Advanced Statistics 4 Physics Analysis

Alberto Garfagnini

Università di Padova

AA 2021/2022 - Lect. 1

Course Timetable

- 6 CFU: 4 hours/week → 12 full weeks
  - Thursday: Aula P3, 8:30 – 10:15, theory lectures
  - Thursday: Lab P104, 12:30 – 14:15, hands-on R laboratory sessions
- Course MOODLE Web Pages:
  - https://elearning.unipd.it/dfa/course/view.php?id=1400

Week		Mon	Tue	Wed	Thu		Fri
1	Feb 28 - Mar 4				Stat01	R01	
2	Mar 7 - 11				R02	R03	
3	Mar 14 - 18				R04	R05	
4	mar 21 - 25				Stat02	RLab01	
5	Mar 28 - Apr 1				Stat03	RLab02	
6	Apr 4 - 8						
7	Apr 11- 15				Stat04	RLab03	
8	Apr 18 - 22				Stat05	RLab04	
9	Apr 25 - 29				Stat06	RLab05	
10	May 2 - 6				Stat07	RLab06	
11	May 9 - 13				Stat08	RLab07	
12	May 16 - 20				Stat09	R06	
13	May 23 - 27				R07	Stat10	
14	May 30 - Jun 3				buffer	buffer	
15	Jun 6 - 10						

Course Program and Structure

➜ Part I:

- introduction to deductive logic and plausible reasoning
- review of discrete probability distributions
- review of continuous probability distributions
- sampling of random variables and Monte Carlo methods

➜ Part II:

- statistical models and inference: parameter estimation
- linear models
- model selections
- Markov Chain Monte Carlo and Gibbs sampling
- Bayesian Networks

➜ Part III (in parallel to I and II):

- R language structure and R libraries will be presented and used to solve exercises complementing the theory part

Laboratory Assignments and Exams

- during each laboratory session, a set of exercises will be assigned
- you will have to complete the assignments at home and deliver them in due time (within 2 weeks after each laboratory session)

Final Exam

- the final exam is made of two parts:
  - a written test with questions and small R exercises
  - a R computational problem will by assigned to each group of two-three students and the used techniques and the obtained results will be discussed in an oral group presentation
- the final vote will be a combination of three parts:

$$\text{Final Mark} = \frac{1}{3}\text{LaboratoryAssignments} + \frac{1}{3}\text{WrittenTest} + \frac{1}{3}\text{GroupProjectOralPresentation}$$

Statistics and Probability

What is the meaning of Statistics ?

Descriptive Statistics

- it refers to methods for summarizing and organizing the information in a data set
- it uses numbers, graphs and tables to describe data sets, as a first step of data analysis

Probability Theory

- Probability theory is the branch of mathematics concerned with probability
- as a mathematical foundation for statistics, probability theory is essential to many activities involving quantitative analysis of data
- Methods of probability theory also apply to descriptions of complex systems given only partial knowledge of their state, as in statistical mechanics

Inference

- statistical inference consists of methods for estimating and drawing conclusions about population characteristics based on the information contained in a subset (sample) of that population

➜ learning by data

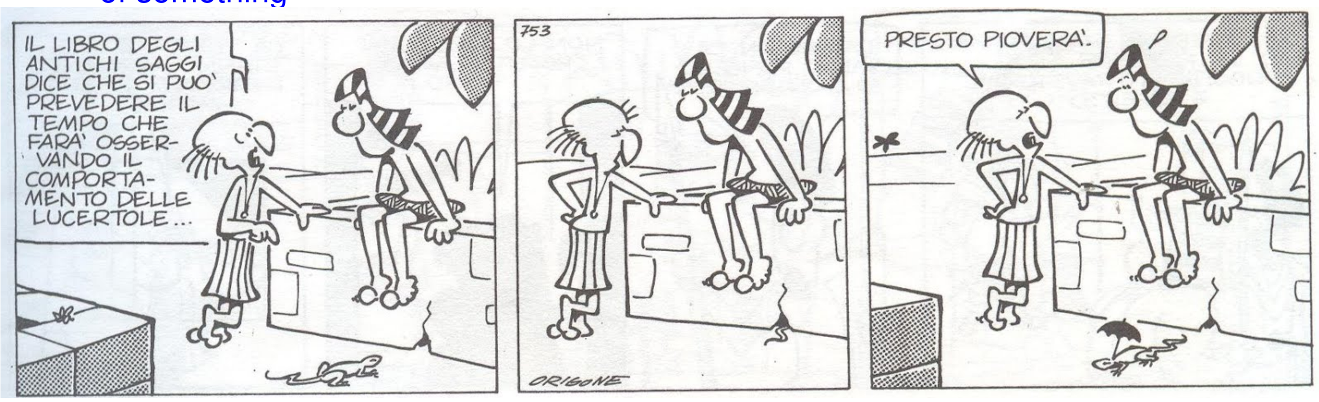
Probability and Inference

- probability theory is the doctrine of chances
- a branch of mathematics that tells us how often different kind of events will happen:
  - what are the chances of a fair coin coming up heads 10 times is a row ?
  - if two six sided dice are rolled, how likely is to get 1-1 ?
- all these probabilistic questions start with a known model of the world and we use the model to perform some calculations
- in probability theory, the model is known but data are not
- statistical questions work the other way around:
  we do not known the truth about the world, and we want to know it from the data:
  - I flip a coint 10 times and get 10 heads, is the coin fair ?
  - while drawing 5 cards from a deck I get all hearts, how likely is that the deck had been shuffled ?

