is not very informative
  - Better way: show the frequency of each output
    - For each possible value of the random variable, count how many times we got that value
  - This is called a **histogram**

```
# Counting all values
from collections import Counter
counts = Counter(x)
for number, count in counts.items():
  print(str(number) + ": " + str(count))
```

```
# Plotting a histogram
plt.title("Throwing a dice: histogram")
plt.hist(x, bins = range(1, 8))
plt.ylabel("Count")
plt.show()
```



Combinatorics

How to count things

Combinatorics

- Combinatorics deals with **counting** objects and groups of objects
- Prerequisites
  - Finite (countable) number of outcomes
  - All outcomes have equal probability
- Examples: gambling games
  - Roulette – all segments are equally likely
  - Card games – all card backs are the same
- Counting rules
  - Rules for computing a **combinatorial probability**
  - Show how many "desired" outcomes exist

Combinatorics (2)

- Notation
  - All outcomes: n
  - All experiment outcomes: k
    - Usually, n is fixed and k depends on the experiment
- Types of samples
  - with repetition / without repetition
  - ordered / unordered
- Example: taking numbered balls out of a box
  - Take a ball, then return it to the box
  - Take a ball without returning it to the box (in this case $k \leq n$)

  - Take balls in a specific order (e.g., if they are numbered or colored)
  - Take balls in no specific order

Counting Rules

- **Rule of sum**
  - m choices for one action, n choices for another action
  - The two actions **cannot be done at the same time**

  $\Rightarrow$There are $m + n$ ways to choose one of these actions
- Example
  - A woman will shop at **one** store in town today
    - North part of town – mall, furniture, jewellery (3 stores)
    - South part of town – clothing, shoes (2 stores)
  - In how many ways she could visit one shop?
  - Answer: 3 + 2 = 5 ways

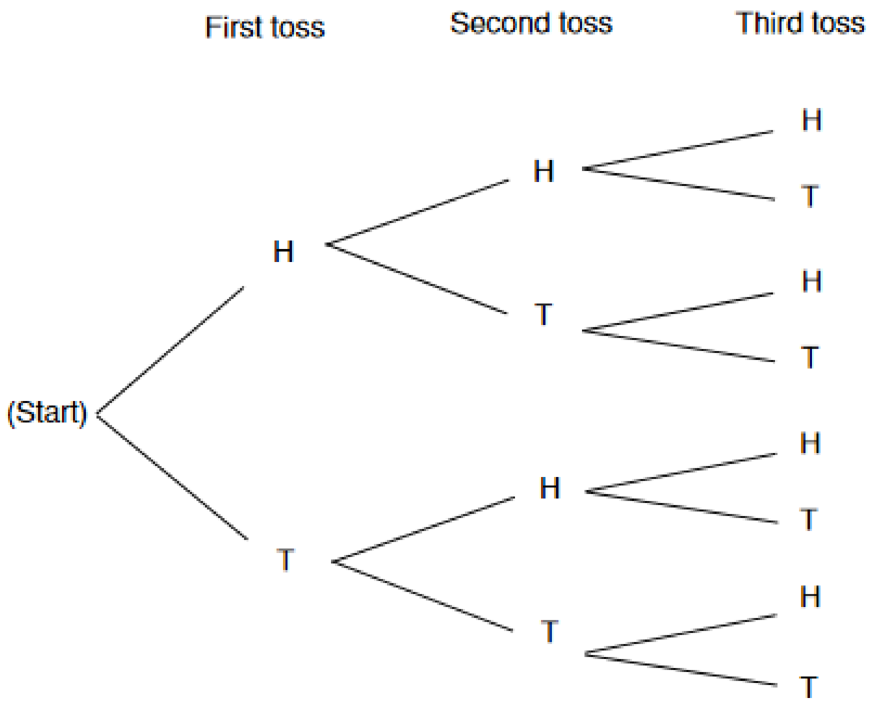
Counting Rules (2)

- **Rule of product**
  - m choices for one action, n choices for another action
  - The two actions are performed **one after the other**

  $\Rightarrow$There are $m.n$ ways to do both actions
- Example
  - You have to decide what to wear
    - Shirts – red, blue, purple (3 colors)
    - Pants – black, white (2 colors)
  - In how many ways can you create one outfit (shirt and pants)?
  - Answer: 3.2 = 6 ways
    - For each choice of shirt, you can choose one color of pants
    - These are **independent**

