# 01 - Probability - 01

Bayesian Statistics

Spring 2022-2023

Josep Fortiana

Matemàtiques - Informàtica UB

Monday, February 13, 2023

# 01 - Probability 01

# 01 - Probability 01

# Conditional probability (elementary definition)

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

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

How likely is event $A$, assuming we know that some antecedent $B$ has happened.

# Interpreting conditional probability

Either $B$ facilitates or hampers the occurrence of $A$.

$$\mathrm{P}(A|B) > \mathrm{P}(A) \;\Rightarrow\; B \text{ facilitates the occurrence of } A.$$

$$\mathrm{P}(A|B) = \mathrm{P}(A) \;\Rightarrow\; A \text{ and } B \text{ are independent.}$$

$$\mathrm{P}(A|B) < \mathrm{P}(A) \;\Rightarrow\; B \text{ hampers the occurrence of } A.$$

# Example

Regular die: each of six possible results has probability $\frac{1}{6}$.

New probabilites conditional to the event:

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

# Example

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

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

# Example

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

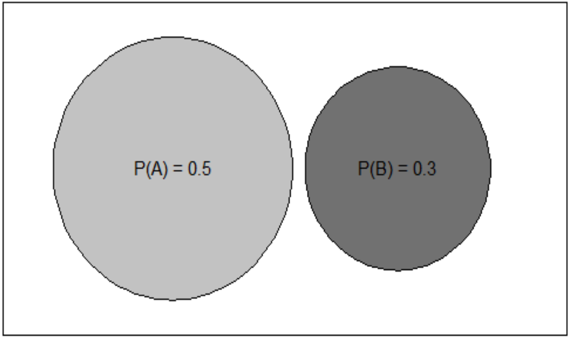
If both $\operatorname{P}(A) > 0$ and $\operatorname{P}(B) > 0$ we can compute both conditional probabilities:

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

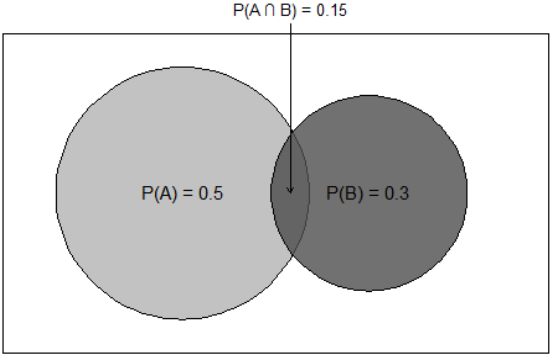
This equality is the source of Bayes' rule.

# 01 - Probability 01

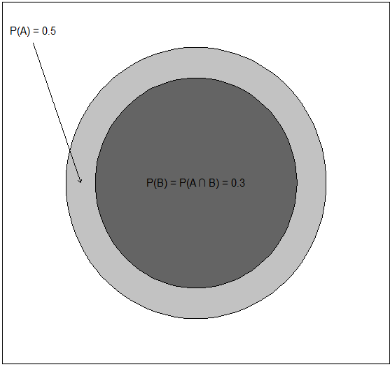
# Diagram 1



# 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:

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

# 01 - Probability 01

Conditional probability

Independent events

Bayes' rule

Bayes' billiard (1763)

# 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 (for probabilities)

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

The *elementary Bayes formula* relates them:

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

Interpretation: evidence that $A$ has occurred turns *prior* probability of $B$ into *posterior* probability.

# 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:

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

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

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

Numerator is the intersection probability $\mathrm{P}(C_j, A)$.

# 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

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

Interpretation: Experimental data turns *a priori* knowledge (or ignorance) of a model into *a posteriori* knowledge, merging both sources of information.

# Bayes reasoning

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

# Bayes' rule with LEGO

Count Bayesie Blog: Probably a Probability Blog.

*A Guide to Bayesian Statistics.*

Bayes' Theorem with Lego.

# Example problem

30% of the people in a city are vaccinated against flu.

Probability of catching flu: $0.01$ for vaccinated individuals and $0.1$ for non-vaccinated individuals.

Probability that a patient with flu has been vaccinated?

Probability that a given individual who has not caught the flu has been vaccinated?

# Solution

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

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

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

# 01 - Probability 01

# Game description

Alice and Bob play a game: the first one to get 6 points wins.

Pool table that players can't see.
An initial (cue) ball is rolled onto the table.

It comes to rest at a random position,
which is marked but not revealed.