➜ statistical inference questions are not the same as probability questions, but they are deeply connected to one another

Logic, Probability and Uncertainty

- most of the situations we deal with in everyday life are not completely predictable
  - Q: will it rain today, at 14:30, when I will be done with my lecture and go home for lunch ?
  - A: gather information : weather forecast, look at the sky, . . .
  - despite the forecast I could get soaked going home
- ▷ but . . . there is always uncertainty
- therefore, since we cannot eliminate uncertainty, we need to model it
- when faced with uncertainty, we use plausible reasoning
- we adjust our belief about something, based on the occurrence or non-occurrence of something



Plausible Reasoning

- the policeman, the gentleman, and the jewelry shop broken window
  [E. T. Jaynes, *Probability Theory, The Logic of Science*, Cambridge Univ. Press., 2003]
- after having seen the masked gentleman, with a bag, crawling out of a broken jewelry shop window, the policeman thinks - with no doubts - that the gentleman is a thief
- BUT . . . the policeman's conclusion is NOT logical deduction from evidence
- there may be a perfectly innocent explanations:
  - the gentleman owns the jewelry shop
  - he just came home from a masquerade party and he had left his shop keys at home
  - a truck just passed and threw a stone on the window shop, breaking it
- ➜ the gentleman was just protecting his own property
- the policeman's reasoning process was not logical deduction, but it had a certain degree of validity
- ➜ the evidence did not make the gentleman's dishonesty certain, but it did make it extremely plausible

Plausible Reasoning

- the formulation of plausible conclusions is a very subtle process
- we have examples of contrast between deductive reasoning and plausible reasoning

  weak syllogisms:

  `if A is true, then B is true`

  but, if `B is true`, therefore `A becomes more plausible`

- in other words:
- the evidence does not prove that `A is true`
- verification of one of its consequences does `give us more confidence in A`

Example

- `A` ≡ it will start to rain today by 14:30 at the latest
- `B` ≡ the sky will become cloudy before 14:30
- observing clouds at 14:15 does not give us a logical certainty that rain will follow
- nevertheless our commons sense induce us to take an umbrella, or stay inside, if the clouds are sufficiently dark

Plausible reasoning

- in spite of the apparent weakness of its argument, we recognize that the policeman's conclusion has a very strong convincing power

➜ our brain, in doing plausible reasoning, evaluates the degree of plausibility, in some way

- the plausibility of `rain at 14:30` depends very much on the `darkness of the clouds at 14:15`
- the brain makes use of old information as well as specific new data ➜ before deciding what to do, we try to recall our past experience with clouds and rain (and of last night's weather forecast on the news)
- the policeman was also making use of the past experience of policemen in general

*In our reasoning we depend very much on prior information to help us evaluating the degree of plausibility in a new problem. This reasoning process goes on unconsciously, almost instantaneously, and we conceal how complicated it really is by calling it **common sense***.

[E. T. Jaynes, *Probability Theory, The Logic of Science*, Cambridge Univ. Press., 2003]

➜ we also learn with new situations and data

Decision theory and *uncertainty*

- we are *u*ncertain whenever we do not know what to do
- we may be uncertain because we do not have enough information to know just what we should do
- sometimes we learn new information and decide that if we had this piece of information at the time, we would have acted differently
- some people say ”if the outcome was unsatisfactory, the decision was obviously wrong”. Their point is the following: ”whether a decision was RIGHT or WRONG, is to be decided entirely on the results of the decision and not on the basis of the information available at the time the decision was taken”

Example

- you get the following offer:
  - you pick up a coin from your pocket and toss it ten times
  - if you get at least one `Head` you get 200 € , if not you pay 1 €
- you accept the bet and you lose. Was it a *w*rong decision ?
- given the chance, would you make the same *w*rong decision again ?

