Lecture 1 - Review of Basic Probability and Introduction to Computational Statistics

Computational Statistics and Applications

Vu Quoc Hoang (vqhoang@fit.hcmus.edu.vn)
Tran Thi Thao Nhi (thaonhitt2005@gmail.com)

FIT - HCMUS

February 5, 2023

Agenda

1. The birthday problem
2. Review of basic probability
3. Probabilistic approximation by simulation
4. The Monty Hall problem
5. Review of conditional probability

Agenda

1. **The birthday problem**
2. Review of basic probability
3. Probabilistic approximation by simulation
4. The Monty Hall problem
5. Review of conditional probability

The birthday problem

Problem Determine the probability p that at least two people in a group of k people will have the same birthday (same day of the same month).

Assumption: each of the 365 days of the year is equally likely to be the birthday of any person in the group, and the birthdays of k people are "unrelated".

Agenda

1. The birthday problem
2. **Review of basic probability**
3. Probabilistic approximation by simulation
4. The Monty Hall problem
5. Review of conditional probability

Sample space and Events

- **Probability theory** is a mathematical framework that allows us to quantify, analyze and interpret **random** events or experiments whose outcome is **uncertain**.
- A **random experiment** is a(n) process/activity/experiment/task/manipulation by which we observe one uncertain **outcome** from the pool of possible outcomes which can be identified ahead of time.
- The collection of all possible outcomes of an experiment is called the **sample space** of the experiment, denoted by S or Ω (omega).
- When an experiment T has been performed, we call a set of possible outcomes E **an event associated with T** if the outcome of the experiment satisfied the conditions that specified the event E. Event E is determined by its **favorable outcomes**

$$E = \{\omega \in \Omega : \omega \text{ makes } E \text{ occurs}\} \subset \Omega.$$

Sample space and Events (cont.)

"Probability theory" uses the "language of set". Given an experiment T which has the sample space Ω and events $E, F \subset \Omega$

- $\omega \in \Omega$: **elementary event**,
- Ω: **certain event**,
- $\varnothing$: **impossible event**,
- $E^c = \Omega \setminus E$: **complement** of E, or event "E does not occur",
- $E \cup F$: **union** of events, or event "either E or F or both occur",
- $E \cap F$: **intersection** of events, or event "both E and F occur",
- $E \setminus F$: **different (subtraction)**, or event "E occurs but F does not",
- $E \subset F$: **containment (imply)**, or event "E occurs then so does F",
- $E = F$: event "either both E and F occur or both do not",
- $E \cap F = \varnothing$: E, F **disjoint (mutually exclusive)**, or event "E and F cannot both occur".

Probability

Given an experiment T with the sample space Ω, a function P, which assigns to each event $E \subset \Omega$ a real number $P(E)$, is called a **probability measure** if it satisfies 3 specific axioms:

1. For every event $E \subset \Omega, 0 \leq P(E) \leq 1$.
2. For every infinite sequence of disjoint events $E_1, E_2, \ldots$ $(E_i \cap E_j = \varnothing, \forall i \neq j)$:

   $$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i),$$

   or $P(E_1 \cup E_2 \cup \ldots) = P(E_1) + P(E_2) + \ldots$
3. $P(\Omega) = 1$.

$P(E)$ indicates the **probability** of E and shows how likely the event E is when the outcomes of T is unknown.

Probability (cont.)

General Properties of probability (from three axioms above):

1. $P(E^c) = 1 - P(E)$
2. $P(\varnothing) = 0$
3. If $E_1 \subset E_2$ thì $P(E_1) \leq P(E_2)$ và $P(E_2 \setminus E_1) = P(E_2) - P(E_1)$
4. If $E_1 \cap E_2 = \varnothing$ thì $P(E_1 \cup E_2) = P(E_1) + P(E_2)$
5. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$ (**addition law of probability**)
6. $P(E_1 \cup E_2 \cup E_3) =$
   $P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$
7. $P\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} P(E_i)$ (**union bound**)
8. $P\left(\bigcap_{i=1}^{\infty} E_i\right) \geq 1 - \sum_{i=1}^{\infty} P(E_i^c)$ (**Bonferroni inequality**)

Simple probability model

When the sample space is finite, $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$, a probability measure on Ω is specified by assigning a propability to each point of Ω (elementary event) $p_i = P(\omega_i)$

- $p_i \geq 0, i = 1, ..., n,$
- $\sum_{i=1}^{n} p_i = 1,$
- $P(E) = \sum_{w_i \in E} p_i$ for any event $E \subset \Omega$.

