Fundamentals of Probability

Prof. Edmund Lam

Department of Electrical and Electronic Engineering
The University of Hong Kong

ELEC2844: Probabilistic Systems Analysis

(Second Semester, 2023–24)

https://www.eee.hku.hk/~elec2844

Meaning of probability

Does "probability" have different meaning depending on the question?

1. I roll a die and there is $\frac{1}{6}$ chance I get ⚅.

Meaning of probability

Does "probability" have different meaning depending on the question?

1. I roll a die and there is $\frac{1}{6}$ chance I get ⚅.
2. There is $\frac{1}{6}$ chance Hang Seng Index will go down tomorrow.

Calibration questions

[Birthday Party problem]

Question: A year has 365 days. In a room with N people, what is the smallest value of N such that the probability of two or more people having birthday on the same day exceeds $\frac{1}{2}$?

Calibration questions

[Birthday Party problem]

Question: A year has 365 days. In a room with N people, what is the smallest value of N such that the probability of two or more people having birthday on the same day exceeds $\frac{1}{2}$?

ANS: 23 (which you should be able to do)

https://en.wikipedia.org/wiki/Birthday_problem

How about three or more people? four or more people?

Calibration questions

I was sent this paper as a "birthday gift"

2017 IEEE Symposium on Computers and Communications (ISCC)

Non-Recursive Computation of the Probability of More than Two People Having the Same Birthday

Michael Kounavis, Sergej Deutsch, David Durham and Saeedeh Komijani
Intel Labs, Intel Corporation, 2111, NE 25th Avenue, Hillsboro, OR 97124
Email: {michael.e.kounavis, sergej.deutsch, david.durham, saeedeh.komijani}@intel.com

***Abstract*—We address a well known problem of computer science, the problem of computing the probability that a given number of people $m > 1$ have the same birthday from among the members of a larger set of cardinality $n \geq m$. The solution to this problem for $m = 2$ is well known and is usually referred to as the 'birthday surprise probability'. A solution for $m = 3$ is also known and appears in the 2004 paper by DasGupta [The matching, birthday and the strong birthday problem: a contemporary review, Journal of Statistical Planning and Inference]. Further approximations to the solution of the related problem of computing the minimum number of people to interview until m people with the same birthday are found are presented in the seminal work by Klamkin and Newman [Extensions on the birthday surprise, Journal of Combinatorial Theory, 1967]. In this paper we present a new non-recursive approximation for the birthday probability applicable to any value of $m > 1$, which yields results that are experimentally proven accurate under the assumption that the number of birthdays is significantly larger than the number of people. Our expression is easy to compute, non-recursive, and applicable to values of m that can be arbitrarily larger than 2 or 3. We verify the validity of our result computing the birthday probability for different values of m, over billions of sets of random values generated using the Intel® RDRAND hardware random number generation instruction. Our solution is based on a novel tree-based description of the event space which, if used, allows for the computation of the birthday probability efficiently and without involving recursions or multinomial distributions.**

I. Introduction

In this paper we address the problem of computing the birthday probability for more than two persons having the same birthday. The birthday problem is a well known problem of computer science first introduced by von Mises in 1932 [1] and later discussed by Harold Davenport [2] in 1960. The birthday question is whether we can compute the probability that the number of elements from a set having the same value is greater or equal than a given threshold. More precisely, the problem considers a set A containing n elements $e_0, e_1, \ldots, e_{n-1}$. Each element i is associated with a value $v_i = v(e_i)$ which is a discrete random variable associated with probability distribution function $p_i(x) = \text{Prob}(v_i = x)$. The question, which is asked is whether we can compute the probability that a set $C \subseteq A$ exists such that its cardinality $|C|$ is greater or equal than a given threshold m and for every pair of elements $e_i, e_j \in C$ $v(e_i) = v(e_j)$. If the elements are people and the values $v_i = v(e_i)$ are birthdays, the question asked is whether we can compute the probability that a given number of people from a group have the same birthday.

The birthday problem is an important problem of computer science. Knowledge of the birthday probability allows us to evaluate how close some sequence of values is to random. Such methodology is useful in many fields. In cryptography [12], for example, systems such as block ciphers or one-way hash functions are required to approximate theoretical constructions that produce random output such as the random oracle [13], or the random permutation [14]. Being able to observe anomalies in the output of cryptographic systems is highly desirable. Other fields where knowledge of the birthday probability could be helpful include bio-informatics, network anomaly detection, data mining and machine learning.

No precise solution among the published literature is known which computes the birthday probability for any value of m and which is computationally tractable to date. Computing the birthday probability for $m = 2$ is well known and such probability is usually referred to as the 'birthday surprise probability'. This is a well known result which states that it requires only 23 people in order to find two among them that have the same birthday with probability 0.5. For $m = 3$, a solution appears in the 2004 paper by DasGupta [4], which involves the use of the multinomial distribution. It is found that the set size required for finding three people with the same birthday with probability 0.5 increases from 23 to 88. For values of m larger than three, Wolfram MathWorld [5] mentions a solution by Finch proposed in 1997, which is associated with exponential complexity and becomes intractable for large values of m. Several bounds on the birthday probability are discussed in reference [15]. On the other hand, the seminal work by Klamkin and Newman [3] discusses only an approximate solution to the related problem of computing the minimum number of people to interview until m people with same birthday are found. On-line birthday probability calculators such as [7], [6] return results only for values of m equal to 2 or 3.

In our work, we present a new analytical result that approximates the birthday probability for any value of $m > 1$ using a non-recursive formula, under the assumption that the number of birthdays is significantly larger than the number of people. Our expression is generic and applicable to values of m that can be arbitrarily larger than 2 or 3. Compared to some previous results [4] [5], our solution is tractable, does not involve recursions and uses only the binomial distribution. The complexity of an algorithm that implements our solution is $O(nm)$. The only assumptions about the elements of the input

978-1-5386-1629-1/17/$31.00 ©2017 IEEE

same bday	number needed
2	23
3	88
4	188
5	313
6	459
7	624
8	800
9	983
10	1182

Calibration questions

"The birthday problem is an important problem of computer science. Knowledge of the birthday probability allows us to evaluate how close some sequence of values is to random. Such methodology is useful in many fields. In cryptography, for example, systems such as block ciphers or one-way hash functions are required to approximate theoretical constructions that produce random output such as the random oracle, or the random permutation. Being able to observe anomalies in the output of cryptographic systems is highly desirable. Other fields where knowledge of the birthday probability could be helpful include bio-informatics, network anomaly detection, data mining and machine learning."

Kounavis et al., "Non-recursive computation of the probability of more than two people having the same birthday," *IEEE Symposium on Computers and Communications*, 2017.

Calibration questions

[Deck shuffling problem]

Question: You are given a deck of cards organized sequentially. How many shuffles does it take to result in a fully random deck?

Calibration questions

[Deck shuffling problem]

Question: You are given a deck of cards organized sequentially. How many shuffles does it take to result in a fully random deck?

