A Primer on Probabilities

Andreas Maletti

SMT — October 28, 2011

Speaker

Andreas Maletti

- Dipl.-Inf. (2002, TU Dresden) — functional programming
- Dr. rer. nat. (2006, TU Dresden) — tree transducer theory
- PostDoc (ICSI, Berkeley, USA) — first exposure to NLP
- Jr. Research Scientist (2008–2011, URV, Tarragona, Spain) — automata theory in NLP
- Jr. Research Group Leader (IMS) — syntax-based MT

Speaker

Andreas Maletti

- Dipl.-Inf. (2002, TU Dresden) — functional programming
- Dr. rer. nat. (2006, TU Dresden) — tree transducer theory
- PostDoc (ICSI, Berkeley, USA) — first exposure to NLP
- Jr. Research Scientist (2008–2011, URV, Tarragona, Spain) — automata theory in NLP
- Jr. Research Group Leader (IMS) — syntax-based MT

Research interests

- Automata theory
- Complexity theory
- Applications in NLP (esp. machine translation)

Lecture Goals

Probabilities

- Basic notions and intuitive understanding
- Basic laws and computing with probabilities
- Independence

Literature:
[Grinstead, Snell: *Introduction to Probability.* AMS 1997]
`http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html`

Lecture Goals

Probabilities

- Basic notions and intuitive understanding
- Basic laws and computing with probabilities
- Independence

Conditional probabilities

- Basic notions
- Bayes law
- Modelling complications

Literature:
[Grinstead, Snell: *Introduction to Probability.* AMS 1997]
`http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html`

Always ask questions right away!

Consultation hour: Monday, 3-4pm

First Intuitive Notions

Probability

The (theoretical) likelihood of an event in an experiment

Notions

First Intuitive Notions

Probability

The (theoretical) likelihood of an event in an experiment

Notions

First Intuitive Notions

Probability

The (theoretical) likelihood of an event in an experiment

Notions

- **Experiment**: Action producing one of a finite number of outcomes discrete probability distribution

First Intuitive Notions

Probability

The (theoretical) likelihood of an event in an experiment

Notions

- Experiment: Action producing one of a finite number of outcomes — discrete probability distribution
- Event: A subset of the outcomes

First Intuitive Notions

Probability

The (theoretical) likelihood of an event in an experiment

Notions

- Experiment: Action producing one of a finite number of outcomes — discrete probability distribution
- Event: A subset of the outcomes
- Likelihood: Fraction of an outcome happening in a huge number of experiments

First Intuitive Notions

Probability

The (theoretical) likelihood of an event in an experiment

Notions

- Experiment: Action producing one of a finite number of outcomes — discrete probability distribution
- Event: A subset of the outcomes
- Likelihood: Fraction of an outcome happening in a huge number of experiments

Example

- Experiment: Rolling a die
- Event: Rolling a '5' or a '6'
- Likelihood of outcome '6': $\frac{1}{6}$ — for a fair die

First Mathematical Notions

Notions

- Experiment: Sample space S and random variable $X \in S$

First Mathematical Notions

Notions

- Experiment: Sample space S and random variable $X \in S$
- Outcome: $e \in S$
- Event: $E \subseteq S$

First Mathematical Notions

Notions

- Experiment: Sample space S and random variable $X \in S$
- Outcome: $e \in S$
- Event: $E \subseteq S$
- Likelihood: $p\colon S \to [0,1]$ such that $\sum_{e \in S} p(e) = 1$

First Mathematical Notions

Notions

- Experiment: Sample space S and random variable $X \in S$
- Outcome: $e \in S$
- Event: $E \subseteq S$
- Likelihood: $p\colon S \to [0,1]$ such that $\sum_{e \in S} p(e) = 1$

Random variable vs. outcome

- Random variable: represents the actual outcome of an experiment — unknown
- Outcome: specific potential outcome of an experiment — known

First Mathematical Notions

Probability distribution

$p\colon S \to [0,1]$ such that $\sum_{e \in S} p(e) = 1$

First Mathematical Notions

Probability distribution

$p\colon S \to [0,1]$ such that $\sum_{e \in S} p(e) = 1$

Example

- Sample space $S = \{1, 2, 3, 4, 5, 6\}$
- Outcome 6
- Event $E = \{5, 6\}$
- $p(6) = \frac{1}{6}$

