Math 11
Calculus-Based Introductory Probability and Statistics

Eddie Aamari
S.E.W. Assistant Professor

eaamari@ucsd.edu
math.ucsd.edu/~eaamari/
AP&M 5880A

Today:

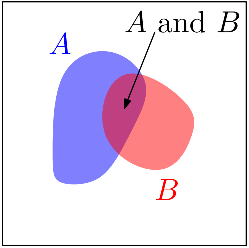
- More Probability Theory

Recap of Last Lecture

- **"or" rule:** If two events A and B are disjoint,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Works for non-disjoint events!

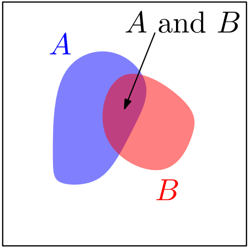


Recap of Last Lecture

- **“or” rule:** If two events A and B are disjoint,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Works for non-disjoint events!



- **“and” rule:** If two events A and B are independent,

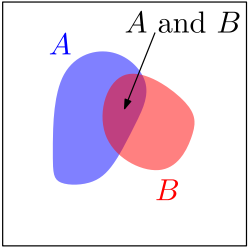
$$P(A \text{ and } B) = P(A) \times P(B).$$

Recap of Last Lecture

- **"or" rule:** If two events A and B are disjoint,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Works for non-disjoint events!



- **"and" rule:** If two events A and B are independent,

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

What if A and B are not independent?

Losing Independence

Losing Independence

Conditional Probability is a tool to handle events that are not independent.

Idea: When computing a probability, you actually have some extra information that you know is true.

Losing Independence

Conditional Probability is a tool to handle events that are not independent.

Idea: When computing a probability, you actually have some extra information that you know is true.

Example:

- $A = \{$ It will rain today in San Diego $\}$
- $B = \{$ You see dark storm clouds in the sky $\}$

Losing Independence

Conditional Probability is a tool to handle events that are not independent.

Idea: When computing a probability, you actually have some extra information that you know is true.

Example:

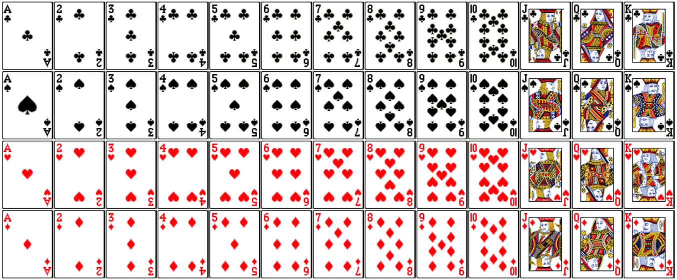
- $A = \{$ It will rain today in San Diego $\}$
- $B = \{$ You see dark storm clouds in the sky $\}$

Although $P(A) \simeq 11.2\%$ is small, you have a much higher chance to see A happen if you know already that B occurred.

Conditional Probability

A card is drawn from a deck. What is the probability that the card is a heart, given that the card is a king?

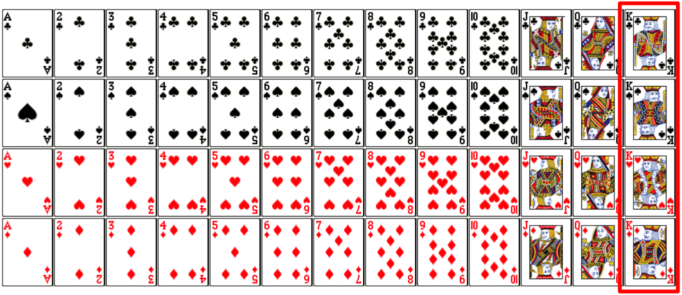
Intuition says 1/4.



Conditional Probability

A card is drawn from a deck. What is the probability that the card is a heart, given that the card is a king?

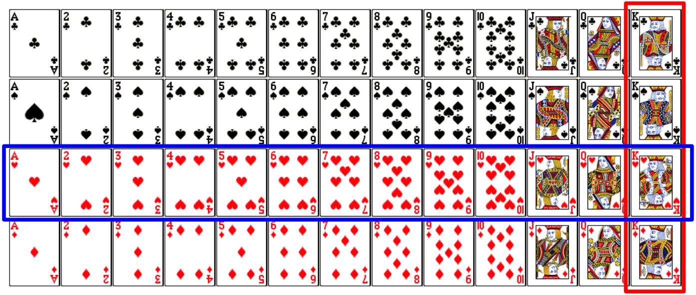
Intuition says 1/4.



Conditional Probability

A card is drawn from a deck. What is the probability that the card is a heart, given that the card is a king?

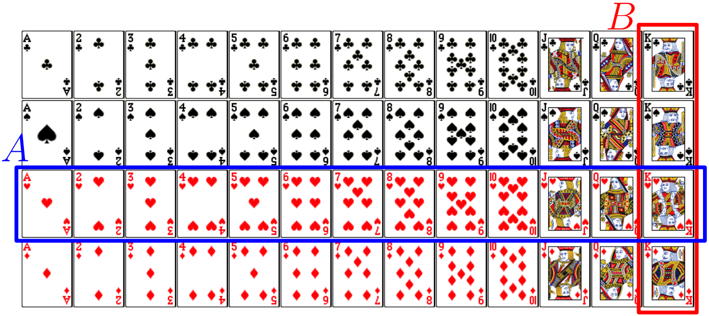
Intuition says 1/4.



Conditional Probability

A card is drawn from a deck. What is the probability that the card is a heart, given that the card is a king?

