Lecture 2: Events and Probability

Foundation of Probability Theory/STA 203

Zhuosong ZHANG

Department of Statistics and Data Science, SUSTech

Fall, 2023

Outcomes, sample space, and events

Sample space and events

In order to define probability model, we first consider sample space and events.

Definition 2 (Sample space)

This set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by Ω.

Example 3

(a) If the outcome of an experiment consists in the determination of the sex of a newborn child, then

$$\Omega = \{g, b\}.$$

(b) If the outcome of an experiment is the order of finish in a race among the 7 horses having post positions 1, 2, 3, 4, 5, 6, and 7, then,

$$\Omega = \{\text{all } 7! \text{ permutations of } (1, 2, 3, 4, 5, 6, 7)\}.$$

Example, cont'd

Example 4 (Your turn)

(a) If the experiment consists of flipping two coins, then the sample space consists of the following four points:

(b) If the experiment consists of measuring (in hours) the lifetime of a Television, then the sample space is :

Definition 5 (Events)

Any subset of the sample space is known as an event, which is denoted by capital letters A, B, E and so on.

Example 6

(a) $A = \{g\}$.

(b) $B = \{$all outcomes in Ω starting with a 3$\}$.

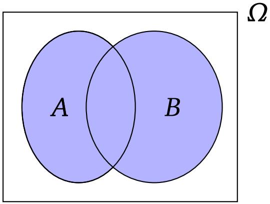
(c) $E = \{x : 0 \leqslant x \leqslant 5\}$.

Event operations: definitions

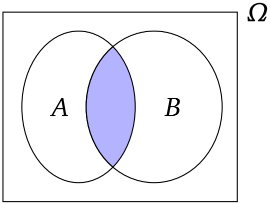
(i) Either A or B happens: Union $A \cup B$.

(ii) Both A and B happens: Intersection $A \cap B$. (In some textbooks, $A \cap B$ is also written as AB for brevity.)

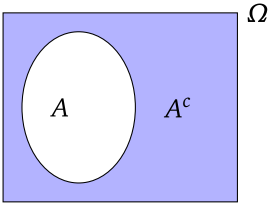
(iii) A does not happen: complement A^c.



Union $A \cup B$



Intersection $A \cap B$



Complement A^c

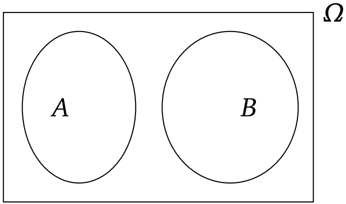
Event operations: definitions, cont'd

(iv) Null event $\varnothing := \Omega^c$.

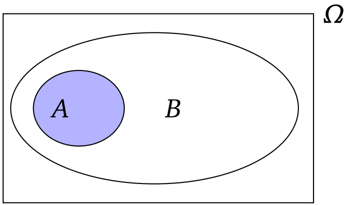
(v) If $A \cap B = \varnothing$, then A and B are said to be mutually exclusive, or disjoint.

(vi) If all of the outcomes in A are also in F, then we say A is contained in B, or A is a subset of B, denoted by $A \subset B$.

(vii) If $A \subset B$ and $B \subset A$, then $A = B$.



Mutually exclusive



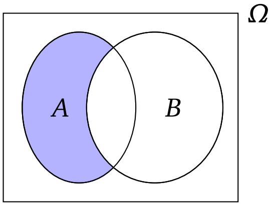
Subset: $A \subset B$

Event operations: definitions, cont'd

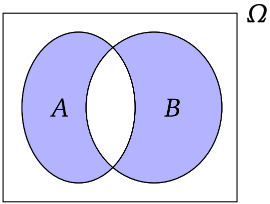
(viii) Difference: $A \setminus B = A \cap B^c$. (In some textbooks, $A \setminus B$ is written as $A - B$.)

(ix) Symmetric difference: $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

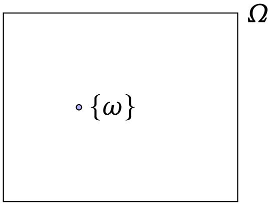
(x) Singleton: $\{\omega\}$.



Difference: $A \setminus B$



Symmetric difference: $A \triangle B$



Singleton

Set Operators and their meanings in Probability Theory

Set Operator	Notation	Description	Meaning in Probability
Union	$A \cup B$	$x \in A$ or $x \in B$	either A or B or both occur
Intersection	$A \cap B$	$x \in A$ and $x \in B$	both A and B occur
Complement	A^c	$x \notin A$	A does not occur
Difference	$A \backslash B$	$x \in A$ but $x \notin B$	A occurs but B does not
Sym. Diff.	$A \Delta B$	$x \in A \cup B$ but $x \notin A \cap B$	either A or B occurs but not both

Table: Set Operators and their meanings in Probability Theory

Event operations: Some propositions

Proposition 7

(i) The operations of forming unions, intersections, and complements of events obey certain rules similar to the rules of algebra. We list a few of these rules:

(a) Commutative laws: $A \cup B = B \cup A$, $A \cap B = B \cap A$.

(b) Associative laws:

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C).$$

(c) Distributive laws:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C),$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Proposition 8

(ii) DeMorgan's laws:

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c,$$

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

Classical probability

Classical probability

- In classical probability, it is assumed that all outcomes in the sample space are equally likely to occur.
- Consider N is a natural number, and $\Omega = \{1, \ldots, N\}$, and $\mathscr{F} = 2^{\Omega}$ is the power set of Ω.
- Key assumption: Assume that

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \cdots = \mathbb{P}(\{N\}).$$

- Then, by axioms of probability, $\mathbb{P}(\{i\}) = 1/N$ for each $1 \leqslant i \leqslant N$ (why?).
- For any $E \in \mathscr{F}$,

$$\mathbb{P}(E) = \frac{\text{\# of elements in } E}{N} = \frac{|E|}{N}.$$

- Here, $|E|$ means the number of elements in E.

