Chapter 2 Probability

Jingrui SUN

Email: sunjr@sustech.edu.cn

Southern University of Science and Technology

September 8, 2024

Outline I

Outline II

Outline

Part 2.1.1: Random Experiments

1. Experiments: An experiment is any situation in which we observe an outcome.

Two types of experiment:

(1) Deterministic: the observed result is not subject to change.

Example: Measure the length of a straight wire by a ruler.

(2) Random: the outcome is always subject to change.

Example: If a coin is tossed, then the outcome can either be "head" or "tail".

2. Random Experiments:

Random experiments are the experiments for which the outcome cannot be predicted with certainty (so, at least two outcomes).

In our course, we shall only consider the random experiments and hereafter refer to "experiments" or "trials".

3. Examples:

(1) Experiment: A coin is tossed one time.

Outcomes: $\{$Head, Tail$\}$.

(2) Experiment: A coin is tossed three times.

Outcomes:

$$\{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}.$$

In the above, "HHT", say, is a particular outcome.

Part 2.1.1: Random Experiments

(3) Experiment: The number of jobs in a print queue of a mainframe computer.

Outcomes: $\{0, 1, 2, 3, 4, 5, \cdots\}$.

(4) Experiment: The number of telephone calls received at a fixed time in ....

Outcomes: $\{0, 1, 2, 3, 4, 5, 6, \cdots\}$.

Part 2.1.1: Random Experiments

(5) Experiment: the length of time between successive earthquakes.

Outcomes: $\{t : t \geq 0\}$.

(6) Experiment: the maximum temperature of a particular (coming) day.

Outcomes: might be $\{x : -10.5 \leq x \leq 30\}$.

More convenient: $\{x : -\infty \leq x \leq \infty\}$.

Examples (1) and (2): Finitely many outcomes;
Examples (3) and (4): A sequence;
Examples (5) and (6): An interval, continuous variable.

Part 2.1.2: Sample Space

1. Definition: For any random experiment, we define the sample space to be the set of all the possible outcomes of the experiment.

2. Notation: The sample space is usually denoted by Ω and a generic element of Ω is denoted by ω.

Part 2.1.2: Sample Space

3. Examples:

In Example (1): $\Omega = \{H, T\}$;

In Example (2):

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\};$$

In Example (3) and (4): $\Omega = \{0, 1, 2, 3, 4, \cdots\}$;

In Example (5): $\Omega = \{t : t \geq 0\}$;

In Example (6): $\Omega = \mathbb{R} = \{x| -\infty \leq x \leq \infty\}$;

Note: Sample space depends upon the experiment.

Part 2.1.3: Events

Recall that the sample space Ω is the set of all the possible outcomes.

1. Definition: An event of the sample space Ω is a (any) subset of Ω.

(However, this is NOT a strict definition, only for convenience at the current stage).

So, any subset of Ω is an event.

2. Example:

In the above Example (2),

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Now, consider: <u>at least two "heads" appear</u>, then it contains the following elements of Ω:

$$\{HHH, HHT, HTH, THH\},$$

this is a subset of Ω and thus <u>an event</u>. We may denote it by, say

$$A = \{HHH, HHT, HTH, THH\},$$

then we say A is an event.

Part 2.1.3: Events

3. Notes:

(1) Since any subset of a set is itself a set, so any event is a set. We may thus use the notations and results in Set Theory, for example, operations.

(2) We usually use capital letters to denote events, for example, A, B, etc.

(3) "Terminology": Suppose A is an event of Ω and we perform the random experiment and if the outcome is in A, we say "A has occurred"; Otherwise "A has not occurred".

(4) Important, "events" depend upon the sample space Ω.

4. Some special events:

(1) Impossible event: $\emptyset$.

Empty set $\emptyset$ is a subset of Ω and so $\emptyset$ is an event.

Here "empty" means "contains no element in the given sample space", i.e. "it can not occur".

For example, in the above Example (2), if we consider "four heads appear", then it is impossible and thus an impossible event.

(2) Certain event (or "Sure event"): Ω.

Note that Ω itself can be viewed as a subset of Ω and so Ω is also an event.

Since Ω contains all the possible outcomes and thus if we perform the experiment, the outcome must be in Ω and thus "Ω always occurs" and so "certain event".

Part 2.1.3: Events

(3) Elementary event:

An elementary event of the sample space is a singleton of Ω corresponding to a particular outcome of the experiment.

For example, in the above Example (2),

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

"$\{THH\}$", for example, is the set of a particular outcome and thus is an elementary event.

Note: We can see, "impossible event $\emptyset$", "Certain event Ω" and "elementary event" all correspond to the given sample space.

Part 2.1.4: Operations of Events

1. Union:

(1) Definition: Suppose A and B are two events of Ω, the union of A and B is the event that either A occurs or B occurs or both occur, denoted by $A \cup B$.

(It is enough just to say either A or B)

(2) Example: In the above Example (2),

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Suppose

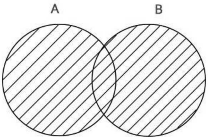
$$A = \{HHH, HHT, HTT, HTH\},$$
$$B = \{HHH, HTH, TTH, THH\},$$

then the union of A and B is the event C, where

$$C = A \cup B = \{HHH, HHT, HTT, HTH, TTH, THH\}.$$

Note that we do not write HHH and HTH two times.

(3) Venn diagram:



(4) Notation:

$$\begin{aligned} A \cup B &= \{\ \omega \in \Omega; \omega \in A \ \underline{\text{or}}\ \omega \in B\ \} \\ &= \{\ \text{either } A \ \underline{\text{or}}\ B \text{ occurs}\ \} \\ &= \{\ \text{either } A \text{ occurs or } B \text{ occurs or } \underline{\text{both}}\ \} \end{aligned}$$

(5) Basic laws:

(i) Commutative law: $A \cup B = B \cup A$;

(ii) Associative law: $(A \cup B) \cup C = A \cup (B \cup C)$;

(iii) $A \cup \emptyset = A$;

(iv) $A \cup \Omega = \Omega$.

All these laws are easily verified.

Part 2.1.4: Operations of Events

2. Intersection:

(1) Definition: Suppose A and B are two events of Ω, then the intersection of A and B is the event that both A and B occur, denoted by $A \cap B$.

$A \cap B$ consists of those outcomes that are <u>common</u> to both A and B).

Pay attention to the difference between the "Union" and "Intersection".

Part 2.1.4: Operations of Events

(2) Example: Suppose

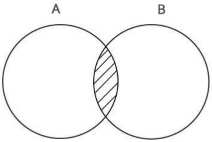
$$A = \{HHH, HHT, HTT, HTH\},$$
$$B = \{HHH, HTH, TTH, THH\},$$

then the intersection of A and B, is the event

$$C = A \cap B = \{HHH, HTH\}.$$

Part 2.1.4: Operations of Events

(3) Venn diagram:



(4) Notation:

$$A \cap B = \{\ \omega \in \Omega; \omega \in A \ \underline{\text{and}}\ \omega \in B\ \}$$
$$= \{\ \text{both } A \ \underline{\text{and}}\ B \text{ occur}\ \}.$$

Part 2.1.4: Operations of Events

(5) Basic laws:

(i) Commutative law: $A \cap B = B \cap A$;

(ii) Associative law: $(A \cap B) \cap C = A \cap (B \cap C)$;

(iii) $A \cap \emptyset = \emptyset$;

(iv) $A \cap \Omega = A$.

