Einführung in die Computerlinguistik
Fraser: Probabilities

Alexander Fraser and Robert Zangenfeind

Center for Information and Language Processing

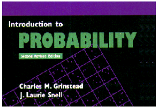
2020-01-08

Dieses Foliensatz wurde von Prof. Dr. Andreas Maletti (Leipzig) erstellt.

Fehler und Mängel sind ausschließlich meine Verantwortung.

Today's goals

- Basic notions and intuitive understanding
- Basic laws and computing with probabilities
- Independence
- Conditional probabilities
- Bayes' law
- Modelling complications



Charles M. Grinstead, J. Laurie Snell
Introduction to Probability
2nd edition, AMS 1997
▸ Website

Organization

Please ask questions immediately!

First notions

§1.1 Definition (probability)

The (theoretical) *probability* of an event E in an experiment $\mathcal{E}$ is the expected relative frequency of occurrence of the event E in many executions of $\mathcal{E}$.

Notions:

- Experiment: activity that yields exactly one of a finite number of outcomes (discrete probability distribution)
- Event: subset of the outcomes

First notions

- Experiment: Dice throw (with a standard fair die)
- Set of outcomes is $\mathcal{E} = \{1, \ldots, 6\}$ (at the same time the safe event)
- the event $E \subseteq \mathcal{E}$ is $E = \{5, 6\}$ (roll of a '5' or '6')
- Probability of the event E is $\frac{2}{6} = \frac{1}{3}$

Mathematical formulation

§1.2 Definition (probability measure)

Let $\mathcal{E}$ be a finite set of outcomes.
Then $p \colon \mathcal{P}(\mathcal{E}) \to [0,1]$ is a probability measure, if

- $p(\mathcal{E}) = 1$ (the safe event always occurs)
- $p(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} p(E_i)$ for all $n \in \mathbb{N}$ and all pairwise disjoint events $E_1, \ldots, E_n \subseteq \mathcal{E}$.

Notes:

- yields the known laws for computing with probabilities
- the values $p(E)$ for singletons $E \subseteq \mathcal{E}$ define p uniquely

Mathematical formulation

- Experiment: 2-fold coin toss, sequential
- Outcomes $\mathcal{E} = \{$, , , $\}$
- Probability measure
  $p(\{$$\}) = p(\{$$\}) = p(\{$$\}) = p(\{$$\}) = \frac{1}{4}$

What is the probability of the event “at least once tails”?

- Event $E = \{$, , $\}$

$p(E) = \sum_{e \in E} p(\{e\}) = p(\{$$\}) + p(\{$$\}) + p(\{$$\}) = \frac{3}{4}$

Mathematical formulation

- Experiment: Coin toss with 2 equal coins, simultaneously
- Outcomes $\mathcal{E} = \{\{$$\}, \{$$,$$\}, \{$$\}\}$ (seen faces)
- Probability measure

  $p(\{\{$$\}\}) = p(\{\{$$\}\}) = \frac{1}{4}$

  $p(\{\{$$,$$\}\}) = \frac{1}{2}$

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

- Event $E = \{\{$$\}\}$

$p(E) = p(\{\{$$\}\}) = \frac{1}{4}$

Note:

$p(\{\{$$\}\}) = p(\{\{$$\}\}) = p(\{\{$$,$$\}\}) = \frac{1}{3}$

also yields a probability measure, but it does not fit the experiment.

First little problems

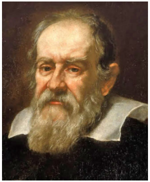
Problem [Galilei, 17th century]

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.

Galileo Galilei (* 1564; † 1641)

- Italian polymath
- Founder of exact natural sciences
- had “beef” with Catholic church

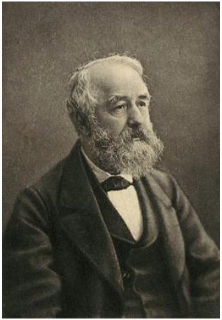


First little problems

Historical experiments:

- Buffon, 18th century: 4,040 coin tosses (2,048 ; 1,992 )
- Wolf, approx. 1884: 100,000 dice rolls
- Weldon, 1894: 26,306 rolls of 12 dice

Georges-Louis Leclerc, Comte de Buffon
(* 1707; † 1788)



Johann Rudolf Wolf
(* 1816; † 1893)

Walter Frank Raphael Weldon
(* 1860; † 1906)

First little problems

Problem [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?

