Probability and Statistics

Module 1: Introduction to Probability

Complete classroom and examination notes

B.Sc. Computer Science • Course Code DASH250011

Based on the NIELIT Deemed to be University 2025 curriculum

Contents

1 Module map 1
2 1. Basic language of probability 1
2.1 Random experiment 1
2.2 Sample-space examples 1
3 2. Events and set operations 1
3.1 Mutually exclusive events 2
3.2 Partition of a sample space 2
4 3. Classical definition of probability 2
4.1 Counting tools used in classical probability 2
5 4. Axiomatic definition and probability laws 2
6 5. Conditional probability 3
6.1 Tree diagrams 3
7 6. Law of total probability 3
8 7. Independent events 4
9 8. Bayes' theorem 4
9.1 Bayes' theorem as an update rule 4
10 9. Applications in data science 4
11 10. Problem-solving checklist 5
12 11. Quick formula sheet 5
13 12. Practice questions with answers 5

1 Module map

The syllabus module contains: sample space and events; classical and axiomatic definitions of probability; conditional probability; independent events; Bayes' theorem; and applications in data science. These notes add the prerequisite counting ideas and worked examples needed to solve problems confidently.

> **Learning outcomes.** After studying this module, you should be able to define a random experiment, construct a sample space, manipulate events, calculate probabilities using the axioms and counting, distinguish independence from mutual exclusivity, apply conditional probability and Bayes' theorem, and interpret probability in data-science settings.

2 1. Basic language of probability

2.1 Random experiment

A **random experiment** is a repeatable process with a well-defined set of possible outcomes, where the exact outcome cannot be predicted with certainty before the experiment. Examples include tossing a coin, selecting a record from a database, or measuring packet delay.

An **outcome** is one possible result. The **sample space**, written Ω, is the set of all possible outcomes. An **event** is a subset of Ω.

Term	Meaning
Experiment	The procedure being observed.
Outcome	A single possible result, often written ω.
Sample space Ω	The set of every possible outcome.
Event A	Any subset of Ω. $A \subseteq \Omega$.
Probability $P(A)$	A number measuring how likely event A is; $0 \le P(A) \le 1$.

2.2 Sample-space examples

- One coin toss: $\Omega = \{\mathrm{H}, \mathrm{T}\}$.
- Two coin tosses: $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$ if order is recorded.
- One die: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Two dice: $\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$, so there are 36 ordered outcomes.
- A measurement such as response time: Ω may be an interval of real numbers, for example $[0, \infty)$.

> **Modelling decision.** Decide whether outcomes are ordered, whether outcomes are equally likely, and whether the experiment is discrete or continuous before writing a formula. Many errors begin with an incorrectly constructed sample space.

3 2. Events and set operations

For events A and B:

- **Complement**: A^c (or $|A|$) means "A does not occur."
- **Union**: $A \cup B$ means "A or B or both."
- **Intersection**: $A \cap B$ means "both A and B."

- **Difference**: $A - B$ means "A occurs but B does not."
- **Empty event**: $\varnothing$ can never occur; $P(\varnothing) = 0$.
- **Certain event**: Ω must occur; $P(\Omega) = 1$.

Important identities:

$$A \cup B = B \cup A \quad ; A \cap B = B \cap A$$
$$(A \cup B)^c = A^c \cap B^c \quad ; (A \cap B)^c = A^c \cup B^c$$
$$A \cup A^c = \Omega \quad ; A \cap A^c = \varnothing$$

3.1 Mutually exclusive events

Events A and B are **mutually exclusive** (disjoint) when they cannot occur together: $A \cap B = \varnothing$. For example, on one die, "even" and "odd" are mutually exclusive. Mutually exclusive events with positive probabilities cannot be independent, because their inter has probability 0.

3.2 Partition of a sample space

Events $B_1, B_2, \ldots, B_n$ form a partition of Ω when they are pairwise disjoint and their union is Ω. Partitions are central to the law of total probability and Bayes' theorem.

4 3. Classical definition of probability

When a finite sample space has equally likely outcomes, the **classical probability** of event A is

$$P(A) = \frac{\text{number of favourable outcomes}}{\text{total number of outcomes}} = \frac{|A|}{|\Omega|}.$$

Example: For two fair dice, the probability of obtaining a sum of 8 is $\frac{5}{36}$, because the favourable ordered pairs are $(2,6), (3,5), (4,4), (5,3), (6,2)$.