Decision theory and Inductive logic

- of course you would bet again!
  - the odds in favor are roughly 1024 to 1
- the conclusion is that you made a good decision, but you had a bad outcome
- if a decision involves a risk, it is always possible that a good decision can lead to a bad outcome. This is the meaning of risk
- if all is known about a situation, our knowledge is deterministic
- on the other hand, if we have some uncertainties, the reasoning is inductive, that is, it requires inductive logic for finding the best solution

→ the important aspect is that we do not know everything, but we do know something

Statistical Inference

- it is of primary interest to physicists and engineers
- and it allows to learn about theory or models from experimental observations

▷ inference has to be based on probability theory

- and summaries of descriptive statistics can be used in cases in which statistical sufficiency holds, i.e. when sample arithmetic average and standard deviation are used instead of the n data points

How to Measure Plausibility

- suppose we try to use numbers to measure plausibility of propositions
- when we change our plausibility for some propositions, based on the occurrence of some other proposition, we are performing an induction

Properties of Plausibility Measures

- degrees of plausibility are represented by non-negative real numbers
- they qualitatively agree with common sense: larger values mean greater plausibility
- if a proposition can be represented in more than one way, all representations must give the same plausibility
- we must always take all the relevant evidence into account
- equivalent states of knowledge are always given the same plausibility

- a sensible way to revise plausibility is by using the rules of probability
- probability is used as a extension of logic to cases where deduction cannot be made

The adopted Language

a Proposition

- must have an unambiguous meaning
- must be of a simple logical type, i.e. `true` or `false`

- given two propositions, A, and B

$$A \cdot B$$

- is called `logical product` or conjunction and denotes the proposition: `both A and B are` **`true`**
- the order does not matter: $A \cdot B$ and $B \cdot A$ say the same thing

$$A + B$$

- is called `logical sum` or disjunction and stands for `at least one of the propositions A, B are` **`true`**
- the order does not matter: $A + B$ brings the same information of $B + A$

$$A = B$$

- means that the propostion on the left side has the same truth value as that on the right side

$$\overline{A}$$

- indicated the denial of a propostion, i.e. if `A is` **`true`**, $\overline{A}$ `is` **`false`**, and vice versa

Probability and Random Experiments

- we borrow examples from theory of games: in a random experiment (i.e. dice draw, coin tossing, ...) the outcome is uncertain

Definitions

- Random experiment : an experiment with an outcome not completely predictable. When we repeat the experiment we may get a different result (e.g. coin tossing)
- Outcome : the result of a single trial of the experiment
- Sample space : the set of all possible outcomes of one single trial of the experiment. It is usually denoted by Ω. The sample space containing everything we are considering in the analysis of the experiment is called Universe
- Event : any set of possible outcomes of the experiment

Given two events, E and F, we can construct

- Union of the events : is the set of outcomes in either E or F (inclusive or)
- Intersection of the events : is the set of the outcomes in both E and F, simultaneously
- Complement of an event : is the set of all outcomes not in E

Axiomatic Definition of Probability

- probabilities are real numbers between 0 and 1
- the higher the probability, the more likely it is to occur
- a probability equals to 1 means the event is certain to occur
- a probability of 0 means that the event cannot possibly occur

You know it all from Probability Theory

The following axioms are satisfied

1) $P(A|I) \geq 0$ for any event E
2) $P(U|I) = 1$, is the probability of the universe (it means that some outcome occurs every time the experiment is performed)
3) $P(AB|I) = P(A|B, I) \cdot P(B|I) = P(B|A, I) \cdot P(A|I)$ **(PRODUCT RULE)**

Other properties, derived from the Axioms

1) $P(\emptyset) = 0$
2) $P(\overline{A}|I) = 1 - P(A|I)$ **(NORMALIZATION)**
3) $P(A + B|I) = P(A|I) + P(B|I) - P(AB|I)$ **(SUM RULE)**
4) $P(A|I) = P(A, B|I) + P(A, \overline{B}|I)$ **(MARGINALIZATION)**

All Probabilities are Conditional

- whenever we talk about probability, we always use the symbol

$$P(A|I)$$

- and never $P(A)$
- I is the background condition, related to information we have
- because it makes no sense to talk about the probability of the truth of the statement A without being explicit about the conditions upon which the assignment of probability is based
- we are engaged in a chain of inductive logic and at each point where an answer is required, we shall report the best inference we can make based upon the evidence available to that point
- as new evidence becomes available, we shall use the same procedure to update our inferences
- it is therefore an iterative process until evidence is so overwhelming that it does not seem worthwhile to pursue the matter any further

A note on the English Language

▷ Ordinary language is not precise

one of the most ambiguous word in the language is OR

- As an example, the statement: *"This is a ceramic or a conductor"* may be represented at least in two ways.
  Let's define:
  - proposition A: "it is a ceramic"
  - proposition B: "it is a conductor"

inclusive

exclusive

- the sentence can be represented by $(A+B)$ or by $\left(A\overline{B}+\overline{A}B\right)$
- because it is not clear if the word *or* is used in the inclusive or exclusive sense
  - if we say $A+B$ and we mean $(A+B)$, the word or is used in the inclusive sense
  - instead if $A+B$ is meant by $\left(A\overline{B}+\overline{A}B\right)$, the word or is used in the exclusive sense
- similarly the equal sign does not mean that two statements mean the same thing, but only that they have the same truth table

