Probability Theory

Ben Gravell

benjamin.gravell@utdallas.edu
The Erik Jonsson School of Engineering and Computer Science
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080

My goal is to give you theory foundations and practical tools for your research

I'll give lots of definitions, but the underlying concepts are typically simple

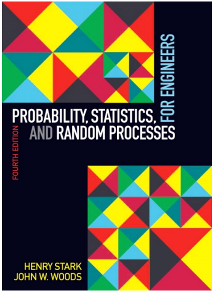
Do the exercises to check your understanding

All referenced Python code is in the `probability_theory` folder

I'm only giving you a small taste of this rich field - take further courses and study on your own!

I will cover material from

- **Stark & Wood's textbook "Probability, Statistics, and Random Processes for Engineers" [1]**
- Assorted other textbooks
- My own experience



Outline

What is probability?

"Probability is a mathematical model to help us study physical systems in an average sense. We have to be able to repeat the experiment many times under the same conditions. Probability then tells us how often to expect the various outcomes." [1]

Why study and use probabilistic models?

"We are forced to use probabilistic models in the real world because we do not know, cannot calculate, or cannot measure all the causes contributing to an effect. The causes may be too complicated, too numerous, or too faint." [1]

Interpretations of probability

Generic
"Probability" means the chance of something

Frequentist
"Probability" means the relative frequency of events

Bayesian
"Probability" means the degree to which we believe something to be true

Axiomatic
"Probability" is a mathematical construct that follows a set of rules

- No interpretation needed - conclusions follow logically from premises
- Be prepared for **counter-intuitive** conclusions

Preliminaries

Set

A **set** is a collection of individual **elements**.

Sets are denoted by braces, with the elements e_i contained inside

$$S = \{e_1, e_2, e_3, \ldots\} \tag{1}$$

Often constructed via set-builder notation

$$S = \{e_i \mid \text{predicate}(e_i)\} \tag{2}$$

"the set of all elements ee-eye such that the predicate holds for ee-eye"

An element e is "in" a set S if S contains e, denoted as $e \in S$.

The **cardinality** of a set is the number of elements in the set.

Examples of Sets

- The set of people reading this slide right now
- The set of hairs on your head
- The **empty set**, denoted $\emptyset$, the set containing nothing at all
  - $\emptyset$ is the only set with cardinality zero
- The set containing the empty set $\{\emptyset\}$
  - This set is not itself empty - it has cardinality one
- The **universal set**, denoted $\mathbb{U}$, the set containing every possible element
- The set of **whole numbers**, denoted $\mathbb{W} = \{0, 1, 2, 3, \ldots\}$
  - It has cardinality $\aleph_0$, a countable infinity
- The set of **real numbers**, denoted $\mathbb{R}$
  - It has cardinality $\mathfrak{c} = 2^{\aleph_0} > \aleph_0$, an uncountable infinity
    - See Cantor's diagonal argument from 1891

Basic mathematical operations that apply to **truth/false statements**

- Just like "standard" math operations that apply to numbers like addition, multiplication, etc.

Let x and y be two truth values

Operation	Notation	Definition
Disjunction	$x \vee y$	x is true **or** y is true
Conjunction	$x \wedge y$	x is true **and** y is true
Negation	$\neg x$	x is **not** true
Equivalence	$x \leftrightarrow y$	x is true **if and only if** y is true

Basic mathematical operations that apply to **sets**

- Defined with Boolean algebra applied to set membership

Let E and F be two sets

Operation	Notation	Definition
Union	$E \cup F$	Set of all elements in E **or** in F
Intersection	$E \cap F$	Set of all elements in E **and** F
Complement	E^c	Set of all elements **not** in E
Difference	$E - F$	Set of all elements in E **and not** in F
Exclusive Union	$E \oplus F$	Set of all elements in E **or** F **and not in both**
Subset	$E \subset F$	Every element in E is also in F
Superset	$E \supset F$	Every element in F is also in E
Equality	$E = C$	Every element in E is also in F and vice versa.