> **Worked example: at least one six**
>
> For two fair dice, $P(\text{at least one six})$ is easier by complement. There are 36 outcomes and 25 outcomes with no six, so $P(\text{at least one six}) = 1 - \frac{25}{36} = \frac{11}{36}$.

4.1 Counting tools used in classical probability

- **Multiplication principle**: if a task has m choices followed by n choices, there are mn combined choices.
- **Permutation**: arrangements of r objects selected from n: $P(n,r) = \frac{n!}{(n-r)!}$.
- **Combination**: selections of r objects from n: $C(n,r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$.
- **Complement counting**: count the total minus the unwanted cases.

Use permutations when order matters; use combinations when order does not matter.

5 4. Axiomatic definition and probability laws

The **Kolmogorov axioms** define a probability measure P on events:

1. **Non-negativity:** $P(A) \geq 0$ for every event A.
2. **Normalization:** $P(\Omega) = 1$.
3. **Countable additivity:** for pairwise disjoint events $A_1, A_2, \ldots$, $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$.

From these axioms follow the most-used laws:

Law	Formula
Complement	$P(A^c) = 1 - P(A)$
Addition / union	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Disjoint addition	If $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$.
Monotonicity	If $A \subseteq B$, then $P(A) \le P(B)$.
Difference	$P(A - B) = P(A) - P(A \cap B)$

Worked example: addition rule

If $P(A) = 0.55$, $P(B) = 0.40$, and $P(A \cap B) = 0.20$, then $P(A \cup B) = 0.55 + 0.40 - 0.20 = 0.75$. The inter is subtracted because it was counted twice.

6 5. Conditional probability

The **conditional probability** of A given that B has occurred is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0.$$

It restricts attention to the outcomes inside B. Rearranging gives the **multiplication rule**:

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

For three events, $P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B)$.

Worked example: defective components

A supplier produces 2% defective components. A screening test flags a defective component with probability 0.95, so $P(F \mid D) = 0.95$. If a component is good, the false-positive probability is 0.04, so $P(F \mid D^c) = 0.04$. The probability of a flag is computed by total probability in the next section.

6.1 Tree diagrams

A probability tree lists stages from left to right. Multiply probabilities along one path to obtain the probability of that path; add probabilities of disjoint paths to obtain the probability of an event. Conditional probabilities belong on branches after the first stage.

7 6. Law of total probability

If $B_1, \dots, B_n$ partition Ω and $P(B_i) > 0$, then for any event A,

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

For two cases: $P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$.

Worked example: screening system

Let D mean "defective" and F mean "flagged." Suppose $P(D) = 0.02$, $P(F \mid D) = 0.95$, and $P(F \mid D^c) = 0.04$. Then $P(F) = 0.95(0.02) + 0.04(0.98) = 0.0582$. Thus about 5.82% of all components are flagged.

8 7. Independent events

Events A and B are **independent** when knowledge that one occurred does not change the probability of the other:

$$P(A \texttt{ mid } B) = P(A) \quad (P(B) > 0).$$

Equivalent tests are $P(B \texttt{ mid } A) = P(B)$ and $P(A \cap B) = P(A)P(B)$.

For independent events, complements are also independent: A and B independent implies A^c and B, A and B^c, and A^c and B^c are independent. For independent events $A_1, \ldots, A_n$, $P(A_1 \cap \ldots \cap A_n) = P(A_1)\ldots P(A_n)$.

> **Do not confuse independence and mutual exclusivity.** Mutually exclusive means the events cannot happen together. Independence means one event gives no information about the other. If both have positive probability, the two properties cannot hold simultaneously.

9 8. Bayes' theorem

For events A and B with positive probabilities,

$$P(A \texttt{ mid } B) = \frac{P(B \texttt{ mid } A)P(A)}{P(B)}.$$

With a partition $B_1, \ldots, B_n$,

$$P(B_j \texttt{ mid } A) = \frac{P(A \texttt{ mid } B_j)P(B_j)}{\sum_i P(A \texttt{ mid } B_i)P(B_i)}.$$

The terms have intuitive names: $P(B_j)$ is the prior, $P(A \texttt{ mid } B_j)$ is the likelihood, $P(A)$ is the evidence, and $P(B_j \texttt{ mid } A)$ is the posterior.