Example: Three Coin Tosses

- Let's explore a graphic method of solving combinatorial problems called a **tree diagram**
  - Draw all intermediate results and the links between them
  - A "path" through the tree represents an outcome
  - Useful when the outcomes are relatively few
- What's the probability of getting 3 tails out of 3 coin tosses?
  - Answer: 1/8
- What's the probability that both of these are true?
  - The first outcome is a head
  - The second outcome is a tail
  - Answer: 1/4



Example 2: Eating at a Restaurant

- A restaurant offers
  - 5 choices of appetizer
  - 10 choices of main course
  - 4 choices of dessert
- You can choose one course, two **different** courses, or all three
- How many possible meals can you make?
  - One course: either appetizer, main course, or dessert: $5 + 10 + 4 = 19$
  - Two courses: total 110
    - Appetizer + main course: $5.10 = 50$
    - Main course + dessert: $10.4 = 40$
    - Appetizer + dessert: $5.4 = 20$
  - Three courses: $5.10.4 = 200$
  - Total: $19 + 110 + 200 = 329$ possible meals

Permutations

- A permutation (without repetition) of a set A is any shuffling of all elements in A
  - The order matters
  - Notation: P_n
- Example:
  - If $A = \{1, 2, 3, 4\}$, some permutations are $\{1, 2, 3, 4\}$; $\{1, 4, 3, 2\}$; $\{2, 3, 4, 1\}$
- Number of permutations of n elements
  - n choices for the first element
  - $n - 1$ for the second one
    - Because the first one is already taken
  - $n - 2$ for the third one
  - 1 for the last one
  - Total: $n! = 1.2.3. \cdots . n$

Variations

- A variation is an ordered subset of k elements from A
- Notation: V_n^k
  - We read this as "Variations of n elements, k^{th} class"
- Example:
  - If $A = \{1, 2, 3, 4\},\ k = 2$, some variations are $\{1, 2\};\ \{4, 3\};\ \{3, 1\};\ \{4, 1\}$
- Number of variations
  - Same technique as in permutations
  - n choices for the first element
  - $n - 1$ for the second one
  - $(n - k + 1)$ for the last one

$$V_n^k = n.(n-1).\cdots.(n-k+1) = \frac{n!}{(n-k)!}$$

Combinations

- A combination is an unordered subset of k elements from A
- Notation: C_n^k
- Example:
  - If $A = \{1, 2, 3, 4\}, \quad k = 2$, some combinations are $\{1, 2\}; \{3, 4\}; \{3, 1\}; \{4, 1\}$
- Number of combinations of n elements
  - Using a similar (but more involved) logic, we can prove that

$$C_n^k = \frac{n!}{(n-k)!k!}$$

  - This is also known as "n choose k" (we choose k elements from n)

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Example Usages

- Shuffle a deck of cards
  - The same as generating a random permutation of 52 (or 54) elements
- Crack a password
  - How many 3-letter passwords are there ($26 + 26$ letters total)? V_{52}^{3}
- Generate all anagrams of a given word
  - Anagram: a different word using the same letters
    - Example: emits → items, mites, smite, times
  - Method:
    - Generate all permutations of the letters
    - For each permutation, find whether it's a valid word (check with a dictionary)
    - Return all valid words
- Make a fruit salad
  - Generate combinations of fruits (the order doesn't matter)
    - Possibly, combinations with repetition (if I love bananas, I'll take a double serving)

Probability Algebra

Sets and probabilities, geometry intuition

Events

- **Event** – a result from the experiment
- **Elementary event**
  - One particular outcome
  - Example: outcomes of two coin flips: $\{HH\}, \{HT\}, \{TH\}, \{TT\}$
- Compound event
  - Consists of many elementary events
  - Example: getting an odd number from a dice
    - Consists of the elementary events $1, 3, 5$
- **Event space** – the set Ω of all possible events
- The algebra of events is the same as the algebra of sets
  - … and we already know these :)

Algebra of Events

- If event A happens with event B, A is a consequence of B: $A \subset B$
- If $A \subset B$ and $B \subset A$, then $A = B$
- Complementary event: $\bar{A}$ happens iff A does **not** happen
- Impossible event: contains no elementary events: $\emptyset$
- Product of events: happens iff **A and B** happen: $C = A \cap B$
  - Incompatible events: $A \cap B = \emptyset$
- Sum of events: happens if **A, B or both** happen: $C = A \cup B$
  - If A and B are incompatible, $C = A + B$
- Observe that
  - Logical relations are the same as set operations (and event operations)
    - **AND**: intersection
    - **OR**: union
    - **NOT**: complement