Approach: 60% of 45 and 15 are 27 and 9, respectively. The probability for at least 27 times tails in 45 coin tosses is 11,6%, whereas at least 9 times tails in 15 tosses is more probable at 30,4%.

Amos Tversky (* 1937; † 1996)

- Israeli psychologist
- systematic analysis of human risk behavior
- would have received the Nobel prize in 2002

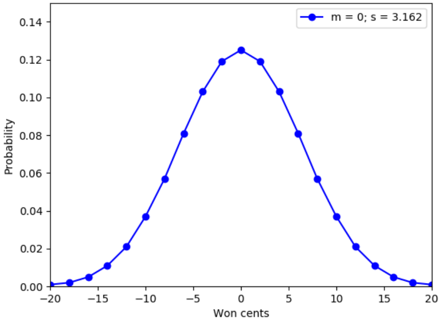
A more difficult problem

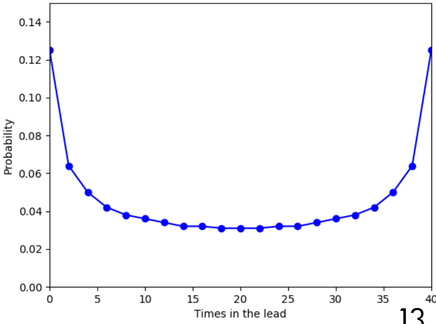
Problem

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





Basic laws

§1.3 Theorem

For every probability measure $p\colon \mathcal{E} \to [0,1]$:

1. $p(E) \geq 0$ for every event $E \subseteq \mathcal{E}$
2. $p(E) \leq p(E')$ for all $E \subseteq E' \subseteq \mathcal{E}$
3. $p(\mathcal{E} \setminus E) = 1 - p(E)$ for every event $E \subseteq \mathcal{E}$
4. $p(E) = \frac{|E|}{|\mathcal{E}|}$ if $p(\{e\}) = \frac{1}{|\mathcal{E}|}$ for every outcome $e \in \mathcal{E}$

Notes:

- If $p(\{e\}) = \frac{1}{|\mathcal{E}|}$ for every outcome $e \in \mathcal{E}$, then p is uniform
- In this case the probability of an event is the relation of positive outcomes to all outcomes.

A dice problem

Problem [Chevalier de Méré, Pascal, Fermat, 1654]

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

Answers:

(a) $[0, \frac{1}{3}]$

(b) $(\frac{1}{3}, \frac{1}{2})$

(c) $\frac{1}{2}$

(d) $(\frac{1}{2}, \frac{2}{3})$

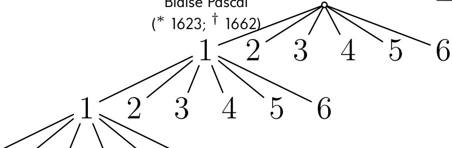
(e) $[\frac{2}{3}, 1]$

It is 52%.

Antoine Gombaud,
Chevalier de Méré
(* 1607; † 1684)

Blaise Pascal
(* 1623; † 1662)

Pierre de Fermat
(* 1601; † 1665)



Independence

§1.4 Definition (independence)

Two events $E, E' \subseteq \mathcal{E}$ are independent if

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

Notes:

- Independence is often obvious
- but can also be tricky

Independence

- Dice rolling: 2 rolls of a die
  Events "no '6' on 1st roll" and "no '6' on 2nd roll"
  are independent.
- Coin toss: 2 tosses of a fair coin
  E_1: "1st toss is heads" and E_2: "Both tosses yield same result"

$p(E_1 \cap E_2) = p(\{$  $\}) = \frac{1}{4}$

$p(E_1) = p(\{$  $\}) + p(\{$  $\}) = \frac{1}{2}$

$p(E_2) = p(\{$  $\}) + p(\{$  $\}) = \frac{1}{2}$

E_1 and E_2 are independent.

Independence

- Linguistics: 1 English sentence from a standard volume
  E_1: "first word is 'I'" and E_2: "second word is 'are'"

  Not independent, because the individual events are reasonably likely whereas it is extremely unlikely that both occur at the same time.