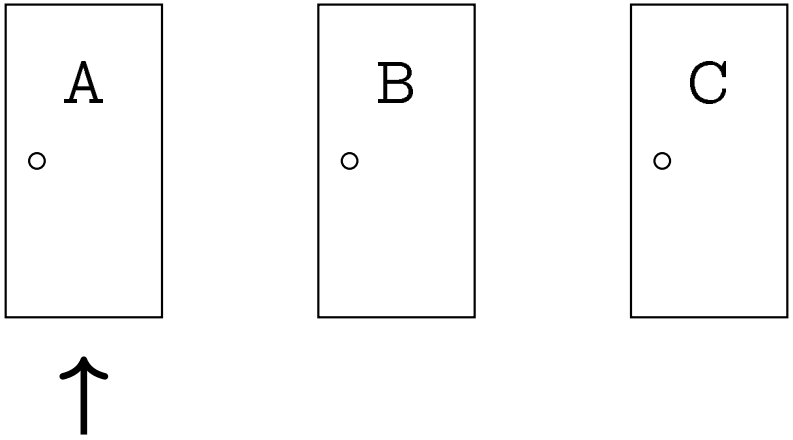
ANS: 7

https://www.math.hmc.edu/funfacts/ffiles/20002.4-6.shtml
https://www.dartmouth.edu/~chance/teaching_aids/Mann.pdf

Calibration questions

[Monty Hall problem]

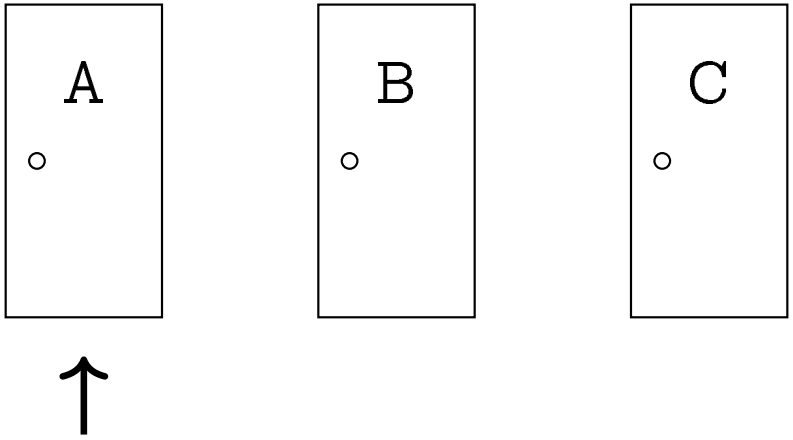
Question: There is one prize hidden behind one of three doors with equal likelihood. You choose one door. The host opens one of the remaining two doors and reveals that there is no prize behind that door. You are now offered a chance to switch. Will you?



Calibration questions

[Monty Hall problem]

Question: There is one prize hidden behind one of three doors with equal likelihood. You choose one door. The host opens one of the remaining two doors and reveals that there is no prize behind that door. You are now offered a chance to switch. Will you?

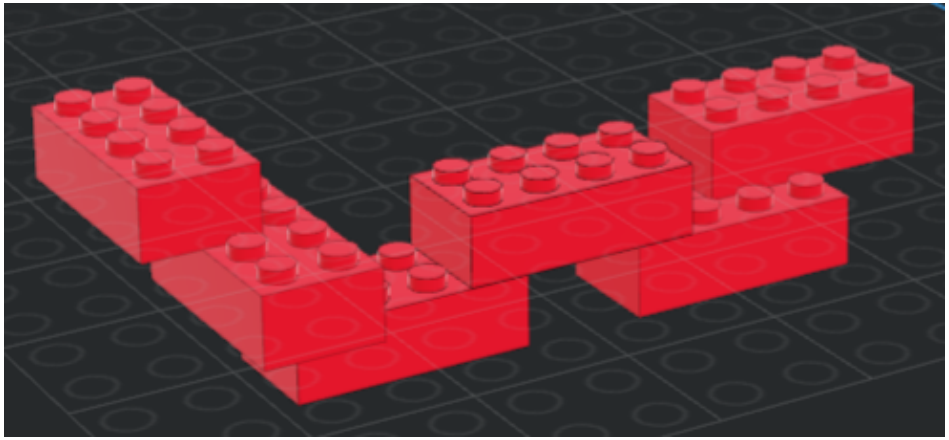


ANS: Yes!

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

Random bricks

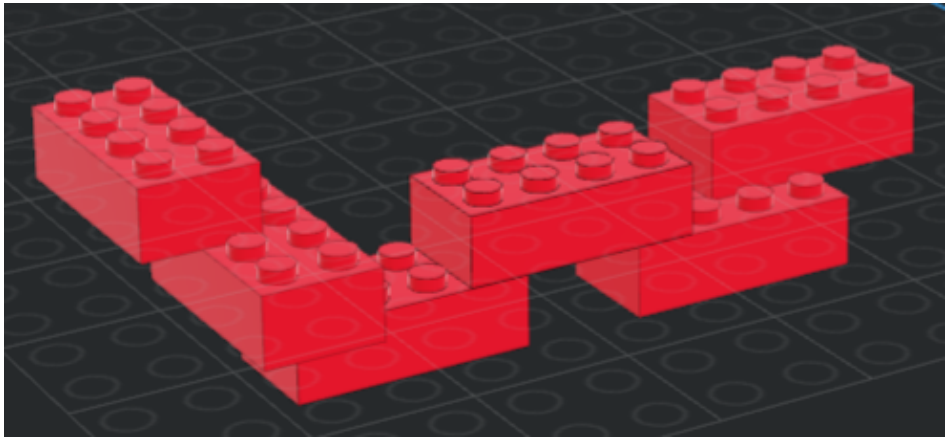
You and I were both given six 2×4 red bricks. I made the following construction:



What is the probability that you and I made the same construction? (Not counting rotation or symmetry)

Random bricks

You and I were both given six 2×4 red bricks. I made the following construction:



What is the probability that you and I made the same construction? (Not counting rotation or symmetry)

With 6 bricks, you can get $915,103,765$ combinations! So the probability is 1/915103765. (This particular construction is #658,916,208 according to Lego. See also `http://web.math.ku.dk/~eilers/lego.html`)

