Conditional Probability

Alexander Shen

LIRMM / CNRS & University of Montpellier

Outline

What is Conditional Probability?

How Reliable Is the Test?

Bayes' Theorem

Conditional Probability: A Paradox

Past and Future

Independence

Demographic Example

Demographic Example

- US: age 65 and over: about 14% (Wikipedia)

Demographic Example

- US: age 65 and over: about 14% (Wikipedia)
- “for a random American the probability to be at least 65 years old is 0.14”

Demographic Example

- US: age 65 and over: about 14% (Wikipedia)
- “for a random American the probability to be at least 65 years old is 0.14”
- random selection: a difficult task for pollsters

Demographic Example

- US: age 65 and over: about 14% (Wikipedia)
- "for a random American the probability to be at least 65 years old is 0.14"
- random selection: a difficult task for pollsters
- 0.75 male/female ratio for age 65 and over $(3 : 4)$

Demographic Example

- US: age 65 and over: about 14% (Wikipedia)
- "for a random American the probability to be at least 65 years old is 0.14"
- random selection: a difficult task for pollsters
- 0.75 male/female ratio for age 65 and over $(3 : 4)$
- $\approx$ 43% males $(3 : 7)$

Demographic Example

- US: age 65 and over: about 14% (Wikipedia)
- "for a random American the probability to be at least 65 years old is 0.14"
- random selection: a difficult task for pollsters
- 0.75 male/female ratio for age 65 and over $(3 : 4)$
- $\approx 43\%$ males $(3 : 7)$
- fraction of males of age 65 and over:

Demographic Example

- US: age 65 and over: about 14% (Wikipedia)
- "for a random American the probability to be at least 65 years old is 0.14"
- random selection: a difficult task for pollsters
- 0.75 male/female ratio for age 65 and over $(3:4)$
- $\approx 43\%$ males $(3:7)$
- fraction of males of age 65 and over: $0.14 \times 0.43 \approx 0.06$

Probability Language

Probability Language

- probability to be at least 65 y.o.: ≈ 0.14

Probability Language

- probability to be at least 65 y.o.: ≈ 0.14
- the *conditional probability* of being male if you are at least 65 y.o.: ≈ 0.43

Probability Language

- probability to be at least 65 y.o.: ≈ 0.14
- the *conditional probability* of being male if you are at least 65 y.o.: ≈ 0.43
- probability to be at least 65 y.o. and male: $0.14 \times 0.43 \approx 0.06$

Probability Language

- probability to be at least 65 y.o.: ≈ 0.14
- the *conditional probability* of being male if you are at least 65 y.o.: ≈ 0.43
- probability to be at least 65 y.o. and male: $0.14 \times 0.43 \approx 0.06$

$$\Pr[\text{male and } \geq 65] = \Pr[\geq 65] \times \Pr[\text{male} \mid \geq 65]$$

$$0.06\ldots \quad = \quad 0.14\ldots \quad \times \quad 0.43\ldots$$

Probability Language

- probability to be at least 65 y.o.: ≈ 0.14
- the *conditional probability* of being male if you are at least 65 y.o.: ≈ 0.43
- probability to be at least 65 y.o. and male: $0.14 \times 0.43 \approx 0.06$

$$\begin{array}{ccccc} \Pr[\text{male and } \geq 65] & = & \Pr[\geq 65] & \times & \Pr[\text{male} \mid \geq 65] \\ 0.06\ldots & = & 0.14\ldots & \times & 0.43\ldots \end{array}$$

$$\Pr[A \text{ and } B] = \Pr[B] \times \Pr[A \mid B]$$

Dice Example

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: "at least 3" $(3, 4, 5, 6)$

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: “at least 3” $(3, 4, 5, 6)$
- B: “even number” $(2, 4, 6)$

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: "at least 3" $(3, 4, 5, 6)$
- B: "even number" $(2, 4, 6)$
- A and B: even number at least 3 $(4, 6)$

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: “at least 3” $(3, 4, 5, 6)$
- B: “even number” $(2, 4, 6)$
- A and B: even number at least 3 $(4, 6)$
- $\Pr[A \text{ and } B] = \Pr[A] \times \Pr[B \mid A]$

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: "at least 3" $(3, 4, 5, 6)$
- B: "even number" $(2, 4, 6)$
- A and B: even number at least 3 $(4, 6)$
- $\Pr[A \text{ and } B] = \Pr[A] \times \Pr[B \mid A]$

$$1/3 = (4/6) \times (2/4)$$

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: "at least 3" $(3, 4, 5, 6)$
- B: "even number" $(2, 4, 6)$
- A and B: even number at least 3 $(4, 6)$
- $\Pr[A \text{ and } B] = \Pr[A] \times \Pr[B \mid A]$

$$1/3 = (4/6) \times (2/4)$$

- $\Pr[B \mid A]$: the fraction of B-outcomes among A-outcomes

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: "at least 3" $(3, 4, 5, 6)$
- B: "even number" $(2, 4, 6)$
- A and B: even number at least 3 $(4, 6)$
- $\Pr[A \text{ and } B] = \Pr[A] \times \Pr[B \mid A]$

$$1/3 = (4/6) \times (2/4)$$

- $\Pr[B \mid A]$: the fraction of B-outcomes among A-outcomes (for equiprobable outcomes)

