Chapter 7
Discrete Probability

Discrete Structures for Computing on September 30, 2021

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang
Faculty of Computer Science and Engineering
University of Technology - VNUHCM
trtanh@hcmut.edu.vn - htnguyen@hcmut.edu.vn

We would like to thank Dr. Nguyen Van Minh Man from Mahidol University, Thailand for his contribution to this chapter.

Contents

Course outcomes

	Course learning outcomes
L.O.1	Understanding of logic and discrete structures
	L.O.1.1 – Describe definition of propositional and predicate logic L.O.1.2 – Define basic discrete structures: set, mapping, graphs
L.O.2	Represent and model practical problems with discrete structures
	L.O.2.1 – Logically describe some problems arising in Computing L.O.2.2 – Use proving methods: direct, contrapositive, induction L.O.2.3 – Explain problem modeling using discrete structures
L.O.3	Understanding of basic probability and random variables
	L.O.3.1 – Define basic probability theory L.O.3.2 – Explain discrete random variables
L.O.4	Compute quantities of discrete structures and probabilities
	L.O.4.1 – Operate (compute/ optimize) on discrete structures L.O.4.2 – Compute probabilities of various events, conditional ones, Bayes theorem

Contents

Motivations

- **Gambling**

- **Real life problems**

- **Computer Science**: cryptology – deals with encrypting codes or the design of error correcting codes

Learning Objectives of Basic Probability

After careful study of basic probability, you should be able to

1. **Compute and interpret** probability of an event (Probability = a chance that an event occurs)
2. **Find** probability of mixed events, including conditional probability
3. **Explain** the concepts of random variable, and determine its components
4. **Understand** expectation and variance of a random variable

Learning Objectives of Basic Probability

Probability in practice

Probabilistic Methods help us to solve realistic problems as follows.

1) In Computing

- Estimate the average computing time of an algorithm
- Compute the risk of security breach to a computer system by (bad) hackers in cyber-space...

2) In Finance Science and Insurance Industry

- Suppose an insurance company B has thousands of customers, and each customer is charged \$500 a year.
- Since the customer's businesses are risky, from the past experience the company B estimates that about 15% of their customers would get fatal trouble (e.g. fire, accident ...) and, as a result they will submit a claim in any given year.
- We assume that the claim will always be \$3000 for each customer.

Our problem: Compute **how much average profit** can the company expect to make *per customer* in a certain year?

Randomness

Which of these are random phenomena?

- The number you receive when rolling a fair dice
- The sequence for lottery special prize (by law!)
- Your blood type (No!)
- You met the red light on the way to school
  - The traffic light is not random. It has timer.
  - The pattern of your riding is random.

So what is special about randomness?

In the **long run**, they are predictable and have relative frequency (fraction of times that the event occurs over and over and over).

Does randomness exist in the above problem? What is it? Could we

1. **Model** the amount of money X that the insurance company believes to obtain from each customer
2. Find the expectation (Expected Value) of that amount, denoted by $E[X]$.

Terminology

- **Experiment** (*thí nghiệm*): a procedure that yields one of a given set of possible outcomes.
  - Tossing a coin to see the face
- **Sample space** (*không gian mẫu*): set of possible outcomes
  - {Head, Tail}
- **Event** (*sự kiện*): a subset of sample space.
  - You see Head after an experiment. {Head} is an event.

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Example

Example (1)

Experiment: Rolling a die. What is the sample space?

Answer: $\{1, 2, 3, 4, 5, 6\}$

Example (2)

Experiment: Rolling two dice. What is the sample space?

Answer: It depends on what we're going to ask!

- The total number?
  $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
- The number of each die?
  $\{(1,1), (1,2), (1,3), \ldots, (6,6)\}$

Which is better?
The latter one, because they are **equally likely outcomes**

The Law of Large Numbers

Definition

The Law of Large Numbers (*Luật số lớn*) states that the **long-run** relative frequency of repeated independent events gets closer and closer to the true relative frequency as the number of trials increases.

Example

Do you believe that the true relative frequency of Head when you toss a coin is 50%?

Let's try!

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Be Careful!

Don't misunderstand the Law of Large Numbers (LLN). It can lead to money lost and poor business decisions.

Example

I had 8 children, all of them are girls. Thanks to LLN (!?), there are high possibility that the next one will be a boy. (**Overpopulation!!!**)

Example

I'm playing *Oysters* - *Crap* - *Lobster* - *Fish* game, the fish has not appeared in recent 5 games, it will be more likely to be fish next game. Thus, I bet all my money in fish. (**Sorry, you lose!**)

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Probability

Probability of an event E, or a phenomenon ... informally is a number $\mathbb{P}[E]$ or $p(E)$ measuring how much chance that the event E would happen.