R codes

- The experiments can be done using numerical methods.
- In this course, we illustrate the methods using R.
- If we want to toss a coin once or 10 times:

```
Omega <- c("head","tail")
sample(x=Omega,size= 1,replace=TRUE)
[1] "tail"
sample(x=Omega,size=10,replace=TRUE)  # why replace should be true?
[1] "head" "tail" "head" "tail" "head" "tail" "head" "tail"
[9] "tail" "head"
```

- Whether your result is as same as mine?
- Use the `set.seed()` function.

```
set.seed(1)  # you can change 1 to your favorate number
```

Example 9 (Dice)

If two dice are rolled, what is the probability that the sum of the upturned faces will equal 7?

Solution.

The sample space is $\Omega = \{(i,j) : 1 \leqslant i,j \leqslant 6\}$. It follows that $N = 36$. Let $\mathscr{F}$ be the power set of Ω. Let $E = \{$the sum of the upturned faces equals to 7$\}$, then

$$E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\},$$

implying that $|E| = 6$ and thus $\mathbb{P}(E) = \dfrac{6}{36} = \dfrac{1}{6} \approx 0.1667.$ ■

Examples

- We can estimate this probability by using repeated experiments.
- Suppose we roll two dice, and the sum is $3 + 5 = 8$:

```
sample(x=1:6,size=2,replace=TRUE)
[1] 3 5
```

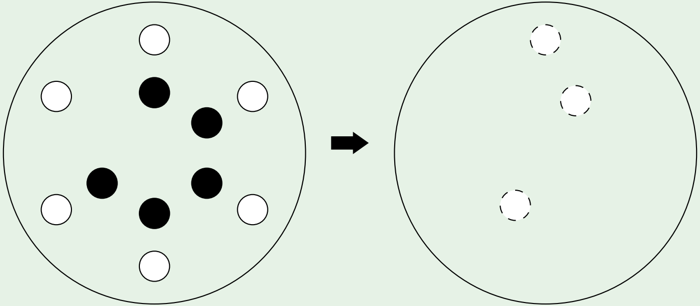
- We can repeat this experiments for 100 times, and count the numbers that the sum is 7:

```
dicesum <- numeric(100)
for (i in 1:100){
  dicesum[i] <- sum(sample(x=1:6,size=2,replace=TRUE))
}
sum(dicesum==7)
[1] 16
```

- Then, the probability can be estimated by $16/100 = 0.16$.

Example 10 (Balls)

If 3 balls are "randomly drawn" from a bowl containing 6 white and 5 black balls, what is the probability that one of the balls is white and the other two black?



Solution.

Regard the outcome of the experiment as the unordered set of drawn balls. Then, $N = \binom{11}{3} = 165$. Let E denote the event that one of the balls is white and the other two black:

$$E = \{\text{one white and two black balls are selected}\},$$

then the number of elements in E is

$$|E| = \binom{6}{1}\binom{5}{2} = 60.$$

If all outcomes are assumed equally likely, then, the desired probability is

$$\mathbb{P}(E) = \frac{|E|}{N} = \frac{60}{165} = \frac{4}{11} \approx 0.3636. \qquad \blacksquare$$

Examples

- Define all the balls in the bowl:

```
balls <- c(rep("White",6), rep("Black",5))
```

- Draw 3 balls without replacement:

```
sample(x=balls,size=3,replace=FALSE)
[1] "White" "Black" "White"
```

- Count the number of white balls:

```
sam <- sample(x=balls,size=3,replace=FALSE)
sum(sam=="White")
[1] 2
```

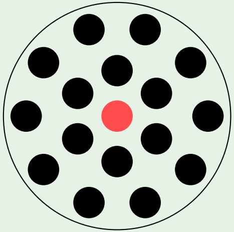
Examples

- Therefore, we can design our codes as follows:

```
M <- 10000
whitenum <- numeric(M)
for (i in 1:M) {
  sam <- sample(x=balls,size=3,replace=FALSE)
  whitenum[i] <- sum(sam=="White")
}
sum(whitenum==1)/M
[1] 0.3638
```

Example 11 (Balls selection)

An urn contains n balls, one of which is special. If k of these balls are withdrawn one at a time, with each selection being equally likely to be any of the balls that remain at the time, what is the probability that the special ball is chosen?



n balls with 1 special

Solution.

Since all of the balls are treated in an identical manner, it follows that the set of k balls selected is equally likely to be any of the $\binom{n}{k}$ sets of k balls. Therefore,

$$\mathbb{P}\{\text{special ball is selected}\} = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}. \qquad \blacksquare$$

Axioms of classical probability

Can we give some general properties for probability?
In classical probability, suppose that the sample space is $\Omega = \{\omega_1, \dots, \omega_N\}$. We have the following proposition:

Proposition 12 (Counting formula)

For any disjoint $A, B \subset \Omega$,

$$|A \cup B| = |A| + |B|.$$

As a consequence,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Finite additivity

Proposition 13 (Finite additivity)

Let $A_1, A_2, \ldots, A_n$ be n disjoint events, then

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

Normalization

The sample space Ω itself is also an event, which is called the certain event (必然事件).

Proposition 14 (Normalization)

We have

$$\mathbb{P}(\Omega) = 1.$$

Nonnegativity

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 15

For any $A \subset \Omega$,

$$\mathbb{P}(A) \geqslant 0.$$

Some basic corollaries

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Corollary 16

Based on Propositions 13, 14 and 15, We have the following results:

(a) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(b) For any A and B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leqslant \mathbb{P}(A) + \mathbb{P}(B)$.

(c) For any A and B, $\mathbb{P}(A \cap B) + \mathbb{P}(A \setminus B) = \mathbb{P}(A)$.

(d) If $B \subset A$, then $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B) \geqslant 0$.