First Mathematical Notions

Probability distribution

$p\colon S \to [0,1]$ such that $\sum_{e \in S} p(e) = 1$

Example

- Sample space $S = \{1, 2, 3, 4, 5, 6\}$
- Outcome 6
- Event $E = \{5, 6\}$
- $p(6) = \frac{1}{6}$

Definition (Probability of an event)

Event $E \subseteq S$ and prob. distribution $p\colon S \to [0,1]$

$$p(E) = \sum_{e \in E} p(e)$$

Introductory Examples

Example (Coin toss)

- Experiment: Toss a coin twice
- Sample space: $\{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$ record sequence
- Prob. distribution: $p(\text{HH}) = p(\text{HT}) = p(\text{TH}) = p(\text{TT}) = \frac{1}{4}$

What is the probability of the event "at least once tails"?

Introductory Examples

Example (Coin toss)

- Experiment: Toss a coin twice
- Sample space: $\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$ record sequence
- Prob. distribution: $p(\mathrm{HH}) = p(\mathrm{HT}) = p(\mathrm{TH}) = p(\mathrm{TT}) = \frac{1}{4}$

What is the probability of the event "at least once tails"?

- Event: $E = \{\mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$
- $p(E) = \sum_{e \in E} p(e) = \frac{3}{4}$

Introductory Examples

Example (Another coin toss)

- Experiment: Toss 2 coins
- Sample space: $\{\{H\}, \{H, T\}, \{T\}\}$ record faces seen
- Prob. distribution:
  $p(\{H\}) = p(\{T\}) = \frac{1}{4}$
  $p(\{H, T\}) = \frac{1}{2}$

What is the probability of the event "no tails"?

Introductory Examples

Example (Another coin toss)

- Experiment: Toss 2 coins
- Sample space: $\{\{H\}, \{H, T\}, \{T\}\}$ record faces seen
- Prob. distribution:
  $p(\{H\}) = p(\{T\}) = \frac{1}{4}$
  $p(\{H, T\}) = \frac{1}{2}$

What is the probability of the event "no tails"?

- Event: $E = \{\{H\}\}$
- $p(E) = p(\{T\}) = \frac{1}{4}$

Introductory Examples

Example (Another coin toss)

- Experiment: Toss 2 coins
- Sample space: $\{\{\mathsf{H}\},\{\mathsf{H},\mathsf{T}\},\{\mathsf{T}\}\}$ record faces seen
- Prob. distribution:
  $p(\{\mathsf{H}\}) = p(\{\mathsf{T}\}) = \frac{1}{4}$
  $p(\{\mathsf{H},\mathsf{T}\}) = \frac{1}{2}$

What is the probability of the event "no tails"?

- Event: $E = \{\{\mathsf{H}\}\}$
- $p(E) = p(\{\mathsf{T}\}) = \frac{1}{4}$

Remark

$p(\{\mathsf{H}\}) = p(\{\mathsf{T}\}) = p(\{\mathsf{H},\mathsf{T}\}) = \frac{1}{3}$ is also a prob. distribution, but it does not capture the imagined experiment!

The First Problems

Example (Galileo, early 1600s)

Does a sum of 10 show up more often than a sum of 9 in a roll of 3 dice?

The First Problems

Example (Galileo, early 1600s)

Does a sum of 10 show up more often than a sum of 9 in a roll of 3 dice?

Hint

There are 25 and 27 triplets summing to 9 and 10, respectively.

The First Problems

Example (Galileo, early 1600s)

Does a sum of 10 show up more often than a sum of 9 in a roll of 3 dice?

Hint

There are 25 and 27 triplets summing to 9 and 10, respectively.

Historical remarks

- Buffon, 18th century: 4,040 coin tosses (2,048 H; 1,992 T)
- Weldon, 1894: 26,306 throws of 12 dice
- Wolf, ≈1884: 100,000 throws of a die

The First Problems

Example (Tversky, 1982)

In a large hospital 45 babies are born each day, and in a smaller hospital 15 babies are born each day. The overall proportion (over the year) of boys is about 50%. Which hospital will have the greater number of days in a year on which more than 60% of the babies born were boys?

The First Problems

Example (Tversky, 1982)

In a large hospital 45 babies are born each day, and in a smaller hospital 15 babies are born each day. The overall proportion (over the year) of boys is about 50%. Which hospital will have the greater number of days in a year on which more than 60% of the babies born were boys?

Answer

(a) the large hospital
(b) the small hospital
(c) neither

The First Problems

Example (Tversky, 1982)

In a large hospital 45 babies are born each day, and in a smaller hospital 15 babies are born each day. The overall proportion (over the year) of boys is about 50%. Which hospital will have the greater number of days in a year on which more than 60% of the babies born were boys?

