Probability theory

B. Jacobs

Institute for Computing and Information Sciences – Digital Security
Radboud University Nijmegen

Version: fall 2014

Outline

Historical background

"Probability" is the part of mathematics which looks for laws governing random events. It has its origins in games of chance i.e. in gambling.
Chevalier de Méré (1607-1684) was a famous gambler and a friend of Blaise Pascal, who started to develop probability theory

Example (Question about rolling dices)

What is more likely to get:

1. at least one 6 in 4 rolls of one dice
2. at least one pair (6,6) in 24 simultaneous rolls of two dice?

Chevalier expected (2), and lost money as a result.

- $p_1 = 1 - (\frac{5}{6})^4 \approx 0.518$ (or 51.8% chance)
- $p_2 = 1 - (\frac{35}{36})^{24} \approx 0.491$

Another de Méré-like challenge (from teacherlink.org)

Would you take the following bet, about repeatedly rolling two dice:

"I will get both a sum 8 and a sum 6,
before you get two sums of 7."

If you take it, I win and you loose

Consider all possible sums as outcomes:

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The catch: the order of the 8 and the 6 are not specified: the probability of (6,8) or (8,6) is higer than the probability of (7,7).

Combinatorics = smart counting

Combinatorics is a branch of mathematics that studies counting, typically in finite structures, of objects satisfying certain criteria.

Example (Counting permutations)

- A permutation of n-objects is a rearrangement in some order
- **Question**: how many different permutations are there of n-objects?
  - Try to think of the answer for $n = 2, 3, 4, \ldots$
- The **answer** is $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$
  - Pronounce: $n!$ as "n factorial"
  - For those who like recursion: $n! = n \cdot (n-1)!$ and $0! = 1$.
- Interestingly, each permutation of n corresponds to a particular ordering of n objects; we will use this later

Fundamental principle of (successive) counting

- Suppose that a task involves a sequence of k successive choices
  - let n_1 be the number of options at the first stage;
  - let n_2 be the number of options at the second stage, after the first stage has occurred;
  - ...
  - let n_k term be the number of options at the k-th stage, after the previous $k-1$ stages have occurred.
- Then the total number of different ways the task can occur is:

$$n_1 \cdot n_2 \cdot \ldots \cdot n_k = \prod_{1 \le i \le k} n_i$$

Simple counting example

A company places a 6-symbol code on each unit of its products, consisting of:

- 4 digits, the first of which is the number 5,
- followed by 2 letters, the first of which is NOT a vowel.

How many different codes are possible?

Using the basic counting principle:

- there are 10 options (decimals) for digits 2, 3, 4
- there are 26 letters in the alphabet, 26 options for letter 2
- 5 of the letters in the alphabet are vowels (a, e, i, o, u), so that means there are 21 options for letter 1

Altogether there are $10 \cdot 10 \cdot 10 \cdot 21 \cdot 26 = 546,000$ different codes.

Samples *(grepen)*

We will study the following four combinations of samples

Samples	Ordered	Unordered
With replacement	I	III
Without replacement	II	IV

Ad I ordered samples with replacement

Question

- Suppose you have n objects, and you take an ordered sample with replacement of r out of them (with $r \le n$)
- This means that the order of the selected r elements matters, and the same element may be selected multiple times
- How many such samples are there?

Example (2-samples out of 3 elements, say $\{1, 2, 3\}$)

- samples: 11, 12, 13, 21, 22, 23, 31, 32, 33
- number of samples: $9 = 3^2$

Lemma

There are n^r ordered samples with replacement

Ad II ordered samples without replacement

- With replacement we can reason as follows
  - for the first item of the sample, there are n options
  - for the second item of the sample, there are still n options
  - etc.

  This gives n^r samples in total
- Without replacement we now reason:
  - for the first item of the sample, there are n options
  - for the second item of the sample, there are only $n-1$ options
  - for the third item of the sample, there are only $n-2$ options
  - etc.

Lemma

There are $n \cdot (n-1) \cdot (n-2) \cdots (n-r+1) = \dfrac{n!}{(n-r)!}$ ordered samples without replacement.

Ad II Example (ordered, without replacement)

In how many ways can 10 people be seated on a bench with 4 seats?

