ORF 418 : OPTIMAL LEARNING

LECTURE 7 : September 24, 2025

BAYES FORMULA

1) Bayes formula for events
2) Example: Testing
3) Bayes formula for partitions.
4) A legal example.

Homework 1 is due

this Sunday 28$^{\underline{th}}$ before midnight

No late homeworks!

No homework will be dropped!

I. BAYES FORMULA FOR EVENTS

Consider a probability space Ω and probability $\mathbb{P}$ defined on the subsets (called events) of Ω.

Let A and B be two events. Then, the conditional probability of A given B is defined as

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \qquad (*)$$

Conditional probability of B given A is also given by

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

Hence,

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

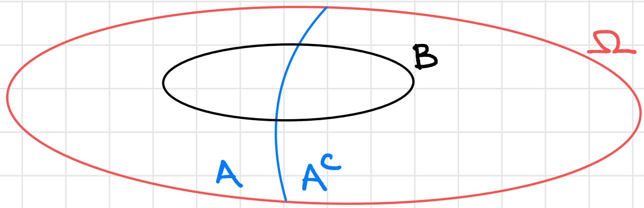
Using these we obtain:

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$$
$$= \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c).$$

Substitute above into $(*)$ yields

$$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}.}$$

Above is the Bayes Formula for events. It states that if we know the probabilities of the events $A \cap B, A, A^c$ and also the conditional probabilities of $B|A$ and $B|A^c$, then we can compute the "opposite" conditional probability of $A|B$. Through examples we will see that this is a very powerful tool. Pictorially:



II: Example: Testing

Suppose we perform a medical test to determine whether someone is infected by a certain desease, say covid. The following two events are relevant:

A = person is infected,

B = test is positive.

It is natural to assume that the probability of infection $\mathbb{P}(A)$ is known. Also by test design the conditional probabilities $\mathbb{P}(B|A)$ and $\mathbb{P}(B|A^c)$

are also known. Suppose that

probability of infection $= \mathbb{P}(A) = 0.03$.

That is, 3% of the population is infected. Also suppose

probability of correct diagnosis $= \mathbb{P}(B|A) = 0.99$,

probability of false positive $= \mathbb{P}(B|A^c) = 0.1$.

In all real tests, the false positive probability is small but never zero, and making it very small is costly.

Now suppose that someone is tested positive. Then, what is the probability that this person is actually infected. Mathematically, we want to compute $\mathbb{P}(A|B)$. This is the classical application of the Bayes Formula:

$$\mathbb{P}(A|B) = \frac{(0.99)(0.03)}{(0.99)(0.03) + (0.1)(0.97)} = 0.234.$$

Note that with a very accurate test (i.e. $\mathbb{P}(B|A) = 0.99$) a positive test implies only a 23% probability of being infected. So the probability increased from 3% to 23%; a substantial increase but far from certainty.

Some terminology: 3% is the prior probability of infection, while 23% is the posterior probability. Hence, the prior is updated after the observation.

<u>Second testing</u>: If one tested positive then the probability of infection is 23%. If this person is tested positive once again, then

$$\mathbb{P}(A \mid \text{two positive tests}) = \frac{(0.99)(0.234)}{(0.99)(0.234) + (0.1)(0.766)}$$

$$= 0.7515$$

Similarey,

$$\mathbb{P}(A \mid 3 \text{ positive tests}) = \frac{(0.99)(0.7515)}{(0.99)(0.7515) + (0.1)(0.2484)}$$

$$= 0.9677$$

III: Bayes Formula for partitions

Suppose that $\{A_1, \dots, A_N\}$ is a partition of Ω:

(1) $A_i \cap A_j = \emptyset$, they are disjoint;

(2) $\bigcup_{i=1}^{N} A_i = \Omega$.

Then, for any $i = 1, \dots, N$,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\,\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B|A_j)\,\mathbb{P}(A_j)}$$

This formula is used in the following way. Suppose we have a prior distribution of $\mathbb{P}(A_1) \dots \mathbb{P}(A_N)$.

Then an observation, event B, occurs. We use this observation to update our prior and obtain the posterior distribution $\mathbb{P}(A_1|B) \dots \mathbb{P}(A_N|B)$.

Since we observed B, our "world" Ω shrinks to B.

IV. Example (People vs Collins, 1968)

A couple with certain characteristics was involved in a purse scratching. It is estimated that probability of a couple having these characteristics is 8.3×10^{-8} or one in 12 million. Few days later, a couple with same characteristics was arrested. Only evidence is these characteristics. They were convicted but Supreme Court ruled it over.

How do we formulate it mathematically?

1) important quantity is this: how likely that there is no other couple with these characteristics.

2) If this probability is low, then conviction is justified.

Let's define

B = the event there is at least one couple with these characteristics.

A = there is at least two.

We are told that any couple having these characteristics is

$$p = \frac{1}{12 \times 10^6} = \text{one in 12 million.}$$

The number of couples in the area is $n = 8$ million

Then,

$$\mathbb{P}(B) = 1 - \mathbb{P}(\text{no such couple}) = 1 - (1-p)^n$$

$$\mathbb{P}(A) = \mathbb{P}(B) - \mathbb{P}(\text{exactly one}) = 1 - (1-p)^n - np(1-p)^{n-1}$$

We calculate that

$\mathbb{P}(B) = 0.487$ $\quad \mathbb{P}(A) = 0.144$ $\quad \mathbb{P}(A|B) = 0.297$.