Answer

60% of 45 and 15 are 27 and 9, respectively.

The First Problems

Example (Tversky, 1982)

In a large hospital 45 babies are born each day, and in a smaller hospital 15 babies are born each day. The overall proportion (over the year) of boys is about 50%. Which hospital will have the greater number of days in a year on which more than 60% of the babies born were boys?

Answer

60% of 45 and 15 are 27 and 9, respectively.

Assuming day-to-day independence, we can compute the probability for more than 60% boys for a single day:

- large hospital: 11.63%
- small hospital: 30.36%

A Difficult Problem

Example

We toss a coin 40 times. Every time heads comes up, you give me a cent, and every time tail comes up, I give you a cent.

(a) What is the most likely amount of cents won by me at the end?

(b) What is the most likely number of times I am in the lead?

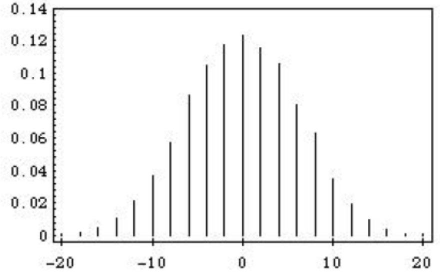
A Difficult Problem

Example

We toss a coin 40 times. Every time heads comes up, you give me a cent, and every time tail comes up, I give you a cent.

(a) What is the most likely amount of cents won by me at the end?

(b) What is the most likely number of times I am in the lead?



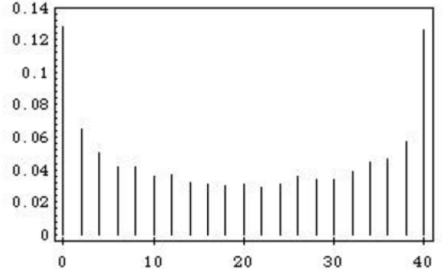
A Difficult Problem

Example

We toss a coin 40 times. Every time heads comes up, you give me a cent, and every time tail comes up, I give you a cent.

(a) What is the most likely amount of cents won by me at the end?

(b) What is the most likely number of times I am in the lead?



Elementary Properties

Theorem

prob. distribution $p\colon S \to [0,1]$

1. $p(E) \geq 0$ *for every event* $E \subseteq S$

Elementary Properties

Theorem

prob. distribution $p\colon S \to [0,1]$

1. $p(E) \geq 0$ *for every event* $E \subseteq S$
2. $p(S) = 1$

Elementary Properties

Theorem

prob. distribution $p\colon S \to [0,1]$

1. $p(E) \geq 0$ *for every event* $E \subseteq S$
2. $p(S) = 1$
3. $p(E) \leq p(E')$ *for all* $E \subseteq E' \subseteq S$

Elementary Properties

Theorem

prob. distribution $p\colon S \to [0,1]$

1. $p(E) \geq 0$ *for every event* $E \subseteq S$
2. $p(S) = 1$
3. $p(E) \leq p(E')$ *for all* $E \subseteq E' \subseteq S$
4. $p(E \cup E') = p(E) + p(E')$ *for all* $E, E' \subseteq S$ *with* $E \cap E' = \emptyset$

Elementary Properties

Theorem

prob. distribution $p\colon S \to [0,1]$

1. $p(E) \geq 0$ *for every event* $E \subseteq S$
2. $p(S) = 1$
3. $p(E) \leq p(E')$ *for all* $E \subseteq E' \subseteq S$
4. $p(E \cup E') = p(E) + p(E')$ *for all* $E, E' \subseteq S$ *with* $E \cap E' = \emptyset$
5. $p(S \setminus E) = 1 - p(E)$ *for every event* $E \subseteq S$

Elementary Properties

Theorem

prob. distribution $p\colon S \to [0,1]$

1. $p(E) \geq 0$ *for every event* $E \subseteq S$
2. $p(S) = 1$
3. $p(E) \leq p(E')$ *for all* $E \subseteq E' \subseteq S$
4. $p(E \cup E') = p(E) + p(E')$ *for all* $E, E' \subseteq S$ *with* $E \cap E' = \emptyset$
5. $p(S \setminus E) = 1 - p(E)$ *for every event* $E \subseteq S$
6. $p(E) = \frac{|E|}{|S|}$ *if* $p(e) = \frac{1}{|S|}$ *for every outcome* $e \in S$

