COMP229: Introduction to Data Science
Lecture 5: Conditional probabilities and Bayes theorem.

Olga Anosova, O.Anosova@liverpool.ac.uk
Autumn 2023, Computer Science department
University of Liverpool, United Kingdom

Reminder: sum and product rules

- Probabilities can be computed after defining a probability space $(\Omega, \mathcal{E}, P)$ satisfying 3 axioms
  A1: positivity $P(A) \geqslant 0$ for any event A
  A2: something always happens $P(\Omega) = 1$
  A3: additivity $P(A \cup B) = P(A) + P(B)$ for mutually exclusive A, B.
- Sum rule: for any events A, B
  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Product rule: for independent events
  $P(A \cap B) = P(A)P(B)$

Fairness in coins

Problem 5.1. Flipping a fair coin, you got 2023 heads. What's the probability to get a head again?

Solution 5.1. The probability to get a head by flipping a fair coin is always 0.5, because this event doesn't depend on previous flips.

Fairness in coins

Problem 5.1. Flipping a fair coin, you got 2023 heads. What's the probability to get a head again?

Solution 5.1. The probability to get a head by flipping a fair coin is always 0.5, because this event doesn't depend on previous flips.
The probability to get 2023 heads in a row is very small: 0.5^{2023}.

Is there a way to adjust our expectations from obtained evidence?

Fair result from a biased coin

A biased coin might have different probabilities of a head and a tail, e.g. $P(H) = 0.4$ and $P(T) = 0.6$.

Here is John von Neumann's method to get fair results when flipping a biased coin:

1) flip the coin twice;

2) if we get HH or TT, flip the coin twice again;

3) if we get HT or TH, ignore the second flip and use the outcome of the first flip.

Fairness explained

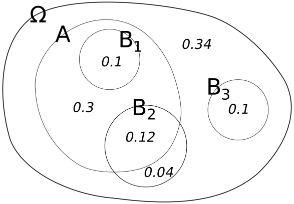
Von Neumann's method gives equally probable outcomes even for a biased coin.

Proof. Suppose that $P(H) = p$, $P(T) = q$ for $p, q \neq 0$. Flips are independent even for a biased coin, so $P(HH) = p^2, P(TT) = q^2$,
$P(HT) = pq = P(TH)$.
The events HT and TH have the same probability even for a biased coin, which works if after every two flips we start again and ignore all previous results.

Conditional probability

Definition 5.2. For events A, B, the **conditional** probability that A happens assuming B has happened is defined as $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

Example.



$P(A) = 0.30 + 0.10 + 0.12 = 0.52$

$P(A|B1) = 1,$

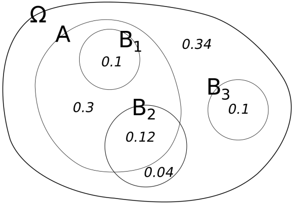
$P(A|B2) = \frac{0.12}{0.12+0.04} = 0.75,$ and

$P(A|B3) =$

Conditional probability

Definition 5.2. For events A, B, the **conditional** probability that A happens assuming B has happened is defined as $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

Example.



$$P(A) = 0.30 + 0.10 + 0.12 = 0.52$$
$$P(A|B1) = 1,$$
$$P(A|B2) = \frac{0.12}{0.12+0.04} = 0.75, \text{ and}$$
$$P(A|B3) = 0.$$

By Gnathan87 - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=15991401

Colour blindness problem

Because males have only one X chromosome and females have two, men are about 16 times more likely to suffer adverse effects of a defective X chromosome and therefore to be *colour blind*.

Problem 5.3. Assume that the male/female ratio is 50/50. About 4.25% of all people are colour blind. Find these conditional probabilities

P(colour blind | male), P(colour blind | female).

Colour blind solution

P(male, colour blind) $=$ P(male) $\times$ P(colour blind) $= 0.5 \times 0.0425 \approx 2\%$ and P(female, colour blind) $=$ P(female) $\times$ P(colour blind) $= 0.5 \times 0.0425 \approx 2\%$, the same.

But we know that males are 16 times more likely to be colour blind than females:
P(colour blind $|$ male) $=$ 16 P(colour blind $|$female)

The product rule applies only to *independent* events.

General product rule

Claim 5.4. General product rule: for any A, B

$P(A \cap B) = P(A|B)P(B)$.

Proof. From the definition of conditional probaility,

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

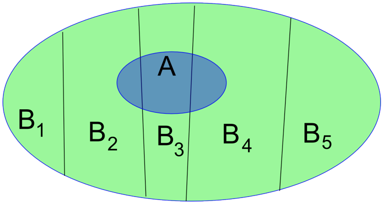
How about independent events C, D?

$$P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{P(C)P(D)}{P(D)} = P(C),$$

and, similarly, $P(D|C) = P(D)$.