Part 2.1.4: Operations of Events

(6) Further:

(v) Distributive laws: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$;

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

The two distributive laws need to be proved, but easy.

3. Complement:

(1) Definition: Suppose A is an event of Ω, then the complement of A is the event that A does not occur and thus consists of all those elements in the sample space that are not in A.

The complement of A is denoted by A^c

(2) Examples: Suppose

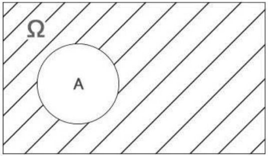
$$\begin{aligned}\Omega &= \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},\\ A &= \{HHH, HHT, HTT, HTH\},\\ B &= \{HHH, HTH, TTH, THH\},\end{aligned}$$

then

$$\begin{aligned}A^c &= \{THH, THT, TTH, TTT\},\\ B^c &= \{HHT, HTT, THT, TTT\}.\end{aligned}$$

Part 2.1.4: Operations of Events

(3) Venn diagram:



(4) Notation:

$$
\begin{aligned}
A^c &= \{\omega \in \Omega, \omega \notin A\} \\
&= \{A \text{ does not occur}\}.
\end{aligned}
$$

(5) Basic laws:

By the above example, we notice that

$$A \cup B = \{HHH, HHT, HTT, HTH, TTH, THH\},$$
$$A^c = \{THH, THT, TTH, TTT\},$$
$$B^c = \{HHT, HTT, THT, TTT\}.$$

Thus

$$(A \cup B)^c = \{THT, TTT\}, \quad A^c \cap B^c = \{THT, TTT\}.$$

So,

$$(A \cup B)^c = A^c \cap B^c.$$

Part 2.1.4: Operations of Events

(5) Basic laws:

Similarly, $A \cap B = \{HHH, HTH\}$,

$$\Rightarrow (A \cap B)^c = \{HHT, HTT, THH, THT, TTH, TTT\},$$

and

$$A^c \cup B^c = \{HHT, HTT, THH, THT, TTH, TTT\},$$

So

$$(A \cap B)^c = A^c \cup B^c.$$

Part 2.1.4: Operations of Events

(5) Basic laws:

These are called the "De-Morgan laws", i.e. for any two events A and B, we have

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

Check it using the Venn diagram!!

(5) Basic laws:

Also easy to see: $\emptyset^c = \Omega$, $\Omega^c = \emptyset$.

"De-Morgan laws" need to be strictly proved, but the proof is easy. Try yourself.

4. Remarks:

Similarly, we can define the union and intersection for finitely many or even a sequence of events. The meaning should be clear.

We usually use "$\cup_{i=1}^{n} A_i$", "$\cap_{i=1}^{n} A_i$", "$\cup_{i=1}^{\infty} A_i$", "$\cap_{i=1}^{\infty} A_i$" to denote these operations. Still, De-Morgan laws apply.

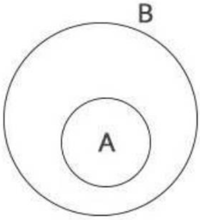
Again, the proof is easy.

Part 2.1.5: Relations among events

1. "Containing":

(1) Suppose A and B are two events, if A is a subset of B, i.e. each element of A is also an element of B, then we say B contains A, or A is included in B.

(2) Venn diagram:



(3) Notation: $A \subset B$ or $B \supset A$.

In other words $A \subset B$ means:

if A occurs, then "B must occur", or "A occurs implies B occurs".

(4) Example: In the above Example 2, let

$$E = \{HHH, HTH\}, G = \{HHH, HHT, HTH, TTT\},$$

then $E \subset G$, which means that

if "E occurs", then "G occurs.

Note that for any event A, we have

$$\emptyset \subset A \subset \Omega.$$

Part 2.1.5: Relations among events

2. "Disjoint":

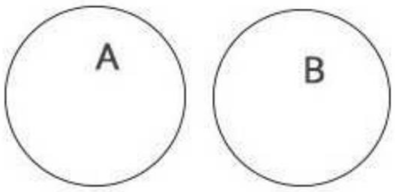
(1) Two events A and B are said to be disjoint if A and B have no outcomes in common.

In other words, the intersection of A and B contains no element (i.e. impossible event),

or "A and B are disjoint" means "$A \cap B = \emptyset$".

Part 2.1.5: Relations among events

(2) Venn diagram:



$$A \cap B = \emptyset.$$

Part 2.1.5: Relations among events

(3) Example: In the above Example 2, if

$$A = \{HHH, HTH\}, \quad B = \{TTT, THH, TTH\},$$

then $A \cap B = \emptyset$, so A and B are disjoint.

(4) Meaning: If A and B are disjoint, then if A occurs, then B can not occur, i.e. A "occurs" implies "B does not occur".

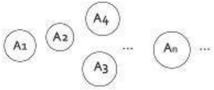
(Also, of course "B occurs" implies "A does not occur".)

(5) Remark: Similarly, we can define several events that are disjoint events. Also, the meaning of "a sequence of events are disjoint" should be clear.

Formally, a sequence of events $A_1, A_2, \cdots$ are called mutually disjoint if any two of them are disjoint.

Part 2.1.5: Relations among events

(5) Remark: In Venn diagram:



In other words, a sequence of events

$$A_1, A_2, \cdots, A_n, \cdots$$

are called mutually disjoint , if for any $i \neq j, A_i \cap A_j = \emptyset$.

(Similarly, for finitely many of events $A_1, A_2, \cdots, A_n$.)

Outline

Part 2.2.1: Definition

1. Definition: For a given sample space Ω, the probability measure (or simply, "probability") is a function $\mathrm{P}(\cdot)$ from the events to $\mathbb{R}$ that satisfies the following axioms:

(i) $\mathrm{P}(\Omega) = 1$;

(ii) for any event A, $\mathrm{P}(A) \geq 0$;

(iii) if A and B are disjoint, then

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B). \tag{2.2.1}$$

(iv) if $A_1, A_2, \cdots, A_n, \cdots$ are mutually disjoint, then

$$\mathrm{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathrm{P}(A_n). \tag{2.2.2}$$

Also (2.2.2) is true for finitely many of disjoint events, and (2.2.2) is usually called "σ-addictive" property.

2. Explanations:

(1) Essentially, probability (or more exactly, probability measure) is a function.

But which function?

The meaning is: for any event A, a real value, denoted by $\mathrm{P}(A)$, is assigned. Thus, for this function,

$$\left.\begin{aligned} &\textit{Domain} : \textit{events} \\ &\textit{Range} : \textit{real values } (\mathbb{R}) \end{aligned}\right\} \textit{set function}$$

Hence, probability P: events $\to \mathbb{R}$.

Part 2.2.1: Definition

(2) For any event A, $\mathrm{P}(A)$ is not an event, it is a real value (and actually non-negative). $\mathrm{P}(A)$ represents the possibility of the occurrence of A (chance of A).

So, A is an event (not a number usually), but $\mathrm{P}(A)$ is a real number (not an event in general).