When the sample space is finite and each outcome is equally likely **(equiprobable outcomes)**, a **simple/classical probability model** is derived as

- $p_i = \frac{1}{n}, i = 1, ..., n,$
- $P(E) = \frac{|E|}{|\Omega|}$ for any event $E \subset \Omega$, ($|X|$ is the cardinality of X)
- Thus, finding probability of A reduces to a **counting** problem in which we need to count how many elements are in E and Ω.

Agenda

Probabilistic approximation by simulation

To appoximate the probability of event E associated with an experiment T, we can execute an analytical calculation below

- Perform the experiment T N times repetitively (and independently) and count how many times event E occurs, m. Then, $f(E) = \frac{m}{N}$ is defined as the frequency of E.
- When we execute this experiment the large number of times, $f(E) \approx P(E)$.
- Performing this experiment N times repetitively can be implemented by a computer simulation program.

The birthday problem - Theoretically

Without loss of generality, we can call the set of days in one year as

$$\mathcal{Y} = \{1, 2, ..., 365\}.$$

The sample space $\Omega = \{(x_1, x_2, ..., x_k) : x_i \in \mathcal{Y}, i = 1, ..., k\} = \mathcal{Y}^k$ with $|\Omega| = 365^k$.
Putting 2 events:

- A: "at least two people must have the same birthday",
- B: "none of them has the same birthday".

Thus, $A = B^c$ and $B = \{k\text{-permulations of } \mathcal{Y}\}$ with $|B| = P_{365}^k$.
Using the simple probability model, we obtain

$$p = P(A) = 1 - P(B) = 1 - \frac{P_{365}^k}{365^k} = 1 - \frac{365!}{(365-k)!365^k}.$$

The birthday problem - Simulation

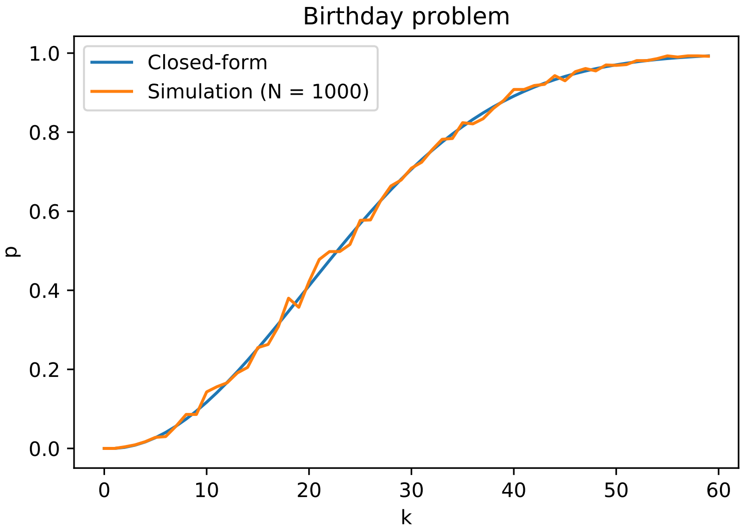
```
def birthday(k):
    return [random.randint(1, 365) for _ in range (k)]

def at_least_2(outcome):
    return len(set(outcome)) < len(outcome)

def relative_frequency(k, N, event):
    m = sum(event(birthday(k)) for _ in range (N))
    return m/N

relative_frequency(50 , 1000, at_least_2)
#0.973
```