Definition

The **probability** (*xác suất*) of an event E of a finite nonempty sample space S of equally likely outcomes is:

$$\mathbb{P}[E] = \frac{|E|}{|S|}.$$

- Note that $E \subseteq S$ so $0 \leq |E| \leq |S|$
- $0 \leq \mathbb{P}[E] \leq 1$
  - 0 indicates impossibility
  - 1 indicates certainty

People often say: "It has a 20% probability of female students in class DS1007" if there are 20 women over a total of 100 students.

Examples (We use both notation $\mathbb{P}[E]$ and $p(E)$)

Example (1)

What is the probability of getting a Head when tossing a coin?

Answer:

- There are $|S| = 2$ possible outcomes
- Getting a Head is $|E| = 1$ outcome, so $p(E) = 1/2 = 0.5 = 50\%$

Example (2)

What is the probability of getting a 7 by rolling two dice?

Answer:

- Product rule: There are a total of 36 equally likely possible outcomes
- There are six successful outcomes: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)
- Thus, $|E| = 6$, $|S| = 36$, $p(E) = 6/36 = 1/6$

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Examples

Example (3)

We toss a coin 6 times. What is probability of H in 6th toss, if all the previous 5 are T?

Answer:
Don't be silly! Still 1/2.

Example (4)

Which is more likely:

- Rolling an 8 when 2 dice are rolled?
- Rolling an 8 when 3 dice are rolled?

Answer:
Two dice: $5/36 \approx 0.139$
Three dice: $21/216 \approx 0.097$

Summary I: Conceptual terms of probability

Experiment is a specific trial/activity (of scientists, human being) whose outcomes possess randomness.

Sample space- set of all possible outcomes, denoted by S.

Event- is subset E of sample space S: $E \subset S$. Usually we include all events into a set $\mathcal{Q} := \{\text{events} \;\; E : E \subset S\}$, called the **event set** or σ-algebra. So $\mathcal{Q} \subset 2^S$, the power set of S.

Probability function: is a map $\mathbb{P} : \mathcal{Q} \to [0, 1]$,

$$E \in \mathcal{Q} \Longrightarrow p(E) = \mathbb{P}[E] = \operatorname{Prob}(E) =$$

probability or chance that the event E occurs.

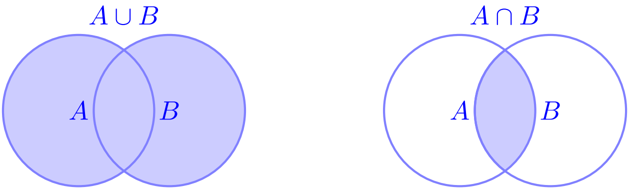
Remarks:

1. S can be a discrete set or continuous set. [DIY: could you invent few examples?]
2. We use 3 notations $p(E), \mathbb{P}[E]$ and $\operatorname{Prob}(E)$ equivalently from now on.

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Event combinations

- **The Union** of two events A, B, denoted as $A \cup B$, consists of all outcomes that are contained in one event *or* the other.
- **The Intersection** of two events A, B, denoted as $A \cap B$, consists of all outcomes that are contained in one event *and* the other.



- **The Complement** of an event B, denoted as B^c, B' or $\overline{B}$, is the set of outcomes in the sample space that are *not* contained in B.

Formal Probability (Axioms, by A. Kolmogorov, 1933)

Axiom A1: A probability is a number between 0 and 1.

$$0 \leq \mathbb{P}[E] \leq 1$$

Axiom A2: The sample space S has probability 1

The probability of S (all possible outcomes of a trial) must be 1.

$$\mathbb{P}[S] = 1$$

Axiom A3: Probabilities of disjoint events $A, B \subset \Omega$, and $A \cap B = \emptyset$:

$$\mathbb{P}[A \cup B] = \mathbb{P}[A \text{ or } B] = \mathbb{P}[A] + \mathbb{P}[B]$$

Axiom A3b: Compliment Rule (obtained from Axioms A2 and A3)

The probability of an event B occurring is 1 minus the probability that it doesn't occur $\mathbb{P}[B] = 1 - \mathbb{P}[\overline{B}]$.

Examples

Example

What is the probability of NOT drawing a heart card from 52 deck cards?

Answer:
Let E be the event of getting a heart from 52 deck cards.
We have $p(E) = 13/52 = 1/4$.

By the compliment rule, the probability of NOT getting a heart card is

$$p(\overline{E}) = 1 - p(E) = 3/4.$$

Extensions of Axiom A3

General Addition Rule, the 1st extension of Axiom A3:

For nonmutually exclusive or overlapping events, $A \cap B \neq \emptyset$,

since $\mathbb{P}[A \text{ and } B] = \mathbb{P}[A \cap B] \neq 0$ then:

$$\mathbb{P}[A \cup B] = \mathbb{P}[A \text{ or } B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

If $A \cap B = \emptyset$, they are disjoint,
which means they can't occur together, then,

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B].$$

What if we have more than two events?