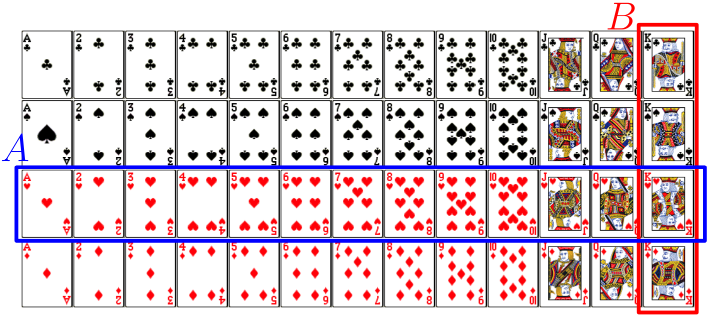
Intuition says 1/4.



Conditional Probability

A card is drawn from a deck. What is the probability that the card is a heart, given that the card is a king?

Intuition says 1/4.

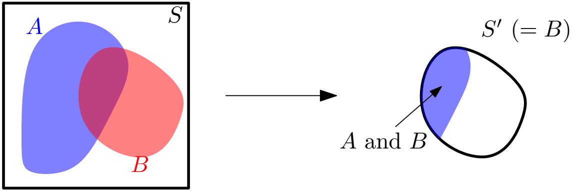


$$P(A \text{ given that } B \text{ occured}) = \frac{P(A \text{ and } B)}{P(B)}$$

Conditional Probability: Definition

For two events A, B the conditional probability of A given B is

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



Computing $P(A|B)$ amounts to do as if B was the sample space.

Understanding Conditional Probability Visually

Go to

http://students.brown.edu/seeing-theory/compound-probability/

and click on Conditional Probability.

Here,

- A is the event "the ball hit ledge A"
- B is the event "the ball hit ledge B"
- (Ignore ledge C)

Note that you can move and stretch any of the ledges.

How can we use the visualization to show $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)}$?

If you get to assume event B has occurred, you can click the B button (lower right) to make it your whole universe. Then $P(A|B)$ is just the fraction of the B ledge covered by the portion of the A ledge in the picture.

Notion of Independence Revisited

By definition, A and B are independent when

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

Notion of Independence Revisited

By definition, A and B are independent when

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

But for any two events A and B, we have

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

Notion of Independence Revisited

By definition, A and B are independent when

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

But for any two events A and B, we have

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

Therefore,

$$\begin{aligned} A \text{ and } B \text{ are independent } &\Leftrightarrow P(A|B) = P(A) \\ &\Leftrightarrow P(B|A) = P(B). \end{aligned}$$

Checking for Independence Revisited

Independent or not?

1. Find $P(A)$
2. Find $P(A$ assuming you know event B has occured)

Do you get the same answer?

- **Yes:** Events are independent
- **No:** Events are NOT independent (= dependent)

Checking for Independence Revisited

Independent or not?

1. Find $P(A)$
2. Find $P(A$ assuming you know event B has occured$) = \mathbf{P(A|B)}$

Do you get the same answer? **= Check if $\mathbf{P(A|B) = P(A)}$**

- **Yes:** Events are independent
- **No:** Events are NOT independent (= dependent)

Social Justice and Independence

Those fighting for social justice often want independence:

$$P(\text{Female}) = P(\text{Female}|\text{Person is a scientist})$$

means the events "Female" and "Person is a scientist" are independent.

$$P(\text{Being pulled over}) = P(\text{Being pulled over}|\text{Being } \textit{insert race here})$$

means that the likelihood to be pulled over should not change based on your race.

Conditional Probability in a Contingency Table

	Junior	Other Levels	Margin totals
Biology Major	41	20	61
Other Major	44	87	131
Margin Totals	85	107	192

What is $P(\text{Junior})$?

Conditional Probability in a Contingency Table

	Junior	Other Levels	Margin totals
Biology Major	41	20	61
Other Major	44	87	131
Margin Totals	85	107	192

What is $P(\text{Junior})$?

$$p = \frac{85}{192} \simeq 44.2\%.$$

Conditional Probability in a Contingency Table

	Junior	Other Levels	Margin totals
Biology Major	41	20	61
Other Major	44	87	131
Margin Totals	85	107	192

What is $P(\text{Junior})$? $p = \dfrac{85}{192} \simeq 44.2\%.$

What is $P(\text{Junior}|\text{Biology Major})$?

Conditional Probability in a Contingency Table

	Junior	Other Levels	Margin totals
Biology Major	41	20	61
Other Major	44	87	131
Margin Totals	85	107	192

What is $P(\text{Junior})$? $$p = \frac{85}{192} \simeq 44.2\%.$$

What is $P(\text{Junior}|\text{Biology Major})$? $$p = \frac{41}{61} \simeq 67.2\%.$$

So the *Level* and *Major* are NOT independent.

Things Work The Same With Probabilities, Not Counts

	Like Video Games	Dislike Video Games	Margin Totals
Like Learning	0.62	0.18	0.8
Dislike Learning	0.08	0.12	0.2
Margin Totals	0.7	0.3	1

Find $P(\text{Dislike Video Games}|\text{Like Learning})$.

Things Work The Same With Probabilities, Not Counts

	Like Video Games	Dislike Video Games	Margin Totals
Like Learning	0.62	0.18	0.8
Dislike Learning	0.08	0.12	0.2
Margin Totals	0.7	0.3	1

Find $P(\text{Dislike Video Games}|\text{Like Learning})$.

$$p = \frac{0.18}{0.8} = 22.5\%.$$

Things Work The Same With Probabilities, Not Counts

	Like Video Games	Dislike Video Games	Margin Totals
Like Learning	0.62	0.18	0.8
Dislike Learning	0.08	0.12	0.2
Margin Totals	0.7	0.3	1

Find $P(\text{Dislike Video Games}|\text{Like Learning})$.

$$p = \frac{0.18}{0.8} = 22.5\%.$$

In general,

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

Natural Language and Probabilities

We study data from 3921 emails sent to one user over 3 months, and describe each email with:

- Spam/Not Spam: Indicator for whether the email was spam
- None/Small/Big: Saying whether there was no number, a small number (under 1 million), or a big number

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

Natural Language and Probabilities

We study data from 3921 emails sent to one user over 3 months, and describe each email with:

- Spam/Not Spam: Indicator for whether the email was spam
- None/Small/Big: Saying whether there was no number, a small number (under 1 million), or a big number

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

For each case:

- Write an expression for the given probability
- Say if this probability is marginal, joint, or conditional

Natural Language and Probabilities

We study data from 3921 emails sent to one user over 3 months, and describe each email with:

- Spam/Not Spam: Indicator for whether the email was spam
- None/Small/Big: Saying whether there was no number, a small number (under 1 million), or a big number

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

For each case:

- Write an expression for the given probability
- Say if this probability is marginal, joint, or conditional

- What percent of messages are spam with no number?

