Probability basics

Last modified: 2026-09-28 23:45:35 (PDT)

Defining probabilities

Definition 0.1 (Sample space)**.** The **sample space** of a random experiment, denoted Ω, is the set of all of its possible outcomes.

Example 0.1 (Sample space of a die roll)**.** Rolling a six-sided die once has sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Definition 0.2 (Event)**.** An **event** is a subset of the sample space Ω: a set of outcomes.

Remark. When Ω is uncountable, such as an interval of real numbers, only the subsets in a designated σ-algebra $\mathcal{S}$ on Ω count as events; every set that arises in these notes is in $\mathcal{S}$. When Ω is finite or countably infinite, every subset of Ω can be an event, because the collection of all subsets of Ω is a σ-algebra (all subsets form a σ-algebra), as in σ-algebras for die rolls.

Example 0.2 (Rolling an even number)**.** In Example 0.1, the event "the roll is even" is $A = \{2, 4, 6\}$.

Definition 0.3 (Occurrence of an event)**.** An event A **occurs** when the outcome of the experiment is in A, and does not occur when the outcome is not in A.

Example 0.3 (Rolling a 4)**.** In Example 0.2, if the die shows 4, the event "the roll is even", $\{2, 4, 6\}$, occurs, because $4 \in \{2, 4, 6\}$, and the event "the roll is at most 2", $\{1, 2\}$, does not occur.

Definition 0.4 (Complement of an event)**.** The **complement** of an event A, denoted $\neg A$, is the event that A does not occur: the set of outcomes in the sample space Ω that are not in A.

$$\neg A \overset{\text{def}}{=} \Omega \setminus A$$

Remark. Other sources write A^c or $\bar{A}$.

Example 0.4 (Rolling an odd number)**.** In Example 0.2, the complement of the event "the roll is even", $A = \{2, 4, 6\}$, is the event "the roll is odd", $\neg A = \{1, 3, 5\}$.

Definition 0.5 (Mutually exclusive events)**.** Finitely or countably many sets $A_1, A_2, \ldots$, such as events, are **mutually exclusive** (also called **disjoint**, **pairwise disjoint**, or **mutually disjoint**) when no two of them share an element:

$$A_i \cap A_j = \emptyset \quad \text{for all } i \neq j$$

Remark. The math notes define the same property for sets in general, as pairwise disjoint.

Example 0.5 (Low and high rolls)**.** In Example 0.1, the events "the roll is at most 2", $\{1, 2\}$, and "the roll is at least 5", $\{5, 6\}$, are mutually exclusive: no roll is in both. The events "the roll is even", $\{2, 4, 6\}$, and "the roll is at most 2", $\{1, 2\}$, are not mutually exclusive: a roll of 2 is in both.

Definition 0.6 (Partition of an event)**.** A **partition** of an event A is a finite or countably infinite collection of events $A_1, A_2, \ldots$ that satisfies:

- The events $A_1, A_2, \ldots$ are mutually exclusive.
- Their union is A: $\bigcup_i A_i = A$.

A partition of the sample space is the case $A = \Omega$.

Remark. Some sources also require each piece A_i to contain at least one outcome.

Example 0.6 (Partitioning the die roll)**.** In Example 0.1, the events "the roll is even", $\{2, 4, 6\}$, and "the roll is odd", $\{1, 3, 5\}$, are mutually exclusive, and their union is Ω, so they partition the sample space. So do the six single-outcome events $\{1\}, \{2\}, \ldots, \{6\}$.

The events $\{1,2\}$ and $\{5,6\}$ from Example 0.5 do not partition Ω: their union leaves out 3 and 4.

Definition 0.7 (Probability measure)**.** A **probability measure** on a sample space Ω, often denoted $\Pr()$ or $\mathrm{P}()$, is a measure on the events of Ω that gives the whole sample space probability 1:

$$\Pr(\Omega) = 1$$

Example 0.7 (Probability measure for a fair die)**.** For the die roll in Example 0.1, define $\Pr(A) \overset{\text{def}}{=} |A|/6$, where $|A|$ is the number of outcomes in A. This function is $1/6$ times the counting measure $\mu(A) = |A|$ on the die rolls (counting elements is a measure), so it is a measure on the events of Ω, and it gives $\Pr(\Omega) = 6/6 = 1$. The event "the roll is even" from Example 0.2 has probability $\Pr(\{2,4,6\}) = 3/6 = 1/2$.

