01b - Probability - 02

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

Josep Fortiana

Wednesday, September 19, 2018

Universitat de Barcelona

01b - Probability 02

01b - Probability 02

The chain rule for probabilities

Total probabilities formula

Bayes' rule

The Monty Hall problem

For two events

From the definition of conditional probability, given events A_1 and A_2 such that $\mathrm{P}(A_1) > 0$, then:

$$\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\ \mathrm{P}(A_2|A_1).$$

For three events

Given events A_1, A_2 and A_3 such that $\mathrm{P}(A_1 \cap A_2) > 0$, then:

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\, \mathrm{P}(A_2|A_1)\, \mathrm{P}(A_3|A_1 \cap A_2).$$

Proof is immediate, by applying the previous property.

In general

For $k > 1$, given events $A_1, A_2, \ldots, A_k$ such that $\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_{k-1}) > 0$, then:

$$
\begin{aligned}
&\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_{k-1} \cap A_k) \\
&\qquad = \mathrm{P}(A_1)\,\mathrm{P}(A_2|A_1)\,\cdots\,\mathrm{P}(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).
\end{aligned}
$$

Proof by recurrence.

01b - Probability 02

The chain rule for probabilities

Total probabilities formula

Bayes' rule

The Monty Hall problem

Total probabilities formula

Given events $C_1, \ldots, C_k$, such that:

$$\Omega = \bigsqcup_{i=1}^{k} C_i, \qquad C_i \cap C_j = \emptyset, \quad i \neq j, \quad \text{all } \mathrm{P}(C_i) > 0.$$

The probability of $A \subset \Omega$ is the sum:

$$\mathrm{P}(A) = \sum_{i=1}^{k} \mathrm{P}(A \cap C_i) = \sum_{i=1}^{k} \mathrm{P}(A|C_i)\,\mathrm{P}(C_i).$$

Example 1

We extract a ball at random from an urn containing 2 red balls, 3 white balls and 2 black balls and deposit it into a second urn already containing 2 white balls and 3 black balls.

What is the probability of extracting now a black ball from the second urn (after mixing)?

Solution to example 1

Consider the partition of Ω into the three subsets:

$$R_1 = \{1^{st} \text{ ball is red}\},$$
$$W_1 = \{1^{st} \text{ ball is white}\},$$
$$B_1 = \{1^{st} \text{ ball is black}\}.$$

We want to compute the probability of:

$$B_2 = \{2^{nd} \text{ ball is black}\}.$$

Solution to example 1

$$
\begin{aligned}
\mathsf{P}(B_2) &= \mathsf{P}(B_2|R_1)\cdot\mathsf{P}(R_1) + \mathsf{P}(B_2|W_1)\cdot\mathsf{P}(W_1) + \\
&\quad + \mathsf{P}(B_2|B_1)\cdot\mathsf{P}(B_1) = \\
&= \frac{3}{6}\times\frac{2}{7} + \frac{3}{6}\times\frac{3}{7} + \frac{4}{6}\times\frac{2}{7} \\
&= \frac{23}{42}.
\end{aligned}
$$

Example 2

An urn contains 6 white balls labelled 1 to 6.

We extract a ball and paint in black as many balls as the label of the extracted ball indicates.

Then (after mixing) we extract a second ball.

What is the probability that this second ball is white?

Solution to example 2

Consider the 6 events from the 1^{st} extraction:

$$C_i = \{\text{Extract the ball labelled } i\}, \qquad 1 \leq i \leq 6.$$

They are a partition of the total Ω.

With the notation:

$$W = \{\text{The second extracted ball is white}\},$$

Solution to example 2

$$\mathrm{P}(W) = \sum_{i=1}^{6} \mathrm{P}(W|C_i) \cdot \mathrm{P}(C_i) =$$

$$= \sum_{i=1}^{6} \frac{6-i}{6} \cdot \frac{1}{6} =$$

$$= \frac{5}{12}.$$

Example 3

Two urns, U_1 and U_2, contain white (w) and red (r) balls in the following proportions:

$$U_1 = (3w, 5r), \qquad U_2 = (2w, 1r).$$

We roll a die. If a 3 or a 6 turns up then a randomly selected ball from U_2 is transferred to U_1 and then (after mixing) we extract at random a ball from U_1.

Otherwise a random ball from U_1 is transferred to U_2. Then we mix and extract at random a ball from U_2.