Each point is decided by rolling a new ball onto the table randomly.

# Problem sketch

If the ball comes to rest to the left of the initial mark, Alice wins the point; if to the right, Bob wins the point.

1-dimensional schematic description:


$$p = \{\text{Probability that Alice gets a point,}\}$$

# Setting

Assume Alice is already winning, 5 points to Bob's 3 points

(so with one more point she has 6 and wins the game).

We are asked to evaluate $\quad \mathrm{P}(\{\text{Bob wins the game}\})$

(by winning in a row the next 3 points).

# Development

If we knew the position of the initial ball:

$$p = \{\text{Probability that Alice gets a point,}\}$$

# Development

If we knew the position of the initial ball:

$$p = \{\text{Probability that Alice gets a point,}\}$$

Then the answer is just $(1-p)^3$.

# Development

If we knew the position of the initial ball:

$$p = \{\text{Probability that Alice gets a point,}\}$$

Then the answer is just $(1-p)^3$.

But we don't know!

# Frequentist approach: Estimate $p$ from data

Given $p$, we can compute:

$$\mathrm{P}(A = 5,\ B = 3 \mid p) = \binom{8}{3}\, p^5\, (1 - p)^3.$$

This is the Likelihood, as a function of $p$.

Its maximum is attained for:

$$p = \widehat{p}_{ML} = \frac{5}{8},$$

the Maximum Likelihood estimate of $p$.

# Frequentist result

With this value, the estimate of Bob's probability of winning is:

$$\mathrm{P}_{FREQ}(\text{Bob wins}) = (\,1 - \widehat{p}\,)^3 = \left(\frac{3}{8}\right)^3 = \frac{27}{512} = 0.0527.$$

The odds are:

$$\text{odds}_{FREQ}(\text{Bob wins}) = \frac{27/512}{1 - 27/512} = \frac{27}{512 - 27} = 0.05567 \approx \frac{1}{18}.$$

# Bayesian approach

Acknowledge $p$ is unknown. Then treat it as such.

# Bayesian approach

Acknowledge $p$ is unknown. Then treat it as such.

The probability of Bob winning will be a weighted average over all possible values of $p$.

# Bayesian approach

Acknowledge $p$ is unknown. Then treat it as such.

The probability of Bob winning will be a weighted average over all possible values of $p$.

$$\mathrm{P}_{BAYES}(\text{Bob wins}) = \int_0^1 (1-p)^3 \cdot \mathrm{P}(p \mid A=5,\ B=3)\, dp.$$

# Bayesian approach

Acknowledge $p$ is unknown. Then treat it as such.

The probability of Bob winning will be a weighted average over all possible values of $p$.

$$\mathrm{P}_{BAYES}(\text{Bob wins}) = \int_0^1 (1-p)^3 \cdot \mathrm{P}(p \mid A=5,\ B=3)\, dp.$$

For each $p$, the weight is the probability $\mathrm{P}(p \mid A=5,\ B=3)$ of this particular value, given the observed data.

# Avoid mistaking one probability for another

This is NOT the quantity we met before, the Likelihood,

$$\mathrm{P}(A = 5,\ B = 3 \mid p),$$

probability of the observed data given some fixed *p* value.

Now we want:

$$\mathrm{P}(p \mid A = 5,\ B = 3),$$

the probability of a *p* value, given the observed data.

This is the Posterior or *"a posteriori"* probability.

# From one probability to the other: Bayes' rule

For any two random quantities, $X$ and $Y$,

$$\mathrm{P}(X \mid Y) = \frac{\mathrm{P}(Y, X)}{\mathrm{P}(Y)} = \frac{\mathrm{P}(Y \mid X) \cdot \mathrm{P}(X)}{\mathrm{P}(Y)} = \frac{\mathrm{P}(Y \mid X) \cdot \mathrm{P}(X)}{\sum_{X'} \mathrm{P}(Y \mid X') \cdot \mathrm{P}(X')}.$$

Here "P" stands for "probability" or pdf or pmf, as appropriate.
For pdf's the summation will be an integral.

*Proof:* Just the definition of conditional probability.