Calibration questions

Problems involving probability can be

- counterintuitive
- ill-defined
- complicated

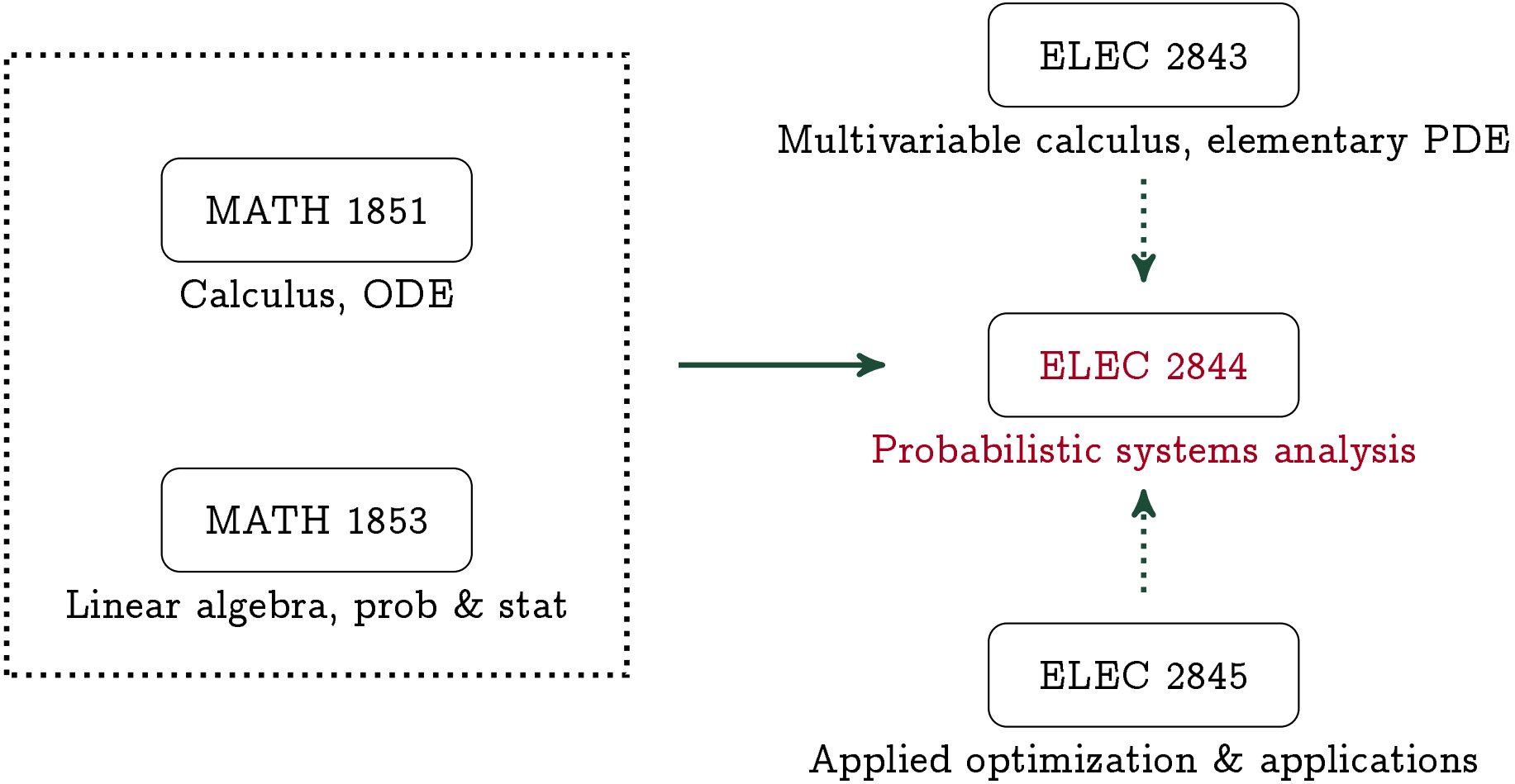
but also

- important
- powerful

especially in this era of big data *and* machine learning,

because probability helps to characterize uncertainty.

A comment on the required mathematics knowledge



1 Sets and probability models

2 Conditional probability

3 Total probability and Bayes' rule

4 Independence

5 Counting

Mathematical foundation

We have some intuitive understanding of "chance", but we need a mathematical formalism so we can agree on more precise meaning.

- I am "110% sure" you want to know this!
- In multiple choice questions, can you be 100% sure about your answer?

We also need mathematical tools to help us manipulate expressions and do various derivations. *Set theory* turns out to be the appropriate mathematics for the task.

Sets — definitions

- Set S with elements x_i:

$$S = \{x_1, x_2, \ldots, x_n\} \quad \text{and also} \quad x_i \in S \text{ for } 1 \leqslant i \leqslant n \tag{1}$$

- A set can be finite, e.g.,

$$S = \{H, T\} \quad \text{or} \quad S = \{1, 2, 3, 4, 5, 6\} \tag{2}$$

- A set can be countably infinite, e.g.,

$$S = \{2, 4, 6, \ldots\} \quad \text{set of all positive even integers} \tag{3}$$

- A set can be uncountable, e.g.,

$$S = \{x \,|\, 0 \leqslant x \leqslant 1\} \tag{4}$$

The symbol "|" means "such that"

Sets — relations and operations

- Set S is a subset of T:

$$S \subset T \quad \text{or, if they can be equal,} \quad S \subseteq T \tag{5}$$

- Null set $\emptyset$ and universal set Ω
- Set complement:

$$S^c = \{x \in \Omega \,|\, x \notin S\} \tag{6}$$

- Set union:

$$S \cup T = \{x \,|\, x \in S \text{ or } x \in T\} \tag{7}$$

$$\bigcup_{n=1}^{\infty} S_n = S_1 \cup S_2 \cup \ldots \tag{8}$$

- Set intersection:

$$S \cap T = \{x \,|\, x \in S \text{ and } x \in T\} \tag{9}$$

$$\bigcap_{n=1}^{\infty} S_n = S_1 \cap S_2 \cap \ldots \tag{10}$$

Sets — algebra

- Some equations:

$$S \cup T = T \cup S \tag{11}$$

$$S \cup (T \cup U) = (S \cup T) \cup U \tag{12}$$

$$S \cap (T \cup U) = (S \cap T) \cup (S \cap U) \tag{13}$$

$$S \cup (T \cap U) = (S \cup T) \cap (S \cup U) \tag{14}$$

$$(S^c)^c = S \tag{15}$$

$$S \cap S^c = \emptyset \tag{16}$$

$$S \cup \Omega = \Omega \tag{17}$$

$$S \cap \Omega = S \tag{18}$$

- De Morgan's theorems:

$$\left(\bigcup_n S_n\right)^c = \bigcap_n {S_n}^c \qquad \left(\bigcap_n S_n\right)^c = \bigcup_n {S_n}^c \tag{19}$$

Probabilistic models

Fundamental math structure is a set:

- Sample space Ω: set of all possible outcomes or events

Law of Probability:

- Let A = a set of possible outcomes
- Map A to a nonnegative number $\mathbb{P}(A)$ called the probability of A
- This number reflects our belief of the "likelihood" of occurrence of such outcomes in an experiment

"Probability is a measure of the size of a set."

Probability axioms

Three axioms: (Kolmogorov axioms)

1. Nonnegativity: for every event A

$$\boxed{\mathbb{P}(A) \geqslant 0} \tag{20}$$

2. Additivity: if A and B are disjoint events

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)} \tag{21}$$

3. Normalization:

$$\boxed{\mathbb{P}(\Omega) = 1} \tag{22}$$

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) = 1 + \mathbb{P}(\emptyset) \quad \Rightarrow \quad \mathbb{P}(\emptyset) = 0$$

Properties of probability laws

Let A, B, and C be events (not necessarily disjoint):

$$A \subset B \Rightarrow \mathbb{P}(A) \leqslant \mathbb{P}(B) \tag{23}$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \tag{24}$$

$$\mathbb{P}(A \cup B) \leqslant \mathbb{P}(A) + \mathbb{P}(B) \tag{25}$$

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B) + \mathbb{P}(A^c \cap B^c \cap C) \tag{26}$$