Independent versus Mutually Exclusive Events

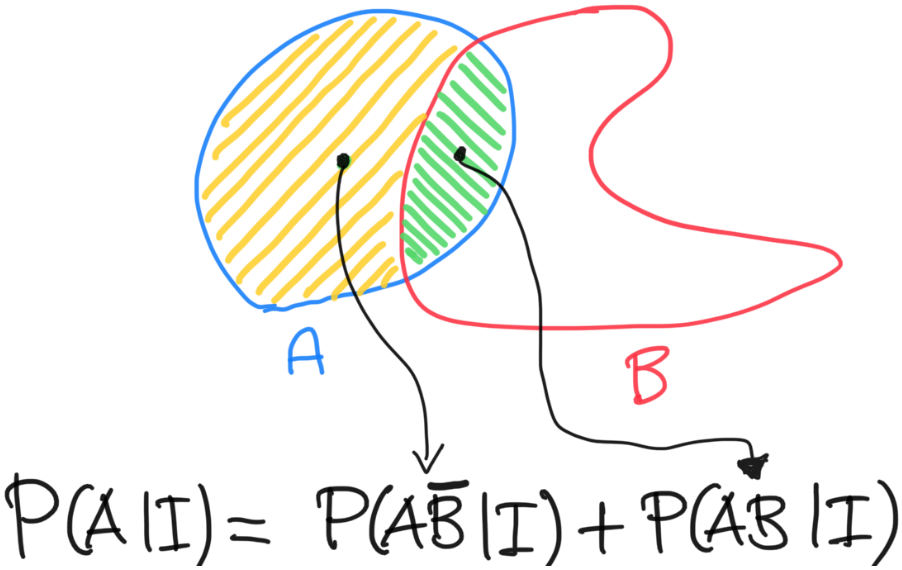
- a semantic confusion may arise because the word independent has several meanings
- primary meaning : two events are independent if the occurrence of the first does not affect the occurrence or non-occurrence of the second
- For independent events, $P(A|B,I)=P(A|I)$, therefore $P(AB|I)=P(A|I)P(B|I)$
- independence of two events is not a property of the events themselves, but it comes from the probabilities of the events and their intersection
  - if A and B are mutually exclusive events, (i.e. they have no outcome in common), $P(A+B|I)=P(A|I)+P(B|I)$
  - mutually exclusive events contain no elements in common, and this is a property of the events

Laplace's "Bayes Theorem"

Marginal Probability

- The marginal probability of one event is found by summing its disjoint parts:

$$P(A|I) = P(AB|I) + P(A\overline{B}|I)$$



Laplace's "Bayes Theorem"

Marginal Probability

- The marginal probability of one event is found by summing its disjoint parts:

$$P(A|I) = P(AB|I) + P(A\overline{B}|I)$$

Bayes' Theorem : 1

- from the definition of the probability of the intersection of two events

$$P(AB|I) = P(A|B,I)P(B|I) = P(B|A,I)P(A|I)$$

- we can re-write one of the two equalities as:

$$P(B|A,I) = \frac{P(AB|I)}{P(A|I)}$$

- and making use of the marginalized expression for $P(A|I)$

$$P(B|A,I) = \frac{P(AB|I)}{P(AB|I) + P(A\overline{B}|I)}$$

Laplace's "Bayes Theorem" : 2

- we now use the product rule to find each of the joint probabilities

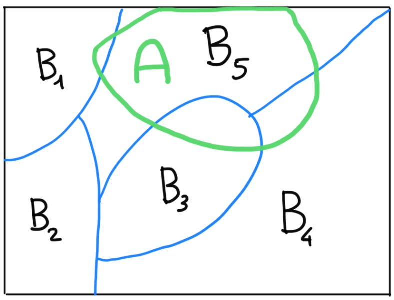
$$P(B|A,I) = \frac{P(A|B,I)P(B|I)}{P(A|B,I)P(B|I) + P(A|\overline{B},I)P(\overline{B}|I)}$$

- Bayes' theorem is a restatement of the original product rule in terms of the $P(B|A)$ probability, where
  1) the probability of A is found as the sum of the probabilities of its disjoints parts AB and $A\overline{B}$
  2) each of the joint probabilities are found using the multiplication rule
- it is important to notice that:
  1) the union of B and $\overline{B}$ is the whole universe
  2) and they are disjoint

Partition and Marginalization

- let's imagine that we have a set of more than two events that partition the universe:

$$U = B_1 \cup B_2 \cup \ldots \cup B_n$$



- and every distinct pair of the events are disjoint

$$B_j \cap B_i = \emptyset \text{ , with } j \neq i$$

- an observable event A will be partitioned into parts

$$A = \bigcup_{j=1}^{n} (A \cap B_j)$$

- and therefore

$$P(A|I) = \sum_{j=1}^{n} P(AB_j|I) = \sum_{j=1}^{n} P(A|B_j,I)P(B_j|I)$$

- where we have used the product rule on each joint probability

Laplace's "Bayes Theorem" : 3

- finally, the conditional probability $P(B_j|A, I)$ is found by dividing each joint probability by the probability of the event A, partitioned over all B_j

$$P(B_j|A, I) = \frac{P(A|B_j, I)P(B_j|I)}{\sum_{j=1}^{n} P(A|B_j, I)P(B_j|I)}$$

  - this is known as Bayes' theorem
  - it was published posthumously in 1763 after the death of its discoverer, Reverend Thomas Bayes
  - Pierre-Simon Laplace reproduced and extended Bayes's results in 1774, apparently unaware of Bayes's work
  - The Bayesian interpretation of probability was developed mainly by Laplace