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Examples

Example (1)

If you choose a number between 1 and 100, what is the probability that it is divisible by either 2 or 5?

Short Answer: $\frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{3}{5}$

Example (2)

There are a survey that about
45% of VN population has Type O blood,
40% type A, 11% type B and
the rest type AB.

What is the probability that a blood donor has Type A or Type B?

Short Answer: 40% + 11% = 51%. WHY?

Axiom A3. Probabilities of disjoint events

- The 2nd extension of Axiom A3, generally, says

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_m) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_m)$$

  for m *mutually disjoint* events, i.e. the intersection $A_i \cap A_j = \emptyset$ when $1 \leq i \neq j \leq m$.

- The 3rd extension, so-called **countably additive of probabilities** is a generalization:

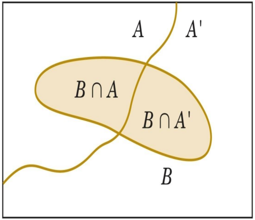
$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

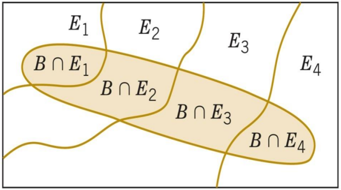
NOTE: Here events $A_i \subset S$, and the sample space S can be either

* discrete if it consists of a *finite* or *countable infinite*.

Partitioning sample space

Observe that $A \cup A' = \Omega$, and $B = B \cap \Omega = B \cap (A \cup A') = ?$.





$B = (B \cap E_1) \cup (B \cap E_2) \cup (B \cap E_3) \cup (B \cap E_4)$

Events $E_1, E_2, \cdots, E_n$ $(n \geq 1)$ ($n = 4$ in the above figure) form a **partition** of set Ω if

$$(E_i \cap E_j) = \emptyset, \ \forall i \neq j, \quad \text{and} \quad E_1 \cup E_2 \cup \cdots \cup E_n = S.$$

General addition rule for n events

If events $E_1, E_2, \cdots, E_n$ $(n \geq 1)$ form a **partition** of set Ω, then for any event B,

$$\mathbb{P}[B] = \sum_{i}^{n} \mathbb{P}[B \cap E_i]. \quad (1)$$

Homework (DIY)

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Example (3)

In a hospital unit there are 8 nurses and 5 physicians; 7 nurses and 3 physicians are females. If a staff person is selected, find the probability that the subject is a nurse or a male.

The sample space is shown here:

Staff	Females	Male	Total
Nurses	7	1	8
Physicians	3	2	5
Total	10	3	13

The probability is:

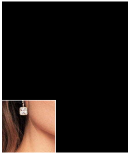
P(nurse or male) = P(nurse) +P(male) - P(male nurse)

$= \frac{8}{13} + \frac{3}{13} - \frac{1}{13} = \frac{10}{13}$

Dependent events

“Knowledge” changes probabilities!

Event A and event B are **dependent** if the appearance of an event is related (in a certain way) to the occurrence of the other event.



We next use **event dependency** to define Independence (*độc lập*), then Conditional Probability (*Xác suất có điều kiện*), given two events A, B, where B occurred.

Independent events

Definition

Events A and B are independent whenever the occurrence of A is *not related to or not influenced by* the occurrence of B, and the other way round.
Then we write

$$p(A \mid B) = p(A).$$

- We can say the outcome of one event does not influence the probability of the other.
- Example: $p($"Head"$|$"It's raining outside"$) = p($"Head"$)$

NOTE: Disjoint $\neq$ Independence

Disjoint events cannot be independent. They have no outcomes in common, so knowing that one occurred means the other did not.

Conditional Probability

Definition: The conditional probability $p(A \mid B)$ = Probability of event A given that event B **has occurred**.
Two cases of interest:

1. When event A is independent of B, meaning the occurrence of A is *not influenced by* B, then obviously

$$p(A \mid B) = p(A).$$

The **joint probability** of two independent events A, B then is defined as

$$p(A \cap B) = p(B) \cdot p(A) = p(A) \cdot p(B). \qquad (2)$$

2. When A is dependent of B (A is influenced by B)

$$p(A \mid B) \neq p(A).$$

How to evaluate $p(A \mid B)$ in the 2nd case?

General Multiplication Rule gives Conditional Probability

In general, [no matter dependent or independent] we always have

$$p(A \cap B) \quad = \quad p(B) \times p(A \mid B) = \; p(A) \times p(B \mid A)$$

Therefore, the **conditional probability** of event A given event B

$$p(A \,|\, B) = \frac{p(A \cap B)}{p(B)}. \tag{3}$$

Example

What is the probability of drawing a red card and then another red card without replacement (*không hoàn lại*)?