Discrete model

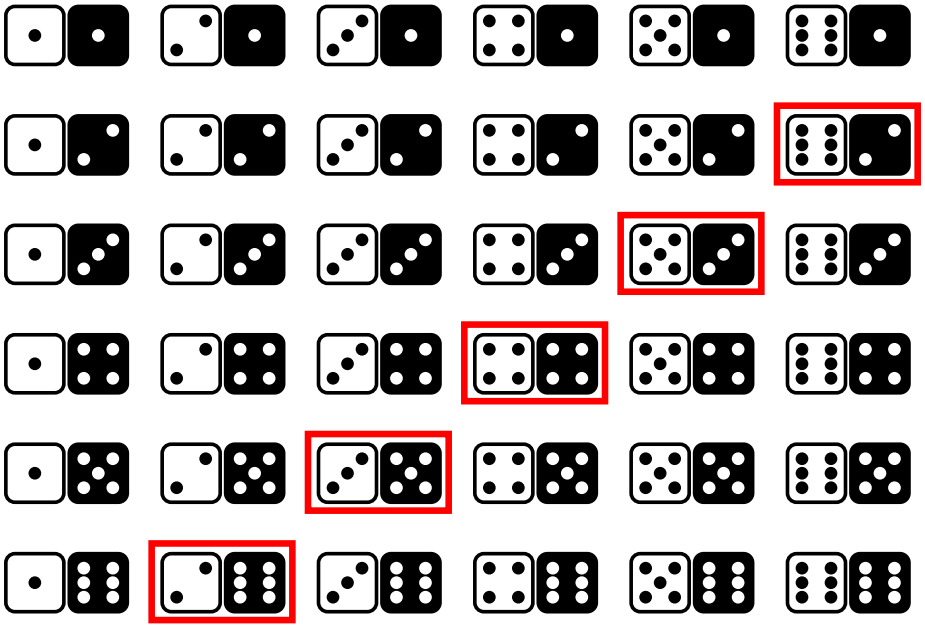
Discrete model: "Count" the probability of individual outcome, and sum them up for the probability of a collection of outcomes (= event)

Example: Roll two dice. What is the probability that the sum is 8?

Discrete model

Discrete model: "Count" the probability of individual outcome, and sum them up for the probability of a collection of outcomes (= event)

Example: Roll two dice. What is the probability that the sum is 8?



Ans: $\mathbb{P}(\text{sum} = 8) = \dfrac{5}{36}$

Continuous model

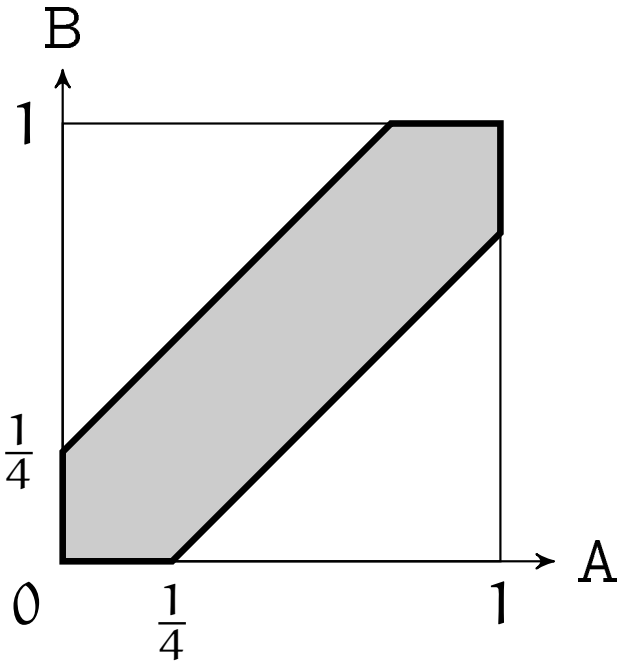
Continuous model: "Measure" the probability of individual interval, and integrate them for the probability of an event

Example: A and B agreed to meet at 3pm, but each will be late somewhere between 0 and 1 hour. *All pairs of delays are equally likely.* Whoever is first to arrive will wait for 15 min, and will leave if the other has not arrived. What is the probability that they meet?

Continuous model

Continuous model: "Measure" the probability of individual interval, and integrate them for the probability of an event

Example: A and B agreed to meet at 3pm, but each will be late somewhere between 0 and 1 hour. *All pairs of delays are equally likely.* Whoever is first to arrive will wait for 15 min, and will leave if the other has not arrived. What is the probability that they meet?



Ans: $\mathbb{P}(\text{A and B will meet}) = \dfrac{7}{16}$

1 Sets and probability models

2 Conditional probability

3 Total probability and Bayes' rule

4 Independence

5 Counting

Conditional probability

Conditional probability gives us a formal way to assess the idea of "information", and how that influences our computation of probability for certain events. Intuitively,

- Some information can be useless
- Some information may tell us everything
- Most interesting cases are in between, i.e., the new knowledge can be regarded as partial information, which adjusts our probability calculations

Definition

Consider a case where we have partial information. e.g. What is the probability of the first die being 3 if the sum of two dice is 8?

$$\mathbb{P}(A \mid B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B} \tag{27}$$

Thus we define conditional probability:

$$\boxed{\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}} \tag{28}$$

To comply with our mathematical formalism, we can show that $\mathbb{P}(A \mid B)$ *defined above follows the probability law.*

Examples

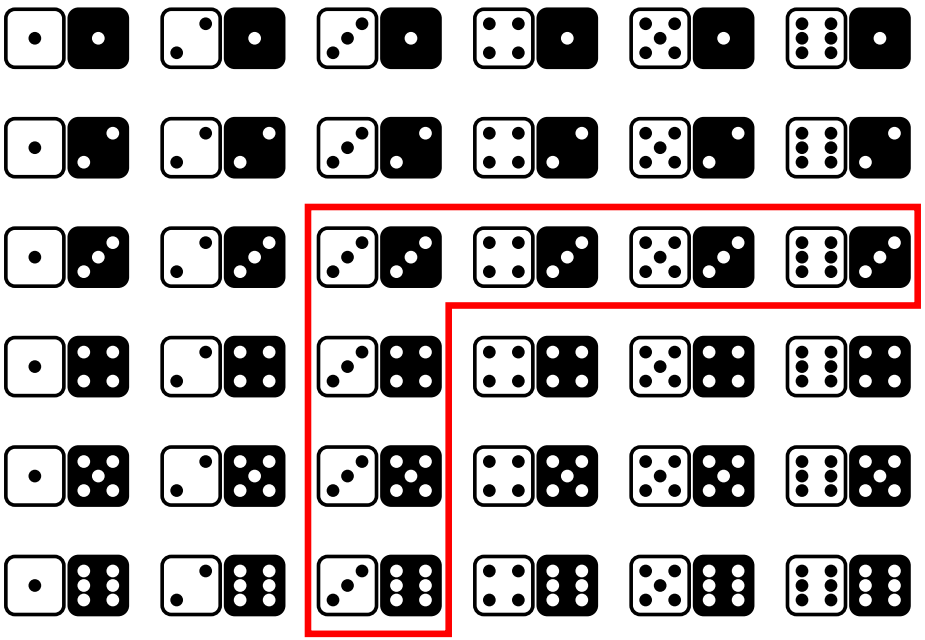
1. Roll two dice, with outcomes X and Y. Let events

$$A = \{\max(X, Y) = 5\} \qquad B = \{\min(X, Y) = 3\}.$$

Find $\mathbb{P}(A \mid B)$.

2. Two students enter a competition. Probability that X wins a prize is $2/3$. Probability that Y wins a prize is $1/2$. Probability that at least one of them wins a prize is $3/4$. Now we know exactly one of them indeed wins a prize. What is the probability that it is X (or Y)?

