

Unit 2: Probability, Random Variables, and Probability Distributions

Unit 2 is where statistics gets its mathematical backbone. It starts with describing the relationship between two categorical variables, then builds the rules of probability from scratch, and ends with the two distributions that run the rest of the course: the binomial and the normal.

AP Statistics

Probability and Random Variables

About 14 minutes to read

How to use this guide

Read it in order the first time. The unit builds from describing data (topics 2.1 to 2.2) to modeling chance (2.3 to 2.7) to the formal machinery of random variables and their distributions (2.8 to 2.12). The probability rules in 2.5 to 2.7 are the ones exam questions test most often, and the normal distribution in 2.11 returns in every later unit.

After the first read, use the trap boxes to review the distinctions that cost the most points, especially independent versus mutually exclusive. Finish with the practice questions, then complete the recall check on the last page out loud and note any items you cannot explain yet.

What this unit is worth. Probability, random variables, and probability distributions make up about 10 to 20 percent of the AP Statistics exam. More important than the percentage is that nearly every later unit assumes you can read a two-way table, apply the probability rules, and work with normal distributions without thinking about it.

2.1 Tabular and Graphical Representations for the Distributions of Two Categorical Variables

A **two-way table**, also called a contingency table, summarizes data for two categorical variables at once. Each cell holds a count (or a proportion) of observations that fall into one combination of categories, like the number of students who are both juniors and in band. The rows show one variable and the columns show the other.

To see the relationship visually, three graphs do slightly different jobs. A **side-by-side bar chart** places the bars for each category of one variable next to each other within each category of the other, so you can compare heights directly. A **segmented bar chart** stacks the bars for one variable inside a single bar per category of the other, so each bar shows the full distribution for that category. A **mosaic plot** works like a segmented bar chart, except each rectangle's area is proportional to its cell frequency, so the tile sizes themselves show the joint distribution.

Two categorical variables are **associated** when the relationship of one variable differs across the levels of the other. If the conditional distribution of band membership looks the same for juniors and seniors, there is no association. If juniors are in band at a much higher rate than seniors, the variables are associated.

Trap. Association is about comparing across categories, not about the size of any single count. A cell with 200 students tells you nothing about association until you compare it to the relevant row or column totals.

2.2 Summary Statistics for Two Categorical Variables

Three relative frequencies answer three different questions about a two-way table. The **joint relative frequency** is a cell frequency divided by the grand total, so it answers "what proportion of everyone is in this cell?" The **marginal relative frequency** is a row or column total divided by the grand total, so it answers "what proportion of everyone is in this row or column?" The **conditional relative frequency** restricts to one category first: a cell frequency divided by its row total (or column total), answering "among just the juniors, what proportion are in band?"

Conditional relative frequency is the one that detects association. Compute the conditional distribution for each level of the explanatory variable and compare them. If they match, the variables are independent. If they differ, the variables are associated.

Trap. Joint and conditional relative frequencies answer different questions, and mixing them up is one of the most common errors in this unit. A joint frequency of 0.12 means 12 percent of all students are junior band members. A conditional frequency of 0.40 means 40 percent of juniors are in band. Always say which denominator you used.

2.3 Estimating Probabilities Using Simulation

A **random process** generates results determined by chance. One performance of it is a **trial**, the result of a trial is an **outcome**, and an **event** is a collection of outcomes. Rolling a die once is a trial. The outcome might be 4. The event "rolling an even number" is the collection $\{2, 4, 6\}$.

A **simulation** models a random process by assigning every possible outcome a value determined by chance, then running many trials and recording the counts. It is the practical answer when the exact probability is hard to compute. The **law of large numbers** is what makes simulation trustworthy: as the number of independent trials grows, the long-run relative frequency of an event settles toward a single value. The **empirical probability** is that observed relative frequency, and it estimates the true probability.

Trap. Simulation needs a large number of trials to be useful. Fifty simulated trials will not give a stable estimate. The law of large numbers is a statement about the long run, not about what happens in the next ten trials.

2.4 Introduction to Probability

The **sample space** is the set of all possible nonoverlapping outcomes, and its total probability is 1.

Theoretical probability applies when every outcome in the sample space is equally likely: the chance of event E is the number of outcomes in E divided by the total number of outcomes. Every probability is between 0 and 1, inclusive.

The **complement of an event**, written E' , $\bar{}$, or E^c , is the event that E does not occur. Its probability is $P(E') = 1 - P(E)$. Whenever a question asks for "at least one" or the direct computation looks messy, check whether the complement is easier.

Trap. The complement shortcut only works when you actually want the opposite of the event. "At least one success" complements to "no successes," which is easy. "Exactly two successes" complements to "anything except exactly two," which is not helpful. Check the complement before you commit to it.