Note that

- the events A and B_j, $(j = 1, \ldots, n)$, are not treated symmetrically :
  - A is an observable event
  - B_j are considered not observable. We never know which one of them occurred
- The marginal probabilities $P(B_j)$ are assumed known before we start and are called our prior probabilities

Interpretation of the Laplace's "Bayes Theorem"

- we often write Bayes' in its proportional form as

$$\underbrace{P(B_j|A, I)}_{\text{posterior}} \propto \underbrace{P(A|B_j, I)}_{\text{likelihood}} \cdot \underbrace{P(B_j|I)}_{\text{prior}}$$

- this gives the relative weights for each of the events B_j, after we know A has occurred

- dividing by the sum of the relative weights it re-scales the relative weights, they now sum to 1 $\rightarrow$ making it a probability distribution

$$P(B_j|A, I) = \frac{P(A|B_j, I)P(B_j|I)}{\sum_{j=1}^{n} P(A|B_j, I)P(B_j|I)}$$

Exercise 1

- a new medical diagnostic test for a specific disease comes to market
- the sensitivity of the test, i.e. the probability the test gives a positive response if a person has the disease is 0.95
- the specificity, i.e. the probability that the test gives a negative response if a person does not have the disease, is 0.90

↘ this allows to evaluate the false positive rate to 0.10 (i.e. the probability that the test gives a positive response if a person does not have the disease)

- we know that 1% of the population, in Italy, has the disease

➜ What is the probability that a person has the disease, given the fact that the test gives a positive response ?

- we define the following propositions:
  $D :=$ a person has a disease
  $T :=$ the test gives a positive result

- and the data tell us
  $P(T|D, I) = 0.95$
  $P(\overline{T}|\overline{D}, I) = 0.90$
  $P(D|I) = 0.01$

Contingency Tables

- it's a method of keeping track of the various probabilities, and seeing how to compute one from another
- the following example shows a 2×2 table of probabilities involving two statements A and B :

	B	$\overline{B}$	
A	$P(AB|I) = \omega_1$	$P(A\overline{B}|I) = \omega_2$	$P(A|I) = \omega_1 + \omega_2$
$\overline{A}$	$P(\overline{A}B|I) = \omega_3$	$P(\overline{A}\overline{B}|I) = \omega_4$	$P(\overline{A}|I) = \omega_3 + \omega_4$
	$P(B|I) = \omega_1 + \omega_3$	$P(\overline{B}|I) = \omega_2 + \omega_4$	1

Exercise 1, Solution

- the initial data give: $P(T|D,I) = 0.95$, $P(\overline{T}|\overline{D},I) = 0.90$ and $P(D|I) = 0.01$

- we build a contingency table with the propositions:
  $D :=$ a person has a disease
  $T :=$ the test gives a positive result

	T	$\overline{T}$	
D	0.0095	0.0005	0.01
$\overline{D}$	0.099	0.891	0.99
	0.1085	0.8915	1

- we deduce: $P(\overline{D}|I) = 1 - P(D|I) = 0.99$ and $P(T|\overline{D},I) = 1 - P(T|D,I) = 0.05$
  - $P(DT|I) = P(T|D,I) \cdot P(D|I) = 0.95 \times 0.01 = 0.0095$
  - $P(D\overline{T}|I) = P(D|I) - P(DT|I) = 0.01 - 0.0095 = 0.0005$
  - $P(\overline{D}T|I) = P(T|\overline{D},I) \cdot P(\overline{D}|I) = P(T|\overline{D},I)(1 - P(D|I) = 0.1 \times 0.99 = 0.099$
  - $P(\overline{DT}|I) = P(\overline{D}|I) - P(\overline{D}T|I) = 0.99 - 0.099 = 0.891$
  - finally,
    $P(T|I) = P(DT|I) + P(\overline{D}T|I) = 0.1085$ and $P(\overline{T}|I) = P(D\overline{T}|I) + P(\overline{D}\,\overline{T}|I) = 0.8915$

Exercise 1, Solution

- the initial data give: $P(T|D,I) = 0.95$, $P(\overline{T}|\overline{D},I) = 0.90$ and $P(D|I) = 0.01$

- we build a contingency table with the propositions:
  $D :=$ a person has a disease
  $T :=$ the test gives a positive result

	T	$\overline{T}$	
D	0.0095	0.0005	0.01
$\overline{D}$	0.099	0.891	0.99
	0.1085	0.8915	1

