Bayes' Theorem

1 Introduction

Bayes' Theorem (or Bayes' Rule) is a key theorem in Probability Theory, and has many practical applications, eg. medical treatment, educational assessment, and machine learning. It is important to learn it well.

conditioned probability.

Given a sample space S, and 2 events A, B, Bayes' Theorem states that:

$$P(B \mid A) = \frac{P(A \mid B)\, P(B)}{P(A)} = \frac{P(AB)}{P(A)} \tag{1}$$

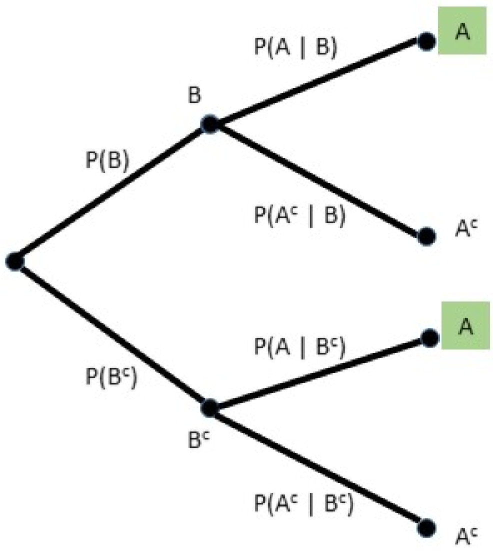
B^c, B': everything not in B

$$\text{Or:} \quad P(B \mid A) = \frac{P(A \mid B)\, P(B)}{P(A \mid B)\, P(B) + P(A \mid B^c)\, P(B^c)} \tag{2}$$

$$\text{because} \quad P(A) = \underbrace{P(A \mid B)\, P(B)}_{= P(AB)} + \underbrace{P(A \mid B^c)\, P(B^c)}_{P(AB^c)}. \tag{3}$$

utilize: $P(B)+P(B^c)=1$.

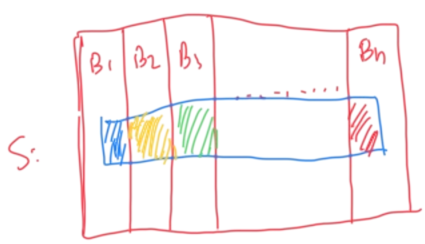
To see how Eq (3) is derived, refer to the possibility tree below, which shows how $P(A)$ is the sum of the probabilities of the green leaves.



We may generalize Bayes' Theorem by first noting that in the above, the sample space S is partitioned into 2 parts: B and B^c. If, instead, S were to be partitioned into n parts, $B_1, \ldots, B_n$, where $n \in \mathbb{Z}_{\geq 2}$, then Eq (2) becomes:

Generalized Bayes' Theorem

$$P(B_k \mid A) = \frac{P(A \mid B_k)\, P(B_k)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n)} \tag{4}$$



2 Example: Covid-19 testing

Suppose you take the Covid-19 test and the test result is positive. Does this mean you have contracted the disease? Maybe, or maybe not. Note that *every* medical test is not perfect but has errors. There are 2 kinds of errors, called *False Negative* and *False Positive*.

- False Negative: This error occurs when the test result is negative (meaning that you don't have the disease) even though you actually have it. This means you are considered healthy and won't be treated, which may lead to the disease worsening, or being spread to other people.
- False Positive: This error occurs when the test result is positive (meaning that you have the disease) but you actually don't. As a result, you may be given unnecessary medical treatment, incurring side effects, money, and physical and emotional pain.

Clearly, both errors are undesirable, and should be made as low as possible. The False Negative Rate (FNR), and False Positive Rate (FPR) measure how frequent such errors occur, respectively. For Covid-19, the FNR is 9.3%, according to [3], while the FPR varies between 0.8% and 4.0%, according to [1]. How do these 2 types of errors affect how we interpret the test result? Bayes' Theorem provides the answer.

Define T to be the event that the "test result is positive", and H to be the event "have contracted Covid-19". We seek to determine the conditional probability: $P(H \mid T)$. Using Bayes' Theorem ,

$$P(H \mid T) = \frac{P(T \mid H)\, P(H)}{P(T \mid H)\, P(H) + P(T \mid H^c)\, P(H^c)} \tag{5}$$

And so we need to determine all the probabilities on the right hand side of the above equation.

- Now, $P(T \mid H)$ is the probability that the test result is positive when the patient actually has Covid-19. By definition, $FNR = P(T^c \mid H)$, and thus $P(T \mid H) = 1 - FNR = 0.907$.
- Likewise, $P(T \mid H^c) = FPR$ by the definition of FPR. Let's take the upper limit for FPR, ie. $FPR = 0.04$.
- $P(H)$ is the probability of contracting Covid-19. One estimate of this is how likely one comes into contact with a Covid-19 patient in the general public. According to [2], the number of active cases of Covid-19 in Singapore on 5^{th} April 2021 is 251. And the population of Singapore is $5.685M$ in June 2020 (which is the most current figure, according to [4]). Hence, we may estimate $P(H) = 251/(5.685 \times 10^6) = 4.42 \times 10^{-5}$. This means $P(H^c) = 1 - P(H) = 0.9999558$.

Substituting all these numbers into Eq (5) gives: $P(H \mid T) \approx 0.001$, ie. about 0.1%. This means that you are *very unlikely* to have Covid-19, despite the positive test result!

References

[1] S. E., N. V., and D. F. False-positive covid-19 results: hidden problems and costs. *The LANCET*, 8(12), 2020.

[2] Gov.Sg. Updates on the covid-19 situation in singapore. `https://www.gov.sg/features/covid-19`, Apr 2021.

[3] K. J.N., Z. N., and C. e. a. MacDonald. False negative rate of covid-19 pcr testing: a discordant testing analysis. *Virology Journal*, 18(13), 2021.

[4] D. of Statistics Singapore. Population and population structure. `https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data`, Sep 2020.