Examples



$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{2/36}{7/36} = \frac{2}{7}$$

Examples

We use uppercase letters for winning, lowercase for losing. There are four outcomes: XY, Xy, xY, xy. We know:

$$\begin{aligned}
\mathbb{P}(XY) + \mathbb{P}(Xy) &= \tfrac{2}{3} && \text{(X wins)} \\
\mathbb{P}(XY) + \mathbb{P}(xY) &= \tfrac{1}{2} && \text{(Y wins)} \\
\mathbb{P}(XY) + \mathbb{P}(Xy) + \mathbb{P}(xY) &= \tfrac{3}{4} && \text{(at least one wins)} \\
\mathbb{P}(XY) + \mathbb{P}(Xy) + \mathbb{P}(xY) + \mathbb{P}(xy) &= 1 && \text{(all cases)}
\end{aligned}$$

So we know $\mathbb{P}(xy) = \frac{1}{4}$, $\mathbb{P}(xY) = \frac{1}{12}$, $\mathbb{P}(Xy) = \frac{1}{4}$, $\mathbb{P}(XY) = \frac{5}{12}$.

$$\mathbb{P}(Xy \,|\, \{xY, Xy\}) = \frac{\frac{1}{4}}{\frac{1}{12} + \frac{1}{4}} = \frac{3}{4}$$

$$\mathbb{P}(xY \,|\, \{xY, Xy\}) = \frac{\frac{1}{12}}{\frac{1}{12} + \frac{1}{4}} = \frac{1}{4}$$

Multiplication rule

A consequence of the conditional probability formula is that

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B)} \tag{29}$$

Further generalization:

$$\boxed{\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_1 \cap A_2)\ldots\mathbb{P}\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right)} \tag{30}$$

Example

1. Draw three cards from an ordinary 52-card deck without replacement. What is the probability that none is a ♠?

2. We divide a class of 16 students — 4 male, 12 female — into 4 groups of 4 students each. What is the probability that each group has exactly one male student?

Example

Define

$$A_i = \{\text{ith card not a } \spadesuit\} \qquad i = 1, 2, 3$$

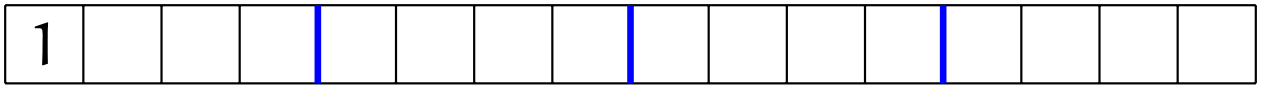
and we know:

$$\begin{aligned} \mathbb{P}(A_1) &= \tfrac{39}{52} \\ \mathbb{P}(A_2 \,|\, A_1) &= \tfrac{38}{51} \\ \mathbb{P}(A_3 \,|\, A_1 \cap A_2) &= \tfrac{37}{50} \end{aligned}$$

Therefore, using the multiplicative rule,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \,|\, A_1)\,\mathbb{P}(A_3 \,|\, A_1 \cap A_2) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}$$

Example



Label the four students 1, 2, 3, 4. Suppose we assign them sequentially.

$$A_1 = \{\text{student 2 not in the same group as student 1}\}$$
$$A_2 = \{\text{student 3 not in the same group as student 1 or 2}\}$$
$$A_3 = \{\text{student 4 not in the same group as student 1 or 2 or 3}\}$$

$$\mathbb{P}(A_1) = \frac{12}{15}$$
$$\mathbb{P}(A_2 \mid A_1) = \frac{8}{14}$$
$$\mathbb{P}(A_3 \mid A_1 \cap A_2) = \frac{4}{13}$$

Therefore, using the multiplicative rule,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_1 \cap A_2) = \frac{12}{15} \cdot \frac{8}{14} \cdot \frac{4}{13}$$

1 Sets and probability models

2 Conditional probability

3 Total probability and Bayes' rule

4 Independence

5 Counting

Total probability theorem

A "divide-and-conquer" approach: assume we partition the sample space into a *disjoint* set $\{A_1, \ldots, A_n\}$ of events, so

$$\boxed{\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(A_1 \cap B) + \ldots + \mathbb{P}(A_n \cap B) && (31) \\ &= \mathbb{P}(A_1)\,\mathbb{P}(B \mid A_1) + \ldots + \mathbb{P}(A_n)\,\mathbb{P}(B \mid A_n) && (32) \end{aligned}}$$

This is useful when the "event" can occur due to a number of possible ways.

Example

A certain switch may be either on or off, and at the end of every 60 seconds, it may stay the same or flip. If it is originally on, there is a 0.8 probability to remain on. If it is originally off, there is a 0.6 probability to remain off. Assume at time zero it was on. What is the probability that it is on after 3 minutes?

Example

Let U_i and D_i be the event that the switch is on or off, respectively, after the ith minute. Note it was on at time zero.

- At time 1, $\mathbb{P}(U_1) = 0.8$ and $\mathbb{P}(D_1) = 0.2$.
- Then, at time 2,

$$\begin{aligned}\mathbb{P}(U_2) &= \mathbb{P}(U_1)\,\mathbb{P}(U_2 \,|\, U_1) + \mathbb{P}(D_1)\,\mathbb{P}(U_2 \,|\, D_1)\\ &= 0.8 \cdot 0.8 + 0.2 \cdot 0.4 = 0.72\\ \mathbb{P}(D_2) &= \mathbb{P}(U_1)\,\mathbb{P}(D_2 \,|\, U_1) + \mathbb{P}(D_1)\,\mathbb{P}(D_2 \,|\, D_1)\\ &= 0.8 \cdot 0.2 + 0.2 \cdot 0.6 = 0.28\end{aligned}$$

- Then, at time 3,

$$\begin{aligned}\mathbb{P}(U_3) &= \mathbb{P}(U_2)\,\mathbb{P}(U_3 \,|\, U_2) + \mathbb{P}(D_2)\,\mathbb{P}(U_3 \,|\, D_2)\\ &= 0.72 \cdot 0.8 + 0.28 \cdot 0.4 = 0.688\\ \mathbb{P}(D_3) &= \mathbb{P}(U_2)\,\mathbb{P}(D_3 \,|\, U_2) + \mathbb{P}(D_2)\,\mathbb{P}(D_3 \,|\, D_2)\\ &= 0.72 \cdot 0.2 + 0.28 \cdot 0.6 = 0.312\end{aligned}$$

Example

Extra (if interested): What about 30 minutes later? 3000 minutes?

Example

Extra (if interested): What about 30 minutes later? 3000 minutes?

We can write

$$\begin{bmatrix}\mathbb{P}(U_2)\\ \mathbb{P}(D_2)\end{bmatrix} = \begin{bmatrix}0.8 & 0.4\\ 0.2 & 0.6\end{bmatrix}\begin{bmatrix}\mathbb{P}(U_1)\\ \mathbb{P}(D_1)\end{bmatrix} = \begin{bmatrix}0.8 & 0.4\\ 0.2 & 0.6\end{bmatrix}^2\begin{bmatrix}\mathbb{P}(U_0)\\ \mathbb{P}(D_0)\end{bmatrix}$$

$$\begin{bmatrix}\mathbb{P}(U_n)\\ \mathbb{P}(D_n)\end{bmatrix} = \begin{bmatrix}0.8 & 0.4\\ 0.2 & 0.6\end{bmatrix}^n\begin{bmatrix}\mathbb{P}(U_0)\\ \mathbb{P}(D_0)\end{bmatrix}$$

As $n \to \infty$, (we will not discuss how to do this calculation)

$$\begin{bmatrix}0.8 & 0.4\\ 0.2 & 0.6\end{bmatrix}^n \to \begin{bmatrix}0.67 & 0.67\\ 0.33 & 0.33\end{bmatrix}$$

So, $\mathbb{P}(U_n) = 0.67$ for a large n *no matter what the initial state is!*

Bayes' rule

Assume we partition the sample space into a *disjoint* set $\{A_1, \ldots, A_n\}$ of events, so

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i)}{\mathbb{P}(B)} \tag{33}$$

$$= \frac{\mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i)}{\mathbb{P}(A_1)\,\mathbb{P}(B \mid A_1) + \ldots + \mathbb{P}(A_n)\,\mathbb{P}(B \mid A_n)} \tag{34}$$

1. For inference or inverse problem: from *output* to *input*
2. $\mathbb{P}(A_i)$ is prior probability
3. $\mathbb{P}(A_i \mid B)$ is posterior probability

Bayes' rule is an extremely important tool for us!