Part 2.2.1: Definition

(3) The probability, i.e. the set function $\mathrm{P}(\cdot)$: events $\rightarrow \mathbb{R}$ must satisfy conditions (i)-(iv).

They are axioms (!!!)

But, of course, reasonable in the meaning of "agreement with the intuition".

Condition (i): Certain event Ω consists of all possible outcomes and thus must occur, hence, the probability is 100%, i.e. 1;

(3) Condition (ii): "Possibility" must be non-negative;

Condition (iii): If two events A and B are disjoint, then the "possibility" of "either A or B occurs" equals to the sum of the possibilities of A and B.

Also, it must be true even for a sequence of disjoint events.

(4) meaning of (2.2.2):

In the left hand side: $\cup_{n=1}^{\infty} A_n$ is also an event and so has a probability (real number).

In the right hand side: a series.

(5) “Probability measure” or “Probability” refers to three objects: sample space, events and the set function $\mathrm{P}(\cdot)$.

Part 2.2.2: Properties of Probability

Property 1: For any event A,

$$\mathrm{P}(A^c) = 1 - \mathrm{P}(A). \tag{2.2.3}$$

Property 2: $\mathrm{P}(\emptyset) = 0$ (The probability of impossible event is zero).

Property 3: If $A \subset B$, then $\mathrm{P}(A) \leq \mathrm{P}(B)$.

Property 4: If A and B are any two events, then

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B). \tag{2.2.4}$$

Property 5: For any event A, we have $\mathrm{P}(A) \leq 1$.

Property 6: If $\{A_n, n \geq 1\}$ is an increasing sequence of events, i.e.,

$$A_1 \subset A_2 \subset A_3 \subset \cdots \subset A_n \subset A_{n+1} \subset \cdots ,$$

then

$$\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right). \tag{2.2.5}$$

Proof of Property 6:

Step 1: Try to construct a sequence of disjoint events. Define

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad \cdots,$$

in general,

$$B_n = A_n \setminus A_{n-1}, \quad \forall n \geq 2. \tag{2.2.6}$$

Then $\{B_n, n \geq 1\}$ are disjoint, and

$$A_1 = B_1, \quad A_2 = B_1 \cup B_2, \quad \cdots$$

in general, $\forall k \geq 1$,

$$A_k = \bigcup_{m=1}^{k} B_m. \qquad (Easy!!!) \tag{2.2.7}$$

Step 2: We show that

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n. \tag{2.2.8}$$

(a) First, for $\forall n \geq 1, B_n \subset A_n$ ($\because B_n = A_n \setminus A_{n-1}$), and thus

$$\bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} A_n. \tag{2.2.9}$$

(b) In order to get (2.2.8), we only need to prove

$$\bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n=1}^{\infty} B_n. \tag{2.2.10}$$

Suppose that $x \in \bigcup_{n=1}^{\infty} A_n$ (Try to show $x \in \bigcup_{n=1}^{\infty} B_n$), then

$$\exists k \geq 1 \quad \text{such that} \quad x \in A_k.$$

But for this fixed k, we have

$$x \in A_k = \bigcup_{n=1}^{k} B_n \subset \bigcup_{n=1}^{\infty} B_n \quad \Longrightarrow \quad x \in \bigcup_{n=1}^{\infty} B_n.$$

Step 3: Prove the conclusion.

Since $\{B_n, n \geq 1\}$ are disjoint, so

$$P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n).$$

But by (2.2.8) we have $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$, and thus

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(B_n) = \lim_{k\to\infty} \sum_{n=1}^{k} P(B_n). \quad (Obviously)$$

Because $\{B_n\}$ are disjoint, we further have

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{k\to\infty} \sum_{n=1}^{k} P(B_n) = \lim_{k\to\infty} P\left(\bigcup_{n=1}^{k} B_n\right).$$

However, $\bigcup_{n=1}^{k} B_n = A_k$ (see (2.2.7)), hence

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{k\to\infty} P(A_k) = \lim_{n\to\infty} P(A_n).$$

The conclusion is proved.

Property 7: If $\{A_n, n \geq 1\}$ is a decreasing sequence of events, i.e.,

$$A_1 \supset A_2 \supset A_3 \supset \cdots \supset A_n \supset A_{n+1} \supset \cdots ,$$

then

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right). \tag{2.2.11}$$

Proof: A_n decreasing $\Rightarrow$ A_n^c increasing. Hence by Property 6,

$$\lim_{n\to\infty} P(A_n^c) = P\left(\bigcup_{n=1}^{\infty} A_n^c\right). \tag{2.2.12}$$

Part 2.2.3: Further Properties of Probability

But for any event B, we have $P(B^c) = 1 - P(B)$ and thus

$$P(A_n^c) = 1 - P(A_n), \quad \forall n \geq 1, \tag{2.2.13}$$

and

$$P\left(\bigcup_{n=1}^{\infty} A_n^c\right) = 1 - P\left(\left(\bigcup_{n=1}^{\infty} A_n^c\right)^c\right). \tag{2.2.14}$$

However, by De-Morgan's law,

$$\left(\bigcup_{n=1}^{\infty} A_n^c\right)^c = \bigcap_{n=1}^{\infty} (A_n^c)^c = \bigcap_{n=1}^{\infty} A_n. \tag{2.2.15}$$

Part 2.2.3: Further Properties of Probability

Substituting (2.2.15) into (2.2.14) yields

$$P\left(\bigcup_{n=1}^{\infty} A_n^c\right) = 1 - P\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Hence (2.2.12) reads

$$\lim_{n\to\infty} P(A_n^c) = 1 - P\left(\bigcap_{n=1}^{\infty} A_n\right).$$

But by (2.2.13) we have

$$\lim_{n\to\infty} [1 - P(A_n)] = 1 - P\left(\bigcap_{n=1}^{\infty} A_n\right),$$

that is,

$$1 - \lim_{n\to\infty} P(A_n) = 1 - P\left(\bigcap_{n=1}^{\infty} A_n\right),$$

or equivalently,

$$\lim_{n\to\infty} P(A_n) = P(\bigcap_{n=1}^{\infty} A_n). \tag{2.2.16}$$

The above (2.2.16) is just what we want to prove, i.e. (2.2.11).

Part 2.2.3: Further Properties of Probability

Remark: We define the following:

- if $A_n \uparrow$ (i.e. increasing), then

$$\lim_{n\to\infty} A_n \triangleq \bigcup_{n=1}^{\infty} A_n;$$

- if $A_n \downarrow$ (i.e. decreasing), then

$$\lim_{n\to\infty} A_n \triangleq \bigcap_{n=1}^{\infty} A_n.$$

Then Properties 6 and 7 can be stated as following:

If $\{A_n, n \geq 1\}$ is a monotone sequence of events, then

$$\lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right).$$

Part 2.2.4: An equivalent definition of Probability

1. **Another definition:** For a given sample space Ω, the probability measure is a function $P(\cdot)$ from the events to the real numbers that satisfies the following axioms:

(a) For any event A, $P(A) \geq 0$;

(b) $P(\emptyset) = 0$, where $\emptyset$ stands for the impossible event;

(c) If $\{A_1, A_2, \cdots, A_n, \cdots\}$ is a sequence of mutually disjoint events, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{2.2.17}$$

(d) $P(\Omega) = 1$.

