01a - Probability - 01

Master in Foundations of Data Science
Bayesian Statistics and Probabilistic Programming
Fall 2018-2019

Josep Fortiana

Monday, September 17, 2018

Universitat de Barcelona

01a - Probability 01

01a - Probability 01

A mugging

A woman is strolling about an expensive shopping zone in a big city, Paseo de Gracia in Barcelona, say. There she is the victim of a mugging, her handbag is stolen from her. Unfortunately, the handbag contains a hefty quantity of money in cash.

Both she and another witness describe the mugger as a very tall man, more than 2 metres, age between 20 and 30 years old, red-haired, and walking with a visible limp.

Police acts

A little later, a police patrol spots a man, whose description coincides with that of the culprit, carrying a quite large HDTV.

He is arrested.

He denies any wrongdoing, but fails to exhibit an alibi, stating that around the time when the crime was commited he was home, all by himself.

Further research shows he paid cash for the TV set.

Prosecutor's statement

Before the judge, the prosecutor explains the, notwithstanding the absence of forensic evidence, such as fingerprints, etc., circumstantial evidence is substantial enough to convict the indicted man.

A criminologist is called to declare as an expert witness.

He displays statistics extracted from demographic tables.

Criminologist's figures

He explains that, in Barcelona, the probability of:

$$\begin{cases} \text{Being male} & \text{is} \quad 0.51, \\ \text{Being taller than 2 metres} & \text{is} \quad 0.025, \\ \text{Being between 20 and 30 years old} & \text{is} \quad 0.25, \\ \text{Being red-haired} & \text{is} \quad 0.037, \\ \text{Walking with a visibly uneven step} & \text{is} \quad 0.017. \end{cases}$$

How likely?

Since all these are independent features, the probability that a given random person in Barcelona has all of them is the product:

$$
\begin{array}{rl}
 & 0.51 \\
\times & 0.025 \\
\times & 0.25 \\
\times & 0.037 \\
\times & 0.017 \\
\hline
= & 0.000002
\end{array}
$$

Prosecutor's closing arguments

The prosecutor explains that, had the crime been commited by a perpetrator with no special characteristics, identification would be like looking for a needle in a haystack, prompting the Police to go after the "usual suspects".

Now, since in our case the defendant meets this unique conjunction of unusual features, we can safely convict him.

This reasoning is wrong. An instance of the well-known *Prosecutor's Fallacy.*

Notations

To understand the origin of the mistake, let us set some notations:

$$\begin{aligned} A = &\quad \text{The defendant meets the given description,} \\ C = &\quad \text{The defendant is guilty.} \end{aligned}$$

The fallacy arises from believing that since A is a highly rare event, $\mathrm{P}(A) = 2 \times 10^{-6}$, then the culprit's identification is very conclusive, which is not.

Mistaking probability for conditional probability

What we are interested in is evaluating the conditional probability:

$$\mathrm{P}(C|A).$$

Meaning: What is the probability that the defendant is guilty given he matches the description?

To fix ideas, asume that $n = 10$ people, from a population of $N = 5 \times 10^6$, have the above mentioned set of features, so that:

$$\mathrm{P}(A) = \frac{n}{N} = 2.0 \times 10^{-6}.$$

Mistaking probability for conditional probability

On the other hand, the probability that randomly selecting an individual from the population we select the culprit is:

$$\mathrm{P}(C) = \frac{1}{N} = 2.0 \times 10^{-7}.$$

The observed evidence is that the culprit does indeed have the given set of features, that is:

$$\mathrm{P}(A|C) = 1.$$

Mistaking probability for conditional probability

Finally, the *conditional probability* that an individual having all these features is indeed the culprit is:

$$\mathrm{P}(C|A) = \frac{\mathrm{P}(A|C)\,\mathrm{P}(C)}{\mathrm{P}(A)} = \frac{1 \times \dfrac{1}{N}}{\dfrac{n}{N}} = \frac{1}{n} = \frac{1}{10}.$$

Clearly, it is far from "very likely" that the defendant is the culprit.

Other fallacies

There are many analogous apparently intuitive misaprehensions, based on mistaking a probability for a conditional probability, or the other way around, for instance the so-called *Defence attorney's fallacy*.

More examples, with data on DNA identification in real trials, can be found in:

Fung, W. K. (2001) *Teaching Statistics Using Forensic Examples,* International Statistical Institute, Invited talk, 53rd Session ISI 2001.

01a - Probability 01

Historical roots

- Blaise Pascal, Pierre Fermat (1654)
- Thomas Bayes (1764)
- Pierre-Simon de Laplace (1820)

Kolmogorov (1933) consistency axioms

Ω, a set of possible results of an experiment,
$\mathscr{E}$, a class of subsets of Ω, *(events, observables)*.