2.5 Mutually Exclusive Events

The **joint probability** $P(A \cap B)$ is the probability that both events occur. Two events are **mutually exclusive**, also called disjoint, when they cannot occur at the same time, in which case $P(A \cap B) = 0$. Drawing a single card cannot be both a heart and a spade. Note that this is a statement about the events themselves, not about the person observing them.

2.6 Conditional Probability

Conditional probability, written $P(A | B)$, is the probability that A occurs given that B has occurred. It is computed as $P(A \cap B) / P(B)$. The condition restricts the sample space to just the outcomes where B happened, and the question becomes what fraction of those also have A.

The **general multiplication rule** rearranges that definition: $P(A \cap B) = P(A) \cdot P(B | A)$. It gives the joint probability whenever you know one event's probability and the conditional probability of the other given the first.

Trap. $P(A | B)$ and $P(B | A)$ are not the same number. The probability of having the disease given a positive test is very different from the probability of testing positive given the disease. Read the order of the condition carefully every time.

2.7 Independent Events and Unions of Events

Events A and B are **independent** if knowing whether A occurred does not change the probability that B occurs. The checks are equivalent: $P(A | B) = P(A)$, $P(B | A) = P(B)$, and $P(A \cap B) = P(A) \cdot P(B)$. Any one of them establishes independence.

The **union of events**, written $P(A \cup B)$, is the probability that A or B (or both) occurs. The **general addition rule** is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. The subtraction corrects for double counting the overlap. Only when the events are mutually exclusive does the overlap vanish and the rule simplify to $P(A \cup B) = P(A) + P(B)$.

Trap. Independent and mutually exclusive are not synonyms, and students mix them up constantly. Mutually exclusive means the events cannot both happen, so $P(A \cap B) = 0$. Independent means knowing one tells you nothing about the other, so $P(A \cap B) = P(A) \cdot P(B)$. In fact, two events with positive probability cannot be both independent and mutually exclusive. If $P(A)$ and $P(B)$ are both nonzero, independence forces $P(A \cap B) = P(A) \cdot P(B) > 0$, which contradicts mutual exclusivity. When a question says "independent," reach for multiplication. When it says "mutually exclusive" or "disjoint," the joint probability is zero.

2.8 Introduction to Random Variables and Probability Distributions

A **random variable** is a variable whose numerical values result from a random phenomenon, like the number of heads in three coin flips. Its **probability distribution** lists every possible value and the chance of each, and those chances sum to 1. It can be shown as a table, a graph, or a function, and it can be worked out exactly or estimated with a simulation.

A **cumulative probability distribution** shows, for each value, the chance of being less than or equal to that value. It is what you use when the question asks for "at most" or "no more than."

2.9 Parameters of Random Variables

A **parameter** of a probability distribution is a single fixed numerical value describing a characteristic of the distribution, such as its center or spread. For a **discrete random variable**, which takes only countable or finite values, the two main parameters have formulas you should know.

The **expected value**, written $E(X)$ or μ_X , is the mean of the random variable: $\mu_X = \sum x_i \cdot P(x_i)$. Interpret it as the long-run average outcome if you repeated the process many times. The **standard deviation**, written $SD(X)$ or σ_X , is $\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 \cdot P(x_i)}$, interpreted as the typical deviation from the mean over the long run. The **variance**, written $V(X)$ or σ_X^2 , is simply the square of the standard deviation.

Trap. Expected value is a long-run average, not a prediction about the next trial. If $E(X) = 2.4$ customers, no single trial produces 2.4 customers. It is the average you would approach over many repetitions.

2.10 The Binomial Distribution

A **binomial random variable** counts the number of successes in n repeated independent trials, where each trial has only two outcomes, success or failure, with success probability p and failure probability $1 - p$. Before using any binomial formula, check that all four conditions hold: a fixed number of trials, independent trials, two outcomes per trial, and a constant p .

The parameters have short formulas. The **mean of a binomial distribution** is $\mu_X = np$. The **standard deviation of a binomial distribution** is $\sigma_X = \sqrt{np(1 - p)}$. The **binomial probability function** gives the chance of exactly x successes: $P(X = x) = C(n, x) \cdot p^x \cdot (1 - p)^{n-x}$, for $x = 0, 1, 2, \dots, n$.

For example, with $n = 10$ free throws and $p = 0.7$, the mean number made is $10 \cdot 0.7 = 7$, the standard deviation is $\sqrt{10 \cdot 0.7 \cdot 0.3} = \sqrt{2.1} \approx 1.45$, and $P(X = 8) = C(10, 8) \cdot 0.7^8 \cdot 0.3^2 \approx 0.233$.

