Systems and Signals 344

2026

Lecture 2

Probability Axioms and Conditional Probability

Sections 1.2-1.3

Recap

Probability is based on a repeatable **experiment** that consists of a *procedure* and *observations* with an associated *model*.

This leads to a *set-theory* representation with a:

- **sample space** (universal set S),
- **outcomes** (s that are elements of S), and
- **events** (A that are sets of outcomes/elements).

To complete the model, we assign a *probability* $P(A)$ to every event, A, in the sample space.

A is a valid event if its probability $P(A)$ is defined.

Recap

- A and B are **disjoint** (or **mutually exclusive**) if their intersection is the empty set, i.e. when $A \cap B = \emptyset$

Comments on Notation

In this module we will primarily use $P(\cdot)$ to indicate the probability of an event within the context of an experiment. The expression in brackets indicates the event.

Examples are $P(A)$, $P(\textbf{roll a 4})$, and $P(\{1,4,5\})$.

Note that the textbook prefers using $\mathbf{P}[A]$ instead, and you will encounter various other notations online.

Comments on Notation

We will also sometimes use some shorthand:

$$P(A \cap B) = P(A, B) = P(AB)$$

$P(A \cup B) = P(A + B)$ a lot less common

I will also sometimes read $\cap$ as *and* and $\cup$ as *or*

The Foundation of Probability Axioms

A probability model assigns a number between 0 and 1 to every event.

The probability of the union of mutually exclusive events is the sum of the probabilities of the events in the union.

Axioms of Probability

A probability measure $P(\cdot)$ is a *function* that maps events in the sample space to real numbers such that:

- **Axiom 1:** For any event A, $P(A) \geq 0$.
- **Axiom 2:** $P(S) = 1$.
- **Axiom 3:** For any countable collection $A_1, A_2, \dots$ of mutually exclusive events:

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

We will build our entire theory of probability on these three axioms.

Relative Frequency

One useful and practical definition of probability that satisfies these axioms is the **relative frequency**:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

where n is the total number of **trials** (times the experiment is repeated) and n_A is the total number of times it resulted in A

Relative Frequency

Python

```
import numpy as np
rng = np.random.default_rng()

num_rolls = 10

rolls = rng.integers(1, 6, num_rolls, endpoint=True)
print(rolls)

for i in range(6):
    print("P({}) = {}".format(i+1,
        np.round(np.count_nonzero(rolls == (i+1))/num_rolls, 3)))
```

Story Time

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was *deeply concerned with issues of discrimination* and social justice, and also participated in anti-nuclear demonstrations.

Story Time

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was *deeply concerned with issues of discrimination* and social justice, and also participated in anti-nuclear demonstrations.

Which one is more likely?

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

Story Time

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was *deeply concerned with issues of discrimination* and social justice, and also participated in anti-nuclear demonstrations.

There are 100 people that fit this description. How many are:

1. Bank tellers?
2. Bank tellers who are active in the feminist movement?

Theorems

For mutually exclusive events A_1 *and* A_2,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

Theorems

For mutually exclusive events A_1 and A_2,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

*If A is **partitioned** such that $A = A_1 \cup A_2 \cup \ldots \cup A_m$ and $A_i \cap A_j = \emptyset \; \forall \; i \neq j$ then*

$$P(A) = \sum_{i=1}^{m} P(A_i)$$

For mutually exclusive events, the probability of the union is simply the sum of the event probabilities.

Theorems

The probability of an event $B = \{s_1, s_2, \ldots, s_m\}$ is the sum of the probabilities of the outcomes obtained in the event: $P(B) = \sum_{i=1}^{m} P(\{s_i\})$

Side Note

- You can only *add* probabilities directly for the *union* ($\cup$) of events that are **mutually exclusive**:
  - $P(A \cup B) = P(A) + P(B)$ when $A \cap B = \emptyset$.
  - In general, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- You can only *multiply* probabilities directly for the *intersection* ($\cap$) of events that are **independent**:
  - $P(A \cap B) = P(A)P(B)$ when A and B are independent.
  - In general, $P(A \cap B) = P(A)P(B \mid A)$ when $P(A) > 0$.

Theorems

The probability measure $P(\cdot)$ *satisfies*

- $P(\bar{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- *For any* A *and* B *(not necessarily mutually exclusive),*
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- *If* $A \subseteq B$*, then* $P(A) \leq P(B)$

Joint Probability

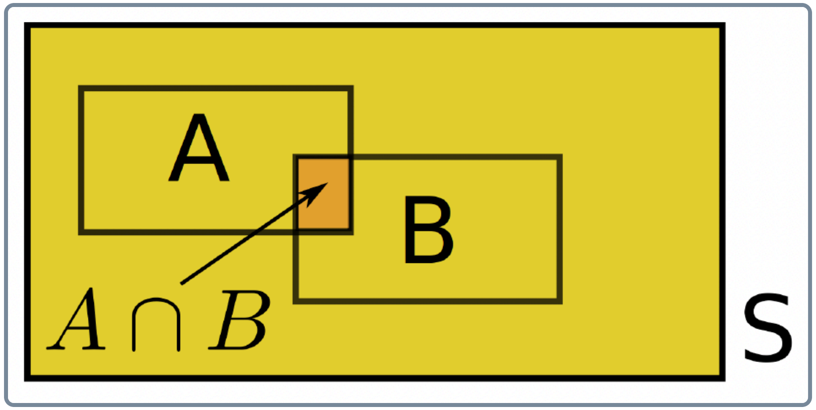
From $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ it follows that

$$P(A \cup B) \leq P(A) + P(B)$$

and

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

The latter is called **joint probability**



Theorems

For an experiment with sample space $S = \{s_1, \ldots, s_n\}$ in which each outcome s_i is equally likely,

$$P(s_i) = \frac{1}{n} \quad 1 \leq i \leq n$$

Conditional Probability

Conditional probabilities correspond to a modified probability model that reflects partial information about the outcome of an experiment. The modified model restricts the sample space to B.