> **Worked example: interpreting a positive test**
>
> Suppose disease prevalence is $P(D) = 0.01$, sensitivity $P(+ \texttt{ mid } D) = 0.95$, and specificity $P(- \texttt{ mid } D^c) = 0.90$. Thus $P(+ \texttt{ mid } D^c) = 0.10$. The evidence is $P(+) = 0.95(0.01) + 0.10(0.99) = 0.1085$. Therefore $P(D \texttt{ mid } +) = \frac{0.0095}{0.1085} \approx 0.0876$, or about 8.76%. A positive result is not the same as a 95% chance of disease; the base rate matters.

9.1 Bayes' theorem as an update rule

1. List mutually exclusive and exhaustive hypotheses.
2. Write their prior probabilities.
3. Write the likelihood of the observed evidence under each hypothesis.
4. Compute total evidence using total probability.
5. Divide the selected hypothesis' prior-times-likelihood by the evidence.
6. Check that posterior probabilities sum to 1.

10 9. Applications in data science

Probability is a language for uncertainty in data-driven systems:

- **Classification**: Bayes classifiers estimate $P(\text{class} \texttt{ mid } \text{features})$.
- **Spam filtering**: update the probability of spam after observing words, sender information, or links.

- **Medical and anomaly detection**: combine base rates with test evidence; precision is strongly affected by prevalence.
- **Reliability**: compute the probability that a system functions when components fail.
- **A/B testing**: model uncertain conversion rates and compare evidence from two groups.
- **Risk scoring**: assign probabilities to events such as churn, fraud, or failure.
- **Sensor fusion**: combine evidence from multiple sensors, accounting for dependence rather than multiplying blindly.

A good data-science probability model states the population, event definitions, assumptions, source of priors, and whether observations are independent.

11 10. Problem-solving checklist

1. Define the experiment and the event in words.
2. Write the sample space or a suitable partition.
3. Decide whether outcomes are equally likely.
4. Use a complement when "at least one" or "none" appears.
5. Use the addition rule for "or," multiplication/conditional probability for "and."
6. Use total probability when cases form a partition.
7. Use Bayes when the question reverses a conditional probability.
8. Check that the final answer lies in $[0, 1]$ and matches the context.

12 11. Quick formula sheet

Concept	Formula
Complement	$P(A^c) = 1 - P(A)$
Union	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Conditional probability	$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
Multiplication	$P(A \cap B) = P(A \mid B)P(B)$
Independence	$P(A \cap B) = P(A)P(B)$
Total probability	$P(A) = \sum_i P(A \mid B_i)P(B_i)$
Bayes	$P(B_j \mid A) = P(A \mid B_j)\frac{P(B_j)}{P(A)}$

13 12. Practice questions with answers

1. A fair die is rolled. Find the probability of a number greater than 4 or even. **Answer:** $P(\{5, 6\} \cup \{2, 4, 6\}) = \frac{4}{6} = \frac{2}{3}$.
2. Two cards are drawn without replacement from a standard deck. Find the probability both are aces. **Answer:** $\left(\frac{4}{52}\right)\left(\frac{3}{51}\right) = \frac{1}{221}$.
3. If $P(A) = 0.6$, $P(B) = 0.5$, and $P(A \cap B) = 0.3$, are A and B independent? **Answer:** Yes, because $0.3 = 0.6(0.5)$.
4. A machine makes 70% of items with defect rate 1% and 30% with defect rate 4%. Find the defect probability. **Answer:** $0.7(0.01) + 0.3(0.04) = 0.019$.

5. In the preceding problem, given a defective item, find the probability it came from the second machine. **Answer:** $0.3\frac{0.04}{0.019} \approx 0.6316$.
6. Explain why a high-accuracy detector can have low precision. **Answer:** If the condition is rare, false positives from the much larger non-condition group can dominate; Bayes' theorem accounts for the base rate.

Exam reminder. Always show the event definitions and the rule used. A correct numerical answer without identifying the sample space, complement, conditioning event, or partition often earns less credit than a clearly justified solution.