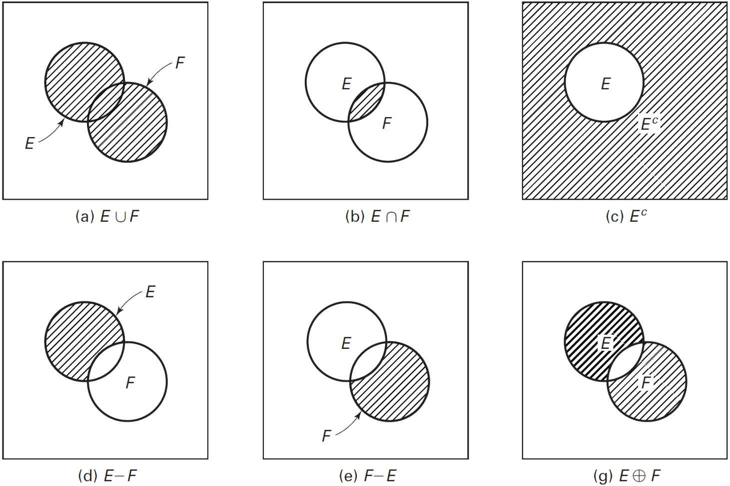


Figure 1: Set operations: (a) Union (b) Intersection (c) Complement (d) Difference (e) Difference (f) Exclusive Union

Set terminology

Let $\{E_i\}$ be a collection of sets

Let A be another set (if unspecified, the universal set $A = \mathbb{U}$ is implied)

- $\{E_i\}$ is **disjoint** or **mutually exclusive** if no elements are shared between any two different sets
- $\{E_i\}$ **collectively exhausts** A if the union of $\{E_i\}$ is A
- $\{E_i\}$ **partitions** A if $\{E_i\}$ is disjoint and collectively exhausts A

Set algebra laws

Set operations are related by simple laws, can be proved using Boolean logic (e.g. truth tables) and definitions

Examples:

- $E = F \quad \leftrightarrow \quad (E \subset F) \wedge (E \supset F)$
- $E \cap E^c = \emptyset$
- $E \cup E^c = \mathbb{U}$
- $E - F = E \cap F^c$
- $E \oplus F = (E - F) \cup (F - E) = (E \cup F) \cap (E \cap F)^c$

More set algebra laws

De Morgan's laws

- $\left[\bigcup_{i=1}^{n} E_i\right]^c = \bigcap_{i=1}^{n} E_i^c$
- $\left[\bigcap_{i=1}^{n} E_i\right]^c = \bigcup_{i=1}^{n} E_i^c$

Associative laws

- $A \cup (B \cup C) = (A \cup B) \cup C$
- $A \cap (B \cap C) = (A \cap B) \cap C$

Distributive laws

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Outcome

A **random experiment** results in **individual outcomes**, denoted as ζ.

Sample space

The **sample space** of a random experiment is the set of all possible outcomes of the experiment, denoted as Ω.

Event

An **event** is a subset of the sample space i.e. a set of outcomes.

In probability

- The sample space plays Ω the role of the universal set $\mathbb{U}$, and is called the **certain event**.
- The empty set $\emptyset$ is called the **null event**.
- Any individual outcome ζ is an element of Ω.

Sigma fields (sigma algebras)

Field

The collection of events $\mathcal{F} = \{E_i\}$ is a **field** if

1. $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$
2. If $E_i \in \mathcal{F}$ for all $i = 1, \dots, n$, then $\bigcup_{i=1}^{n} E_i \in \mathcal{F}$ and $\bigcap_{i=1}^{n} E_i \in \mathcal{F}$
   - "Closed under **finite** union and intersection"
3. If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$
   - "Closed under complement"

If condition 2 further holds with n countably infinite i.e. "closed under **countably infinite** union and intersection", then $\mathcal{F}$ is a **sigma (σ) field**.

