

Probabilistic Method and Random Graphs

Lecture 1. Elementary probability theory with applications ¹

Xingwu Liu

Institute of Computing Technology
Chinese Academy of Sciences, Beijing, China

¹The slides are mainly based on Chapters 1 and 2 of *Probability and Computing*.

Important information

Course homepage

<https://probabilityandgraphs.github.io/>

Teaching assistant

Mengying Guo (guomengying@ict.ac.cn)

Zhenyu Sun (sunzhenyu19s@ict.ac.cn)

Office hours: TBD

Homework

Submit in **PDF** to: probabilitygraphs@163.com

Deadline: 9:00am, Thursday

Grading policy

Homework+**Participation**: 50%

Final exam (Open book): 50%

Warning: **Enrolling in this course is at your own risk!**

The brilliant history of probability theory

Gamblers As long as human history?

Cardano 1564, "Book on Games of Chance", a founder of modern prob., Informal LLN, independence

Fermat&Pascal 1654, math. theory of probabilities, division of stakes, expected values, argument of belief in God

Huygens 1657, *On Reasoning in Games of Chance*, systematic treatise, **division of stakes, expected values**

Jacob Bernoulli 1713, *Ars Conjectandi*, a sound mathematical footing, Law of large numbers, Bernoulli trials

de Moivre 1718, *The Doctrine of Chances*, a sound mathematical footing

Gauss 18xx, application in astronomy, normal distribution

Laplace 1812, *Theorie analytique des probabilites*, fundamental results: PGF, MLS, CLT

...

Kolmogorov 1933, *Foundations of the Theory of Probability*, modern axiomatic foundations

...

Laplace(1745-1827)

Probability theory is nothing but
a formulation of common sense



Advice from this book: Part of the research process in random processes is first to **understand what is going on at a high level** and then to use this understanding in order to develop formal mathematical proofs. ...To gain insight, you should perform experiments based on **writing code to simulate** the processes.

Why probability in CS: two fundamental ways

Algorithm design

- Randomized
- Probability-theory-based: statistical, derandomized ...
- Quantum computing

Algorithm analysis

- Average complexity
- Smoothed complexity:
Spielman and Teng
- Learning theory



No probability, no viability!

Probability axioms and basic properties

A probability space (modeling a random process) has 3 elements

Sample space $\Omega \neq \emptyset$ The set of possible outcomes

Event family $\mathcal{F} \subseteq 2^\Omega$ The set of eligible events, a σ -algebra

Prob. function $\Pr : \mathcal{F} \rightarrow \mathbb{R}$ The *likelihood* of the events

\Pr satisfies 3 conditions:

- $\text{Range}(\Pr) \subseteq [0, 1]$
- $\Pr(\Omega) = 1$
- $\Pr(\bigcup_{i \geq 1} E_i) = \sum_{i \geq 1} \Pr(E_i)$ if the **countably many** events are **mutually disjoint**

Remarks

- We mainly consider the discrete case with $\mathcal{F} = 2^\Omega$
- Events are sets, so Venn diagrams will be used for intuition

Coin flip

- $\Omega = \{H, T\}$
- $\mathcal{F} = 2^\Omega$
- $\Pr(\{H\}) = p, \Pr(\{T\}) = 1 - p$
 $\Pr(\Omega) = 1, \Pr(\emptyset) = 0$



Coin flip

- $\Omega = \{H, T\}$
- $\mathcal{F} = 2^\Omega$
- $\Pr(\{H\}) = p, \Pr(\{T\}) = 1 - p$
 $\Pr(\Omega) = 1, \Pr(\emptyset) = 0$



An example probability space

Coin flip

- $\Omega = \{H, T\}$
- $\mathcal{F} = 2^\Omega$
- $\Pr(\{H\}) = p, \Pr(\{T\}) = 1 - p$
 $\Pr(\Omega) = 1, \Pr(\emptyset) = 0$



$p = 1/2$ if the coin is unbiased.

