Probability

A guided account of events, conditional reasoning, and random variables

Ratiomera Statistics

Table of contents

Course-Page Summary

Probability is the formal language for quantifying uncertainty. The frequentist interpretation uses long-run relative frequency, the classical definition applies to equally likely outcomes, and the subjective interpretation represents a degree of belief.

The complement, addition, and multiplication rules provide a consistent system for combining event probabilities. Conditional probability and Bayes' theorem describe how evidence changes probabilities, with direct applications to diagnostic testing and base-rate reasoning.

Random variables formalize numerical outcomes. Discrete variables use probability functions, while continuous variables use density functions and assign probability to areas over intervals. The binomial and normal distributions are two important models, and z-scores connect a normal distribution to a common reference scale.

Sampling distributions describe how statistics vary across repeated samples. The standard error of the mean decreases as sample size grows. Under independence and finite-variance conditions, the central limit theorem can justify a normal approximation for sufficiently large samples. How large is sufficient depends on the population shape and the analysis. Greater precision also does not guarantee representativeness, because undercoverage, self-selection, and nonresponse can bias even a very large sample.

Assumptions and Practical Checks

Table 1: Assumptions and checks for the main probability tools.

Model or calculation	What must be true	Practical check
Classical probability	The listed outcomes are genuinely equally likely	Justify equal likelihood from the design, not convenience
Conditional probability	The conditioning event has positive probability	State the restricted group and use its size as the denominator
Binomial distribution	Fixed n, two outcomes per trial, constant π, and independent trials	Check the design before applying the formula
Normal model	The distribution being modeled is sufficiently well approximated by a normal curve	Inspect shape, boundaries, mixtures, and unusual observations
Standard error and central limit theorem	Independent sampling or an appropriate design, finite variance, and enough information for the approximation	Examine the sampling process and population shape, not only sample size

What You Can Do Now

You can now translate ordinary uncertainty questions into event notation, select and apply the relevant probability rule, work with discrete and continuous probability models, standardize normal values, and explain what sampling distributions and standard errors say about repeated-sample variation. You can also state the assumptions behind these tools, recognize when a calculation is

only an approximation, and communicate why a precise estimate can still be biased by the way the sample was selected.

Definitions and Core Concepts

Table 2: Core probability and sampling concepts.

Concept	Definition
Sample space (Ω)	The set of all possible outcomes of a random experiment
Event (A)	Any subset of the sample space
Complement (A')	All outcomes in Ω not in A
Classical probability	$P(A) = m/n$, where m is the number of favorable outcomes and n is the total number of equally likely outcomes
Frequentist probability	Long-run relative frequency over many repetitions
Mutually exclusive	Two events that cannot both occur ($A \cap B = \emptyset$)
Independent	$P(A \cap B) = P(A)P(B)$: occurrence of one does not change the probability of the other
Population	All potentially observable units sharing the characteristic of interest
Sample	A subset drawn from the population
Simple random sample	Every possible sample of a fixed size has equal probability of selection
Sampling frame	The operational list or mechanism through which population units can be selected
Parameter	A fixed population quantity, such as μ or π, that is usually unknown
Statistic	A quantity computed from a sample, such as $\bar{x}$ or $\hat{p}$, that varies across samples
Selection bias	Systematic error caused when inclusion or response is related to the quantity being estimated

---

Probability Rules

Table 3: Probability rules and their uses.

Rule	Formula	When to use
Complement	$P(A') = 1 - P(A)$	Computing the probability of the opposite event
Addition (disjoint)	$P(A \cup B) = P(A) + P(B)$	Events cannot co-occur
Addition (general)	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	Events may co-occur
Many disjoint events	$P(A_1 \cup \cdots \cup A_k) = \sum P(A_i)$	All pairs of events are disjoint
Conditional probability	$P(A \mid B) = P(A \cap B)/P(B)$	Probability of A given B has occurred
Multiplication (general)	$P(A \cap B) = P(A \mid B) \cdot P(B)$	Finding the joint probability
Multiplication (independent)	$P(A \cap B) = P(A) \cdot P(B)$	When events are confirmed independent
Bayes' theorem	$P(A \mid B) = \frac{P(B\|A)\cdot P(A)}{P(B\|A)\cdot P(A)+P(B\|A')\cdot P(A')}$	Updating probability with new evidence

---

Random Variables and Distributions

Table 4: Discrete and continuous distribution objects.