Law of total probability

If the sample space Ω (i.e. all possible outcomes) can be partitioned into pairwise disjoint events $\Omega = \sum_n B_n$, then for any event A



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

hence

$$P(A) = \sum_n P(A \cap B_n) = \sum_n P(A|B_n)P(B_n).$$

Reminder: Colour blindness problem

Because males have only one X chromosome and females have two, men are about 16 times more likely to suffer adverse effects of a defective X chromosome and therefore to be *colour blind*.

Problem 5.3. Assume that the male/female ratio is 50/50. About 4.25% of all people are colour blind. Find these conditional probabilities

P(colour blind | male), P(colour blind | female).

Finding conditional probabilities

Solution 5.3. 4.25%=P(colour blind) $=$
P(colour blind, male) $+$ P(colour blind, female) $=$
P(colour blind | male) $\times$ P(male) $+$
$+$ P(colour blind | female) $\times$ P(female)$=$
0.5(P(colour blind | male)$+$P(colour blind |female))

$=$

$=$0.5(16$+$1)P(colour blind |female), hence

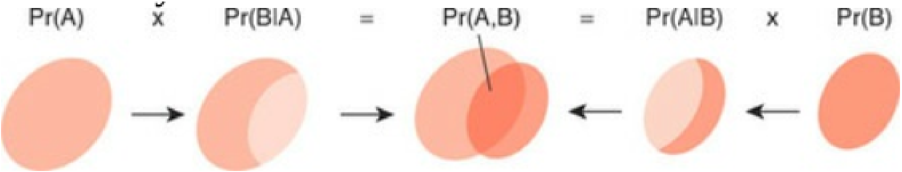
$$\text{P(colour blind | female)} = \frac{4.25\%}{0.5 \times 17} = 0.5\%.$$

$$\text{P(colour blind | male)} = 16 \times 0.5\% = 8\%.$$

The Bayes-Price-Laplace formula

Theorem 5.5. $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$.

Proof. $P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$.



The diagram illustrates the proof with another notation for the intersection:
$P(A \cap B) = P(A, B)$. $\square$

Applying the Bayes formula

Problem 5.6. Let P(colour blind | male)=8%, P(colour blind |male)=0.5%, P(colour blind)= 4.25%. A random person turns out to be colour blind. What's the probability that this is a male?

Solution 5.6. P(male | colour blind)=

Applying the Bayes formula

Problem 5.6. Let P(colour blind | male)=8%, P(colour blind |male)=0.5%, P(colour blind)= 4.25%. A random person turns out to be colour blind. What's the probability that this is a male?

Solution 5.6. P(male | colour blind)=

$$\frac{P(\text{colour blind}|\text{male})P(\text{female})}{P(\text{colour blind})} =$$

Applying the Bayes formula

Problem 5.6. Let P(colour blind | male)=8%, P(colour blind |male)=0.5%, P(colour blind)= 4.25%. A random person turns out to be colour blind. What's the probability that this is a male?

Solution 5.6. P(male | colour blind)=

$$\frac{P(\text{colour blind}|\text{male})P(\text{female})}{P(\text{colour blind})} = \frac{8\% \times 0.5}{4.25\%} = \frac{16}{17}.$$

Similarly P(female | colour blind)=

Applying the Bayes formula

Problem 5.6. Let P(colour blind | male)=8%, P(colour blind |male)=0.5%, P(colour blind)= 4.25%. A random person turns out to be colour blind. What's the probability that this is a male?

Solution 5.6. P(male | colour blind)=

$$\frac{P(\text{colour blind}|\text{male})P(\text{female})}{P(\text{colour blind})} = \frac{8\%\times 0.5}{4.25\%} = \frac{16}{17}.$$

Similarly P(female | colour blind)= $\frac{0.5\%\times 0.5}{4.25\%} = \frac{1}{17}$.

Rediscovering science

frontiers in
PSYCHOLOGY

METHODS ARTICLE
published: 16 November 2010
doi: 10.3389/fpsyg.2010.00192

Bayes' rule for clinicians: an introduction

Chris F. Westbury*

Department of Psychology, University of Alberta, Edmonton, AB, Canada

Edited by:
Edward Callus, IRCCS Policlinico San Donato, Italy

Reviewed by:
Gianluca Castelnuovo, Università Cattolica del Sacro Cuore, Italy
Gian Mauro Manzoni, Istituto Auxologico Italiano IRCCS, Italy
Francesco Pagnini, University of Bergamo, Italy

***Correspondence:**
Chris F. Westbury, Department of Psychology, University of Alberta, P220 Biological Sciences Building, Edmonton, AB, Canada, T6G 2E9.
e-mail: chrisw@ualberta.ca

Bayes' Rule is a way of calculating conditional probabilities. It is difficult to find an explanation of its relevance that is both mathematically comprehensive and easily accessible to all readers. This article tries to fill that void, by laying out the nature of Bayes' Rule and its implications for clinicians in a way that assumes little or no background in probability theory. It builds on Meehl and Rosen's (1955) classic paper, by laying out algebraic proofs that they simply allude to, and by providing extremely simple and intuitively accessible examples of the concepts that they assumed their reader understood.

