COMP170
Discrete Mathematical Tools for Computer Science

Independence

Version 2.0: Last updated, May 13, 2007

Discrete Math for Computer Science
K. Bogart, C. Stein and R.L. Drysdale
Section 5.3, pp. 236-247

Conditional Probability and Independence

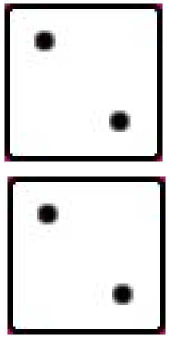
- Conditional Probability

- Independence

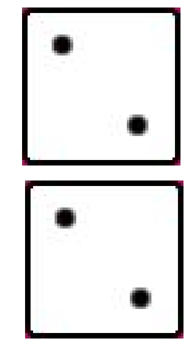
- Independent Trials Processes

Conditional Probability

Suppose we've thrown two fair dice. The probability of seeing "double-twos" is $\frac{1}{36}$.



Conditional Probability

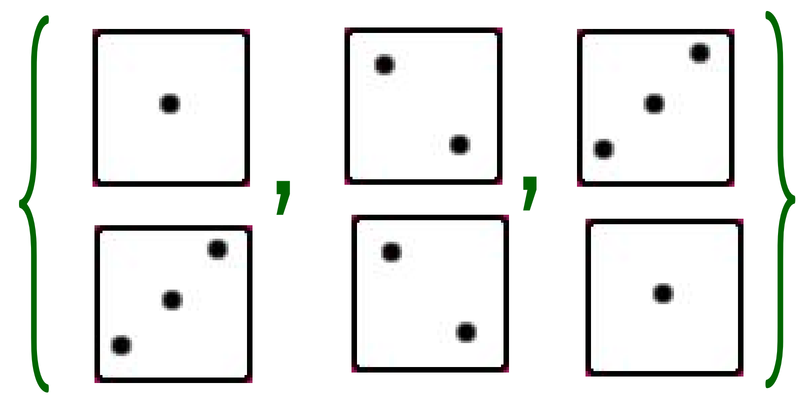


Suppose we've thrown two fair dice. The probability of seeing "double-twos" is $\frac{1}{36}$.

Now suppose that we don't see the dice but know that the event

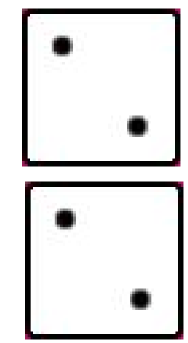
"the dice sum up to 4"

has occured. What is the probability that "double-twos" occurred given that "the dice sum up to 4"?



Conditional Probability

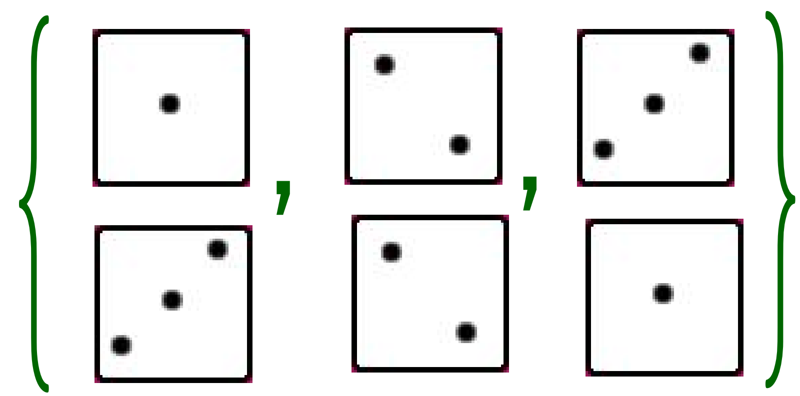
Suppose we've thrown two fair dice. The probability of seeing "double-twos" is $\frac{1}{36}$.



Now suppose that we don't see the dice but know that the event

"the dice sum up to 4"

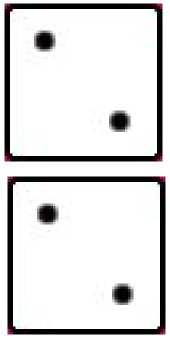
has occured. What is the probability that "double-twos" occurred given that "the dice sum up to 4"?



Answer "should be" $\frac{1}{3}$, shouldn't it?

Conditional Probability

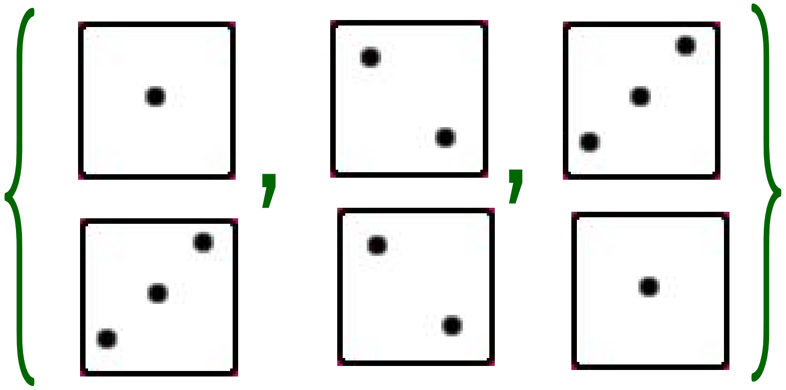
Suppose we've thrown two fair dice. The probability of seeing "double-twos" is $\frac{1}{36}$.



Now suppose that we don't see the dice but know that the event

"the dice sum up to 4"

has occured. What is the probability that "double-twos" occurred given that "the dice sum up to 4"?

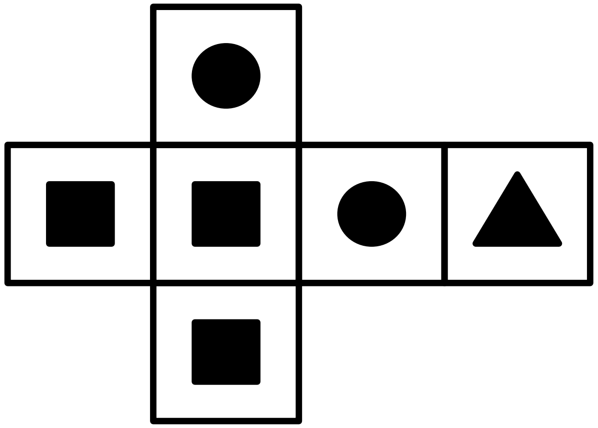


Answer "should be" $\frac{1}{3}$, shouldn't it?

This lecture formalizes this intuition.