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: "at least 3" $(3, 4, 5, 6)$
- B: "even number" $(2, 4, 6)$
- A and B: even number at least 3 $(4, 6)$
- $\Pr[A \text{ and } B] = \Pr[A] \times \Pr[B \mid A]$

  $$1/3 = (4/6) \times (2/4)$$
- $\Pr[B \mid A]$: the fraction of B-outcomes among A-outcomes (for equiprobable outcomes)
- $\Pr[B \mid A] = \Pr[A \text{ and } B] / \Pr[A]$

Dice Example

- probability space: 6 outcomes $1, 2, 3, 4, 5, 6$
- A: "at least 3" $(3, 4, 5, 6)$
- B: "even number" $(2, 4, 6)$
- A and B: even number at least 3 $(4, 6)$
- $\Pr[A \text{ and } B] = \Pr[A] \times \Pr[B \mid A]$

  $$1/3 = (4/6) \times (2/4)$$
- $\Pr[B \mid A]$: the fraction of B-outcomes among A-outcomes (for equiprobable outcomes)
- $\Pr[B \mid A] = \Pr[A \text{ and } B] / \Pr[A]$ (in general)

A, B, C form a queue in the random order (all orderings are equiprobable). What is the probability of the events X = "A is the second", Y = "A is before B in the queue"? What are the probabilities $\Pr[X \mid Y]$ and $\Pr[Y \mid X]$?

Outline

What is Conditional Probability?

How Reliable Is the Test?

Bayes' Theorem

Conditional Probability: A Paradox

Past and Future

Independence

Be Careful

Be Careful

- disease: 1% of population

Be Careful

- disease: 1% of population
- test: false negative rate 10% (negative for 1/10 of people with the disease)

Be Careful

- disease: 1% of population
- test: false negative rate 10% (negative for 1/10 of people with the disease)
- false positive rate 10% (positive for 1/10 of people without the disease)

Be Careful

- disease: 1% of population
- test: false negative rate 10% (negative for 1/10 of people with the disease)
- false positive rate 10% (positive for 1/10 of people without the disease)
- fraction of having the disease among the test-positive = ?

Be Careful

- disease: 1% of population
- test: false negative rate 10% (negative for $1/10$ of people with the disease)
- false positive rate 10% (positive for $1/10$ of people without the disease)
- fraction of having the disease among the test-positive = ?
- $\Pr[D] = 0.01$; $\Pr[T \mid D] = 0.9$; $\Pr[T \mid \text{not } D] = 0.1$.

Be Careful

- disease: 1% of population
- test: false negative rate 10% (negative for $1/10$ of people with the disease)
- false positive rate 10% (positive for $1/10$ of people without the disease)
- fraction of having the disease among the test-positive = ?
- $\Pr[D] = 0.01$; $\Pr[T \mid D] = 0.9$; $\Pr[T \mid \text{not } D] = 0.1$. $\quad \Pr[D \mid T] = ?$

Be Careful

- disease: 1% of population
- test: false negative rate 10% (negative for $1/10$ of people with the disease)
- false positive rate 10% (positive for $1/10$ of people without the disease)
- fraction of having the disease among the test-positive = ?
- $\Pr[D] = 0.01$; $\Pr[T \mid D] = 0.9$; $\Pr[T \mid \text{not } D] = 0.1$. $\quad \Pr[D \mid T] = ?$
- test is quite reliable, isn't it?

Keep Calm and Compute

Keep Calm and Compute

- 90% positive among 1% ill = 0.9% positive and ill

Keep Calm and Compute

- 90% positive among 1% ill =
  0.9% positive and ill
- 10% positive among 99% healthy =
  9.9% positive and healthy

Keep Calm and Compute

- 90% positive among 1% ill = 0.9% positive and ill
- 10% positive among 99% healthy = 9.9% positive and healthy
- $0.9 : 9.9 = 1 : 11$

Keep Calm and Compute

- 90% positive among 1% ill = 0.9% positive and ill
- 10% positive among 99% healthy = 9.9% positive and healthy
- $0.9 : 9.9 = 1 : 11$
- only $1/12$ of test-positives are ill

The Same Computation

The Same Computation

- $\Pr[D] = 0.01; \Pr[T \mid D] = 0.9;$
  $\Pr[T \mid \text{not } D] = 0.1.$

The Same Computation

- $\Pr[D] = 0.01; \Pr[T \mid D] = 0.9;$ $\Pr[T \mid \text{not } D] = 0.1. \quad \Pr[D \mid T] = ?$

The Same Computation

- $\Pr[D] = 0.01$; $\Pr[T \mid D] = 0.9$; $\Pr[T \mid \text{not } D] = 0.1$. $\quad \Pr[D \mid T] = ?$
- $\Pr[T \text{ and } D] = \Pr[D] \Pr[T \mid D] =$ $0.01 \cdot 0.9 = 0.009$.