The birthday problem - Result

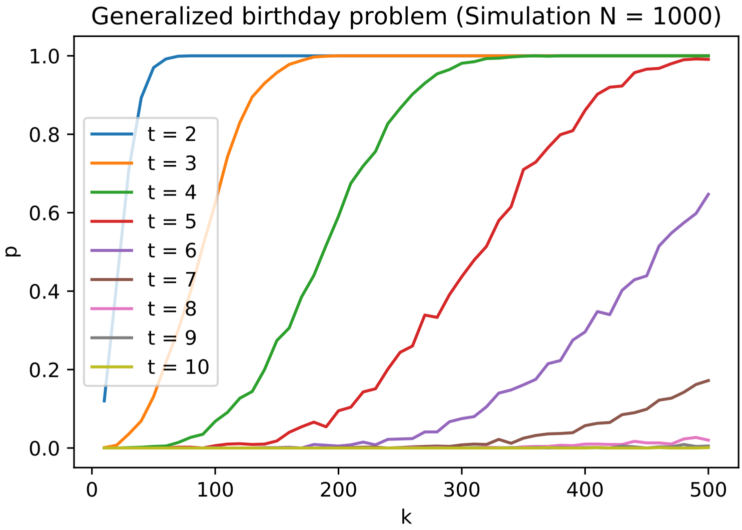


The birthday problem - Extension

An extension Determine the probability p that at least t people in a group of k will have the same birthday (same day of the same month)?

- Theory calculation: obtain a proper result. However, it is hard and nearly impossible to compute for many cases.
- Computer simulation: easy in coding modification, but time and resource-consuming and provides an approximation result only.
  (See the Python codes in Notebook)

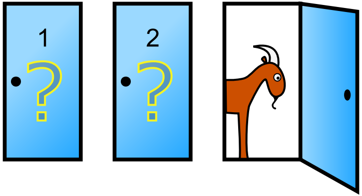
The birthday problem - Extension (cont.)



Agenda

The Monty Hall problem

Monty Hall problem. Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say #1, and the host, who knows what's behind the doors, opens another door, say #3, which has a goat. He then says to you, "Do you want to pick door #2?" Is it to your advantage to switch your choice?



(https://en.wikipedia.org/wiki/Monty_Hall_problem)

Agenda

1. The birthday problem
2. Review of basic probability
3. Probabilistic approximation by simulation
4. The Monty Hall problem
5. **Review of conditional probability**

Conditional probability

A need of modifying or updating the probabilities of events associated with an experiment T as any addition information of T has been observed

- The addition information of T relates to the occurrence of certain event(s).

The updated probability of event A after we learn that event B has occurred is the conditional probability of A given B, which is denoted by $P(A|B)$ and computed as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ (if } P(B) > 0).$$

- $A \cap B$ means "A in/intersect B",
- Divided by $P(B)$ is to normalize the probability,
- $P(.|B)$ calculates probability "on a new sample space" B, which is completely legal.

Multiplication rule

Multiplication rule

$$P(A \cap B) = P(B)P(A|B) \text{ (if } P(B) > 0),$$

$$P(A \cap B) = P(A)P(B|A) \text{ (if } P(A) > 0).$$

In some experiments, certain conditional probability $P(A|B)$ is relatively easy to compute rather than $P(A \cap B)$.

General multiplication rule Suppose that $A_1, ..., A_n$ are n events such that $P(A_1 \cap ... \cap A_n) > 0$, then

$$P(A_1 \cap ... \cap A_n) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1, A_2) \times ... \times P(A_n|A_1, A_2, ..., A_{n-1}).$$

Law of total probability

$B_1, B_2, ..., B_n$ form a **countably infinite partition** (or **partition**) of Ω if

1. $B_i \cap B_j = \varnothing, \forall i \neq j$,
2. $\Omega = B_1 \cup B_2 \cup ... \cup B_n$.