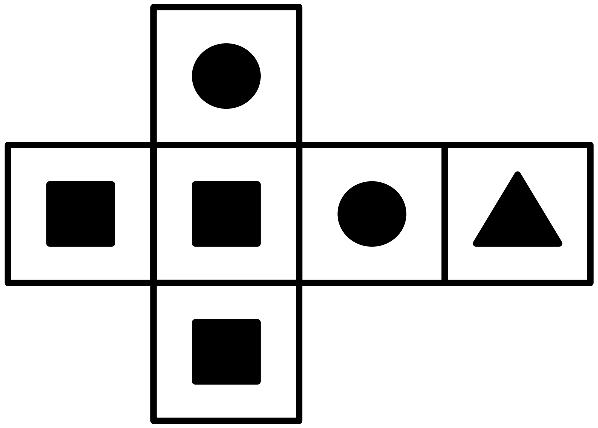
A more complicated example

A more complicated example

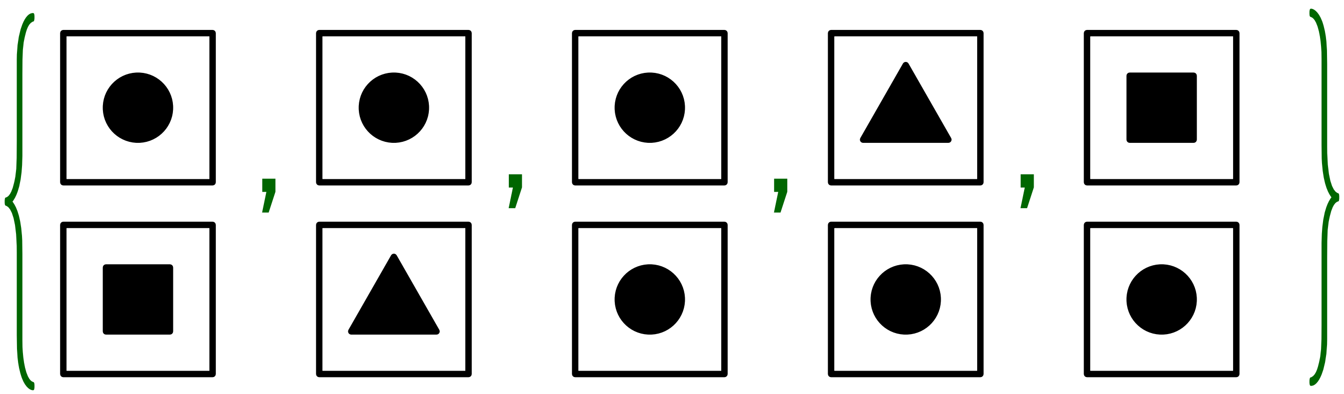
2 dice: 

A more complicated example

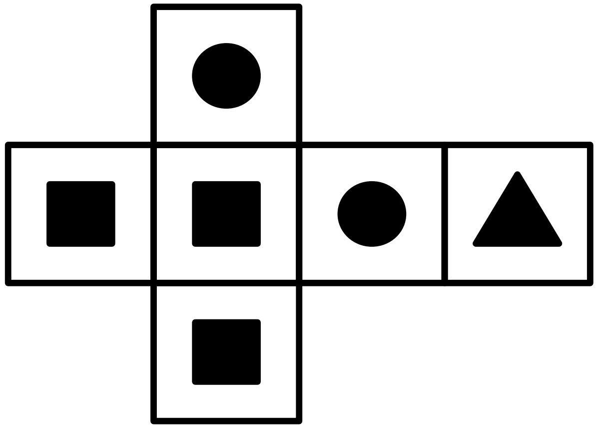
2 dice:



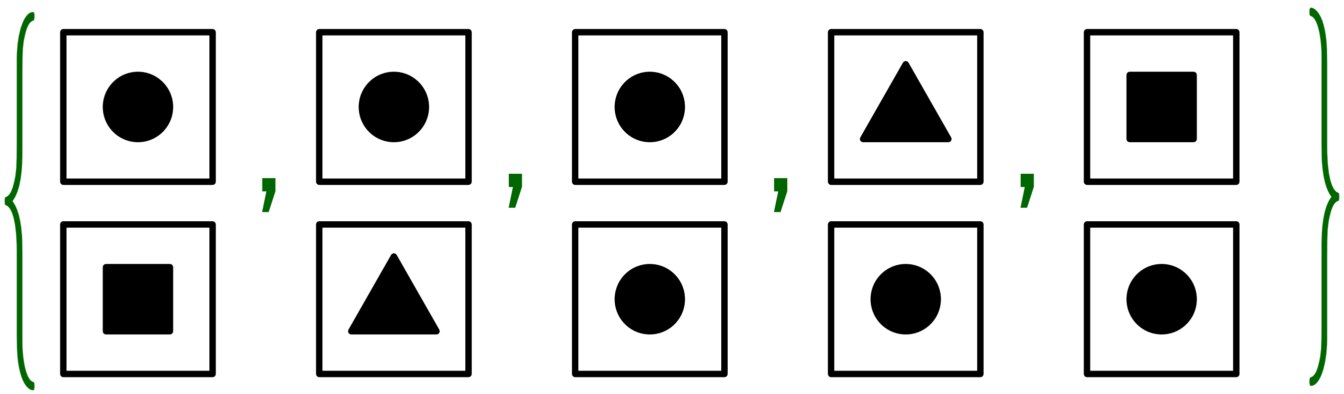
Event *"at least one circle on top"* is:



A more complicated example

2 dice: 

Event *"at least one circle on top"* is:



Applying **principle of inclusion and exclusion**: probability of seeing a circle on at least one top when we roll the dice is

$$\frac{1}{3} + \frac{1}{3} - \frac{1}{9} = \frac{5}{9}$$

Suppose you are told that the two dice have been rolled and *both top shapes are the same*?

Suppose you are told that the two dice have been rolled and *both top shapes are the same*?

What is the probability that at least one top shape (and now therefore both top shapes) is a circle?

Suppose you are told that the two dice have been rolled and *both top shapes are the same*?

What is the probability that at least one top shape (and now therefore both top shapes) is a circle?

Originally, (i) chance of getting (two) circles was 4 times chance of getting (two) triangles and (ii) chance of getting (two) squares was 9 times chance of getting (two) triangles so, given that both top shapes are the same intuitively, we "should" have

Suppose you are told that the two dice have been rolled and *both top shapes are the same*?

What is the probability that at least one top shape (and now therefore both top shapes) is a circle?

Originally, (i) chance of getting (two) circles was 4 times chance of getting (two) triangles and (ii) chance of getting (two) squares was 9 times chance of getting (two) triangles so, given that both top shapes are the same intuitively, we "should" have

$$P(\text{circles}) = 4P(\text{triangles})$$

Suppose you are told that the two dice have been rolled and *both top shapes are the same*?

What is the probability that at least one top shape (and now therefore both top shapes) is a circle?

Originally, (i) chance of getting (two) circles was 4 times chance of getting (two) triangles and (ii) chance of getting (two) squares was 9 times chance of getting (two) triangles so, given that both top shapes are the same intuitively, we "should" have

$$P(\text{circles}) = 4P(\text{triangles})$$

and

$$P(\text{squares}) = 9P(\text{triangles})$$

$$P(\text{circles}) = 4P(\text{triangles})$$

and

$$P(\text{squares}) = 9P(\text{triangles})$$

Let $p = P(\text{triangles})$. Since probabilities sum to 1,

$$P(\text{circles}) = 4P(\text{triangles})$$

and

$$P(\text{squares}) = 9P(\text{triangles})$$

Let $p = P(\text{triangles})$. Since probabilities sum to 1,

$$p + 4p + 9p = 1 \text{ or } p = \frac{1}{14}, \text{ and}$$

$$P(\text{two circles if both tops are the same}) = 4p = \frac{2}{7}.$$

Do these analyses make sense?

Do these analyses make sense?