The Same Computation

- $\Pr[D] = 0.01$; $\Pr[T \mid D] = 0.9$; $\Pr[T \mid \text{not } D] = 0.1$. $\Pr[D \mid T] = ?$
- $\Pr[T \text{ and } D] = \Pr[D]\Pr[T \mid D] = 0.01 \cdot 0.9 = 0.009$.
- $\Pr[T \text{ and not } D] = \Pr[\text{ not } D]\Pr[T \mid \text{not } D] = 0.99 \cdot 0.1 = 0.099$.

The Same Computation

- $\Pr[D] = 0.01; \Pr[T \mid D] = 0.9;$ $\Pr[T \mid \text{not } D] = 0.1. \quad \Pr[D \mid T] = ?$
- $\Pr[T \text{ and } D] = \Pr[D] \Pr[T \mid D] =$ $0.01 \cdot 0.9 = 0.009.$
- $\Pr[T \text{ and not } D] = \Pr[\text{ not } D] \Pr[T \mid \text{not } D] =$ $0.99 \cdot 0.1 = 0.099.$
- $\Pr[T] = \Pr[T \text{and } D] + \Pr[T \text{and not } D] =$ $0.009 + 0.099 = 0.108$

The Same Computation

- $\Pr[D] = 0.01; \Pr[T \mid D] = 0.9;$ $\Pr[T \mid \text{not } D] = 0.1. \quad \Pr[D \mid T] = ?$
- $\Pr[T \text{ and } D] = \Pr[D] \Pr[T \mid D] =$ $0.01 \cdot 0.9 = 0.009.$
- $\Pr[T \text{ and not } D] = \Pr[\text{ not } D] \Pr[T \mid \text{not } D] =$ $0.99 \cdot 0.1 = 0.099.$
- $\Pr[T] = \Pr[T \text{and } D] + \Pr[T \text{and not } D] =$ $0.009 + 0.099 = 0.108$
- $\Pr[D \mid T] = \Pr[D \text{and } T] / \Pr[T] =$ $0.009/0.108 = 9/108 = 1/12 =$ $0.0833\ldots \approx 8.3\%$

The Law of Total Probability

The Law of Total Probability

- probability space is split into mutually exclusive cases

The Law of Total Probability

- probability space is split into mutually exclusive cases
- $X = B_1 \cup B_2 \cup \ldots \cup B_n$

The Law of Total Probability

- probability space is split into mutually exclusive cases
- $X = B_1 \cup B_2 \cup \ldots \cup B_n$
- $B_i \cap B_j = \varnothing$

The Law of Total Probability

- probability space is split into mutually exclusive cases
- $X = B_1 \cup B_2 \cup \ldots \cup B_n$
- $B_i \cap B_j = \emptyset$
- than A is split into mutually exclusuve "A and B_1", "A and B_2",..., "A and B_n"

The Law of Total Probability

- probability space is split into mutually exclusive cases
- $X = B_1 \cup B_2 \cup \ldots \cup B_n$
- $B_i \cap B_j = \varnothing$
- than A is split into mutually exclusuve "A and B_1", "A and B_2",..., "A and B_n"
- $\Pr[A] = \Pr[A \text{ and } B_1] + \ldots + \Pr[A \text{ and } B_n]$

The Law of Total Probability

- probability space is split into mutually exclusive cases
- $X = B_1 \cup B_2 \cup \ldots \cup B_n$
- $B_i \cap B_j = \varnothing$
- than A is split into mutually exclusuve "A and B_1", "A and B_2",..., "A and B_n"
- $\Pr[A] = \Pr[A \text{ and } B_1] + \ldots + \Pr[A \text{ and } B_n]$=

$$= \Pr[B_1] \cdot \Pr[A \mid B_1] + \ldots + \Pr[B_n] \cdot \Pr[A \mid B_n]$$

The Law of Total Probability

- probability space is split into mutually exclusive cases
- $X = B_1 \cup B_2 \cup \ldots \cup B_n$
- $B_i \cap B_j = \varnothing$
- than A is split into mutually exclusuve "A and B_1", "A and B_2",..., "A and B_n"
- $\Pr[A] = \Pr[A \text{ and } B_1] + \ldots + \Pr[A \text{ and } B_n] =$

$$= \Pr[B_1] \cdot \Pr[A \mid B_1] + \ldots + \Pr[B_n] \cdot \Pr[A \mid B_n]$$

$$= \Pr[B] \cdot \Pr[A \mid B] + (1 - \Pr[B]) \cdot \Pr[A \mid \text{not } B]$$

Outline

What is Conditional Probability?

