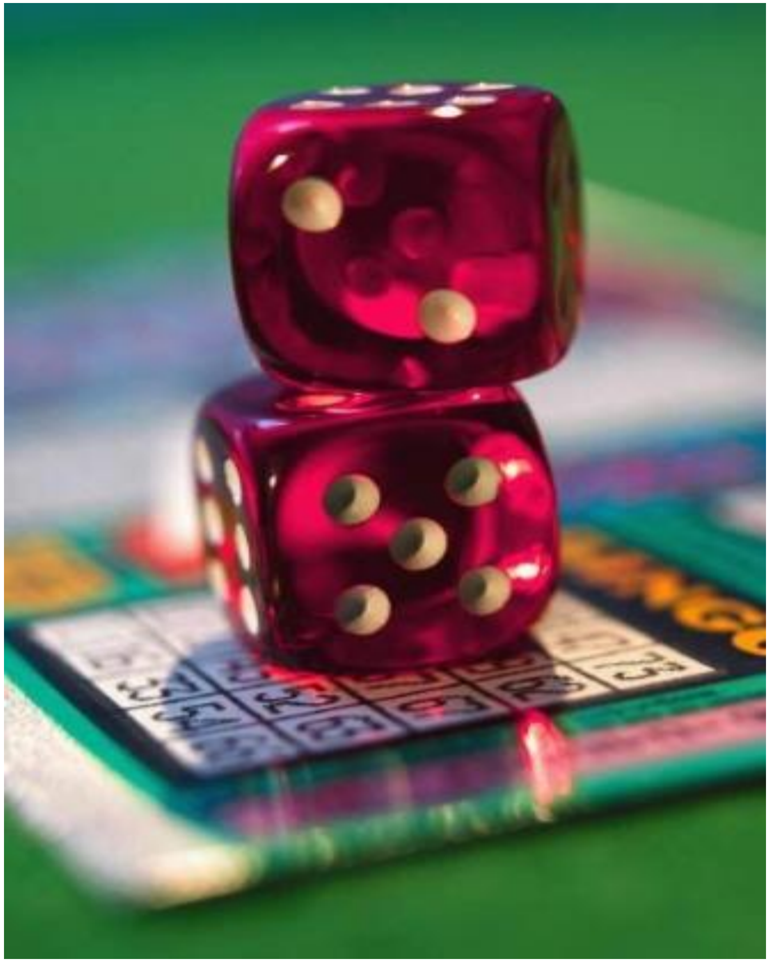
Probability

Chapter 2: Probability

Defining probability

Random processes

- A *random process* is a situation in which we know what outcomes could happen, but we don't know which particular outcome will happen.
- Examples: coin tosses, die rolls, iTunes shuffle, whether the stock market goes up or down tomorrow, etc.
- It can be helpful to model a process as random even if it is not truly random.



Probability

There are several possible interpretations of probability but they (almost) completely agree on the mathematical rules probability must follow.

- $P(A)$ = Probability of event A
- $0 \leq P(A) \leq 1$

Frequentist interpretation:

- The probability of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

Bayesian interpretation:

- A Bayesian interprets probability as a subjective degree of belief: For the same event, two separate people could have different viewpoints and so assign different probabilities.
- Largely popularized by revolutionary advance in computational technology and methods during the last twenty years.

Random Phenomena

- Outcome is Uncertain
- In the short-run outcome are highly random
- In the long-run, outcomes are very predictable
- Probability quantifies long-run randomness

Law of large numbers

Law of large numbers states that as more observations are collected, the proportion of occurrences with a particular outcome, $\hat{p}_n$, converges to the probability of that outcome, p.

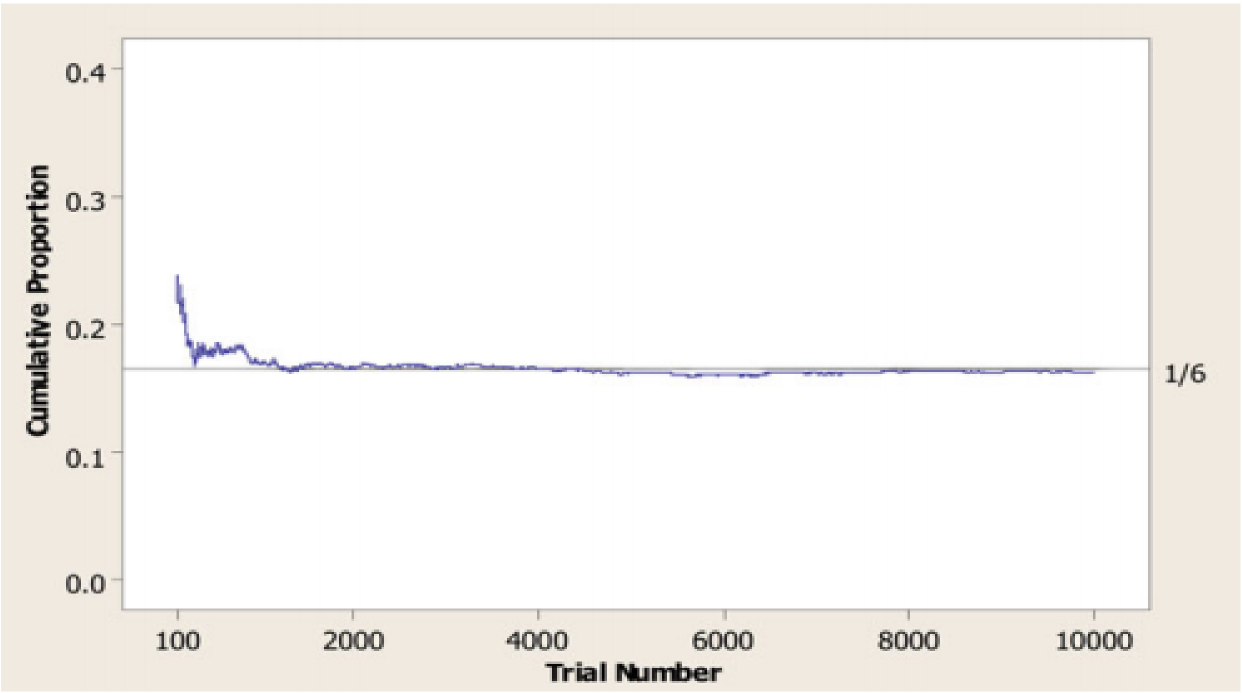


Figure The Cumulative Proportion of Times that a 6 Occurs for a Simulation of 10,000 Rolls of a Fair Die

As # of trials increase, the proportion of an outcome approaches a particular number (probability): In the long run, 1/6 of die Tosses is 6

Law of large numbers (cont.)

