Lecture 2: Kolmogorov's Axioms for Probability

We have seen that though for finite sample spaces the idea of defining probability through counting works pretty well, we run into rough weather the moment we try to extend such an idea to an infinite sample space setting.
It was the great mathematicians, Emile Borel, Henri Lebesgue, Felix Hausdorff and Cantelli who realized that at a much deeper level the notion of probability is intimately linked with notion of measure of a set and thus notion of probability needed a fresh mathematization, which finally ~~boi~~ completed by Kolmogorov.
Borel and his co-workers gained a tremendous insight into the nature of probability by considering questions about infinite coin tosses and thus each instance can be represented as a point in the interval [0,1]. This finally led to the strong law of large numbers which we will mention later in the course. Roughly what happens is the following. When we make an arbitrarily large number of trials the proportion of ~~half is~~ heads is nearly $1/2$ and remains there forever with probability one.
Borel and his co-workers understood the fact that every subset of an infinite sample space may not be an event. It was believed that events must have the following property

i) Ω & ϕ must be events (sure event & null event must be there)

ii) If A is an event then so is A^c.

iii) If $A_1, A_2 \dots A_n \dots$ be a countable sequence of events then so is $\bigcup_{i=1}^{\infty} A_i$.

The class of such subsets of Ω is often denoted by $\mathcal{F}$ and $\mathcal{F}$ is often referred to as the σ-algebra or σ-fields of events.

One might be worried that we spoke of a countably infinite sequence of events but did not say anything about finite number events. But it is simple to show that if $A_1 \ldots A_k$ are a finite number of events then $\bigcap_{i=1}^{k} A_i$ is also an event, i.e. a member of $\mathcal{F}$. Set $A_{k+1} = \phi$, $A_{k+2} = \phi$ . . . .

Since by i) ϕ is an event, we have

$$\bigcup_{i=1}^{k} A_i = A_1 \cup A_2 \cup \ldots \cup A_k \cup \phi \cup \phi \cup \phi \cup \ldots \ldots \cup \phi \ldots$$

is also an event.

Further if $A_1 \ldots \ldots A_n \ldots$ is a countable sequence of events then $\bigcap_{i=1}^{\infty} A_i$ is also an event (Try this out as an assignment).

Kolmogorov assumed or rather viewed that given the pair $(\Omega, \mathcal{F})$, we can define a function $P : \mathcal{F} \to \cancel{[0,1]}\ \mathbb{R}$ (A set function actually) which satisfies the following rules

i) For any $A \in \mathcal{F}$, $0 \leq P(A) \leq 1$ (This is clear from defn of P)

ii) $P(\Omega) = 1$

iii) If $\{A_n\}_{n=1}^{\infty}$ is a sequence of mutually exclusive events (i.e. $A_i \cap A_j = \phi$, if $i \neq j$), then

$$P\left(\bigcup_{n=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

~~The third rule or axi~~ The above rules are often called Kolmogorov's axioms of probability. The third rule is a central one and provides us with some important facts about the function P which is called the probability function or probability measure

Our first result which is a consequence of the axiom is as follows,

$$\boxed{P(\emptyset) = 0}$$

One might wonder why this was not mentioned in the axiom. In fact $P(\emptyset) = 0$ is a consequence of the third axiom, since if we set $A_i = \phi$, $\forall i \in \mathbb{N}$, (Here $\mathbb{N}$ denotes the set of natural numbers), Then ~~A_0~~ $\{A_n\}_{n=1}^{\infty}$ is a mutually disjoint sequence of sets, and iii) gives us

$$P(\emptyset) = \sum_{n=1}^{\infty} P(\emptyset)$$

$$P(\emptyset) = P(\emptyset) + P(\emptyset) + \dots + P(\emptyset) + \dots\dots$$

This equation is only possible if $P(\emptyset) = 0$ which is what we already know from the classical approach, but deriving this fact using Kolmogorov's axioms does not put any restriction on the sample space Ω.

Our next result is the following, which is again a consequence of the third rule and the fact that $P(\emptyset) = 0$. Let $A_1 A_2 \dots A_k$ be k, mutually disjoint events, then

$$P\left(\bigcup_{n=1}^{k} A_n\right) = \sum_{n=1}^{k} P(A_k).$$

The strategy is the same; Set ~~$\cup A_k$~~ $A_{k+1} = \phi$, $A_{k+2} = \phi \dots\dots$

Thus

$$\bigcup_{n=1}^{k} A_n = A_1 \cup A_2 \cup A_3 \dots \cup A_k \cup \emptyset \cup \emptyset \dots\dots \cup \emptyset$$

Thus using (iii) we have

$$P\left(\bigcup_{n=1}^{k} A_n\right) = \sum_{n=1}^{k} P(A_n) + \sum_{n=k+1}^{\infty} P(\emptyset)$$

$$= \sum_{n=1}^{k} P(A_n) \quad (\because P(\emptyset) = 0).$$

Thus we have that if $A_1 \cdots A_k$ are mutually disjoint events, then

$$\boxed{P\left(\bigcup_{n=1}^{k} A_n\right) = \sum_{n=1}^{k} P(A_n).} \longrightarrow (\#)$$

Now as $A \cup A^c = \Omega$, using (#) we have

$$P(\Omega) = P(A \cup A^c)$$
$$= P(A) + P(A^c) \text{ by } (\#)$$

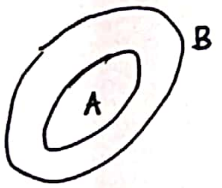
$\therefore$ $P(A) + P(A^c) = 1$ (since $P(\Omega) = 1$, from second axiom)

Hence irrespective of the nature of Ω we have

$$\boxed{P(A^c) = 1 - P(A)}$$

We will now derive some more results about the probability function based on the Kolmogorov's axioms.