# Applying Bayes' rule

$$\mathrm{P}(p \mid A=5,\ B=3)=\frac{\mathrm{P}(A=5,\ B=3 \mid p)\cdot \mathrm{P}(p)}{\displaystyle\int_0^1 \mathrm{P}(A=5,\ B=3 \mid p)\cdot \mathrm{P}(p)\, dp}$$

# Applying Bayes' rule

$$P(p \mid A = 5,\ B = 3) = \frac{P(A = 5,\ B = 3 \mid p) \cdot P(p)}{\displaystyle\int_0^1 P(A = 5,\ B = 3 \mid p) \cdot P(p)\, dp}$$

$P(p)$, both in numerator and denominator, is the prior or *"a priori"* pdf, the probability of a given $p$ before recording any data.

# Applying Bayes' rule

$$\mathrm{P}(p \mid A=5,\ B=3) = \frac{\mathrm{P}(A=5,\ B=3 \mid p) \cdot \mathrm{P}(p)}{\displaystyle\int_0^1 \mathrm{P}(A=5,\ B=3 \mid p) \cdot \mathrm{P}(p)\, dp}$$

$\mathrm{P}(p)$, both in numerator and denominator, is the prior or *"a priori"* pdf, the probability of a given $p$ before recording any data.

Since $p$ is uniform on $[0, 1]$, $\mathrm{P}(p)$ is a constant,
thus it simply cancels out.

# Putting everything together

Substituting $\mathrm{P}(A = 5,\ B = 3 \mid p) = \binom{8}{3}\, p^5\, (1 - p)^3$ in the above formula and then in:

$$\mathrm{P}_{BAYES}(\text{Bob wins}) = \int_0^1 (1 - p)^3 \cdot \mathrm{P}(p \mid A = 5,\ B = 3)\, dp.$$

we have our result:

$$\mathrm{P}_{BAYES}(\text{Bob wins}) = \frac{\displaystyle\int_0^1 p^5 \cdot (1 - p)^6\, dp.}{\displaystyle\int_0^1 p^5 \cdot (1 - p)^3\, dp.}$$

# The Beta function

The *Beta function* is defined as:

$$\mathrm{B}(x, y) = \int_0^1 t^{x-1} \cdot (1-t)^{y-1}\, dt,$$

for $x > 0, y > 0$.

$$\mathrm{P}_{BAYES}(\text{Bob wins}) = \frac{\mathrm{B}(6, 7)}{\mathrm{B}(6, 4)}.$$

# The Gamma function

Beta function values can be obtained from:

$$\mathrm{B}(x, y) = \frac{\Gamma(x) \cdot \Gamma(y)}{\Gamma(x + y)},$$

$\Gamma(\,\cdot\,)$ is the *Gamma* function.

For a positive integer $n$, $\Gamma(\,\cdot\,)$ is the factorial function:

$$\Gamma(n) = (n - 1)!$$

# Result

$$P_{BAYES}(\text{Bob wins}) = \frac{B(6,7)}{B(6,4)} = \frac{6!\cdot 9!}{12!\cdot 3!} = \frac{1}{11} = 0.09091.$$

The odds are:

$$\text{odds}_{BAYES}(\text{Bob wins}) = \frac{1}{10} = 0.1.$$

# Result

$$P_{BAYES}(\text{Bob wins}) = \frac{B(6,7)}{B(6,4)} = \frac{6!\cdot 9!}{12!\cdot 3!} = \frac{1}{11} = 0.09091.$$

The odds are:

$$\text{odds}_{BAYES}(\text{Bob wins}) = \frac{1}{10} = 0.1.$$

Quite different from the classical result!

# Result

$$P_{BAYES}(\text{Bob wins}) = \frac{B(6,7)}{B(6,4)} = \frac{6!\cdot 9!}{12!\cdot 3!} = \frac{1}{11} = 0.09091.$$

The odds are:

$$\text{odds}_{BAYES}(\text{Bob wins}) = \frac{1}{10} = 0.1.$$

Quite different from the classical result!

Which one is right?

# References


Downey, Allen (2015), *Bayesian Billiards*
Thursday, June 25, 2015 entry in his Blog *Probably Overthinking It*

Vanderplas, Jake (2014) *Frequentism and Bayesianism: What's The Big Deal?* YouTube of talk at SciPy 2014

Eddy, Sean R. (2004) *What is Bayesian statistics?* Nature Biotechnology, 22(9):1177-1178.

David Spiegelhalter (2018) People of Science with Brian Cox (Vimeo)

# And, of course

Bayes, T. (1763). *An Essay Towards Solving a Problem in the Doctrine of Chances.* Philosophical Transactions of the Royal Society of London, 53:370-418.