Answer:

- We have $n = 10$, from which we take samples of size $r = 4$
- The order matters, and people who are already seated cannot be seated again: no replacement
- Number of options: $10 \cdot 9 \cdot 8 \cdot 7 = 5040 = \dfrac{10!}{6!} = \dfrac{10!}{(10-4)!}$

Ad IV unordered samples without replacement

Recall two things:

- there are $r!$ ways to order/permute r items
- there are $\frac{n!}{(n-r)!}$ ordered samples without replacement

Combining these two yields:

Lemma

There are $\dfrac{n!}{r!(n-r)!}$ *unordered samples without replacement.*

One writes $\dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$. *This is called the binomial coefficient.*

It is pronounced as "n choose r" or as "n over r".

An unordered sample is sometimes called a combination.

Ad IV Examples (of unordered samples without replacement)

Example (Lotto with 49 numbered balls)

How many possible outcomes are there if we consecutively take out 6 balls?

Answer: $\binom{49}{6} = 13,983,816$

Example

Find the number of ways to form a committee of 5 people from a set of 9.

Answer: $\binom{9}{5} = 126$. (what is the difference with the bench example?)

Example

How many symmetric keys are needed so that n people can all communicate directly with each other?

Answer: $\binom{n}{2} = \frac{n(n-1)}{2} = (n-1) + (n-2) + \cdots + 2 + 1$

Calculation rules for binomial coefficients

1. $\binom{n}{r} = \binom{n}{n-r}$
2. $\displaystyle\sum_{r=0}^{r=n} \binom{n}{r} = 2^n$
3. $\binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r}$

Recall also Pascal's triangle

$$
\begin{array}{ccccc}
 & & \binom{0}{0} & & \\
 & \binom{1}{0} & & \binom{1}{1} & \\
\binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} \\
\vdots & & \vdots & & \vdots
\end{array}
$$

Binomial expansion of powers of sums

- Recall: $(x+y)^2 = x^2 + 2xy + y^2$
  $= \binom{2}{0}x^2y^0 + \binom{2}{1}x^1y^1 + \binom{2}{2}x^0y^2$
- Similarly:
  $(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$
  $= \binom{3}{0}x^3y^0 + \binom{3}{1}x^2y^1 + \binom{3}{2}x^1y^2 + \binom{3}{3}x^0y^3$

Lemma

For arbitrary $n \in \mathbb{N}$,

$$(x+y)^n = \sum_{i=0}^{i=n} \binom{n}{i} x^{n-i} y^i$$

Ad III unordered samples with replacement

Now the number of options is: $\binom{n+r-1}{r}$

Example (Lotto with 10 numbered balls, pick and replace 2)

- How many outcomes xx? 10
- How many outcomes $xy \sim yx$? $45 = \frac{10 \cdot 9}{2} = \binom{10}{2}$

Total: $10 + 45 = 55 = \frac{11 \cdot 10}{2} = \binom{11}{2} = \binom{10+2-1}{2}$ indeed!

Note that with the earlier calculation rules:

$$\binom{11}{2} = \binom{10}{2} + \binom{10}{1} = 45 + 10 = 55$$

Ad III Example (unordered samples with replacement)

Example (Lotto with 10 numbered balls, pick and replace 3)

- How many outcomes xxx? 10
- How many outcomes $xyy \sim yxy \sim yyx$? $10 \cdot 9 = 90$
- How many $xyz \sim xzy \sim yxz \sim yzx \sim zxy \sim zyx$? $\frac{10 \cdot 9 \cdot 8}{6} = \binom{10}{3} = 120$

Total: $10 + 90 + 120 = 220 = \frac{12 \cdot 10 \cdot 11}{3 \cdot 2} = \binom{12}{3} = \binom{10+3-1}{3}$ Indeed!

Again with the earlier calculation rules:

$$
\begin{aligned}
\binom{12}{3} &= \binom{11}{3} + \binom{11}{2} \\
&= \binom{10}{3} + \binom{10}{2} + \binom{10}{2} + \binom{10}{10} \\
&= 120 + 45 + 45 + 10.
\end{aligned}
$$

Birthday paradox

1. What is the probability that at least 2 of r randomly selected people have the same birthday?
2. How large must r be so that the probability is greater than 50%?