Natural Language and Probabilities

We study data from 3921 emails sent to one user over 3 months, and describe each email with:

- Spam/Not Spam: Indicator for whether the email was spam
- None/Small/Big: Saying whether there was no number, a small number (under 1 million), or a big number

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

For each case:

- Write an expression for the given probability
- Say if this probability is marginal, joint, or conditional

- What percent of messages are spam with no number?
  $P($ spam and no number $)$, joint probability

Natural Language and Probabilities

We study data from 3921 emails sent to one user over 3 months, and describe each email with:

- Spam/Not Spam: Indicator for whether the email was spam
- None/Small/Big: Saying whether there was no number, a small number (under 1 million), or a big number

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

For each case:

- Write an expression for the given probability
- Say if this probability is marginal, joint, or conditional

- What percent of messages are spam with no number?
  $P($ spam and no number $)$, joint probability

$$P(\text{ spam and no number }) = \frac{149}{3921} \simeq 3.8\%$$

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What is the probability that a spam message will have a small number?

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What is the probability that a spam message will have a small number?
  $P($ small number | spam $)$, conditional probability

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What is the probability that a spam message will have a small number?
  $P($ small number | spam $)$, conditional probability

$$P(\text{ small number} \mid \text{spam }) = \frac{168}{149 + 168 + 50} \simeq 45.8\%$$

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What is the probability that a spam message will have a small number?
  $P($ small number $\mid$ spam $)$, conditional probability

$$P(\text{ small number } \mid \text{ spam }) = \frac{168}{149 + 168 + 50} \simeq 45.8\%$$

- What is the likelihood that a randomly-selected message with a big number will not be spam?

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What is the probability that a spam message will have a small number?
  $P($ small number | spam $)$, conditional probability

$$P(\text{ small number} \mid \text{spam }) = \frac{168}{149 + 168 + 50} \simeq 45.8\%$$

- What is the likelihood that a randomly-selected message with a big number will not be spam?
  $P($ not spam | big number $)$, conditional probability

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What is the probability that a spam message will have a small number?
  $P($ small number | spam $)$, conditional probability

$$P(\text{ small number } | \text{ spam }) = \frac{168}{149 + 168 + 50} \simeq 45.8\%$$

- What is the likelihood that a randomly-selected message with a big number will not be spam?
  $P($ not spam | big number $)$, conditional probability

$$P(\text{ not spam } | \text{ big number }) = \frac{495}{495 + 50} \simeq 90.8\%$$

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What percent of messages do not contain a small number, ignoring the categorization of spam?

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What percent of messages do not contain a small number, ignoring the categorization of spam?
  $P($ not small $)$, marginal probability

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What percent of messages do not contain a small number, ignoring the categorization of spam?
  $P($ not small $)$, marginal probability

$$P(\text{ not small }) = \frac{400 + 149 + 495 + 50}{3921} \simeq 27.9\%$$

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What percent of messages do not contain a small number, ignoring the categorization of spam?
  $P($ not small $)$, marginal probability

$$P(\text{ not small }) = \frac{400 + 149 + 495 + 50}{3921} \simeq 27.9\%$$

- What fraction of emails would we expect to be non-spam with small or big numbers?

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What percent of messages do not contain a small number, ignoring the categorization of spam?
  $P($ not small $)$, marginal probability

$$P(\text{ not small }) = \frac{400 + 149 + 495 + 50}{3921} \simeq 27.9\%$$

- What fraction of emails would we expect to be non-spam with small or big numbers?
  $P($ not spam and not none $)$, joint probability

Natural Language and Probabilities

	None	Small	Big
Not Spam	400	2659	495
Spam	149	168	50

- What percent of messages do not contain a small number, ignoring the categorization of spam?
  $P($ not small $)$, marginal probability

$$P(\text{ not small }) = \frac{400 + 149 + 495 + 50}{3921} \simeq 27.9\%$$

- What fraction of emails would we expect to be non-spam with small or big numbers?
  $P($ not spam and not none $)$, joint probability

$$P(\text{ not spam and not none }) = \frac{2659 + 495}{3921} \simeq 80.4\%$$

Dependent VS Independent Events

A box contains 2 red balls and 3 green ones.
You pick two balls **with replacement** (= you pick a ball, note its color, put it back in the box, and then pick a second one).
What is the probability that both balls are red?

Dependent VS Independent Events

A box contains 2 red balls and 3 green ones.
You pick two balls **with replacement** (= you pick a ball, note its color, put it back in the box, and then pick a second one).
What is the probability that both balls are red? Write

$$A = \text{“the first ball is red”}$$
$$B = \text{“the second ball is red”}$$

Dependent VS Independent Events

A box contains 2 red balls and 3 green ones.
You pick two balls **with replacement** (= you pick a ball, note its color, put it back in the box, and then pick a second one).
What is the probability that both balls are red? Write

$$A = \text{“the first ball is red”}$$
$$B = \text{“the second ball is red”}$$

<u>As you replace the ball after the first draw</u>, B is unaffected by A.

Dependent VS Independent Events

A box contains 2 red balls and 3 green ones.
You pick two balls **with replacement** (= you pick a ball, note its color, put it back in the box, and then pick a second one).
What is the probability that both balls are red? Write

$$A = \text{"the first ball is red"}$$
$$B = \text{"the second ball is red"}$$

As you replace the ball after the first draw, B is unaffected by A.