How Reliable Is the Test?

Bayes' Theorem

Conditional Probability: A Paradox

Past and Future

Independence

Crime and evidence

Crime and evidence

- e-mail: `MESSAGE FROM BANK OF AMERICA`

Crime and evidence

- e-mail: `MESSAGE FROM BANK OF AMERICA`
- From: `<info4@water.ocn.ne.jp>` (Japan)

Crime and evidence

- e-mail: `MESSAGE FROM BANK OF AMERICA`
- From: `<info4@water.ocn.ne.jp>` (Japan)
- probably a scam, i thought…

Crime and evidence

- e-mail: `MESSAGE FROM BANK OF AMERICA`
- From: `<info4@water.ocn.ne.jp>` (Japan)
- probably a scam, i thought… why?

Crime and evidence

- e-mail: `MESSAGE FROM BANK OF AMERICA`
- From: `<info4@water.ocn.ne.jp>` (Japan)
- probably a scam, i thought… why?
- many scam messages use foreign e-mail address

Crime and evidence

- e-mail: `MESSAGE FROM BANK OF AMERICA`
- From: `<info4@water.ocn.ne.jp>` (Japan)
- probably a scam, i thought… why?
- many scam messages use foreign e-mail address
- foreign e-mail address is rather unusual in general

Crime and evidence

- e-mail: `MESSAGE FROM BANK OF AMERICA`
- From: `<info4@water.ocn.ne.jp>` (Japan)
- probably a scam, i thought… why?
- many scam messages use foreign e-mail address
- foreign e-mail address is rather unusual in general
- scam messages are quite frequent now

Crime and evidence

- e-mail: `MESSAGE FROM BANK OF AMERICA`
- From: `<info4@water.ocn.ne.jp>` (Japan)
- probably a scam, i thought… why?
- many scam messages use foreign e-mail address
- foreign e-mail address is rather unusual in general
- scam messages are quite frequent now
- three components of Bayes' reasoning

Bayes' Formula

Bayes' Formula

- H for hypothesis, E for evidence

Bayes' Formula

- H for hypothesis, E for evidence (arbitrary two events having a common probability space)

Bayes' Formula

- H for hypothesis, E for evidence (arbitrary two events having a common probability space)
- $\Pr[H \mid E] = \dfrac{\Pr[H \text{ and } E]}{\Pr[E]} = \dfrac{\Pr[E \mid H] \Pr[H]}{\Pr[E]}$

Bayes' Formula

- H for hypothesis, E for evidence (arbitrary two events having a common probability space)
- $\Pr[H \mid E] = \dfrac{\Pr[H \text{ and } E]}{\Pr[E]} = \dfrac{\Pr[E \mid H] \Pr[H]}{\Pr[E]}$
- $\Pr[H \mid E] = \dfrac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$ (Bayes)

Bayes' Formula

- H for hypothesis, E for evidence (arbitrary two events having a common probability space)
- $\Pr[H \mid E] = \dfrac{\Pr[H \text{ and } E]}{\Pr[E]} = \dfrac{\Pr[E \mid H] \Pr[H]}{\Pr[E]}$
- $\Pr[H \mid E] = \dfrac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$ (Bayes)
- condition E multiplies the probability of H by factor $\Pr[E \mid H] / \Pr[E]$

Bayes' Formula

- H for hypothesis, E for evidence (arbitrary two events having a common probability space)
- $\Pr[H \mid E] = \dfrac{\Pr[H \text{ and } E]}{\Pr[E]} = \dfrac{\Pr[E \mid H] \Pr[H]}{\Pr[E]}$
- $\Pr[H \mid E] = \dfrac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$ (Bayes)
- condition E multiplies the probability of H by factor $\Pr[E \mid H] / \Pr[E]$ that measures how much the condition H increases the probability of E

Theory and Practice

Theory and Practice

$$\Pr[H \mid E] = \frac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$$

Theory and Practice

$$\Pr[H \mid E] = \frac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$$

- $H:$ message is a scam;
  $E:$ message uses foreign address

Theory and Practice

$$\Pr[H \mid E] = \frac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$$

- H: message is a scam;
  E: message uses foreign address
- $\Pr[H \mid E]$ is high, since:

Theory and Practice

$$\Pr[H \mid E] = \frac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$$

- H: message is a scam;
  E: message uses foreign address
- $\Pr[H \mid E]$ is high, since:
  - $\Pr[E \mid H]$ is quite high;

Theory and Practice

$$\Pr[H \mid E] = \frac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$$

- H: message is a scam;
  E: message uses foreign address
- $\Pr[H \mid E]$ is high, since:
  - $\Pr[E \mid H]$ is quite high;
  - $\Pr[E]$ is low;