Solution, part I

- Assume that no one is born on Feb. 29 and that all birthdays are equally distributed.
  - $n = 365$
  - we look at samples of r, which are ordered, with replacement (once a birthday occurs, it is not excluded, since it can occur again)
  - $n^r = 365^r$ birthday options for r people
- Look at r birthdays, all at different days
  - number of options: $365 \cdot 364 \cdots (366 - r) = \frac{365!}{(365-r)!} = \binom{365}{r} r!$
  - take fraction: the probability that r people have their birthday on different days is:

$$\frac{\frac{365!}{(365-r)!}}{365^r} = \frac{365!}{(365-r)! \cdot 365^r}$$

- Therefore, the probability that at least 2 people out of r have their birthday on the same day is $p(r) = 1 - \frac{365!}{(365-r)! \cdot 365^r}$

Solution, part II

Some values for $p(r) = 1 - \frac{365!}{(365-r)! \cdot 365^r}$, depending on r.

r	$p(r)$
10	0.117
20	0.411
23	0.507
30	0.706
50	0.97
57	0.99

Hence for $r = 23$ the probability of birthday-coincidence is $\geq 50\%$.

Application: Birthday attacks on hash functions

- SHA1 with a 160 bit output requires brute-force work of at most 2^{80} operations
  - (although because of weaknesses in SHA1 collisions are found already in around 2^{60} steps)
- In general hash functions used for signature schemes should have the number of output bits n large enough such that $2^{n/2}$ computations are impractical

Note: With 8M budget an 80-bit key can be retrieved in a year (2011).

Experiments and their sample spaces

- An experiment is called random if the result will vary even if the conditions are the same
- A sample space consists of all possible outcomes of a random experiment, usually denoted with the letter S or Ω

Example (What are the relevant sample spaces?)

1. coin tossing once: $S = \{T, H\}$
2. coin tossing twice: $S = \{TT, HT, TH, HH\}$
3. die tossing: $S = \{1, 2, 3, 4, 5, 6\}$
4. lifetime of a bulb: $S = \{t \mid 0 \leq t \leq 1\,year\}$

(Oxford dictionary: Historically, dice is the plural of die, but in modern standard English dice is both the singular and the plural)

Events

Definition

An event is a subset of outcomes of a random experiment, that is, a subset of the sample space.

We write the powerset $\mathcal{P}(S) = \{A \mid A \subseteq S\}$ for the set of events.

Example (for sample space S)

- the entire subset $S \subseteq S$ is the "certain" event
- $\emptyset \subseteq S$ is the impossible event
- two events A and B are mutually exclusive if $A \cap B = \emptyset$.

Probability measure

Definition

A probability measure P for a sample space S is a function that gives for each event $A \subseteq S$ a probability $P(A) \in [0,1]$, with:

1. Axiom 1: $P(S) = 1$
2. Axiom 2: $P(A \cup B) = P(A) + P(B)$ for mutually exclusive events $A, B \subseteq S$, that is, when $A \cap B = \emptyset$

A probability measure on S is thus a function $P\colon \mathcal{P}(S) \to [0,1]$ satisfying (1) and (2).

It is called discrete if the sample space S is finite; this implies that there only finitely many events.

(Officially, discrete spaces can also be countable, but we shall not use those here)

Properties of probability measures

Theorem

Let P be a probability measure on space S, and let A, A_i, B, be events. Then:

1. $A \subseteq B \Rightarrow P(A) \leq P(B)$
2. $P(\emptyset) = 0$
3. $P(\neg A) = 1 - P(A)$, *where* $\neg A = S - A = \{s \in S \mid s \notin A\}$
4. *For mutually exclusive events* $A_1, A_2, \ldots, A_n$, *where* $A_i \cap A_j = \emptyset$, *for all* $i \neq j$, *one has* $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. $P(A) = P(A \cap B) + P(A \cap \neg B)$

The points can all be derived from the axioms (1) and (2) for a probability measure P.

Example: proof of point (1)

Proof.

- Assume $A \subseteq B$; RTP: $P(A) \leq P(B)$
  - RTP = "Required To Prove"
- We can write B as disjoint union $B = A \cup (B - A)$, where:
  - $B - A = B \cap \neg A = \{s \in S \mid s \in B \text{ and } s \notin A\}$
  - $A \cap (B - A) = \emptyset$
- By Axiom 2 we get: $P(B) = P(A) + P(B - A)$
- Since $P(B - A) \in [0,1]$, by definition, we get $P(B) \geq P(A)$.

□

Discrete sample space example

Recall that a sample space S is called discrete if it is finite

Example (One dice)

- $S = \{1,2,3,4,5,6\}$, with events $A \subseteq S$
- The probability measure $P\colon \mathcal{P}(S) \to [0,1]$ is easy:
  - $P(\{1,3,5\}) = \frac{1}{2}$
  - $P(\{1,6\}) = \frac{1}{3}$
- We see that P is determined by what it does on singleton events $\{i\} \subseteq S$
- This is typical for finite (and countable) sample spaces.