By independence, $P(A \text{ and } B) = P(A) \times P(B) = \frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$.

Dependent VS Independent Events

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement** (= you pick a ball, note its color, do not put it back in the box, and then pick a second one).
What is the probability that both balls are red?

Dependent VS Independent Events

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement** (= you pick a ball, note its color, do not put it back in the box, and then pick a second one).
What is the probability that both balls are red? Write

$$A = \text{“the first ball is red”}$$
$$B = \text{“the second ball is red”}$$

Dependent VS Independent Events

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement** (= you pick a ball, note its color, do not put it back in the box, and then pick a second one).
What is the probability that both balls are red? Write

$$A = \text{“the first ball is red”}$$
$$B = \text{“the second ball is red”}$$

B is affected by A, since if A occurs, there is only 1 red ball left among the 4 balls remaining, so B is less likely to happen.

Dependent VS Independent Events

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement** (= you pick a ball, note its color, do not put it back in the box, and then pick a second one).
What is the probability that both balls are red? Write

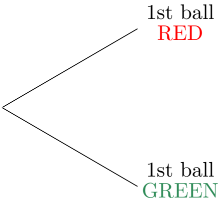
$$A = \text{"the first ball is red"}$$
$$B = \text{"the second ball is red"}$$

B is affected by A, since if A occurs, there is only 1 red ball left among the 4 balls remaining, so B is less likely to happen.

$$P(A \text{ and } B) = P(B|A) \times P(A) = \frac{1}{4} \times \frac{2}{5} = \frac{1}{10}.$$

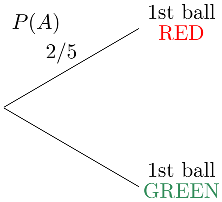
Tree Diagram

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



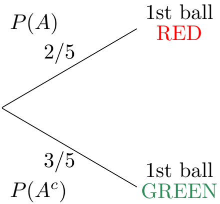
Tree Diagram

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



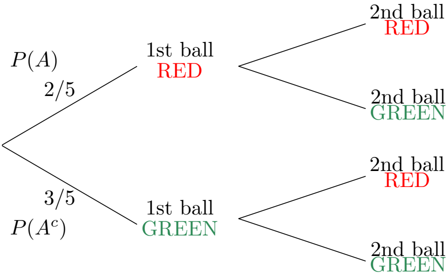
Tree Diagram

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



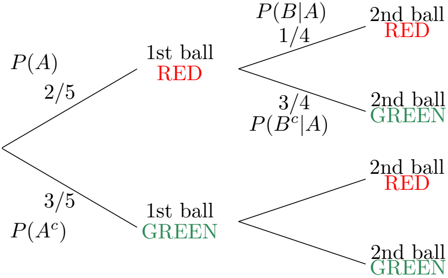
Tree Diagram

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



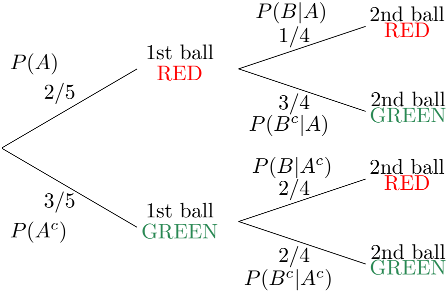
Tree Diagram

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



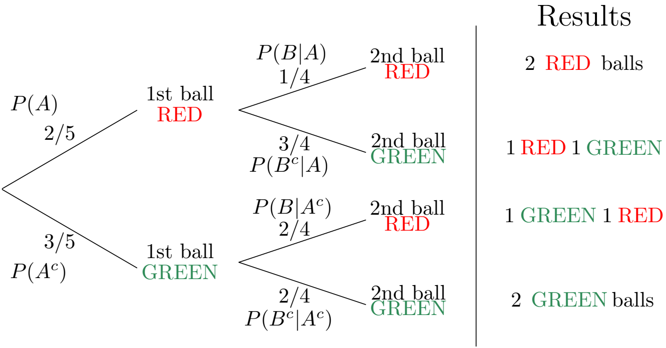
Tree Diagram

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



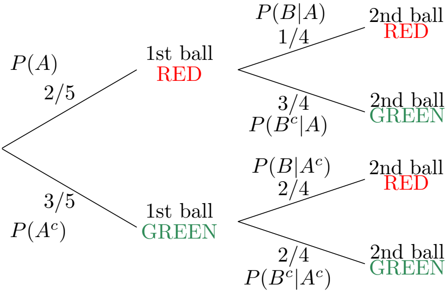
Tree Diagram

A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



Tree Diagram

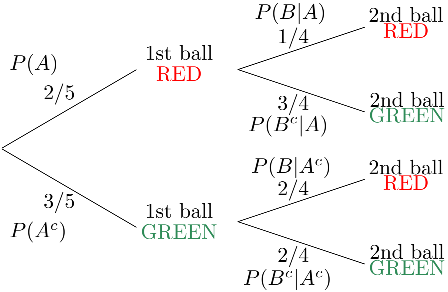
A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



Results	Probability
2 RED balls	$\frac{2}{5} \times \frac{1}{4}$
1 RED 1 GREEN	$\frac{2}{5} \times \frac{3}{4}$
1 GREEN 1 RED	$\frac{3}{5} \times \frac{2}{4}$
2 GREEN balls	$\frac{3}{5} \times \frac{2}{4}$

Tree Diagram

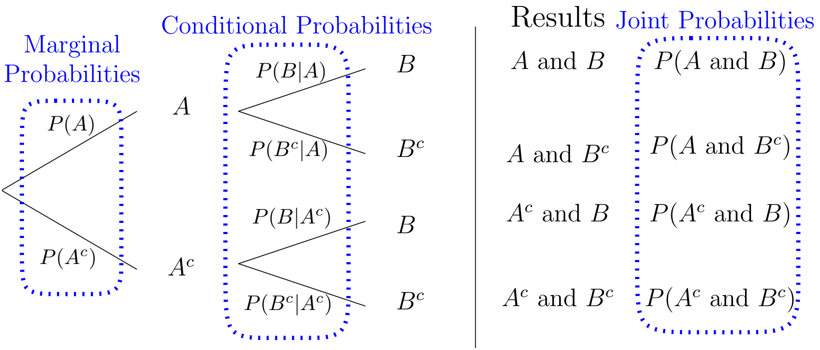
A box contains 2 red balls and 3 green ones.
You pick two balls **without replacement**.