- Let $A \subseteq B$, then $P(A) \leq P(B)$.



In this case we have

$$B = A \cup (B \setminus A)$$

$$\therefore \quad P(B) = P(A) + P(B \setminus A), \quad (\text{Using } (\#))$$

$$\therefore \quad P(B) - P(A) = P(B \setminus A)$$

But by i) $P(B \setminus A) \geq 0 \Rightarrow P(B) \geq P(A)$.

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, for any $A \,\&\, B \in \mathcal{F}$.

Now $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$

We observe that above three sets on the left are mutually exclusive.

Hence using (#) we have

$$P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) \longrightarrow (*)$$

Again

$$A = (A \setminus B) \cup (A \cap B).$$

But as $(A \setminus B)$ and $(A \cap B)$ are disjoint, i.e. mutually exclusive we have using (#)

$$P(A) = P(A \setminus B) + P(A \cap B)$$

$$\therefore \quad P(A \setminus B) = P(A) - P(A \cap B)$$

Further

$$B = (B \setminus A) \cup (A \cap B)$$

and by similar arguments as before

$$P(B) = P(B \setminus A) + P(A \cap B)$$

$$\therefore \quad P(B \setminus A) = P(B) - P(A \cap B).$$

Putting these expressions back in (*) we have

$$P(A \cup B) = P(A) + P(B) - 2P(A \cap B) + P(A \cap B)$$

$$\therefore \boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)} \longrightarrow (++)$$

We can immediately conclude that

$$\boxed{P(A \cup B) \leq P(A) + P(B)} \quad \text{since } P(A \cap B) \geq 0$$

We can extend this using mathematical ~~way~~ induction to the following result. If $A_1 \ldots A_k$ are ~~are~~ elements of $\mathcal{F}$ then

$$\boxed{P\left(\bigcup_{n=1}^{k} A_n\right) \leq \sum_{n=1}^{k} P(A_n)}$$

This is called Boole's Inequality.

The equality (++) can be easily generalized for more than two events and is given as:-

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k) - \sum_{k_1<k_2}^{n} P(A_{k_1} \cap A_{k_2})$$

$$+ \sum_{k_1<k_2<k_3}^{n} P(A_{k_1} \cap A_{k_2} \cap A_{k_3})$$

$$- \sum_{k_1<k_2<k_3<k_4}^{n} P(A_{k_1} \cap A_{k_2} \cap A_{k_3}) + \cdots$$

$$\cdots + (-1)^{n+1} P\left(\bigcap_{k=1}^{n} A_n\right)$$

This is sometimes referred to ~~the~~ as the "Principal of Inclusion and Exclusion".

Since the Kolmogorov's axiom involves, infinite sequences of events let us ~~as~~ look into some properties of such sequences.

- Consider the set of non-decreasing events; ie $A_n \subset A_{n+1}$. The set $A = \bigcup_{n=1}^{\infty} A_n$ is then called the limit of $\{A_n\}$
- Similarly consider the sequence of non-increasing events, i.e. $A_{n+1} \subset A_n$. Then $A = \bigcap_{n=1}^{\infty} A_n$ is call the limit of $\{A_n\}$

Let us look at the following result, which is important enough to be mentioned as a theorem

<u>Theorem 2·1:</u> Consider a sequence $\{A_n\}$ of non-decreasing event. Then

$$\lim_{n \to \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

Proof: Let us set $A = \bigcup_{n=1}^{\infty} A_n$. We can write A as follows

$$A = A_n + \bigcup_{j=n}^{\infty} (A_{j+1} - A_j)$$

Again using the third axiom of Kolmogorov, (Think why??) we have

$$P(A) = P(A_n) + \sum_{j=n}^{\infty} P(A_{j+1} - A_j)$$

Note that

$$0 \leq \sum_{j=1}^{\infty} P(A_{j+1} - A_j) \leq 1$$

Thus $\sum_{j=1}^{\infty} P(A_{j+1} - A_j)$ is a convergent series of non-negative numbers. ~~that~~ Thus

$$\lim_{n \to \infty} \sum_{j=n}^{\infty} P(A_{j+1} - A_j) = 0$$

Hence $n \to \infty$

$$P(A) = \lim_{n \to \infty} P(A_n) + \lim_{n \to \infty} \sum_{j=n}^{\infty} P(A_{j+1} - A_j)$$

$$\therefore \boxed{\lim_{n \to \infty} P(A_n) = P(A)}$$

— x —