AP STATISTICS

UNIT 4

Probability, Random Variables, and Probability Distributions

AP

10–20%
AP EXAM WEIGHTING

~18–20
CLASS PERIODS

Remember to go to **AP Classroom** to assign students the online **Personal Progress Check** for this unit.

Whether assigned as homework or completed in class, the **Personal Progress Check** provides each student with immediate feedback related to this unit's topics and skills.

Personal Progress Check 4

Multiple-choice: ~45 questions
Free-response: 2 questions

- Probability
- Investigative Task

UNIT
4

10–20% AP EXAM WEIGHTING

~18–20 CLASS PERIODS

Probability, Random Variables, and Probability Distributions

BIG IDEA 1
Variation and Distribution **VAR**

- How can an event be both random and predictable?

BIG IDEA 2
Patterns and Uncertainty **UNC**

- About how many rolls of a fair six-sided die would we anticipate it taking to get three 1s?

Developing Understanding

Probabilistic reasoning allows statisticians to quantify the likelihood of random events over the long run and to make statistical inferences. Simulations and concrete examples can help students to understand the abstract definitions and calculations of probability. This unit builds on understandings of simulated or empirical data distributions and fundamental principles of probability to represent, interpret, and calculate parameters for theoretical probability distributions for discrete random variables. Interpretations of probabilities and parameters associated with a probability distribution should use appropriate units and relate to the context of the situation.

Building Course Skills

2.B **3.A** **3.B** **4.B**

Probability is a notoriously difficult topic for students to grasp because it's difficult to conceptualize future outcomes in concrete ways. Before introducing new formulas, teachers can help students get an intuitive feel for why the formulas (and related notation) make sense. For example, the probability formulas for $P\left(A \text{ or } B\right)$ and for $P\left(A \mid B\right)$ can be presented intuitively with two-way tables. Simulations can also help students internalize what it means to quantify random behavior. To help students understand when to apply different probability rules, teachers can use explicit strategies such as matching verbal scenarios to their corresponding probability formulas.

Students frequently misinterpret probability distributions and parameters for random variables. Teachers can reinforce that a complete interpretation will include context and units. A common misconception later in the course is that every question involving probability requires a significance test. Students should practice making predictions and decisions based on probability alone to avoid this misconception early on. They should revisit these problems in later units to practice differentiating between inference and probability problems.

Preparing for the AP Exam

To help students prepare for the AP Exam, teachers can model showing all steps in probability calculations and expect students to do the same. Calculations on the AP Exam should include presentation of an appropriate expression that communicates the structure of the formula, substitution of relevant values extracted from the problem, and an answer. In **2017 FRQ 3**, for example, a student who writes "$P(G)=P(G \mid J)\cdot P(J)+P(G \mid K)\cdot P(K)=(0.2119)(0.7)+(0.8413)(0.3)=0.4007$" has communicated the products in the multiplication rule, the sum in the addition rule, and an understanding that the events are mutually exclusive—all components of a complete response. To avoid a common error, students who present the same work using a tree diagram should practice using probabilities in the diagram correctly. Students importing incorrect solutions from one part of a multipart question to solve another will not be penalized a second time, unless the subsequent result is not a reasonable value (like a probability less than 0 or greater than 1).

UNIT AT A GLANCE

Enduring Understanding	Topic	Skills	Class Periods
			~18–20 CLASS PERIODS
VAR-1	4.1 Introducing Statistics: Random and Non-Random Patterns?	**1.A** Identify the question to be answered or problem to be solved *(not assessed)*.	
UNC-2	4.2 Estimating Probabilities Using Simulation	**3.A** Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
VAR-4	4.3 Introduction to Probability	**3.A** Determine relative frequencies, proportions, or probabilities using simulation or calculations. **4.B** Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.4 Mutually Exclusive Events	**4.B** Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.5 Conditional Probability	**3.A** Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
	4.6 Independent Events and Unions of Events	**3.A** Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
VAR-5	4.7 Introduction to Random Variables and Probability Distributions	**2.B** Construct numerical or graphical representations of distributions. **4.B** Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.8 Mean and Standard Deviation of Random Variables	**3.B** Determine parameters for probability distributions. **4.B** Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.9 Combining Random Variables	**3.B** Determine parameters for probability distributions. **3.C** Describe probability distributions.	