Theory and Practice

$$\Pr[H \mid E] = \frac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$$

- H: message is a scam;
  E: message uses foreign address
- $\Pr[H \mid E]$ is high, since:
  - $\Pr[E \mid H]$ is quite high;
  - $\Pr[E]$ is low;
  - $\Pr[H]$ is not low;

Theory and Practice

$$\Pr[H \mid E] = \frac{\Pr[E \mid H]}{\Pr[E]} \Pr[H]$$

- H: message is a scam;
  E: message uses foreign address
- $\Pr[H \mid E]$ is high, since:
  - $\Pr[E \mid H]$ is quite high;
  - $\Pr[E]$ is low;
  - $\Pr[H]$ is not low;
- "foreign address make the scam hypothesis much more probable, because it appears in scam messages much more often than in general"

Disease and Test Revisited

Disease and Test Revisited

- $\Pr[D \mid T] = \dfrac{\Pr[T \mid D]}{\Pr[T]} \Pr[D]$

Disease and Test Revisited

- $\Pr[D \mid T] = \dfrac{\Pr[T \mid D]}{\Pr[T]} \Pr[D]$
- $\Pr[D \mid T] = \dfrac{0.9}{0.108} \cdot 0.01 = 0.0833\ldots$

Disease and Test Revisited

- $\Pr[D \mid T] = \dfrac{\Pr[T \mid D]}{\Pr[T]} \Pr[D]$
- $\Pr[D \mid T] = \dfrac{0.9}{0.108} \cdot 0.01 = 0.0833\ldots$
- Indeed, $\Pr[T \mid D]$ is much bigger than $\Pr[T]$, but $\Pr[D]$ is so small that $\Pr[D \mid T]$ is still rather low

Disease and Test Revisited

- $\Pr[D \mid T] = \dfrac{\Pr[T \mid D]}{\Pr[T]} \Pr[D]$
- $\Pr[D \mid T] = \dfrac{0.9}{0.108} \cdot 0.01 = 0.0833\ldots$
- Indeed, $\Pr[T \mid D]$ is much bigger than $\Pr[T]$, but $\Pr[D]$ is so small that $\Pr[D \mid T]$ is still rather low
- take-home message from Bayes: *if condition B increases the probability of A by factor k, then condition A increases the probability of B by the same factor k*

Outline

What is Conditional Probability?

How Reliable Is the Test?

Bayes' Theorem

Conditional Probability: A Paradox

Past and Future

Independence

Two children

Two children

- “Mary has two children. At least one of them is a girl. What is the probability that she has two daughters?”

Two children

- “Mary has two children. At least one of them is a girl. What is the probability that she has two daughters?”
- “Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?”

Two children

- “Mary has two children. At least one of them is a girl. What is the probability that she has two daughters?”
- “Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?”
- solution attempt:

Two children

- “Mary has two children. At least one of them is a girl. What is the probability that she has two daughters?”
- “Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?”
- solution attempt:
    - either first or the second bit was a tail

Two children

- “Mary has two children. At least one of them is a girl. What is the probability that she has two daughters?”
- “Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?”
- solution attempt:
    - either first or the second bit was a tail
    - may assume w.l.o.g. it was the first

Two children

- "Mary has two children. At least one of them is a girl. What is the probability that she has two daughters?"
- "Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"
- solution attempt:
  - either first or the second bit was a tail
  - may assume w.l.o.g. it was the first
  - two outcomes: TH, TT

Two children

- "Mary has two children. At least one of them is a girl. What is the probability that she has two daughters?"
- "Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"
- solution attempt:
  - either first or the second bit was a tail
  - may assume w.l.o.g. it was the first
  - two outcomes: TH, TT
  - probability $1/2$

An Alternative Solution

An Alternative Solution

“Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?”

An Alternative Solution

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- 4 equiprobable outcomes HH, HT, TH, TT

An Alternative Solution

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- 4 equiprobable outcomes HH, HT, TH, TT
- event C: *at least one tail*

An Alternative Solution

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- 4 equiprobable outcomes HH, HT, TH, TT
- event C: *at least one tail*
- contains HT, TH, TT: probability $3/4$

An Alternative Solution

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- 4 equiprobable outcomes HH, HT, TH, TT
- event C: *at least one tail*
- contains HT, TH, TT: probability $3/4$
- event E: *two tails*

An Alternative Solution

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- 4 equiprobable outcomes HH, HT, TH, TT
- event C: *at least one tail*
- contains HT, TH, TT: probability $3/4$
- event E: *two tails*
- contains TT: probability $1/4$

An Alternative Solution

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- 4 equiprobable outcomes HH, HT, TH, TT
- event C: *at least one tail*
- contains HT, TH, TT: probability $3/4$
- event E: *two tails*
- contains TT: probability $1/4$
- conditional probability:

$$\Pr[E \mid C] = \frac{\Pr[E \text{ and } C]}{\Pr[C]} = \frac{\Pr[E]}{\Pr[C]} = \frac{1}{3}$$