2. **Conclusion**: The two definitions for the probability measure given in Sections 2.2.1 and 2.2.4 are **equivalent**.

Proof: Definition (2.2.1) $\Rightarrow$ Definition (2.2.4).
We only need to show that $P(\emptyset) = 0$.
But this has been shown above. See Property 2.
Definition (2.2.4) $\Rightarrow$ Definition (2.2.1).
We only need to show that for **finitely many disjoint events** $\{A_1, A_2, \cdots, A_n\}$, where $n \geq 2$, we have

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k). \tag{2.2.18}$$

But this is easy.

Indeed, for the given finitely many events $\{A_1, A_2, \cdots, A_n\}$, we add infinity many impossible events to get $\{A_1, A_2, \cdots, A_n, A_{n+1}A_{n+2}, \cdots\}$, where $A_k = \emptyset$ for all $k \geq n+1$.

Thus this is a sequence of disjoint events and thus by (2.2.17), we have

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{n} P(A_k) + \sum_{k=n+1}^{\infty} P(A_k). \qquad (2.2.19)$$

Part 2.2.4: An Equivalent Definition of Probability

But

$$\bigcup_{k=1}^{\infty} A_k = \left(\bigcup_{k=1}^{n} A_k\right) \cup \left(\bigcup_{k=n+1}^{\infty} A_k\right) = \left(\bigcup_{k=1}^{n} A_k\right) \cup \left(\bigcup_{k=n+1}^{\infty} \emptyset\right) = \bigcup_{k=1}^{n} A_k,$$

and thus (2.2.19) reads

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{n} P(A_k) + \sum_{k=n+1}^{\infty} P(\emptyset).$$

However, by (b) we know $P(\emptyset) = 0$ and thus (2.2.18) is proved.

Part 2.2.4: An Equivalent Definition of Probability

3. Remark

In more advanced courses, we usually use the definition 2.2.4.

Also, if only the first three, i.e. (a), (b), (c) are required, then this function is called a **measure** (but in this case, we use other terms to replace "events" and "probability").

If, furthermore, (d) is also required, then this measure is called a probability measure. For details, see later.

Part 2.2.5: Summary

Probability is a set function defined for all events that satisfies the following properties:

1. Non-negative and bounded, i.e. for any event A, $0 \leq P(A) \leq 1$.
2. Monotone, i.e. $A \subset B \Rightarrow P(A) \leq P(B)$.
3. Additive, i.e. If $A_1, A_2, \cdots, A_n, \cdots$ are disjoint (finitely many or a sequence of them), then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Part 2.2.5: Summary

4. $P(\emptyset) = 0$, $P(\Omega) = 1$.
5. $P(A^c) = 1 - P(A)$.
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
7. If $\{A_n, n \geq 1\}$ is an increasing sequence of events, then

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

8. If $\{A_n, n \geq 1\}$ is a decreasing sequence of events, then

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Part 2.2.6: Measure and Probability

1. **Definition of σ-algebra:** Let Ω be an arbitrary non-empty set. A collection $\mathcal{F}$ of subsets of Ω is called a σ-algebra on Ω if this collection $\mathcal{F}$ satisfies the following conditions:

(a) $\Omega \in \mathcal{F}$.

(b) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$. ($\mathcal{F}$ is closed under complement)

(c) $\forall i = 1, 2, \cdots, A_i \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$. ($\mathcal{F}$ is closed under countable union)

2. **Properties of σ-algebra:** Let Ω be an arbitrary non-empty set and $\mathcal{F}$ is a σ-algebra on Ω. Then

(a) $\emptyset \in \mathcal{F}$.

(b) $\forall i = 1, 2, \cdots, A_i \in \mathcal{F} \quad \Rightarrow \quad \cap_{i=1}^{\infty} A_i \in \mathcal{F}$.

(c) $\forall i = 1, 2, \cdots, n, A_i \in \mathcal{F} \quad \Rightarrow \quad \cap_{i=1}^{n} A_i \in \mathcal{F}$ and $\cup_{i=1}^{n} A_i \in \mathcal{F}$.

(d) If $A \in \mathcal{F}, B \in \mathcal{F}$, then $A \setminus B \in \mathcal{F}$.

Part 2.2.6: Measure and Probability

In short, a σ-algebra on Ω is closed under the operations of finite union, finite intersection, countable union, countable intersection, complement, difference and symmetric difference.

(The symmetric difference between the two sets A and B is defined by $A \triangle B \equiv (A \setminus B) \cup (B \setminus A)$.)

However, usually, σ-algebra is not closed under the set operation of uncountable union and uncountable intersection.

3. **Measurable space**

Let Ω be an arbitrary non-empty set and $\mathcal{F}$ is a σ-algebra on Ω. Then the pair $(\Omega, \mathcal{F})$ is called a measurable space.

4. **Definition of Measure**

Let $(\Omega, \mathcal{F})$ be a measurable space. A set function μ, whose domain of definition is the σ-algebra $\mathcal{F}$, is called a measure on $(\Omega, \mathcal{F})$ if

(a) For any $B \in \mathcal{F}$, $\mu(B) \geq 0$.

(b) $\mu(\emptyset) = 0$.

(c) For each infinite sequence $\{A_i\}$ of disjoint sets that belong to $\mathcal{F}$, we have $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

Part 2.2.6: Measure and Probability

A measure is called finite if $\mu(\Omega) < \infty$.

Furthermore, if the measure μ satisfies the condition $\mu(\Omega) = 1$, then this measure is called a probability measure, or more simply, is called a probability.

For probability measure, we usually use P to denote it. For probability measure, the set Ω is usually called a sample space and the σ-algebra $\mathcal{F}$ on Ω is called the set of events.

5. Definition of Probability Measures

Definition: Let Ω be a sample space, and $\mathcal{F}$ denotes the set of events. Then $(\Omega, \mathcal{F})$ is called a **measurable space**. Let $(\Omega, \mathcal{F})$ be a measurable space, a set function P on $\mathcal{F}$ is called a **probability measure**, if

(a) For any $B \in \mathcal{F}$, $P(B) \geq 0$;

(b) $P(\emptyset) = 0$;

(c) $P(\Omega) = 1$;

(d) for each infinite sequence $\{A_i\}$ of disjoint sets that belong to $\mathcal{F}$, we have $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Then $(\Omega, \mathcal{F}, P)$ is called a **probability space**, or a probability triple.