Example

How do we know when to use Bayes' rule to solve a problem? Let's consider this example:

A person is infected with one of three possible diseases, which cause identical symptoms. The probability of it being disease A is 0.5, B is 0.25, and C is also 0.25. A certain drug is effective against A with probability 0.3, against B with probability 0.4, and against C with probability 0.5. The drug eventually is shown to work. What is the probability that it was disease A?

Example

Analysis of the problem:

- The question asks for: the probability that "it was disease A" *given* "the drug eventually is shown to work"
- *Before* knowing the result of the drug: those are the prior probabilities
- *After* knowing the result of the drug: those are the posterior probabilities (and what we need to find)
- If we flip the question and ask: the probability that "the drug eventually is shown to work" *given* "it was disease A" — this is straightforward and is, in fact, given in the question

The above are strong indications that Bayes' rule would be applicable. Now let's put them into symbols.

Example

Let E denote effectiveness. Then,

$$\mathbb{P}(A) = 0.5, \qquad \mathbb{P}(B) = 0.25, \qquad \mathbb{P}(C) = 0.25,$$

and

$$\mathbb{P}(E \mid A) = 0.3, \qquad \mathbb{P}(E \mid B) = 0.4, \qquad \mathbb{P}(E \mid C) = 0.5.$$

We can then directly apply Bayes' rule,

$$\begin{aligned}\mathbb{P}(A \mid E) &= \frac{\mathbb{P}(A)\,\mathbb{P}(E \mid A)}{\mathbb{P}(A)\,\mathbb{P}(E \mid A) + \mathbb{P}(B)\,\mathbb{P}(E \mid B) + \mathbb{P}(C)\,\mathbb{P}(E \mid C)} \\ &= \frac{0.5 \cdot 0.3}{0.5 \cdot 0.3 + 0.25 \cdot 0.4 + 0.25 \cdot 0.5} = 0.4\end{aligned}$$

Example

Another example:

A test for a disease is correct 95% of the time. Probability of a random person having that disease is 0.001. If one is just tested positive, what is the probability of having that disease?

Note: When there are only two possibilities, given the probability of one of them implies the other!

Example

Let A = has disease, and B = tested positive. We are interested in $\mathbb{P}(A \mid B)$.

We know:

$$\mathbb{P}(A) = 0.001 \qquad \mathbb{P}(B \mid A) = 0.95$$
$$\mathbb{P}(A^c) = 0.999 \qquad \mathbb{P}(B \mid A^c) = 0.05$$

Therefore,

$$\begin{aligned}\mathbb{P}(A \mid B) &= \frac{\mathbb{P}(A)\,\mathbb{P}(B \mid A)}{\mathbb{P}(A)\,\mathbb{P}(B \mid A) + \mathbb{P}(A^c)\,\mathbb{P}(B \mid A^c)} \\ &= \frac{0.001 \cdot 0.95}{0.001 \cdot 0.95 + 0.999 \cdot 0.05} = 0.0187\end{aligned}$$

Further question: What if the test is 99% correct? 100% correct?

Example: The Three Prisoners problem (from Supplementary Book)

Once upon a time, there were three prisoners A, B, and C. One day, the king decided to pardon two of them and sentence the last one. Prisoner A heard the news and wanted to ask a friendly guard about his situation. The guard was allowed to tell him that prisoner B or C would be pardoned, but he could not tell A whether he would be pardoned. Prisoner A reasoned:

- Without asking, his probability of being pardoned is $2/3$.
- However, if he asks the guard, this probability reduces to $1/2$ because the guard would tell him which one of the two other prisoners would be pardoned. Then, only he and the other are left.

Should prisoner A ask the guard? What has gone wrong with his reasoning?

Example: The Three Prisoners problem

Let X_A, X_B, X_C be the events of sentencing prisoners A, B, C, respectively.

- Without asking:

$$\mathbb{P}(X_A) = \frac{1}{3}, \qquad \mathbb{P}(X_B) = \frac{1}{3}, \qquad \mathbb{P}(X_C) = \frac{1}{3}$$

- Suppose A asked. Let G_B be the event that the guard says that the prisoner B is released. $\mathbb{P}(G_B \mid X_B) = 0$ because the guard is honest, and $\mathbb{P}(G_B \mid X_C) = 1$ because the guard has no choice. Also, $\mathbb{P}(G_B \mid X_A) = 1/2$ because the guard is equally likely to say B or C.

What we are really after is $\mathbb{P}(X_A \mid G_B)$.

Example: The Three Prisoners problem

Applying Bayes' rule,

$$\mathbb{P}(X_A \mid G_B) = \frac{\mathbb{P}(X_A)\,\mathbb{P}(G_B \mid X_A)}{\mathbb{P}(X_A)\,\mathbb{P}(G_B \mid X_A) + \mathbb{P}(X_B)\,\mathbb{P}(G_B \mid X_B) + \mathbb{P}(X_C)\,\mathbb{P}(G_B \mid X_C)}$$
$$= \frac{\frac{1}{3}\cdot\frac{1}{2}}{\frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1} = \frac{1}{3}$$

So, $\mathbb{P}(X_A \mid G_B) = \mathbb{P}(X_A) = \frac{1}{3}$.

Therefore, asking the guard would not reduce A's chance of being pardoned. It makes no difference.

1 Sets and probability models

2 Conditional probability

3 Total probability and Bayes' rule

4 Independence

5 Counting

Definition

Independence of two events A and B means

$$\boxed{\mathbb{P}(A \mid B) = \mathbb{P}(A)} \tag{35}$$

as well as

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)} \tag{36}$$

- Literally: Knowing B gives no information about A
- They are equivalent (assuming $\mathbb{P}(B) \neq 0$) since

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

- Relationship is symmetric:

$$A \text{ is independent of } B \Leftrightarrow B \text{ is independent of } A$$

Example

We roll two dice.

1. Are these two events independent?

$$A = \{1\text{st roll} = 1\} \qquad B = \{\text{sum} = 7\}$$

2. Are these two events independent?

$$C = \{1\text{st roll} = 1\} \qquad D = \{\text{sum} = 6\}$$

3. Are these two events independent?

$$E = \{\text{maximum of the two} = 4\} \qquad F = \{\text{minimum of the two} = 2\}$$

Example

1. We know $\mathbb{P}(A) = \frac{1}{6}$, $\mathbb{P}(B) = \frac{6}{36}$, and $\mathbb{P}(A \cap B) = \frac{1}{36}$. So they are independent.
2. We know $\mathbb{P}(C) = \frac{1}{6}$, $\mathbb{P}(D) = \frac{5}{36}$, and $\mathbb{P}(C \cap D) = \frac{1}{36}$. So they are not independent.
3. We know $\mathbb{P}(E) = \frac{7}{36}$, $\mathbb{P}(F) = \frac{9}{36}$, and $\mathbb{P}(E \cap F) = \frac{2}{36}$. So they are not independent.