- let's apply Bayes' theorem:

$$P(D|T,I) = \frac{P(T|D,I)P(D|I)}{P(T|I)}$$

- the normalization factor is

$$\begin{aligned} P(T|I) &= P(T|D,I)P(D|I) + P(T|\overline{D},I)P(\overline{D}|I) \\ &= 0.95 \times 0.01 + 0.10 \times 0.99 = 0.1085 \end{aligned}$$

- and the final answer is:

$$P(D|T,I) = \frac{0.0095}{0.1085} = 0.0876$$

Exercise 2

- a new medical screening procedure for a specific cancer is introduced
- the screening has:
  sensitivity = 0.90, and specificity = 0.95
- suppose the underlying rate of the cancer in the population is 0.001

➜ What is the probability that a person has the disease given the results of the screening is positive ?

➜ Does this show that screening is effective in detecting this cancer ?

- we define the following propositions:

  $B :=$ a person has a specific cancer

  $A :=$ the screening has positive result

- and the data tell us

  $P(A|B, I) = 0.90$

  $P(\overline{A}|\overline{B}, I) = 0.95$

  $P(B|I) = 0.001$

Exercise 2, Solution

- the initial data give: $P(A|B, I) = 0.90$, $P(\overline{A}|\overline{B}, I) = 0.95$ and $P(B|I) = 1 \cdot 10^{-3}$

$B :=$ a person has a specific cancer

$A :=$ the screening has positive result

	A	$\overline{A}$	
B	$9 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$10 \cdot 10^{-4}$
$\overline{B}$	0.050	0.949	0.999
	0.051	0.949	1

Calculations

- given sensitivity and specificity, $P(A|\overline{B}, I) = 1 - 0.95 = 0.05$
- $P(AB|I) = P(A|B, I)P(B|I) = 0.90 \times 0.001 = 9 \cdot 10^{-4}$
- $P(A\overline{B}|I) = P(A|\overline{B}, I)P(\overline{B}|I) = 0.05 \times 0.999 = 0.050$

$$P(A|I) = P(AB|I)P(B|I) + P(A\overline{B}|I)P(\overline{B}|I)$$
$$= 9 \cdot 10^{-4} \times 10 \cdot 10^{-4} + 0.04995 \times 0.999 = 0.051$$

- Bayes' theorem tells us that

$$P(B|A, I) = \frac{P(A|B, I)P(B|I)}{P(A|I)} = \frac{9 \cdot 10^{-4}}{0.05} = 0.018$$

Odds Ratios

- another way of dealing with uncertain events we are modeling is to form the odd ratio of the event:

$$odds(A|I) = \frac{P(A|I)}{P(\overline{A}|I)}$$

- since $P(\overline{A}|I) = 1 - P(A|I)$,

$$odds(A|I) = \frac{P(A|I)}{1 - P(A|I)}$$

- if we are using prior probabilities, we get the prior odds ratio
- if, instead, if we are using posterior probabilities, we get the posterior odds ratio

- solving the equation for the probability of the event A, we get

$$P(A|I) = \frac{odds(A|I)}{1 + odds(A|I)}$$

Assigning probabilities: 1

- so far we have derived consistent rules that can be used to manipulate probabilities, associated to real numbers
- but nothing tell us what actual numerical value we should assign at the beginning of a problem

- Ssuppose we have n propositions A_j, $j = 1, \ldots, n$, and at least one of them is `true`
- In addition, let's suppose that they are mutually exclusive, i.e.

$$P(A_i A_j|B) = P(A_i|B)\delta_{ij}$$

- therefore

$$P(A_1 + A_2 + \ldots + A_m|B) = \sum_i^m P(A_i|B), \text{ with} 1 \leq m \leq n$$

- we add the hypothesis that the propositions are also exhaustive, the background information B stipulates that one and only one of them must be true. Therefore:

$$\sum_{i=1}^n P(A_i|B) = 1$$

Assigning probabilities: 2

Problem I

- we have a set of mutually exclusive and exhaustive propositions: $\{A_1, \ldots, A_n\}$ and we seek to evaluate $P(A_i|B)_I$, with $i = 1, 2, \ldots, n$
- note that the labels are arbitrary: it makes no difference which proposition is called A_1 and which A_2, and so on

Problem II

- we have a set of mutually exclusive and exhaustive propositions: $\{A'_1, \ldots, A'_n\}$, given by
$$A'_1 \equiv A_2, \quad A'_2 \equiv A_1 \text{ and } A'_k = A_k, 3 \leq k \leq n$$
- and we seek to evaluate $P(A_i|B)_{II}$, with $i = 1, 2, \ldots, n$
- since the background information B is the same for both problems, we must have
$$P(A_1|B)_I \equiv P(A'_2|B)_{II} \text{ and } P(A_2|B)_I \equiv P(A'_1|B)_{II}$$
- problem I and II are not only related, but entirely equivalent, therefore
$$P(A_i|B)_I \equiv P(A'_i|B)_{II} \; \forall i = 1, 2, \ldots, n$$