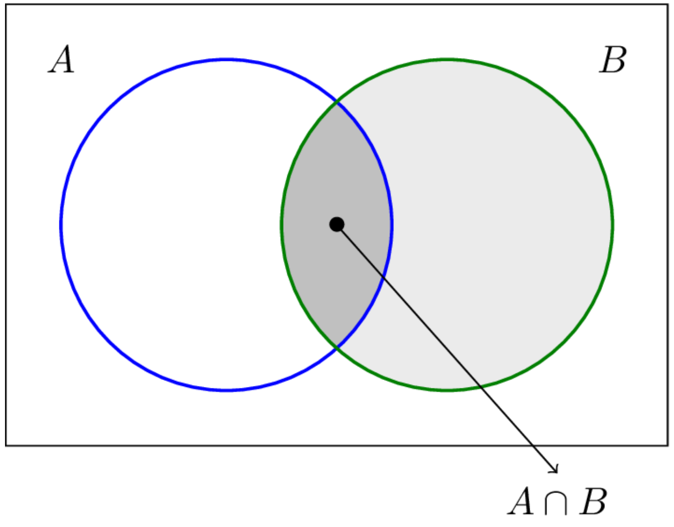
Conditional Probability

- Additional information about the experiment outcome can change the probabilities
  - "a priori" → "a posteriori"
- Example:
  - "Hidden dice": someone rolls a dice and doesn't tell us the result
  - Probabilities: $1/6$ for every number
    - These are also called "a priori" probabilities
  - Now we know the number is even
    - This changes all outcome probabilities: $\left\{1 \to 0;\ 2 \to \frac{1}{3}; 3 \to 0; 4 \to \frac{1}{3}; 5 \to 0; 6 \to \frac{1}{3}\right\}$
      - These are called "a posteriori" probabilities
- Conditional probability
  - Probability of event A given event B
  - Math notation: $P(A|B)$

Conditional Probability (2)

- More formally
  - If we know B happened, the probability $P(A|B)$ corresponds to the part of the set B which is shared between A and B

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



- In our example
  - Event A: number on a fair dice
    - $A = \{1, 2, 3, 4, 5, 6\}$
  - Event B: the number is even
    - $B = \{2, 4, 6\}$
  - $A \cap B = \{2, 4, 6\}$
  - $P(1|\text{even}) = 0; P(2|\text{even}) = \frac{1}{3}; \ldots$

Event Independence

- Sometimes, an event doesn't influence another event
  - They are called independent events
- If two events are independent, knowledge of one **does not tell us anything** about the other
- More formally, $P(A \cap B) = P(A).P(B)$
  - If $P(B) \neq 0$, $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A).P(B)}{P(B)} = P(A)$
  - The same can be applied to A if $P(A) \neq 0$
- Example
  - 99% of all people who died of cancer, have consumed pickles
  - 99,8% of all soldiers have eaten pickles
    - http://www.pleacher.com/mp/mhumor/pickles.html
  - http://www.dhmo.org/facts.html

Bayes' Theorem

- The theorem tells us how to update the probabilities when we know some evidence

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

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

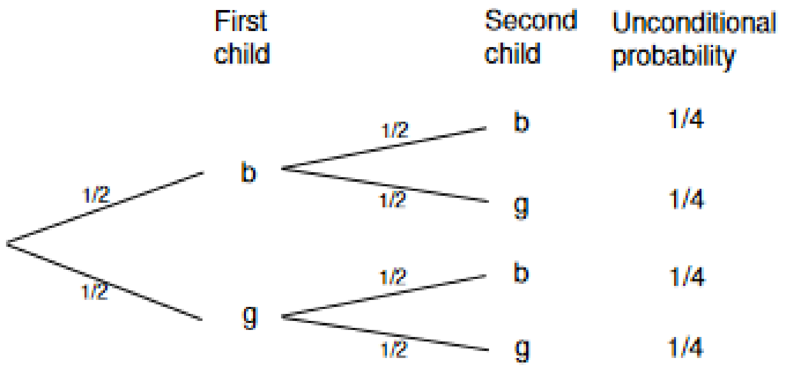
$$A \cap B = B \cap A \Rightarrow P(A|B)P(B) = P(B|A)P(A)$$