Definition

Two events A and B can have conditional independence given event C:

$$\boxed{\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\,\mathbb{P}(B \mid C)} \tag{37}$$

or alternatively,

$$\boxed{\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid C)} \tag{38}$$

- Literally: *When C is known, then* knowing B gives no information about A
- They are equivalent (assuming $\mathbb{P}(B \mid C) \neq 0$)
- Relationship is symmetric:

$$A \mid C \text{ is independent of } B \mid C \Leftrightarrow B \mid C \text{ is independent of } A \mid C$$

Independence vs conditional independence

A and B are independent does not imply they are conditional independent, and vice versa

1. We have two coin tosses.

$$A = \{\text{1st toss is head}\}$$
$$B = \{\text{2nd toss is head}\}$$
$$C = \{\text{two tosses have different results}\}$$

   Are A and B independent? How about $A \mid C$ and $B \mid C$?

2. We have a blue coin (B) and a red coin (R), equally likely to be chosen and then tossed twice, and that $\mathbb{P}(H \mid B) = 0.99$ and $\mathbb{P}(H \mid R) = 0.01$. Let

$$D_i = \{\text{ith toss is H}\}.$$

   Are D_1 and D_2 independent? How about $D_1 \mid B$ and $D_2 \mid B$?

Example

1. We can check

$$\mathbb{P}(A) = \tfrac{1}{2} \qquad \mathbb{P}(B) = \tfrac{1}{2} \qquad \mathbb{P}(A \cap B) = \tfrac{1}{4}$$
$$\mathbb{P}(A \mid C) = \tfrac{1}{2} \qquad \mathbb{P}(B \mid C) = \tfrac{1}{2} \qquad \mathbb{P}(A \cap B \mid C) = 0$$

So A and B are independent, but $A \mid C$ and $B \mid C$ are not.

2. We can compute, for $i = 1, 2$,

$$\begin{aligned}
\mathbb{P}(D_i) &= \mathbb{P}(B)\,\mathbb{P}(D_i \mid B) + \mathbb{P}(B^c)\,\mathbb{P}(D_i \mid B^c) \\
&= \tfrac{1}{2} \cdot 0.99 + \tfrac{1}{2} \cdot 0.01 = \tfrac{1}{2} \\
\mathbb{P}(D_1 \cap D_2) &= \mathbb{P}(B)\,\mathbb{P}(D_1 \cap D_2 \mid B) + \mathbb{P}(B^c)\,\mathbb{P}(D_1 \cap D_2 \mid B^c) \\
&= \tfrac{1}{2} \cdot 0.99 \cdot 0.99 + \tfrac{1}{2} \cdot 0.01 \cdot 0.01 = 0.4901 \\
\mathbb{P}(D_i \mid B) &= 0.99 \\
\mathbb{P}(D_1 \cap D_2 \mid B) &= 0.99 \cdot 0.99
\end{aligned}$$

So D_1 and D_2 are not independent, but $D_1 \mid B$ and $D_2 \mid B$ are.

Three of more events

When we have three events A_1, A_2, and A_3, independence means satisfying all four conditions:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2) \tag{39}$$
$$\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2)\,\mathbb{P}(A_3) \tag{40}$$
$$\mathbb{P}(A_3 \cap A_1) = \mathbb{P}(A_3)\,\mathbb{P}(A_1) \tag{41}$$
$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\,\mathbb{P}(A_3) \tag{42}$$

Eqs. (39)–(41) stipulate pairwise independence.

More generally, for events $A_1, A_2, \ldots, A_n$, they are independent if

$$\mathbb{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i) \qquad \text{for every subset } S \text{ of } \{1, 2, \ldots, n\} \tag{43}$$

Example

Examples:

1. We have two coin tosses.

$$A = \{\text{1st toss is head}\}$$
$$B = \{\text{2nd toss is head}\}$$
$$C = \{\text{two tosses have different results}\}$$

   Do A, B, and C satisfy Eqs. (39)–(42)? Are they independent?

2. We have two dice rolls.

$$D = \{\text{1st roll is one of } \{1, 2, 3\}\}$$
$$E = \{\text{1st roll is one of } \{3, 4, 5\}\}$$
$$F = \{\text{sum of two rolls is equal to } 9\}$$

   Do D, E, and F satisfy Eqs. (39)–(42)? Are they independent?

Example

1. We can check

$$\mathbb{P}(A) = \frac{1}{2} \qquad \mathbb{P}(B) = \frac{1}{2} \qquad \mathbb{P}(A \cap B) = \frac{1}{4}$$
$$\mathbb{P}(C) = \frac{1}{2} \qquad \mathbb{P}(A \cap C) = \frac{1}{4} \qquad \mathbb{P}(B \cap C) = \frac{1}{4}$$
$$\mathbb{P}(A \cap B \cap C) = 0$$

So A, B, and C are pairwise independent, but fail Eq. (42).

2. We can check

$$\mathbb{P}(D) = \frac{1}{2} \qquad \mathbb{P}(E) = \frac{1}{2} \qquad \mathbb{P}(F) = \frac{4}{36}$$
$$\mathbb{P}(D \cap E) = \frac{1}{6} \qquad \mathbb{P}(E \cap F) = \frac{3}{36} \qquad \mathbb{P}(D \cap F) = \frac{1}{36}$$
$$\mathbb{P}(D \cap E \cap F) = \frac{1}{36}$$

So D, E, and F satisfy Eq. (42), but fail pairwise independence.

Example

The DNA of a bacteria is a length 1 million circular sequence containing the nucleotides A, G, C and T. We sequence the DNA by sampling $100,000$ reads randomly from the DNA, where each read is a length 100 substring of the genome. We are interested in particular in a specific gene, which is a specific length 1000 substring of the DNA.

1. What is the probability that the first two reads overlap?
2. What is the probability that a specific nucleotide of the DNA is covered by at least one read?
3. What is the probability that there is at least one read that is completely contained in the gene of interest?

Example

ANS: Let $X_i \in \{1, \ldots, 1000000\}$ be the starting location of the ith read, where $i = 1, \ldots, 100000$.

1. By symmetry, since the sequence is circular, we can assume X_1 to be any number, so let's pick $X_1 = 100$. Then we need $X_2 \in [1, 199]$. So probability is $199/1000000$.

Example

ANS: Let $X_i \in \{1, \ldots, 1000000\}$ be the starting location of the ith read, where $i = 1, \ldots, 100000$.

1. By symmetry, since the sequence is circular, we can assume X_1 to be any number, so let's pick $X_1 = 100$. Then we need $X_2 \in [1, 199]$. So probability is 199/1000000.
2. Now assume the specific nucleotide is located at position 100, and we are interested in the event "$X_i \leqslant 100$ for some i".

$$\mathbb{P}(X_i \leqslant 100 \text{ for some } i) = 1 - \left(\frac{1000000 - 100}{1000000}\right)^{100000} \approx 1 - e^{-10}$$

since $e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$, with $x = -10$.

Example

ANS: Let $X_i \in \{1, \ldots, 1000000\}$ be the starting location of the ith read, where $i = 1, \ldots, 100000$.