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

Back to the dice problem

Problem [Chevalier de Méré, Pascal, Fermat, 1654]

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

Answer:

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

- E': never a '6'
- E'_1: no '6' on 1st roll
- E'_2: no '6' on 2nd roll

Back to the dice problem

Problem [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:

$$0.5 \geq (\tfrac{35}{36})^n \qquad \to \qquad n \geq \log_{\frac{35}{36}} 0.5 = \frac{-\log 2}{\log 35 - \log 36} \approx 24.6$$

- Although 4 rolls are sufficient for a single six and a pair of sixes is exactly 6 times less likely
- $4 \cdot 6 = 24$ rolls are still insufficient for a favorable game
- as Chevalier de Méré determined empirically

$\to$ Failure of Mathematics

Conditional probability

Problem

The fridge does not work.

Analysis:

Reason	probability
not plugged	0,40
fuse blown	0,20
motor broken	0,1
coolant leak	0,1
cable or plug broken	0,10
alien sabotage	0,05
...	...

Observation: The inside light still works.

Conditional probability

§1.5 Definition (conditional probability)

Given events $E, E' \subseteq \mathcal{E}$, the probability of E given that E' already happened is

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

1 die:

- $E = \{6\}$ (roll of a '6')
- $E' = \{4, 5, 6\}$ (roll of '4' or more)

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

§1.6 Theorem

Two events $E, E' \subseteq \mathcal{E}$ are independent if and only if

- both $p(E)$ and $p(E')$ are positive and $p(E \mid E') = p(E)$, or
- $p(E) = 0$ or $p(E') = 0$.

Intuition:
The occurrence of an independent event E'
does not influence the probability of the event E.

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	1,019

Estimations:

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

$p(++ \mid d_1) = \frac{2{,}110}{3{,}215} = 65.63\%$ $\quad p(++ \mid d_2) = \frac{396}{2{,}125} = 18.64\%$

$p(++ \mid d_3) = \frac{510}{4{,}660} = 10.94\%$

What is the most likely disease of the woman?

Bayes' rule

§1.7 Theorem (Bayes' rule)

For two events $E, E' \subseteq \mathcal{E}$, we call $p(E')$ the prior (probability of E' before E) and $p(E' \mid E)$ the posterior (probability of E' after E).

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

Thomas Bayes (* 1701; † 1761)

- English statistician and minister
- Bayes' rule not published by him
- Fellow of the Royal Society

Back to the sick woman

$$p(d_1 \mid ++) = \frac{p(++ \mid d_1) \cdot p(d_1)}{p(++)} = \frac{2,110/3,215 \cdot 0.3215}{0.3016} = 69.96\%$$

$$p(d_2 \mid ++) = \frac{p(++ \mid d_2) \cdot p(d_2)}{p(++)} = \frac{396/2,125 \cdot 0.2125}{0.3016} = 13.13\%$$

$$p(d_3 \mid ++) = \frac{p(++ \mid d_3) \cdot p(d_3)}{p(++)} = \frac{510/4,660 \cdot 0.4660}{0.3016} = 16.91\%$$

Most likely she suffers from disease d_1.

Beware of low priors

Problem

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: $\mathcal{E} = \{(♋, -), (♋, +), (\checkmark, -), (\checkmark, +)\}$

Events:

- $♋ = \{(♋, -), (♋, +)\}$
- …

Interpretation:

- 99% sensitive $\rightarrow p(+ \mid ♋) = 0.99$
- 95% specific $\rightarrow p(- \mid \checkmark) = 0.95$

$$p(♋ \mid +) = \frac{p(+ \mid ♋) \cdot p(♋)}{p(+)}$$

$$= \frac{p(+ \mid ♋) \cdot p(♋)}{p(+ \mid ♋) \cdot p(♋) + p(+ \mid \checkmark) \cdot p(\checkmark)}$$

Beware of low priors

Problem

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 the test is useless?

$$p(\checkmark \mid -) = \frac{p(- \mid \checkmark) \cdot p(\checkmark)}{p(-)}$$

$$= \frac{p(- \mid \checkmark) \cdot p(\checkmark)}{p(- \mid ♋) \cdot p(♋) + p(- \mid \checkmark) \cdot p(\checkmark)}$$

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

Based on a negative test result you can be 99.9989% sure to be unaffected.

Final problem

Problem [vos Savant, 1996]

On the night before the final exam, two students were partying in another state and did not 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?

Answers:

(a) $\frac{1}{16}$ (6.25%)

(b) $\frac{1}{4}$ (25%)

(c) $\frac{1}{2}$ (50%)

Marilyn vos Savant (* 1946)

- American author
- Highest Guinness book IQ (228)
- Category suspended since

That’s all folks!

Thank you for the attention.