Example 3 (continued)

Compute the probabilities:

1. That both balls are red.
2. That both balls are white.

Example 3 - Preparation

$$\begin{aligned} \text{Notation:} \quad \alpha \quad &= \text{“a 3 or a 6 turns up”}. \\ \text{Then,} \quad P(\alpha) &= 1/3, \quad P(\alpha^c) = 2/3. \end{aligned}$$

We construct a tree with all possible results.

Probabilities written along the branches are conditional (on the node from which the branch sprouts).

Example 3 - The tree

$U_1 = (3w, 5r)$
$U_2 = (2w, 1r)$

- α (1/3): Ball from U_2 to U_1
  - w (2/3): $U_1 = (4w, 5r)$
  - r (1/3): $U_1 = (3w, 6r)$
- α^c (2/3): Ball from U_1 to U_2
  - w (3/8): $U_2 = (3w, 1r)$
  - r (5/8): $U_2 = (2w, 2r)$

Example 3 - Solution

Probability of extracting two red balls:

$$
\begin{aligned}
\mathrm{P}(2r) &= \mathrm{P}(2r|\alpha)\cdot\mathrm{P}(\alpha)+\mathrm{P}(2r|\alpha^c)\cdot\mathrm{P}(\alpha^c)\\
&= (1/3\times 6/9)\times 1/3+(5/8\times 1/2)\times 2/3\\
&= 61/216.
\end{aligned}
$$

Example 3 - Solution

Probability of extracting two white balls

$$
\begin{aligned}
P(2w) &= P(2w|\alpha) \cdot P(\alpha) + P(2w|\alpha^c) \cdot P(\alpha^c) \\
&= (2/3 \times 4/9) \times 1/3 + (3/8 \times 3/4) \times 2/3 \\
&= 371/1296.
\end{aligned}
$$

01b - Probability 02

Rev.Thomas Bayes (c.1701 – 7 April 1761)

Posthumous Essay:

Thomas Bayes (1763),

An essay towards solving a problem in the doctrine of chances,

Philosophical Transactions of the Royal Society of London, 53(0), 370-418.

Bayes' rule (elementary, for probabilities)

If $\mathrm{P}(A) > 0$ and $\mathrm{P}(B) > 0$, then both $\mathrm{P}(A|B)$ and $\mathrm{P}(B|A)$ are well defined.

The *elementary Bayes formula* relates them:

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

Interpretation: evidence that event A has occurred modifies the *prior* probability of its posible antecedent or cause B.

Inverse probability

If an event A has $k > 1$ possible antecedents or causes, $C_1, \ldots, C_k$, and we know the conditional probabilities:

$$\mathrm{P}(A|C_i), \qquad 1 \leq i \leq k,$$

and we acquire *the evidence* that A has happened,

We can compute the *inverse probability* of each of the possible causes.

Requirements for Bayes' rule

The events $C_1, \ldots, C_k$ must be a partition:

$$\Omega = \bigsqcup_{i=1}^{k} C_i, \qquad C_i \cap C_j = \emptyset, \quad i \neq j.$$

Needed conditions: $\mathrm{P}(A) > 0$ and all $\mathrm{P}(C_i) > 0$.

Bayes' rule

For the j-th cause, $1 \leq j \leq k$,

$$\mathrm{P}(C_j|A) = \frac{\mathrm{P}(A|C_j)\ \mathrm{P}(C_j)}{\sum_{i=1}^{k} \mathrm{P}(A|C_i)\ \mathrm{P}(C_i)}.$$

Proof of Bayes' rule

The denominator is the total probability $\mathrm{P}(A)$.

The numerator is the intersection probability $\mathrm{P}(C_j, A)$, so this formula reduces to the one above.

Bayes' rule in statistical practice

A model consists of the k possible "causes" C_j of the observed data.

Their *a priori* or *initial* probabilities $\mathrm{P}(C_i)$, before the observation.

A posteriori o *final* probabilities $\mathrm{P}(C_i|A)$, blending in the *information* or *evidence* that A has been observed.

Bayes' rule in statistical practice

$$\mathrm{P}(\text{Model}|\text{Data}) = \frac{\mathrm{P}(\text{Data}|\text{Model})\,(\text{a priori }\ \mathrm{P}(\text{Model}))}{\mathrm{P}(\text{Data})}.$$

Interpretation: *A priori* knowledge (or ignorance) of a model is modified by experimental data, to obtain *a posteriori* knowledge, merging both sources of information.