Discrete sample spaces

Let S be a discrete (ie. finite) sample space, with probability measure $P\colon \mathcal{P}(S) \to [0,1]$.

- An event $A \subseteq S$ is then also finite, say $A = \{x_1, \ldots, x_n\}$
- Hence we can write it as disjoint union of singletons:

$$A = \{x_1\} \cup \cdots \cup \{x_n\}$$

- Hence $P(A) = P(\{x_1\}) + \cdots + P(\{x_n\})$, by Axiom 2.
- Thus, P is entirely determined by its values $P(\{x\})$ on singletons, for $x \in S$.
- The function $f\colon S \to [0,1]$ with $f(x) = P(\{x\})$ is called the underlying distribution
- It satisfies $\sum_{x \in S} f(x) = 1$ since:

$$\textstyle\sum_{x \in S} f(x) = \sum_{x \in S} P(\{x\}) = P(\bigcup_{x \in S}\{x\}) = P(S) = 1$$

The uniform distribution

Fix a number $n \in \mathbb{N}$ and take as sample space $S = \{1, 2, \ldots, n\}$.

- The simplest distribution is the uniform distribution $u_n\colon S \to [0,1]$, which assigns the same probability to each $i \in S$
- Since the sum of probabilities must we 1, the only option is:

$$u_n(i) = \frac{1}{n}$$

- More generaly, on each finite set X we can define $u\colon X \to [0,1]$ as $u(x) = \frac{1}{\#X}$, where $\#X \in \mathbb{N}$ is the number of elements of X.

The binomial distribution

Fix $n \in \mathbb{N}$ with $S = \{0, 1, \dots, n\}$ and $p \in [0, 1]$.

- Define the binomial distribution $b\colon S \to [0, 1]$ as:

$$b(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Read $b(k)$ as:

  the probability of exactly k successes after n trials, each with chance p

  Briefly: $b(k) = P(k \text{ out of } n)$.
- This is well-defined distribution by binomial expansion:

  $\sum_k b(k) = \sum_k \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1$

Example binomial expansion

Suppose we have a biased coin, which comes up head with probability $p \in [0, 1]$.

Example (Toss the coin $n = 5$ times)

What is the probability of getting head k times (for $0 \leq k \leq 5$)?

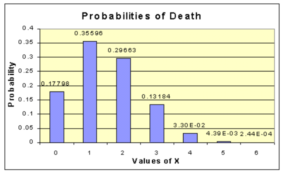
- If $k = 0$, then: $(1-p)^5$
  - via the formula: $b(0) = \binom{5}{0} p^0 (1-p)^{5-0} = (1-p)^5$
- If $k = 1$, then: $5p(1-p)^4$
  - $b(1) = \binom{5}{1} p^1 (1-p)^{5-1} = 5p(1-p)^4$
- In general: $b(k) = \binom{5}{k} p^k (1-p)^{5-k}$.

What happens if $p = \frac{1}{2}$?

Another binomial distribution example

Hospital records show that of patients suffering from a certain disease, 75% die of it. What is the probability that of 6 randomly selected patients, 4 will recover?

- We have $n = 6$, with recovery probability $p = \frac{1}{4}$.
- Hence $b(4) = \binom{6}{4} (\frac{1}{4})^4 (\frac{3}{4})^2 \simeq 0,0329595$
- Picture of all (recovery) probabilities in a histogram

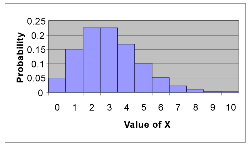


(source: intmath.com)

Other distributions

There are many other standard distributions, like:

- Normal distribution (see later in the continuous case)
- Hypergeometric distribution
- Poisson distribution
  - for independent occurrences, where some average μ is known
  - then $p(k) = e^{-\mu} \cdot \frac{\mu^k}{k!}$, for $k \in \mathbb{N}$. For instance, for $\mu = 3$,



We will not discuss these distributions here. Look up the details, later in your life, when you need them.

Conditional probability intro

Example (Suppose you throw one dice)

- Of course, the probability of 4 is $\frac{1}{6}$
- But what is the probability of 4, if you already know that the outcome is even?
- Intuitively it is clear it should be: $\frac{1}{3}$.
- We write $P(4) = \frac{1}{6}$ and $P(4 \mid \text{even}) = \frac{1}{3}$

Conditional probability is about updating probabilities in the light of given (aka. prior) information.