How can we replace intuitive calculations with a formula that we can use in similar situations?

Do these analyses make sense?

How can we replace intuitive calculations with a formula that we can use in similar situations?

WARNING

There are situations where our intuitive idea of probability does not always agree with what the rules of probability give us!

Rolling our two unusual dice

Original sample space with probabilities

$$\{\texttt{TT},\ \texttt{TC},\ \texttt{TS},\ \texttt{CT},\ \texttt{CC},\ \texttt{CS},\ \texttt{ST},\ \texttt{SC},\ \texttt{SS}\}.$$

$$\frac{1}{36} \quad \frac{1}{18} \quad \frac{1}{12} \quad \frac{1}{18} \quad \frac{1}{9} \quad \frac{1}{6} \quad \frac{1}{12} \quad \frac{1}{6} \quad \frac{1}{4}$$

Rolling our two unusual dice

Original sample space with probabilities

$$\{\texttt{TT},\ \texttt{TC},\ \texttt{TS},\ \texttt{CT},\ \texttt{CC},\ \texttt{CS},\ \texttt{ST},\ \texttt{SC},\ \texttt{SS}\}.$$

$$\frac{1}{36}\quad \frac{1}{18}\quad \frac{1}{12}\quad \frac{1}{18}\quad \frac{1}{9}\quad \frac{1}{6}\quad \frac{1}{12}\quad \frac{1}{6}\quad \frac{1}{4}$$

We know that event $\{\texttt{TT},\ \texttt{CC},\ \texttt{SS}\}$ happened.

Rolling our two unusual dice

Original sample space with probabilities

$$\{\texttt{TT},\ \texttt{TC},\ \texttt{TS},\ \texttt{CT},\ \texttt{CC},\ \texttt{CS},\ \texttt{ST},\ \texttt{SC},\ \texttt{SS}\}.$$

$$\frac{1}{36} \quad \frac{1}{18} \quad \frac{1}{12} \quad \frac{1}{18} \quad \frac{1}{9} \quad \frac{1}{6} \quad \frac{1}{12} \quad \frac{1}{6} \quad \frac{1}{4}$$

We know that event $\{\texttt{TT},\ \texttt{CC},\ \texttt{SS}\}$ happened.

Thus, although this event used to have probability

$$\frac{1}{36} + \frac{1}{9} + \frac{1}{4} = \frac{14}{36} = \frac{7}{18}$$

it now has probability 1.

Rolling our two unusual dice

Original sample space with probabilities

$$\{\texttt{TT},\ \texttt{TC},\ \texttt{TS},\ \texttt{CT},\ \texttt{CC},\ \texttt{CS},\ \texttt{ST},\ \texttt{SC},\ \texttt{SS}\}.$$

$$\frac{1}{36} \quad \frac{1}{18} \quad \frac{1}{12} \quad \frac{1}{18} \quad \frac{1}{9} \quad \frac{1}{6} \quad \frac{1}{12} \quad \frac{1}{6} \quad \frac{1}{4}$$

We know that event $\{\texttt{TT},\ \texttt{CC},\ \texttt{SS}\}$ happened.

Thus, although this event used to have probability

$$\frac{1}{36} + \frac{1}{9} + \frac{1}{4} = \frac{14}{36} = \frac{7}{18}$$

it now has probability 1.

Given this, what probability should we assign event of seeing a circle ($\texttt{CC}$)?

New Sample Space $\{\texttt{TT, CC, SS}\}$

New Sample Space $\{\mathtt{TT},\ \mathtt{CC},\ \mathtt{SS}\}$

Probabilities in old sample space $\frac{1}{36}$ $\frac{1}{9}$ $\frac{1}{4}$

New Sample Space $\{\mathtt{TT},\ \mathtt{CC},\ \mathtt{SS}\}$

Probabilities in old sample space $\frac{1}{36}$ $\frac{1}{9}$ $\frac{1}{4}$ Sum is $\frac{7}{18}$

New Sample Space $\{\texttt{TT, CC, SS}\}$

Probabilities in old sample space $\frac{1}{36}$ $\frac{1}{9}$ $\frac{1}{4}$ Sum is $\frac{7}{18}$

Multiply all three old probabilities by $18/7$:
new probabilities will preserve ratios and sum to 1.

New Sample Space $\{\texttt{TT},\ \texttt{CC},\ \texttt{SS}\}$

Probabilities in old sample space $\frac{1}{36}$ $\frac{1}{9}$ $\frac{1}{4}$ Sum is $\frac{7}{18}$

Multiply all three old probabilities by $18/7$:
new probabilities will preserve ratios and sum to 1.

$$P(\text{two circles}) = \frac{1}{9} \cdot \frac{18}{7} = \frac{2}{7}$$

New Sample Space $\{\mathtt{TT},\ \mathtt{CC},\ \mathtt{SS}\}$

Probabilities in old sample space	$\frac{1}{36}$	$\frac{1}{9}$	$\frac{1}{4}$	Sum is $\frac{7}{18}$
New Probabilities	$\frac{1}{14}$	$\frac{2}{7}$	$\frac{9}{14}$	Sum is 1

Multiply all three old probabilities by $18/7$:
new probabilities will preserve ratios and sum to 1.

$$P(\text{two circles}) = \frac{1}{9} \cdot \frac{18}{7} = \frac{2}{7}$$

New Sample Space	$\{\mathtt{TT},\ \mathtt{CC},\ \mathtt{SS}\}$			
Probabilities in old sample space	$\frac{1}{36}$	$\frac{1}{9}$	$\frac{1}{4}$	Sum is $\frac{7}{18}$
New Probabilities	$\frac{1}{14}$	$\frac{2}{7}$	$\frac{9}{14}$	Sum is 1

Multiply all three old probabilities by $18/7$:
new probabilities will preserve ratios and sum to 1.

$$P(\text{two circles}) = \frac{1}{9} \cdot \frac{18}{7} = \frac{2}{7}$$

We now capture this reasoning process in a formula!

Definition:

The **conditional probability** of E given F, denoted by $P(E|F)$ *(read as "the probability of E given F")* is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Definition:

The **conditional probability** of E given F, denoted by $P(E|F)$ *(read as "the probability of E given F")* is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Example: From previous page

Definition:

The **conditional probability** of E given F, denoted by $P(E|F)$
(read as "the probability of E given F") is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Example: From previous page

F is event $\{\mathtt{TT},\ \mathtt{CC},\ \mathtt{SS}\}$

Definition:

The **conditional probability** of E given F, denoted by $P(E|F)$ *(read as "the probability of E given F")* is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Example: From previous page

F is event $\{\texttt{TT}, \texttt{CC}, \texttt{SS}\}$

E is event a circle on top

Definition:

The **conditional probability** of E given F,
denoted by $P(E|F)$
(read as "the probability of E given F") is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Example: From previous page

F is event $\{\texttt{TT, CC, SS}\}$

E is event a circle on top

$$E \cap F = \{CC\}$$

Definition:

The **conditional probability** of E given F, denoted by $P(E|F)$ *(read as "the probability of E given F")* is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Example: From previous page

F is event $\{\texttt{TT, CC, SS}\}$

$P(F) = \frac{7}{18}$

E is event a circle on top

$$E \cap F = \{CC\}$$

$P(E \cap F) = \frac{1}{9}$

Definition:

The **conditional probability** of E given F, denoted by $P(E|F)$ *(read as "the probability of E given F")* is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Example: From previous page

F is event $\{\texttt{TT, CC, SS}\}$

E is event a circle on top

$$E \cap F = \{CC\}$$

$$P(F) = \tfrac{7}{18}$$

$$P(E \cap F) = \tfrac{1}{9}$$

$$\Rightarrow \quad P(E|F) = \frac{1}{9} / \frac{7}{18} = \frac{2}{14}$$

Definition:

The **conditional probability** of E given F, denoted by $P(E|F)$ *(read as "the probability of E given F")* is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Note: This definition doesn't make sense when $P(F) = 0$.
In this case we define $P(E|F) = E$.
This makes sense, since if event F can not occur then it occuring gives us no information (since this can't happen).