Elementary Properties

Theorem

prob. distribution $p\colon S \to [0,1]$

1. $p(E) \geq 0$ *for every event* $E \subseteq S$
2. $p(S) = 1$
3. $p(E) \leq p(E')$ *for all* $E \subseteq E' \subseteq S$
4. $p(E \cup E') = p(E) + p(E')$ *for all* $E, E' \subseteq S$ *with* $E \cap E' = \emptyset$
5. $p(S \setminus E) = 1 - p(E)$ *for every event* $E \subseteq S$
6. $p(E) = \frac{|E|}{|S|}$ *if* $p(e) = \frac{1}{|S|}$ *for every outcome* $e \in S$

Remark

- If $p(e) = \frac{1}{|S|}$ for every outcome $e \in S$, then p is uniform
- in that case: probability of an event is ratio of positive outcomes to all outcomes

A Dice Problem

Example (Chevalier de Méré, Pascal, Fermat 1654)

What is the probability of at least one '6' in 4 rolls of a die?

Answers

(a) $\ll 50\%$ ($\leq \frac{2}{6} = \frac{1}{3}$)

(b) $< 50\%$

(c) $= 50\%$

(d) $> 50\%$

(e) $\gg 50\%$ ($\geq \frac{4}{6} = \frac{2}{3}$)

A Dice Problem

Example (Chevalier de Méré, Pascal, Fermat 1654)

What is the probability of at least one '6' in 4 rolls of a die?

Answers

(a) $\ll 50\%$ ($\leq \frac{2}{6} = \frac{1}{3}$)

(b) $< 50\%$

(c) $= 50\%$

(d) $> 50\%$

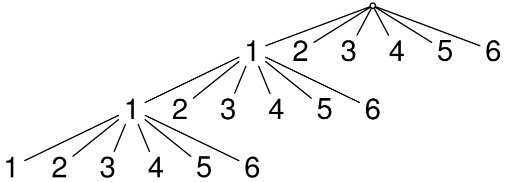
(e) $\gg 50\%$ ($\geq \frac{4}{6} = \frac{2}{3}$)

The probability is 0.52 (or 52%).

A Dice Problem

Example (Chevalier de Méré, Pascal, Fermat 1654)

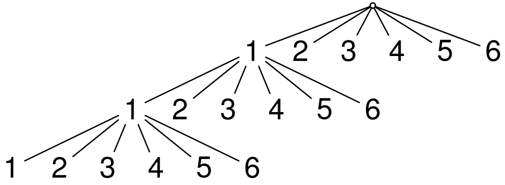
What is the probability of at least one '6' in 4 rolls of a die?



A Dice Problem

Example (Chevalier de Méré, Pascal, Fermat 1654)

What is the probability of at least one '6' in 4 rolls of a die?



1st roll $= 6$: $6^3 = 216$ positive outcomes remaining rolls irrelevant

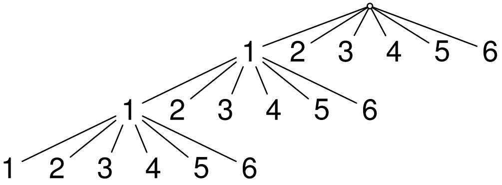
1st roll $\neq 6$:

- 2nd roll $= 6$: $6^2 = 36$ positive outcomes
- 2nd roll $\neq 6$: 5 outcomes
  - 3rd roll $= 6$: 6 positive outcomes
  - 3rd roll $\neq 6$: 5 outcomes
    - 4th roll $= 6$: 1 positive outcome
    - 4th roll $\neq 6$: 0 positive outcomes 5 outcomes

A Dice Problem

Example (Chevalier de Méré, Pascal, Fermat 1654)

What is the probability of at least one '6' in 4 rolls of a die?



positive outcomes:

$$p(E) = p(E_1) + p(E_2) + p(E_3) + p(E_4)$$
$$= \frac{216 + 5 \cdot 36 + 5^2 \cdot 6 + 5^3 \cdot 1}{1296} = \frac{671}{1296} = 0.5177$$

- E_1: '6' on 1st roll
- E_2: '6' on 2nd roll, but no '6' on 1st roll
- ...