Assigning probabilities: 3

- since equivalent state of knowledge must represent equivalent plausibility assignments
- propositions A_1 and A_2 must be assigned equal probabilities in Problem I and II
- the argument just given is the 'baby' version of the group invariance principle
- more generally, $\{A_1", \ldots, A_n"\}$ may be any permutation of $\{A_1, \ldots, A_n\}$
- if the background information B is indifferent among all propositions. we are exactly in the same state of knowledge
- we obtain n equation of the form $P(A_i|B)_I = P(A_k|B)_I$ and the relations must hold whatever the particular permutation we used in our problem
- we come to the conclusion that all the $P(A_i|B)$ must be equal
- using the initial assumption that the $\{A_1, \ldots, A_n\}$ are exhaustive, i.e. $\sum_{i=1}^{m} P(A_i|B) = 1$, the only possibility is
$$P(A_j|B) = \frac{1}{n} \text{ with } 1 \leq j \leq n$$

> following Keynes (1921), we shall call this result the principle of indifference

Boolean Algebra

- Given two propositions **A** and **B**, we define two binary operations:

> **A**$\cdot$ **B** as the logical product or conjunction (also called `AND`). This will be `true`, if 'both A and B are `true`'

> **A**$+$ **B** as the logical sum or disjunction; it stands for 'at least one of the propositions A, B is `true`'

- and one unary operation:

> $\overline{A}$ indicates the denial of a proposition

> and the relation between A and $\overline{A}$ is reciprocal: $\overline{\overline{A}} = A$

A	B	NOT $\overline{A}$	NOT $\overline{B}$	AND $A \wedge B$	OR $A \vee B$
T	T	F	F	T	T
T	F	F	T	F	T
F	T	T	F	F	T
F	F	T	T	F	F

Boolean Algebra: basic identities

$A \cdot A = A$	$A + A = A$	idempotence
$A \cdot B = B \cdot A$	$A + B = B + A$	commutativity
$A\,(B \cdot C) = (A \cdot B)\,C$	$A + (B + C) = (A + B) + C$	associativity
$A\,(B + C) = A \cdot B + A \cdot C$	$A + B \cdot C = (A + B)\,(A + C)$	distributivity
$C = A \cdot B \rightarrow \overline{C} = \overline{A} + \overline{B}$	$D = A + B \rightarrow \overline{D} = \overline{A} \cdot \overline{B}$	De Morgan's
$A \Longrightarrow B$	to be read as **A implies B**	implication

A implies B does not assert that either A or B is `true`, but it means that $A\overline{B}$ is `false` or that $\left(\overline{A} + B\right)$ is `true`.
That is if A is `true`, then B must be `true`; or if B is `false`, then A must be `false`.

Boolean Algebra: adequate set

- so far we have defined four operations by which, starting from two propositions, A and B, other propositions may be defined
  - the logical product: $A \cdot B$ (also written AB)
  - the logical sum: $A + B$
  - the negation: $\overline{A}$
  - the implication: $A \Longrightarrow B$

NOTE : in ordinary language, 'A implies B' ($A \Longrightarrow B$) means that B is logically deducible from A. In formal logic, 'A implies B' means only that the propositions A and B have the same truth value

- it is possible to demonstrate that the three basic operations of Boolean Algebra, AND, OR and NOT constitute and 'adequate set', i.e. they suffice to generate all possible logic functions

Example

- Boolean algebra rules can be used, for instance in enumerating all possible cases of an event
- let's define the following propositions:
  - ▷ E_1 : a coin is tossed once
  - ▷ E_2 : a coin is tossed for a second time
  - ▷ H_1 : the coin lands 'tail' on the first toss
  - ▷ H_2 : the coin lands 'tail' on the second toss
- we can write: $E_1 = H_1 + \overline{H_1}$ and $E_2 = H_2 + \overline{H_2}$ using the multiplication rule:

$$\begin{aligned} E_1 E_2 &= \left(H_1 + \overline{H_1}\right)\left(H_2 + \overline{H_2}\right) \\ &= H_1 H_2 + H_1 \overline{H_2} + \overline{H_1} H_2 + \overline{H_1}\,\overline{H_2} \end{aligned}$$

Boolean Algebra: alternative adequate set

- it is possible to demonstrate that there are operations which, alone, would constitute ad adequate set: these are NAND and NOR

> NAND is the negation of AND:

$$A \uparrow B \equiv \overline{AB} = \overline{A} + \overline{B}$$

> and it can generate the other operations:

- $\overline{A} = A \uparrow A$
- $AB = (A \uparrow B) \uparrow (A \uparrow B)$
- $A + B = (A \uparrow A) \uparrow (B \uparrow B)$

> NOR is the negation of OR:

$$A \downarrow B \equiv \overline{A + B} = \overline{A}\,\overline{B}$$

> and it can generate the other operations:

- $\overline{A} = A \downarrow A$
- $A + B = (A \downarrow B) \downarrow (A \downarrow B)$
- $AB = (A \downarrow A) \downarrow (B \downarrow B)$

NAND formulas demonstration

NAND

- $\overline{A} = A \uparrow A$ $\square$

since $\overline{A} = \overline{A \cdot A}$ using idempotence

- $AB = \overline{\overline{AB}}$ since $\overline{\overline{A}} = A$

$= \overline{\overline{A \cdot B} \cdot \overline{A \cdot B}}$ using idempotence

$= (A \uparrow B) \uparrow (A \uparrow B)$ $\square$

- $A + B = \overline{\overline{A + B}} = \overline{\overline{A} \cdot \overline{B}}$ using De Morgan's

$= \overline{\overline{A \cdot A} \cdot \overline{B \cdot B}}$ using idempotence

$= (A \uparrow A) \uparrow (B \uparrow B)$ $\square$

The Extended Sum Rule

$$
\begin{aligned}
P(A+B|C) &= 1 - P\left(\overline{A+B}|C\right) = 1 - P\left(\overline{A}\cdot\overline{B}|C\right) \\
&= 1 - P\left(\overline{A}|C\right)\cdot P\left(\overline{B}|\overline{A}C\right) = 1 - P\left(\overline{A}|C\right)\left[1 - P\left(B|\overline{A}C\right)\right] \\
&= 1 - P\left(\overline{A}|C\right) + P\left(\overline{A}|C\right)P\left(B|\overline{A}C\right) \\
&= P(A|C) + P\left(\overline{A}B|C\right) = P(A|C) + P(B|C)\,P\left(\overline{A}|BC\right) \\
&= P(A|C) + P(B|C)\left[1 - P(A|BC)\right] \\
&= P(A|C) + P(B|C) - P(B|C)\,P(A|BC) \\
&= P(A|C) + P(B|C) - P(AB|C)
\end{aligned}
$$

Summary of our mathematical tools

PRODUCT RULE	$P(AB|C) = P(A|BC)\,P(B|C) = P(B|AC)\,P(A|C)$
SUM RULE	$P(A|B) + P(\overline{A}|B) = 1$
EXTENDED SUM RULE	$P(A+B|C) = P(A|C) + P(B|C) - P(AB|C)$
BAYES' THEOREM	$P(B_j|A,I) = \dfrac{P(A|B_j,I)P(B_j|I)}{\sum_{j=1}^{n} P(A|B_j,I)P(B_j|I)}$
PRINCIPLE of INDIFFERENCE	Given $H_1,\ldots,H_n$, mutually exclusive and exhaustive. If B does not favor any of them: $P(H_j|B) = 1/n$

Some useful transformations

- given $P(A|B$, three probability transformations are particularly useful and will be employed later in our calculations:

- Odds :

$$O(A|E) = P(A|E)/P(\overline{A}|E)$$

- Evidence :

$$ev(A|E) = K \ln (O(A|E)) = K(\ln (P(A|E)) - \ln (P(\overline{A}|E)))$$

- Surprisal :

$$u(A|E) = -K \ln (P(A|E))$$

Useful formulas

definition	range	denial rule	joint rule
Probability $p(A\|C)$	[0,1]	$P(\overline{A}\|C) = 1 - P(A\|C)$	$P(AB\|C) = P(A\|BC)P(B\|C)$
Odds $O(A\|C) = \frac{P(A\|C)}{P(\overline{A}\|C)}$	$[0, \infty]$	$O(\overline{A}\|C) = \frac{1}{O(\overline{A}\|C)}$	$O(AB\|C) = \frac{O(A\|BC)O(B\|C)}{1+O(A\|BC)+O(B\|C)}$
Surprisal $u(A\|C) = -k \ln P(A\|C)$	$[\infty, 0]$	$u(\overline{A}\|C) = -k \ln (1 - \exp \frac{-u(A\|C)}{k})$	$u(AB\|C) = u(A\|BC) + u(B\|C)$
Evidence $ev(A\|C) = k \ln \frac{P(A\|C)}{P(\overline{A}\|C)}$	$[-\infty, \infty]$	$ev(\overline{A}\|C) = -ev(A\|C)$	*

$$^{*}\ ev(AB|C) = ev(A|BC) + ev(B|C) - k \ln \left[1 + \exp \frac{ev(A|BC)}{k} + \exp \frac{ev(B|C)}{k}\right]$$

References

Books

- Tribus M., *Rational Descriptions, Decisions and Designs*, Pergamon Press, 1969
- Bolstad W. M., *Introduction to Bayesian Statistics*, John Wiley & Sons, 2007, 978-0-470-14115-1
- D'Agostini G., *Bayesian reasoning in data analysis - A critical introduction*, World Scientific Publishing, 2003, ISBN 978-981444795-9
- Jaynes E. T., *Probability Theory, The Logic of Science*, Cambridge University Press, 2003, ISBN 978-0-521-59271-0