Law of total probability. Suppose that the events $B_1, B_2, ..., B_n$ form a partition of the sample space Ω and $P(B_i) > 0$ $(i = 1, ..., n)$. Then, for every event A in Ω,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(B_i)P(A|B_i).$$

Especially,

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c).$$

Bayes' theorem

Bayes' theorem (Bayes's rule). Let the events $B_1, B_2, ..., B_n$ form a partition such that $P(B_i) > 0$ $(i = 1, ..., n)$ and let A be an event such that $P(A) > 0$. Then, for all $i = 1, ..., n$,

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}$$

- $P(B_i)$: **prior probability** of B_i,
- $P(B_i|A)$: **posterior probability** of B_i given A,
- $P(A|B_i)$: **likelihood** of A given B_i,
- Note, $P(A)$ does not rely on B_i, thus

$$P(B_i|A) \propto P(B_i)P(A|B_i) \text{ (the symbol } \propto \text{ means "proportional to").}$$

Independent events

Two events $\{A, B\}$ are **independent** (statistically independent, stochastically independent) if

$$P(A \cap B) = P(A) \times P(B).$$

Similarly, $P(A|B) = P(A)$ $(P(B) > 0)$ or $P(B|A) = P(B)$ $(P(A) > 0)$.

Theorem. If two events $\{A, B\}$ are independent, then the others $\{A^c, B\}, \{A, B^c\}, \{A^c, B^c\}$ are also independent.

Three events $\{A, B, C\}$ are (mutually) independent if all of the pairwise $\{A, B\}, \{A, C\}, \{B, C\}$ are also independent respectively, and

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C).$$

Statistical procedure “independent replication”

The events $\{A_1, A_2, \ldots\}$ are independent if, for every non-empty and countable subcollection $\{B_1, B_2, \ldots, B_k\}$ of these events,

$$P\left(\bigcap_{i=1}^{k} B_i\right) = \prod_{i=1}^{k} P(B_i).$$

- From right to left: to demonstrate/prove the independence,
- From left to right: to calculate the simple probability based on the indenpendence.

Statistical procedure “independent replication”: performing the experiment T repetitively and independently the large number of times, suppose that A_i is the “associated with the i-th replicate” event,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

Conditional independence

Given an event C with $P(C) > 0$, we say that the events $\{A_1, A_2, ...\}$ are **conditionally independent** given C if, for every non-empty and countable subcollection $\{B_1, B_2, ..., B_k\}$ of these events,

$$P\left(\bigcap_{i=1}^{k} B_i | C\right) = \prod_{i=1}^{k} P(B_i|C).$$

Two events $\{A, B\}$ are conditionally independent given C if and only if

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

Similarly,

$$P(A|B, C) = P(A|C) \ (P(B|C) > 0) \text{ or } P(B|A, C) = P(B|C) \ (P(A|C) > 0).$$

Probabilistic approximation by simulation

To approximate the independent probability of event B given event A associated with an experiment T, we can execute an analytical calculation below

- Perform the experiment T N times repetitively (and independently), count how many times event A occurs, m, and B occurs also whenever A occurs, p. Then, $f(B|A) = \frac{p}{m}$ is defined as the frequency of B given A.
- When we execute this experiment the large number of times, $f(B|A) \approx P(B|A)$.
- Performing this experiment N times repetitively can be implemented by a computer simulation program.

The Monty Hall problem - Theoretically

Such the final stratergy is as above (player picked door #1, the host opened #3), since all of them share the same type of result.

Consider the event A_i, indicating that "the car is behind door number $\#i$" $(1 \leq i \leq 3)$ and event B_j which is "the host opens the door number $\#j$" $(1 \leq j \leq 3)$. Player picked door number #1, then

- $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$.
- $P(B_3|A_1) = \frac{1}{2}$ (Player picked door #1, which contains the car, so the host can open both of the others #2, #3).
- $P(B_3|A_2) = 1$ (Player picked door #1, the car is in #2, so the host only can open #3).
- $P(B_3|A_3) = 0$ (the car is in #3 so #3 could not be opened).

The Monty Hall problem - Theoretically (cont.)

