Bayes' Rule
Chapter 2

Kevin Lutz
PhD Candidate of Statistics
Wednesday, August 31, 2022

Department of Mathematical Sciences
The University of Texas at Dallas

Table of Contents

Review: Conditional and Joint Probability of A and B

The probability of $A \cap B$ can be written by one of the following:

$$\boxed{P(A \cap B) = P(A) \cdot P(B|A)}$$

$$\boxed{P(A \cap B) = P(B) \cdot P(A|B)}$$

- $P(A \cap B)$ is a joint probability.
- $P(A|B)$ and $P(B|A)$ are conditional probabilities.
- $P(A)$ and $P(B)$ are marginal probabilities.

Bayes' Rule

- Derived by combining the joint probability formulas:

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

Bayes' Rule

$$\boxed{P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}} \text{ or } \boxed{P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}}$$

- $P(A|B)$: The conditional probability of event A occurring given that B occurs.
- $P(B|A)$: The conditional probability of event B occurring given that A occurs.
- $P(A)$ and $P(B)$: probabilities of observing A and B independently of each other.

Background and Terminology

- Thomas Bayes (1702-1761)
- Usefulness: conditional probability under multiple conditions.
- Applications: Testing procedures for software, pregnancy tests, COVID tests, etc.
- Terminology for $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$:
  1. You need to know the conditional $P(B|A)$ to find $P(A|B)$.
  2. $P(A)$ is called the prior because $P(B|A)$ is conditional on already knowing A.
  3. The denominator term $P(B)$ is called the normalizing constant (explained later).

Example 1

In a particular pain clinic, 10% of patients are prescribed narcotic pain killers. Overall, 5% of the clinic's patients are addicted to narcotics (including pain killers and illegal substances). Out of all the people prescribed pain pills, 8% are addicts. If a patient is an addict, what is the probability that they will be prescribed pain pills? See handwritten solutions.

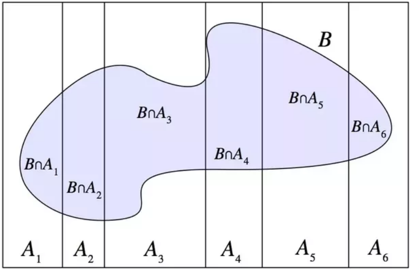
Example 2 - Your Turn

75% of the students at a particular school have a dog, and 30% have a cat. Given that 60% of those that have cat also have a dog, what percent of those that have a dog also have a cat?

Table of Contents

Law of Total Probability

- If events A_i are mutually exclusive and partition Ω, then



$$P(B) = P(B|A_1)\cdot P(A_1) + P(B|A_2)\cdot P(A_2) + \cdots + P(B|A_n)\cdot P(A_n)$$
$$= \sum_{i=1}^{n} P(B|A_i)P(A_i) \text{ or equivalently } \sum_{i=1}^{n} P(B\cap A_i)$$

Example 3

Under good weather conditions, 90% of flights arrive on time. During bad weather, only 20% of flights arrive on time. Tomorrow, the chance of good weather is 60%.

- What is the probability that your flight will arrive on time?
- What is the probability that your flight will not arrive on time?

See handwritten solutions.

Example 4 - Your Turn

A system may become infected by spyware through the internet or email. Spyware arrives via the internet 70% of the time; spyware arrives via email 30% of the time. If spyware enters via the internet, the system detects it with probability of 0.6; if spyware enters via email, it is detected with probability 0.8. What percentage of times is this spyware detected by the system?

Table of Contents

Bayes' Theorem (or Rule)

- Law of Total Probability:

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

- Let B be an event with non-zero probability, then the statement of Bayes' Theorem is:

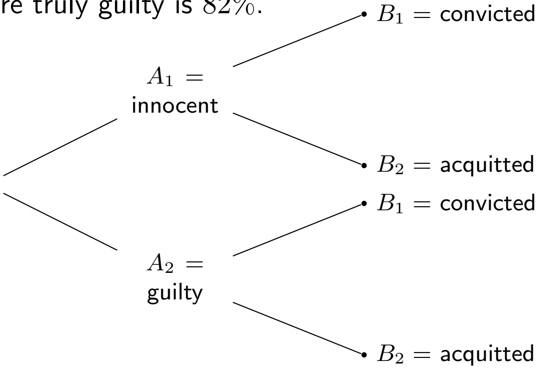
$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \boxed{\frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}}$$

or

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \boxed{\frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}}$$

