Lecture 3: Conditional Probability, Independence & Bayes Theorem

Sec 1: Conditional Probability

In this lecture we shall introduce the reader to a very important notion of probability theory called "Conditional Probability". Mathematically stated it means the following: Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and events $A, B \in \mathcal{F}$, we ask the question as to what is the ~~probability~~ probability of the event A given the fact or conditioned on the fact that the event B has already occured"?

We ask such questions pretty often asked in real-life scenarios. For example given that a train has departed from the originating station with a ten minute delay, and we may be interested to know the probability that it will reach 'its' destination on time.

The probability of an event A, given that the event B has occured is called the conditional probability of A given that B, and is denoted by $P(A|B)$.

**Definition 3.1**: Conditional Probability: Given $(\Omega, \mathcal{F}, \mathbb{P})$ to be a probability space, and $\mathbb{P}(B) > 0$; where $B \in \mathcal{F}$. Then the conditional probability $P(A|B)$, for any $A \in \mathcal{F}$ is given as

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

- Let us view this definition from the classical angle. Let $\Omega$ be a sample space, with N possible out comes. The ~~$n_B$ number~~ that some event B has occured me means $n_B$ that $\omega \in B$ has occured. Now what is the probability that A will occur. So now our total number of possibilities have shrunk to only those which represent the event B. So among those ~~events~~ possibilities we need to what is proportion that is common to ~~A~~ the event A, i.e. the sample points which represent the occurence $A \cap B$

$$\therefore P(A|B) = \frac{n_{A\cap B}}{n_B} = \frac{\frac{n_{A\cap B}}{N}}{\frac{n_B}{N}} = \frac{P(A\cap B)}{P(B)}, \text{ where } P(B) > 0.$$

**Example 3.1** Let us consider an urn having 5 red & 5 black balls. Find the probability that a red ball appears on the 2nd draw given that the first draw resulted in a black ball. The first ball is not replaced back in the urn.

Let $B$ be the event that the first ball drawn was black. Let $A$ be the event that second ball drawn is red.

$$\therefore \quad P(A|B) = \frac{5}{9}$$

How did we come to this value. We have already drawn a black ball. So before making the second draw we have only 9 balls left so that is exactly our ~~total~~ possible outcomes and we have 5 balls are red, which gives the above value.

Observe that from the definition of ~~P(B|A) as~~ $P(A|B)$ we have

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

So in Example 3.1 we have

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

Here $P(B) = \frac{5}{10} = \frac{1}{2}$ $\quad \therefore \boxed{P(A \cap B) = \frac{5}{9} \times \frac{1}{2} = \frac{5}{18}}$

Of course $A \cap B$ is the event that a black ball appears in the first draw and a red in the second.

Given any $B \in \mathcal{F}$ with $P(B) > 0$, we define the conditional probability function given $B$ as $P(\cdot\,|B) : \mathcal{F} \to \cancel{[0,1]}\,\mathbb{R}$, i.e.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \forall A \in \mathcal{F}.$$

Now let us see if the axioms of Kolmogorov are satisfied. For any $A \in \mathcal{F}$, it is obvious that $P(A|B) \geq 0$. Further as $A \cap B \subset B$, $\Rightarrow P(A \cap B) \leq P(B)$. Thus $P(A|B) \leq 1$.

Now if $A = \Omega$ we have

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Let $\{A_n\}$ be a sequence of mutually ~~exclusive~~ disjoint events, ~~i.e~~

then
$$P\left(\bigcup_{n=1}^{\infty} A_n \mid B\right) = \frac{P\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right)}{P(B)}$$

$$= \frac{P\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right)}{P(B)}$$

As $\biguplus \{A_n \cap B\}$, forms a ~~dis~~ mutually disjoint sequence of events. ~~Such that~~ Hence

$$P\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right) = \sum_{n=1}^{\infty} P(A_n \cap B)$$

$$\therefore \quad P\left(\bigcup_{n=1}^{\infty} A_n \mid B\right) = \sum_{n=1}^{\infty} \frac{P(A_n \cap B)}{P(B)}$$

$$= \sum_{n=1}^{\infty} P(A_n \mid B).$$

Thus all the three laws of Kolmogorov holds. Thus we have shown that for any given event B, with $P(B)>0$, $P(\cdot \mid B)$ is a probability measure or probability function.

We shall now state an important result called the "Theorem of Total Probabilities" which will lead to do computations with the famous Bayes Theorem, which we will discuss in the next section.