Reading the Problem Statement

Reading the Problem Statement

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

Reading the Problem Statement

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- probability space: HH, HT, TH, TT

Reading the Problem Statement

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- probability space: HH, HT, TH, TT
- event C: *at least one tail* [condition]

Reading the Problem Statement

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- probability space: HH, HT, TH, TT
- event C: *at least one tail* [condition]
- contains HT, TH, TT : probability $3/4$

Reading the Problem Statement

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- probability space: HH, HT, TH, TT
- event C: *at least one tail* [condition]
- contains HT, TH, TT : probability $3/4$
- event E: *two tails*

Reading the Problem Statement

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- probability space: HH, HT, TH, TT
- event C: *at least one tail* [condition]
- contains HT, TH, TT : probability $3/4$
- event E: *two tails* (part of C)

Reading the Problem Statement

"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- probability space: HH, HT, TH, TT
- event C: *at least one tail* [condition]
- contains HT, TH, TT : probability $3/4$
- event E: *two tails* (part of C)
- contains TT : probability $1/4$

Reading the Problem Statement

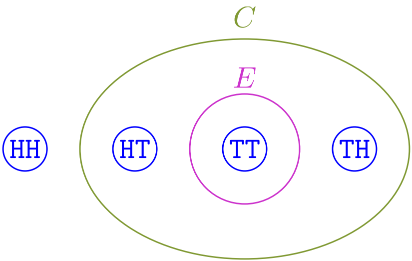
"Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"

- probability space: HH, HT, TH, TT
- event C: *at least one tail* [condition]
- contains HT, TH, TT: probability $3/4$
- event E: *two tails* (part of C)
- contains TT: probability $1/4$
- conditional probability:

$$\Pr[E \mid C] = \frac{\Pr[E \text{ and } C]}{\Pr[C]} = \frac{1}{3}$$

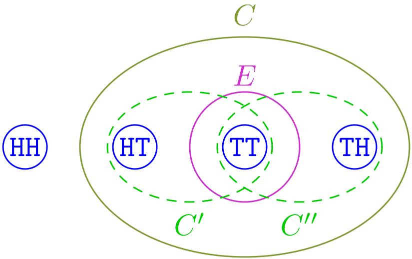
What the First “Solution” Says

What the First "Solution" Says



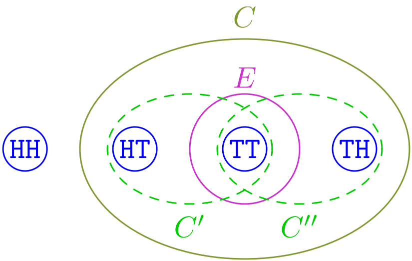
- space, condition, event

What the First “Solution” Says



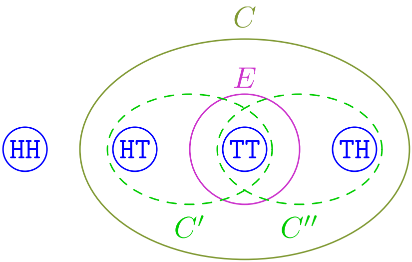
- space, condition, event
- two events in the first ‘solution’

What the First "Solution" Says



- space, condition, event
- two events in the first 'solution'
- together they make C, but $\Pr[E \mid C'] = \Pr[E \mid C''] = 1/2$ while $\Pr[E \mid C] = 1/3$.

What the First "Solution" Says



- space, condition, event
- two events in the first 'solution'
- together they make C, but
  $\Pr[E \mid C'] = \Pr[E \mid C''] = 1/2$ while
  $\Pr[E \mid C] = 1/3$.
- information $\neq$ condition !

Outline

What is Conditional Probability?

How Reliable Is the Test?

Bayes' Theorem

Conditional Probability: A Paradox

Past and Future

Independence

Does the History Matter?

A dice is rolled twice. Does the probability of six in the second experiment

increase / decrease / remain unchanged

if the first time we get six?

Does the History Matter?

A dice is rolled twice. Does the probability of six in the second experiment

increase / decrease / remain unchanged

if the first time we get six?

- $\Pr[\, y = 6 \mid x = 6 \,] = \Pr[\, y = 6 \,]$?

Does the History Matter?

A dice is rolled twice. Does the probability of six in the second experiment

increase / decrease / remain unchanged

if the first time we get six?

- $\Pr[\,y = 6 \mid x = 6\,] = \Pr[\,y = 6\,]$?
- model of 36 equiprobable outcomes:

Does the History Matter?

A dice is rolled twice. Does the probability of six in the second experiment

increase / decrease / remain unchanged

if the first time we get six?