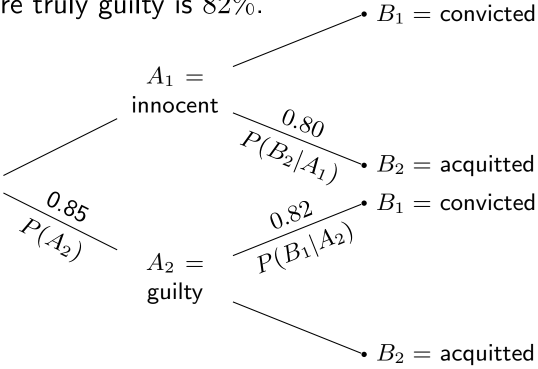
Example 5: Law and Order

- Suppose that 85% of all defendants are truly guilty.
- The probability that a defendant is acquitted given that they are truly innocent is 80%.
- The probability that a defendant is convicted given that they are truly guilty is 82%.



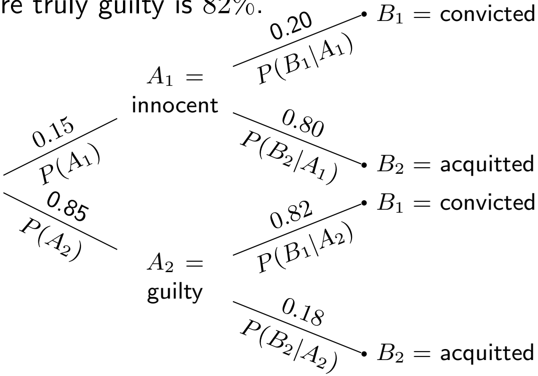
Example 5: Law and Order

- Suppose that 85% of all defendants are truly guilty.
- The probability that a defendant is acquitted given that they are truly innocent is 80%.
- The probability that a defendant is convicted given that they are truly guilty is 82%.

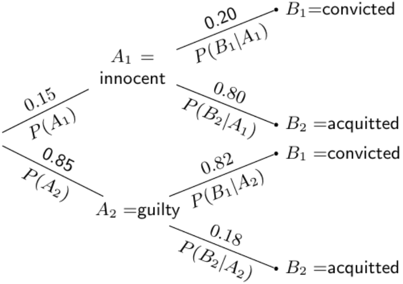


Example 5: Law and Order

- Suppose that 85% of all defendants are truly guilty.
- The probability that a defendant is acquitted given that they are truly innocent is 80%.
- The probability that a defendant is convicted given that they are truly guilty is 82%.



Example 5: Law and Order



1. What is the probability a defendant is innocent given that they were acquitted?
2. What is the probability a defendant is innocent given that they were convicted?
3. What is the probability a defendant is guilty given that they were convicted?

See handwritten solutions.

Example 6: Classification/Medical Testing

1. Prevalence: the probability of having a disease.

$$\implies P(\text{disease})$$

2. Sensitivity: the probability that a person tests positive when they actually have the disease (true positive TP)

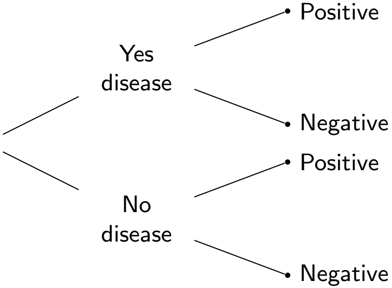
$$\implies P(\text{positive}|\text{have the disease})$$

3. Specificity: the probability that a person tests negative when they actually don't have the disease (true negative TN).

$$\implies P(\text{negative}|\text{does not have the disease})$$

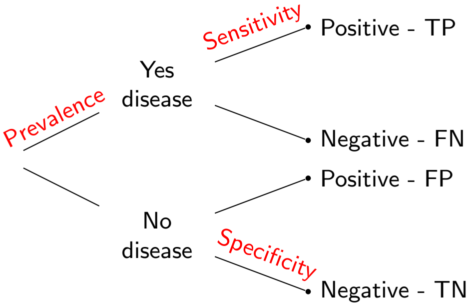
Example 6: Classification/Medical Testing

1. Prevalence: the probability of having a disease.
2. Sensitivity: the probability that a person tests positive when they actually have the disease (true positive TP)
3. Specificity: the probability that a person tests negative when they actually don't have the disease (true negative TN).



