Introduction to Probability, Statistics and Random Processes

Chapter 1: Basic Concepts

Anduo Wang
Temple University
Email: anduo.wang@gmail.com
https://cis-linux1.temple.edu/∼tug29203/25fall-2033/index.html

Cars and goats: the Monty Hall dilemma

- On Sunday September 9, 1990, the following question appeared in the “Ask Marilyn” column in Parade, a Sunday supplement to many newspapers across the United States:

Cars and goats: the Monty Hall dilemma

- On Sunday September 9, 1990, the following question appeared in the "Ask Marilyn" column in Parade, a Sunday supplement to many newspapers across the United States:

  > *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 No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice? — Craig F. Whitaker, Columbia, Md.*

Cars and goats: the Monty Hall dilemma

- On Sunday September 9, 1990, the following question appeared in the "Ask Marilyn" column in Parade, a Sunday supplement to many newspapers across the United States:

  > *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 No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice? — Craig F. Whitaker, Columbia, Md.*

- Marilyn's answer — one should switch

To change or not to change

- Why it matters?
  - Why the odds not always a fifty-fifty?
  - To stress the point of switching, consider a generalization of the problem: suppose there are 10 000 doors, behind one is a car and behind the rest, goats. After you make your choice, the host will open 9998 doors with goats, and offers you the option to switch. To change or not to change, that's the question!
  - Switching doubles the likelihood (probability) of winning the car!

What is the probility of seeing the car behind the other door.

- The Monty Hall dilemma asks ...

  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 No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice? — Craig F. Whitaker, Columbia, Md.

- The host will always open a door with a goat. With probability $2/3$ your initial choice was wrong, and with probability $1/3$ it was right: So seeing the car behind the remaining door is $2/3$. You should switch.

Statistics versus intelligence agencies

- During World War II: to obtain more reliable estimates of German war production (potentials), experts from the Economic Warfare Division of the American Embassy and the British Ministry of Economic Warfare started to analyze markings and serial numbers obtained from captured German equipment ...

Statistics versus intelligence agencies: Analyzing the serial numbers

- Analyze tires taken from German aircraft shot over Britain and from supply dumps of aircraft and motor vehicle tires captured in North Africa. The marking on each tire contained the **maker's name**, a serial number, and a two-letter code for the **date** of manufacture.
- First step: breaking the two-letter date code

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Dunlop	T	I	E	B	R	A	P	O	L	N	U	D
Fulda	F	U	L	D	A	M	U	N	S	T	E	R
Phoenix	F	O	N	I	X	H	A	M	B	U	R	G
Sempirit	A	B	C	D	E	F	G	H	I	J	K	L

Figure: The 12 letter variations used by four different manufacturers.

- We now have for **each month** serial numbers recoded to numbers running from 1 to some unknown largest number N.

Statistics versus intelligence agencies: estimate the production (potential) based on observed serial numbers

- Next, estimate the largest number N for each month and each manufacturer separately by means of the observed (recoded) serial numbers.
  - Option 1: only the maximum observed (recoded) serial number
  - Option 2: the average observed (recoded) serial number

Type of tire	Estimated production	Actual production
Truck and passenger car	147 000	159 000
Aircraft	28 500	26 400
Total	175 500	186 100

Figure: With a sample of about 1400 tires from five producers, individual monthly output figures were obtained for almost all months over a period from 1939 to mid-1943.

2033: Computational Probability and Statistics

- In this course: more on the probability theory

2033: Computational Probability and Statistics

- In this course: more on the probability theory
- Contact: anduo.wang@gmail.com, office hours in SERC 342.

2033: Computational Probability and Statistics

- In this course: more on the probability theory
- Contact: anduo.wang@gmail.com, office hours in SERC 342.
- All course materials will be published on THIS website.

2033: Computational Probability and Statistics

- In this course: more on the probability theory
- Contact: anduo.wang@gmail.com, office hours in SERC 342.
- All course materials will be published on THIS website.
- Logistics:
  https://cis-linux1.temple.edu/∼tug29203/25fall-2033/index.html

What is Probability Theory?

- A mathematical framework that allows us to analyze random phenomena.
- **Probability theory** provides us such a framework.
- But what do we mean by **random phenomena** and **probability**? How can we express **randomness**?

Randomness

- We define **random phenomena** as events and experiments whose outcomes we cannot predict with certainty.
- Example: Flipping a fair coin; Throwing a die. We cannot predict whether the outcome would be heads or tails.
- Remember these experiments. We will come back to them a lot.

More on Randomness

- We can think about randomness as a way to express what we do not know.
- Let's go back to the coin experiment again. What if we knew more about the force with which the coin was flipped, the initial orientation, etc.
- With all the information provided we might be able to predict the outcome of the coin flip every time.
- When our knowledge about the outcome is limited, we say that it is random.

Interpretations of Probability

- If we flip the fair coin many times, without prior information about the flip, we say the probability of heads is 50% or $\frac{1}{2}$. What does this mean?
- **Relative frequency**: If we flip the coin a large number of times, it will come to heads about $\frac{1}{2}$ the time.
- **Subjective personal belief**: Probability is the quantification of our belief that something would happen. For example: What we think is the chance of rain today?