	Discrete	Continuous
Values	Countable (0, 1, 2, …)	Any value in a range
Described by	Probability function $P(X = x)$	Density function $f(x)$
Expected value	$E(X) = \sum x \cdot P(X = x)$	$E(X) = \mu$
Variance	$\text{Var}\,(X) = \sum (x - E(X))^2 \cdot P(X = x)$	$\text{Var}\,(X) = \sigma^2$
CDF	$F(x) = P(X \leq x)$, sum up to x	$F(x) = P(X \leq x)$, area up to x
Key example	Binomial $B(n, \pi)$	Normal $N(\mu, \sigma^2)$

The Binomial Distribution

Table 5: Binomial model reference.

Property	Formula	Interpretation
Notation	$X \sim B(n, \pi)$	n trials, success probability π per trial
Probability function	$P(X = x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$	Probability of exactly x successes
Expected value	$E(X) = n\pi$	Long-run average number of successes
Variance	$\text{Var}\,(X) = n\pi(1-\pi)$	Spread of the distribution
Symmetric only when	$\pi = 0.5$	Otherwise skewed

The Normal Distribution and Sampling

Table 6: Normal-model and sampling-distribution reference.

Property	Formula or value	Interpretation
Normal distribution	$X \sim N(\mu, \sigma^2)$	Characterized by mean and variance
Empirical rule: $\mu \pm \sigma$	$\approx 68\%$	Most values near the center
Empirical rule: $\mu \pm 2\sigma$	$\approx 95\%$	Very few beyond two SDs
Empirical rule: $\mu \pm 3\sigma$	$\approx 99.7\%$	Extreme values very rare
Z-score	$z = (x - \mu)/\sigma$	Standardized distance from the mean
Standard normal	$Z \sim N(0, 1)$	Reference after z-transformation
Standard error	$\mathrm{SE} = \sigma/\sqrt{n}$	Precision of the sample mean
Central limit theorem	$\bar{X} \approx N(\mu, \sigma^2/n)$ for large n	Sample means are approximately normal

Expanded Reference

Purpose and foundations

Probability provides a language for situations in which an outcome is not known in advance. The **sample space**, usually written Ω, is the set of every possible outcome under consideration. An **event** is a collection of outcomes from that sample space. A single outcome either belongs to the event or does not. This set-based view matters because probability rules operate on events before they operate on numbers.

Imagine selecting one labeled tile from the numbers 1 through 10. The sample space is $\Omega = \{1, 2, ..., 10\}$. Event A might contain tiles 1, 2, and 3, while event D contains tiles 2, 3, and 7. A probability assigns each event a number from 0 to 1. Zero means the event cannot occur within the stated sample space. One means it must occur. Values between zero and one represent degrees of uncertainty under the model being used.

Probability can be approached through equally likely outcomes, long-run relative frequencies, or a stated model. In each approach, define the process and sample space before calculating. A probability is never detached from its conditions. The chance of an event may change when you learn new information, when the selection mechanism changes, or when the population under discussion changes.

Set idea	Notation	Meaning in words
Union	$A \cup D$	Outcomes in A, in D, or in both
Intersection	$A \cap D$	Outcomes that belong to both A and D
Complement	A^c	Outcomes in the sample space that are not in A
Disjoint events	$A \cap B = \varnothing$	Events with no shared outcome

Core ideas

The complement rule is useful when the opposite event is easier to count. The addition rule prevents the overlap of two events from being counted twice. Conditional probability narrows the reference set: $P(A \mid D)$ asks for the probability of A among situations in which D is known to have occurred. The vertical bar is read “given.” The denominator is therefore the probability of the condition, not the probability of the full sample space.

Independence has a precise meaning. Events A and D are independent when learning that D occurred does not change the probability of A. This is different from being disjoint. If two events are disjoint and one occurs, the other cannot occur, so that information changes its probability. Except in special zero-probability cases, disjoint events are not independent.

Bayes’ rule reverses a conditional probability. It combines the probability of a result under a condition with the prior frequency of that condition. Base rates matter: even a result that occurs more often in one group can correspond to a modest probability of group membership when that group is rare. Write every event in words before inserting numbers so the direction of the condition remains visible.

A **random variable** assigns a numerical value to each outcome of a chance process. A discrete random variable has separate countable values, such as the number of high-anxiety responses in a group. A continuous random variable can take values across an interval, such as a measured score. A discrete probability mass function assigns probability to individual values. A continuous probability density describes how probability is distributed across intervals; the probability of an interval is represented by area under the density curve.

Model or idea	What it describes	Main reading question
Binomial distribution	Number of successes in a fixed number of independent trials with constant success probability	Are the trial count, two outcomes, independence, and constant probability defensible?
Normal distribution	A symmetric bell-shaped model described by a mean and standard deviation	Does the model fit the variable and the question being asked?
Sampling distribution	How a statistic varies across repeated samples from the same process	How much sample-to-sample uncertainty should be expected?
Expected value	Probability-weighted long-run center of a random variable	What average would emerge across many repetitions of the model?