Bayes reasoning

1. Initially, the *a priori* probability $\mathsf{P}(B)$ is known.
2. We blend in the *evidence* that A has occurred,
3. The initial probability is transformed into the *final*, *a posteriori*, probability $\mathsf{P}(B|A)$.

Bayes' rule with LEGO

Count Bayesie Blog: Probably a Probability Blog.

A Guide to Bayesian Statistics.

Bayes' Theorem with Lego.

Example 4

30% of the people in a city are vaccinated against flu. The probability of catching flu is 0.01 for vaccinated individuals and 0.1 for non-vaccinated individuals.

What is the probability that a patient with flu has been vaccinated?

What is the probability that a given individual who has not caught the flu has been vaccinated?

Solution to example 4

Notation: a randomly selected individual:

$$V = \{\text{has been vaccinated}\},$$

$$F = \{\text{has caught flu}\}.$$

From the statement,

$$\mathrm{P}(V) = \frac{3}{10}, \quad \mathrm{P}(F|V) = \frac{1}{100}, \quad \mathrm{P}(F|V^c) = \frac{1}{10}.$$

Solution to example 4

$$\mathrm{P}(V|F) = \frac{\mathrm{P}(F|V)\,\mathrm{P}(V)}{\mathrm{P}(F|V)\,\mathrm{P}(V) + \mathrm{P}(F|V^c)\,\mathrm{P}(V^c)} = \frac{3}{73}.$$

$$\mathrm{P}(V|F^c) = \frac{\mathrm{P}(F^c|V)\,\mathrm{P}(V)}{\mathrm{P}(F^c|V)\,\mathrm{P}(V) + \mathrm{P}(F^c|V^c)\,\mathrm{P}(V^c)} = \frac{33}{103}.$$

Example 5

A bag contains 8 red, 15 white, and 5 yellow balls.

The experimenter extracts a ball at random and registers its color (does not tell).

If the ball is red or yellow it is returned to the bag, adding 2 more balls of the same color; if the ball is white it is kept out of the bag.

Then a second ball is extracted.

Example 5

Compute the probability:

1. That the second ball is red.

2. That the first ball was not red if we observe the second ball is red.

Example 5 - Preparation

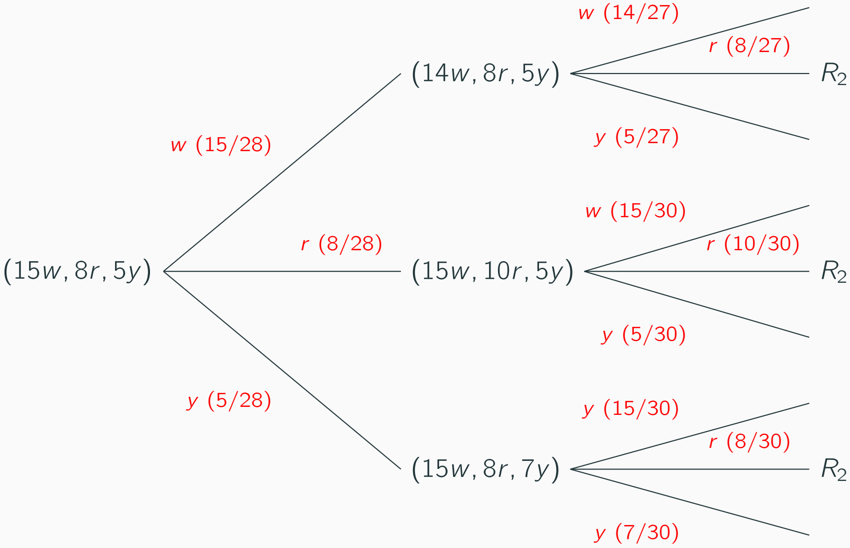
Notations:

$$\begin{aligned} W_1, R_1, Y_1 &= \text{White, red, yellow ball on the first extraction,} \\ R_2 &= \text{Red ball on the second extraction.} \end{aligned}$$

Questions are:

$$\mathrm{P}(R_2), \qquad \mathrm{P}(R_1^c | R_2).$$

Exemple 5 - Tree



Example 5 - Solution (1)

$$
\begin{aligned}
\mathrm{P}(R_2) &= \mathrm{P}(R_2|W_1)\cdot\mathrm{P}(W_1) + \mathrm{P}(R_2|R_1)\cdot\mathrm{P}(R_1)\\
&\quad + \mathrm{P}(R_2|Y_1)\cdot\mathrm{P}(Y_1)\\
&= \frac{8}{27}\times\frac{15}{28} + \frac{10}{30}\times\frac{8}{28} + \frac{8}{30}\times\frac{5}{28}\\
&= \frac{10}{63} + \frac{6}{63} + \frac{3}{63} = \frac{19}{63}.
\end{aligned}
$$