Using Bayes' rule, we have the probability that the player wins the car if staying at door #1 is

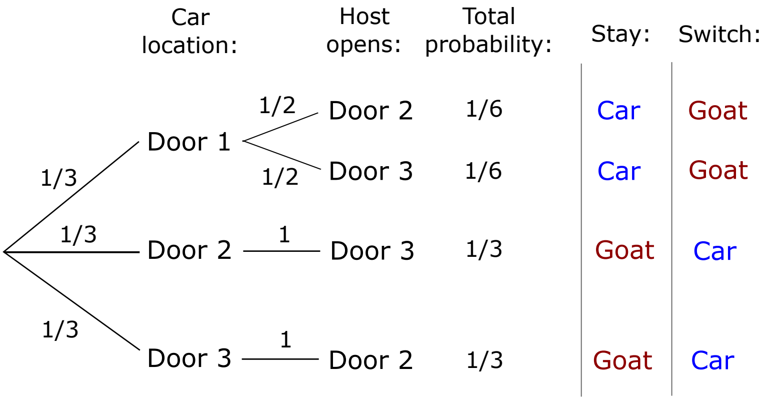
$$\begin{aligned} P(A_1|B_3) &= \frac{P(A_1)P(B_3|A_1)}{P(A_1)P(B_3|A_1) + P(A_2)P(B_3|A_2) + P(A_3)P(B_3|A_3)} \\ &= \frac{\frac{1}{3}\frac{1}{2}}{\frac{1}{3}\frac{1}{2} + \frac{1}{3}1 + \frac{1}{3}0} = \frac{1}{3} \end{aligned}$$

and the probability that player wins by switching is

$$\begin{aligned} P(A_2|B_3) &= \frac{P(A_2)P(B_3|A_2)}{P(A_1)P(B_3|A_1) + P(A_2)P(B_3|A_2) + P(A_3)P(B_3|A_3)} \\ &= \frac{\frac{1}{3}1}{\frac{1}{3}\frac{1}{2} + \frac{1}{3}1 + \frac{1}{3}0} = \frac{2}{3} \end{aligned}$$

In conclusion, it is never to the contestant's disadvantage to switch, as the conditional probability of winning by switching is twice as much as staying.

The Monty Hall problem - Tree diagram



(https://en.wikipedia.org/wiki/Monty_Hall_problem#Conditional_probability_by_direct_calculation)

The Monty Hall problem - Simulation

```
def Monty_Hall(doors={"#1", "#2", "#3"}):
    car_door = random.choice(list(doors))
    choice_door = random.choice(list(doors))
    open_door = random.choice(list(doors - {choice_door,
        car_door}))
    op_door = random.choice(list(doors - {choice_door,
        open_door}))

    return car_door == choice_door, car_door == op_door
```

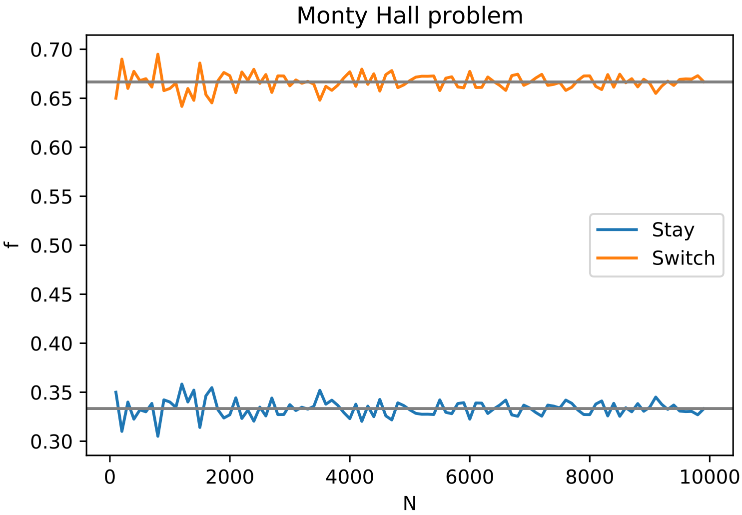
The Monty Hall problem - Simulation (cont.)

```
N = 10000
results = [Monty_Hall() for _ in range(N)]

stay_freq = sum([v for v, _ in results])/N
switch_freq = sum([v for _, v in results])/N

print(stay_freq, switch_freq)
#0.3317 0.6683
```

The Monty Hall problem - Simulation (cont.)



References

Chapter 1-2. Morris H. DeGroot, Mark J. Schervish. *Probability and Statistics*. Addison-Wesley, 2012.

Chapter 1-2. H. Pishro-Nik. "*Introduction to probability, statistics, and random processes*", available at `https://www.probabilitycourse.com`. Kappa Research LLC, 2014.