Results	Probability
2 RED balls	$\frac{2}{5} \times \frac{1}{4}$
1 RED 1 GREEN	$\frac{2}{5} \times \frac{3}{4}$
1 GREEN 1 RED	$\frac{3}{5} \times \frac{2}{4}$
2 GREEN balls	$\frac{3}{5} \times \frac{2}{4}$

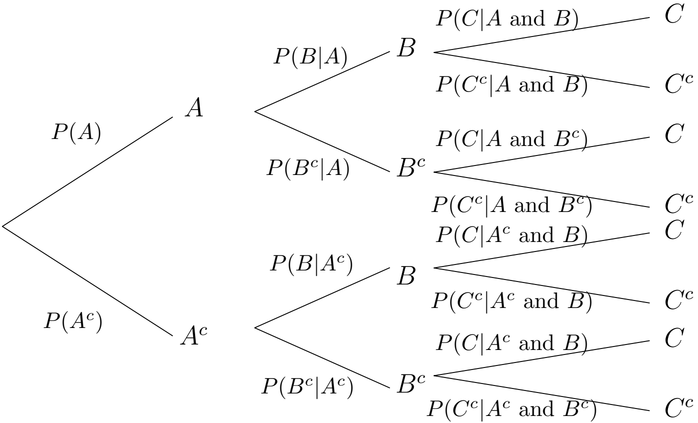
$$P(2 \text{ red}) = \frac{2}{20}, \quad P(1 \text{ red and } 1 \text{ green}) = \frac{12}{20}, \quad P(2 \text{ green}) = \frac{6}{20}.$$

Tree Diagram: General Case



Bigger Tree Diagrams?

If you deal with tree events A, B and C, you can still make a tree.

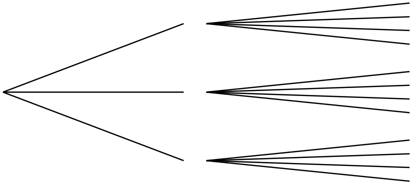


Examples:

- Draw 3 balls/cards without replacement
- Studying the weather of 3 consecutive days
- ...

Richer Tree Diagrams?

If you deal with variables having more than 2 outcomes, you can branch more broadly.



Examples:

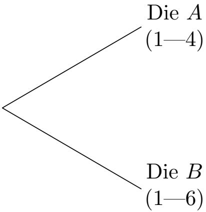
- Outcome of a die roll (6 branches)
- Color of a ball in a box containing red,green and blue ones (3 branches)
- ...

Exercise

Consider a board game in which you need to kill an enemy that has 3 health points. You begin by spinning a dial that comes up A with probability 2/3, and B with probability 1/3. In the event of A, you roll a four-sided die (1—4); In the event of B, you roll a six-sided die (1—6). Your roll is the amount of damage you deal the opponent. What is the probability the enemy dies?

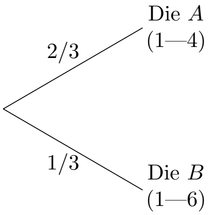
Exercise

Consider a board game in which you need to kill an enemy that has 3 health points. You begin by spinning a dial that comes up A with probability $2/3$, and B with probability $1/3$. In the event of A, you roll a four-sided die (1—4); In the event of B, you roll a six-sided die (1—6). Your roll is the amount of damage you deal the opponent. What is the probability the enemy dies?



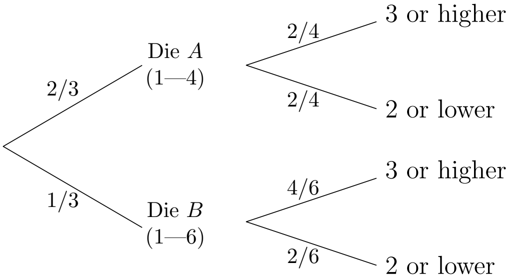
Exercise

Consider a board game in which you need to kill an enemy that has 3 health points. You begin by spinning a dial that comes up A with probability 2/3, and B with probability 1/3. In the event of A, you roll a four-sided die (1—4); In the event of B, you roll a six-sided die (1—6). Your roll is the amount of damage you deal the opponent. What is the probability the enemy dies?



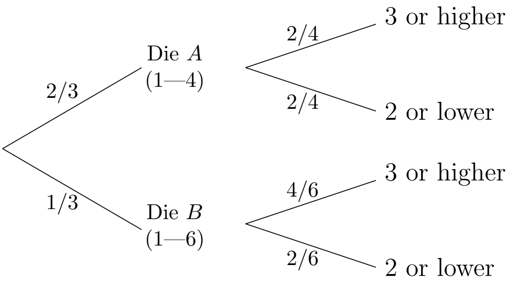
Exercise

Consider a board game in which you need to kill an enemy that has 3 health points. You begin by spinning a dial that comes up A with probability 2/3, and B with probability 1/3. In the event of A, you roll a four-sided die (1—4); In the event of B, you roll a six-sided die (1—6). Your roll is the amount of damage you deal the opponent. What is the probability the enemy dies?



Exercise

Consider a board game in which you need to kill an enemy that has 3 health points. You begin by spinning a dial that comes up A with probability $2/3$, and B with probability $1/3$. In the event of A, you roll a four-sided die (1—4); In the event of B, you roll a six-sided die (1—6). Your roll is the amount of damage you deal the opponent. What is the probability the enemy dies?