Conditional probability example

Assume a group of students for which:

- The probability that a student does mathematics and computer science is $\frac{1}{10}$
- The probability that a student does computer science is $\frac{3}{4}$.

Question: What is the probability that a student does mathematics, given that we know that (s)he does computer science?

Answer: We have $P(M \cap CS) = \frac{1}{10}$ and $P(CS) = \frac{3}{4}$.

We seek the conditional probability $P(M \mid CS) =$ "M, given CS"

The formula is:

$$P(M \mid CS) = \frac{P(M \cap CS)}{P(CS)} = \frac{\frac{1}{10}}{\frac{3}{4}} = \frac{4}{30} = \frac{2}{15}.$$

Basic definitions

Definition

For two events A, B, the conditional probability $P(A \mid B) =$ "the probability of A, given B", is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Alternatively, $P(A \mid B) \cdot P(B) = P(A \cap B)$.

Definition

Two events A, B are independent if $P(A \cap B) = P(A) \cdot P(B)$.

Equivalently, $P(A \mid B) = P(A)$.

Election example

Assume there are three candidates: A, B, C; only one can win

- the probability $P(A)$ that A wins is the same as for B
- $P(C)$ is half of $P(A)$.

Question 1: What are $P(A), P(B)$ and $P(C)$?

Answer 1: Solving $P(A) + P(B) + P(C) = 1$, $P(A) = P(B)$ and $P(C) = \frac{1}{2}P(A)$ yields: $P(A) = P(B) = \frac{2}{5}, P(C) = \frac{1}{5}$.

Question 2: Assume A withdraws; what are the chances of B, C now?

Answer 2: Think first what they would be intuitively!

$$P(B \mid \neg A) = \frac{P(B \cap \neg A)}{P(\neg A)} = \frac{P(B)}{1 - P(A)} = \frac{\frac{2}{5}}{\frac{3}{5}} = \frac{2}{3}$$
$$P(C \mid \neg A) = \frac{P(C \cap \neg A)}{P(\neg A)} = \frac{P(C)}{1 - P(A)} = \frac{\frac{1}{5}}{\frac{3}{5}} = \frac{1}{3}.$$

Conditional probability, for multiple events

- Recall $P(A_1 \cap A_2) = P(A_1 \mid A_2) \cdot P(A_2)$
- Hence
$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1 \mid A_2 \cap A_3) \cdot P(A_2 \cap A_3) \\ &= P(A_1 \mid A_2 \cap A_3) \cdot P(A_2 \mid A_1) \cdot P(A_1). \end{aligned}$$
- Alternatively:
$$P(A_1 \mid A_2 \cap A_3) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_2 \cap A_3)} = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_2 \mid A_1) \cdot P(A_1)}$$
- This can be generalised to $A_1, \ldots, A_n$.

Partitions

Definition

A partition of a sample space S is a collections of events $A_1, \ldots, A_n \subseteq S$ with both:

$$A_1 \cup \cdots \cup A_n = S \qquad \text{and} \qquad A_i \cap A_j = \emptyset, \text{ for } i \neq j$$

A binary partion is given by $A, \neg A$.

Partitions and the total probability lemma

Lemma (Total probability)

For a partition $A_1, \ldots, A_n$ and arbitrary event B,

$$P(B) = P(B \mid A_1) \cdot P(A_1) + \cdots + P(B \mid A_n) \cdot P(A_n).$$

Because:
$$\begin{aligned} & P(B \mid A_1) \cdot P(A_1) + \cdots + P(B \mid A_n) \cdot P(A_n) \\ &= P(B \cap A_1) + \cdots + P(B \cap A_n) \\ &= P((B \cap A_1) \cup \cdots \cup (B \cap A_n)) \\ &= P(B \cap (A_1 \cup \cdots \cup A_n)) \\ &= P(B \cap S) \\ &= P(B). \end{aligned}$$

Total probability illustration

Example (Two boxes with long & short bolts)