Ensures any union, intersection, and complement of any set of events is well-defined (by construction).

If Ω is continuous and thus uncountable, e.g. $\Omega = \mathbb{R}$, we can generate a sigma field from the set of all open and closed intervals in Ω.

- In this case the sigma field is called the **Borel field**.

We can compute sigma fields of finite and discrete Ω using combinatorics

- See `sigma_field.py`

Axiomatic definition of probability

Probability is a function that maps events to real numbers $P[\cdot] : \mathcal{F} \to [0, 1]$ that satisfies three axioms

1. $P[E] \geq 0$
2. $P[\Omega] = 1$
3. $P[E \cup F] = P[E] + P[F]$ if $P[EF] = \emptyset$

From the axioms we can establish the additional properties

4. $P[\emptyset] = 0$
5. $P[E - F] = P[E] - P[E \cap F]$
6. $P[E^c] = 1 - P[E]$
7. $P[E \cup F] = P[E] + P[F] - P[EF]$

Probability formulation example

Example: Single coin flip

- Sample space is $\Omega = \{H, T\}$ where $H =$ heads, $T =$ tails
- There are 2^2 possible events, $\emptyset$, H, T, Ω
    - Consider events H and T with equal probability
- σ-field is $\mathcal{F} = \{\emptyset, H, T, \Omega\}$

Example: Die roll

- Sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$
- There are 2^6 possible events, each one containing, or not, each of the 6 possible outcomes
    - Consider events $\{1, 3\}$ and $\{2, 3, 4\}$
    - Consider each singleton event equally probable i.e. $P[\{i\}] = 1/6$
- σ-field is...tedious - see Example 1.4-9 [1]

Probability of a union

Probability of a union of disjoint events

Let $\{E_i\}_{i=1}^{n}$ be a set of mutually disjoint events, i.e.
$E_i \cap E_j = \phi$ for all $i \neq j$.
Then

$$P\left[\bigcup_{i=1}^{n} E_i\right] = \sum_{i=1}^{n} P\left[E_i\right]. \tag{3}$$

Proof: Use mathematical induction with Axiom 3.

Union bound (Boole's inequality)

Let $\{E_i\}_{i=1}^n$ be a set of events.
Then

$$P\left[\bigcup_{i=1}^n E_i\right] \leq \sum_{i=1}^n P\left[E_i\right]. \tag{4}$$

Proof: Use mathematical induction with Axiom 7.

Note: The only difference vs the previous result is that the events E_i are not assumed disjoint - the union bound always applies!

Probability of a union

Bonferroni inequality

Let $\{E_i\}_{i=1}^{n}$ be a set of events. Define the sums

$$S_m = \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} P\left[\bigcap_{j=1}^{m} E_{i_j}\right] \tag{5}$$

Then for any $k \in \{1, \ldots, n\}$

$$P\left[\bigcup_{i=1}^{n} E_i\right] \quad \begin{Bmatrix} \leq & \text{if } k \text{ odd} \\ \geq & \text{if } k \text{ even} \\ = & \text{if } k = n \end{Bmatrix} \quad \sum_{j=1}^{k} (-1)^{j-1} S_j \tag{6}$$

Proof: Use mathematical induction, see Theorem 1.5-1 in [1].

Note: Bonferroni is more tedious, but gives tighter bounds than Boole

Let A and B be two events with nonzero probability.

Joint probability

The **joint probability** of events A and B is the probability of their intersection $P[A \cap B]$.

Intuitively, it is the probability that both A and B will occur.

Conditional probability

The **conditional probability** of event A given B is the ratio

$$P[A|B] = \frac{P[A \cap B]}{P[B]}. \tag{7}$$

Intuitively, it is the probability that event A will occur, given the knowledge that event B already occurred.