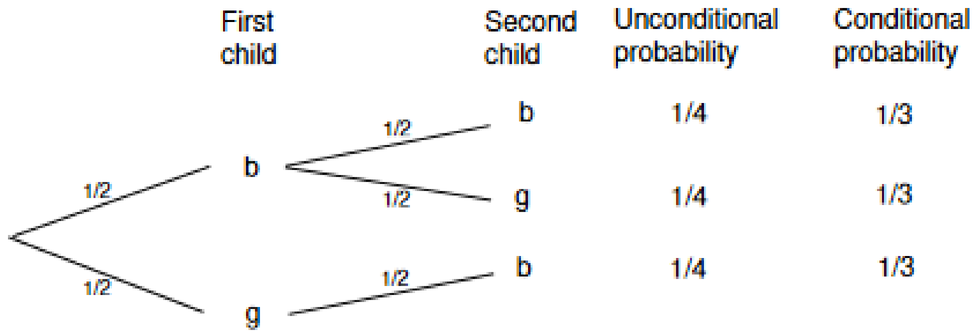
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Example usage: spam detection
  - Consider each word w; compute number of emails which contain it
    - m spam emails containing w; n total emails containing w:
    - "Spamminess" of word: frequency $P(\text{word}|\text{spam}) = m/n$
    - "Spamminess" of email: $P(\text{spam}|\text{all words})$

Example: Family Paradox 1

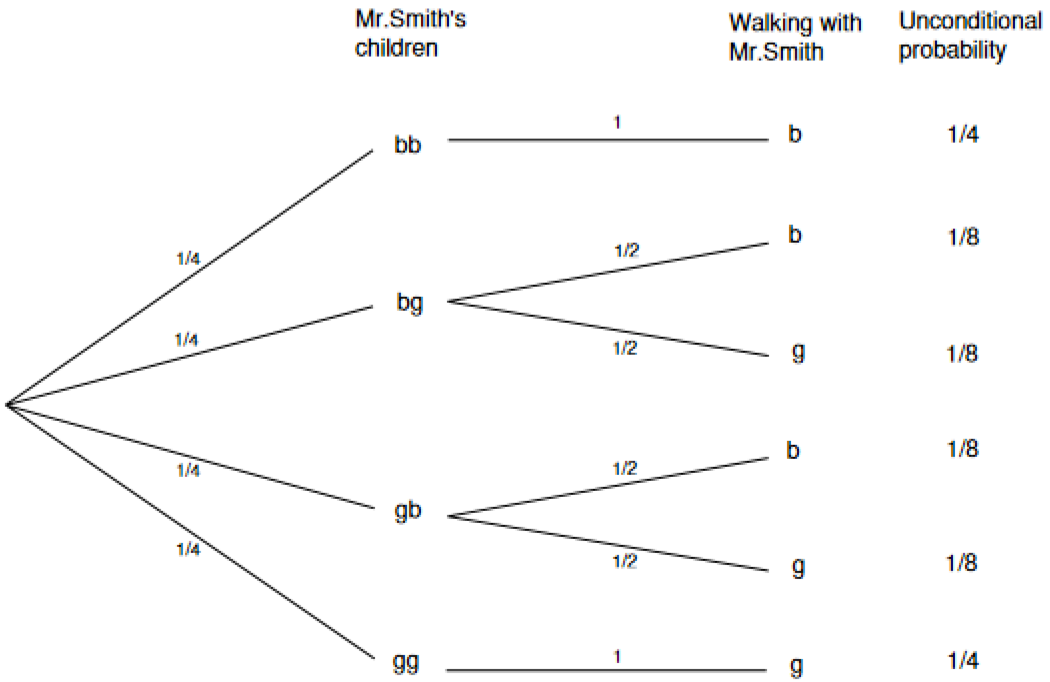
- A family has two children
  - One of them is a boy
  - What is the probability that both children are boys?
    - A child has a 0,5 chance of being a boy or a girl
- Intuitive answer: 0,25
  - But wait… let's exhaust all possibilities