Probability and Certainty

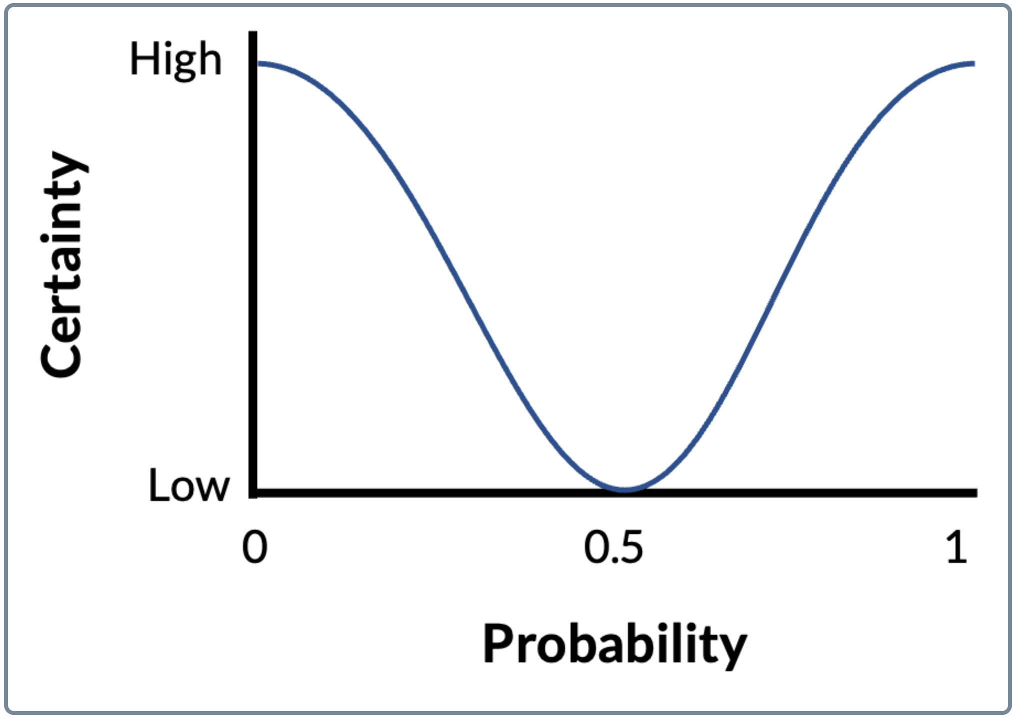
It is sometimes useful to interpret $P(A)$ as our *knowledge* of the occurrence of event A *before* an experiment takes place.

Probability and Certainty

It is sometimes useful to interpret $P(A)$ as our *knowledge* of the occurrence of event A *before* an experiment takes place.

Probability is an indication of our *certainty* about some event.

The probability of an event is a number between 0 and 1.



Probability and Certainty

Thus $P(A)$ reflects our knowledge of the occurrence of A *prior* to performing an experiment.

Sometimes, we refer to $P(A)$ as the ***a priori*** probability, or the **prior probability**, of A.

If we know the outcomes of the experiment, we refer to this as ***a posteriori*** probability.

Conditional Probability

Conditional probability describes our knowledge of A when we know that B has occurred but we still don't know the precise outcome of the experiment.

Conditional Probability

Conditional probability describes our knowledge of A when we know that B has occurred but we still don't know the precise outcome of the experiment.

We use new notation to indicate this:

$$P(A|B)$$

The probability of A *given* B.

Conditional Probability

Conditional probability describes our knowledge of A when we know that B has occurred but we still don't know the precise outcome of the experiment.

We use new notation to indicate this:

$$P(A|B)$$

The probability of A *given* B.

Or: The probability of A occurring given that we know that B occurs.

Conditional Probability

The conditional probability of the event A given the occurrence of the event B is

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

Conditional Probability

The conditional probability of the event A given the occurrence of the event B is

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

This ratio is defined only when $P(B) > 0$. When $P(B) = 0$, the ratio formula is undefined; a zero-probability event is not necessarily impossible.

Conditional Probability

A conditional probability measure $P(A|B)$ has the following properties that correspond to the axioms of probability:

- **Axiom 1:** $P(A|B) \geq 0$.
- **Axiom 2:** $P(B|B) = 1$.
- **Axiom 3:** If $A = A_1 \cup A_2 \cup \ldots$ with $A_i \cap A_j = \emptyset \; \forall \; i \neq j$

$$P(A|B) = P(A_1|B) + P(A_2|B) + \ldots$$

Conditional Probability: Side Note

Correct

$$P(A|B) = 1 - P(\bar{A}|B)$$

Incorrect

$$P(A|B) = 1 - P(A|\bar{B})$$

Example

You roll two fair 6-sided dice. What is the probability of:

- Event A, that the sum is greater than 6.
- Event B, that the first die is greater than 3.
- The sum being greater than 6, given that the first die was greater than 3.