- $\Pr[y = 6 \mid x = 6] = \Pr[y = 6]$?
- model of 36 equiprobable outcomes: $\Pr[y = 6 \mid x = 6] = \frac{\Pr[x=6, y=6]}{\Pr[x=6]} = \frac{1/36}{6/36} = \frac{1}{6} = \Pr[y = 6]$.

Models and Real World

Models and Real World

- if you see 999 heads while tossing a coin, what about the next coin tossing?

Models and Real World

- if you see 999 heads while tossing a coin, what about the next coin tossing?
- the model says "still $1/2$"
  ($2^{-1000}/2^{-999} = 1/2$)

Models and Real World

- if you see 999 heads while tossing a coin, what about the next coin tossing?
- the model says “still $1/2$” ($2^{-1000}/2^{-999} = 1/2$)
- would you agree?

Models and Real World

- if you see 999 heads while tossing a coin, what about the next coin tossing?
- the model says “still $1/2$” ($2^{-1000}/2^{-999} = 1/2$)
- would you agree?
- probably not: head looks “more probable”

Models and Real World

- if you see 999 heads while tossing a coin, what about the next coin tossing?
- the model says "still $1/2$" ($2^{-1000}/2^{-999} = 1/2$)
- would you agree?
- probably not: head looks "more probable"
- for a good reason: probably the model does not fit this experiment

Models and Real World

- if you see 999 heads while tossing a coin, what about the next coin tossing?
- the model says "still $1/2$" ($2^{-1000}/2^{-999} = 1/2$)
- would you agree?
- probably not: head looks "more probable"
- for a good reason: probably the model does not fit this experiment
- statistics: finding a good model

Models and Real World

- if you see 999 heads while tossing a coin, what about the next coin tossing?
- the model says “still $1/2$” ($2^{-1000}/2^{-999} = 1/2$)
- would you agree?
- probably not: head looks “more probable”
- for a good reason: probably the model does not fit this experiment
- statistics: finding a good model (two heads coin?)

The Mystery of Edgar Poe

The Mystery of Edgar Poe

- Edgar Allan Poe (1809–1849): famous author (short stories, poems)

The Mystery of Edgar Poe

- Edgar Allan Poe (1809–1849): famous author (short stories, poems)
- believed that probability of six decreases if we already had six

The Mystery of Edgar Poe

- Edgar Allan Poe (1809–1849): famous author (short stories, poems)
- believed that probability of six decreases if we already had six
- more: he believed (?) that this is what probability theory says

The Mystery of Edgar Poe

- Edgar Allan Poe (1809–1849): famous author (short stories, poems)
- believed that probability of six decreases if we already had six
- more: he believed (?) that this is what probability theory says
- and complained that ‘general readers’ do not agree with that

The Mystery of Edgar Poe

- Edgar Allan Poe (1809–1849): famous author (short stories, poems)
- believed that probability of six decreases if we already had six
- more: he believed (?) that this is what probability theory says
- and complained that ‘general readers’ do not agree with that
- crazy statistician brings the bomb to a plane because he thinks two bombs on the same plane are quite improbable

Edgar Poe Says

Nothing… is more difficult than to convince the merely general reader that the fact of sixes having been thrown twice in succession…is sufficient cause for betting the largest odds that sixes will not be thrown in the third attempt. A suggestion to this effect is usually rejected by the intellect at once. It does not appear that the two throws which have been completed, and which lie now absolutely in the Past, can have influence upon the throw which exists only in the Future. The chance for throwing sixes seems to be precisely as it was at any ordinary time… And this is a reflection which appears so exceedingly obvious that attempts to controvert it are received more frequently with a derisive smile than with any thing like respectful attention. The error here involved – a gross error redolent of mischief – I cannot pretend to expose within the limits assigned me at present… (Edgar Poe, 'The Mystery of Marie Roget', 1850)

Outline

What is Conditional Probability?

How Reliable Is the Test?

Bayes' Theorem

Conditional Probability: A Paradox

Past and Future

Independence

Mathematical Notion of Independence

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$
- $\Pr[A \text{ and } B] / \Pr[B] = \Pr[A]$

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$
- $\Pr[A \text{ and } B] / \Pr[B] = \Pr[A]$
- $\boxed{\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]}$

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$
- $\Pr[A \text{ and } B] / \Pr[B] = \Pr[A]$
- $\boxed{\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]}$
- symmetry

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$
- $\Pr[A \text{ and } B] / \Pr[B] = \Pr[A]$
- $\boxed{\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]}$
- symmetry
- ‘product rule’

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$
- $\Pr[A \text{ and } B] / \Pr[B] = \Pr[A]$
- $\boxed{\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]}$
- symmetry
- ‘product rule’—not a rule

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$
- $\Pr[A \text{ and } B] / \Pr[B] = \Pr[A]$
- $\boxed{\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]}$
- symmetry
- ‘product rule’—not a rule
- only a definition of independence