Solution: Let B be the event of drawing the first red card, and A be the event of drawing the second red card. Clearly, $p(B) = 26/52 = 1/2$, the conditional $p(A \mid B) = 25/51$.
So the event of drawing a red card and then another red card is the **joint probability**
$p(A \cap B) = p(B) \times p(A \mid B) = 1/2 \times 25/51 = 25/102.$

Example

Example (1)

The probability that Sam parks in a no-parking zone and gets a parking ticket is 0.06, and the probability that Sam cannot find a legal parking space and has to park in the no-parking zone is 0.20. On Tuesday, Sam arrives at school and has to park in a no-parking zone.
Find the probability that he will get a parking ticket.

Solution

N: *parking in a no-parking zone*
T: *get a ticket*
$p(T|N) = \frac{p(N \cap T)}{p(N)} = \frac{0.06}{0.2} = 0.3$
Hence, Sam has a 0.3 probability of getting a parking ticket, given that he parked in a no-parking zone.

REMARK: In general, $p(A \mid B) \neq p(B \mid A)$,
the conditional probability is not symmetric.

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Bayes's Theorem

Reminder:

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}.$$

Note also that

$$p(B) \cdot p(A \mid B) = p(A \cap B) = p(B \cap A) = p(A) \cdot p(B \mid A)$$

Theorem (Bayes's Theorem- basic version)

We always have the following, for any pair of events A, B, where we assume B occurred.

$$p(A \mid B) = \frac{p(A) \cdot p(B \mid A)}{p(B)}. \tag{4}$$

Recall $\overline{A} = A^c = A' = S \setminus A$, the complement of event A.

Bayes's Theorem II

Theorem (Bayes's Theorem- full version)

Given two events B, A where we assume B occurred.

$$p(A \mid B) = \frac{p(A) \cdot p(B \mid A)}{p(A) \cdot p(B \mid A) + p(\overline{A}) \cdot p(B \mid \overline{A})}$$

Proof: Rule (1) gives
$p(B) = p(B \cap E_1) + p(B \cap E_2)$, with $E_1 = A, E_2 = \overline{A}$.

Example

If we know that the probability that a person has tuberculosis (TB) is $p(\mathrm{TB}) = 0.0005$, also know $p(+|\ \mathrm{TB}) = 0.999$ and $p(-|\ \overline{\mathrm{TB}}) = 0.99$. What is $p(\mathrm{TB}|\ +)$ and $p(\overline{\mathrm{TB}}|\ -)$?

$$p(TB|+) = \frac{p(+|TB)p(TB)}{p(+|TB)p(TB) + p(+|\overline{TB})p(\overline{TB})}$$
$$= \frac{0.999 \times 0.0005}{0.999 \times 0.0005 + (1 - 0.99) \times (1 - 0.0005)} = 0.0476$$

$p(\overline{TB}|-) = 0.99$

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Random Variable: Concepts

From now on, given a random experiment, we also write Ω for the sample space (previously denoted S) of that experiment. Then we perform observation or measurement on elements of Ω.

Definition

A *random variable* X is a map (function or measurement) from Ω to the reals $\mathbb{R}$. That is for $w \in \Omega$ then $X(w) \in \mathbb{R}$.

- The **domain** of variable X is the sample space Ω.
- The **range** of variable X is the set of all observed values

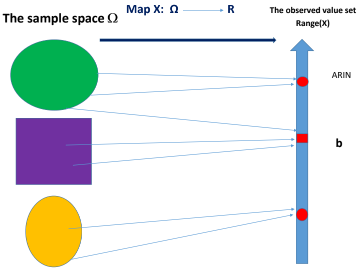
$$R_X = \text{Range}(X) = \{X(w)\} \subseteq \mathbb{R}.$$

For any value $b \in \mathbb{R}$, the preimage

$$A := X^{-1}(b) = \{w \in \Omega : \ X(w) = b\} \subset \Omega$$

is an event, we define $\mathbb{P}[X = b] = \text{Prob}\{X = b\} := \text{Prob}(A)$.

Random Variable: Visualization of a random variable X



In Entertainment industry, a random experiment can be defined as observing certain students of HCMUT, to get a sample space Ω of students who like music (or art and drama).
Next we ask every $w \in \Omega$ about his/her unique favorite singer, to make a random variable "fan of"

$$X : \Omega \longrightarrow \mathrm{Range}(X) = \{x : x \text{ is good singer in the world}\}.$$

Here specifically $X(w) = b$ means b is the observed value at w.
Then event $\{X = b\}$ is $A := X^{-1}(b)$, the violet square...