Example

Example

When we roll two ordinary dice, what is the
probability that the sum is even,
given that the sum is greater than or equal to 10?

Example

When we roll two ordinary dice, what is the
probability that the sum is even,
given that the sum is greater than or equal to 10?

Sample space is ordered pairs, each of weight $1/36$.

Example

When we roll two ordinary dice, what is the probability that the sum is even, given that the sum is greater than or equal to 10?

Sample space is ordered pairs, each of weight $1/36$.

Let E be event that "sum is even".

Example

When we roll two ordinary dice, what is the probability that the sum is even, given that the sum is greater than or equal to 10?

Sample space is ordered pairs, each of weight $1/36$.

Let E be event that "sum is even".

Let F be event that "sum is ≥ 10".

Example

When we roll two ordinary dice, what is the probability that the sum is even, given that the sum is greater than or equal to 10?

Sample space is ordered pairs, each of weight $1/36$.

Let E be event that "sum is even". $P(F) = 1/6$

Let F be event that "sum is ≥ 10".

Example

When we roll two ordinary dice, what is the probability that the sum is even, given that the sum is greater than or equal to 10?

Sample space is ordered pairs, each of weight $1/36$.

Let E be event that "sum is even". $P(F) = 1/6$

Let F be event that "sum is ≥ 10".

$E \cap F$ is the event that "sum is either 10 or 12".

Example

When we roll two ordinary dice, what is the probability that the sum is even, given that the sum is greater than or equal to 10?

Sample space is ordered pairs, each of weight $1/36$.

Let E be event that "sum is even". $P(F) = 1/6$

Let F be event that "sum is ≥ 10".

$E \cap F$ is the event that "sum is either 10 or 12". $P(E \cap F) = 1/9$

Example

When we roll two ordinary dice, what is the probability that the sum is even, given that the sum is greater than or equal to 10?

Sample space is ordered pairs, each of weight $1/36$.

Let E be event that "sum is even". $P(F) = 1/6$

Let F be event that "sum is ≥ 10".

$E \cap F$ is the event that "sum is either 10 or 12". $P(E \cap F) = 1/9$

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{1/9}{1/6} = \frac{2}{3}.$$

Sometimes we are explicitly given information about conditional probabilities:

Sometimes we are explicitly given information about conditional probabilities:

Example

Sometimes we are explicitly given information about conditional probabilities:

Example

If a student knows 80% of the material in a course, you expect her to get a grade of around 80% on a well designed exam.

Sometimes we are explicitly given information about conditional probabilities:

Example

If a student knows 80% of the material in a course, you expect her to get a grade of around 80% on a well designed exam.

What is the probability that she answers a question correctly on a 100-question **true-false** test if she guesses at each question for which she does not know the answer?

Sometimes we are explicitly given information about conditional probabilities:

Example

If a student knows 80% of the material in a course, you expect her to get a grade of around 80% on a well designed exam.

What is the probability that she answers a question correctly on a 100-question **true-false** test if she guesses at each question for which she does not know the answer?

The answer might surprise you!

$R =$ she gets the right answer.

$R =$ she gets the right answer.

$K =$ she knows that right answer.

$R =$ she gets the right answer.

$K =$ she knows that right answer.

$\overline{K} =$ she guesses.

$R =$ she gets the right answer.

$K =$ she knows that right answer.

$\overline{K} =$ she guesses.

Then $R = (R \cap K) \cup (R \cap \overline{K})$.

$R =$ she gets the right answer.

$K =$ she knows that right answer.

$\overline{K} =$ she guesses.

Then $R = (R \cap K) \cup (R \cap \overline{K})$.

$P(\text{she gets the right answer} \mid \text{she knows the answer}) = P(R|K) = 1.$

$R =$ she gets the right answer.

$K =$ she knows that right answer.

$\overline{K} =$ she guesses.

Then $R = (R \cap K) \cup (R \cap \overline{K})$.

$P($she gets the right answer | she knows the answer$) = P(R|K) = 1.$

$P($she gets the right answer | she does not know the answer$)$
$$= P(R|\overline{K}) = \tfrac{1}{2}.$$

$R =$ she gets the right answer.

$K =$ she knows that right answer. $P(K) = .8$

$\overline{K} =$ she guesses. $P(\overline{K}) = .2$

Then $R = (R \cap K) \cup (R \cap \overline{K})$.

$P(\text{she gets the right answer} \mid \text{she knows the answer}) = P(R|K) = 1$.

$P(\text{she gets the right answer} \mid \text{she does not know the answer})$
$= P(R|\overline{K}) = \frac{1}{2}$.

$R =$ she gets the right answer.

$K =$ she knows that right answer. $P(K) = .8$

$\overline{K} =$ she guesses. $P(\overline{K}) = .2$

Then $R = (R \cap K) \cup (R \cap \overline{K})$.

$P($she gets the right answer $|$ she knows the answer$) = \boxed{P(R|K) = 1}$.

$P($she gets the right answer $|$ she does not know the answer$)$

$= \boxed{P(R|\overline{K}) = \frac{1}{2}}$.

Use conditional probabilities:

$R =$ she gets the right answer.

$K =$ she knows that right answer. $\quad P(K) = .8$

$\overline{K} =$ she guesses. $\quad P(\overline{K}) = .2$

Then $R = (R \cap K) \cup (R \cap \overline{K})$.

$P($she gets the right answer | she knows the answer$) = \boxed{P(R|K) = 1}$.

$P($she gets the right answer | she does not know the answer$)$
$= \boxed{P(R|\overline{K}) = \frac{1}{2}}$.

Use conditional probabilities:

$$\begin{aligned} P(R) &= P(R \cap K) + P\left(R \cap \overline{K}\right) \\ &= P(R|K)P(K) + P\left(R|\overline{K}\right) P\left(\overline{K}\right) \\ &= 1 \cdot .8 + .5 \cdot .2 = .9. \end{aligned}$$

$R =$ she gets the right answer.

$K =$ she knows that right answer. $P(K) = .8$

$\overline{K} =$ she guesses. $P(\overline{K}) = .2$

Then $R = (R \cap K) \cup (R \cap \overline{K})$.

$P($she gets the right answer | she knows the answer$) = \boxed{P(R|K) = 1}$.

$P($she gets the right answer | she does not know the answer$)$
$= \boxed{P(R|\overline{K}) = \frac{1}{2}}$.