Corollary 0.1 (Probability measures are finitely additive)**.** *Every probability measure is finitely additive: for any mutually exclusive events $A_1, \ldots, A_n$,*

$$\Pr(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \Pr(A_i)$$

Proof

Proof. A probability measure is a measure (Definition 0.7), so measures are finitely additive applies to it. □

Theorem 0.1 (Continuity of probability)**.** *If $A_1 \supseteq A_2 \supseteq \cdots$ are events with $\bigcap_{i=1}^{\infty} A_i = \emptyset$, then:*

$$\lim_{n\to\infty} \Pr(A_n) = 0$$

Proof

Proof. For each i, let $B_i \overset{\text{def}}{=} A_i \setminus A_{i+1}$, the outcomes in A_i but not in A_{i+1}. Each $B_i = A_i \cap (\Omega \setminus A_{i+1})$ is an event, by closure properties of a σ-algebra.
The B_i are pairwise disjoint. For $i < j$, $B_j \subseteq A_j \subseteq A_{i+1}$, while B_i has no outcomes in A_{i+1}, so $B_i \cap B_j = \emptyset$.
$A_n = \bigcup_{i=n}^{\infty} B_i$ *for each n.* For $i \geq n$, $B_i \subseteq A_i \subseteq A_n$, so the union is contained in A_n.

Conversely, let $\omega \in A_n$. Since $\bigcap_i A_i = \emptyset$, ω is not in every A_i; since the A_i are nested, the indices i with $\omega \in A_i$ are $1, \ldots, m$ for some $m \geq n$. Then $\omega \in A_m$ and $\omega \notin A_{m+1}$, so $\omega \in B_m$.

$\Pr(A_1)$ *is finite.* The events A_1 and $\Omega \setminus A_1$ are disjoint with union Ω, so Corollary 0.1 applies to them:

$$\begin{aligned} \Pr(A_1) &\leq \Pr(A_1) + \Pr(\Omega \setminus A_1) && (\Pr(\Omega \setminus A_1) \geq 0) \\ &= \Pr(\Omega) && (\text{finite additivity of } \Pr) \\ &= 1 && (\text{definition of a probability measure}) \end{aligned}$$

The limit. For each n:

$$\begin{aligned} \Pr(A_n) &= \Pr\left(\bigcup_{i=n}^{\infty} B_i\right) && \left(A_n = \bigcup_{i=n}^{\infty} B_i\right) \\ &= \sum_{i=n}^{\infty} \Pr(B_i) && (\text{countable additivity; the } B_i \text{ are pairwise disjoint}) \end{aligned}$$

With $n = 1$, this says the series $\sum_{i=1}^{\infty} \Pr(B_i)$ has the finite total $\Pr(A_1)$. So, for $n \geq 2$:

$$\begin{aligned} \Pr(A_n) &= \sum_{i=1}^{\infty} \Pr(B_i) - \sum_{i=1}^{n-1} \Pr(B_i) && (\text{split off the first } n-1 \text{ terms; the total is finite}) \\ &= \Pr(A_1) - \sum_{i=1}^{n-1} \Pr(B_i) && (\text{the case } n = 1) \end{aligned}$$

As $n \to \infty$, the partial sums $\sum_{i=1}^{n-1} \Pr(B_i)$ converge to the total $\Pr(A_1)$, so $\Pr(A_n) \to \Pr(A_1) - \Pr(A_1) = 0$. □

Remark. Requiring countable additivity, not just finite additivity, enables results such as Theorem 0.1, and it is what makes sums over countably infinite partitions valid.

Theorem 0.2 (Kolmogorov axioms). *A function* $\Pr$ *that assigns a real number* $\Pr(A)$ *to each event* A *of a sample space* Ω *is a probability measure if and only if it satisfies:*

1. *For any event* A, $\Pr(A) \geq 0$.
2. *The probability of the whole sample space is 1:*
$$\Pr(\Omega) = 1$$
3. $\Pr$ *is countably additive: for any mutually exclusive events* $A_1, A_2, \ldots$,
$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$$