Union bound

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$$

Inclusion-exclusion principle

$$\Pr\left(\bigcup_{i \geq 1}^n E_i\right) = \sum_{l=1}^n (-1)^{l-1} \sum_{i_1 < i_2 < \dots < i_l} \Pr\left(\bigcap_{r=1}^l E_{i_r}\right)$$

Union bound (Boole's Inequality)

$$\Pr\left(\bigcup_{i \geq 1} E_i\right) \leq \sum_{i \geq 1} \Pr(E_i)$$

Bonferroni Inequalities

- $\Pr\left(\bigcup_{i \geq 1}^n E_i\right) \leq \sum_{l=1}^r (-1)^{l-1} \sum_{i_1 < i_2 < \dots < i_l} \Pr\left(\bigcap_{r=1}^l E_{i_r}\right)$ for odd r
- $\Pr\left(\bigcup_{i \geq 1}^n E_i\right) \geq \sum_{l=1}^r (-1)^{l-1} \sum_{i_1 < i_2 < \dots < i_l} \Pr\left(\bigcap_{r=1}^l E_{i_r}\right)$ for even r

Definition: independent events

- $\Pr(E \cap F) = \Pr(E) \Pr(F)$
- Events E_1, E_2, \dots, E_k are **mutually** independent if for any $I \subseteq [1, k]$, $\Pr(\bigcap_{i \in I} E_i) = \prod_{i \in I} \Pr(E_i)$

Independence and conditional probability

Definition: independent events

- $\Pr(E \cap F) = \Pr(E) \Pr(F)$
- Events E_1, E_2, \dots, E_k are **mutually** independent if for any $I \subseteq [1, k]$, $\Pr(\bigcap_{i \in I} E_i) = \prod_{i \in I} \Pr(E_i)$

Definition: conditional probability

- $\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)}$, well-defined if $\Pr(F) \neq 0$
- Conditioning changes/restricts the sample space
- Probability changes when more information is available

Independence and conditional probability

Definition: independent events

- $\Pr(E \cap F) = \Pr(E) \Pr(F)$
- Events E_1, E_2, \dots, E_k are **mutually** independent if for any $I \subseteq [1, k]$, $\Pr(\bigcap_{i \in I} E_i) = \prod_{i \in I} \Pr(E_i)$

Definition: conditional probability

- $\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)}$, well-defined if $\Pr(F) \neq 0$
- Conditioning changes/restricts the sample space
- Probability changes when more information is available

Corollary

- $\Pr(E|F) = \Pr(E)$ if E and F are independent
- Independence means that the probability of one event is not affected by the information on the other
- Chain rule: $\Pr(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n \Pr(A_i | \bigcap_{j=1}^{i-1} A_j)$

Law of total probability

If E_1, E_2, \dots, E_n are mutually disjoint and $\bigcup_{i=1}^n E_i = \Omega$, then

$$\Pr(B) = \sum_{i=1}^n \Pr(B \cap E_i) = \sum_{i=1}^n \Pr(B|E_i) \Pr(E_i).$$

Example

Find the probability that the sum of n dice is divisible by 6.

Law of total probability

If E_1, E_2, \dots, E_n are mutually disjoint and $\bigcup_{i=1}^n E_i = \Omega$, then
$$\Pr(B) = \sum_{i=1}^n \Pr(B \cap E_i) = \sum_{i=1}^n \Pr(B|E_i) \Pr(E_i).$$

Example

Find the probability that the sum of n dice is divisible by 6.

Solution:

- X_k : the result of the k -th roll for $1 \leq k \leq n$
- $Y_k = \sum_{i=1}^k X_i$ for $1 \leq k \leq n$
- $\Pr(Y_n \equiv 0 \pmod{6}) = \sum_{i=1}^6 \Pr((Y_n \equiv 0 \pmod{6}) \cap (X_n = i))$
- Claim: $\Pr((Y_n \equiv 0 \pmod{6}) \cap (X_n = i))$
$$= \Pr(Y_{n-1} \equiv 6 - i \pmod{6}) \Pr(X_n = i)$$