A *probability* $(\Omega, \mathscr{E})$ is a map $\mathrm{P} : \mathscr{E} \to [0, 1]$ such that:

1. $\mathrm{P}(\Omega) = 1$.
2. If $A \cap B = \emptyset$, then $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$.

Frequentist vs Bayesian

Interpretation of probability:

Frequentist: Relative frequency in a potentially infinite sequence of repetitions.

Bayesian: A measure of certainty/uncertainty. A degree of belief.

Subjective probability

A number of axiomatizations since Keynes (1921).

- Bruno de Finetti (1906–1985)
- Edwin T. Jaynes (1922–1998)
- Richard T. Cox (1889–1991)

What about statistics?

Classical: (X, Θ), constants, variables, parameters.

Bayesian: All quantities represented as random variables.

01a - Probability 01

Conditional probability (elementary definition)

Elementary definition of conditional probability:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)},$$

where A, B are events in a probabilty space such that $\mathrm{P}(B) > 0$.

This definition can be interpreted as a measure of how likely is the event A, assuming we know that some antecedent circumstance B has happened.

Interpreting conditional probability

Either B facilitates or hampers the occurrence of A.

$\mathrm{P}(A|B) > \mathrm{P}(A), \implies$ B facilitates the occurrence of A.

$\mathrm{P}(A|B) = \mathrm{P}(A), \implies$ A and B are independent events.

$\mathrm{P}(A|B) < \mathrm{P}(A), \implies$ B hampers the occurrence of A.

Example 1

In a regular dice each of the six possible results has probability $\frac{1}{6}$.

Now we evaluate the probabilites conditional to the event:

$$A = \{\text{the result is even}\} = \{2, 4, 6\},$$

Example 1

$$\mathrm{P}(\{1\}|A) = \frac{\mathrm{P}(\{1\} \cap A)}{\mathrm{P}(A)} = \frac{\mathrm{P}(\emptyset)}{\mathrm{P}(A)} = 0,$$

$$\mathrm{P}(\{2\}|A) = \frac{\mathrm{P}(\{2\} \cap A)}{\mathrm{P}(A)} = \frac{\mathrm{P}(\{2\})}{\mathrm{P}(A)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Example 1

Similarly:

$$\mathrm{P}(\{1\}|A) = \mathrm{P}(\{3\}|A) = \mathrm{P}(\{5\}|A) = 0,$$

$$\mathrm{P}(\{2\}|A) = \mathrm{P}(\{4\}|A) = \mathrm{P}(\{6\}|A) = \frac{1}{3}.$$

From conditioning (both sides) to Bayes' rule

If both $\mathrm{P}(A) > 0$ and $\mathrm{P}(B) > 0$ we can compute both conditional probabilities:

$$\mathrm{P}(A \cap B) = \mathrm{P}(B|A)\,\mathrm{P}(A) = \mathrm{P}(A|B)\,\mathrm{P}(B).$$

This equality is the source of Bayes' rule.