Random Variable: the probability density function

Now fix another singer with name ARIN, we then view the event $\{X = \text{ARIN}\}$ in the meaning of the set $A := X^{-1}(\text{ARIN})$, it is the green oval consisting of fans of ARIN, obviously $A \subseteq \Omega$, so the probability $\mathrm{Prob}(A)$ does exist.

Definition

In general, for a random variable $X : \Omega \longrightarrow \mathrm{Range}(X)$, fix any value $x \in \mathrm{Range}(X)$, then $A := X^{-1}(x)$ is an event and the probability

$$\mathbb{P}[X = x] := \mathrm{Prob}(A) = \frac{|A|}{|\Omega|}.$$

- $\mathbb{P}[X = x]$ indicates **how much chance observed value x occurred**, and
- $f(x) := \mathbb{P}[X = x]$ called the **probability density function** (**p.d.f**) of variable X at observation (or observed value) x.

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

Random Variable: Example revisited with calculation

- Sample space Ω = the set of HCMUT 'musical' students, ask each student $w \in \Omega$ to know his (her) most liked singer x in the globe, $X(w) = x$, [x is the most liked singer of w.]
- If the university has 30000 students, and 1500 students like - say singer $x = `ARIN'$ (called observed value or measured value) then event

$$A := X^{-1}(\text{`ARIN'}) = \{w \in \Omega : \ X(w) = \text{`ARIN'}\} \subset \Omega$$

has cardinality (i.e. the number of students) 1500, and

$$\mathbb{P}[X = \text{`ARIN'}] := \mathrm{Prob}(A) = \frac{|A|}{|\Omega|} = \frac{1500}{30000} = 1/20.$$

$f(\text{`ARIN'}) := \mathbb{P}[X = \text{`ARIN'}] = 1/20$ is
the **probability density or mass** at the value 'singer ARIN'.

Distinguish random variables and observed values

Notation is used to distinguish between a random variable and the real observed values.

- A random variable is denoted by an **uppercase letter** such as X (it is a map).
- After a measurement (observation) is conducted, the *measured value* of the random variable is denoted generally by a **lowercase letter** x (it is number or specific value), as

  $x =$ ADEL in Western entertainment industry,
  $x =$ DONAL Trump in Politics,

  $x = 70$ milliamperes in Electrical engineering,
  $x = 1.7$ meter in Public Health,

  $x = \text{`}success\text{'}$ outcome in Business administration,

  $x = \text{`}fatal\ accident\text{'}$ in Medical insurance...

Discrete random variable and Expected Value

Definition (Discrete random variable - biến ngẫu nhiên rời rạc)

(1) A rand. variable $X(.)$ is discrete if it has a **discrete range**, meaning that the set of observed values consists of no more than *a countable number of elements*, i.e. $|S_X| \leq |\mathbb{N}|$.

In **finite set** case, $|S_X| < |\mathbb{N}|$, we usually write set of values

$$S_X := \text{Range}(X) = \{x_1, x_2, x_3, \cdots, x_{n-1}, x_n\}, \; n \in \mathbb{N}.$$

In **countably infinite set** case, $|S_X| = |\mathbb{N}|$, we write

$$S_X := \text{Range}(X) = \{x_1, x_2, x_3, \cdots, x_{n-1}, x_n, \ldots\}$$

(2) Expected value (giá trị kỳ vọng) of X

$$\mathbf{E}[X] = \sum_{x \in S_X} \; x \cdot p(X = x)$$

where $p(X = x) := \mathbb{P}[X = x]$ is the density (mass) at x.

The probability distribution table of a discrete variable

The finite $S_X = \{x_1, x_2, x_3, \cdots, x_n\}$ helps us to compute the **probability distribution table** of X, structured by

X	x_1	$\cdots$	x_{n-1}	x_n
$p_k := \mathbb{P}[X = x_k]$	p_1	$\cdots$	p_{n-1}	p_n.

Of course

$$p_k \geq 0 \text{ and } \sum_{x_k \in S_X} p_k = 1.$$

The probability distribution table in *Entertainment industry*:

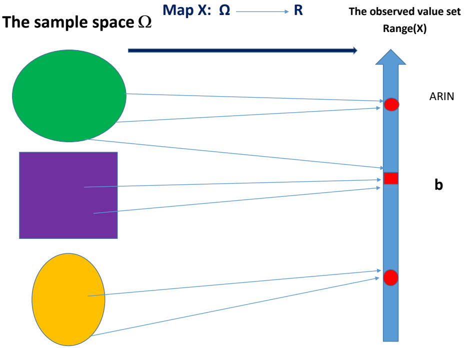
$A := X^{-1}(\text{ARIN})$ is the green oval, with $|A| = 1500$,
$E := X^{-1}(b)$ is the violet square on the left, with $|E| = 21000$,
and $C := X^{-1}(\text{c})$ is the yellow oval, with $|C| = 7500$ students,

then we can find the densities p_k and finally
check that $\sum_{x_k \in S_X} p_k = 1$?