When tossing a fair coin, if heads comes up on each of the first 10 tosses, what do you think the chance is that another head will come up on the next toss? 0.5, less than 0.5, or more than 0.5?

$$\underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{?}$$

- The probability is still 0.5, or there is still a 50% chance that another head will come up on the next toss.

$$P(H \text{ on } 11^{th} \text{ toss}) = P(T \text{ on } 11^{th} \text{ toss}) = 0.5$$

- The coin is not “due” for a tail.
- The common misunderstanding of the LLN is that random processes are supposed to compensate for whatever happened in the past; this is just not true and is also called *gambler’s fallacy* (or *law of averages*).

Disjoint and non-disjoint outcomes

Disjoint (mutually exclusive) outcomes: Cannot happen at the same time.

- The outcome of a single coin toss cannot be a head and a tail.
- A student both cannot fail and pass a class.
- A single card drawn from a deck cannot be an ace and a queen.

Addition Rule of disjoint outcomes

If A_1 and A_2 represent two disjoint outcomes, then the probability that one of them occurs is given by

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2)$$

- Note: For disjoint events P(A and B) = 0, so the above formula simplifies to P(A or B) = P(A) + P(B).

Probability of Union of Two Events

Non-disjoint outcomes: Can happen at the same time.

- A student can get an A in Stats and A in Econ in the same semester.

Addition Rule for the union of any two events,

For the union of two events P(A or B) = P(A) + P(B) – P(A and B)

If disjoint, P(A and B) = 0, so

P(A or B) = P(A) + P(B)

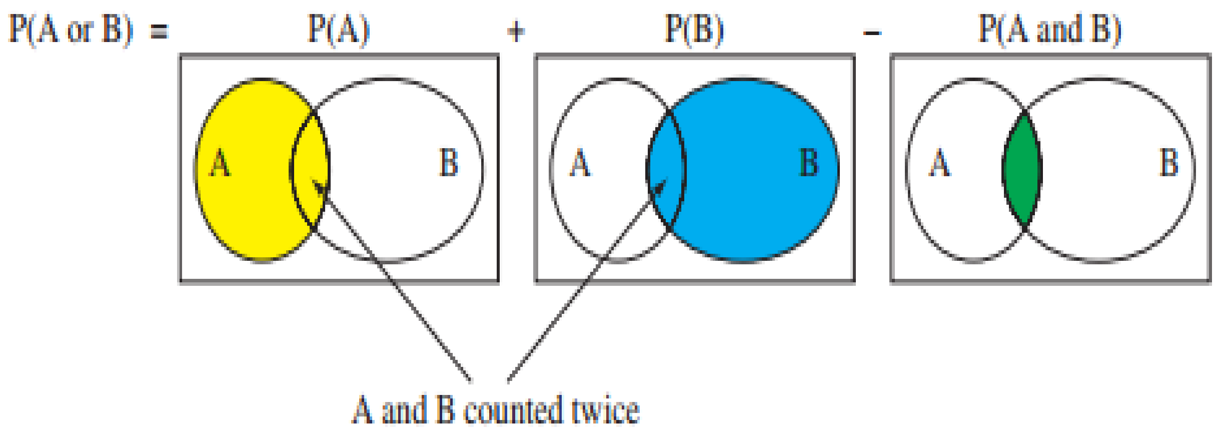


Figure The Probability of the Union, Outcomes in A or B or Both

Intersection & Union of A and B

Intersection is all outcomes in both A and B
Union is all outcomes in either A or B or both

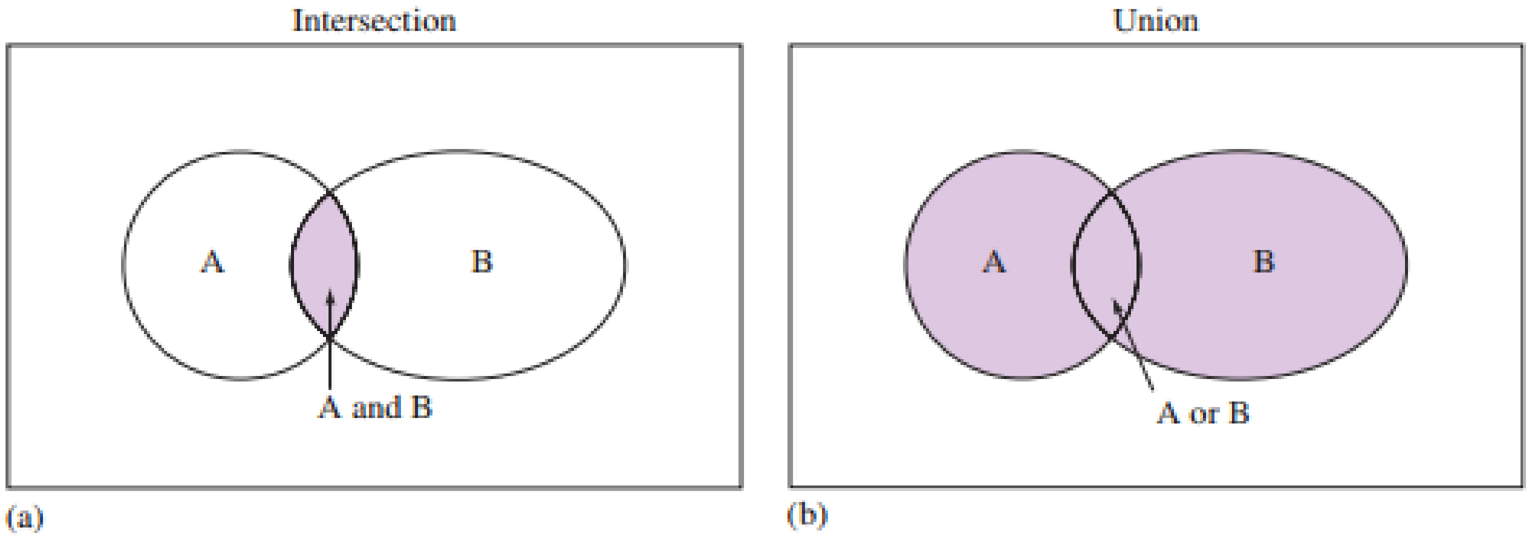


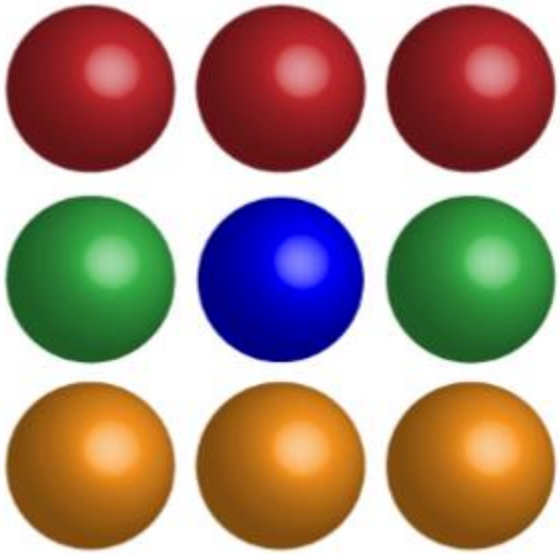
Figure The Intersection and the Union of Two Events. Intersection means A occurs and B occurs, denoted "A and B."