A sampling distribution is not the distribution of individual scores. It is the distribution of a statistic, such as a sample mean, across hypothetical repeated samples. Its spread is measured by a **standard error**. Larger samples usually produce less variable sample means when the underlying process remains the same. This idea connects probability to confidence intervals and hypothesis tests.

Formula guide

For any event A, its complement contains all outcomes outside A. Their probabilities add to one:

$$P(A^c) = 1 - P(A)$$

For two events, add their probabilities and subtract their overlap once. The subtraction corrects the double counting created when both events are included in the first two terms:

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

If the events are disjoint, their intersection is empty and the overlap term is zero. Do not use the shortened disjoint rule until the sample space confirms that both events cannot occur together.

Conditional probability restricts attention to the condition D. It requires $P(D) > 0$:

$$P(A \mid D) = \frac{P(A \cap D)}{P(D)}$$

The multiplication rule follows from the same relationship. It also shows the condition needed for independence. If A and D are independent, then $P(A \mid D) = P(A)$ and their joint probability factors:

$$P(A \cap D) = P(A \mid D)P(D) = P(A)P(D)$$

Bayes' rule reverses the condition by using the joint event in a different order:

$$P(A \mid D) = \frac{P(D \mid A)P(A)}{P(D)}$$

When A and A^c cover all possibilities, the denominator can be built with the law of total probability:

$$P(D) = P(D \mid A)P(A) + P(D \mid A^c)P(A^c).$$

This denominator keeps the base rate visible. A natural-frequency table expresses the same update with counts and is often the safest way to distinguish sensitivity, false-positive probability, and the probability of A after observing D.

For a discrete random variable, the probability mass function is $p(x) = P(X = x)$ and the cumulative distribution function is

$$F(x) = P(X \leq x) = \sum_{u \leq x} p(u).$$

If the possible values are $x_1, ..., x_m$, its expected value and variance are

$$E(X) = \sum_{j=1}^{m} x_j p(x_j), \qquad Var(X) = \sum_{j=1}^{m} (x_j - E(X))^2 p(x_j).$$

The expected value is the long-run balance point of the probability model, not a promise that one observation will equal it.

For a continuous random variable with density f, probability is the area under the density over an interval. The cumulative distribution function gives that interval probability without requiring calculus notation:

$$P(a < X \leq b) = F(b) - F(a).$$

A density height is not itself a probability, and $P(X = x) = 0$ for one exact point in a continuous model.

For a binomial random variable X with n trials and success probability p, the probability of exactly k successes is:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

The coefficient $\binom{n}{k}$ counts how many trial sequences contain exactly k successes. Use this model only after checking its assumptions, not because the outcome happens to be a count.

The binomial model also gives

$$E(X) = np, \qquad Var(X) = np(1 - p).$$

An upper tail such as $P(X > k)$ can be calculated through its complement, $1 - P(X \leq k)$. The model requires a fixed number of trials, two outcomes per trial, constant p, and independent trials.

For a normal variable $X \sim N(\mu, \sigma^2)$, standardize a boundary with

$$Z = \frac{X - \mu}{\sigma}.$$

Lower tails use $P(X \leq x) = \Phi(z)$, upper tails use $1 - \Phi(z)$, and an interval subtracts two cumulative areas. An inverse question begins with a cumulative probability q, finds $z_q = \Phi^{-1}(q)$, and returns to the original scale through $x_q = \mu + z_q \sigma$.

For independent observations with population mean μ and variance σ^2, the sampling distribution of the sample mean satisfies

$$E(\bar{X}) = \mu, \qquad Var(\bar{X}) = \frac{\sigma^2}{n}, \qquad SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

When σ is unknown, $s/\sqrt{n}$ estimates the standard error. If the population is normal, the sample mean is exactly normally distributed. For suitable nonnormal populations, its distribution can become approximately normal as n grows. The adequacy of that approximation depends on the population shape; no single sample-size cutoff guarantees it.