Proof

Proof. Suppose Pr is a probability measure. It is a measure, so its values lie in the extended non-negative reals $[0, \infty]$, which gives axiom 1, and it is countably additive, which is axiom 3. Axiom 2 is the condition $\Pr(\Omega) = 1$ in Definition 0.7.
Conversely, suppose Pr satisfies the three axioms. By axiom 1, Pr takes values in $[0, \infty]$, so axiom 3 says Pr is countably additive in the sense of the definition of countable additivity. Then, by countable additivity and the empty set, $\Pr(\emptyset)$ is 0 or ∞; $\Pr(\emptyset)$ is a real number, so $\Pr(\emptyset) = 0$. So Pr satisfies both conditions of the definition of a measure ($\Pr(\emptyset) = 0$ and countable additivity) and is a measure on the events of Ω. With axiom 2, Pr is a probability measure (Definition 0.7). □

Remark. Many sources define a probability measure by these three axioms instead; the theorem shows that the two definitions agree. The axioms are named for Kolmogorov, who introduced them in 1933 (see Wikipedia: Probability axioms). The axiom form does not list $\Pr(\emptyset) = 0$; the proof above derives it.

Theorem 0.3 (Probability of a subset's intersection)**.** *If A and B are events and $A \subseteq B$, then* $\Pr(A \cap B) = \Pr(A)$.

Proof

Proof. Since $A \subseteq B$, every outcome in A is also in B, so $A \cap B = A$, and therefore $\Pr(A \cap B) = \Pr(A)$. □

Theorem 0.4 (An event and its complement sum to 1)**.** *For any event A and its complement $\neg A$:*

$$\Pr(A) + \Pr(\neg A) = 1$$

Proof

Proof. The events A and $\neg A$ are disjoint, and their union is Ω.

$$\begin{aligned} \Pr(A) + \Pr(\neg A) &= \Pr(A \cup \neg A) && \text{(additivity of probability for disjoint events)} \\ &= \Pr(\Omega) && (A \cup \neg A = \Omega) \\ &= 1 && \text{(probability of the sample space is 1)} \end{aligned}$$

□

Corollary 0.2 (Complement rule)**.** *For any event A:*

$$\Pr(\neg A) = 1 - \Pr(A)$$

Proof

Proof. Subtract $\Pr(A)$ from both sides of Theorem 0.4. $\square$

Corollary 0.3 (Complement rule in probability (π) notation)**.** *If the probability of an event A is $\Pr(A) = \pi$, then the probability that A does not occur is:*

$$\Pr(\neg A) = 1 - \pi$$

Proof

Proof.

$$\begin{aligned} \Pr(\neg A) &= 1 - \Pr(A) && \text{(complement rule)} \\ &= 1 - \pi && \text{(substitute } \Pr(A) = \pi) \end{aligned}$$

$\square$

Example 0.8 (Probability of not rolling a six)**.** For a fair die, the event "the roll is a six" has probability $\pi = 1/6$, so by Corollary 0.3 the probability of not rolling a six is $1 - 1/6 = 5/6$.

Conditional probability

Definition 0.8 (Conditional probability)**.** For two events A and B with $\Pr(B) > 0$, the **conditional probability** of A given B, denoted $\Pr(A \mid B)$, is:

$$\Pr(A \mid B) \stackrel{\text{def}}{=} \frac{\Pr(A \cap B)}{\Pr(B)}$$

Example 0.9 (Rolling a six, given an even roll)**.** For a fair die (Example 0.7), let $A = \{6\}$ and $B = \{2, 4, 6\}$. Then $A \cap B = \{6\}$, so:

$$
\begin{aligned}
\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} && \text{(definition of conditional probability)} \\
&= \frac{1/6}{3/6} && \text{(substitute the probabilities)} \\
&= \frac{1}{3} && \text{(simplify)}
\end{aligned}
$$

Theorem 0.5 (Law of conditional probability)**.** *For any two events A and B with $\Pr(B) > 0$:*

$$\Pr(A \cap B) = \Pr(A \mid B) \cdot \Pr(B)$$

Proof

Proof. Rearranging Definition 0.8:

$$
\begin{aligned}
\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} && \text{(definition of conditional probability)} \\
\Pr(A \cap B) &= \Pr(A \mid B) \cdot \Pr(B) && \text{(multiply both sides by } \Pr(B))
\end{aligned}
$$