Keywords: probability, diagnosis, Bayes theory, base rates

Bayes' Rule is a way of calculating conditional probabilities. Although it is simple in its conception, Baye's Rule can be fiendishly difficult for beginners to understand and apply. In part this is because it forces us to confront and overcome strong biases in our natural way of thinking and in part it is because it is not easy to be specific about exactly where Bayes' Rule will apply, or how it may apply in any particular case. The purpose of this paper is

The first section consists of a general introduction to understanding conditional probabilities. The second section introduces Bayes' Rule itself, in an historical and mathematical setting. The third section lays out some implications of Bayes' Rule that follow as a direct result of its definition.

CONDITIONAL PROBABILITIES

EDITOR'S CHOICE

A qualitative approach to Bayes' theorem

Mitchell A Medow,[1] Catherine R Lucey[2]

10.1136/ebm-2011-0007

[1]Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts, USA
[2]Internal Medicine, The Ohio State University College of Medicine, Columbus, Ohio, USA

Correspondence to:
Dr Mitchell A Medow
Section of General Internal Medicine, Boston University School of Medicine, Crosstown Center Building, 2nd Floor, 801 Massachusetts Avenue, Boston, MA 02118, USA;
mitchell.medow@bmc.org

Accepted 5 July 2011

Abstract

While decisions made according to Bayes' theorem are the academic normative standard, the theorem is rarely used explicitly in clinical practice. Yet the principles can be followed without intimidating mathematics. To do so, one can first categorise the prior-probability of the disease being tested for as very unlikely (less likely than 10%), unlikely (10–33%), uncertain (34–66%), likely (67–90%) or very likely (more likely than 90%). Usually, for disorders that are very unlikely or very likely, no further testing is needed. If the prior probability is unlikely, uncertain or likely, a test and a Bayesian-inspired update process incorporating the result can help. A positive result of a good test increases the probability of the disorder by one likelihood category (eg, from uncertain to likely) and a negative test decreases the probability by one category. If testing is needed to escape the extremes of likelihood (eg, a very unlikely but particularly dangerous condition or in the circumstance of population screening, or a very likely condition with a particularly noxious treatment), two tests may be needed to achieve. Negative results of tests with sensitivity ≥99% are sufficient to rule-out a diagnosis; positive results of tests with specificity ≥99% are sufficient to rule-in a diagnosis. This method overcomes some common heuristic errors: ignoring the base rate, probability adjustment errors and order effects. The simplicity of the method

was before the test was done (the pretest probability, also referred to as the prevalence or prior-probability), the test result (positive or negative) and the ability of the test to discriminate between those afflicted and not afflicted with the disease (test characteristics expressed as sensitivity and specificity, or likelihood ratios). A simple formula, Bayes' theorem, combines these elements to produce the post-test probability of the disease. A positive test increases confidence in a diagnosis, but usually does not indicate certainty. Whether this confidence exceeds a treatment (or action) threshold[9] remains a decision for the clinician and patient. Likewise, a negative test decreases confidence in a diagnosis, but rarely rules it out completely. It is up to those involved to decide if further action is warranted.

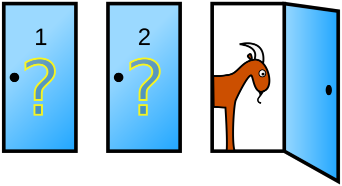
What if an easy, non-mathematical method to apply these concepts were available? Could the application of Bayes' theorem find its appropriate place in clinical practice and not be relegated to academic exercises for medical students and residents? Could its benefits in clinical practice finally be realised? There is a simple, qualitative or categorical application of Bayes' theorem that might ease the application of Bayes' underlying precepts. The method is based on categorising the pretest probability and handling a small set of probabilistic categories instead of the full spectrum of continuous probabilities, thus eliminating the need for mathematical calculations.

Time to revise and ask questions

- Sum: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cap B) = P(A|B)P(B)$, and only for independent events $P(A \cap B) = P(A)P(B)$
- Independence is not mutual exclusivity!

Monty Hall

“Let’s Make a Deal” show, hosted by Monty Hall, allowed to choose one of three doors. Behind one of the doors was a prize, behind the other two was goats. The contestant chooses one of the doors, but before opening it, Monty Hall always opens a door with a goat, and asks:



“Do you want to switch doors or stick to your original choice?”

How will you answer?