Huynh Tuong Nguyen, Tran Tuan Anh, Nguyen An Khuong, Le Hong Trang

The set of observed values (also called **value space**)
$S_X = \{x_1, x_2, x_3\} = \{\text{ARIN}, b, c\}$ gives

X	$x_1 = \text{ARIN}$	$x_2 = b$	$x_3 = c$
$p_i = \mathbb{P}[X = x_i]$	$p_1 = \frac{1}{20}$	$p_2 = \frac{7}{10}$	$p_3 = \frac{1}{4}$



Expectation (Expected Value) to Variability

Generally, assume that random variable X is discrete, with the value space $S_X = \mathrm{Range}(X) = \{x_1, x_2, \ldots, x_n\}$, having the density $p_k = p(x_k)$, then we knew the expected value (expectation or average) μ of X is given by

$$\mu = \mathbf{E}[X] = \sum_{x_k \in S_X} x_k\, p_k.$$

Definition

The **variability** of the random variable X is expressed via two concepts.

- Variance of X is

$$\mathbf{V}[X] = \sigma^2 := \mathbf{E}[(X-\mu)^2] = \sum_{x_k \in S_X} [x_k - \mu]^2\, p_k.$$

- Standard deviation of X is $\sigma_X = \sigma = \sqrt{\mathbf{V}[X]}$.

Expected Value for Central Tendency of X

Example: An insurance company $LifeInsu$ charges \$50 a year. Can the company make a profit? Assuming that it made a research on 1000 people and have following table:

Outcome	**Payroll** (m)	**Probability** $p(X = x) = p(M = m)$
Death	10,000	$\frac{1}{1000}$
Disability	5000	$\frac{2}{1000}$
Good	0	$\frac{997}{1000}$

Or horizontally

X	Death	Disability	Good
M (Payroll)	10,000	5000	0
$p_i = \mathbb{P}[X = x_i]$	$p_1 = \frac{1}{1000}$	$p_2 = \frac{2}{10000}$	$p_3 = \frac{997}{1000}$

X is a discrete random variable, so is the payroll M.
Besides, $p_i = \mathbb{P}[X = x_i] = \mathbb{P}[M = m_i]$, where $m_1 = 10,000$, $m_2 = 5000$, and $m_3 = 0$.

Can company make a profit?

- X is discrete since customer's injury level has three values (outcomes) Death, Disability and Good (nothing happen).
- The company **expects** that they have to pay for **each customer**:

$$\mathbf{E}[X] = \sum x \cdot \mathbb{P}(X = x) = \sum m \cdot \mathbb{P}(M = m)$$

$$= \$10,000\ (\frac{1}{1000}) + \$5000\ (\frac{2}{1000}) + \$0\ (\frac{997}{1000}) = \$20.$$

X M (Payroll)	Death 10,000	Disability 5000	Good 0
$p_i = \mathbb{P}[M = m_i]$	$p_1 = \frac{1}{1000}$	$p_2 = \frac{2}{10000}$	$p_3 = \frac{997}{1000}$

Expected money amount $\mathbf{E}[X] = \mathbf{E}[M] = \mu = 20$ USD per customer. Well, shall we say $LifeInsu$ gets profit $50 - 20 = \$30$ per customer, in average?

Variance: capture the Spreading, Risk or Variability

- Of course, the expected value \$20 will not happen in reality
- There will be **variability**. Let's calculate!

- Variance (*phương sai*)
  $\mathbf{V}[X] = \sum (x - \mathbf{E}[X])^2 \cdot p(X = x)$

$$\mathbf{V}[X] = \sum (m - \mathbf{E}[M])^2 \cdot p(M = m) = \mathbf{V}[M]$$

  - $\mathbf{V}[M] = 9980^2(\frac{1}{1000}) + 4980^2(\frac{2}{1000}) + (-20)^2(\frac{997}{1000}) = 149,600$
- Standard deviation (*độ lệch chuẩn*)
  $SD(M) = \sqrt{\mathbf{V}[M]}$
  - $SD(M) = \sqrt{149,600} \approx \$386.78$

Comment

The company expects to pay out \$20, and make \$30. However, the standard deviation of \$386.78 indicates that it's no sure thing. That's pretty big spread (and risk) for an average profit of \$30.

Example

Example

One thousand tickets are sold at 1 each for a color television valued at 350. What is the expected value of the gain if you purchase one ticket?