Probability of an Event

- Probability of event A, P(A), is sum of outcome probabilities in event

When all outcomes
are equally likely:

$$P(A) = \frac{\#\,\text{outcomes in A}}{\#\,\text{outcomes in Sample Space}}$$



Probability distributions

- **Probability distributions**

A *probability distribution* lists all possible events and the probabilities with which they occur.

- The probability distribution for the gender of one kid:

Event	Male	Female
Probability	0.5	0.5

- Rules for probability distributions:
  1. The events listed must be disjoint
  2. Each probability must be between 0 and 1
  3. The probabilities must total 1

Probability distributions

Income range ($1000s)	0-25	25-50	50-100	100+
(a)	0.18	0.39	0.33	0.16
(b)	0.38	-0.27	0.52	0.37
(c)	0.28	0.27	0.29	0.16

Figure Proposed distributions of US household incomes

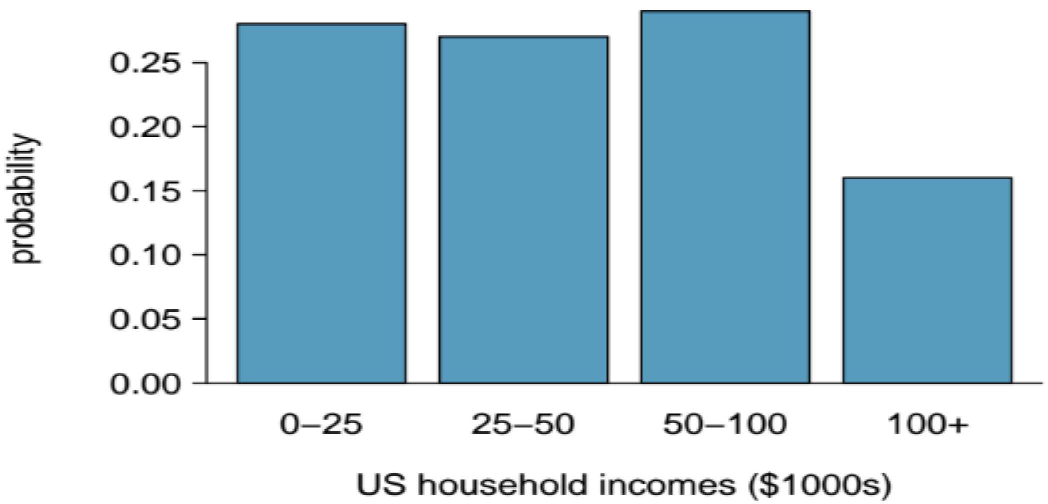


Figure The probability distribution of US household income.

Sample space and complements

Sample space is the collection of all possible outcomes of a trial.

- A couple has one kid, what is the sample space for the gender of this kid? S = {M, F}
- A couple has two kids, what is the sample space for the gender of these kids?

S = {MM, FF, FM, MF}

Complementary events are two mutually exclusive events whose probabilities that add up to 1.

- A couple has one kid. If we know that the kid is not a boy, what is gender of this kid? { ~~M~~, *F* } Boy and girl are *complementary* outcomes.
- A couple has two kids, if we know that they are not both girls, what are the possible gender combinations for these kids?

S = { MM, ~~FF~~, FM, MF }

Complement of Event A

All outcomes in sample space not in A

$P(A) + P(A^C) = 1$

So, P(Ac) = 1 – P(A)

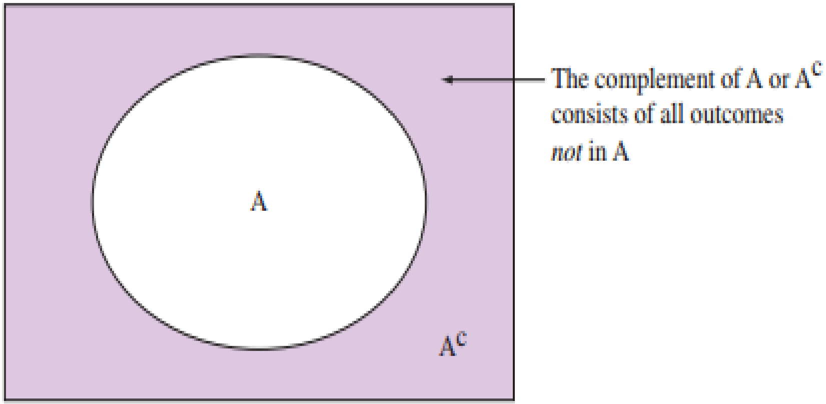


Figure Venn Diagram Illustrating an Event A and Its Complement A^C.

Complement

The complement of event A is denoted A^c, and A^c represents all outcomes not in A. A and A^c are mathematically related:

$$P(A) + P(A^c) = 1, \quad \text{i.e.} \quad P(A) = 1 - P(A^c)$$

Independence

Two processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

- Knowing that the coin landed on a head on the first toss does not provide any useful information for determining what the coin will land on in the second toss.
  >> Outcomes of two tosses of a coin are independent.
- Knowing that the first card drawn from a deck is an ace does provide useful information for determining the probability of drawing an ace in the second draw.
  >> Outcomes of two draws from a deck of cards (without replacement) are dependent.

Checking for independence

If P(A occurs, given that B is true) = $P(A \mid B) = P(A)$,

then A and B are independent.

For instance, on a quiz with only two questions, the instructor found the following proportions for the actual responses of her students (I=incorrect, C=correct) :

Outcome:	II	IC	CI	CC
Probability:	0.26	0.11	0.05	0.58

Let A denote {first question correct}
and let B denote {second question correct}. Based on these probabilities,

$$P(A) = P(\{CI, CC\}) = 0.05 + 0.58 = 0.63$$

$$P(B) = P(\{IC, CC\}) = 0.11 + 0.58 = 0.69$$

$$P(A \text{ and } B) = P(\{CC\}) = 0.58.$$

If A and B were independent, then

$$P(A \text{ and } B) = P(A) \times P(B) = 0.63 \times 0.69 = 0.43.$$

Don't assume independence!