Interpretations of Probability

- The two interpretations often coincide since personal beliefs are based on the assessment of relative frequency of events.
- The beauty of probability theory is that it is applicable in both cases.
- It provides a solid framework to study random phenomena and starts with the **axioms of probability**.

Relative frequency

- Consider an experiment with a finite number of **mutually exclusive** outcomes which are **equiprobable** ("equally likely")

Relative frequency

- Consider an experiment with a finite number of **mutually exclusive** outcomes which are **equiprobable** ("equally likely")
- Let A denote some event associated with the possible outcomes of the experiment. Then the probability $P(A)$ of the event A is defined as the fraction of the outcomes in which A occurs. More exactly
$$P(A) = \frac{N(A)}{N}$$

Relative frequency

- Consider an experiment with a finite number of **mutually exclusive** outcomes which are **equiprobable** ("equally likely")
- Let A denote some event associated with the possible outcomes of the experiment. Then the probability $P(A)$ of the event A is defined as the fraction of the outcomes in which A occurs. More exactly
$$P(A) = \frac{N(A)}{N}$$
- N: the total number of outcomes of the experiment

Relative frequency

- Consider an experiment with a finite number of **mutually exclusive** outcomes which are **equiprobable** ("equally likely")
- Let A denote some event associated with the possible outcomes of the experiment. Then the probability $P(A)$ of the event A is defined as the fraction of the outcomes in which A occurs. More exactly
$$P(A) = \frac{N(A)}{N}$$
- N: the total number of outcomes of the experiment
- $N(A)$: the number of outcomes leading to the occurrence of the event A.

Example

- In throwing a pair of dice, there are $N = 36$ mutually exclusive equiprobable events, each represented by an ordered pair (a,b), where a is the number of spots showing on the first die and b the number showing on the second die. Let A be the event that both dice show the same number of spots. What is the probability of A?

Example

- In throwing a pair of dice, there are $N = 36$ mutually exclusive equiprobable events, each represented by an ordered pair (a,b), where a is the number of spots showing on the first die and b the number showing on the second die. Let A be the event that both dice show the same number of spots. What is the probability of A?
- A occurs whenever $a = b$, i.e., $n(A) = 6$. Therefore $P(A) = 6/36$.

Relative frequency and Probability

- In an experiment with a finite number of **mutually exclusive** outcomes which are **equiprobable**, the relative frequencies n(A)/n observed in different series of trials are virtually the same for large n:

$$P \sim \frac{n(A)}{n}$$

De Mere's paradox

- As a result of extensive observation of dice games, the French gambler de Mere noticed that the total number of spots showing on three dice thrown simultaneously turns out to be 11 (the event A_1,) more often than it turns out to be 12 (the event A_2), although from his point of view both events should occur equally often: A_1 occurs in just six ways (6:4:1, 6:3:2, 5:5:1, 5:4:2, 5:3:3, 4:4:3), and A_2 also occurs in just six ways (6:5:1, 6:4:2, 6:3:3, 5:5:2, 5:4:3, 4:4:4)

De Mere's paradox

- As a result of extensive observation of dice games, the French gambler de Mere noticed that the total number of spots showing on three dice thrown simultaneously turns out to be 11 (the event A_1,) more often than it turns out to be 12 (the event A_2), although from his point of view both events should occur equally often: A_1 occurs in just six ways (6:4:1, 6:3:2, 5:5:1, 5:4:2, 5:3:3, 4:4:3), and A_2 also occurs in just six ways (6:5:1, 6:4:2, 6:3:3, 5:5:2, 5:4:3, 4:4:4)
- Do you see the fallacy *(found by Pascal)*? Can you calculate $N(A_1)$ and $N(A_2)$?

De Mere's paradox

- As a result of extensive observation of dice games, the French gambler de Mere noticed that the total number of spots showing on three dice thrown simultaneously turns out to be 11 (the event A_1,) more often than it turns out to be 12 (the event A_2), although from his point of view both events should occur equally often: A_1 occurs in just six ways (6:4:1, 6:3:2, 5:5:1, 5:4:2, 5:3:3, 4:4:3), and A_2 also occurs in just six ways (6:5:1, 6:4:2, 6:3:3, 5:5:2, 5:4:3, 4:4:4)
- Do you see the fallacy *(found by Pascal)*? Can you calculate $N(A_1)$ and $N(A_2)$?
- **Hint:** There are six distinct outcomes leading to the combination 6:4: 1, namely (6,4,1), (6,1,4), (4,6,1), (4,1,6), (1,6,4) and (1,4,6)...

Random Experiments

- A **random experiment** is the process of observing something uncertain. For example: rolling a die.

Random Experiments

- A **random experiment** is the process of observing something uncertain. For example: rolling a die.
    - When we repeat a random experiment several times, we call each one a trial.