continued on next page

UNIT AT A GLANCE *(cont'd)*

Enduring Understanding	Topic	Skills	Class Periods
			~18–20 CLASS PERIODS
UNC-3	**4.10** **Introduction to the Binomial Distribution**	**3.A** Determine relative frequencies, proportions, or probabilities using simulation or calculations.	
	4.11 **Parameters for a Binomial Distribution**	**3.B** Determine parameters for probability distributions. **4.B** Interpret statistical calculations and findings to assign meaning or assess a claim.	
	4.12 **The Geometric Distribution**	**3.A** Determine relative frequencies, proportions, or probabilities using simulation or calculations. **3.B** Determine parameters for probability distributions. **4.B** Interpret statistical calculations and findings to assign meaning or assess a claim.	

Go to **AP Classroom** to assign the **Personal Progress Check** for Unit 4.
Review the results in class to identify and address any student misunderstandings.

SAMPLE INSTRUCTIONAL ACTIVITIES

The sample activities on this page are optional and are offered to provide possible ways to incorporate various instructional approaches into the classroom. They were developed in partnership with teachers from the AP community to share ways that they approach teaching some of the topics in this unit. Please refer to the Instructional Approaches section beginning on p. 207 for more examples of activities and strategies.

Activity	Topic	Sample Activity
1	4.3 4.5 4.8	**Error Analysis** Using 2015 FRQ 3, provide students with several answers containing errors for each part. Provide some responses with incorrect notation, incorrect work, missing work, work that shows calculator commands only, an incorrect formula or approach, and an incorrect final answer. Then ask students to identify the errors.
2	4.3 4.5 4.6	**Think-Pair-Share** Provide students with a set of five probability questions: one for the complement rule, the conditional probability formula, the general multiplication rule, the multiplication rule for independent events, and the general addition rule. Ask students to individually identify the formula needed to solve each problem, without doing the final calculations. Then have them share their thoughts with a partner.
3	4.5 4.6	**Create Representations** Provide students with the scenario from 2018 FRQ 3. Ask them to create a tree diagram to organize the information in the problem. Then ask them to use the information in the problem to set up a hypothetical 100,000 table (to make the decimals easy to work with), such as the one below. Encourage students to try both representations when solving probability questions in the future.

	Multiple Birth	Single Birth	Total
Left handed	770	10,615	11,385
Right handed	2,730	85,885	88,615
Total	3,500	96,500	100,000

Activity	Topic	Sample Activity
4	4.10 4.12	**Odd One Out** After modeling an odd one out example, have students form groups of four and give each of them a description of either a binomial or a geometric random variable. Explain that three of their variables follow the same probability distribution and one is different. Have students work together in their groups to determine whose is the odd one out and explain why.
5	4.2 4.12	**Predict and Confirm** Ask students to consider couples who plan to continue having children until they have one girl and predict how many children they think these couples will have, on average. Then ask each student to perform 10 trials using a coin toss where Heads = Girl and Tails = Boy. A trial is finished once one girl is observed and the number of total children is recorded. Combine the class results and calculate the average. Confirm with the geometric mean formula once it is discussed.

TOPIC 4.1

Introducing Statistics: Random and Non-Random Patterns?

SKILL

Selecting Statistical Methods

1.A

Identify the question to be answered or problem to be solved.

Required Course Content

ENDURING UNDERSTANDING

VAR-1

Given that variation may be random or not, conclusions are uncertain.

LEARNING OBJECTIVE

VAR-1.F

Identify questions suggested by patterns in data. **[Skill 1.A]**

ESSENTIAL KNOWLEDGE

VAR-1.F.1

Patterns in data do not necessarily mean that variation is not random.

SKILL

Using Probability and Simulation

3.A

Determine relative frequencies, proportions, or probabilities using simulation or calculations.

AVAILABLE RESOURCES

- Classroom Resources >
  - **Graphing Calculator Simulations Simplified**
  - **Three Calculator Simulation Activities**

ILLUSTRATIVE EXAMPLES

An outcome:

- Rolling a particular value on a six-sided number cube is one of six possible outcomes.

An event:

- When rolling two six-sided number cubes, an event would be a sum of seven. The corresponding collection of outcomes would be (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1), where the ordered pairs indicate (face value on one cube, face value on the other cube).

TOPIC 4.2

Estimating Probabilities Using Simulation

Required Course Content

ENDURING UNDERSTANDING

UNC-2

Simulation allows us to anticipate patterns in data.