Object	What varies	Spread to report
Individual-value distribution	Individual observations	Population or sample standard deviation
Sampling distribution of $\bar{X}$	Sample means across repeated samples	Standard error $\sigma/\sqrt{n}$ or estimate $s/\sqrt{n}$
Biased achieved sample	Cases admitted by a defective frame or response process	A smaller standard error does not repair selection bias

Reading the explanatory figure

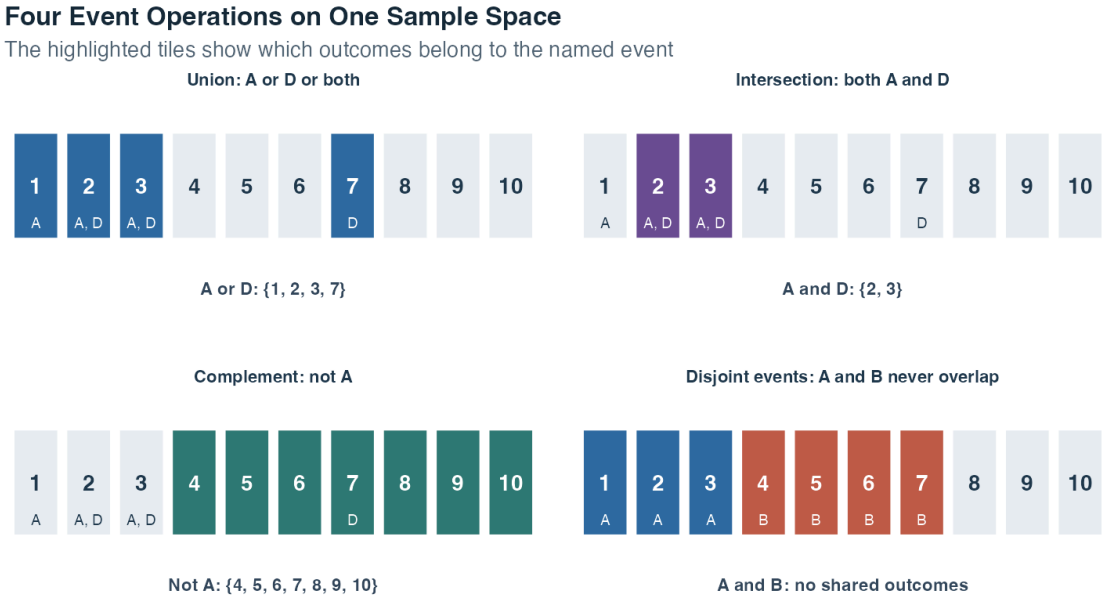

Figure 1: Four panels use numbered tiles to show union, intersection, complement, and disjoint events within the same ten-outcome sample space.

Begin with the upper-left panel. Event A contains 1, 2, and 3, while event D contains 2, 3, and 7. Their union highlights 1, 2, 3, and 7 because “or” includes outcomes that appear in either event and outcomes shared by both. The upper-right panel shows the intersection. Only 2 and 3 are highlighted because those are the shared outcomes.

The lower-left panel shows the complement of A. Tiles 4 through 10 are highlighted because every outcome in the sample space must be either in A or outside A. The lower-right panel introduces event B, containing 4, 5, 6, and 7. A and B share no tiles, so they are disjoint. The colors encode membership, not probability size. If the ten tiles were equally likely, a highlighted count could be divided by ten. If the outcomes were not equally likely, counting tiles would not be enough and their assigned probabilities would be needed.

This figure helps you check notation before using a formula. First identify the relevant tiles, then translate that set into a probability. That order reduces common errors such as treating “or” as exclusive, forgetting the overlap in an addition rule, or confusing disjointness with independence.

Interpretation checklist

Define the random process, the sample space, and each event in words. Check whether outcomes are equally likely before using counts. For conditional probability, name the condition and use it as the reference group. Draw a set, table, or probability tree when the direction of a condition feels

uncertain. Distinguish disjoint events from independent events. For a random variable, state whether it is discrete or continuous and identify what one value represents.

When choosing a distribution, match its assumptions to the process. Report whether a probability came from a theoretical model, an observed relative frequency, or a simulation. Remember that a simulation approximates the consequences of its stated rules; it does not repair unsuitable assumptions. For a sampling distribution, keep individual observations separate from sample statistics and state what would vary across repeated samples.

How this topic connects

Descriptive statistics summarized the dataset that was observed. Probability now describes how results could vary under a chance process. Statistical inference brings these ideas together: it uses the sampling distribution of a statistic to judge how compatible an observed result is with a population claim and to quantify uncertainty around an estimate.

Later models also rely on probability. A correlation or regression coefficient changes from sample to sample. Confidence intervals and tests describe that variability with probability models. Analysis of variance compares systematic and residual variation through an F distribution. The notation becomes richer, but the core questions remain familiar: What are the possible outcomes, what conditions are being assumed, and which uncertainty does the probability represent?

Document ID: `topic-02-probability-summary-en`

Notes 1 of 3

Probability | Topic 2

Notes 2 of 3

Probability | Topic 2

Notes 3 of 3

Probability | Topic 2