The inclusive-exclusive property

Corollary 17 (The inclusive-exclusive property)

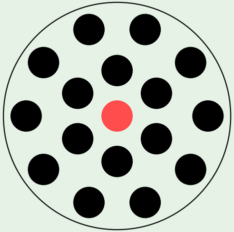
For any $E_1, \ldots, E_n \subset \Omega$,

$$\begin{aligned}
&\mathbb{P}(E_1 \cup E_2 \cup \cdots \cup E_n) \\
&\quad = \sum_{i=1}^{n} \mathbb{P}(E_i) + \cdots + (-1)^{r+1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) \\
&\quad\quad + \cdots + (-1)^{n+1} \mathbb{P}(E_1 \cap \cdots \cap E_n).
\end{aligned}$$

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 18 (Ball replacement)

An urn contains $n-1$ black balls and 1 red ball. At each time, a ball is randomly drawn from the urn and replaced with a black ball. What is the probability of drawing a black ball on the kth draw?



$n-1$ black balls and 1 red ball

Solution.

Let A be the event that a black ball is drawn on the kth draw, and it follows that A^c is the event that the red ball is draw on the kth draw. Note that there is only one red ball in the urn, and thus

$$A^c = \{\text{black balls on the first } k-1\text{th draws, and the red ball on the } k\text{th draw.}\}$$

Therefore,

$$\mathbb{P}(A^c) = \frac{(n-1)^{k-1}\cdot 1}{n^k} = \frac{1}{n}\left(1-\frac{1}{n}\right)^{k-1},$$

and by (a) in Corollary 16,

$$\mathbb{P}(A) = 1-\mathbb{P}(A^c) = 1-\frac{1}{n}\left(1-\frac{1}{n}\right)^{k-1}. \quad \blacksquare$$

Example 19 (Poker)

A deck of poker cards consists 52 cards (without Jokers). A poker hand consists of 5 cards. If the cards have distinct consecutive values and are not all of the same suit, we say that the hand is a straight.

For instance, a poker hand consisting of the following are all straights:

5♠ 6♦ 7♠ 8♥ 9♣

A♣ 2♦ 3♠ 4♥ 5♥

10♠ J♠ Q♠ K♠ A♥

What is the probability that one is dealt a straight?

Solution.

The total number of possible poker hands is

$$N = \binom{52}{5} = 2,598,960.$$

Let A be the event that the five cards have distinct consecutive values, and B be the event that the cards are all of the same suit. Let E be the event that the hand of cards is a straight. Then,

$$E = A \cap B^c.$$

Then, by Proposition (e),

$$\mathbb{P}(E) = \mathbb{P}(A) - \mathbb{P}(A \cap B).$$

Note that

$$\mathbb{P}(A) = \frac{|A|}{N} = \frac{10 \times 4^5}{N}, \quad \mathbb{P}(A \cap B) = \frac{|A \cap B|}{N} = \frac{10 \times 4}{N},$$

then,

$$\mathbb{P}(E) = \frac{10,200}{2,598,960} \approx 0.0039. \qquad \blacksquare$$

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 20 (Dou dizhu, 斗地主)

Dou dizhu (fighting the landlord, 斗地主) is one of the most popular card games played in China. It is played among three people with one pack of cards, including the two differentiated jokers. A shuffled pack of 54 cards is dealt to three players, Each "peasant (农民)" is dealt 17 cards, while the "landlord (地主)" is dealt 20 cards.

A Rocket (火箭) is the Colored Joker (大王) and black-and-white Joker (小王), and a bomb (炸弹) is 4 cards of the same rank.

What is the probability that the landlord receive a rocket but no bombs?

Example 21 (The matching problem)

Suppose that each of n men at a party throws his hat into the center of the room. The hats are first mixed up, and then each man randomly selects a hat. What is the probability that none of the men selects his own hat?

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Solution.

We first calculate the complementary probability of at least one man's selecting his own hat. Denote by E_i the event that the ith selects his own hat, where $i = 1, \ldots, n$. Then, by the inclusion-exclusion property,

$$
\begin{aligned}
&\mathbb{P}(E_1 \cup E_2 \cup \cdots \cup E_n) \\
&\quad = \sum_{i=1}^{n} \mathbb{P}(E_i) + \cdots + (-1)^{r+1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) \\
&\quad\quad + \cdots + (-1)^{n+1} \mathbb{P}(E_1 \cap \cdots \cap E_n).
\end{aligned}
$$

Examples

For any r, we have $\mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_r}) = \frac{(n-r)!}{n!}$, $\sum_{1 \leqslant i_1 < \cdots < i_r \leqslant n} \mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_r}) = \binom{n}{r} \frac{(n-r)!}{n!} = \frac{1}{r!}$.
Therefore,

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i \right) = 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + (-1)^{n+1} \frac{1}{n!}.$$

Let $E = \{\text{none of the men selects his own hat}\}$, and it follows that $E = \left(\bigcup_{i=1}^{n} E_i \right)^c$, which gives

$$\mathbb{P}(E) = 1 - 1 + \frac{1}{2!} + \cdots + (-1)^n \frac{1}{n!} \to e^{-1} \quad \text{if } n \to \infty.$$

■

Probability on a general set

Example 22

Let Ω be a countable set: $\Omega = \{\omega_j, j \in J\}$, where J is the collection of finite or countable index set. Let $\mathscr{F}$ be the power set of Ω. Choose any sequence of numbers $\{p_j, j \in J\}$ satisfying that

$$\forall j \in J : p_j \geqslant 0; \quad \sum_{j \geqslant 1} p_j = 1.$$

Define a set function $\mathbb{P} : \mathscr{F} \longrightarrow [0,1]$ as $\forall E \in \mathscr{F} : \mathbb{P}(E) = \sum_{\omega_j \in E} p_j$. Then, $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space.

Remark

In words, we assign p_j as the value of probability of the singleton $\{\omega_j\}$, and for an arbitrary set E of ω_j's we assign a probability the sum of all the probabilities assigned to its elements.

A measure of belief

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 23 (Probability in real life)

(i) It is 40% probable that 吴承恩 actually wrote the novel 《西游记》.

(ii) The probability that Oswald acted alone in assassinating Kennedy is 0.8.

(iii) You are 90% sure that you will receive an A+ as long as you work hard in this course.

- The most simple and natural interpretation is that the probabilities referred to are measures of the individual's degree of belief in the statements that he or she is making.
- Whether we interpret probability as a measure of belief or as a long-run frequency of occurrence, its mathematical properties remain unchanged.