Part 2.2.6: Measure and Probability

6. Terminology

$(\Omega, \mathcal{F}, P)$	probability space or probability triple
Ω	sample space
$\omega \in \Omega$	sample point
$\mathcal{F}$	σ-filed, the family of events

7. Properties of Probability (including the ones in the definition)

Three Groups: $(\Omega, \mathcal{F}, P)$, a probability space

① Group A: Inequality

$$\begin{aligned}
&0 \leq P(A) \leq 1, && \forall A \in \mathcal{F} \\
&P(A) \leq P(B), && \forall A \in \mathcal{F}, B \in \mathcal{F}, A \subset B \\
&P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n), && \forall A_n \in \mathcal{F} \\
&P\left(\bigcup_{n=1}^{m} A_n\right) \leq \sum_{n=1}^{m} P(A_n), && A_1, \cdots, A_m \in \mathcal{F}.
\end{aligned}$$

② Group B: Equality. Let $A, B, A_i \in \mathcal{F}$.

$$
\begin{aligned}
&P(\emptyset) = 0, \qquad P(\Omega) = 1, \\
&P(B \backslash A) = P(B) - P(A), \quad \text{if } A \subset B, \\
&P(A^c) = 1 - P(A), \\
&P(A \cup B) = P(A) + P(B) - P(A \cap B), \\
&P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j\leq n} P(A_i \cap A_j) + \cdots \\
&\qquad\qquad\qquad + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n).
\end{aligned}
$$

Part 2.2.6: Measure and Probability

② Group B: Equality

$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$, for disjoint sequence $\{A_n\}$ in $\mathcal{F}$;

$P(\bigcup_{n=1}^{m} A_n) = \sum_{n=1}^{m} P(A_n)$, for disjoint $\{A_n\}$ in $\mathcal{F}$.

③ Group C: Limiting property

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \text{ for increasing } \{A_n\} \in \mathcal{F};$$
$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \text{ for decreasing } \{A_n\} \in \mathcal{F}.$$

Outline

Part 2.3.1: Equally Liked Outcomes (Special case)

1. Definition:
If the sample space has only a finite number of outcomes and each particular outcome (elementary event) has the same probability, then it is called the equally likely outcome case.

2. Example:
A fair coin is thrown twice.

$$\Omega = \{HH, HT, TH, TT\}.$$

Question: Event A: exactly one head appears. $P(A) =?$
Answer: Reasonably, $P(A) = \frac{2}{4} = \frac{1}{2}$.

Part 2.3.1: Equally Liked Outcomes (Special case)

2. Example:

Since: Equally likely outcomes, so

$$P(A) = \frac{\text{The number of ways } A \text{ can occur}}{\text{The total number of outcomes}}.$$

3. **Conclusion**: Suppose the sample space has n elements, $\{e_1, e_2, \dots, e_n\}$ say, and suppose that each elementary event $\{e_i\}$ has the same probability $\frac{1}{n}$.

Then the probability of any event is the number of ways this event can occur over the total number n, i.e. for any event A,

$$P(A) = \frac{\text{The number of ways } A \text{ can occur}}{\text{The total number of outcomes}}.$$

Part 2.3.1: Equally Liked Outcomes (Special case)

4. Note:

(1) "Equally likely outcomes" is a special case.

For example, for unfair coin, the above formula is not true.

(2) We need the method to calculate the number of outcomes.

Part 2.3.2: Permutations and Combinations

1. Problems:

(1) Suppose that from 5 children, 3 are to be chosen and lined up. How many different lines are possible?

(2) Suppose that from 5 children, 3 are to be chosen to form a team. How many ways can this be done?

2. Idea:

(1) Difference between Problems 1 and 2?

Problem 1: Ordered!!

Problem 2: Unordered!!

(2) For Problem 1:

The first position: 5 different ways;

The second position: 4 different ways;

The third position: 3 different ways;

Altogether: $5 \times 4 \times 3$ different ways.

For Problem 2, we consider as follows:

First, assume ordered (line up!!), then $5 \times 4 \times 3$ different ways!!

Secondly, but actually “ordered is no use.

(3) For Problem 1:

Then we fix three children, then it is only one team!

However, this team can line up for $3 \times 2 \times 1$ different ways (Think why here!!)

In other words,

(Number of teams)$\times(3 \times 2 \times 1) =$ Number of ways to line up.

(4) For Problem 2: So

$$\begin{aligned}\text{The number of teams} &= \frac{5\times 4\times 3}{3\times 2\times 1} = \frac{5\times 4\times 3}{3!}\\ &= \frac{5\times 4\times 3\times 2\times 1}{3!\times 2\times 1}\\ &= \frac{5!}{3!\times 2!}.\end{aligned}$$

Part 2.3.2: Permutations and Combinations

3. **Conclusions:** Using the above idea, we can easily get the following conclusions.

(1) **Proposition 1:**

For a set of size n and a sample of size r, there are

$$n(n-1)(n-2)\cdots(n-r+1)$$

different ordered samples.

(2) **Proposition 2:**

The number of unordered samples of r objects from n objects (where $r \leq n$) is $\frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}$.

The second conclusion is mostly often used!

The first (special case): Permutation;

The second: Combination.

Part 2.3.2: Permutations and Combinations

Note that, the number of $\frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}$ can be written as

$$\frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n(n-1)\cdots(n-r+1)(n-r)\cdots 1}{r!(n-r)\cdots 1}$$
$$= \frac{n!}{r!(n-r)!}.$$

4. Notation and Terminology:

(1) we define $\binom{n}{r}$, for $r \leq n$ by

$$\binom{n}{r} = \frac{n!}{(n-r)!r!},$$

and say that $\binom{n}{r}$ represents the number of possible combinations of n objects taken r at a time.

$\star$ Thus $\binom{n}{r}$ represents the number of different groups of size r that could be selected from a set of n objects when the order of selection is not considered relevant. $\star$

(2) Other notation: $\binom{n}{r}$, $\underline{\underline{C_n^r}}$, ${}^nC_r$.

For example $\binom{5}{3}$, $\underline{\underline{C_5^3}}$, ${}^5C_3$.

(3) Notes:

By convention, 0! is defined to be 1, thus $\binom{n}{0} = \binom{n}{n} = 1$.

Also, in calculation, we usually use the original one, i.e.

$$\binom{n}{r} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{1\cdot 2\cdots r}.$$

5. Important Application: The Binomial Theorem

(1) Question: $(a+b)^n =?$ (n is a positive integer)

(2) Idea: $(1+x)^n =?$

It must be

$$\begin{aligned}(1+x)^n &= \overbrace{(1+x)(1+x)\cdots(1+x)} \\ &= 1 + b_1 x + b_2 x^2 + \cdots + b_r x^r + \cdots + x^n.\end{aligned}$$

$b_r =?$

Easy to see: $b_r = \binom{n}{r}$ (Think why here),

$$\Rightarrow (1+x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r.$$

Now

$$\begin{aligned}(a+b)^n &= b^n \left(1+\frac{a}{b}\right)^n \\ &= b^n \cdot \sum_{r=0}^{n} \binom{n}{r} \left(\frac{a}{b}\right)^r \quad \left(\text{let } x = \frac{a}{b}!!\right) \\ &= \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}.\end{aligned}$$

(3) **Conclusion:**

The Binomial Theorem: For any positive integer n, we have

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r} \equiv \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r$$
$$= a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \cdots + b^n.$$

The values $\binom{n}{r}$ are often referred to as <u>binomial coefficients</u>.

(4) Simple properties: (Easily proved by definition)

(i) $\binom{n}{0} = \binom{n}{n} = 1;$

(ii) $\binom{n}{r} = \binom{n}{n-r};$

(iii) $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$, where $1 \le r \le n$.