Independent Events

Definition

Events $E, E' \subseteq S$ are independent if

$$p(E \cap E') = p(E) \cdot p(E')$$

Independent Events

Definition

Events $E, E' \subseteq S$ are independent if

$$p(E \cap E') = p(E) \cdot p(E')$$

Remarks

- Independence is often obvious
- but can also be tricky

Independent Events

Definition

Events $E, E' \subseteq S$ are independent if

$$p(E \cap E') = p(E) \cdot p(E')$$

Example (Die roll)

- E_1': no '6' on 1st roll
- E_2': no '6' on 2nd roll

E_1' and E_2' are independent.

Independent Events

Definition

Events $E, E' \subseteq S$ are independent if

$$p(E \cap E') = p(E) \cdot p(E')$$

Example (Coin toss)

2 tosses of a (fair) coin

- E_1: first toss is a head
- E_2: both tosses yield the same

$$\begin{aligned} p(E_1 \cap E_2) &= p(\mathrm{HH}) &&= \tfrac{1}{4} \\ p(E_1) &= p(\mathrm{HH}) + p(\mathrm{HT}) &&= \tfrac{1}{2} \\ p(E_2) &= p(\mathrm{HH}) + p(\mathrm{TT}) &&= \tfrac{1}{2} \end{aligned}$$

Independent Events

Definition

Events $E, E' \subseteq S$ are independent if

$$p(E \cap E') = p(E) \cdot p(E')$$

Example (Coin toss)

2 tosses of a (fair) coin

- E_1: first toss is a head
- E_2: both tosses yield the same

$$\begin{aligned} p(E_1 \cap E_2) &= p(\mathrm{HH}) &&= \tfrac{1}{4} \\ p(E_1) &= p(\mathrm{HH}) + p(\mathrm{HT}) &&= \tfrac{1}{2} \\ p(E_2) &= p(\mathrm{HH}) + p(\mathrm{TT}) &&= \tfrac{1}{2} \end{aligned}$$

E_1 and E_2 are independent.

Independent Events

Definition

Events $E, E' \subseteq S$ are independent if

$$p(E \cap E') = p(E) \cdot p(E')$$

Example (Sentence utterance)

1 spoken English sentence

- E_1: first word is 'I'
- E_2: second word is 'are'

Not independent!

Both individually rather likely, but together extremely unlikely.

$$p(E_1 \cap E_2) \neq p(E_1) \cdot p(E_2)$$

Back to the Dice Problem

Example (Chevalier de Méré, Pascal, Fermat 1654)

What is the probability of at least one '6' in 4 rolls of a die?

Answer

$p(E) = 1 - p(E') = 1 - p(E'_1) \cdot p(E'_2) \cdot p(E'_3) \cdot p(E'_4) = 1 - (\frac{5}{6})^4$

- E': no '6' on any roll
- E'_1: no '6' on 1st roll
- E'_2: no '6' on 2nd roll
- ...

Spice it up!

4 rolls of a single die expecting at least once a '6' is favorable.

Example (Chevalier de Méré, Pascal, Fermat 1654)

How many rolls of 2 dice are needed for a favorable game, when we expect to see at least once a pair of sixes?

Spice it up!

4 rolls of a single die expecting at least once a '6' is favorable.

Example (Chevalier de Méré, Pascal, Fermat 1654)

How many rolls of 2 dice are needed for a favorable game, when we expect to see at least once a pair of sixes?

Answer

- A '6' shows in $\frac{1}{6}$ of the cases
- A pair of sixes shows in $\frac{1}{36}$ of the cases.

4 rolls are sufficient to make the 1-die game favorable. Since the probability of a pair of sixes is 6 times as unlikely, we need 6 times more rolls, which gives $4 \cdot 6 = 24$ rolls.

Spice it up!

4 rolls of a single die expecting at least once a ‘6’ is favorable.

Example (Chevalier de Méré, Pascal, Fermat 1654)

How many rolls of 2 dice are needed for a favorable game, when we expect to see at least once a pair of sixes?

Answer

- A ‘6’ shows in $\frac{1}{6}$ of the cases
- A pair of sixes shows in $\frac{1}{36}$ of the cases.

4 rolls are sufficient to make the 1-die game favorable. Since the probability of a pair of sixes is 6 times as unlikely, we need 6 times more rolls, which gives $4 \cdot 6 = 24$ rolls.

Problem

But Chevalier de Méré felt that he was losing and at least 25 rolls were needed. → Failure of Mathematics

Roadmap

Basic Notions

Conditional Probabilities

Conditional probabilities — Intuition

Problem

Your fridge does not work.

Conditional probabilities — Intuition

Problem

Your fridge does not work.

Analysis

Cause	**Probability**
disconnected	0.4
fuse blown	0.2
motor broken	0.1
coolant leak	0.1
plug or cord broken	0.1
alien sabotage	0.05
...	...