Examples

Example 24

Suppose that, in a 7-horse race, you feel that

- each of the first 2 horses has a 20 percent chance of winning,
- horses 3 and 4 each have a 15 percent chance, and
- the remaining 3 horses have a 10 percent chance each.

Would it be better for you to wager at even money that the winner will be one of the first three horses or to wager, again at even money, that the winner will be one of the horses 1, 5, 6, and 7?

Proof.

The sample space is $\Omega = \{1, 2, \ldots, 7\}$, $\mathscr{F} = 2^{\Omega}$, and $\mathbb{P} : \mathscr{F} \to [0, 1]$ is defined as

$$\begin{aligned} \mathbb{P}(\{1\}) &= \mathbb{P}(\{2\}) = 0.2, \\ \mathbb{P}(\{3\}) &= \mathbb{P}(\{4\}) = 0.15, \\ \mathbb{P}(\{5\}) &= \mathbb{P}(\{6\}) = \mathbb{P}(\{7\}) = 0.1. \end{aligned}$$

It can be shown that $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space.
Let E be the event that one of the first three horses wins, and F that one of the horses 1, 5, 6, 7 wins: $E = \{1, 2, 3\}$, $F = \{1, 5, 6, 7\}$. Then,

$$\begin{aligned} \mathbb{P}(E) &= \mathbb{P}(\{1\} \cup \{2\} \cup \{3\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) + \mathbb{P}(\{3\}) = 0.55, \\ \mathbb{P}(F) &= \mathbb{P}(\{1\} \cup \{5\} \cup \{6\} \cup \{7\}) = 0.5. \end{aligned}$$

Hence, it is better to wager that the winner is among the first 3 horses. ■

Example 25

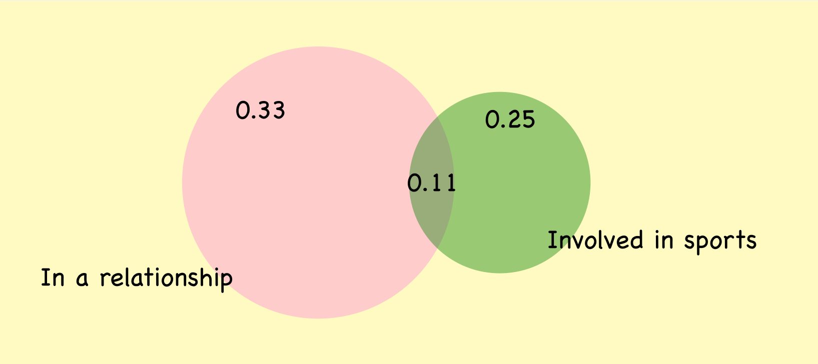
A survey of college students asked the questions: "Are you currently in a relationship?" and "Are you involved in intercollegiate or club sports?" The survey found that 33% were currently in a relationship, and 25% were involved in sports. 11% responded "yes" to both.

What's the probability that a randomly selected student either is in a relationship or is involved in athletics?

Proof.

Let E be the event that a randomly selected college student is in a relationship, and F be the event that he/she is involved in sports. Then,

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) = 0.33 + 0.25 - 0.11 = 0.47. \quad \blacksquare$$



College students

Probability on intervals

Let $\mathscr{U} = (0,1]$, and let $\mathscr{C} = \{(a,b] : 0 < a < b \leqslant 1\}$. Let $\mathscr{B}$ be the minimal σ-field containing $\mathscr{C}$, which is also known as the Borel set. Define the Lebesgue measure m on $\mathscr{C}$ as

$$m((a,b]) = b - a.$$

Let $\mathscr{B}_0$ be the collection of subsets of $\mathscr{U}$ each of which is the union of a countable number of members of $\mathscr{C}$, say, a typical event B in $\mathscr{B}_0$ is of the form

$$B = \bigcup_{j=1}^{\infty} (a_j, b_j] \quad a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n < \ldots$$

has the probability $m(B) = \sum_{j=1}^{\infty} (b_j - a_j)$. Then, $(\mathscr{U}, \mathscr{B}, m)$ is a probability space.

Geometric Probability

Definition 26 (Geometric proabililty)

Geometric Probability involves geometric measures such as length, area, and volume to describe the probability of an event.

Equally likely probability

If all outcomes are equally likely, the geometric probability of an event A is given by:

$$\mathbb{P}(A) = \frac{\text{Geometric measure of } A}{\text{Geometric measure of the sample space } \Omega}$$

Example 27 (Dart game)

Consider a dart game where darts are thrown at a square dartboard with a circle inscribed in it. What is the geometric probability that a dart lands inside the circle?

Solution.

- Assume that the radius is 1.
- The area of the square (sample space) is $2 \times 2 = 4$ square units.
- The area of the inscribed circle (event A) is $\pi \times 1^2 = \pi$ square units.
- Therefore, the geometric probability that a dart lands inside the circle is $\frac{\pi}{4}$.

Example 28

In a random experiment of drawing a point from a line segment of length 1, what is the geometric probability that the drawn point lies in the middle third of the line segment?

Solution.

- The length of the line segment (sample space) is 1 unit.
- The length of the middle third of the line segment (event A) is $\frac{1}{3}$ unit.
- Therefore, the geometric probability that the drawn point lies in the middle third of the line segment is $\frac{1/3}{1} = \frac{1}{3}$.

Examples

Example 29 (Waiting time)

Suppose that a traffic light has a period of 1 minute, where the red light lasts for 30 seconds, the green light lasts for 25 seconds, and the yellow light lasts for 5 seconds. If you arrive at the traffic light at a random time, what is the probability that the light is green?

Solution.

Let $\Omega = (0,1]$, let $\mathscr{F}$ be the Borel set of Ω and $\mathbb{P}$ be the Lebesgue measure. Let E be the event that the light is green, then $E = \{t : 1/2 < t \leqslant 11/12\}$, and thus $\mathbb{P}(E) = 11/12 - 1/2 = 5/12$. ∎

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 30 (Romeo and Juliet)

Romeo and Juliet have a date at a given time, and each will arrive at the meeting place with a delay between 0 and 1 hour, will all pairs of delays being equally likely. The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived. What is the probability that they will meet?