Product rule for independent events

Multiplication Rule for intersection of two independent events, A and B,

P(A and B) = P(A) x P(B)

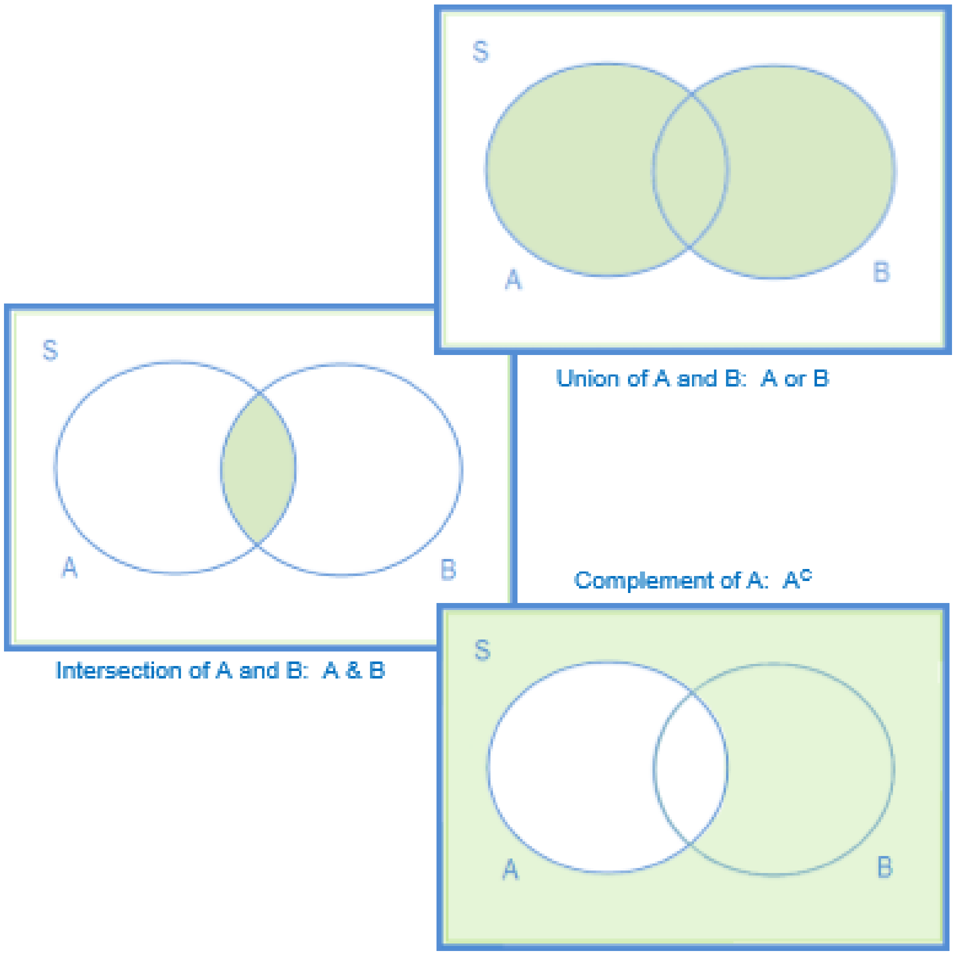
P(A and B) = P(A) x P(B)

Or more generally, $P(A_1$, and, … and $A_k) = P(A_1) \times \ldots \times P(A_k)$

You toss a coin twice, what is the probability of getting two tails in a row?

P(T on the first toss) x P(T on the second toss) = ½ x ½ =¼

Probability Rules of Pairs of Events



- In either or both events (**Union**)
- In both events (**Intersection**)
- Not in the event (**Complement**)

Example

Picture the Scenario

Your statistics instructor decides to give an unannounced quiz with three multiple-choice questions. Each question has five options, and the student's answer is either correct (C) or incorrect (I). If a student answered the first two questions correctly and the last question incorrectly, the student's outcome on the quiz can be symbolized by CCI.

Question to Explore

What is the sample space for the correctness of a student's answers on this pop quiz?

Conditional probability

Marginal probability

Researchers randomly assigned 72 chronic users of cocaine into three groups: desipramine (antidepressant), lithium (standard treatment for cocaine) and placebo. Results of the study are summarized below.

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

What is the probability that a patient relapsed?

P(relapsed) = 48 / 72 ~ 0.67

Joint probability

What is the probability that a patient received the antidepressant (desipramine) <u>and</u> relapsed?

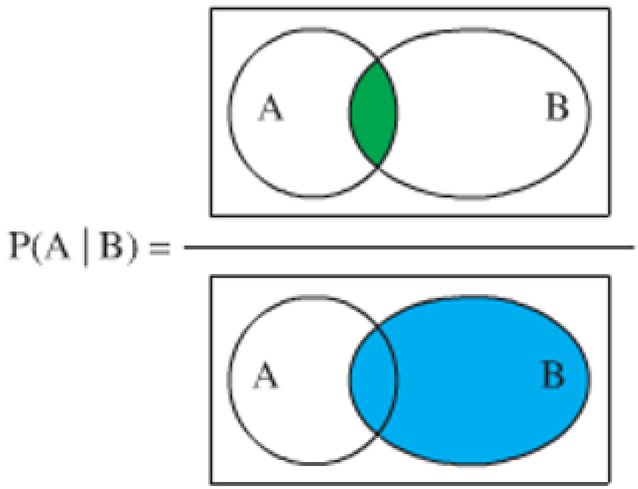
	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(relapsed and desipramine) = 10 / 72 ~ 0.14

Conditional probability

The conditional probability of the outcome of interest A given condition B is calculated as

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$



The vertical slash represents the word “given”. Of the times that B occurs, P(A|B) is the proportion of times that A also occurs

Conditional probability

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

$$P(relapse|desipramine) = \frac{P(relapse \text{ and } desipramine)}{P(desipramine)}$$

$$= \frac{10/72}{24/72} \qquad = \frac{10}{24} \qquad = 0.42$$

Conditional probability

If we know that a patient received the antidepressant (desipramine), what is the probability that they relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

$P(relapse \mid desipramine) = 10 / 24 \sim 0.42$

$P(relapse \mid lithium) = 18 / 24 \sim 0.75$

$P(relapse \mid placebo) = 20 / 24 \sim 0.83$

Conditional probability (cont.)