(5) Keep in mind:

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1 \\
 & \cdots & \cdots & & \cdots & & \cdots & \cdots &
\end{array}
$$

The $(n+1)$st row is just

$$\binom{n}{0} \quad \binom{n}{1} \quad \binom{n}{2} \quad \cdots \quad \binom{n}{n}.$$

Outline

Part 2.4.1: Definition

1. Motivation: An example.

Total: 135 patients.

High blood concentration (Positive test);

Low blood concentration (Negative test);

Toxicity (disease present);

No toxicity (disease absent).

	Disease Present	Disease Absent	Total
Positive test	25	14	39
Negative test	18	78	96
Total	43	92	135

	Disease Present	Disease Absent	Total
Positive test	25	14	39
Negative test	18	78	96
Total	43	92	135

Now, choose a patient "at random" (meaning:equally likely!!) from the 135 patients.

Event A: disease present, then

$$P(A) = ?$$

Easy! $P(A) = \frac{\#(\text{ Disease Present })}{\#(\text{ Patients })} = \frac{43}{135} \approx 0.3185.$

Part 2.4.1: Definition

	Disease Present	Disease Absent	Total
Positive test	25	14	39
Negative test	18	78	96
Total	43	92	135

Now if a doctor knows that the test for the chosen person was positive (Event B), what is the probability of disease present given this knowledge?

$$\cdots = \frac{\#(\text{ Disease Present and Positive})}{\#(\text{ Positive })} = \frac{25}{39} \approx 0.6410.$$

Of course $0.3185 \neq 0.6410$.

Reason: for the second one: B has occurred, which affects the probability of A.

The second probability is called the probability of event A under the condition that B has occurred, or simply called: "the conditional probability of A given B, and is denoted by $P(A \mid B)$".

We can see that usually $P(A) \neq P(A \mid B)$.

But we can see that

$$
\begin{aligned}
P(A \mid B) &= \frac{25}{39} = \frac{\text{The number of Disease Present and Positive}}{\text{The number of Positive}} \\
&= \frac{\frac{25}{135}}{\frac{39}{135}} = \frac{\frac{\text{The number of Disease Present and Positive}}{\text{The total Number}}}{\frac{\text{The number of Positive}}{\text{The total Number}}} \\
&= \frac{P(A \cap B)}{P(B)}.
\end{aligned}
$$

Here, we need the condition that $P(B) > 0$, otherwise undefined.

Part 2.4.1: Definition

2. Definition:

Let A and B be two events with $P(B) \neq 0$. The conditional probability of A given B is defined to be

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{2.5.1}$$

Part 2.4.2: Multiplication Law

1. **Conclusion:** Let A and B be two events with $P(B) \neq 0$. Then
$$P(A \cap B) = P(B) \cdot P(A \mid B). \tag{2.5.2}$$
2. Proof: By (2.5.1) directly.
3. Also if $P(A) > 0$, then we can get
$$P(A \cap B) = P(A) \cdot P(B \mid A).$$

4. Application: Usually, $P(A \cap B)$ may be quite hard.
   But $P(B)$ and $P(A \mid B)$ are easy.

5. Example: An urn contains 3 red balls and 1 blue ball. Two balls are selected without replacement.

   What is the probability that they are both red.

Method 1: Without using the conditional probability.

Total number of outcomes: 4×3;

Total number of "Two reds": 3×2;

$$\Rightarrow \text{ Prob. } = \frac{3 \times 2}{4 \times 3} = \frac{1}{2}.$$

($\because$ Equally likely!)

Part 2.4.2: Multiplication Law

Method 2: Using the conditional probability.

A: the event that the first one is red.

B: the event that the second one is red.

Then the event that both are red is : $A \cap B$.

Easy to see $P(A) = \frac{3}{4}$ ($\because$ total 4; Red 3).

$$P(B \mid A) = ?$$

" A has occurred" $\Leftrightarrow$ "the first one is red" $\Leftrightarrow$ "3 left with 2 red".

$$\Rightarrow P(B \mid A) = \frac{2}{3}.$$

Now $P(A \cap B) = P(A) \cdot P(B \mid A) = \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2}$.

We get the same result: Both methods work.

However, the following example shows that without using conditional probability, the question would be very difficult.

6. More interesting example: Pólya's urn scheme.

   Suppose, originally we have m blue balls and n red balls.

   We draw a ball and note its color, then we replace it and add one more ball of the same color.

   What is the probability that the first and the second balls are both red?

Part 2.4.2: Multiplication Law

If we do not use the conditional probability, the problem seems very difficult.

Let's try to use the method of conditional probability.

Answer: Let

$$A: \text{"First red"}, \quad B: \text{"Second red"}.$$

$$P(A \cap B) = ? \quad \text{Not easy!!}$$

But: $P(A) = \frac{n}{m+n}$ (very easy).

$P(B \mid A) = \frac{n+1}{m+n+1}$ ((also easy).

$$\Rightarrow P(A \cap B) = P(A) \cdot P(B \mid A) = \frac{n}{m+n} \cdot \frac{n+1}{m+n+1}.$$

7. Remarks:

   (1)

$$\begin{aligned} P(A \cap B) &= P(B) \cdot P(A \mid B) \qquad && (\text{if } P(B) \neq 0) \\ &= P(A) \cdot P(B \mid A) \qquad && (\text{if } P(A) \neq 0). \end{aligned}$$

But $P(A \cap B) = P(A) \cdot P(A \mid B)$ **is wrong**.

Part 2.4.2: Multiplication Law

(2) If $P(B) = 0$, one cannot use the formula

$$P(A \cap B) = P(B) \cdot P(A \mid B).$$

Then

$$P(A \cap B) = ?$$

Answer: $P(A \cap B) = 0$.

Reason:

$$\begin{aligned} A \cap B \subset B \quad & \Rightarrow \quad 0 \leq P(A \cap B) \leq P(B) \\ & \Rightarrow \quad 0 \leq P(A \cap B) \leq 0 \\ & \Rightarrow \quad P(A \cap B) = 0. \end{aligned}$$

Part 2.4.2: Multiplication Law

(3) How to choose A and B?

According to the convenience !!

(4) $P(A \cap B \cap C) = ?$

Let $A \cap B = D$, then

$$\begin{aligned} P(A \cap B \cap C) &= P(D) \cdot P(C \mid D) \\ &= P(A \cap B) \cdot P(C \mid A \cap B) \\ &= P(A) \cdot P(B \mid A) \cdot P(C \mid A \cap B). \end{aligned}$$

Part 2.4.2: Multiplication Law

(4) In more general,

$$
\begin{aligned}
P\left(\cap_{i=1}^{n} A_i\right) &= P(A_1 \cap A_2 \cap \cdots \cap A_n) \\
&= P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1 \cap A_2) \times \cdots \\
&\quad \times P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}).
\end{aligned}
$$

For example,

$$
\begin{aligned}
P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1) \times P(A_2 \mid A_1) \times P(A_3 \mid A_1 \cap A_2) \\
\times P(A_4 \mid A_1 \cap A_2 \cap A_3).
\end{aligned}
$$

Part 2.4.3: Law of Total Probability

1. Idea: An example:

A school boy has 5 blue and 4 white marbles in his left pocket and 4 blue and 5 white marbles in his right pocket.