## Theorem of Total Probability

**Theorem 3·1·1 :** Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{B_1 \ldots B_n\}$ be a collection of $n$ mutually disjoint collection of events such that $\Omega = \bigcup_{i=1}^{n} B_i$. Further assume that for each $i \in \{1,2,\ldots n\}$, $P(B_i)>0$.

Then for any $A \in \mathcal{F}$

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

Proof: Note that since $\{B_1 \ldots B_n\}$ forms a mutually disjoint partition of $\Omega$, then A either occurs jointly with $B_1$ or $B_2$ or .... or $B_n$, i.e.

$$A = \bigcup_{i=1}^{n} (A \cap B_i)$$

while we know that $(A \cap B_1) \ldots (A \cap B_n)$ are mutually disjoint. Then

$$P(A) = P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right)$$

$$= \sum_{i=1}^{n} P(A \cap B_i) \quad \left[ \because A \cap B_i,\ i=1 \ldots n \text{ are mutually exclusive} \right]$$

By using exclusion-inclusion principle.

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

This theorem remains true if $n = \infty$ Think how.

Section: Bayes Theorem

Thomas Bayes was a priest, an English priest. ~~He is~~ The theorem that now bears his name, was published after his notes were edited by Robert Price. Thomas Bayes had not published his result. The idea of Bayes later led to the field of Bayesian statistics which has huge applications in modern medicine and machine learning, for example.

Bayes Theorem Let us consider the probability space $(\Omega, \mathcal{F}, P)$. Let $\{B_1 \ldots B_n\}$ be a finite sequence of events such that $P(B_i) > 0$, $\forall\, i=1 \ldots n$ & $\Omega = \bigcup_{i=1}^{n} B_i$. Let $A \in \mathcal{F}$ such that $P(A) > 0$, then

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

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

This theorem also holds for $n=\infty$. Now we shall provide the simple proof.

$$P(B_i \mid A) = \frac{P(B_i \cap A)}{P(A)}$$

$$P(B_i \cap A) = \cancel{P(A)}\, P(B_i)\, P(A \mid B_i)$$

while by the theorem of total probability

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A \mid B_i).$$

A more simple version of the Bayes Theorem can be provided with two events $A$ & $B \in \mathcal{F}$, with $P(A) > 0$, $P(B) > 0$.

Note that if $B \in \mathcal{F} \Rightarrow \cancel{B \in \mathcal{F}}\ B^c \in \mathcal{F}$ and we know that

$$\Omega = B \cup B^c$$

$$\therefore \quad P(A) = P(A \cap B) + P(A \cap B^c)$$

$$= P(B) P(A \mid B) + P(B^c) P(A \mid B^c)$$

$$\therefore \quad \boxed{P(B \mid A) = \frac{P(B)\, \cancel{P(B \mid A)}\, P(A \mid B)}{P(B) P(A \mid B) + P(B^c) P(A \mid B^c)}}$$

What the Bayes theorem does is as follows. It can help us answering the following questions. For example in an urn there $n_W$ white balls and $n_B$ black balls. Suppose balls are drawn at random without replacement. Suppose we draw the balls twice and suppose that the second ball is white: Then we can ask the question as to what is the probability that the first ball drawn was white or the first ball drawn is black. Bayes theorem helps us in answering such questions.

If $B$ be the event that the first ball drawn is black then $B^c$ is naturally the event that the first ball drawn is white. Let $W$ be the event that the second ball drawn is white. So

$$P(B \mid W) = \frac{P(B) P(W \mid B)}{P(W)} \quad \text{(By Bayes Theorem).}$$

Now $P(B) = \dfrac{n_B}{\cancel{n_B}+n_W}$

$$P(W|B) = \frac{n_W}{\cancel{n_W+n_B-1}} \cdot \frac{n_W}{n_B+n_W-1}$$

While

$$P(W) = P(W\cap B) + P(W \cap B^c)$$

$$= P(B)\,P(W|B) + P(B^c)\,P(W|B^c)$$

$$= \cancel{\frac{n_B}{(n_B+n_B)} \cdot n_W}$$

$$= \frac{n_B}{n_B+n_W} \cdot \frac{n_W}{n_B+n_W-1} + \frac{n_W}{n_B+n_W} \cdot \frac{n_W-1}{n_B+n_W-1}$$