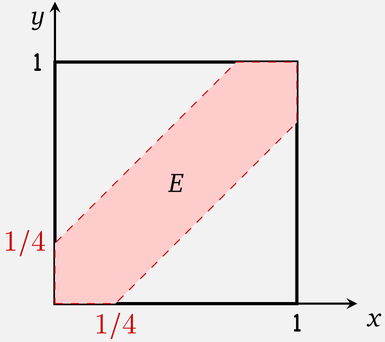
Solution.

Let $\Omega = (0,1]\times(0,1]$, and $\mathscr{F}$ be the Borel σ-field that is generated by $(a_1, b_1] \times (a_2, b_2]$. For any $A \in \mathscr{F}$, let $\mathbb{P}(A)$ be the Lebesgue measure (or, intuitively, the area) of A. Let E be the event that Romeo and Juliet meet, then

$$E = \{(x,y) : |x-y| \leqslant 1/4, 0 \leqslant x, y \leqslant 1\}.$$

Then,

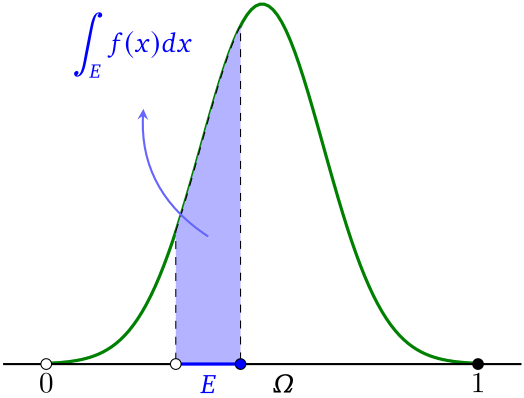
$$\mathbb{P}(E) = \text{area of } E = \frac{7}{16}. \qquad \blacksquare$$



Continuous event with different weights

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Suppose that the sample space is an interval, say $\Omega = (0, 1]$. On this interval we define a weighting function $f(x)$ where $f(x_0)$ specifies the weight for x_0. Because Ω is an interval, events defined on this Ω must also be intervals. For example, for an event $E \subset (0, 1]$, the probability of E is $\mathbb{P}(E) = \int_E f(x)dx$.



SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 31 (Bertrand's paradox)

Consider a circle with radius 1. What is the probability that a randomly chosen chord of the circle is greater than the side of the inscribed equilateral triangle of the circle?

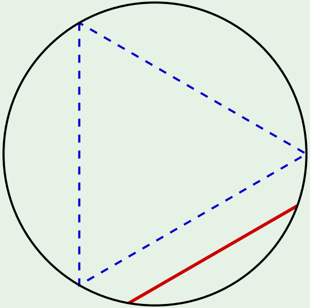


Figure: Joseph Bertrand (1822-1900) and a randomly chosen chord

Solution 1.

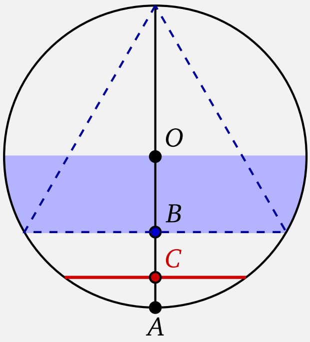
We take a radius of the circle such as OA, and we chose a point C on the radius, with all points being equally likely. We then draw the chord through C that is orthogonal to OA. For elementary geometry, OA intersects the triangle at the midpoint of OA, say B.

Let $\Omega = [0,1]$ be the points on OA, where $x \in \Omega$ represents the point C such that $OC = x$, $\mathscr{F} = \mathscr{B}([0,1])$, and $\mathbb{P}$ be the Lebesgue measure. Let E be the event that the chord is greater than the side of the triangle, then

$$E = \{x : 0 \leqslant x \leqslant \frac{1}{2}\}.$$

Then,

$$\mathbb{P}(E) = \frac{1}{2}. \quad \blacksquare$$



Solution 2.

We take a point on the circle, such as the vertex V, we draw the tangent to the circle through V, and draw a line through V that forms a random angle Θ with the tangent.

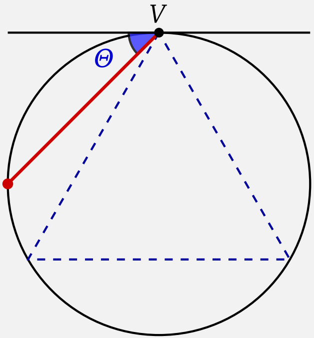
Let $\Omega = (0, \pi]$, where $\alpha \in \Omega$ represents the chord whose angle equals to α. Let $\mathscr{F}$ be the Borel σ-field of Ω, and let

$$\mathbb{P}((\alpha, \beta]) = \frac{\beta - \alpha}{\pi}.$$

Let E be the event that the chord is longer than the side of the inscribed equilateral triangle, then $E = \{\alpha : \frac{\pi}{3} \leqslant \alpha \leqslant \frac{2\pi}{3}\}$. Therefore,

$$\mathbb{P}(E) = \frac{1}{3}.$$

■



What shall we do?

- We need to specify unambiguously a probability model.
- The answer depends on the precise meaning of “randomly chosen”. The two methods lead to contradictory results.
- The model should reflect the real world.
- “Simple” but “incorrect” models, or “Correct” but “non-tractable” models? That is a trade-off.
- Sometimes, a model is chosen on the basic of historical data or past outcomes of similar experiments, using statistical inference methods.

Problems

1. In geometric probability, the sample space is a geometric figure and the probability of an event is defined as the ratio of the areas (or volumes, or lengths, depending on the context) of two regions.

   A. True

   B. False

Problems

2. The probability of an event is always equal to the ratio of the number of favorable outcomes to the total number of outcomes.