If he transfers one marble at random from his left to his right pocket, what is the probability of his then drawing a blue from his right pocket.

Original:

	Blue	White
Left	5	4
Right	4	5

One ball from left to right at random.

Part 2.4.3: Law of Total Probability

Let A be the event that drawing blue from right after transferring, $P(A) =$? Complicated?

But how about if we know the result of transferring? Easy. isnt it ? Indeed,

$$B_1 : \text{ the result of transferring being blue.}$$

$$P(A \mid B_1) = \frac{4+1}{9+1} = \frac{5}{10}.$$

$$B_2 : \text{ the result of transferring being white.}$$

$$P(A \mid B_2) = \frac{4}{9+1} = \frac{4}{10}.$$

Part 2.4.3: Law of Total Probability

However, relation between $P(A)$ and $P(A \mid B_i)$ etc?

Note that $B_1 \cup B_2 = \Omega, \quad B_1 \cap B_2 = \emptyset.$

$$\begin{aligned}\Rightarrow A = A \cap \Omega &= A \cap (B_1 \cup B_2) \\ &= (A \cap B_1) \cup (A \cap B_2).\end{aligned}$$

Easy to see $A \cap B_1$ and $A \cap B_2$ are disjoint ($\because$ B_1 and B_2 are!!)

$$\begin{aligned}\Rightarrow P(A) &= P(A \cap B_1) + P(A \cap B_2) \quad \text{(Think why here!)} \\ &= P(B_1) \cdot P(A \mid B_1) + P(B_2) \cdot P(A \mid B_2).\end{aligned}$$

How about $P(B_1)$ and $P(B_2)$?

Easy! $P(B_1) = \frac{5}{9}$ and $P(B_2) = \frac{4}{9}$.

($\because$ In left pocket: 5 Blue + 4 White !!)

Now:

$$P(A) = \frac{5}{9} \cdot \frac{5}{10} + \frac{4}{9} \cdot \frac{4}{10} = \frac{25 + 16}{90} = \frac{41}{90}.$$

2. **Conclusion:** Law of Total Probability:

(a) Let $B_1, B_2, \cdots, B_n$ be events such that

$$P(B_i) > 0, \quad \cup_{i=1}^n B_i = \Omega, \quad B_i \cap B_j = \emptyset, \quad \forall i \neq j.$$

Then for any event A, we have

$$P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A \mid B_i). \tag{2.5.3}$$

Part 2.4.3: Law of Total Probability

(b) **Proof:**

$$\because A = A \cap \Omega = A \cap \left(\cup_{i=1}^{n} B_i\right) = \cup_{i=1}^{n} \left(A \cap B_i\right),$$

therefore,

$$\begin{aligned} P(A) &= P\Big(\cup_{i=1}^{n} (A \cap B_i)\Big) \\ &= \sum_{i=1}^{n} P(A \cap B_i) \quad ((A \cap B_i) \text{ are disjoint !}) \\ &= \sum_{i=1}^{n} P(B_i) \cdot P(A \mid B_i). \end{aligned}$$

3. Notes:

(a) $\{B_i; i = 1, 2, \cdots, n\}$ is called a partition of Ω if

$$\cup_{i=1}^{n} B_i = \Omega, \quad B_i \cap B_j = \emptyset, \quad \forall i \neq j.$$

(b) The law is still true if the partition is "a sequence of events".

(c) In application, the most important thing is to find a suitable partition. This very important method is called "conditioning".

4. Example again:

In a certain population 5% of the females and 8% of the males are left-handed; and 48% of the population are males.

What is the probability that a randomly chosen member of the population is left-handed?

Analysis: Let A be the event: "the chosen member is left-handed".

$P(A) =$? Conditioning on what?

Certainly "gender"! (Since if we know the gender, then the conditional probability is easy!)

Solution: Let

$$\begin{aligned} A &: \text{ the event "left-handed"}, \\ B_1 &: \text{ the event "male"}, \\ B_2 &: \text{ the event "female"}. \end{aligned}$$

then

$$P(B_1) = 0.48, \qquad\qquad P(A \mid B_1) = 0.08,$$
$$P(B_2) = 1 - 0.48 = 0.52, \qquad\qquad P(A \mid B_2) = 0.05.$$

Part 2.4.3: Law of Total Probability

Now, by the law of total probability,

$$
\begin{aligned}
P(A) &= P\left(A \cap B_1\right) + P\left(A \cap B_2\right) \\
&= P\left(B_1\right) \cdot P\left(A \mid B_1\right) + P\left(B_2\right) \cdot P\left(A \mid B_2\right) \\
&= 0.48 \times 0.08 + 0.52 \times 0.05 = 0.0644.
\end{aligned}
$$

(Check: $B_1 \cup B_2 = \Omega$, $B_1 \cap B_2 = \emptyset$!!)

Part 2.4.4: Bayes' Rule

1. Example:

Return to the "left-handed" problem. We want to ask the following question:

Suppose a member has been chosen and found left-handed. What's probability that the person is male?

Analysis: A has occurred, we want to find $P(B_1 \mid A)$.

What can we do? (whenever in doubt about a conditional probability, try the definition.)

$$P(B_1 \mid A) = \frac{P(B_1 \cap A)}{P(A)}, \quad (\text{ Does this help? })$$

1. Example

Sure! Try to find the denominator $P(A)$ and the numerator

$$P(B_1 \cap A) = P(B_1) \cdot P(A \mid B_1)$$

Solution: Let

$$\begin{aligned} A &: \text{ the event "left-handed"}, \\ B_1 &: \text{ the event "male"}, \\ B_2 &: \text{ the event "female"}. \end{aligned}$$

Then

$$\begin{aligned} P(B_1) &= 0.48, & P(A \mid B_1) &= 0.08, \\ P(B_2) &= 1 - 0.48 = 0.52, & P(A \mid B_2) &= 0.05. \end{aligned}$$

Part 2.4.4: Bayes' Rule

Now

$$
\begin{aligned}
P\left(B_1 \mid A\right) &= \frac{P\left(A \cap B_1\right)}{P(A)} \\
&= \frac{P\left(B_1\right) \cdot P\left(A \mid B_1\right)}{P\left(B_1\right) \cdot P\left(A \mid B_1\right) + P\left(B_2\right) \cdot P\left(A \mid B_2\right)} \\
&= \frac{0.48 \times 0.08}{0.48 \times 0.08 + 0.52 \times 0.05} = 0.596.
\end{aligned}
$$

Similarly, we can get $P(B_2 \mid A)$.

2. Generalization:

$\{B_k\}$ is a partition of Ω; A is another event.

$P(A \mid B_k)$ etc. are easy to get.