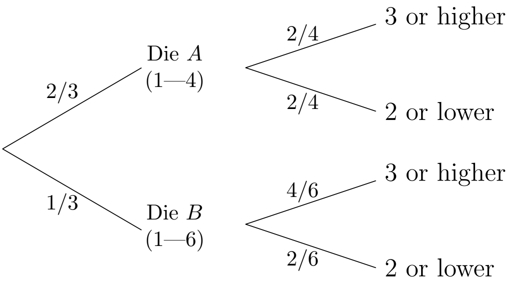


$$\frac{2}{3} \times \frac{2}{4} = \frac{4}{12}$$

$$\frac{1}{3} \times \frac{4}{6} = \frac{4}{18}$$

Exercise

Consider a board game in which you need to kill an enemy that has 3 health points. You begin by spinning a dial that comes up A with probability 2/3, and B with probability 1/3. In the event of A, you roll a four-sided die (1—4); In the event of B, you roll a six-sided die (1—6). Your roll is the amount of damage you deal the opponent. What is the probability the enemy dies?



$$\frac{2}{3} \times \frac{2}{4} = \frac{4}{12}$$

$$\frac{1}{3} \times \frac{4}{6} = \frac{4}{18}$$

$$P(\text{enemy dies}) = \frac{4}{12} + \frac{4}{18} = \frac{5}{9}.$$

Proceeding Without a Tree Diagram

Find the probability of pulling two red cards from a standard deck of 52 cards on your first two pulls.

Proceeding Without a Tree Diagram

Find the probability of pulling two red cards from a standard deck of 52 cards on your first two pulls.

- Let A be the event that the first card is red.
- Let B be the event that the second card is red.

Proceeding Without a Tree Diagram

Find the probability of pulling two red cards from a standard deck of 52 cards on your first two pulls.

- Let A be the event that the first card is red.
- Let B be the event that the second card is red.

Note that A and B are dependent events. Use a tree diagram or conditional probability.

Proceeding Without a Tree Diagram

Find the probability of pulling two red cards from a standard deck of 52 cards on your first two pulls.

- Let A be the event that the first card is red.
- Let B be the event that the second card is red.

Note that A and B are dependent events. Use a tree diagram or conditional probability.

From before, $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)}$,
so $P(A \text{ and } B) = P(A) \times P(B|A)$

Proceeding Without a Tree Diagram

Find the probability of pulling two red cards from a standard deck of 52 cards on your first two pulls.

- Let A be the event that the first card is red.
- Let B be the event that the second card is red.

Note that A and B are dependent events. Use a tree diagram or conditional probability.

From before, $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)}$,
so $P(A \text{ and } B) = P(A) \times P(B|A)$

$$\begin{aligned} P(A \text{ and } B) &= P(A) \times P(B|A) \\ &= \frac{26}{52} \times \frac{25}{51} \simeq 24.5\% \end{aligned}$$

“and” rule: General Case

- If two events A and B are independent,

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

"and" rule: General Case

- If two events A and B are <u>independent</u>,

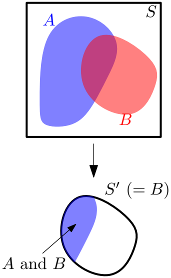
$$P(A \text{ and } B) = P(A) \times P(B).$$

- In general,

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

Can also write

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



“and” rule: General Case

- If two events A and B are independent,

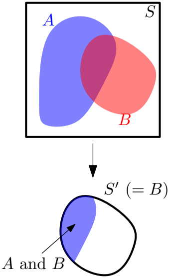
$$P(A \text{ and } B) = P(A) \times P(B).$$

- In general,

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

Can also write

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



Remark: The second formula coincides with the first one when A and B are independent, since $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Bayes' Rule

Bayes' Rule

Remember that

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

Bayes' Rule

Remember that

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

Switching the roles of A and B

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

Bayes' Rule

Remember that

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

Switching the roles of A and B

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

Since $P(B \text{ and } A) = P(A \text{ and } B)$, we get

$$P(A|B) \times P(B) = P(B|A) \times P(A).$$

Bayes' Rule

Remember that

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

Switching the roles of A and B

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

Since $P(B \text{ and } A) = P(A \text{ and } B)$, we get

$$P(A|B) \times P(B) = P(B|A) \times P(A).$$

Dividing by $P(B)$ gives **Bayes' rule**:

$$\boxed{P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}.}$$

Bayes' Rule: Example

In a Math 11 class, 30% of students got an A on Midterm I. 40% of students studied a long time. Suppose the probability that you got an A given that you studied a long time is 70%.
What is the probability that you studied a long time if you got an A?

Bayes' Rule: Example

In a Math 11 class, 30% of students got an A on Midterm I. 40% of students studied a long time. Suppose the probability that you got an A given that you studied a long time is 70%.
What is the probability that you studied a long time if you got an A?

$$\begin{aligned} A &= \text{“earning an A grade”} \\ S &= \text{“studying a long time”} \end{aligned}$$

Bayes' Rule: Example

In a Math 11 class, 30% of students got an A on Midterm I. 40% of students studied a long time. Suppose the probability that you got an A given that you studied a long time is 70%.
What is the probability that you studied a long time if you got an A?

$$\begin{aligned} A &= \text{“earning an A grade”} \\ S &= \text{“studying a long time”} \end{aligned}$$

$P(A) = 0.3$

Bayes' Rule: Example

In a Math 11 class, 30% of students got an A on Midterm I. 40% of students studied a long time. Suppose the probability that you got an A given that you studied a long time is 70%.
What is the probability that you studied a long time if you got an A?

$$A = \text{ "earning an A grade"}$$
$$S = \text{ "studying a long time"}$$

$P(A) = 0.3$ $\qquad$ $P(S) = 0.4$

Bayes' Rule: Example

In a Math 11 class, 30% of students got an A on Midterm I. 40% of students studied a long time. Suppose the probability that you got an A given that you studied a long time is 70%.
What is the probability that you studied a long time if you got an A?