Conditional probabilities — Intuition

Problem

Your fridge does not work.

Analysis

Cause	Probability
disconnected	0.4
fuse blown	0.2
motor broken	0.1
coolant leak	0.1
plug or cord broken	0.1
alien sabotage	0.05
...	...

Observation

The light inside is still on.

Conditional probabilities — Intuition

Problem

Your fridge does not work.

Analysis

Cause	**Probability**
disconnected	0
fuse blown	0
motor broken	0.1
coolant leak	0.1
plug or cord broken	0
alien sabotage	0.05
...	...

Observation

The light inside is still on.

Conditional probabilities

Definition

Given events $E, E' \subseteq S$, the probability that E happened given that we already observed E' is $p(E|E')$.

Conditional probabilities

Definition

Given events $E, E' \subseteq S$, the probability that E happened given that we already observed E' is $p(E|E')$.

Definition (Formally...)

$$p(E|E') = \frac{p(E \cap E')}{p(E')} \qquad\qquad p(E') \neq 0$$

Conditional probabilities

Definition
Given events $E, E' \subseteq S$, the probability that E happened given that we already observed E' is $p(E|E')$.

Definition (Formally...)

$$p(E|E') = \frac{p(E \cap E')}{p(E')} \qquad p(E') \neq 0$$

Example (Dice again)
uniform prob. distribution $p\colon \{1,2,3,4,5,6\} \to [0,1]$

- $E = \{6\}$ rolling a '6'
- $E' = \{4,5,6\}$ rolling at least a '4'

$$p(E|E') = \frac{p(E \cap E')}{p(E')} = \frac{1/6}{1/2} = \frac{1}{3}$$

Independence and Conditional Probabilities

Theorem

Two events E, E' are independent if and only if

- *both $p(E)$ and $p(E')$ are positive and $p(E|E') = p(E)$, or*
- *$p(E)$ or $p(E')$ is 0.*

Independence and Conditional Probabilities

Theorem

Two events E, E' are independent if and only if

- *both* $p(E)$ *and* $p(E')$ *are positive and* $p(E|E') = p(E)$*, or*
- $p(E)$ *or* $p(E')$ *is* 0*.*

Intuitive reading

For independent events, knowledge of one event does not affect the probability of the other.

Another Problem

Example

A sick woman sees the doctor, who runs 2 positive tests ($++$) and then looks up his clinical studies:

disease	affected	$++$	$+-$	$-+$	$--$
d_1	3,215	2,110	301	704	100
d_2	2,125	396	132	1,187	410
d_3	4,660	510	3,568	73	509
total	10,000	3,016	4,001	1,964	1019

Another Problem

Example

A sick woman sees the doctor, who runs 2 positive tests ($++$) and then looks up his clinical studies:

disease	affected	$++$	$+-$	$-+$	$--$
d_1	3,215	2,110	301	704	100
d_2	2,125	396	132	1,187	410
d_3	4,660	510	3,568	73	509
total	10,000	3,016	4,001	1,964	1019

Estimation

$p(d_1) = 32.15\%$ $\quad p(d_2) = 21.25\%$ $\quad p(d_3) = 46.60\%$

$p(++|d_1) = \frac{2110}{3215} = 65.63\%$ $\quad p(++|d_2) = \frac{396}{2125} = 18.64\%$

$p(++|d_3) = \frac{510}{4660} = 10.94\%$

Another Problem

Example

A sick woman sees the doctor, who runs 2 positive tests ($++$) and then looks up his clinical studies:

disease	affected	$++$	$+-$	$-+$	$--$
d_1	3,215	2,110	301	704	100
d_2	2,125	396	132	1,187	410
d_3	4,660	510	3,568	73	509
total	10,000	3,016	4,001	1,964	1019

Estimation

$p(d_1) = 32.15\%$ $\quad p(d_2) = 21.25\%$ $\quad p(d_3) = 46.60\%$

$p(++|d_1) = \frac{2110}{3215} = 65.63\%$ $\quad p(++|d_2) = \frac{396}{2125} = 18.64\%$

$p(++|d_3) = \frac{510}{4660} = 10.94\%$

What is the most likely disease of the woman?