$$= \frac{n_B n_W + n_W(n_W-1)}{(n_B+n_W)(n_B+n_W-1)}$$

$$= \frac{n_B n_W + n_W^2 - n_W}{(n_B+n_W)(n_B+n_W-1)}$$

$$\therefore P(B|W) = \frac{\dfrac{n_B}{n_B+n_W} \cdot \dfrac{n_W}{n_B+n_W-1}}{\dfrac{n_B n_W + n_W^2 - n_W}{(n_B+n_W)(n_B+n_W-1)}}$$

$$\therefore \quad P(B|W) = \frac{n_B n_W}{n_B n_W + n_W^2 - n_W}$$

Since $n_W \in \mathbb{N}$, i.e. $n_W \geq 1 \Rightarrow n_W^2 \geq n_W \Rightarrow n_W^2 - n_W \geq 0$

$\Rightarrow n_B n_W + n_W^2 - n_W \geq n_B n_W$

$\Rightarrow P(B|W) \leq 1$, and thus fulfilling the basic requirement of the probability function.

$$\boxed{P(B|W) = \frac{n_B n_W}{n_B n_W + n_W^2 - n_W}}$$

- We urge the reader to compute $P(B^c|A) = ??$

## Sec 3

## Independence of Events

I do not even remember which year but I remember reading a book called "The Enigmas of Chance", by the famous mathematician Mark Kac. Where he discussed an important idea called "Independence of Events" i.e. two events A and B, are called independent if and only if

$$P(A \mid B) = P(A), \quad \text{if } P(B)>0$$

$$P(B \mid A) = P(B), \quad \text{if } P(A)>0$$

i.e. $\boxed{P(A \cap B) = P(A).P(B)} \longrightarrow$ The key fact

This idea as I read in the book was due to Hugo Stienhaus, who developed this notion, when he was in hiding during the second world war. Stienhaus was a famous Polish mathematician, who made major contributions to an area of mathematics.

Let us see how this can be applied. Consider the simple situation where you toss a coin twice. What is the probability that head (H) appears in the second toss for the first time

Let A be the event that head appears in the second toss for the first time, and B be the event that tail appears. Thus we are trying to find the probability the configuration HT. We know that $P(HT) = \frac{1}{4}$ from direct computation. But

$HT = B \cap A$

$$\therefore P(HT) = P(B \cap A) = \frac{1}{4}$$

$$\therefore P(B \cap A) = \frac{1}{2} \cdot \frac{1}{2} = P(B).P(A)$$

Thus B & A are independent events and which appears to be intuitively true.

Let us now introduce the notion of independence of three events $A, B, C \in \mathcal{F}$. The definition of independence now means the following: Three events A, B & C are independent events if and only if

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

It is important that pairwise independence of events do not mean they are independent. The following example from the book [1] by Moods, Graybill and Boes, ~~shows~~ establishes what we have just mentioned above.

Example: Pairwise independence does not mean independence.

Let

$A_1$: Event that odd face appears in the first die

$A_2$: Event that odd face appear in the second die.

$A_3$: The sum of the two numbers in a random throw of two die is odd.

$$P(A_1) = \frac{3}{6} = \frac{1}{2} \quad \& \quad P(A_2) = \frac{3}{6} = \frac{1}{2}$$

$$P(A_1 \cap A_2) = \frac{9}{36} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) P(A_2).$$

We urge the reader to show that

$$P(A_1 \cap A_3) = P(A_1) P(A_3)$$
$$P(A_2 \cap A_3) = P(A_2) P(A_3)$$

$$P(A_3) = \frac{18}{36} = \frac{1}{2} \Big( A_3 = \{(1,2),(2,1),(1,4),(4,1),(1,6)(6,1),(2,3),(3,2)$$
$$(5,6),(6,5),(3,6),(6,3),(4,3),(3,4)$$
$$(5,4),(4,5),(2,5),(5,2)\} \Big)$$

$$P(A_1 \cap A_3) = \frac{9}{36} = \frac{1}{4} \quad \text{(Finding the cases in } A_3 \text{ with odd first term)}$$
$$= \frac{1}{2} \cdot \frac{1}{2} = P(A_1) . P(A_3)$$

We leave $P(A_2 \cap A_3) = P(A_2) P(A_3)$ verification to the reader. It is symmetric.

$$P(A_1 \cap A_2 \cap A_3) = 0.$$

But $P(A_1) \cap P(A_2) \cap P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

$$\therefore \boxed{P(A_1 \cap A_2 \cap A_3) \neq P(A_1) P(A_2) P(A_3)}$$

—x—