Trap. The binomial formulas assume a fixed n and a constant p . If the trials are not independent, like drawing cards without replacement from a small deck, or if p changes between trials, the binomial model does not apply. State the conditions and check them before computing.

2.11 The Normal Distribution

A **continuous random variable** can take any value in an interval, and probabilities come from areas under its density curve. The **normal distribution** is the most important one: a continuous, unimodal, bell-shaped, symmetric curve identified by its mean μ and standard deviation σ . A smaller σ makes the curve taller and more concentrated around the mean. A larger σ makes it shorter and more spread out.

The **standard normal distribution** is the normal distribution with $\mu = 0$ and $\sigma = 1$. Any normal value converts to a z -score by $z = (x - \mu) / \sigma$, which measures how many standard deviations the value sits from the mean. For roughly normal data, the **empirical rule** estimates that about 68% of observations fall within 1 standard deviation of the mean, about 95% within 2, and about 99.7% within 3.

Normal probability as area under the curve means that $P(a < X < b)$ is the area under the normal curve between a and b , with total area 1. Going the other direction, **determining interval boundaries from a given area** uses z -scores and a standard normal table (or technology): find the z -value that captures the desired area, then convert back with $x = \mu + z\sigma$.

Trap. The empirical rule only applies to approximately normal distributions. Do not use 68-95-99.7 on skewed data or on a distribution whose shape you have not checked. And a z -score is unitless, so a z of 2 means the same relative position whether the data are test scores or heights.

2.12 Sampling Distributions and the Central Limit Theorem

A **sampling distribution** is the distribution of a statistic over all possible samples of a given size from a population. It describes how the statistic varies from sample to sample, which is the foundation of inference in Units 3 through 5.

The **central limit theorem** says the sampling distribution of a sample mean can be approximated by a normal distribution, and the approximation improves as the sample size grows. This is what justifies using normal-based methods on means even when the underlying population is not normal.

When the exact sampling distribution is out of reach, simulation fills the gap. A **randomization distribution** comes from repeatedly reallocating response values to treatment groups at random and recording the statistic each time, which approximates the sampling distribution under the assumption of no treatment effect. **Simulating a sampling distribution** more generally means drawing many random samples from a population with assumed parameter values, recording the statistic for each, and using the resulting distribution as the approximation.

Trap. The central limit theorem is about the distribution of the sample mean, not about the population or a single sample. A skewed population does not become normal as n grows. What becomes approximately normal is the distribution of $\bar{x}$ across many samples.

Confusions That Cost Points

PAIR	HOW TO KEEP THEM STRAIGHT
Independent vs mutually exclusive	Independent: $P(A \cap B) = P(A) \cdot P(B)$, knowing one tells you nothing about the other. Mutually exclusive: $P(A \cap B) = 0$, both cannot happen. Events with positive probability cannot be both.
Joint vs conditional relative frequency	Joint divides by the grand total. Conditional divides by the row or column total. Name the denominator before you compute.
$P(A \mid B)$ vs $P(B \mid A)$	The condition comes second in the notation and restricts the sample space. Swapping them answers a different question.
Addition rule vs multiplication rule	Addition is for "or": $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Multiplication is for "and": $P(A \cap B) = P(A) \cdot P(B \mid A)$.
Expected value vs observed value	Expected value is the long-run average $\mu_X = \sum x_i \cdot P(x_i)$. It need not be a value the variable can take.
Binomial vs normal	Binomial counts successes in a fixed number of trials and is discrete. Normal is continuous and symmetric, and its probabilities are areas under the curve.
Sampling distribution vs population distribution	The population distribution describes individuals. The sampling distribution describes a statistic across many samples. The CLT connects them for means.

Practice Questions

Original questions written for this guide in the style of the AP exam. Answers and explanations are on the next page, so complete the questions before checking them.

1. A two-way table shows 200 students cross-classified by grade level (junior, senior) and club membership (in a club, not in a club). Among juniors, 40% are in a club. Among seniors, 40% are in a club. Which conclusion is supported?
 - A. Grade level and club membership are associated
 - B. Grade level and club membership are independent
 - C. More juniors than seniors are in a club
 - D. The joint relative frequency of junior club members is 0.40
2. Events A and B satisfy $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.2$. Which statement is true?
 - A. A and B are mutually exclusive
 - B. A and B are independent
 - C. $P(A \cup B) = 0.9$
 - D. $P(A | B) = 0.5$
3. A basketball player makes 70% of free throws, with each attempt independent. In 10 attempts, what is the probability she makes exactly 8?
 - A. $C(10, 8) \cdot 0.7^8 \cdot 0.3^2$
 - B. 0.7^8
 - C. $8 / 10 \cdot 0.7$
 - D. $1 - C(10, 2) \cdot 0.3^2 \cdot 0.7^8$
4. SAT math scores are approximately normal with mean 520 and standard deviation 110. Using the empirical rule, about what percent of scores fall between 410 and 630?
 - A. 68%
 - B. 95%
 - C. 99.7%
 - D. 50%