If we know that a patient relapsed, what is the probability that they received the antidepressant (desipramine)?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

$P(desipramine \mid relapse)$ = 10 / 48 ~ 0.21

$P(lithium \mid relapse)$ = 18 / 48 ~ 0.38

$P(placebo \mid relapse)$ = 20 / 48 ~ 0.42

General multiplication rule

- Earlier we saw that if two events are independent, their joint probability is simply the product of their probabilities. If the events are not believed to be independent, the joint probability is calculated slightly differently.
- If A and B represent two outcomes or events, then

  $$P(A \text{ and } B) = P(A \mid B) \times P(B)$$

  Note that this formula is simply the conditional probability formula, rearranged.
- It is useful to think of A as the outcome of interest and B as the condition.

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is 60 / 100 = 0.6.
  - The probability that a randomly selected student is a social science major given that they are female is 30 / 50 = 0.6.
    - Since *P(SS | M)* also equals 0.6, major of students in this class does not depend on their gender: *P(SS | F) = P(SS)*.

Independence and conditional probabilities (cont.)

Generically, if *P(A | B) = P(A)* then the events *A* and *B* are said to be independent.

- Conceptually: Giving *B* doesn't tell us anything about *A*.

Mathematically: We know that if events *A* and *B* are independent, *P(A and B) = P(A) x P(B)*. Then,

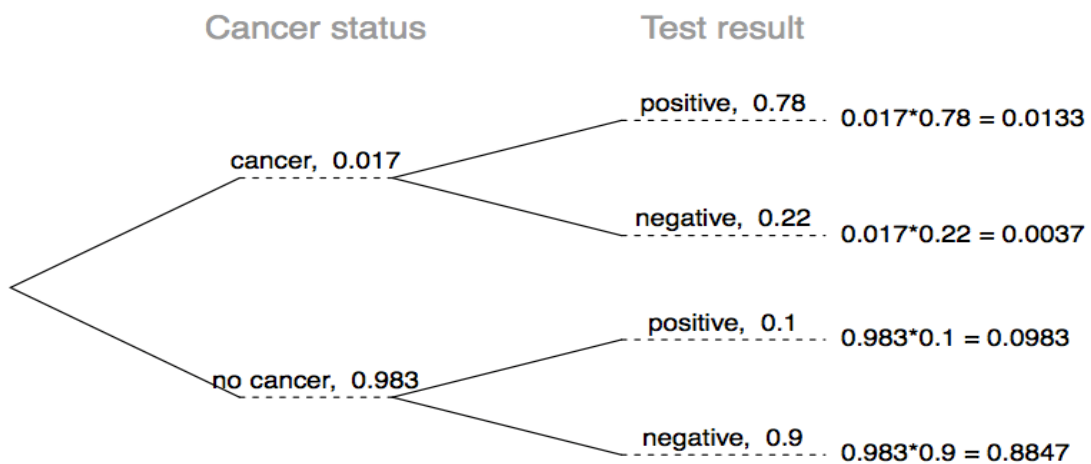
$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A)$$

Breast cancer screening

- American Cancer Society estimates that about 1.7% of women have breast cancer. *http://www.cancer.org/cancer/cancerbasics/cancer-prevalence*
- Susan G. Komen For The Cure Foundation states that mammography correctly identifies about 78% of women who truly have breast cancer. *http://ww5.komen.org/BreastCancer/AccuracyofMammograms.html*
- An article published in 2003 suggests that up to 10% of all mammograms result in false positives for patients who do not have cancer. *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1360940*

Inverting probabilities

When a patient goes through breast cancer screening there are two competing claims: patient had cancer and patient doesn't have cancer. If a mammogram yields a positive result, what is the probability that patient actually has cancer?



$$P(C|+) = \frac{P(C \text{ } and \text{ } +)}{P(+)} = \frac{0.0133}{0.0133 + 0.0983} = 0.12$$

Practice

2-What is the probability that this woman has cancer if this second mammogram also yielded a positive result?

$$P(C|+) = \frac{P(C\ and\ +)}{P(+)} = \frac{0.0936}{0.0936 + 0.088} = 0.52$$

Bayes' Theorem

The conditional probability formula we have seen so far is a special case of the Bayes' Theorem, which is applicable even when events have more than just two outcomes.

$$P(\textit{outcome A of variable } 1 \mid \textit{outcome B of variable } 2)$$

$$= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)}$$

where $A_2, \cdots, A_k$ represent all other possible outcomes of variable 1.

Sampling from a small population

Sampling with replacement

When sampling with replacement, you put back what you just drew.

- Imagine you have a bag with 5 red, 3 blue and 2 orange chips in it. What is the probability that the first chip you draw is blue?

5 O, 3 O, 2 O

$$Prob(1^{st} \text{ chip } B) = \frac{3}{5+3+2} = \frac{3}{10} = 0.3$$

- Suppose you did indeed pull a blue chip in the first draw. If drawing with replacement, what is the probability of drawing a blue chip in the second draw?

1st Draw - 5 O, 3 O, 2 O

2nd Draw - 5 O, 3 O, 2 O

P(2nd chip B | 1st chip B) = 3 / 10 = 0.3

Sampling with replacement (cont.)

- Suppose you actually pulled an orange chip in the first draw. If drawing with replacement, what is the probability of drawing a blue chip in the second draw?

1st Draw - 5 **O**, 3 **O**, 2 **O**

2nd Draw - 5 **O**, 3 **O**, 2 **O**

P(2nd chip B | 1st chip O) = 3 / 10 = 0.3

- If drawing with replacement, what is the probability of drawing two blue chips in a row?

1st Draw - 5 **O**, 3 **O**, 2 **O**

2nd Draw - 5 **O**, 3 **O**, 2 **O**

P(1st chip B) x P(2nd chip B | 1st chip B) = 0.3 x 0.3 = 0.09

Sampling with replacement (cont.)

When drawing with replacement, probability of the second chip being blue does not depend on the color of the first chip since whatever we draw in the first draw gets put back in the bag.

$$Prob(B \mid B) = Prob(B \mid O)$$

In addition, this probability is equal to the probability of drawing a blue chip in the first draw, since the composition of the bag never changes when sampling with replacement.

$$Prob(B \mid B) = Prob(B)$$

When drawing with replacement, draws are independent

Sampling without replacement

When drawing without replacement you do not put back what you just drew.

- Suppose you pulled a blue chip in the first draw. If drawing without replacement, what is the probability of drawing a blue chip in the second draw?