The problem can be set up as follows:

	Win	Lose
Gain	349	-1
probability	$\frac{1}{1000}$	$\frac{999}{1000}$

The solution, then, is

$$\mathbf{E}[X] = \$349(\frac{1}{1000}) + (-1)(\frac{999}{1000}) = -0.65$$

SUMMARY II

- Probability theory: Concepts and operations

- Random variable X with 5 key components
  1. The observed value set $\text{Range}(X) = S_X$

  2. The probability density function (pdf) $f(x)$
  3. The probability cumulative function (or cdf) $F(x)$

  4. Expectation (or average) $\mathbf{E}[X] = \mu$ of X
  5. Variance $\mathbf{V}[X] = \sigma^2$ of X, and the Standard deviation σ.

The last section is spent for **Probability Models**, which play an extremely important role nowadays in modeling of phenomena in Science, Engineering and Technology, in particular because of high uncertainty of our world (war, epidemic, pandemic, risks in cyberspace, calamity in environment, climate change...).
See SUMMARY III of emerging applications at the end for more.

Bernoulli Trials

Bernoulli variable or trial describes a random variable B that can take only two possible values, i.e. $S_B = \{0, 1\}$. Its probability density function is given by probability of success $\mathbb{P}(B = 1) = p$, and

$$\mathbb{P}(B = 0) = 1 - p \text{ for some } p \in [0, 1].$$

Expectation and variance $\mathbf{E}[B] = p; \quad \mathbf{V}[B] = p(1 - p).$

Example

Some people madly drink Coca-Cola, hoping to find a ticket to see Big Bang. Let's call tearing a bottle's label **trial** (*phép thử*):

- There are only possible outcomes (congrats or good luck)
- The probability of success, p, is the same on every trial, say 0.06 So

$$\mathbb{P}(B = 1) = p = 0.06,\ \mathbb{P}(B = 0) = 1 - p = 0.94$$

Geometric Model (Mô hình hình học)

Question: How long it will take us to achieve a success, given p, the probability of success?

Definition (Geometric probability model: Geom(p))

p = probability of success ($q = 1 - p$ = probability of failure)
X = number of trials until the first success occurs

$$p(X = x) = q^{x-1}p$$

Expected value: $\mu = \frac{1}{p}$

Standard deviation: $\sigma = \sqrt{\frac{q}{p^2}}$

Geometric Model: Example

Example

If the probability of finding a Sound Fest ticket is $p = 0.06$, how many bottles do you expect to open before you find a ticket? What is the probability that the first ticket is in one of the first four bottles?

Solution

Let X *= number of trials until a ticket is found*
We can model X *with Geom(0.06).*
$E(X) = \frac{1}{0.06} \approx 16.7$

$$\begin{aligned} P(X \leq 4) &= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\ &= (0.06) + (0.94)(0.06) + (0.94)^2(0.06) \\ &\quad +(0.94)^3(0.06) \\ &\approx 0.2193 \end{aligned}$$

Conclusion: *We expect to open 16.7 bottles to find a ticket. About 22% of time we'll find one within the first 4 bottles.*

Binomial Model (Mô hình nhị thức)

Previous Question: How long it will take us to achieve a success, given p, the probability of success?

New Question: You buy 5 Coca-Cola. What's the probability you get exactly 2 Sound Fest tickets?

Definition (Binomial probability model: Binom(n, p))

n = number of trials
p = probability of success ($q = 1 - p$ = probability of failure)
X = number of successes in n trials

$$p(X = x) = \binom{n}{x} p^x q^{n-x}$$

Expected value: $\mu = np$
Standard deviation: $\sigma = \sqrt{npq}$

Binomial Model: Example

Example

Suppose you buy 20 Coca-Cola bottles. What are the mean and standard deviation of the number of winning bottles among them? What is the probability that there are 2 or 3 tickets?

Solution

Let X = *number of tickets among* $n = 20$ *bottles*
We can model X *with Binom(20, 0.06).*
$E(X) = np = 20(0.06) = 1.2$
$SD(X) = \sqrt{npq} = \sqrt{20(0.06)(0.94)} \approx 1.96$

$$\begin{aligned} P(X = 2 \ or \ 3) &= P(X = 2) + P(X = 3) \\ &= \binom{20}{2}(0.06)^2(0.94)^{18} + \binom{20}{3}(0.06)^3(0.94)^{17} \\ &\approx 0.2246 + 0.0860 = 0.3106 \end{aligned}$$

Conclusion: *In 20 bottles, we expect to find an average of 1.2 tickets, with a sd of 1.06. About 31% of the time we'll find 2 or 3 tickets among 20 bottles.*

LECTURE 7's PROBLEMS

This part discusses Probability Theory with problem solving skills. ASSUME that students have already studied Probability theory at basic level in Lecture 7 at home, we practice the following contents today.