$$\begin{aligned} A &= \text{"earning an A grade"} \\ S &= \text{"studying a long time"} \end{aligned}$$

$P(A) = 0.3$ $\qquad$ $P(S) = 0.4$ $\qquad$ $P(A|S) = 0.7$

Bayes' Rule: Example

In a Math 11 class, 30% of students got an A on Midterm I. 40% of students studied a long time. Suppose the probability that you got an A given that you studied a long time is 70%.
What is the probability that you studied a long time if you got an A?

$$\begin{aligned} A &= \text{“earning an A grade”} \\ S &= \text{“studying a long time”} \end{aligned}$$

$P(A) = 0.3$ $\qquad$ $P(S) = 0.4$ $\qquad$ $P(A|S) = 0.7$

$$P(S|A) =$$

Bayes' Rule: Example

In a Math 11 class, 30% of students got an A on Midterm I. 40% of students studied a long time. Suppose the probability that you got an A given that you studied a long time is 70%.
What is the probability that you studied a long time if you got an A?

$$\begin{aligned} A = &\text{ “earning an A grade”} \\ S = &\text{ “studying a long time”} \end{aligned}$$

$P(A) = 0.3$ $\quad$ $P(S) = 0.4$ $\quad$ $P(A|S) = 0.7$

$$P(S|A) = \frac{P(A|S) \times P(S)}{P(A)} =$$

Bayes' Rule: Example

In a Math 11 class, 30% of students got an A on Midterm I. 40% of students studied a long time. Suppose the probability that you got an A given that you studied a long time is 70%.
What is the probability that you studied a long time if you got an A?

$$\begin{aligned} A &= \text{“earning an A grade”} \\ S &= \text{“studying a long time”} \end{aligned}$$

$P(A) = 0.3$ $\qquad$ $P(S) = 0.4$ $\qquad$ $P(A|S) = 0.7$

$$P(S|A) = \frac{P(A|S) \times P(S)}{P(A)} = \frac{0.7 \times 0.4}{0.3} \simeq 0.93.$$

Bayes' Rule: Level 2

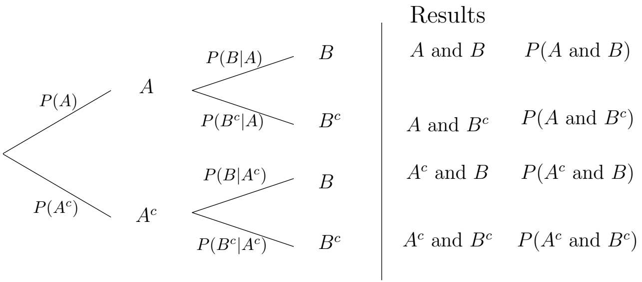
One cannot use only Bayes' rule if $P(B)$ is unknown:

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

Bayes' Rule: Level 2

One cannot use only Bayes' rule if $P(B)$ is unknown:

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



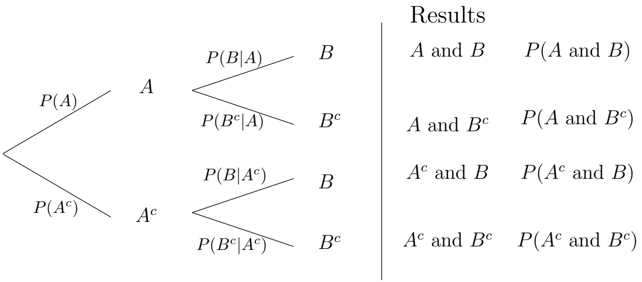
Looking at the tree, we see that:

$$P(B) = P(B|A) \times P(A) + P(B|A^c) \times P(A^c).$$

Bayes' Rule: Level 2

One cannot use only Bayes' rule if $P(B)$ is unknown:

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



Looking at the tree, we see that:

$$P(B) = P(B|A) \times P(A) + P(B|A^c) \times P(A^c).$$

(Exercise for CS/Math majors: prove this rigorously!)

Bayes' Rule: Level 2

We just derived the **advanced Bayes' rule**:

$$\boxed{P(A|B) = \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A^c) \times P(A^c)}.}$$

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

Name your events:

$+$	"the test gives a positive result"
-	"the test gives a negative result"
yes	"the person does have HIV"
no	"the person does not have HIV"

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

Name your events:

$+$	"the test gives a positive result"
-	"the test gives a negative result"
yes	"the person does have HIV"
no	"the person does not have HIV"

The questions amounts to find $P(yes|+)$.

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

$P(+|yes) = 0.999$

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

$P(+|yes) = 0.999$ $\qquad$ $P(\text{-}|no) = 0.99$

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

$P(+|yes) = 0.999$ $\qquad$ $P(\text{-}|no) = 0.99$ $\qquad$ $P(yes) = 0.000172$

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

$P(+|yes) = 0.999$ $\qquad P(\text{-}|no) = 0.99$ $\qquad P(yes) = 0.000172$

$$P(yes|+) =$$

Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

$P(+|yes) = 0.999$ $\quad$ $P(\text{-}|no) = 0.99$ $\quad$ $P(yes) = 0.000172$

$$P(yes|+) = \frac{P(+|yes) \times P(yes)}{P(+)} =$$

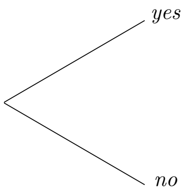
Another Example

Doctors wanted to study the effectiveness of a new HIV test. When they used the test on people known to have HIV, it gave a positive result 99.9% of the time. On those known to not have HIV, it gave negative result 99% of the time. Supposing someone from San Diego (HIV rate: 0.0172%) tests positive on this new test, what is the probability the person actually has HIV?