   A. True

   B. False

Problems

3. In general probability, not all outcomes have to be equally likely.

   A. True

   B. False

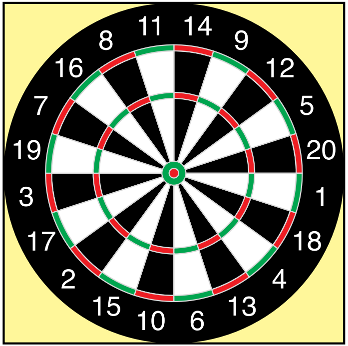
4. In a dart game where darts are thrown at a square dartboard with a circle inscribed in it, what is the geometric probability that a dart lands inside the circle?

A. $\frac{1}{4}$

B. $\frac{\pi}{4}$

C. $\frac{1}{\pi}$

D. $\frac{4}{\pi}$



5. In a random experiment of drawing a point from a line segment of length 1, what is the geometric probability that the drawn point lies in the middle third of the line segment?

A. $\frac{1}{3}$

B. $\frac{1}{2}$

C. $\frac{2}{3}$

D. 1

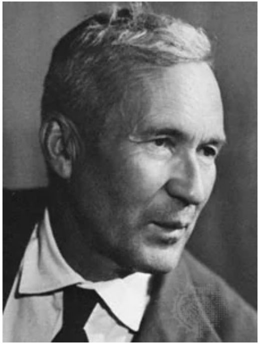
Axioms of Probability

Events and probability: Still some gaps

- Mathematically, probability is a function that assigns probabilities to events.
- We wish to find a good collection of events, on which we can properly define probabilities.

The fundamental theory of modern probability theory was built by the Soviet mathematician Kolmogorov (【前苏联】柯尔莫戈洛夫).

Andrey Kolmogorov (1903–1987). ➡



Collection of events

Let $\mathscr{A}$ be a nonempty collection of events of Ω, and it may have certain "closure properties":

(0) $\Omega \in \mathscr{A}$.

(i) $E \in \mathscr{A} \implies E^c \in \mathscr{A}$.

(ii) $E_1 \in \mathscr{A}, E_2 \in \mathscr{A} \implies E_1 \cup E_2 \in \mathscr{A}$.

(iii) $E_1 \in \mathscr{A}, E_2 \in \mathscr{A} \implies E_1 \cap E_2 \in \mathscr{A}$.

(iv) $\forall\, n \geqslant 2$:
$E_j \in \mathscr{A}, 1 \leqslant j \leqslant n \implies \bigcup_{j=1}^n E_j \in \mathscr{A}$.

(v) $\forall\, n \geqslant 2$:
$E_j \in \mathscr{A}, 1 \leqslant j \leqslant n \implies \bigcap_{j=1}^n E_j \in \mathscr{A}$.

(vi) $E_j \in \mathscr{A}; E_j \subset E_{j+1}, 1 \leqslant j < \infty \implies \bigcup_{j=1}^\infty E_j \in \mathscr{A}$.

(vii) $E_j \in \mathscr{A}; E_j \supset E_{j+1}, 1 \leqslant j < \infty \implies \bigcap_{j=1}^\infty E_j \in \mathscr{A}$.

(viii) $E_j \in \mathscr{A}, 1 \leqslant j < \infty \implies \bigcup_{j=1}^\infty E_j \in \mathscr{A}$.

(ix) $E_j \in \mathscr{A}, 1 \leqslant j < \infty \implies \bigcap_{j=1}^\infty E_j \in \mathscr{A}$.

(x) $E_1, E_2 \in \mathscr{A}, E_1 \subset E_2 \implies E_2 \setminus E_1 \in \mathscr{A}$.

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 32 (Fields and sigma fields)

A collection $\mathscr{F}$ of subsets of Ω is called

(i) a field iff (0), (i) and (ii) hold.

(ii) a σ-field (sigma field, sigma field) iff (0), (i) and (viii) hold.

Remark

(a) The condition (0) ensures that a field is not nonempty.

(b) A field is also called an algebra, and a σ-field is also known as a σ-algebra.

More fields

Definition 33

(i) The collection of all subsets of Ω is a σ-field called the power set, denoted by 2^{Ω}.

(ii) The collection of the two sets $\{\emptyset, \Omega\}$ is a σ-field called the trivial sigma field.

Examples of Sigma-Fields

Example 34 (Trivial Sigma-Field)

Given a sample space $\Omega = \{a, b, c\}$, a trivial sigma-field is $\mathscr{F} = \{\varnothing, \Omega\} = \{\varnothing, \{a, b, c\}\}$. It contains only the empty set and the entire sample space.

Example 35 (Sigma-Field Generated by a Single Event)

Given a sample space $\Omega = \{a, b, c\}$ and an event $A = \{a\}$, the sigma-field generated by A is $\mathscr{F} = \{\varnothing, A, A^c, \Omega\} = \{\varnothing, \{a\}, \{b, c\}, \{a, b, c\}\}$.

Examples of Sigma-Fields (cont.)

Example 36 (Sigma-Field Generated by Two Events)

Given a sample space $S = \{a, b, c\}$, and two events $A = \{a\}$ and $B = \{b\}$, the sigma-field generated by A and B is $\mathscr{F} = \{\emptyset, A, B, A \cup B, A^c, B^c, (A \cup B)^c, S\}$.

Example 37 (The Power Set)

Given a sample space $S = \{a, b, c\}$, the power set of S is a sigma-field, $\mathscr{F} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$.

Sample space and Events, revisited

Actually, not all subsets of Ω can be given a probability. To define the axioms of probability, we first define the events, at an axiom view.

Definition 38

Let Ω is a sample space, and $\mathscr{F}$ is a σ-field of Ω. We say E is an event if $E \in \mathscr{F}$. The σ-field is also called an event space. The pair $(\Omega, \mathscr{F})$ is also called a sample space somewhere else.

Definition 39

We shall assume that, for each event E in the sample space $(\Omega, \mathscr{F})$, there exists a value $\mathbb{P}(E)$, referred to as the probability of E, which satisfies the following axioms:

(i) Non-negativity: $\mathbb{P}(E) \geqslant 0$ for all $E \in \mathscr{F}$.

(ii) Normalization: $\mathbb{P}(\Omega) = 1$.