1st Draw - 5 **O**, 3 **O**, 2 **O**

2nd Draw - 5 **O**, <u>2</u> **O**, 2 **O**

P(2nd chip B | 1st chip B) = 2 / 9 = 0.22

- If drawing without replacement, what is the probability of drawing two blue chips in a row?

1st Draw - 5 **O**, 3 **O**, 2 **O**

2nd Draw - 5 **O**, <u>2</u> **O**, 2 **O**

P(1st chip B) x P(2nd chip B | 1st chip B) = 0.3 x 0.22 = 0.066

Sampling without replacement (cont.)

When drawing without replacement, the probability of the second chip being blue given the first was blue is not equal to the probability of drawing a blue chip in the first draw since the composition of the bag changes with the outcome of the first draw.

$$Prob(B \mid B) \neq Prob(B)$$

When drawing without replacement, draws are not independent.

This is especially important to take note of when the sample sizes are small. If we were dealing with, say, 10,000 chips in a (giant) bag, taking out one chip of any color would not have as big an impact on the probabilities in the second draw.

Random variables

Random variables

A *random variable* is a numeric quantity whose value depends on the outcome of a random event

- We use a capital letter, like X, to denote a random variable
- The values of a random variable are denoted with a lowercase letter, in this case x
- For example, $P(X = x)$

There are two types of random variables:

- *Discrete random variables* often take only integer values
    - Example: Number of credit hours, Difference in number of credit hours this term vs last
- *Continuous random variables* take real (decimal) values
    - Example: Cost of books this term, Difference in cost of books this term vs last

Expectation

- We are often interested in the average outcome of a random variable.
- We call this the *expected value* (mean), and it is a weighted average of the possible outcomes

$$\mu = E(X) = \sum_{i=1}^{k} x_i P(X = x_i)$$

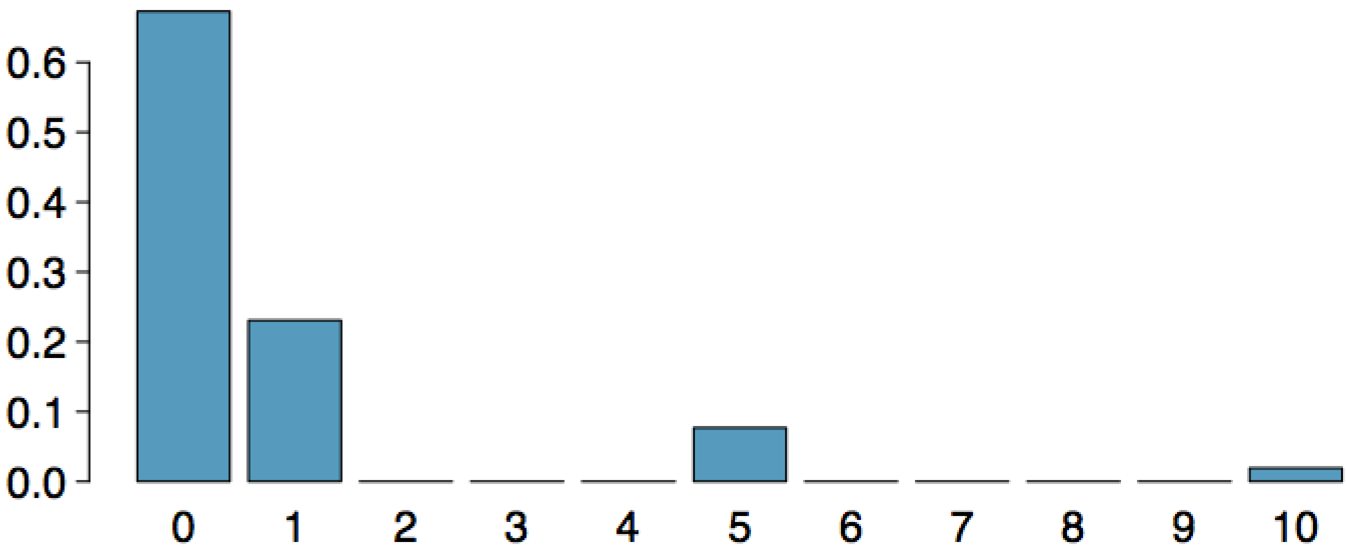
Expected value of a discrete random variable

In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw. Write the probability model for your winnings, and calculate your expected winning.

Event	X	$P(X)$	$X\,P(X)$
Heart (not ace)	1	$\frac{12}{52}$	$\frac{12}{52}$
Ace	5	$\frac{4}{52}$	$\frac{20}{52}$
King of spades	10	$\frac{1}{52}$	$\frac{10}{52}$
All else	0	$\frac{35}{52}$	0
Total			$E(X) = \frac{42}{52} \approx 0.81$

Expected value of a discrete random variable (cont.)

Below is a visual representation of the probability distribution of winnings from this game:



Variability

We are also often interested in the variability in the values of a random variable.

$$\sigma^2 = Var(X) = \sum_{i=1}^{k}(x_i - E(X))^2 P(X = x_i)$$

$$\sigma = SD(X) = \sqrt{Var(X)}$$

For the previous card game example, how much would you expect the winnings to vary from game to game?

X	$P(X)$	$X\,P(X)$	$(X - E(X))^2$	$P(X)\,(X - E(X))^2$
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$	$(1 - 0.81)^2 = 0.0361$	$\frac{12}{52} \times 0.0361 = 0.0083$
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$	$(5 - 0.81)^2 = 17.5561$	$\frac{4}{52} \times 17.5561 = 1.3505$
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.0889 = 1.6242$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$	$(0 - 0.81)^2 = 0.6561$	$\frac{35}{52} \times 0.6561 = 0.4416$
		$E(X) = 0.81$		$V(X) = 3.4246$ $SD(X) = \sqrt{3.4246} = 1.85$

Linear combinations

- A *linear combination* of random variables X and Y is given by

$$aX + bY$$

  where a and b are some fixed numbers.
- The average value of a linear combination of random variables is given by

$$E(aX + bY) = a \times E(X) + b \times E(Y)$$

On average you take 10 minutes for each statistics homework problem and 15 minutes for each chemistry homework problem. This week you have 5 statistics and 4 chemistry homework problems assigned. What is the total time you expect to spend on statistics and physics homework for the week?

$$\begin{aligned} E(5S + 4C) &= 5 \times E(S) + 4 \times E(C) \\ &= 5 \times 10 + 4 \times 15 \\ &= 50 + 60 \\ &= 110\ min \end{aligned}$$

Linear Combination

The variability of a linear combination of two independent random variables is calculated as:

$$V(aX + bY) = a^2 \times V(X) + b^2 \times V(Y)$$

The standard deviation of the linear combination is the square root of the variance.