Then how to get $P(B_k \mid A)$?

$$P(B_k \mid A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(B_k) \cdot P(A \mid B_k)}{\sum_n P(B_n) \cdot P(A \mid B_n)}.$$

Part 2.4.4: Bayes' Rule

3. Bayes' formula:

(1) Let $B_1, B_2, \cdots, B_n$ be events such that

$$P(B_i) > 0, \quad \cup_{i=1}^{n} B_i = \Omega, \quad B_i \cap B_j = \emptyset, \quad \forall i \neq j.$$

Then for any other event A and any B_i in the partition,

$$P(B_i \mid A)$$
$$= \frac{P(B_i)P(A \mid B_i)}{P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + \cdots + P(B_n)P(A \mid B_n)}.$$

(2) Proof:

$$
\begin{aligned}
P(B_i \mid A) &= \frac{P(A \cap B_i)}{P(A)} \quad \text{(Definition !!)} \\
&= \frac{P(B_i) \cdot P(A \mid B_i)}{P(A)} \quad \text{(Multiplication rule)} \\
&= \frac{P(B_i) \cdot P(A \mid B_i)}{\sum_{k=1}^{n} P(B_k) \cdot P(A \mid B_k)} \quad \text{(Total law of Probability).}
\end{aligned}
$$

4. Notes:

(1) Bayes' rule also holds true if the partition of Ω, $\{B_k\}$ is a sequence of events. We usually write it as

$$P(B_i \mid A) = \frac{P(B_i) \cdot P(A \mid B_i)}{\sum_k P(B_k) \cdot P(A \mid B_k)} \tag{2.5.4}$$

(The denominator in (2.5.4) is either a sum of finite terms or a series.)

(2) Memory: Using the “Proof” !!

Desired probability: $P(B_i \mid A)$.

Numerator: the reverse conditional probability $P(B_i \mid A)$ times the probability of the corresponding event.

Denominator: the sum of all possible terms like the numerator.

Outline

Part 2.5.1: Independence of Two Events

1. Motivation and Idea:

"Independence" is a very important concept in Probability Theory and Statistics.

Suppose A and B are two events, we know usually

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

Part 2.5.1: Independence of Two Events

However, in some cases, they might be the same.

In these cases, "given B occurred" does not affect the probability of event A. We then say "A and B are independent".

For some reason, we give another equivalent definition.

Note that if $P(A) = P(A \mid B) = \frac{P(A \cap B)}{P(B)}$, then

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

Part 2.5.1: Independence of Two Events

2. **Definition:** Two events A and B are called independent events if

$$P(A \cap B) = P(A) \cdot P(B). \tag{2.6.1}$$

3. Notes:

(1) If A and B are independent, then so is B and A. Also, condition $P(B) \neq 0$ is not needed.

(2) Condition (2.6.1) is convenient for checking the independence.

4. Example:

Experiment: A card is selected randomly from a deck.

Event A: "It is an ace"; Event B: "It is a diamond",

$$\Rightarrow P(A) = \frac{4}{52} = \frac{1}{13}, \quad P(B) = \frac{13}{52} = \frac{1}{4}$$

$A \cap B$: "It is a diamond ace".

$$P(A \cap B) = \frac{1}{52} = \frac{1}{13} \cdot \frac{1}{4}$$

$$\Rightarrow A, B \text{ are independent.}$$

5. Property:

Theorem 2.6.1. If A and B are two independent events, then the following pairs of events are also independent.

(1) A and B^c;

(2) A^c and B;

(3) A^c and B^c.

Proof: Easy and thus omitted.

Part 2.5.2: Independence of More than Two events

1. Independent of Three Events:

(1) Definition: Three events A, B and C, are called (mutually) independent if

   (i) $P(A \cap B) = P(A) \cdot P(B), P(A \cap C) = P(A) \cdot P(C),$
   $P(B \cap C) = P(B) \cdot P(C),$

   (ii) $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C).$

(2) Note: Only (i) holds true can not imply A, B and C, are independent. ((i) only is usually called pair-wise independent). Also, only (ii) is not enough for "independence".

Part 2.5.2: Independence of More than Two events

2. Independent of n events:

Definition: n events $A_1, A_2, \cdots, A_n$ are called (mutually) independence if the following hold:

(1) for <u>all pairs</u> A_i and A_j $(i \neq j)$,

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j),$$

(2) for <u>all triples</u> A_i, A_j, A_k $(i, j, k$ all different$)$,

$$P(A_i \cap A_j \cap A_k) = P(A_i) \cdot P(A_j) \cdot P(A_k).$$

(3) for all quadruples A_i, A_j, A_k, A_l (i, j, k, l are all different),

$$P(A_i \cap A_j \cap A_k \cap A_l) = P(A_i) \cdot P(A_j) \cdot P(A_k) \cdot P(A_l),$$

(until finally)

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n).$$

Part 2.5.2: Independence of More than Two events

3. Independence of infinitely many events:

We define an infinite set of events to be independent if every finite subset of these events is independent.

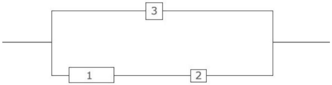
4. Remark: If $A_1, A_2, \cdots, A_n$ are independent, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n)$$

i.e.

$$P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i). \tag{2.6.2}$$

5. Example: Consider a circuit with three relays:



Assume that three relays are mutually independent and the working probability of each relay is p. What is the probability that current flows through the circuit.

Analysis: Let

$$A_i = \text{the event that the } i\text{th relay works } (i = 1, 2, 3),$$
$$F = \text{the event that current flows through the circuit.}$$

Then $F = A_3 \cup (A_1 \cap A_2)$ (Think why here!), and hence

$$P(F) = P(A_3) + P(A_1 \cap A_2) - P(A_1 \cap A_2 \cap A_3) = p + p^2 - p^3.$$

Outline

Part 2.6.1: Basic Concept

1. Sample Space: Ω.
2. Events: Impossible event $\emptyset$, Certain event Ω, Elementary event; General event.
3. Probability Measure: Set function: events $\to \mathbb{R}$.
4. Independence: ....
5. Conditional Probability: ...
6. Disjoint Events: ...
7. Partition of Ω: ...

Part 2.6.2: Operations of Events

1. Union: $A \cup B = \{\text{Either } A \text{ or } B \text{ occurs}\}$.
2. Intersection: $A \cap B = \{\text{Both } A \text{ and } B \text{ occur}\}$.
3. Complement: $A^c = \{A \text{ does not occur}\}$.
4. $\cup_{k=1}^{n} A_k$ and $\cup_{k=1}^{\infty} A_k, \cap_{k=1}^{n} A_k$ and $\cap_{k=1}^{\infty} A_k$.

Part 2.6.3: Properties of Probability

1. $0 \leq P(A) \leq 1, \quad \forall A.$
2. $P(\emptyset) = 0, \quad P(\Omega) = 1.$
3. $A \subset B \Rightarrow P(A) \leq P(B).$
4. $\{B_k\}$ disjoint $\Rightarrow P(\cup_k B_k) = \sum_k P(B_k).$
5. $\{B_k\}$ independent $\Rightarrow P(\cap_{k=1}^n B_k) = \prod_{k=1}^n P(B_k).$

Part 2.6.4: Important Formulae

1. $P(A^c) = 1 - P(A)$.
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. $P(A \cap B) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$.
4. $P(A \cap B) = P(A) \cdot P(B)$ if A, B independent.
5. If $\{B_k\}$ is a partition of Ω, then for any A,

$$P(A) = \sum_k P(B_k) \cdot P(A \mid B_k),$$

$$P(B_n \mid A) = \frac{P(B_n) \cdot P(A \mid B_n)}{\sum_k P(B_k) \cdot P(A \mid B_k)}.$$