(iii) σ-additivity: For any sequence of mutually exclusive events $E_1, E_2, \ldots,$

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

The triple $(\Omega, \mathscr{F}, \mathbb{P})$ is said to be a probability space.

Why these three axioms?

(i) Axiom I (Non-negativity) ensures that probability is never negative.

(ii) Axiom II (Normalization) ensures that probability is never greater than 1.

(iii) Axiom III (Additivity) allows us to add probabilities when two events do not overlap.

Properties of probability: I

These 3 axioms imply the following consequences.

Proposition 40

(a) $\mathbb{P}(\emptyset) = 0$.

(b) Finite additivity: For any finite sequence of mutually exclusive events $E_1, \ldots, E_n$,

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathbb{P}(E_i).$$

(c) $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$ for all $E \in \mathscr{F}$.

(d) $\mathbb{P}(E) \leqslant 1$ for all $E \in \mathscr{F}$.

(e) If $E \subset F$, then $\mathbb{P}(E) = \mathbb{P}(F) - P(F \setminus E) \leqslant \mathbb{P}(F)$.

Proof of (a).

Consider a sequence of events $E_1, E_2, \ldots$, where $E_1 = \Omega$, and $E_i = \varnothing$ for $i \geqslant 2$. Then, these events are mutually exclusive and $\Omega = \cup E_i$. Therefore, by Axiom (iii),

$$\mathbb{P}(\Omega) = \sum_{i=1}^{\infty} \mathbb{P}(E_i) = \mathbb{P}(\Omega) + \sum_{i=2}^{\infty} \mathbb{P}(\varnothing),$$

which implies that $\mathbb{P}(\varnothing) = 0$. ■

Proof of (b).

Consider $E_1, E_2, \ldots$ are a sequence of mutually exclusive events with $E_j = \varnothing$ for $j > n$, then

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i) = \sum_{i=1}^{n} \mathbb{P}(E_i) + 0.$$ ■

Proof of (c).

Let $E_1 = E$ and $E_2 = E^c$. By (b) with $n = 2$,

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(E \cup E^c) = \mathbb{P}(E) + P(E^c). \qquad \blacksquare$$

Proof of (d).

Since $\mathbb{P}(E) = 1 - \mathbb{P}(E^c)$ and $\mathbb{P}(E^c) \geqslant 0$, then

$$\mathbb{P}(E) \leqslant 1. \qquad \blacksquare$$

Proof of (e).

Since $F = E \cup (F \setminus E)$, and E and $F \setminus E$ are mutually exclusive, then by (b),

$$\mathbb{P}(F) = \mathbb{P}(E) + \mathbb{P}(F \setminus E) \geqslant \mathbb{P}(E). \qquad \blacksquare$$

Proposition of probability: II

Proposition (Cont'd)

(f) General addition rule: For any $E, F \in \mathscr{F}$,

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F).$$

(g) The inclusion–exclusion identity: For any $E_1, E_2, \ldots, E_n \in \mathscr{F}$ (not necessarily mutually exclusive),

$$\begin{aligned}
&\mathbb{P}(E_1 \cup E_2 \cup \cdots \cup E_n) \\
&\quad = \sum_{i=1}^{n} \mathbb{P}(E_i) - \sum_{1 \leqslant i_1 < i_2 \leqslant n} \mathbb{P}(E_{i_1} \cap E_{i_2}) + \ldots \\
&\qquad + (-1)^{r+1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) + \ldots \\
&\qquad + (-1)^{n+1} \mathbb{P}(E_1 \cap \cdots \cap E_n).
\end{aligned}$$

Proof of (f).

Note that $E \cup F = E \cup (F \setminus E)$, then by (e),

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F \setminus E),$$

similarly, $F = (F \setminus E) \cup (E \cap F)$, by (e) again,

$$\mathbb{P}(F) = \mathbb{P}(E \cap F) + \mathbb{P}(F \setminus E) \implies \mathbb{P}(F \setminus E) = \mathbb{P}(F) - \mathbb{P}(E \cap F),$$

then

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F).$$

■

Proof of (g).

Can be proved by (f) and recursive arguments. ■

Properties of probability: III

Proposition (Cont'd)

(h) Boole's inequality: For any $E_1, E_2, \ldots \in \mathscr{F}$ (not necessarily mutually exclusive),

$$\sum_{i=1}^{\infty} \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i E_j) \leqslant \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \leqslant \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

Monotone property

Increasing events

In the context of probability, increasing events are events that become more likely to occur as additional information is given.

Definition 41

A sequence of events $\{E_n, n \geqslant 1\}$ is said to be an increasing sequence if

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \ldots.$$

Example 42

An example of increasing events can be rolling a fair six-sided die:

- Event E_1: The outcome is less than or equal to 3.
- Event E_2: The outcome is less than or equal to 4.
- Event E_3: The outcome is less than or equal to 5.

Limit of increasing events

Definition 43

If $\{E_n, n \geqslant 1\}$ is an increasing sequence of events, then $\lim_{n\to\infty} E_n$ is defined by

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i.$$

The following proposition is the so called monotone property:

Proposition 44

If $\{E_n, n \geqslant 1\}$ is an increasing sequence of events with $E_\infty = \lim_{n\to\infty} E_n$, then

$$\mathbb{P}\big(E_\infty\big) = \lim_{n\to\infty} \mathbb{P}(E_n).$$

Proof.