Bayes Rule

Given $p(E|E')$, $p(E)$, and $p(E')$ we want to compute $p(E'|E)$

Bayes Rule

Given $p(E|E')$, $p(E)$, and $p(E')$ we want to compute $p(E'|E)$

Bayes formula

$p(E')$ prior — probability before E happened
$p(E'|E)$ posterior — probability after E happened

$$p(E'|E) = \frac{p(E' \cap E)}{p(E)} \qquad p(E) \neq 0$$

$$= \frac{p(E \cap E') \cdot p(E')}{p(E) \cdot p(E')} \qquad p(E) \cdot p(E') \neq 0$$

$$=$$

Bayes Rule

Given $p(E|E')$, $p(E)$, and $p(E')$ we want to compute $p(E'|E)$

Bayes formula

$p(E')$ prior — probability before E happened
$p(E'|E)$ posterior — probability after E happened

$$p(E'|E) = \frac{p(E' \cap E)}{p(E)} \qquad p(E) \neq 0$$

$$= \frac{p(E \cap E') \cdot p(E')}{p(E) \cdot p(E')} \qquad p(E) \cdot p(E') \neq 0$$

$$= \frac{p(E|E') \cdot p(E')}{p(E)} \qquad p(E) \cdot p(E') \neq 0$$

Bayes Rule

Given $p(E|E')$, $p(E)$, and $p(E')$ we want to compute $p(E'|E)$

Bayes formula

$p(E')$ prior — probability before E happened
$p(E'|E)$ posterior — probability after E happened

$$p(E'|E) = \frac{p(E' \cap E)}{p(E)} \qquad p(E) \neq 0$$

$$= \frac{p(E \cap E') \cdot p(E')}{p(E) \cdot p(E')} \qquad p(E) \cdot p(E') \neq 0$$

$$= \frac{p(E|E') \cdot p(E')}{p(E)} \qquad p(E) \cdot p(E') \neq 0$$

$$p(E'|E) = \frac{p(E|E') \cdot p(E')}{p(E)} \qquad p(E) \neq 0,\ p(E') \neq 0$$

Let us Help The Sick Lady

$$p(d_1|{+}{+}) = \frac{p({+}{+}|d_1) \cdot p(d_1)}{p({+}{+})} = \frac{2110/3215 \cdot 0.3215}{0.3016} = 69.96\%$$

$$p(d_2|{+}{+}) = \frac{p({+}{+}|d_2) \cdot p(d_2)}{p({+}{+})} = \frac{396/2125 \cdot 0.2125}{0.3016} = 13.13\%$$

$$p(d_3|{+}{+}) = \frac{p({+}{+}|d_3) \cdot p(d_3)}{p({+}{+})} = \frac{510/4660 \cdot 0.4660}{0.3016} = 16.91\%$$

Let us Help The Sick Lady

$$p(d_1|{+}{+}) = \frac{p({+}{+}|d_1) \cdot p(d_1)}{p({+}{+})} = \frac{2110/3215 \cdot 0.3215}{0.3016} = 69.96\%$$

$$p(d_2|{+}{+}) = \frac{p({+}{+}|d_2) \cdot p(d_2)}{p({+}{+})} = \frac{396/2125 \cdot 0.2125}{0.3016} = 13.13\%$$

$$p(d_3|{+}{+}) = \frac{p({+}{+}|d_3) \cdot p(d_3)}{p({+}{+})} = \frac{510/4660 \cdot 0.4660}{0.3016} = 16.91\%$$

Most likely she suffers from d_1.

Beware of Low Priors

Example

A doctor tests for a specific cancer that 1 in a 1,000 people suffer from with a test that is 99% sensitive and 95% specific. What is the probability of cancer given a positive test result?

Beware of Low Priors

Example

A doctor tests for a specific cancer that 1 in a 1,000 people suffer from with a test that is 99% sensitive and 95% specific. What is the probability of cancer given a positive test result?

Analysis

$S = \{(\text{cancer}, -), (\text{cancer}, +), (\text{clean}, -), (\text{clean}, +)\}$

Events:

- $\text{Cancer} = \{(\text{cancer}, -), (\text{cancer}, +)\}$
- $\ldots$

Interpretation:

- 99% sensitive $\rightarrow p(+|\text{Cancer}) = 0.99$
- 95% specific $\rightarrow p(-|\text{Clean}) = 0.95$

Beware of Low Priors

Example

A doctor tests for a specific cancer that 1 in a 1,000 people suffer from with a test that is 99% sensitive and 95% specific. What is the probability of cancer given a positive test result?