01a - Probability 01

A tale of two probabilities

Bayesian probability

Conditional probability

Independent events

Diagram 1

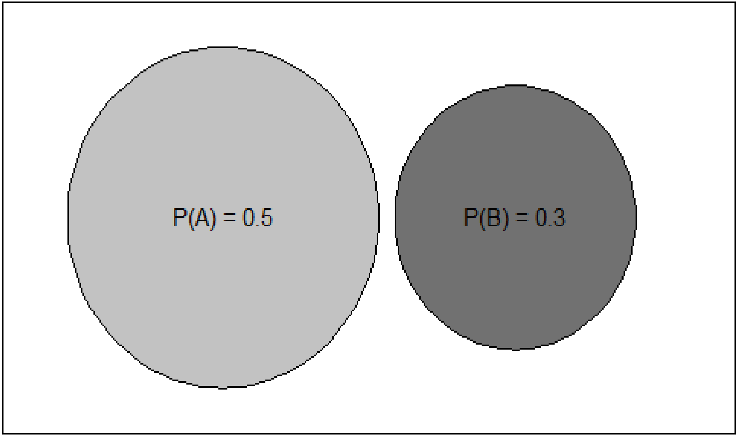


Diagram 2

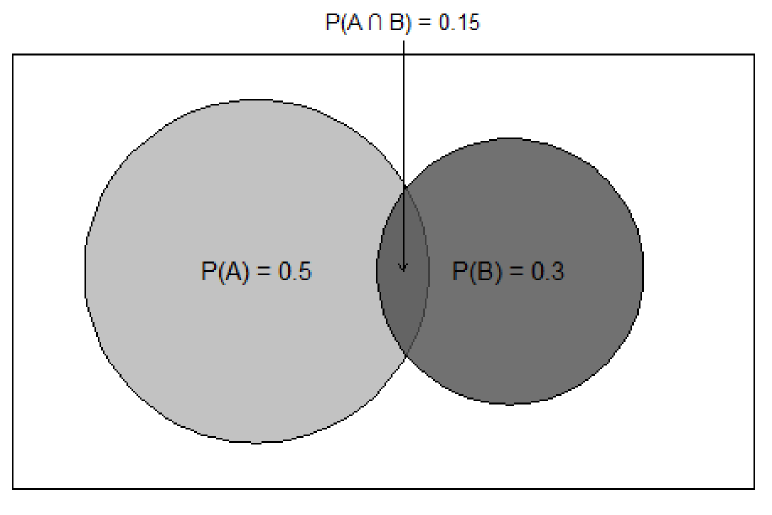
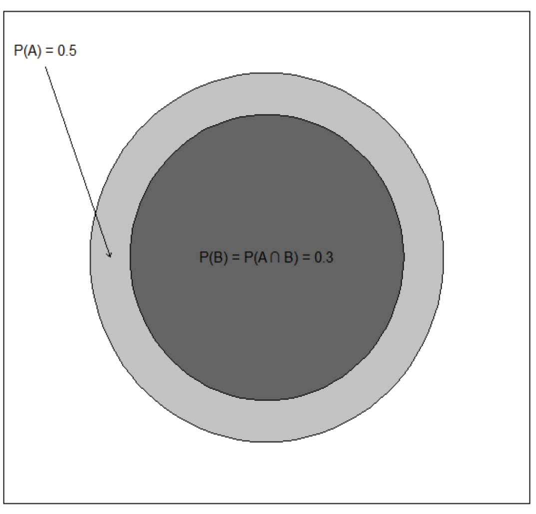


Diagram 3



Independent events: definition

Two events A and B are *independent* if

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B).$$

Notation: $A \perp\!\!\!\perp B$.

When $\mathrm{P}(A) > 0$ this is equivalent to:

$$\mathrm{P}(B|A) = \mathrm{P}(B).$$

Example 2

A Psychophysiology experiment, measuring responses to 4 sound stimuli: (S_1, S_2, S_3, S_4), which are emitted with equal probability $1/4$.

There are 4 possible responses: (R_1, R_2, R_3, R_4). In each case the subject's choice is registered.

Results are tabulated as a 4×4 table where each i-th row contains the 4 probabilites of the 4 responses to the i-th stimulus $(1 \leq i \leq 4)$.

Example 2 (continued)

	R_1	R_2	R_3	R_4
S_1	$3/4$	$1/4$	0	0
S_2	$1/5$	$3/5$	$1/5$	0
S_3	0	0	$4/5$	$1/5$
S_4	0	0	$1/6$	$5/6$

Example 2 (continued)

Questions:

1. Check that stimuli and responses are not independent.
2. A subject issues the R_2 response. What is the probability that stímulus S_2 was emitted.

Example 2 - Data description

Firstly, note that values on the table are conditional probabilities.

Namely, the j-th entry in the i-th row is:

$$\mathrm{P}(R_j|S_i), \qquad\qquad (1 \leq j \leq 4).$$

Each row adds up to 1.

Example 2 - Probabilities of intersections

We can compute the corresponding intersection probabilities:

$$\mathrm{P}(R_j, S_i) \equiv \mathrm{P}(R_j \cap S_i) = \mathrm{P}(R_j | S_i) \cdot \mathrm{P}(S_i), \quad 1 \leq j \leq 4,$$

multiplying each row i by $\mathrm{P}(S_i) = 1/4$.

The result is the *table of joint probabilities.*

Example 2 - Table of joint probabilities

	R_1	R_2	R_3	R_4	
S_1	3/16	1/16	0	0	1/4
S_2	1/20	3/20	1/20	0	1/4
S_3	0	0	4/20	1/20	1/4
S_4	0	0	1/24	5/24	1/4
	19/80	17/80	7/24	31/120	1

Example 2 - Properties of the joint table

- The total sum of the joint table is 1.
- The column of row totals contains the *(marginal)* probabilities $\mathrm{P}(S_i)$ of the 4 stimuli.
- The row of column totals de columnes contains the *(marginal)* probabilities $\mathrm{P}(R_j)$ dels 4 responses.
- Both vectors of *(marginal probabilities)* have sum 1.

Example 2 - Solution

Immediately from the table we see that:

$$\mathrm{P}(S_i \cap R_j) \neq \mathrm{P}(S_i) \cdot \mathrm{P}(R_j),$$

for all pairs (i, j), hence stimuli and responses are not independent.

From the definition of conditional probability:

$$\mathrm{P}(S_2|R_2) = \frac{\mathrm{P}(S_2 \cap R_2)}{\mathrm{P}(R_2)} = \frac{3/20}{17/80} = \frac{12}{17}.$$