Use conditional probabilities:

$$\begin{aligned} P(R) &= P(R \cap K) + P\left(R \cap \overline{K}\right) \\ &= P(R|K)P(K) + P\left(R|\overline{K}\right) P\left(\overline{K}\right) \\ &= 1 \cdot .8 + .5 \cdot .2 = .9. \end{aligned}$$

Which implies that she wil get a 90% on the exam!

Conditional Probability and Independence

- Conditional Probability
- Independence
- Independent Trials Processes

E is **independent** of F if $\quad P(E|F) = P(E)$.

$\boxed{E \text{ is } \textbf{independent} \text{ of } F \text{ if } \quad P(E|F) = P(E).}$

Example

When we roll two dice, one red and one green,
$E =$ "total sum is odd" is independent of
$F =$ "red dice shows an odd number of dots".

$\boxed{E \text{ is } \textbf{independent} \text{ of } F \text{ if} \quad P(E|F) = P(E).}$

Example

When we roll two dice, one red and one green,
$E =$ "total sum is odd" is independent of
$F =$ "red dice shows an odd number of dots".

$$P(E) = P(\text{total sum is odd}) = \frac{1}{2}.$$

$\boxed{E \text{ is } \textbf{independent} \text{ of } F \text{ if } \quad P(E|F) = P(E).}$

Example

When we roll two dice, one red and one green,
$E =$ "total sum is odd" is independent of
$F =$ "red dice shows an odd number of dots".

$$P(E) = P(\text{total sum is odd}) = \tfrac{1}{2}.$$

$P(E|F)$

$\boxed{E \text{ is } \textbf{independent} \text{ of } F \text{ if } \quad P(E|F) = P(E).}$

Example

When we roll two dice, one red and one green,
$E =$ "total sum is odd" is independent of
$F =$ "red dice shows an odd number of dots".

$$P(E) = P(\text{total sum is odd}) = \tfrac{1}{2}.$$

$$P(E|F) = P(\text{total sum is odd} \mid \text{red die is odd})$$

$\boxed{E \text{ is } \textbf{independent} \text{ of } F \text{ if} \quad P(E|F) = P(E).}$

Example

When we roll two dice, one red and one green,
$E =$ "total sum is odd" is independent of
$F =$ "red dice shows an odd number of dots".

$$P(E) = P(\text{total sum is odd}) = \tfrac{1}{2}.$$

$$\begin{aligned} P(E|F) &= P(\text{total sum is odd} \mid \text{red die is odd}) \\ &= P(\text{green die is even}). \\ &= \tfrac{3}{6} = \tfrac{1}{2} \end{aligned}$$

$\boxed{E \text{ is } \textbf{independent} \text{ of } F \text{ if } \quad P(E|F) = P(E).}$

Example

When we roll two dice, one red and one green,
$E =$ "total sum is odd" is independent of
$F =$ "red dice shows an odd number of dots".

$$P(E) = P(\text{total sum is odd}) = \frac{1}{2}.$$

$$\begin{aligned} P(E|F) &= P(\text{total sum is odd} \mid \text{red die is odd}) \\ &= P(\text{green die is even}). \\ &= \frac{3}{6} = \frac{1}{2} \end{aligned}$$

Thus, by definition of independence, "total sum is odd" and "red dice shows an odd number of dots" are independent.

Theorem 5.5
(Product Principle for Independent Probabilities)
Suppose E and F are events in a sample space. Then

E is independent of F if and only if $P(E \cap F) = P(E)P(F)$

Theorem 5.5
(Product Principle for Independent Probabilities)
Suppose E and F are events in a sample space. Then

E is independent of F if and only if $P(E \cap F) = P(E)P(F)$

Proof:

Theorem 5.5
(Product Principle for Independent Probabilities)
Suppose E and F are events in a sample space. Then

E is independent of F if and only if $P(E \cap F) = P(E)P(F)$

Proof:

Case 1: F is empty.

Theorem 5.5
(Product Principle for Independent Probabilities)
Suppose E and F are events in a sample space. Then

E is independent of F if and only if $P(E \cap F) = P(E)P(F)$

Proof:

Case 1: F is empty.

Then $P(E) = P(E|F)$ so E is independent of F.

Theorem 5.5
(Product Principle for Independent Probabilities)
Suppose E and F are events in a sample space. Then

E is independent of F if and only if $P(E \cap F) = P(E)P(F)$

Proof:

Case 1: F is empty.

Then $P(E) = P(E|F)$ so E is independent of F.

Also $P(E)P(F) = 0 = P(E \cap F)$.

Theorem 5.5
(Product Principle for Independent Probabilities)
Suppose E and F are events in a sample space. Then

E is independent of F if and only if $P(E \cap F) = P(E)P(F)$

Proof:

Case 1: F is empty.

Then $P(E) = P(E|F)$ so E is independent of F.

Also $P(E)P(F) = 0 = P(E \cap F)$.

So in this case,
E is independent of F and $P(E \cap F) = P(E)P(F)$.

Case 2: F is nonempty.

Case 2: F is nonempty.

E is independent of F $\quad \Leftrightarrow \quad P(E|F) = P(E)$.

Case 2: F is nonempty.

E is independent of F $\quad \Leftrightarrow \quad P(E|F) = P(E)$.

Starting with RHS:

$$P(E|F) = P(E)$$

Case 2: F is nonempty.

E is independent of F $\quad\Leftrightarrow\quad P(E|F) = P(E)$.

Starting with RHS:

$$P(E|F) = P(E)$$

$$\Leftrightarrow \frac{P(E \cap F)}{P(F)} = P(E) \qquad \text{(by definition)}$$

Case 2: F is nonempty.

E is independent of F $\quad\Leftrightarrow\quad$ $P(E|F) = P(E)$.

Starting with RHS:

$$P(E|F) = P(E)$$

$$\Leftrightarrow \frac{P(E \cap F)}{P(F)} = P(E) \qquad \text{(by definition)}$$

$$\Leftrightarrow P(E \cap F) = P(E)P(F)$$

Case 2: F is nonempty.

E is independent of $F \quad \Leftrightarrow \quad P(E|F) = P(E)$.

Starting with RHS:

$$P(E|F) = P(E)$$

$$\Leftrightarrow \frac{P(E \cap F)}{P(F)} = P(E) \qquad \text{(by definition)}$$

$$\Leftrightarrow P(E \cap F) = P(E)P(F)$$

So in this case as well,
E is independent of F if and only if $P(E \cap F) = P(E)P(F)$.

Theorem 5.5
(Product Principle for Independent Probabilities)
Suppose E and F are events in a sample space. Then

E is independent of F if and only if $P(E \cap F) = P(E)P(F)$

Corollary 5.6
E is independent of F if and only if F is independent of E.

Theorem 5.5
(Product Principle for Independent Probabilities)
Suppose E and F are events in a sample space. Then

E is independent of F if and only if $P(E \cap F) = P(E)P(F)$

Corollary 5.6
E is independent of F if and only if F is independent of E.

Example:
We already saw, when throwing a red, green die
that "total sum is odd" is independent of "red die is odd"
$\Rightarrow$ "red die is odd" is independent of "total sum is odd".

Coin Flipping

Coin Flipping

When flipping a coin twice, we think of second outcome as being independent of first.

Coin Flipping

When flipping a coin twice, we think of second outcome as being independent of first.