Example 5 - Solution (2)

$$
\begin{aligned}
\mathrm{P}(R_1^c | R_2) &= \frac{\mathrm{P}(R_1^c \cap R_2)}{\mathrm{P}(R_2)} = \frac{\mathrm{P}((W_1 \cup Y_1) \cap R_2)}{\mathrm{P}(R_2)} \\
&= \frac{\mathrm{P}(W_1 \cap R_2) + \mathrm{P}(Y_1 \cap R_2)}{\mathrm{P}(R_2)} \\
&= \frac{\mathrm{P}(R_2 | W_1) \cdot \mathrm{P}(W_1) + \mathrm{P}(R_2 | Y_1) \cdot \mathrm{P}(Y_1)}{\mathrm{P}(R_2)} \\
&= \frac{10/63 + 3/63}{19/63} = \frac{13}{19}.
\end{aligned}
$$

Example 6

A box A contains 9 cards numbered 1 to 9 and another box B contains 5 cards numbered 1 to 5.

A box is chosen at random and a card is extracted from it, also at random.

If the card number is even, whithout returning it, a second card is extracted from the same box.

If the card number is odd, a card is extracted from the other box.

Example 6 (continued)

1. What is the probability that both card numbers are even?
2. If both card numbers are even, what is the probability that they came from box A?
3. What is the probability that both card numbers are odd?

Example 6 - Preparation

We write a description of all possible paths, annotating on it the transition *(= changes in configuration)* probabilities.

Instead of a tree format here we use a table format. Just an equivalent visual alternative.

Example 6 - Transitions and probabilities

$e =$ even; $o =$ odd.

1^{st} box selection	1^{st} card extraction	2^{nd} box selection	2^{nd} card extraction	
A $(1/2)$	e $(4/9)$	A	e $(3/8)$	$1/12$
		$(3e{,}5o)$	o $(5/8)$	$5/36$
$(4e{,}5o)$	o $(5/9)$	B	e $(2/5)$	$1/9$
		$(2e{,}3o)$	o $(3/5)$	$1/6$
B $(1/2)$	e $(2/5)$	B	e $(1/4)$	$1/20$
		$(1e{,}3o)$	o $(3/4)$	$3/20$
$(2e{,}3o)$	o $(3/5)$	A	e $(4/9)$	$2/15$
		$(4e{,}5o)$	o $(5/9)$	$1/6$

Example 6 - Computations

Directly from the table, the answers (1) and (3) are:

$$\mathrm{P}(ee) = \frac{1}{12} + \frac{1}{20} = \frac{2}{15},$$

$$\mathrm{P}(oo) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Example 6 - Applying Bayes' rule

In (2) we are asked the conditional probability:

$$\mathrm{P}(AA|ee) = \frac{\mathrm{P}(ee|AA)\cdot\mathrm{P}(AA)}{\mathrm{P}(ee|AA)\cdot\mathrm{P}(AA)+\mathrm{P}(ee|BB)\cdot\mathrm{P}(BB)}$$

$$= \frac{\mathrm{P}(ee|AA)\cdot\mathrm{P}(AA)}{\mathrm{P}(ee)}$$

$$= \frac{1/12}{2/15} = \frac{5}{8}.$$

Example 7: the Kahneman cab problem

A cab was involved in a hit and run accident at night.

Two cab companies, the Green and the Blue, operate in the city. 85% of the cabs in the city are Green and 15% are Blue.

A witness identified the cab as Blue.

Kahneman, Daniel, *Thinking, fast and slow*,
New York: Farrar, Straus and Giroux (2013).

Example 7: the Kahneman cab problem (continued)

The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colours 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue rather than Green knowing that this witness identified it as Blue?