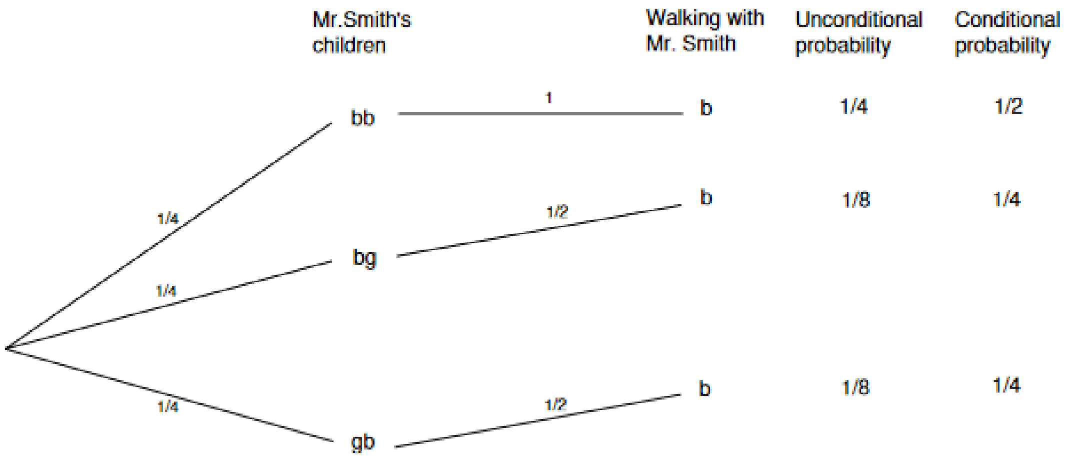




Example: Family Paradox 2

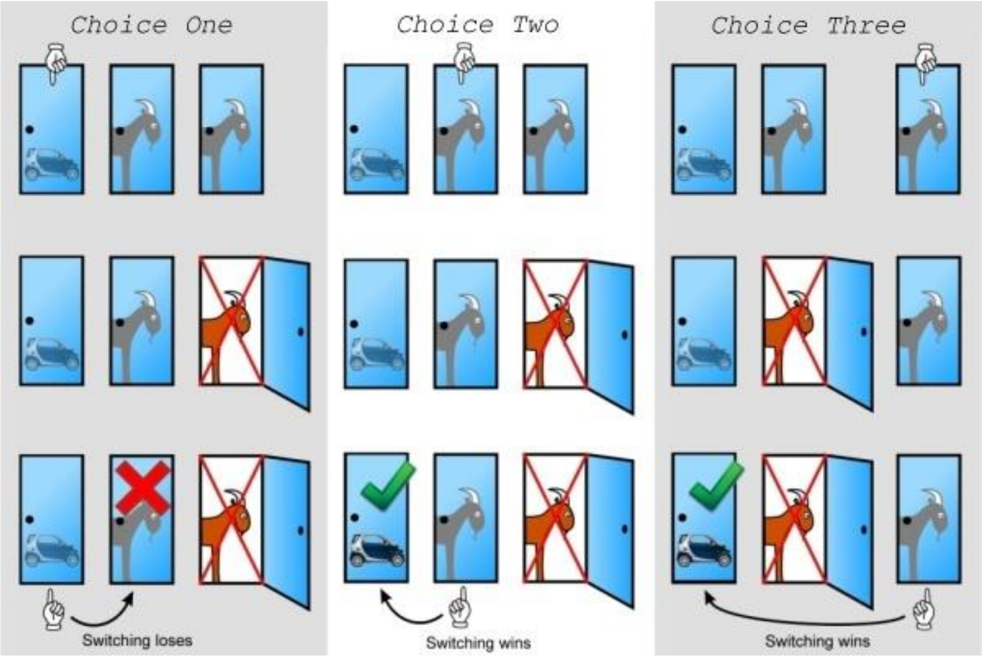
- Mr. Smith is the father of two children
  - When we meet him on the street, he introduces one as his son
  - What's the probability that the other child is a boy?
- Assumption
  - He is equally likely to take any child to a walk





Example: Monty Hall Problem

- In a game show, you have to choose between three doors
  - Behind one is a car, behind the other two – goats
- You pick a door
- The host reveals one of the two other doors – it's always a goat
- You have the option to keep your choice or switch doors
  - Which is the winning strategy?
- It turns out that the winning strategy is to always switch
  - This gives you 2/3 chance of winning the car
- More details: [Quora]