Define the events F_n for $n \geqslant 1$ by

$$F_1 = E_1, \quad F_2 = E_2 \setminus E_1, \quad \dots, \quad F_n = E_n \setminus E_{n-1}, \quad \dots.$$

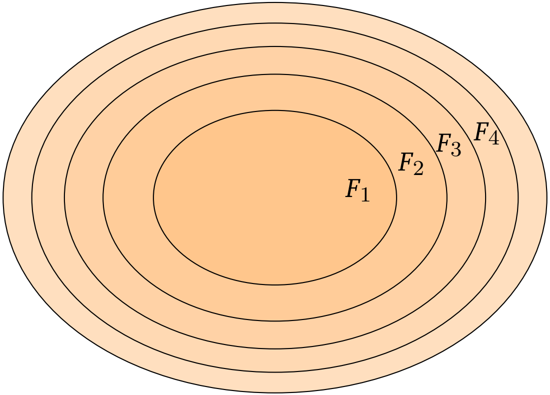
In words, F_n consists of those outcomes in E_n which are not in any of the earlier $E_j, j < n$. It is easy to verify that F_n are mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i = E_n, \quad \text{for all } n \geqslant 1 \text{ and } n = \infty.$$

Then,

$$\begin{aligned} \mathbb{P}(E_\infty) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(F_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(F_i) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} F_i\right) \\ &= \lim_{n\to\infty} \mathbb{P}(E_n). \end{aligned}$$

■


F_1
F_2
F_3
F_4

Decreasing events

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 45

A sequence $\{E_n, n \geqslant 1\}$ is said to be a decreasing seqeunce if $E_1 \supset E_2 \supset \ldots$. Its limit is defined by

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_n.$$

Proposition 46

If $\{E_n, n \geqslant 1\}$ is decreasing with $E_\infty = \lim_{n\to\infty} E_n$, then

$$\mathbb{P}(E_\infty) = \lim_{n\to\infty} \mathbb{P}(E_n).$$

Axioms of continuity

Proposition 47 (Axioms of continuity)

If $E_n \downarrow \varnothing$, then $\mathbb{P}(E_n) \to 0$ as $n \to \infty$.

Remark

This proposition is a special case of the monotone property.

Theorem 48

The axioms of finite additivity and continuity together are equivalent to the axiom of countable additivity.

$$\boxed{\text{Finite Additivity \& Continuity}} \xleftrightarrow{\text{equivalent}} \boxed{\text{Countable Additivity}}$$

Proof.

Step 1. Proof of "Countable additivity" $\Longrightarrow$ "Finite Additivity & Continuity". Proved.

Step 2. Proof of "Finite Additivity & Continuity" $\Longrightarrow$ "Countable additivity". Let $\{E_n, n \geqslant 1\}$ be pairwise disjoint, then $F_n := \cup_{k=n+1}^{\infty} E_k \downarrow \varnothing$. By the "Continuity" property, $\lim_{n\to\infty} \mathbb{P}(F_n) = 0$. If "Finite additivity" is assumed, then

$$1 \geqslant \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) + \mathbb{P}(F_n) = \sum_{i=1}^{n} \mathbb{P}(E_i) + \mathbb{P}(F_n).$$

Let $a_n = \sum_{i=1}^{n} \mathbb{P}(E_i)$. It follows that $a_n \uparrow$ and bounded by 1 (why?), and thus the limit $\lim_{n\to\infty} a_n$ exists. Taking limits on both sides yields

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} \mathbb{P}(F_n) = \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

■

Example: Tossing a Coin

Example 49 (Problem Statement)

Consider an experiment where a fair coin is tossed until the first head appears. Let A_i be the event that the first head appears on or before the i-th toss. As i increases, A_i forms an increasing sequence of events.

Application of Continuity Property

According to the first continuity property, we can say that the probability of getting a head eventually is the limit of the probabilities of A_i as i goes to infinity, i.e.,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i) = \lim_{i\to\infty} (1 - 2^{-i}) = 1.$$

Probability zero sets

- $\varnothing$ has probability 0, but the inverse is not correct:
- Not all of probability zero sets are empty.
- For example, in the probability space $(\mathscr{U}, \mathscr{B}, m)$,

$$m(\{0.5\}) = m\left(\bigcap_{n=1}^{\infty}(0.5-\frac{1}{2n}, 0.5]\right) = \lim_{n\to\infty} m((0.5-\frac{1}{2n}, 0.5]) = 0.$$

- Intuitively, the set $\{0.5\}$ has length 0, and then the probability of $\{0.5\}$ is 0.
- As a result,

$$m([a,b]) = m((a,b)),$$

because $m(\{a\}) = m(\{b\}) = 0$.

A formal definition of probability zero sets

Definition 50

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A set $E \in \mathscr{F}$ is said to have probability zero if for any $\varepsilon > 0$, there exists a countable number of subsets E_n such that $E \subset \cup_{n=1}^{\infty} E_n$, and

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \varepsilon.$$

Example 51 (The rational number set has probability zero)

In the probability space $(\mathscr{U}, \mathscr{B}, m)$, let $E = \mathbb{Q} \cap (0, 1]$ be the collection of all rational number in $\mathscr{U} = (0, 1]$. Then, $\mathbb{P}(E) = 0$.

Almost surely

When we make probabilistic claims without considering the measure zero sets, we say that an event happens almost surely.

Definition 52 (Almost surely)

An event E is said to hold almost surely (a.s.) if $\mathbb{P}(E) = 1$.

Example 53 (Irrational numbers)

In the probability space $(\mathscr{U}, \mathscr{B}, m)$, let E be the event containing all of the irrational numbers. Then

$$\mathbb{P}(E) = 1.$$

Problems

1. If $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ then

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \inf_{i \geqslant 1} \mathbb{P}(A_i)$$

A. True

B. False

Problems

2. If $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, then

$$\mathbb{P}\Big(\bigcap_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} \mathbb{P}(A_i)$$

A. True

B. False

Further reading

[1] Sheldon M. Ross (谢尔登·M. 罗斯).

A first course in probability (概率论基础教程): Chapters 1 and 2.

10th edition (原书第十版), 机械工业出版社

[2] Sheldon M. Ross (谢尔登·M. 罗斯).

Introduction to Probability Models (概率模型导论): Chapter 1.

12th edition (原书第十二版), 人民邮电出版社

[3] Kai-Lai Chung (钟开莱).

A course in probability theory (概率论教程): Chapter 2.

3rd edition (原书第三版), 机械工业出版社

Further reading

[4] Dimitri P. Bertsekas and John N. Tsitsiklis.

Introduction to Probability.

2nd Edition. MIT.

[5] Stanley H. Chan.

Introduction to Probability for Data Science.

Michigan Publishing. (FREE on website)