Note: If the random variables are not independent, the variance calculation gets a little more complicated and is beyond the scope of this course.

Linear combinations

The standard deviation of the time you take for each statistics homework problem is 1.5 minutes, and it is 2 minutes for each chemistry problem. What is the standard deviation of the time you expect to spend on statistics and physics homework for the week if you have 5 statistics and 4 chemistry homework problems assigned?

$$\begin{aligned} V(5S + 4C) &= 5^2 \times V(S) + 4^2 \times V(C) \\ &= 25 \times 1.5^2 + 16 \times 2^2 \\ &= 56.25 + 64 \\ &= 120.25 \end{aligned}$$

Fair game

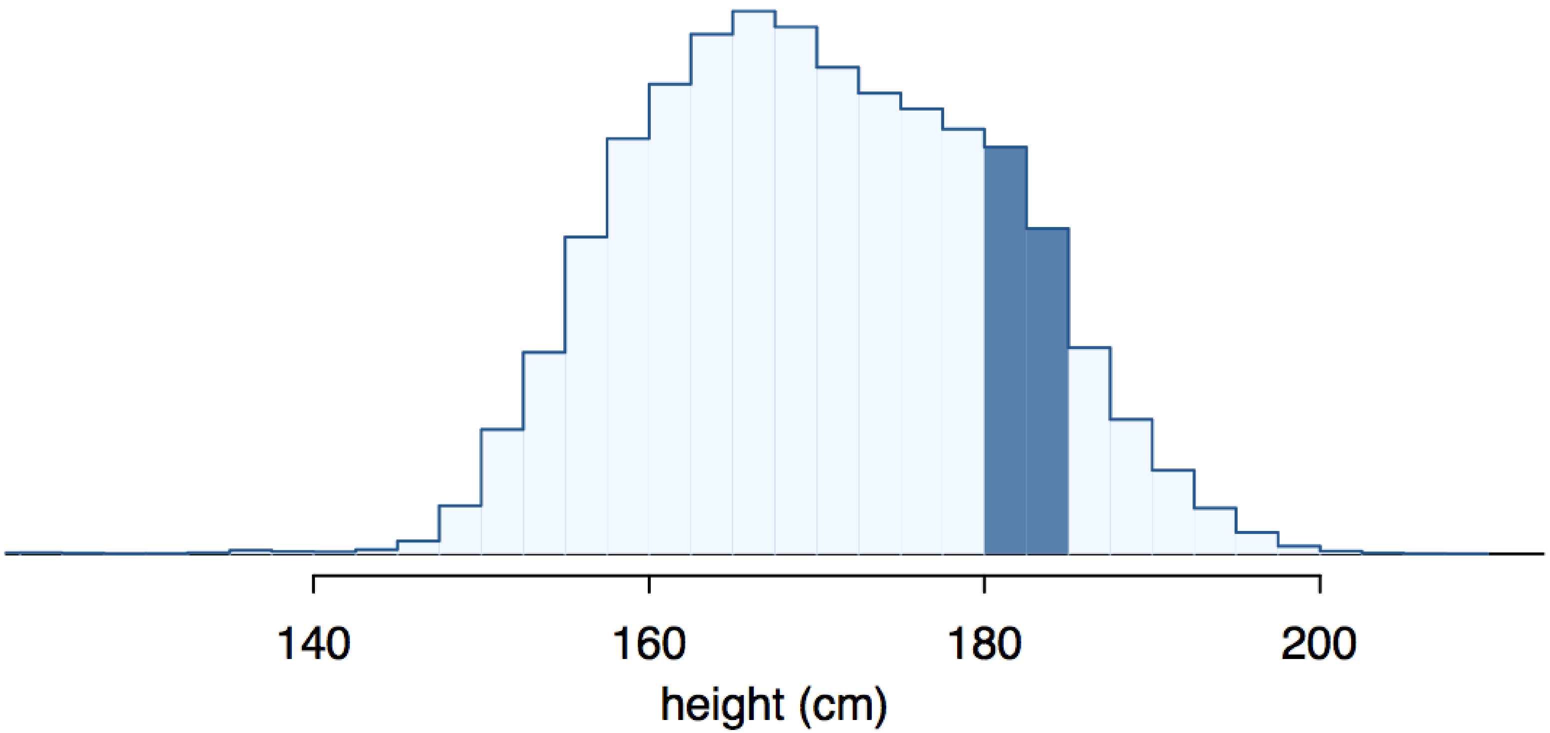
A *fair* game is defined as a game that costs as much as its expected payout, i.e. expected profit is 0.

Do you think casino games in Vegas cost more or less than their expected payouts?

If those games cost less than their expected payouts, it would mean that the casinos would be losing money on average, and hence they wouldn't be able to pay for all this:

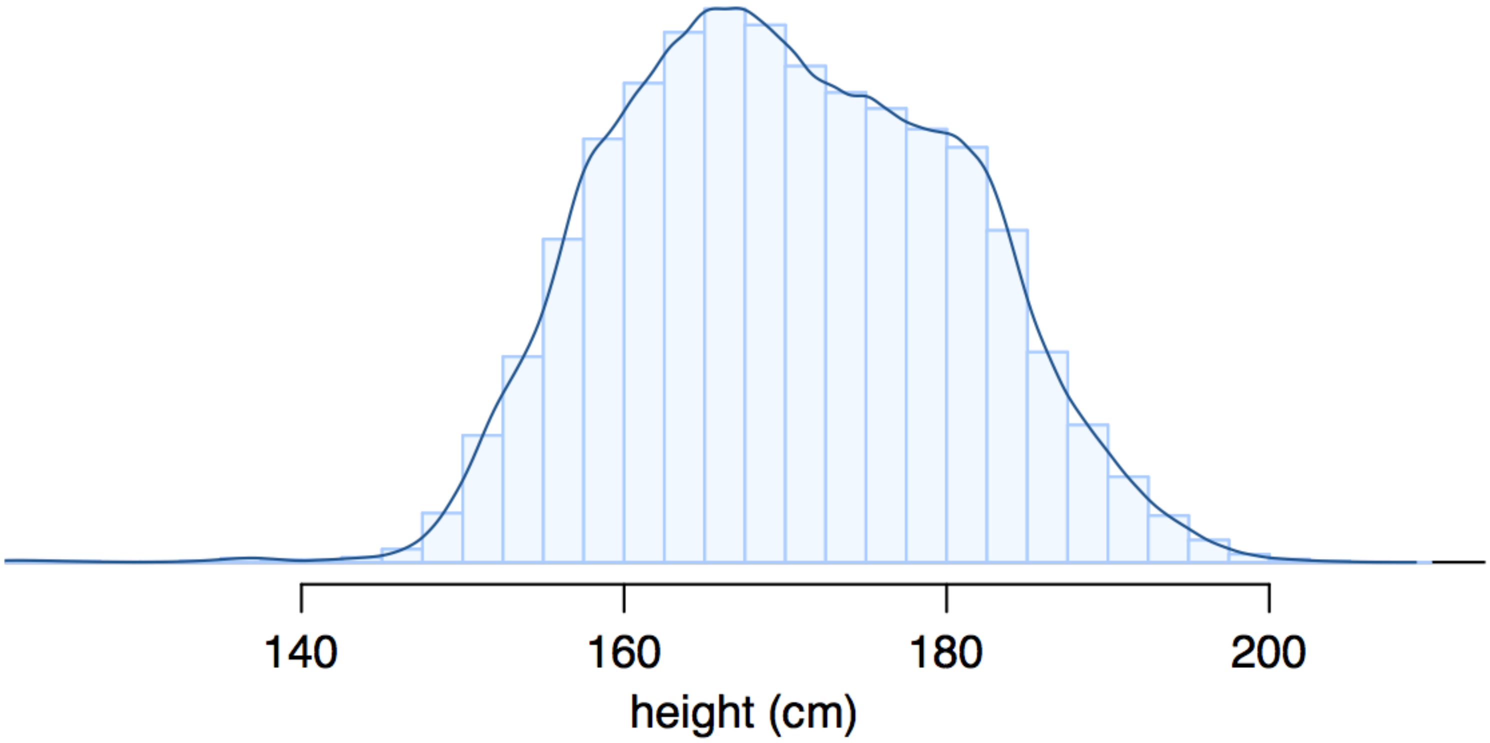
Continuous distributions

- Below is a histogram of the distribution of heights of US adults.
- The proportion of data that falls in the shaded bins gives the probability that a randomly sampled US adult is between 180 cm and 185 cm (about 5'11" to 6'1").



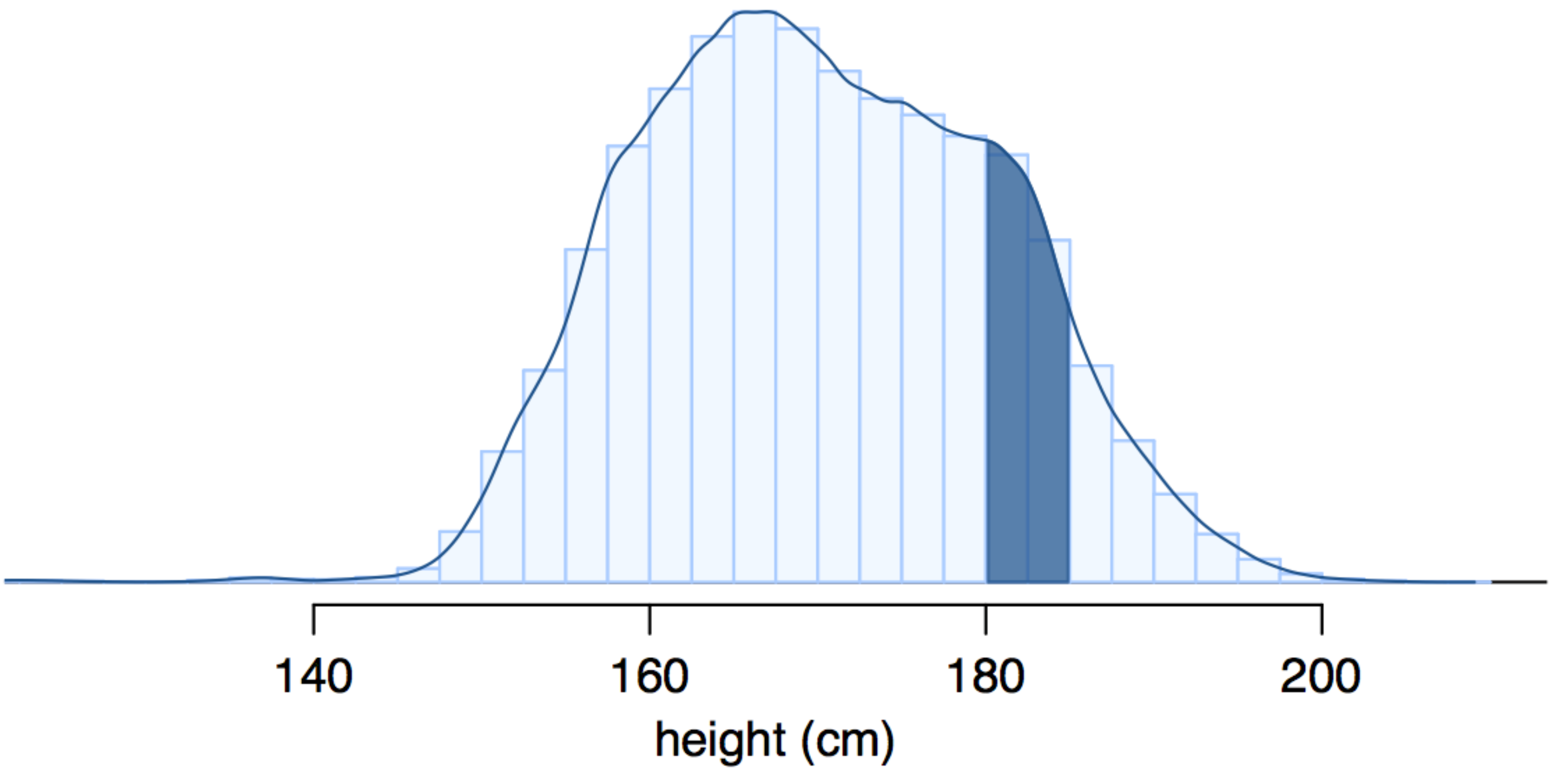
Continuous distributions

Since height is a continuous numerical variable, its probability density function is a smooth curve.



From histograms to continuous distributions

Therefore, the probability that a randomly sampled US adult is between 180 cm and 185 cm can also be estimated as the shaded area under the curve.



Probabilities from continuous distributions

Since continuous probabilities are estimated as “the area under the curve”, the probability of a person being exactly 180 cm (or any exact value) is defined as 0.