Example 7: the Kahneman cab problem (notations)

$$G = \text{“The guilty cab is Green”}$$

$$B = \text{“The guilty cab is Blue”}$$

$$W_G = \text{“Witness says the guilty cab is Green”}$$

$$W_B = \text{“Witness says the guilty cab is Blue”}$$

Example 7: the Kahneman cab problem (Data)

A priori probabilities:

$$\begin{aligned} \mathrm{P}(G) &= 0.85, \\ \mathrm{P}(B) &= 0.15, \end{aligned}$$

Likelihood of each possible witness response, given each possible factual situation:

$$\begin{aligned} \mathrm{P}(W_B|B) &= 0.80, & \mathrm{P}(W_G|B) &= 0.20, \\ \mathrm{P}(W_B|G) &= 0.20, & \mathrm{P}(W_G|G) &= 0.80. \end{aligned}$$

Example 7: the Kahneman cab problem (solution)

Applying Bayes' rule:

$$
\begin{aligned}
\mathrm{P}(B|W_B) &= \frac{\mathrm{P}(B \cap W_B)}{\mathrm{P}(W_B)} \\
&= \frac{\mathrm{P}(W_B|B) \cdot \mathrm{P}(B)}{\mathrm{P}(W_B|B) \cdot \mathrm{P}(B) + \mathrm{P}(W_B|G) \cdot \mathrm{P}(G)} \\
&= \frac{0.80 \times 0.15}{0.80 \times 0.15 + 0.20 \times 0.85} = 0.4139.
\end{aligned}
$$

01b - Probability 02

Setting

A TV game show. Choose one of three closed doors.

One of them hides a hefty prize (a car).

If you choose it you win the car.

Behind each of the other two doors there is a goat.

If you choose either of them you win nothing.

Procedure

You choose a door. For the time being, it remains closed.

The game show host (Monty Hall), who knows where the car is hidden, opens one of the unselected doors, showing a goat.

Then you have the opportunity to maintain your initial choice or to switch to the remaining closed door.

Which is the best strategy?

Stick to the first choice or change?

Or, perhaps, it is indifferent?

Once a door has been selected, for sure at least one of the other two hides a goat, hence opening one of them supplies no new information.

Thus we can conclude that switching will not affect chances of winning.

Which is the best strategy?

Stick to the first choice or change?

Or, perhaps, it is indifferent?

Once a door has been selected, for sure at least one of the other two hides a goat, hence opening one of them supplies no new information.

Thus we can conclude that switching will not affect chances of winning.

Or does it?

Experiment

Perform an experiment, simulating many repetitions of the game with each strategy.

We observe results are VERY different.

Switching doors, the prize is won
DOUBLE number of times
as in the conservative strategy.

A Bayesian solution: 1. Hypotheses, prior probabilities

$\Longrightarrow$ Notebooks for *Think Bayes* by Roger Labbe

Three hypotheses H:

$$H = A, \quad H = B, \quad \text{and} \quad H = C,$$

respectively representing the (unknown but possible) fact that the car is behind the namesake door.

Initially, the *prior probabilities* $\mathsf{P}(H)$ are:

$$\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = 1/3.$$

2. Observed data

AFTER picking your door, relabel the three doors so that A is your choice.

Assume the observed data is:

$$D = \boxed{\begin{array}{c}\text{Monty has opened door } B\\ \text{AND}\\ \text{inside door } B \text{ there is a goat.}\end{array}}$$

3. Likelihood

Conditional probabilities $P(D|H)$ of the observed data D, assuming each possible hypothesis H as being true:

	Prior $P(H)$	Likelihood $P(D\|H)$
A	$1/3$	$1/2$
B	$1/3$	0
C	$1/3$	1

4. Joint probabilities

Multiply each $\mathrm{P}(D|H)$ times the prior $\mathrm{P}(H)$, giving the Joint or intersection probabilities $\mathrm{P}(D \cap H) \equiv \mathrm{P}(D, H)$:

	Prior $\mathrm{P}(H)$	Likelihood $\mathrm{P}(D\|H)$	Joint $\mathrm{P}(D, H)$
A	1/3	1/2	1/6
B	1/3	0	0
C	1/3	1	1/3

5. Bayes' formula $\Longrightarrow$ Posterior probabilities

Divide $\mathrm{P}(D,H)$ by their sum $\mathrm{P}(D) = 1/2$, giving the posterior probabilities $\mathrm{P}(H|D) = \mathrm{P}(D,H)\,/\mathrm{P}(D)$:

	Prior $\mathrm{P}(H)$	Likelihood $\mathrm{P}(D\vert H)$	Joint $\mathrm{P}(D,H)$	Posterior $\mathrm{P}(H\vert D)$
A	$1/3$	$1/2$	$1/6$	$1/3$
B	$1/3$	0	0	0
C	$1/3$	1	$1/3$	$2/3$