Does definition of independence capture this intuitive idea? Let's compute relevant probabilities to see if it does!

Flipping a coin twice

Flipping a coin twice

Sample space with their probabilities

$$\{\texttt{HH},\ \texttt{HT},\ \texttt{TH},\ \texttt{TT}\}.$$

$$\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4}$$

Flipping a coin twice

Sample space with their probabilities

$$\{\texttt{HH},\ \texttt{HT},\ \texttt{TH},\ \texttt{TT}\}.$$
$$\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4}$$

$P(\texttt{H}\text{ first}) = 1/4 + 1/4 = 1/2$

Flipping a coin twice

Sample space with their probabilities

$$\{\texttt{HH},\ \texttt{HT},\ \texttt{TH},\ \texttt{TT}\}.$$

$$\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4}$$

$$P(\texttt{H}\ \text{first}) = 1/4 + 1/4 = 1/2$$

$$P(\texttt{H}\ \text{second}) = 1/4 + 1/4 = 1/2$$

Flipping a coin twice

Sample space with their probabilities

$$\{\texttt{HH},\ \texttt{HT},\ \texttt{TH},\ \texttt{TT}\}.$$

$$\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4}$$

$$P(\texttt{H}\ \text{first}) = 1/4 + 1/4 = 1/2$$

$$P(\texttt{H}\ \text{second}) = 1/4 + 1/4 = 1/2$$

$$P(\texttt{H}\ \text{first and}\ \texttt{H}\ \text{second}) = 1/4$$

Flipping a coin twice

Sample space with their probabilities $\{\mathtt{HH},\ \mathtt{HT},\ \mathtt{TH},\ \mathtt{TT}\}.$

$$\begin{matrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{matrix}$$

$$P(\mathtt{H}\text{ first}) = 1/4 + 1/4 = 1/2$$

$$P(\mathtt{H}\text{ second}) = 1/4 + 1/4 = 1/2$$

$$P(\mathtt{H}\text{ first and }\mathtt{H}\text{ second}) = 1/4$$

$$P(\mathtt{H}\text{ first})P(\mathtt{H}\text{ second}) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} = P(\mathtt{H}\text{ first and }\mathtt{H}\text{ second}).$$

Flipping a coin twice

Sample space with their probabilities $\{\texttt{HH},\ \texttt{HT},\ \texttt{TH},\ \texttt{TT}\}.$

$$\frac{1}{4} \qquad \frac{1}{4} \qquad \frac{1}{4} \qquad \frac{1}{4}$$

$P(\texttt{H}\text{ first}) = 1/4 + 1/4 = 1/2$

$P(\texttt{H}\text{ second}) = 1/4 + 1/4 = 1/2$

$P(\texttt{H}\text{ first and }\texttt{H}\text{ second}) = 1/4$

$P(\texttt{H}\text{ first})P(\texttt{H}\text{ second}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(\texttt{H}\text{ first and }\texttt{H}\text{ second}).$

By Theorem 5.5, "H second" is independent of "H first".

Flipping a coin twice

Sample space with their probabilities $\{\mathtt{HH},\ \mathtt{HT},\ \mathtt{TH},\ \mathtt{TT}\}.$

$$\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4}$$

$$P(\mathtt{H}\text{ first}) = 1/4 + 1/4 = 1/2$$

$$P(\mathtt{H}\text{ second}) = 1/4 + 1/4 = 1/2$$

$$P(\mathtt{H}\text{ first and }\mathtt{H}\text{ second}) = 1/4$$

$$P(\mathtt{H}\text{ first})P(\mathtt{H}\text{ second}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(\mathtt{H}\text{ first and }\mathtt{H}\text{ second}).$$

By Theorem 5.5, "H second" is independent of "H first".

Similarly
"T second" is independent of "T first".
"T second" is independent of "H first".
"H second" is independent of "T first".

Example:

Recall the sample space and probability weights we used when discussing hashing.

Example:

Recall the sample space and probability weights we used when discussing hashing.

If n keys to hash into table of size k,
sample space consists of
all n-tuples of numbers between 1 and k.
Each n-tuple had equal weight k^{-n}.

Example:

Recall the sample space and probability weights we used when discussing hashing.

If n keys to hash into table of size k,
sample space consists of
all n-tuples of numbers between 1 and k.
Each n-tuple had equal weight k^{-n}.

We will now show that the two events
"i hashes to position r" and
"j hashes to position q"
are *independent* when $i \neq j$.

We will now show that the two events

"i hashes to position r" and "j hashes to position q"

are *independent* when $i \neq j$.

We will now show that the two events

"i hashes to position r" and "j hashes to position q"

are *independent* when $i \neq j$.

- "i hashes to position r"

  consists of all n-tuples with r in the ith position.

We will now show that the two events

"i hashes to position r" and "j hashes to position q"

are *independent* when $i \neq j$.

- "i hashes to position r"

  consists of all n-tuples with r in the ith position.

  So its probability is $k^{n-1}/k^n = 1/k$.

We will now show that the two events

"i hashes to position r" and "j hashes to position q"

are *independent* when $i \neq j$.

- "i hashes to position r"
  consists of all n-tuples with r in the ith position.
  So its probability is $k^{n-1}/k^n = 1/k$.
- Probability that "j hashes to position q" is also $1/k$.

We will now show that the two events

"i hashes to position r" and "j hashes to position q"

are *independent* when $i \neq j$.

- "i hashes to position r"

  consists of all n-tuples with r in the ith position.

  So its probability is $k^{n-1}/k^n = 1/k$.

- Probability that "j hashes to position q" is also $1/k$.

- If $i \neq j$,

  "i hashes to r and j hashes to q"

  has probability $k^{n-2}/k^n = 1/k^2$.

We will now show that the two events

"i hashes to position r" and "j hashes to position q"

are *independent* when $i \neq j$.

- "i hashes to position r"

  consists of all n-tuples with r in the ith position.

  So its probability is $k^{n-1}/k^n = 1/k$.

- Probability that "j hashes to position q" is also $1/k$.

- If $i \neq j$,

  "i hashes to r and j hashes to q"

  has probability $k^{n-2}/k^n = 1/k^2$.

- This is the product of the probabilities that

  "i hashes to r" and "j hashes to q".

We will now show that the two events

“i hashes to position r” and “j hashes to position q”

are *independent* when $i \neq j$.

- “i hashes to position r”

  consists of all n-tuples with r in the ith position.

  So its probability is $k^{n-1}/k^n = 1/k$.

- Probability that “j hashes to position q” is also $1/k$.

- If $i \neq j$,

  “i hashes to r and j hashes to q”

  has probability $k^{n-2}/k^n = 1/k^2$.

- This is the product of the probabilities that

  “i hashes to r” and “j hashes to q”.

- Therefore, these two events are independent.

Are the two events

“i hashes to position r” and “j hashes to position q”

independent when $i = j$?

Are the two events
"i hashes to position r" and "j hashes to position q"
independent when $i = j$?

If $i = j$, probability of
"i hashes to r and j hashes to q"
is 0, unless $r = q$, in which case it is 1.

Are the two events

"i hashes to position r" and "j hashes to position q"

independent when $i = j$?

If $i = j$, probability of

"i hashes to r and j hashes to q"

is 0, unless $r = q$, in which case it is 1.