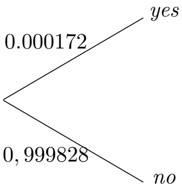
$P(+|yes) = 0.999 \qquad P(\text{-}|no) = 0.99 \qquad P(yes) = 0.000172$

$$P(yes|+) = \frac{P(+|yes) \times P(yes)}{P(+)} = \frac{0.999 \times 0.000172}{??}$$

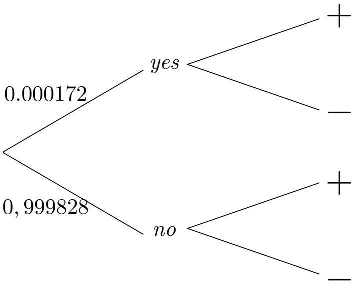
Another Example



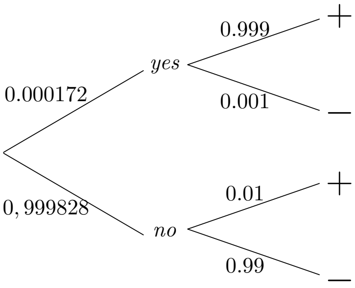
Another Example



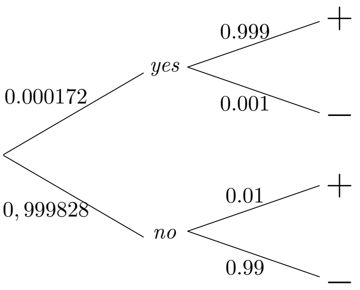
Another Example



Another Example



Another Example

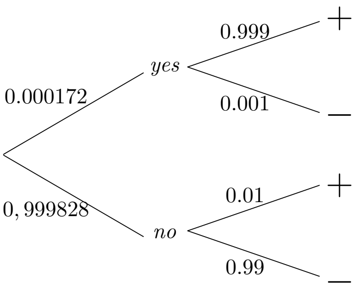


$$0.000172 \times 0.999 = 0,000171828$$

$$0.999828 \times 0.01 = 0.00999828$$

$$P(yes|+) =$$

Another Example

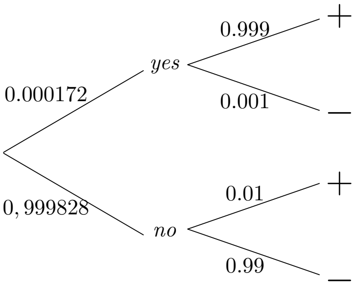


$$0.000172 \times 0.999 = 0,000171828$$

$$0.999828 \times 0.01 = 0.00999828$$

$$P(yes|+) = \frac{P(+|yes) \times P(yes)}{P(+)} =$$

Another Example



$0.000172 \times 0.999 = 0,000171828$

$0.999828 \times 0.01 = 0.00999828$

$$P(yes|+) = \frac{P(+|yes) \times P(yes)}{P(+)} = \frac{0.999 \times 0.000172}{0.000171828 + 0.00999828} \simeq 1.7\%.$$

What's Happening There?

How is it that a test so accurate could give such a terrible result?

What's Happening There?

How is it that a test so accurate could give such a terrible result?

Here, we have a great mismatch between

- The accuracy of the test (only 99.9% and 99%)
- The extreme rareness of the disease (0.0172%) in the population.

For positive test results to be useful (that is, for $\mathrm{P}(\text{yes}|+)$ to be high), you need the orders of magnitude of "test accuracy" and "disease prevalence" to be better matched.

Dependence/Independence in Small/Large Populations

Dependence/Independence in Small/Large Populations

You pick 5 balls in a box containing green and red balls balls.
Study the event $A =$ "all 5 balls are green".

- 3 red 7 green (10 total)
  - **With replacement:** $P(A) = \left(\frac{7}{10}\right)^5 \simeq 0.168.$
  - **Without replacement:** $P(A) = \frac{7}{10}\frac{6}{9}\frac{5}{8}\frac{4}{7}\frac{3}{6} \simeq 0.083.$

Dependence/Independence in Small/Large Populations

You pick 5 balls in a box containing green and red balls balls.
Study the event $A =$ "all 5 balls are green".

- 3 red 7 green (10 total)
  - **With replacement:** $P(A) = \left(\frac{7}{10}\right)^5 \simeq 0.168.$
  - **Without replacement:** $P(A) = \frac{7}{10}\frac{6}{9}\frac{5}{8}\frac{4}{7}\frac{3}{6} \simeq 0.083.$
- 30 red 70 green (100 total)
  - **With replacement:** $P(A) = \left(\frac{70}{100}\right)^5 \simeq 0.168.$
  - **Without replacement:** $P(A) = \frac{70}{100}\frac{69}{99}\frac{68}{98}\frac{67}{97}\frac{66}{96} \simeq 0.161.$

Dependence/Independence in Small/Large Populations

You pick 5 balls in a box containing green and red balls balls.
Study the event $A =$ "all 5 balls are green".

- 3 red 7 green (10 total)
    - **With replacement:** $P(A) = \left(\frac{7}{10}\right)^5 \simeq 0.168.$
    - **Without replacement:** $P(A) = \frac{7}{10}\frac{6}{9}\frac{5}{8}\frac{4}{7}\frac{3}{6} \simeq 0.083.$
- 30 red 70 green (100 total)
    - **With replacement:** $P(A) = \left(\frac{70}{100}\right)^5 \simeq 0.168.$
    - **Without replacement:** $P(A) = \frac{70}{100}\frac{69}{99}\frac{68}{98}\frac{67}{97}\frac{66}{96} \simeq 0.161.$

When you draw a sample without replacement, your choices are always dependent.
<u>Although dependence is still present</u>, it tends to disappear when population is large.

Lesson About Sampling

Moral (to be used later in the course):

If your sample size is less than 10% of the population you are sampling from, you may assume the choices are independent (even though they aren't).

Said differently: for samples $< 10\%$ population size, the distinction between "with replacement" and "without replacement" is unnecessary.