Random Experiments

- A **random experiment** is the process of observing something uncertain. For example: rolling a die.
  - When we repeat a random experiment several times, we call each one a trial.
- An **outcome** is a result of a random experiment.

Random Experiments

- A **random experiment** is the process of observing something uncertain. For example: rolling a die.
  - When we repeat a random experiment several times, we call each one a trial.
- An **outcome** is a result of a random experiment.
- The set of all possible outcomes is called the **sample space**, and in this context is called the universal set Ω.

Random Experiments

- A **random experiment** is the process of observing something uncertain. For example: rolling a die.
  - When we repeat a random experiment several times, we call each one a trial.
- An **outcome** is a result of a random experiment.
- The set of all possible outcomes is called the **sample space**, and in this context is called the universal set Ω.
- An **event** is a subset of the sample space.

Random Experiments

- A **random experiment** is the process of observing something uncertain. For example: rolling a die.
  - When we repeat a random experiment several times, we call each one a trial.
- An **outcome** is a result of a random experiment.
- The set of all possible outcomes is called the **sample space**, and in this context is called the universal set Ω.
- An **event** is a subset of the sample space.
  - An event A occurs if the outcome of the experiment is an element of the set A.

Sample spaces

- Sample spaces are simply sets whose elements describe the outcomes of the experiment in which we are interested.
- Example: Consider an experiment of tossing a coin. Assuming that we will never see the coin land on its rim.
    - What is the sample space Ω?

Sample spaces

- In another experiment we ask the next person we meet on the street in which month her birthday falls.
  - What is the sample space Ω?

Sample spaces

- In a third experiment, we find on our doormat three envelopes, sent to us by three different persons, and we look in which order the envelopes lie on top of each other. Coding them 1, 2, and 3.

Sample spaces

- In a third experiment, we find on our doormat three envelopes, sent to us by three different persons, and we look in which order the envelopes lie on top of each other. Coding them 1, 2, and 3.
  - $\Omega = \{123, 132, 213, 231, 312, 321\}$

Events

- Recall subsets of the sample space are called events.
  - We say that an event A occurs if the outcome of the experiment is an element of the set A.
- For example, in the birthday experiment we can ask for the outcomes that correspond to a long month (a month with 31 days).
- This is the event $L = \{Jan, Mar, May, Jul, Aug, Oct, Dec\}$.

Events

- In the birthday experiment, if R is the event that corresponds to the months that have the letter r in their (full) name, what is R?
- What is the event that corresponds to long months that contain the letter r?
    - $L \cap R$
- Events may be combined according to the usual set operations.

Identical (equivalent) events

- Given two events A and B, suppose A occurs if and only if B, occurs. Then A, and B are said to be identical (or equivalent), and we write A=B.
- Example: In throwing a pair of dice, let A be the event that "the total number of spots is even" and B the event that "both dice turn up even or both dice turn up odd." Then $A = B$.

Mutually exclusive events

- Two events A, and B are said to be mutually exclusive or incompatible if the occurrence of one event precludes the occurrence of the other, i.e., if A, and B cannot occur simultaneously.

Mutually exclusive events

- Two events A, and B are said to be mutually exclusive or incompatible if the occurrence of one event precludes the occurrence of the other, i.e., if A, and B cannot occur simultaneously.

Union, intersection, difference, complementary, implication

- $A \cup B$, the *union* of A and B: the event consisting of the occurrence of *at least one of* the events A and B
- $A \cap B$, the *intersection* of two events A and B: the event consisting of the occurrence of both events
- $A - B$, the *difference*: the event in which A occurs but not B
- $A \subset B$, the occurrence of the event A implies that of the event B

Example

- In throwing a pair of dice, let U be the event that "the total number of spots is even"; A the event that "both dice turn up even", and B the event that "both dice turn up odd."
- A and B are mutually exclusive
- $U = A \cup B$
- $A = U - B$

Events and sets

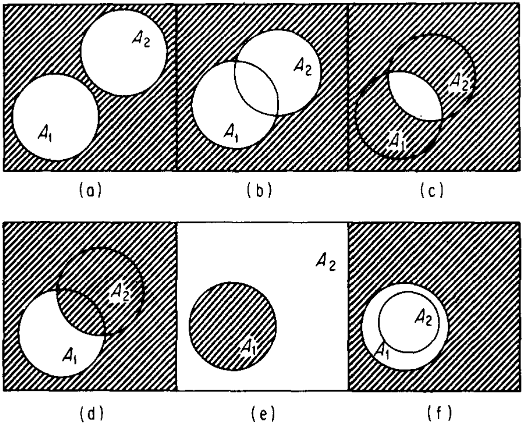

Figure: (a) mutually exclusive; (b) the unshaded figure represents the union ; (c) The unshaded figure represents the intersection; (d) The unshaded figure represents the difference; (e) The shaded and unshaded events are complements of each other; (f) Event A_2 implies event A_1.

Repeated experiments

- Basic to statistics is that one usually does not consider one experiment, but that the same experiment is performed several times.
- For example, throw a coin two times. What is the sample space associated with this new experiment?
- $\Omega = \{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\}$