Statistical Distributions

Seeing the results of our experiments

Distributions

- We saw that random variables can be treated as functions
  - But they look funky
    - Don't have derivatives at most points
    - Difficult to work with
- We can instead take functions of these functions
  - Like we counted each outcome
    - Instead of graphing the real function, we made a histogram of counts
    - This gives us a much better idea what the random variable looks like
- These functions of functions are called **distributions**
  - In our example, we looked at the frequency distribution

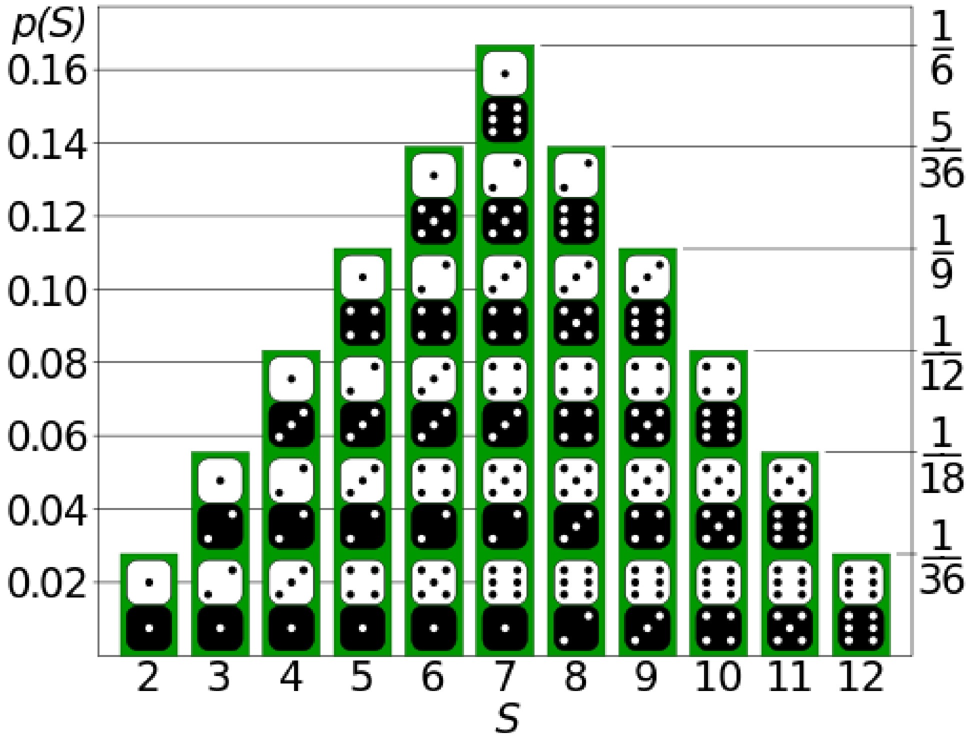
Discrete Distribution

- Probability distribution function
  - A table which maps each outcome of a random variable to a probability: $\mathrm{p_X}(x_i) = P(X = x_i)$
  - Also called **probability mass function** (pmf)
- Example: two die rolls
  - Random variable: sum of numbers
  - Outcomes: $\{2, 3, \ldots, 12\}$
  - Probabilities:

$$P(2) = P(\{1,1\}) = 1/36$$

$$P(3) = P(\{1,2\}) + P(\{2,1\}) = 2/36$$

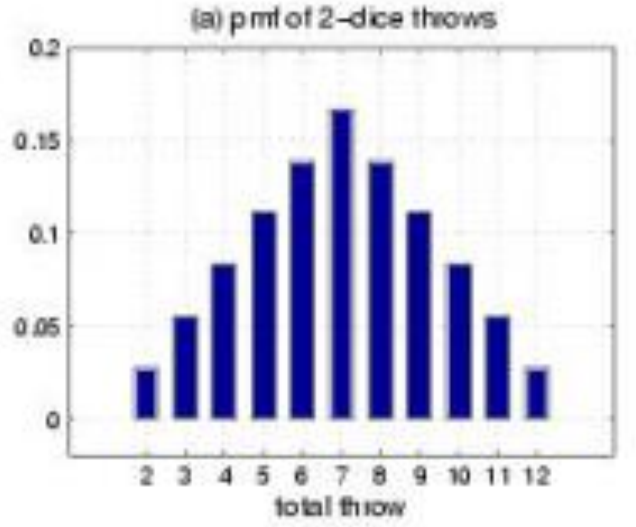
$$\vdots$$

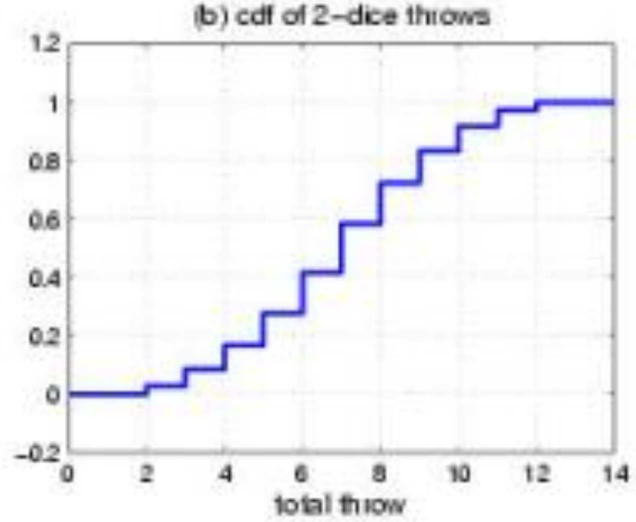


Discrete Distribution (2)

- **Cumulative distribution function**
  - A table which maps each outcome of a random variable
    to the probability of its value
    being less than or equal to
    a given number

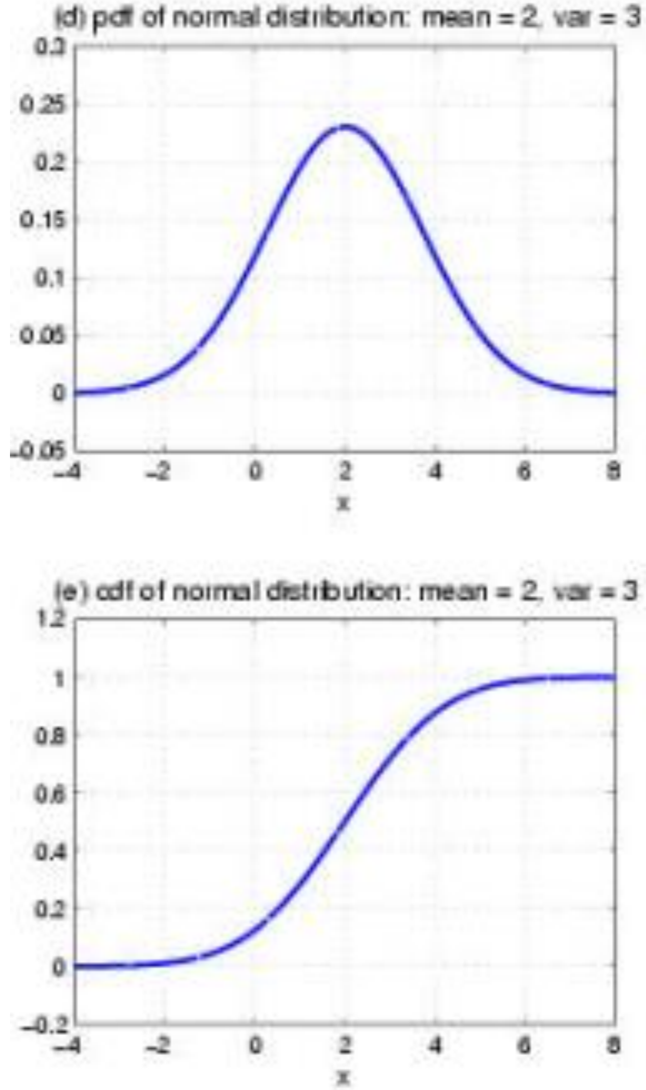
    $$F_X(x_i) = P(X \leq x_i)$$
  - Also called
    cumulative mass function (cmf)
    or **cumulative density function** (cdf)
  - Every cmf is non-decreasing
    - Usually starts at 0
    - Always ends at 1





Continuous Distribution

- **Cumulative density function** (cdf)
  - Defined in the same way as the cmf: $F(x) = P(X \leq x)$
- **Probability density function**
  - Derivative of the cdf:
    $$f(x) = \frac{dF(x)}{dx}$$
  - Meaning: the probability of the function taking values in an infinitely small interval around x
  - **The probability of observing any single value a is exactly 0**
    - The number of outcomes is ∞
    - $p(a) = \left[\frac{\#\text{ of values } a}{\infty}\right] = 0$