Thus, these two events are **not** independent.

Conditional Probability and Independence

- Conditional Probability
- Independence
- Independent Trials Processes

Independent Trials Processes

So far, we've considered *static* sample sets.
That is, we assumed that our sample space contains all possible outcomes that can happen. Many problems, though, are modelled using dynamic processes.

For example, we flip coins one-by-one. After flipping 5 coins, we might do something, and then flip the 6^{th}. Our intuition is that the sixth flip should be independent of the outcomes of the first five.

As another example, we don't hash n keys all at once. We usually hash the first key, then the second, then the third, etc.. Our intuition is that the hashing of the k^{th} key should also be independent of the hashing of the first $(k-1)$ keys.

We formalize this idea with the introduction of Independent Trials Processes.

Examples:

Coin Flipping and Hashing

Examples:

Coin Flipping and Hashing

The Process Proceeds in Stages:

Examples:

Coin Flipping and Hashing

The Process Proceeds in Stages:

x_i: outcome at stage i.

Examples:

Coin Flipping and Hashing

The Process Proceeds in Stages:

x_i: outcome at stage i. ex: $x_i = \mathtt{H}$.

Examples:

Coin Flipping and Hashing

The Process Proceeds in Stages:

x_i: outcome at stage i. ex: $x_i = \mathtt{H}$.

S_i : set of possible outcomes of stage i.

Examples:
Coin Flipping and Hashing

The Process Proceeds in Stages:

x_i: outcome at stage i. ex: $x_i = \mathtt{H}$.

S_i : set of possible outcomes of stage i.

ex: $S_i = \{\mathtt{H,T}\}, 1 \le i \le n$.

Examples:
Coin Flipping and Hashing

The Process Proceeds in Stages:

x_i: outcome at stage i. ex: $x_i = \mathtt{H}$.

S_i : set of possible outcomes of stage i.

ex: $S_i = \{\mathtt{H},\mathtt{T}\}, 1 \leq i \leq n$.

A process that occurs in stages is called an **independent trials process** if

$$P(x_i = a_i | x_1 = a_1, \ldots, x_{i-1} = a_{i-1}) = P(x_i = a_i)$$

for each sequence $a_1, a_2, \ldots, a_n$, with $a_i \in S_i$.

Formally, let E_i be the event that $x_i = a_i$. Then

$$P(x_i = a_i | x_1 = a_1, \ldots, x_{i-1} = a_{i-1}) = P(x_i = a_i)$$

can be rewritten as

Formally, let E_i be the event that $x_i = a_i$. Then

$$P(x_i = a_i | x_1 = a_1, \ldots, x_{i-1} = a_{i-1}) = P(x_i = a_i)$$

can be rewritten as

$$P(E_i | E_1 \cap E_2 \cap \ldots \cap E_{i-1}) = P(E_i).$$

Formally, let E_i be the event that $x_i = a_i$. Then

$$P(x_i = a_i | x_1 = a_1, \ldots, x_{i-1} = a_{i-1}) = P(x_i = a_i)$$

can be rewritten as

$$P(E_i | E_1 \cap E_2 \cap \ldots \cap E_{i-1}) = P(E_i).$$

In words:

An independent trials process has the property that outcome of stage i is independent of outcomes of stages 1 through $i-1$.

Formally, let E_i be the event that $x_i = a_i$. Then

$$P(x_i = a_i | x_1 = a_1, \ldots, x_{i-1} = a_{i-1}) = P(x_i = a_i)$$

can be rewritten as

$$P(E_i | E_1 \cap E_2 \cap \ldots \cap E_{i-1}) = P(E_i).$$

In words:

An independent trials process has the property that outcome of stage i is independent of outcomes of stages 1 through $i-1$.

By product principle for independent probabilities (Theorem 5.5), this is equivalent to

$$P(E_1 \cap E_2 \cap \ldots \cap E_{i-1} \cap E_i) = P(E_1 \cap E_2 \cap \ldots \cap E_{i-1}) P(E_i).$$

Theorem 5.7 In an independent trials process, the probability of a sequence $a_1, a_2, \ldots, a_n$ of outcomes is

$$P(\{a_1\})\, P(\{a_2\}) \cdots P(\{a_n\}).$$

Theorem 5.7 In an independent trials process, the probability of a sequence $a_1, a_2, \ldots, a_n$ of outcomes is

$$P(\{a_1\})\,P(\{a_2\})\cdots P(\{a_n\}).$$

Proof:

Apply mathematical induction and use

$$P(E_1 \cap E_2 \cap \ldots \cap E_{i-1} \cap E_i) = P(E_1 \cap E_2 \cap \ldots \cap E_{i-1}) P(E_i).$$

Relation of independent trials

Relation of independent trials to coin flipping:

Relation of independent trials to coin flipping:

Sample space consists of sequences of n $\mathtt{H}$'s and $\mathtt{T}$'s.

Relation of independent trials to coin flipping:

Sample space consists of sequences of n $\mathtt{H}$'s and $\mathtt{T}$'s.

Probability of "$\mathtt{H}$ on ith flip" is $\dfrac{2^{n-1}}{2^n} = \dfrac{1}{2}$

.

Relation of independent trials to coin flipping:

Sample space consists of sequences of n `H`'s and `T`'s.

Probability of "`H` on ith flip" is $\frac{2^{n-1}}{2^n} = \frac{1}{2}$

Probability of "`H` on the ith flip, given a particular sequence on the first $i-1$ flips", is .

Relation of independent trials to coin flipping:

Sample space consists of sequences of n H's and T's.

Probability of "H on ith flip" is

$$\frac{2^{n-1}}{2^n} = \frac{1}{2}$$

Probability of "H on the ith flip, given a particular sequence on the first $i-1$ flips", is

$$\frac{2^{n-(i-1)-1}}{2^{n-(i-1)}} = \frac{1}{2}$$

Then "H (or T) on ith flip" is independent of "H (or T) on each of first $i-1$ flips".

Relation of independent trials

Relation of independent trials to hashing:

Relation of independent trials to hashing:

List of n keys to hash into a table of size k.
Sample space consists of
all n^k n-tuples of numbers between 1 and k.

Relation of independent trials to hashing:

List of n keys to hash into a table of size k.
Sample space consists of
all n^k n-tuples of numbers between 1 and k.

$P(\text{key } i \text{ hashes to } r\ ) = \frac{k^{n-1}}{k^n} = k^{-1}$

Relation of independent trials to hashing:

List of n keys to hash into a table of size k.
Sample space consists of
all n^k n-tuples of numbers between 1 and k.

$P(\text{key } i \text{ hashes to } r\ ) = \frac{k^{n-1}}{k^n} = k^{-1}$

$$P\left(\begin{array}{l}\text{key } i \text{ hashes to } r \qquad\qquad\qquad \text{and} \\ \text{keys } 1 \text{ through } i-1 \text{ hash to } q_1, q_2, \ldots, q_{i-1}\end{array}\right)$$

Relation of independent trials to hashing:

List of n keys to hash into a table of size k.
Sample space consists of
all n^k n-tuples of numbers between 1 and k.

$$P(\text{key } i \text{ hashes to } r\ ) = \frac{k^{n-1}}{k^n} = k^{-1}$$

$$P\left(\begin{array}{l}\text{key } i \text{ hashes to } r \qquad\qquad\qquad \text{and} \\ \text{keys } 1 \text{ through } i-1 \text{ hash to } q_1, q_2, \ldots, q_{i-1}\end{array}\right) = \frac{k^{n-i}}{k^n} = k^{-i}.$$

Relation of independent trials to hashing:

List of n keys to hash into a table of size k.
Sample space consists of
all n^k n-tuples of numbers between 1 and k.