Product Rule for events

Product Rule for events

The joint probability of events A and B can be computed as

$$P[A \cap B] = P[B|A]P[A] \tag{8}$$

When the events are independent we recover the

Proof: Follows by rearranging the definition of conditional probability.

Sum Rule for events

Sum Rule for events

Suppose the events $\{A_i\}_{i=1}^n$ are disjoint and collectively exhaustive, i.e.

- $A_i \cap A_j = \emptyset$ for any $i \neq j$
- $\bigcup_{i=1}^n A_i = \Omega$

Then the **total probability** of event B can be computed as

$$P[B] = \sum_{i=1}^n P[B|A_i]P[A_i] = \sum_{i=1}^n P[B \cap A_i] \tag{9}$$

Proof: Follows by the product rule and the assumptions on the A_i's.

The sum rule is useful when the conditional probabilities or intersection probabilities are readily available but the total probability is not.

The sum rule is also known as the **law of total probability**.

The total probability is also known as the **marginal probability**, since we are *marginalizing out* the other events A_i.

Total probability - example

Microchip factories

Given information:

1. Factory A makes 4000 chips/day with defect rate of 5%
2. Factory B makes 2000 chips/day with defect rate of 2%
3. Chips from both factories are mixed together at the end of each day then sent to a lab for testing

Question:
What is the probability of getting a defective chip at the lab?

Solution:

Denote the following events:

- D: Chip is defective
- A: Chip is from factory A
- B: Chip is from factory B

First compute base probabilities from frequency of occurrence:

$$P[A] = \frac{4000}{4000+2000} = 66.7\% \tag{10}$$

$$P[B] = \frac{2000}{4000+2000} = 33.3\% \tag{11}$$

Now use the law of total probability:

$$\begin{aligned} P[D] &= P[D|A]P[A] + P[D|B]P[B] && (12) \\ &= (5\%)(66.7\%) + (2\%)(33.3\%) && (13) \\ &= \boxed{4\%} && (14) \end{aligned}$$

Statistical independence

Statistical independence

Two events A and B are **statistically independent** if and only if

$$P[A \cap B] = P[A]P[B]. \tag{15}$$

Equivalently, the conditional and unconditional probabilities of A and B are equal:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A]P[B]}{P[B]} = P[A] \tag{16}$$

$$P[B|A] = \frac{P[B \cap A]}{P[A]} = \frac{P[B]P[A]}{P[A]} = P[B] \tag{17}$$

Intuitively, the outcome B has no effect on the chance of A occurring, and vice versa.

What if there are more than 2 events?

Joint statistical independence

The events $\{A_i\}_{i=1}^{n}$ are **jointly statistically independent** if and only if for all $k \in \{1, 2, \ldots, n\}$

$$P\left[\bigcap_{1 \leq i_1 < i_2 < \cdots \leq i_k} A_{i_k}\right] = \prod_{1 \leq i_1 < i_2 < \cdots \leq i_k} P\left[A_{i_k}\right] \tag{18}$$

Note: pairwise independence does not suffice!

- See e.g. this note `http://faculty.washington.edu/fm1/394/Materials/2-3indep.pdf`

Statistical independence

Pit-stop to build your intuition

Question: Can two disjoint events A and B with $P[A] > 0$, $P[B] > 0$ be statistically independent?

Think about it for a moment

Claim: No, A and B **must be dependent**

Explanation:

1. A, B disjoint means $A \cap B = \emptyset$ which implies $P[A \cap B] = 0$
2. $P[A] > 0$, $P[B] > 0$ implies $P[A]P[B] > 0$
3. Therefore $P[A \cap B] \neq P[A]P[B]$ and the claim follows

Intuition: If we know we flipped heads on a coin, that tells us we did not flip tails.