Common Distributions

Probability and Statistics Playing Together

Bernoulli and Uniform Distributions

- Bernoulli distribution
  - The simplest distribution of a random variable
    - Value 0 with probability p
    - Value 1 with probability $q = 1 - p$
  - The two events are incompatible (mutually exclusive)
  - Example: coin flip (fair coin: $p = 0{,}5$)
  - … Not so interesting on its own
    - But takes part in other distributions
- Uniform distribution
  - All values in some range $[a; b]$ are equally likely
  - Example: number on a fair dice
    - Also generalizes to continuous variables

Binomial Distribution

- n Bernoulli trials
  - Each trial has a "success" probability p
  - $n = 1 \Rightarrow$ Bernoulli distribution
- Discrete distribution
- Notation: $X \sim B(n, p)$
  - "X follows the binomial distribution with parameters n and p"
- Probability mass function

$$f(k; n, p) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Cumulative function

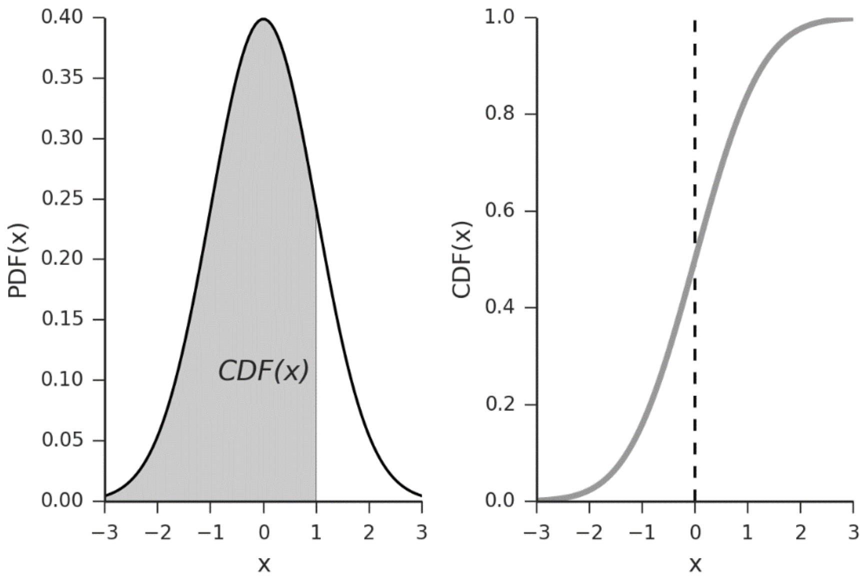
$$F(k; n, p) = P(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$$

Normal Distribution ♥

- Origin: random errors in measurements
  - When we perform an experiment, there are many sources of error
- Example: throwing a dart at the origin of the (x, y)-plane
  - We aim at the origin
  - Random errors prevent us from hitting it every time
  - Sources of error
    - Hand shaking, air currents, distribution of mass inside the arrow, different viewing angles… and many more, some of which we can't even know
- Assumptions
  - The errors don't depend on the orientation of the coordinate system
  - The errors in x and y directions are independent: one doesn't influence the other
  - Large errors are less likely than small errors

Normal Distribution (2)

- We can derive the distribution of errors
  - Distances from the origin
- Normal (Gaussian) distribution
  - pdf: $p(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
  - μ, σ – parameters
    - We'll see their real meaning next time
  - cdf: doesn't exist as a function, we can integrate numerically
- Complete derivation of the formula: [here]
- **Standard normal distribution:** $\mu = 0,\ \sigma = 1$
  - Mainly for convenience



Central Limit Theorem

- The sum of many independent random variables tends to a normal distribution even if the original random variables are not normally distributed
  - In other words: The sampling distribution of the mean of any independent random variable will be normal or nearly normal if the sample is large enough
  - Large enough?
    - $n \in [30; 40]$ for most statisticians, but more is better
- Example: customers in a shop
  - Every customer has their own behavior, reasons, money, etc.
    - We can treat them as random variables with unknown distributions
  - The shop's earnings are approximately normally distributed
    - If there are many customers
  - We **don’t even care** about the many sources of error: they will produce a normal distribution anyway

Summary

- Probability
- Combinatorics
- Algebra of events
- Conditional probability
- Applications
- Statistical distributions
- Central limit theorem

Questions?