LEARNING OBJECTIVE

UNC-2.A

Estimate probabilities using simulation. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

UNC-2.A.1

A random process generates results that are determined by chance.

UNC-2.A.2

An outcome is the result of a trial of a random process.

UNC-2.A.3

An event is a collection of outcomes.

UNC-2.A.4

Simulation is a way to model random events, such that simulated outcomes closely match real-world outcomes. All possible outcomes are associated with a value to be determined by chance. Record the counts of simulated outcomes and the count total.

UNC-2.A.5

The relative frequency of an outcome or event in simulated or empirical data can be used to estimate the probability of that outcome or event.

UNC-2.A.6

The law of large numbers states that simulated (empirical) probabilities tend to get closer to the true probability as the number of trials increases.

TOPIC 4.3

Introduction to Probability

Required Course Content

ENDURING UNDERSTANDING

VAR-4

The likelihood of a random event can be quantified.

LEARNING OBJECTIVE

VAR-4.A

Calculate probabilities for events and their complements. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

VAR-4.A.1

The sample space of a random process is the set of all possible non-overlapping outcomes.

VAR-4.A.2

If all outcomes in the sample space are equally likely, then the probability an event E will occur is defined as the fraction:

$$\frac{\text{number of outcomes in event E}}{\text{total number of outcomes in sample space}}$$

VAR-4.A.3

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

VAR-4.A.4

The probability of the complement of an event E, E' or E^C, (i.e., not E) is equal to $1 - P(E)$.

VAR-4.B

Interpret probabilities for events. **[Skill 4.B]**

VAR-4.B.1

Probabilities of events in repeatable situations can be interpreted as the relative frequency with which the event will occur in the long run.

SKILLS

Using Probability and Simulation

3.A

Determine relative frequencies, proportions, or probabilities using simulation or calculations.

Statistical Argumentation

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

SKILL

Statistical Argumentation

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

TOPIC 4.4

Mutually Exclusive Events

Required Course Content

ENDURING UNDERSTANDING

VAR-4

The likelihood of a random event can be quantified.

LEARNING OBJECTIVE

VAR-4.C

Explain why two events are (or are not) mutually exclusive. **[Skill 4.B]**

ESSENTIAL KNOWLEDGE

VAR-4.C.1

The probability that events A and B both will occur, sometimes called the joint probability, is the probability of the intersection of A and B, denoted $P(A \cap B)$.

VAR-4.C.2

Two events are mutually exclusive or disjoint if they cannot occur at the same time. So $P(A \cap B) = 0$.

TOPIC 4.5

Conditional Probability

SKILL

Using Probability and Simulation

3.A

Determine relative frequencies, proportions, or probabilities using simulation or calculations.

Required Course Content

ENDURING UNDERSTANDING

VAR-4

The likelihood of a random event can be quantified.

LEARNING OBJECTIVE

VAR-4.D

Calculate conditional probabilities. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

VAR-4.D.1

The probability that event A will occur given that event B has occurred is called a conditional probability and denoted

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

VAR-4.D.2

The multiplication rule states that the probability that events A and B both will occur is equal to the probability that event A will occur multiplied by the probability that event B will occur, given that A has occurred. This is denoted $P(A \cap B) = P(A) \cdot P(B \mid A)$.

SKILL

Using Probability and Simulation

3.A

Determine relative frequencies, proportions, or probabilities using simulation or calculations.

TOPIC 4.6

Independent Events and Unions of Events

Required Course Content

ENDURING UNDERSTANDING

VAR-4

The likelihood of a random event can be quantified.

LEARNING OBJECTIVE

VAR-4.E

Calculate probabilities for independent events and for the union of two events. **[Skill 3.A]**

ESSENTIAL KNOWLEDGE

VAR-4.E.1

Events A and B are independent if, and only if, knowing whether event A has occurred (or will occur) does not change the probability that event B will occur.

VAR-4.E.2

If, and only if, events A and B are independent, then $P(A \mid B) = P(A)$, $P(B \mid A) = P(B)$, and $P(A \cap B) = P(A) \cdot P(B)$.

VAR-4.E.3

The probability that event A or event B (or both) will occur is the probability of the union of A and B, denoted $P(A \cup B)$.