□

Example 0.10 (Applying the law of conditional probability)**.** Suppose 30% of adults exercise regularly ($\Pr(E) = 0.30$), and among adults who exercise regularly, 60% have low blood pressure ($\Pr(L \mid E) = 0.60$).
Then, by Theorem 0.5, the probability that a randomly selected adult both exercises regularly and has low blood pressure is:

$$
\begin{aligned}
\Pr(L \cap E) &= \Pr(L \mid E) \cdot \Pr(E) && \text{(law of conditional probability)} \\
&= 0.60 \cdot 0.30 && \text{(substitute the given values)} \\
&= 0.18 && \text{(multiply)}
\end{aligned}
$$

Theorem 0.6 (Law of total probability)**.** *If $B_1, B_2, \ldots$ is a partition of the sample space, with $\Pr(B_i) > 0$ for every i, then for any event A:*

$$\Pr(A) = \sum_i \Pr(A \mid B_i) \cdot \Pr(B_i)$$

ℹ Proof

Proof. Since $B_1, B_2, \ldots$ partition the sample space, the events $A \cap B_1, A \cap B_2, \ldots$ are mutually exclusive and their union is A. By countable additivity, and then by Theorem 0.5:

$$\begin{aligned}
\Pr(A) &= \sum_i \Pr(A \cap B_i) && \text{(countable additivity for partition of } A\text{)} \\
&= \sum_i \Pr(A \mid B_i) \cdot \Pr(B_i) && \text{(law of conditional probability; } \Pr(B_i) > 0\text{)}
\end{aligned}$$

□

Theorem 0.7 (Bayes' theorem)**.** *For any two events A and B with $\Pr(A) > 0$ and $\Pr(B) > 0$:*

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)}$$

ℹ Proof

Proof. By Definition 0.8 and Theorem 0.5:

$$\begin{aligned}
\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} && \text{(definition of conditional probability)} \\
&= \frac{\Pr(B \cap A)}{\Pr(B)} && \text{(intersection is commutative: } A \cap B = B \cap A\text{)} \\
&= \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)} && \text{(law of conditional probability applied to } \Pr(B \cap A)\text{)}
\end{aligned}$$

□

Example 0.11 (Positive predictive value of a medical test)**.** Suppose a disease test has 99% sensitivity and 99% specificity, and the prevalence of the disease in the population is 7%.
Let D be the event "person has the disease" and $+$ be the event "test is positive". Then:

- $\Pr(+ \mid D) = 0.99$ (sensitivity)
- $\Pr(\neg + \mid \neg D) = 0.99$ (specificity), so the false positive rate is $\Pr(+ \mid \neg D) = 1 - 0.99 = 0.01$
- $\Pr(D) = 0.07$ (prevalence)

By Theorem 0.7, with the denominator expanded by the law of total probability over the partition $\{D, \neg D\}$:

$$
\begin{aligned}
\Pr(D \mid +) &= \frac{\Pr(+ \mid D) \cdot \Pr(D)}{\Pr(+)} && \text{(Bayes' theorem)} \\
&= \frac{\Pr(+ \mid D) \cdot \Pr(D)}{\Pr(+ \mid D) \cdot \Pr(D) + \Pr(+ \mid \neg D) \cdot \Pr(\neg D)} && \text{(law of total probability)} \\
&= \frac{0.99 \cdot 0.07}{0.99 \cdot 0.07 + 0.01 \cdot 0.93} && \text{(substitute the given values)} \\
&= \frac{0.0693}{0.0693 + 0.0093} && \text{(multiply each term in the numerator and denomina} \\
&= \frac{0.0693}{0.0786} && \text{(add the denominator's two terms)} \\
&\approx 0.88 && \text{(divide)}
\end{aligned}
$$

Even with a highly accurate test (99% sensitive and 99% specific), only about 88% of people who test positive actually have the disease, because the disease prevalence is relatively low (7%).

Video lecture

Hutchinson's Probability Refresher (27 min) covers conditional distributions, the law of total probability, the chain rule of probability, and Bayes' rule (Hutchinson, n.d.). The login for the video site is posted on Canvas.

References

Hutchinson, Brian. n.d. *DATA 471/571 (Machine Learning) and CSCI 481/581 (Deep Learning) Video Lectures*. Western Washington University. Accessed September 28, 2026. https://facultyweb.cs.wwu.edu/~hutchib2/video_lectures/data371/.