$P(\text{key } i \text{ hashes to } r\ ) = \frac{k^{n-1}}{k^n} = k^{-1}$

$$P\begin{pmatrix} \text{key } i \text{ hashes to } r \text{ and} \\ \text{keys } 1 \text{ through } i-1 \text{ hash to } q_1, q_2, \ldots, q_{i-1} \end{pmatrix} = \frac{k^{n-i}}{k^n} = k^{-i}.$$

$P(\text{keys } 1 \text{ through } i-1 \text{ hash to } q_1, q_2, \ldots, q_{i-1})$

Relation of independent trials to hashing:

List of n keys to hash into a table of size k.
Sample space consists of
all n^k n-tuples of numbers between 1 and k.

$$P(\text{key } i \text{ hashes to } r\ ) = \frac{k^{n-1}}{k^n} = k^{-1}$$

$$P\begin{pmatrix} \text{key } i \text{ hashes to } r \qquad\qquad\qquad \text{and} \\ \text{keys } 1 \text{ through } i-1 \text{ hash to } q_1, q_2, \ldots, q_{i-1} \end{pmatrix} = \frac{k^{n-i}}{k^n} = k^{-i}.$$

$$P(\text{keys } 1 \text{ through } i-1 \text{ hash to } q_1, q_2, \ldots, q_{i-1}) = \frac{k^{n-(i-1)}}{k^n} = k^{1-i}.$$

By definition of conditional probability,

$$P\left(\begin{array}{l}\text{key } i \text{ hashes to } r \\ \quad \mid \text{ keys } 1 \text{ through } i-1 \text{ hash to } q_1, q_2, \ldots, q_{i-1}\end{array}\right)$$

$$= \frac{k^{-i}}{k^{1-i}} = \frac{1}{k}.$$

By definition of conditional probability,

$$P\left(\begin{array}{l}\text{key } i \text{ hashes to } r \\ \quad \Big| \text{ keys } 1 \text{ through } i-1 \text{ hash to } q_1, q_2, \ldots, q_{i-1}\end{array}\right)$$

$$= \frac{k^{-i}}{k^{1-i}} = \frac{1}{k}.$$

Since this is equal to

$$P(\text{key } i \text{ hashes to } r\ ) = \frac{k^{n-1}}{k^n} = k^{-1}$$

our model of hashing is an independent trials process.

Suppose we draw a card from a standard deck of 52 cards, replace it, draw another card, and continue for a total of ten draws.

Is this an independent trials process?

Suppose we draw a card from a standard deck of 52 cards, replace it, draw another card, and continue for a total of ten draws.

Is this an independent trials process?

Yes.

Suppose we draw a card from a standard deck of 52 cards, replace it, draw another card, and continue for a total of ten draws.

Is this an independent trials process?

Yes.

Because the probability that we draw a given card at one stage does *not* depend on the cards we drawn in earlier stages.

Suppose we draw a card from a standard deck of 52 cards, discard it (i.e., we do not replace it), draw another card, and continue for a total of ten draws.

Is this an independent trials process?

Suppose we draw a card from a standard deck of 52 cards, discard it (i.e., we do not replace it), draw another card, and continue for a total of ten draws. Is this an independent trials process?

No.

Suppose we draw a card from a standard deck of 52 cards, discard it (i.e., we do not replace it), draw another card, and continue for a total of ten draws. Is this an independent trials process?

No.

In the first draw, we have 52 cards to draw from

Suppose we draw a card from a standard deck of 52 cards, discard it (i.e., we do not replace it), draw another card, and continue for a total of ten draws. Is this an independent trials process?

No.

In the first draw, we have 52 cards to draw from

In the second draw, we have only 51.

Suppose we draw a card from a standard deck of 52 cards, discard it (i.e., we do not replace it), draw another card, and continue for a total of ten draws. Is this an independent trials process?

No.

In the first draw, we have 52 cards to draw from

In the second draw, we have only 51.

In particular, we do not have the same cards to draw from on the second draw as on the first.

Suppose we draw a card from a standard deck of 52 cards, discard it (i.e., we do not replace it), draw another card, and continue for a total of ten draws.
Is this an independent trials process?

No.

In the first draw, we have 52 cards to draw from

In the second draw, we have only 51.

In particular, we do not have the same cards to draw from on the second draw as on the first.

So, the probabilities for each possible outcome on the second draw depend on the outcome of the first draw.

Example:

Draw two cards.
What is the probability that you are holding two aces?

Example:

Draw two cards.
What is the probability that you are holding two aces?

(1) Drawing with replacement (first case):

$$\frac{4^2}{52^2} = \frac{1}{13^2} \approx .0059$$

Example:

Draw two cards.
What is the probability that you are holding two aces?

(1) Drawing with replacement (first case):

$$\frac{4^2}{52^2} = \frac{1}{13^2} \approx .0059$$

(2) Drawing without replacement (second case):

$$\frac{4 \cdot 3}{52 \cdot 51} = \frac{3}{13 \cdot 51} \approx .0045$$

More Examples:

More Examples:

Suppose we flip n coins and want to calculate the probability that *at least one coin shows a H*. One way to do this would be to use the inclusion-exclusion principle. Now that we know that coin tosses are independent trials, though, another easier way is as follows:

More Examples:

Suppose we flip n coins and want to calculate the probability that *at least one coin shows a H*. One way to do this would be to use the inclusion-exclusion principle. Now that we know that coin tosses are independent trials, though, another easier way is as follows:

Let E_i be the event of "$\mathtt{T}$ on ith flip"

$$Pr(E_i) = \frac{1}{2}$$

More Examples:

Suppose we flip n coins and want to calculate the probability that *at least one coin shows a H*. One way to do this would be to use the inclusion-exclusion principle. Now that we know that coin tosses are independent trials, though, another easier way is as follows:

Let E_i be the event of "T on ith flip"

$$Pr(E_i) = \frac{1}{2}$$

So, the probability that *all coins show a T is*

$$Pr(E_1 \cap E_2 \cap \cdots \cap E_n) = Pr(E_1) \cdot Pr(E_2) \cdots Pr(E_n) = \frac{1}{2^n}$$

More Examples:

Suppose we flip n coins and want to calculate the probability that *at least one coin shows a H*. One way to do this would be to use the inclusion-exclusion principle. Now that we know that coin tosses are independent trials, though, another easier way is as follows:

Let E_i be the event of "T on ith flip"

$$Pr(E_i) = \frac{1}{2}$$

So, the probability that *all coins show a T is*

$$Pr(E_1 \cap E_2 \cap \cdots \cap E_n) = Pr(E_1) \cdot Pr(E_2) \cdots Pr(E_n) = \frac{1}{2^n}$$

and the probability that *at least one coin shows an H* is

$$1 - \frac{1}{2^n}$$