Analysis

$$
\begin{aligned}
p(\text{Cancer}|+) &= \frac{p(+|\text{Cancer}) \cdot p(\text{Cancer})}{p(+)} \\
&= \frac{p(+|\text{Cancer}) \cdot p(\text{Cancer})}{p(+|\text{Cancer}) \cdot p(\text{Cancer}) + p(+|\text{Clean}) \cdot p(\text{Clean})} \\
&= \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.05 \cdot 0.999} = 0.019
\end{aligned}
$$

Beware of Low Priors

Example

A doctor tests for a specific cancer that 1 in a 1,000 people suffer from with a test that is 99% sensitive and 95% specific. What is the probability of cancer given a positive test result?

Analysis

$$\begin{aligned}
p(\text{Cancer}|+) &= \frac{p(+|\text{Cancer}) \cdot p(\text{Cancer})}{p(+)} \\
&= \frac{p(+|\text{Cancer}) \cdot p(\text{Cancer})}{p(+|\text{Cancer}) \cdot p(\text{Cancer}) + p(+|\text{Clean}) \cdot p(\text{Clean})} \\
&= \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.05 \cdot 0.999} = 0.019
\end{aligned}$$

Only 1.9% of positively tested people have cancer; 98.1% are false positives.

Beware of Low Priors

Example

A doctor tests for a specific cancer that 1 in a 1,000 people suffer from with a test that is 99% sensitive and 95% specific. What is the probability of cancer given a positive test result?

So this test is useless?

Beware of Low Priors

Example

A doctor tests for a specific cancer that 1 in a 1,000 people suffer from with a test that is 99% sensitive and 95% specific. What is the probability of cancer given a positive test result?

So this test is useless?

$$
\begin{aligned}
p(\text{Clean}|-) &= \frac{p(-|\text{Clean}) \cdot p(\text{Clean})}{p(-)} \\
&= \frac{p(-|\text{Clean}) \cdot p(\text{Clean})}{p(-|\text{Cancer}) \cdot p(\text{Cancer}) + p(-|\text{Clean}) \cdot p(\text{Clean})} \\
&= \frac{0.95 \cdot 0.999}{0.01 \cdot 0.001 + 0.95 \cdot 0.999} = 0.999989
\end{aligned}
$$

Beware of Low Priors

Example

A doctor tests for a specific cancer that 1 in a 1,000 people suffer from with a test that is 99% sensitive and 95% specific. What is the probability of cancer given a positive test result?

So this test is useless?

$$p(\text{Clean}|-) = \frac{p(-|\text{Clean}) \cdot p(\text{Clean})}{p(-)}$$

$$= \frac{p(-|\text{Clean}) \cdot p(\text{Clean})}{p(-|\text{Cancer}) \cdot p(\text{Cancer}) + p(-|\text{Clean}) \cdot p(\text{Clean})}$$

$$= \frac{0.95 \cdot 0.999}{0.01 \cdot 0.001 + 0.95 \cdot 0.999} = 0.999989$$

So with a negative test, you can be 99.9989% certain to be cancer-free.

Final Problem

Example (vos Savant 1996)

On the night before the final exam, two students were partying in another state and didn't get back until it was over. Their excuse was that they had a flat tire, and they asked if they could take a make-up test. The professor agreed, wrote out a test and sent them to separate rooms. The first question was worth 5 points, and they answered it easily. The second question, worth 95 points, was: *'Which tire was it?'*

What is the probability that both students answer equally?

Final Problem

Example (vos Savant 1996)

On the night before the final exam, two students were partying in another state and didn't get back until it was over. Their excuse was that they had a flat tire, and they asked if they could take a make-up test. The professor agreed, wrote out a test and sent them to separate rooms. The first question was worth 5 points, and they answered it easily. The second question, worth 95 points, was: *'Which tire was it?'*

What is the probability that both students answer equally?

Answer

(a) $\frac{1}{16}$ (6.25%)
(b) $\frac{1}{4}$ (25%)
(c) $\frac{1}{2}$ (50%)

Final Problem

Example (vos Savant 1996)

On the night before the final exam, two students were partying in another state and didn't get back until it was over. Their excuse was that they had a flat tire, and they asked if they could take a make-up test. The professor agreed, wrote out a test and sent them to separate rooms. The first question was worth 5 points, and they answered it easily. The second question, worth 95 points, was: *'Which tire was it?'*

What is the probability that both students answer equally?

Answer

(a) $\frac{1}{16}$ (6.25%)
(b) $\frac{1}{4}$ (25%)
(c) $\frac{1}{2}$ (50%)

That's all folks!

Thank you for the attention.