VAR-4.E.4

The addition rule states that the probability that event A or event B or both will occur is equal to the probability that event A will occur plus the probability that event B will occur minus the probability that both events A and B will occur. This is denoted $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

TOPIC 4.7

Introduction to Random Variables and Probability Distributions

Required Course Content

ENDURING UNDERSTANDING

VAR-5

Probability distributions may be used to model variation in populations.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-5.A Represent the probability distribution for a discrete random variable. **[Skill 2.B]**	**VAR-5.A.1** The values of a random variable are the numerical outcomes of random behavior. **VAR-5.A.2** A discrete random variable is a variable that can only take a countable number of values. Each value has a probability associated with it. The sum of the probabilities over all of the possible values must be 1. **VAR-5.A.3** A probability distribution can be represented as a graph, table, or function showing the probabilities associated with values of a random variable. **VAR-5.A.4** A cumulative probability distribution can be represented as a table or function showing the probability of being less than or equal to each value of the random variable.
VAR-5.B Interpret a probability distribution. **[Skill 4.B]**	**VAR-5.B.1** An interpretation of a probability distribution provides information about the shape, center, and spread of a population and allows one to make conclusions about the population of interest.

SKILLS

Data Analysis

2.B

Construct numerical or graphical representations of distributions.

Statistical Argumentation

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

ILLUSTRATIVE EXAMPLES

Outcomes of trials of a random process:

- The sum of the outcomes for rolling two dice
- The number of puppies in a randomly selected litter for a certain breed of dog

SKILLS

Using Probability and Simulation

3.B

Determine parameters for probability distributions.

Statistical Argumentation

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

TOPIC 4.8

Mean and Standard Deviation of Random Variables

Required Course Content

ENDURING UNDERSTANDING

VAR-5

Probability distributions may be used to model variation in populations.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-5.C Calculate parameters for a discrete random variable. **[Skill 3.B]**	**VAR-5.C.1** A numerical value measuring a characteristic of a population or the distribution of a random variable is known as a parameter, which is a single, fixed value. **VAR-5.C.2** The mean, or expected value, for a discrete random variable X is $\mu_X = \sum x_i \cdot P(x_i)$. **VAR-5.C.3** The standard deviation for a discrete random variable X is $\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 \cdot P(x_i)}$.
VAR-5.D Interpret parameters for a discrete random variable. **[Skill 4.B]**	**VAR-5.D.1** Parameters for a discrete random variable should be interpreted using appropriate units and within the context of a specific population.

TOPIC 4.9

Combining Random Variables

SKILLS

Using Probability and Simulation

3.B

Determine parameters for probability distributions.

3.C

Describe probability distributions.

Required Course Content

ENDURING UNDERSTANDING

VAR-5

Probability distributions may be used to model variation in populations.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
VAR-5.E Calculate parameters for linear combinations of random variables. **[Skill 3.B]**	**VAR-5.E.1** For random variables X and Y and real numbers a and b, the mean of $aX + bY$ is $a\mu_x + b\mu_y$. **VAR-5.E.2** Two random variables are independent if knowing information about one of them does not change the probability distribution of the other. **VAR-5.E.3** For independent random variables X and Y and real numbers a and b, the mean of $aX + bY$ is $a\mu_x + b\mu_y$, and the variance of $aX + bY$ is $a^2\sigma^2_x + b^2\sigma^2_y$.
VAR-5.F Describe the effects of linear transformations of parameters of random variables. **[Skill 3.C]**	**VAR-5.F.1** For $Y = a + bX$, the probability distribution of the transformed random variable, Y, has the same shape as the probability distribution for X, so long as $a > 0$ and $b > 0$. The mean of Y is $\mu_y = a + b\mu_x$. The standard deviation of Y is $\sigma_y = \lvert b \rvert \sigma_x$.

SKILL

Using Probability and Simulation

3.A

Determine relative frequencies, proportions, or probabilities using simulation or calculations.

TOPIC 4.10

Introduction to the Binomial Distribution

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.A Estimate probabilities of binomial random variables using data from a simulation. **[Skill 3.A]**	**UNC-3.A.1** A probability distribution can be constructed using the rules of probability or estimated with a simulation using random number generators. **UNC-3.A.2** A binomial random variable, X, counts the number of successes in n repeated independent trials, each trial having two possible outcomes (success or failure), with the probability of success p and the probability of failure $1-p$.
UNC-3.B Calculate probabilities for a binomial distribution. **[Skill 3.A]**	**UNC-3.B.1** The probability that a binomial random variable, X, has exactly x successes for n independent trials, when the probability of success is p, is calculated as $P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}, x=0, 1, 2, \ldots, n$. This is the binomial probability function.