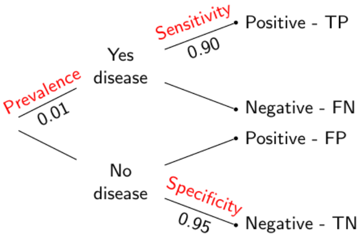
Example 6: Classification/Medical Testing

1. Prevalence: the probability of having a disease.
2. Sensitivity: the probability that a person tests positive when they actually have the disease (true positive TP)
3. Specificity: the probability that a person tests negative when they actually don't have the disease (true negative TN).



Example 6: Classification/Medical Testing

(a) A medical test has sensitivity 90% and specificity 95%. 1% of the population has the disease. What is the probability of actually having the disease given that your test result is positive?

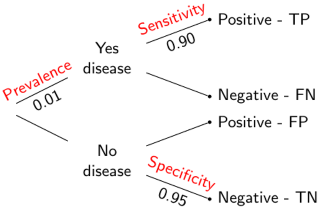


$P(\text{Yes disease}|\text{Positive result}):$

$$
\begin{aligned}
&= \frac{P(pos|yes)P(yes)}{P(pos|yes)P(yes)+P(pos|no)P(no)} \\
&= \frac{(0.90)(0.01)}{(0.90)(0.01)+(0.05)(0.99)} \\
&= 0.1538
\end{aligned}
$$

Example 6: Classification/Medical Testing

(b) A medical test has sensitivity 90% and specificity 95%. 1% of the population has the disease. What is the probability you don't have the disease if the test result is negative?



P(No disease|Negative result) :

$$\begin{aligned} &= \frac{P(Neg|No)P(No)}{P(Neg|Yes)P(Yes) + P(Neg|No)P(No)} \\ &= \frac{(0.95)(0.99)}{(0.10)(0.01) + (0.95)(0.99)} \\ &= 0.9989 \end{aligned}$$

Example 7: More than 2 conditions

A new computer programs consists of two modules. The first modules contains an error with probability of 0.2. An error in the first module alone causes the system to crash with probability 0.5. The second module is independent of the first module and contains an error with probability of 0.4. An error in the second module alone causes the system to crash with probability 0.8. If there are errors in both modules, the system crashes with probability 0.9. Suppose the program crashed. What is the probability of errors in both modules?

- Let $C =$ system crashes. This is the condition.
- Let $A =$ errors in first module $\implies P(A) = 0.2$
- $B =$ errors in second module $\implies P(B) = 0.4.$
- Let $A \cap B =$ errors in both modules $\implies P(A \cap B) = (0.2)(0.4) = 0.08$ since A and B are independent.
- What does "A alone" or "B alone" mean?

Example 7: More than 2 conditions

Ω is partitioned 4 ways into mutually exclusive events:

- "A alone" $\implies A \setminus B \implies P(A \setminus B) = P(A) - P(A \cap B)$
- "B alone" $\implies B \setminus A \implies P(B \setminus A) = P(B) - P(A \cap B)$
- "Both A and B" $\implies P(A \cap B)$
- "No errors" $\implies P(\overline{A \cup B})$

Location	Error	Crash
Both A and B	$P(A \cap B) = 0.08$	$P(C \mid A \cap B) = 0.9$
A alone	$P(A \setminus B) = 0.2 - 0.08 = 0.12$	$P(C \mid A \setminus B) = 0.5$
B alone	$P(B \setminus A) = 0.4 - 0.08 = 0.32$	$P(C \mid B \setminus A) = 0.8$
No errors	$P(\overline{A \cup B}) = 0.48$	$P(C \mid \overline{A \cup B}) = 0$

What is the probability of errors in both modules given that the program crashed? (see handwritten solutions for the answer)