Bayes' theorem

Derivation from definition of conditional probabilities:

$$P[A|B] = \frac{P[A \cap B]}{P[B]}, \tag{19}$$

$$P[B|A] = \frac{P[A \cap B]}{P[A]} \tag{20}$$

Notice the numerators of the right sides are the same!

Rearrange first line into

$$P[A \cap B] = P[A|B]P[B] \tag{21}$$

and put it into the second line to get Bayes' theorem

$$\boxed{P[B|A] = \frac{P[A|B]P[B]}{P[A]}} \tag{22}$$

Intuition: Lets us reason about conditional probability of “flipped” events

Cancer test

Denote the events

- A: test says patient has cancer
- B: patient actually has cancer

Given information:

- Test has an accuracy of 95%
  - 95% of the time when the test says the patient has cancer, they actually do
  - 95% of the time when the test says the patient does not have cancer, they actually do not
- The cancer rate in the population is 0.5%

Question: The patient being tested for cancer cares about the chance they actually have cancer given the test says they do.
What is this probability?

Solution:
Translate given information into math:

$$P[A|B] = P[A^c|B^c] = 95\%, \qquad P[B] = 0.5\% \tag{23}$$

Use the law of total probability to find $P[A]$, the probability of the test saying a patient has cancer:

$$\begin{aligned} P[A] &= P[A|B]P[B] + P[A|B^c]P[B^c] && (24) \\ &= (95\%)(0.5\%) + (100\% - 95\%)(100\% - 0.5\%) && (25) \\ &= 5.45\% && (26) \end{aligned}$$

Now use Bayes' theorem:

$$P[B|A] = \frac{P[A|B]P[B]}{P[A]} = \frac{(95\%)(0.5\%)}{5.45\%} \approx \boxed{8.72\%} \tag{27}$$

How do we resolve this counter-intuitive result?

Even though the test is highly accurate (95%), the chance of actually having cancer is low (8.72%), despite a positive test result. This is because the base rate of cancer is very small, only 0.5%.

On the other hand, conditioning on a positive test result makes the chance of cancer increase dramatically in a relative sense from 0.5% to 8.72%.

From the standpoint of the designer of the cancer test, the smaller the base rate of cancer, the more accurate the test has to be to yield the same probability of a patient actually having cancer.

Homework P1-1:
Consider the previous example. Compute the probability that a patient has cancer, given a negative test result.

Homework P1-2: (1.33 in [1])
A large class in probability theory is taking a multiple-choice test. For a particular question on the test, the fraction of examinees who know the answer is p; $1-p$ is the fraction that will guess. The probability of answering a question correctly is unity for an examinee who knows the answer and $1/m$ for a guessee; m is the number of multiple-choice alternatives.

1. Compute the probability that an examinee knew the answer to a question given that he or she has correctly answered it in terms of m and p.
2. Then evaluate this probability for the specific choice $m = 4$ and $p = 50\%$.

Homework P1-3: (1.35 in [1])
Assume there are three machines A, B, and C in a semiconductor manufacturing facility that make chips. They manufacture, respectively, 25, 35, and 40 percent of the total semiconductor chips there. Of their outputs, respectively, 5, 4, and 2 percent of the chips are defective. A chip is drawn randomly from the combined output of the three machines and is found defective. What is the probability that this defective chip was manufactured by machine A? by machine B? by machine C?

Homework P1-4: (1.55 in [1])
An automatic breathing apparatus (B) used in anesthesia fails with probability P_B. A failure means death to the patient unless a monitor system (M) detects the failure and alerts the physician. The monitor system fails with probability P_M. The failures of the system components are independent events. Professor X, an M.D. at Hevardi Medical School, argues that if $P_M > P_B$ installation of M is useless. Compute the probability of a patient dying with and without the monitor system in place. Take $P_M = 0.1 = 2P_B$. Is Professor X correct in his assessment?

[1] John Woods and Henry Stark.
Probability, Statistics, and Random Processes for Engineers.
Pearson Higher Ed, 4 edition, 2011.