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$
- $\Pr[A \text{ and } B] / \Pr[B] = \Pr[A]$
- $\boxed{\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]}$
- symmetry
- ‘product rule’—not a rule
- only a definition of independence
- two dice: A=“the first…”;
  B=“the second…”

Mathematical Notion of Independence

- A and B are *independent* if $\Pr[A \mid B] = \Pr[A]$
- $\Pr[A \text{ and } B] / \Pr[B] = \Pr[A]$
- $\boxed{\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]}$
- symmetry
- ‘product rule’—not a rule
- only a definition of independence
- two dice: A=“the first…”;
  B=“the second…”
- zero probability case

Recalling Bayes

Recalling Bayes

- dependent events:
  $\Pr[A \text{ and } B] \neq \Pr[A] \cdot \Pr[B]$

Recalling Bayes

- dependent events:
  $\Pr[A \text{ and } B] \neq \Pr[A] \cdot \Pr[B]$
- say, $\Pr[A \text{ and } B] > \Pr[A] \cdot \Pr[B]$

Recalling Bayes

- dependent events:
  $\Pr[A \text{ and } B] \neq \Pr[A] \cdot \Pr[B]$
- say, $\Pr[A \text{ and } B] > \Pr[A] \cdot \Pr[B]$
- then $\Pr[A \mid B] > \Pr[A]$

Recalling Bayes

- dependent events:
  $\Pr[A \text{ and } B] \neq \Pr[A] \cdot \Pr[B]$
- say, $\Pr[A \text{ and } B] > \Pr[A] \cdot \Pr[B]$
- then $\Pr[A \mid B] > \Pr[A]$
- "condition B makes A more probable"

Recalling Bayes

- dependent events:
  $\Pr[A \text{ and } B] \neq \Pr[A] \cdot \Pr[B]$
- say, $\Pr[A \text{ and } B] > \Pr[A] \cdot \Pr[B]$
- then $\Pr[A \mid B] > \Pr[A]$
- "condition B makes A more probable"
- then condition A makes B more probable (symmetry)

Recalling Bayes

- dependent events:
  $\Pr[A \text{ and } B] \neq \Pr[A] \cdot \Pr[B]$
- say, $\Pr[A \text{ and } B] > \Pr[A] \cdot \Pr[B]$
- then $\Pr[A \mid B] > \Pr[A]$
- "condition B makes A more probable"
- then condition A makes B more probable (symmetry)
- Bayes' formula: the factor is the same:

$$\Pr[B \mid A] = \frac{\Pr[A \mid B]}{\Pr[A]} \Pr[B]$$

Mathematics $\neq$ Common Sense

Mathematics $\neq$ Common Sense

- two dice (x, y): A = 'x is a multiple of 2', B = 'y is a multiple of 3'

Mathematics $\neq$ Common Sense

- two dice (x, y): A = 'x is a multiple of 2', B = 'y is a multiple of 3'
- 'x is a multiple of 2', 'x is a multiple of 3': $\{2, 4, 6\} \cap \{3, 6\} = \{6\}$

Mathematics $\neq$ Common Sense

- two dice (x, y): A = 'x is a multiple of 2', B = 'y is a multiple of 3'
- 'x is a multiple of 2', 'x is a multiple of 3': $\{2, 4, 6\} \cap \{3, 6\} = \{6\}$
- mathematical independence: $1/6 = (1/2) \times (1/3)$

Mathematics $\neq$ Common Sense

- two dice (x, y): A = 'x is a multiple of 2', B = 'y is a multiple of 3'
- 'x is a multiple of 2', 'x is a multiple of 3': $\{2, 4, 6\} \cap \{3, 6\} = \{6\}$
- mathematical independence: $1/6 = (1/2) \times (1/3)$
- an independent commitee repeats the test, but no independence in mathematical sense

Mathematics $\neq$ Common Sense

- two dice (x, y): A = 'x is a multiple of 2', B = 'y is a multiple of 3'
- 'x is a multiple of 2', 'x is a multiple of 3': $\{2, 4, 6\} \cap \{3, 6\} = \{6\}$
- mathematical independence: $1/6 = (1/2) \times (1/3)$
- an independent commitee repeats the test, but no independence in mathematical sense : correlation $\neq$ causation

Mathematics $\neq$ Common Sense

- two dice (x, y): A = 'x is a multiple of 2', B = 'y is a multiple of 3'
- 'x is a multiple of 2', 'x is a multiple of 3': $\{2, 4, 6\} \cap \{3, 6\} = \{6\}$
- mathematical independence: $1/6 = (1/2) \times (1/3)$
- an independent commitee repeats the test, but no independence in mathematical sense : correlation $\neq$ causation
- $\Pr[\text{ ill } \mid \text{visiting a doctor}] > \Pr[\text{ ill }]$