TOPIC 4.11

Parameters for a Binomial Distribution

SKILLS

Using Probability and Simulation

3.B

Determine parameters for probability distributions.

Statistical Argumentation

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.C Calculate parameters for a binomial distribution. **[Skill 3.B]**	**UNC-3.C.1** If a random variable is binomial, its mean, μ_x, is np and its standard deviation, σ_x, is $\sqrt{np(1-p)}$.
UNC-3.D Interpret probabilities and parameters for a binomial distribution. **[Skill 4.B]**	**UNC-3.D.1** Probabilities and parameters for a binomial distribution should be interpreted using appropriate units and within the context of a specific population or situation.

TOPIC 4.12

The Geometric Distribution

SKILLS

Using Probability and Simulation

3.A

Determine relative frequencies, proportions, or probabilities using simulation or calculations.

3.B

Determine parameters for probability distributions.

Statistical Argumentation

4.B

Interpret statistical calculations and findings to assign meaning or assess a claim.

Required Course Content

ENDURING UNDERSTANDING

UNC-3

Probabilistic reasoning allows us to anticipate patterns in data.

LEARNING OBJECTIVE	ESSENTIAL KNOWLEDGE
UNC-3.E Calculate probabilities for geometric random variables. **[Skill 3.A]**	**UNC-3.E.1** For a sequence of independent trials, a geometric random variable, X, gives the number of the trial on which the first success occurs. Each trial has two possible outcomes (success or failure) with the probability of success p and the probability of failure $1-p$. **UNC-3.E.2** The probability that the first success for repeated independent trials with probability of success p occurs on trial x is calculated as $P(X=x)=(1-p)^{x-1}p, x=1,2,3,....$ This is the geometric probability function.
UNC-3.F Calculate parameters of a geometric distribution. **[Skill 3.B]**	**UNC-3.F.1** If a random variable is geometric, its mean, μ_x, is $\frac{1}{p}$ and its standard deviation, σ_x, is $\frac{\sqrt{(1-p)}}{p}$.
UNC-3.G Interpret probabilities and parameters for a geometric distribution. **[Skill 4.B]**	**UNC-3.G.1** Probabilities and parameters for a geometric distribution should be interpreted using appropriate units and within the context of a specific population or situation.

QUICK REFERENCE FOR NOTATION AND FORMULAS FOR PROBABILITY DISTRIBUTIONS

Distribution	Notes	Parameter(s)	Random Variable	Conditions	Mean for Distribution	Standard Deviation for Distribution
Probability distribution for a random variable	▪ Represent discrete random variables using frequency/ relative frequency tables or histograms ▪ Represent continuous random variables with density functions.	μ_x, σ_x	X	▪ All probabilities must be between 0 and 1. ▪ $\sum \text{Probabilities} = 1$.	$\mu_X = \sum x_i \cdot P(x_i)$ "expected value"	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 \cdot P(x_i)}$
Sum or difference of independent random variables	See Unit 5 for distributions of other linear transformations of random variables.	$\mu_X, \sigma_X, \mu_Y, \sigma_Y$	$X+Y$ or $X-Y$	To calculate the variance or standard deviation of the difference, the random variables must be independent.	$\mu_{X+Y} = \mu_X + \mu_Y$ or $\mu_{X-Y} = \mu_X - \mu_Y$	Variance, $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$ Variance, $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$
Binomial probability distribution	Binomial probability function: $P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$ $x = 0,1,2,3,\ldots,n.$	n and p	X	▪ n is predetermined. ▪ Binary ▪ Independent ▪ p is the same for each trial.	$\mu_x = np$	$\sigma_X = \sqrt{np(1-p)}$
Geometric probability distribution	Geometric probability formula: $P(X=x) = (1-p)^{x-1} p,$ $x = 1,2,3,\ldots.$	p	X	▪ n is not predetermined. ▪ Binary ▪ Independent ▪ p is the same for each repetition (random).	$\mu = \frac{1}{p}$ expected number of trials to get the first success	$\sigma = \frac{\sqrt{1-p}}{p}$

Note: Other notation could also be correct if properly defined. Incorrect notation will result in lost points on the AP exam.