Law of total probability

If E_1, E_2, \dots, E_n are mutually disjoint and $\bigcup_{i=1}^n E_i = \Omega$, then

$$\Pr(B) = \sum_{i=1}^n \Pr(B \cap E_i) = \sum_{i=1}^n \Pr(B|E_i) \Pr(E_i).$$

Bayes' Law

If E_1, E_2, \dots, E_n are mutually disjoint and $\bigcup_{i=1}^n E_i = \Omega$, then

$$\Pr(E_j|B) = \frac{\Pr(B|E_j) \Pr(E_j)}{\Pr(B)} = \frac{\Pr(B|E_j) \Pr(E_j)}{\sum_{i=1}^n \Pr(B|E_i) \Pr(E_i)}.$$

It is time to solve a BIG Problem!

- Monty Hall problem
- First appeared at *Ask Marilyn* column of Parade, 9.9.1990
- See the demo
- Named after the celebrated TV host Monty Hall
- Confusing, so that formal proofs are not convincing (Paul Erdos & Andrew Vazsonyi)
- What's your answer?

Marilyn in 2017



Monty in 1970'



Solution to Monty Hall problem

Proof

- Reference for a formal proof: The Monty Hall Problem, by Afra Zomorodian, 1998
- An intuitive proof: keeping for one door but switching for two

Solution to Monty Hall problem

Proof

- Reference for a formal proof: The Monty Hall Problem, by Afra Zomorodian, 1998
- An intuitive proof: keeping for one door but switching for two

God is fair: smart Miss Marilyn made silly mistakes

- January 22, 2012: How likely are you chosen over one year?
- May 5, 2013: How many 4-digit briefcase combinations contain a particular digit?
- June 22, 2014: How many work hours is necessary?
6 together, but a 4-hour gap for each
- January 25, 2015: Which salary options do you prefer?
Annual \$1000 or semi-annual \$300 raises

Random variables and expectation

Random variable

- A real-valued function on the sample space of a probability space, $X : \Omega \rightarrow R$
- Random variables on this same probability space have both functional operations and probability operations

Random variables and expectation

Random variable

- A real-valued function on the sample space of a probability space, $X : \Omega \rightarrow R$
- Random variables on this same probability space have both functional operations and probability operations

Probability of a random variable

- $X = a$ stands for the event $\{s \in \Omega | X(s) = a\}$
- $\Pr(X = a) = \sum_{s \in \Omega: X(s)=a} \Pr(s)$

Random variables and expectation

Random variable

- A real-valued function on the sample space of a probability space, $X : \Omega \rightarrow R$
- Random variables on this same probability space have both functional operations and probability operations

Probability of a random variable

- $X = a$ stands for the event $\{s \in \Omega | X(s) = a\}$
- $\Pr(X = a) = \sum_{s \in \Omega: X(s)=a} \Pr(s)$

Independent random variables

- $\Pr((X = x) \cap (Y = y)) = \Pr(X = x) \Pr(Y = y)$
- Gengerally, $\Pr(\bigcap_{i \in I} (X_i = x_i)) = \prod_{i \in I} \Pr(X_i = x_i)$ for any I

Expectation: a basic characteristic

Definition

- $\mathbb{E}[X] = \sum_{i \in \text{Range}(X)} i * \Pr(X = i)$
- It's finite if $\sum_{i \in \text{Range}(X)} |i| * \Pr(X = i)$ converges

Expectation: a basic characteristic

Definition

- $\mathbb{E}[X] = \sum_{i \in \text{Range}(X)} i * \Pr(X = i)$
- It's finite if $\sum_{i \in \text{Range}(X)} |i| * \Pr(X = i)$ converges

Linearity of expectation

- $\mathbb{E}[\sum_{i=1}^n a_i X_i] = \sum_{i=1}^n a_i \mathbb{E}[X_i]$
- No independence is required
- The only condition is that each $\mathbb{E}[X_i]$ is bounded
- The most important property of expectation!