- In box 1, there are 60 short bolts and 40 long bolts. In box 2, there are 10 short bolts and 20 long bolts. Take a box at random, and pick a bolt. What is the probability that you chose a short bolt?
- Write B_i for the event that box i is chosen, for $i = 1, 2$
- The solution is:

$$\begin{aligned} P(short) &= P(short \mid B_1)P(B_1) + P(short \mid B_2)P(B_2) \\ &= \frac{60}{100} \cdot \frac{1}{2} + \frac{10}{30} \cdot \frac{1}{2} \\ &= \frac{3}{10} + \frac{1}{6} \\ &= \frac{7}{15}. \end{aligned}$$

Bayes' Rule/Theorem

Theorem

For events E, H we have:

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}.$$

Terminology:

- E = *evidence*, H = *hypothesis*
- $P(H)$ = *prior probability*, $P(H \mid E)$ = *posterior probability*

Proof

$$P(E \mid H) \cdot P(H) = P(E \cap H) = P(H \cap E) = P(H \mid E) \cdot P(E).$$

Bayes' Rule/Theorem for partitions

Theorem

Suppose we have a partition $H_1, \ldots, H_n$. Then:

$$P(H_i \mid E) = \frac{P(E \mid H_i) \cdot P(H_i)}{\sum_j P(E \mid H_j) \cdot P(H_j)}.$$

Proof Since $P(E) = \sum_j P(E \mid H_j) \cdot P(H_j)$ by the total probability lemma.

Machine example

Setting and question

- There are 3 machines M_1, M_2, M_3 producing items, with defect probabilities 0,01, 0,02, 0,03 respectively.
- 20% of items come from M_1, 30% from M_2, 50% from M_3
- Find the probability that a defect item comes from M_1.

Solution

- We have $P(M_1) = 0,2$, $P(M_2) = 0,3$, $P(M_3) = 0,5$ and $P(D \mid M_1) = 0,01$, $P(D \mid M_2) = 0,02$, $P(D \mid M_3) = 0,03$
- Via the total probability lemma we compute $P(D)$ as:

$$P(D \mid M_1) \cdot p(M_1) + P(D \mid M_2) \cdot p(M_2) + P(D \mid M_3) \cdot p(M_3)$$
$$= 0,01 \cdot 0,2 + 0,02 \cdot 0,3 + 0,03 \cdot 0,5 \ = 0,023$$

- Then: $P(M_1 \mid D) = \frac{P(D|M_1) \cdot P(M_1)}{P(D)} = \frac{0{,}01 \cdot 0{,}2}{0{,}023} = 0,087$

Rain and umbrella example

Setting

- Prior knowledge $P(rain) = \frac{1}{5}$
- $P(umbrella \mid rain) = \frac{7}{10}$ and $P(umbralla \mid \neg rain) = \frac{1}{10}$
- Suppose you see someone with an umbrella. What is the probability that it rains?

Answer

$$\begin{aligned} & P(rain \mid umbrella) \\ = & \frac{P(umbrella \mid rain) \cdot P(rain)}{P(umbrella \mid rain) \cdot P(rain) + P(umbrella \mid \neg rain) \cdot P(\neg rain)} \\ = & \frac{\frac{7}{10} \cdot \frac{1}{5}}{\frac{7}{10} \cdot \frac{1}{5} + \frac{1}{10} \cdot \frac{4}{5}} = \frac{\frac{7}{50}}{\frac{7}{50} + \frac{4}{50}} = \frac{7}{11} \approx 0,64. \end{aligned}$$

Inference: learning from iterated observation

- In the previous example we started from $P(rain) = \frac{1}{5}$, and computed $P(rain \mid umbrella) = \frac{7}{11}$.
- Thus after observing this umbrella we may update our prior knowledge to $P'(rain) = \frac{7}{11}$.
- What if we see another, second umbrella? Surely, the probability of rain is even higher. How to compute it?
- We can play the same game with the updated rain probability $P'(rain) = \frac{7}{10}$.

$$\begin{aligned} & P(rain \mid 2umbrellas) \\ = & \frac{P(umbrella|rain) \cdot P'(rain)}{P(umbrella|rain) \cdot P'(rain) + P(umbrella|\neg rain) \cdot P'(\neg rain)} \\ = & \frac{\frac{7}{10} \cdot \frac{7}{11}}{\frac{7}{10} \cdot \frac{7}{11} + \frac{1}{10} \cdot \frac{4}{11}} = \frac{\frac{49}{110}}{\frac{49}{110} + \frac{4}{110}} = \frac{49}{53} \approx 0,92. \end{aligned}$$

- See courses on AI (esp. Machine Learning) for more information, esp. on Bayesian networks (graphical models)!