Answer Key

- 1. B.** The conditional distributions match: 40% in a club for juniors and 40% for seniors. When the conditional distribution of one variable is the same across levels of the other, the variables are independent (no association). A claims association, which requires the conditionals to differ. C confuses a conditional percentage with a count; 40% of each group says nothing about which group contributes more members. D mistakes the conditional relative frequency (0.40 among juniors) for a joint relative frequency, which would divide by the grand total of 200.
- 2. B.** Check independence: $P(A) \cdot P(B) = 0.4 \cdot 0.5 = 0.2$, which equals $P(A \cap B)$, so the events are independent. A is wrong because mutually exclusive requires $P(A \cap B) = 0$. C applies the addition rule without subtracting the overlap; the correct union is $0.4 + 0.5 - 0.2 = 0.7$. D computes $P(A \mid B)$ as 0.5, but $P(A \mid B) = P(A \cap B) / P(B) = 0.2 / 0.5 = 0.4$, which equals $P(A)$, confirming independence.
- 3. A.** This is a binomial setting: fixed $n = 10$, independent attempts, two outcomes, constant $p = 0.7$. The probability of exactly 8 successes is $C(10, 8) \cdot 0.7^8 \cdot 0.3^2$. B omits the combinatorial factor and the failure probability, as if only 8 attempts mattered. C is not a probability computation at all. D computes the complement of the wrong event; the complement of "exactly 8" is not "exactly 2."
- 4. A.** The interval 410 to 630 is $\mu \pm 110$, exactly one standard deviation on each side of the mean ($520 - 110 = 410$, $520 + 110 = 630$). The empirical rule assigns about 68% to the ± 1 standard deviation band. B would require ± 2 standard deviations (300 to 740). C would require ± 3 . D confuses the symmetric two-sided interval with a one-sided probability.

One-Page Recall Check

- Explain what a two-way table shows and how a mosaic plot differs from a segmented bar chart.
- Define association between two categorical variables in terms of conditional distributions.
- Distinguish joint, marginal, and conditional relative frequency, and say which denominator each uses.
- Define trial, outcome, and event, and explain how simulation estimates a probability.
- State the law of large numbers and explain why it justifies simulation.
- Define the complement of an event and give its probability formula.
- State the general addition rule and say when it simplifies.
- State the general multiplication rule and the three equivalent checks for independence.
- Explain why two events with positive probability cannot be both independent and mutually exclusive.
- Define a random variable and its probability distribution, including what the probabilities sum to.
- Write the formulas for the mean and standard deviation of a discrete random variable and interpret each.
- State the four binomial conditions and the formulas for the binomial mean, standard deviation, and $P(X = x)$.
- Convert a normal value to a z-score and state the empirical rule percentages.
- Explain how to find a probability from a normal curve and how to find an interval boundary from a given area.
- State the central limit theorem and explain what becomes approximately normal.

Where to go next. Open the AP Statistics deck at rycal.web.app/apstats and drill the Unit 2 cards. The deck covers the terms in this guide, and its practice questions target the same traps named here. If you have a test date, add it in the Test Planner. You can also start your next review with a Brain Dump, then check what you missed against this guide.

Key terms for this unit

Two-way table (contingency table), Side-by-side bar chart, Segmented bar chart, Mosaic plot, Association (between categorical variables), Joint relative frequency, Marginal relative frequency, Conditional relative frequency, Random process, Trial, Outcome, Event, Simulation, Law of large numbers, Empirical probability, Sample space, Theoretical probability, Complement of an event, Joint probability, Mutually exclusive (disjoint) events, Conditional probability, General multiplication rule, Independent events, Union of events, General addition rule, Random variable, Probability distribution, Cumulative probability distribution, Parameter (of a probability distribution), Expected value (mean of a random variable), Standard deviation of a random variable, Variance of a random variable, Discrete random variable, Binomial random variable, Mean of a binomial distribution, Standard deviation of a binomial distribution, Binomial probability function, Continuous random variable, Normal distribution, Standard normal distribution, Empirical rule (68–95–99.7 rule), Normal probability as area under the curve, Determining interval boundaries from a given area (inverse normal), Sampling distribution, Randomization distribution, Central limit theorem (CLT), Simulating a sampling distribution.

About this guide. Written for Rycal and aligned to the College Board AP Statistics course framework, Unit 2. All questions and explanations are original Rycal writing. Rycal is independent and is not affiliated with or endorsed by the College Board.