Expectation: a basic characteristic

Definition

- $\mathbb{E}[X] = \sum_{i \in \text{Range}(X)} i * \Pr(X = i)$
- It's finite if $\sum_{i \in \text{Range}(X)} |i| * \Pr(X = i)$ converges

Linearity of expectation

- $\mathbb{E}[\sum_{i=1}^n a_i X_i] = \sum_{i=1}^n a_i \mathbb{E}[X_i]$
- No independence is required
- The only condition is that each $\mathbb{E}[X_i]$ is bounded
- The most important property of expectation!

Product Counterpart

$\mathbb{E}[X * Y] = \mathbb{E}[X]\mathbb{E}[Y]$ if they are **independent**.

Bernoulli random variable

- $\Pr(X = 1) = p, \Pr(X = 0) = 1 - p$
- Modeling coin flip
- $\mathbb{E}[X] = p * 1 + (1 - p) * 0 = p$
- $X^k = X$

Bernoulli distribution

Bernoulli random variable

- $\Pr(X = 1) = p, \Pr(X = 0) = 1 - p$
- Modeling coin flip
- $\mathbb{E}[X] = p * 1 + (1 - p) * 0 = p$
- $X^k = X$

An example

How many triangles among 4 nodes when the links appear independently randomly?

Binomial random variable

- The number of successes in n independent trials of the Bernoulli experiment with success probability p
- For any $0 \leq i \leq n$, $\Pr(X = i) = C_n^i p^i (1 - p)^{n-i}$
- $X = \sum_{i=1}^n X_i$
- $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = np$

Binomial distribution

Binomial random variable

- The number of successes in n independent trials of the Bernoulli experiment with success probability p
- For any $0 \leq i \leq n$, $\Pr(X = i) = C_n^i p^i (1 - p)^{n-i}$
- $X = \sum_{i=1}^n X_i$
- $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = np$

Application

Router packets sampling

Geometric distribution

The story of Farmer & Rabbit



Geometric distribution

The story of Farmer&Rabbit

Geometric random variable

- The number of independent trials until success, where each trial has success probability p
- $\Pr(X = i) = (1 - p)^{i-1}p$ for $i \geq 1$
- $\mathbb{E}[X] = \sum_{i \geq 1} i(1 - p)^{i-1}p = 1/p$

Geometric distribution

The story of Farmer&Rabbit

Geometric random variable

- The number of independent trials until success, where each trial has success probability p
- $\Pr(X = i) = (1 - p)^{i-1}p$ for $i \geq 1$
- $\mathbb{E}[X] = \sum_{i \geq 1} i(1 - p)^{i-1}p = 1/p$

Suppose the daily probability that God throws a rabbit at the trunk is 10^{-4} . How many years does the farmer has to wait?

Geometric distribution

The story of Farmer&Rabbit

Geometric random variable

- The number of independent trials until success, where each trial has success probability p
- $\Pr(X = i) = (1 - p)^{i-1}p$ for $i \geq 1$
- $\mathbb{E}[X] = \sum_{i \geq 1} i(1 - p)^{i-1}p = 1/p$

Suppose the daily probability that God throws a rabbit at the trunk is 10^{-4} . How many years does the farmer has to wait?

Memoryless: particular to geometric distribution

For geometric random variable X , if $n > 0$,
 $\Pr(X = n + k | X > k) = \Pr(X = n)$

Application: coupon collector's problem

Problem statement

The # of boxes of milk to buy before you have all coupon types?

Application: coupon collector's problem

Problem statement

The # of boxes of milk to buy before you have all coupon types?

- X_i : the number of boxes bought while you have $i - 1$ types of coupons until you get the i th type
- $X = \sum_{i=1}^n X_i$
- X_i : geometric random variable with parameter $p_i = 1 - \frac{i-1}{n}$
- $\mathbb{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$

Application: coupon collector's problem

Problem statement

The # of boxes of milk to buy before you have all coupon types?

- X_i : the number of boxes bought while you have $i - 1$ types of coupons until you get the i th type
- $X = \sum_{i=1}^n X_i$
- X_i : geometric random variable with parameter $p_i = 1 - \frac{i-1}{n}$
- $\mathbb{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = n \sum_{i=1}^n \frac{1}{i} = n \ln n + \Theta(n).$$

Thanks!