1. GROUP A: Experiment and sample space
2. GROUP B: Compute probability of an event
3. GROUP C: Independent events
4. GROUP D: Probability Models

GROUP A: Experiment and sample space

We consider an experiment in information transmission in banking sector.

The experiment is to **select a sequence of 5 letters for transmission** of a code in a money transfer operation.
Let $a_1, a_2, \ldots, a_5$ denote the first, ..., fifth letter chosen.
The sample space Ω is the set of all possible sequences of five letters. Formally,

$$\Omega = \{(a_1, a_2, \ldots, a_5) : \; a_i \in \{a, b, c, \cdots, x, y, z\}, \quad i = 1, \cdots, 5\}.$$

* This is a **finite sample space** containing 26^5 possible sequences of 5 letters, due to the multiplication rule in Equation **??**.
* A sample point (sample unit or element) is any such sequence $(a_1, a_2, \ldots, a_5)$ in Ω.

Quiz: Let E be the event that all the 5 letters in the sequence are the same. Describe E and find $\mathbb{P}[E]$.

GROUP B: Compute probability of an event

We knew

Rule 1: $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$ or $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

Rule 2: If events A and B are *mutually exclusive*, then

$$\mathbb{P}(A \,\mathrm{or}\, B) = \mathbb{P}(A) + \mathbb{P}(B)$$

But how to find $\mathbb{P}(A)$, for any $A \subset \Omega$?

Mostly use Counting Techniques:

1. Multiplication rule
2. Permutation rule
3. Combination rule

Each has its special purpose that must be applied properly – **the right tool for the right job**!

GROUP C: Independent events

Five identical departments are designed in a given commercial bank.
Let $E_1, E_2, \cdots, E_5$ be the events that these five departments comply with the quality specifications (non-defective, non bugs...).
Under the model of **mutual independence** the probability that all the five departments are indeed non-defective is

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_5) = \mathbb{P}(E_1)\,\mathbb{P}(E_2) \cdots \mathbb{P}(E_5).$$

Since these departments come from the same production process (i.e. construction company), we can assume that
$\mathbb{P}(E_i) = p$, all $i = 1, \cdots, 5$.
Thus, the probability that all the 5 departments are non-defective is p^5.

What is the probability that one department is defective and all the other four are non-defective?

Independent events - HINTS

What is the probability that one department is defective and all the other four are non-defective?

Let A_1 be the event that one out of five parts is defective.
In order to simplify the notation, we write the intersection of events as their product. Thus,

$$A_1 = E_1^c\ E_2\ E_3\ E_4\ E_5\ \cup E_1 E_2^c\ E_3\ E_4\ E_5\ \cup E_1 E_2 E_3^c\ E_4\ E_5$$
$$\cup\ E_1\ E_2\ E_3\ E_4^c\ E_5\ \cup E_1\ E_2\ E_3\ E_4\ E_5^c.$$

A_1 is the union of five **disjoint** events. Therefore

$$\mathbb{P}(A_1) = \mathbb{P}(E_1^c\ E_2\ E_3\ E_4\ E_5) + \cdots +$$
$$\mathbb{P}(E_1\ E_2\ E_3\ E_4\ E_5^c) = 5p^4(1-p).$$

Can you explain why?

GROUP D: Probability Models in Production Industry

Wafer Contamination in PC - Electronic Manufacturing

- Let the random variable X denote the number of wafers that need to be analyzed to detect a **large particle of contamination**.
- Assume that the probability that a wafer contains a large particle is $p = 0.01$, and that the wafers are independent.
- Determine the probability distribution of X.

QUESTION

How do we choose **suitable probability distribution**? Bernoulli, Binomial or Geometric?

Determine the probability distribution

Let y denote a wafer in which a large particle is present & let n denote a wafer in which it is absent.

The set of sample wafers we took: $\Omega = \{y, ny, nny, nnny, \ldots\}$

- The range of the values of X is: $S_X = \{1, 2, 3, 4, \ldots\} = \mathbb{N}$; S_X a countable set, associated with Ω.
- Hence $X \sim \mathbf{Geom}(p)$, the probability distribution table is

Probability Distribution

$P(X = 1) =$	0.01	0.01
$P(X = 2) =$	(0.99)*0.01	0.0099
$P(X = 3) =$	$(0.99)^2$*0.01	0.009801
$P(X = 4) =$	$(0.99)^3$*0.01	0.009703

Probability mass (density) f[i] = P[X = i]

====== THE END OF LECTURE 7 ==========

SUMMARY III: Emerging Applications and Reminiscence

1. Artificial intelligence
2. Climate change with Butterfly effect
3. Game theory
4. Life game to Cellular Automata
5. Uncertainty engineering: history, current and future

DISCLAIMER: this summary collects views of many generations of computing students in HCMC- Vietnam to Discrete Mathematics. Thank all of beloved students!