1. By symmetry, since the sequence is circular, we can assume X_1 to be any number, so let's pick $X_1 = 100$. Then we need $X_2 \in [1, 199]$. So probability is 199/1000000.
2. Now assume the specific nucleotide is located at position 100, and we are interested in the event "$X_i \leqslant 100$ for some i".

$$\mathbb{P}(X_i \leqslant 100 \text{ for some } i) = 1 - \left(\frac{1000000 - 100}{1000000}\right)^{100000} \approx 1 - e^{-10}$$

since $e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$, with $x = -10$.

3. Now assume the gene of interest is at location 1–1000, so the event is "$X_i \leqslant 901$ for some i". Similar to the above,

$$\mathbb{P}(X_i \leqslant 901 \text{ for some } i) = 1 - \left(\frac{1000000 - 901}{1000000}\right)^{100000} \approx 1 - e^{-90.1}$$

1 Sets and probability models

2 Conditional probability

3 Total probability and Bayes' rule

4 Independence

5 Counting

Probability by counting

Many situations require us to "count" to obtain the probability:

1. An event A consists of a finite number of equally likely outcomes, each having probability p,

$$\mathbb{P}(A) = p \cdot (\text{number of elements of } A)$$

2. There are a finite number of equally likely outcomes in A, so

$$\mathbb{P}(A) = \frac{\text{number of elements of } A}{\text{number of elements of } \Omega}$$

Some notations

- Factorial:

$$\boxed{n! = n(n-1)(n-2)\cdots 2\cdot 1} \tag{44}$$

- k-permutation:

$$\boxed{n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}} \tag{45}$$

- Combination:

$$\boxed{\binom{n}{k} = \frac{n!}{k!\,(n-k)!}} \tag{46}$$

Some (useful) formulas

$$0! = 1$$

Stirling's approx. $\quad n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \quad$ for large n

Gamma function $\quad n! = \Gamma(n+1) \quad$ where $\quad \Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$

$$\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} = 1$$

$$\sum_{k=0}^{n}\binom{n}{k} = 2^n$$

$$\sum_{k=1}^{n}k\binom{n}{k} = n2^{n-1}$$

Bernoulli trials

A sequence of independent trials with two possible outcomes, e.g., Head vs Tail, is called Bernoulli trials, where

$$\mathbb{P}(\text{Head}) = p \qquad \mathbb{P}(\text{Tail}) = 1 - p$$

Probability of seeing k heads followed by $n - k$ tails is

$$p^k (1-p)^{n-k}$$

There are $\binom{n}{k}$ possibilities for the k heads to appear in the sequence.

Overall, the probability of having k heads in n trials is

$$\boxed{p(k) = \binom{n}{k} p^k (1-p)^{n-k}} \tag{47}$$

Counting outcomes

Examples:

1. Number of possible 4-digit code:

Counting outcomes

Examples:

1. Number of possible 4-digit code:

$$10 \times 10 \times 10 \times 10 = 10^4$$

2. Number of possible 4-digit code that cannot have duplicate:

Counting outcomes

Examples:

1. Number of possible 4-digit code:

$$10 \times 10 \times 10 \times 10 = 10^4$$

2. Number of possible 4-digit code that cannot have duplicate:

$$10 \times 9 \times 8 \times 7 = \frac{10!}{6!}$$

3. Number of ways to arrange 4 digits:

Counting outcomes

Examples:

1. Number of possible 4-digit code:

$$10 \times 10 \times 10 \times 10 = 10^4$$

2. Number of possible 4-digit code that cannot have duplicate:

$$10 \times 9 \times 8 \times 7 = \frac{10!}{6!}$$

3. Number of ways to arrange 4 digits:

$$4 \times 3 \times 2 \times 1 = 4!$$

4. Number of ways to pick 4 numbers without duplicate:

Counting outcomes

Examples:

1. Number of possible 4-digit code:

$$10 \times 10 \times 10 \times 10 = 10^4$$

2. Number of possible 4-digit code that cannot have duplicate:

$$10 \times 9 \times 8 \times 7 = \frac{10!}{6!}$$

3. Number of ways to arrange 4 digits:

$$4 \times 3 \times 2 \times 1 = 4!$$

4. Number of ways to pick 4 numbers without duplicate:

$$\frac{10 \times 9 \times 8 \times 7}{4!} = \binom{10}{4} = \binom{10}{6}$$

Partitions

Begin with a set containing n elements; divide it into r disjoint subsets, with subset i having a size of n_i, and that $\sum_{i=1}^{r} n_i = n$. How many ways can we do it?

Partitions

Begin with a set containing n elements; divide it into r disjoint subsets, with subset i having a size of n_i, and that $\sum_{i=1}^{r} n_i = n$. How many ways can we do it?

Ans: First partition into two sets with n_1 elements and $n - n_1$ elements. Then, partition the latter into n_2 and $(n - n_1) - n_2$. Then, partition the latter into n_3 and $(n - n_1 - n_2) - n_3 \ldots$

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r}$$

$$= \frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1-\cdots-n_{r-1})!}{(n-n_1-\cdots-n_r)!n_r!}$$

$$= \frac{n!}{n_1!\,n_2!\cdots n_r!} = \binom{n}{n_1, n_2, \cdots, n_r}$$

Example

How many ways to arrange the letters in HONGKONGYOU?

Example

How many ways to arrange the letters in HONGKONGYOU?

Ans: 11 letters mean 11! permutations, but there are 3! permutations of "O", 2! permutations of "N", and 2! permutations of "G", hence

$$\frac{11!}{3!\,2!\,2!}$$

Example

How many ways to arrange the letters in HONGKONGYOU?

Ans: 11 letters mean 11! permutations, but there are 3! permutations of "O", 2! permutations of "N", and 2! permutations of "G", hence

$$\frac{11!}{3!\,2!\,2!}$$

Ans: There are 11 positions for the letters. We partition them into a set of 3 (where we put the "O"), two sets of 2 (where we use one for "N", one for "G"), and 4 sets of 1 ("H", "K", "Y", "U"), hence

$$\binom{11}{3,2,2,1,1,1,1} = \frac{11!}{3!\,2!\,2!\,1!\,1!\,1!\,1!}$$

Example

We divide a class of 16 students — 4 male, 12 female — into 4 groups of 4 students each. What is the probability that each group has exactly one male student? (We solved before, but here we use partitions)

Example

We divide a class of 16 students — 4 male, 12 female — into 4 groups of 4 students each. What is the probability that each group has exactly one male student? (We solved before, but here we use partitions)

Ans: Together, there are $\binom{16}{4,4,4,4} = \frac{16!}{4!\,4!\,4!\,4!} = A$ partitions.

For the male, there are $\binom{4}{1,1,1,1} = \frac{4!}{1!\,1!\,1!\,1!} = B$ partitions.

For the female, there are $\binom{12}{3,3,3,3} = \frac{12!}{3!\,3!\,3!\,3!} = C$ partitions.

Overall probability $= \frac{B \cdot C}{A} = \frac{4!\,12!}{3!\,3!\,3!\,3!} \cdot \frac{4!\,4!\,4!\,4!}{16!} = \frac{12 \cdot 8 \cdot 4}{15 \cdot 14 \cdot 13}$

Summary

We have outlined the basic mathematical foundation for calculating probabilities. Next, we will concentrate our attention to assigning numerical values to different outcomes, and proceed with developing tools to compute and interpret those values. We will divide broadly into two situations:

- Discrete probability
- Continuous probability

and introduce the notion of *random variables*.