PROBABILITY FROM THE MEASURE THEORETIC VIEWPOINT

ARJUN AGARWAL

1. Introduction

Probability theory is concerned with assigning probabilities to events. These events can be viewed as sets, and the probability can be viewed as a function which takes a set from a sample space at input and assign a real number in the closed interval $[0, 1]$. Probability functions can therefore be interpreted as a measure. $\sigma-$algebras and measure theory can provide mathematical rigor to the study of probability. In probability theory, zero-one laws are laws that assign the probability of an event to be either 0 or 1. Kolmogorov's zero-one law is one such law relating to the tail field of sequences of independent events. In this expository paper, I review some basic basic theory and definitions, followed by proof of the Kolmogorov's zero-one law using the tools of measure theory.

2. Preliminary Definitions

This section will review some basic definitions.

Definition 2.1. Given a set Ω, a $\sigma-$algebra $\mathcal{A}$ is a collection of subsets of Ω satisfying

- $\Omega, \phi \in \mathcal{A}$
- If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$
- If $A_1, A_2, \ldots, A_n$ is a countable collection of elements of $\mathcal{A}$ then

$$\bigcup_{i=1}^{n} A_i \in \mathcal{A} \text{ and } \bigcap_{i=1}^{n} A_i \in \mathcal{A}$$

Therefore, $\mathcal{A}$ is closed under countable unions and intersections.

Definition 2.2. Given a set Ω and a σ-algebra on Ω, a measure μ on $(\Omega, \mathbb{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ satisfying

- $\mu(\phi) = 0 If \mathrm{A}_1, A_2, \ldots$ is a finite or countable collection of pairwise disjoint elements of $\mathcal{A}$ then

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

Definition 2.3. Let A be a collection of subsets of Ω. The smallest σ-algebra containing A, denoted as $\sigma(A)$, is defined as the intersection of all σ-algebras containing A. Since it is the intersection of σ-algebra's, $\sigma(A)$ is also a σ-algebra. $\sigma(A)$ is also called the σ-algebra generated by A. $\sigma(A)$ is also unique. Let $A_1 = \sigma(A)$ and $A_2 = \sigma(A)$. A_1 is the intersection of σ-algebra's containing A and A_2 contains A which implies that $A_1 \subseteq A_2$. Similarly, we can show $A_2 \subseteq A_1$. This implies that $A_1 = A_2$ and $\sigma(A)$ is unique.

With the above definitions, we are ready to define a probability space.

Definition 2.4. A probability space is a triple $(\Omega, \mathcal{A}, \mathbf{P})$ where

- Ω is the sample space
- $\mathcal{A}$ is a σ-algebra of Ω
- $\mathbf{P}$ is a measure $\mathbf{P} : \mathcal{A} \to [0, 1]$ satisfying $\mathbf{P}(\Omega) = 1$

For example, for a simple case when we were studying the roll of a single die, then we can define the probability space with $\Omega = \{1, 2, 3, 4, 5, 6\}$, a σ-algebra $\mathcal{A} := \{A : A \subseteq \Omega\}$ and $\mathbf{P}$ defined as $\mathbf{P}(A) = \dfrac{\#A}{6}$. The formalism of σ-algebra and measure theory is needed in order to prevent additivity of uncountable sets. For example, consider $\Omega = [0, 1]$ and $\mathbf{P}$ be the Lebesgue measure on Ω. If addivity of uncountable sets was allowed then

$$\mathbf{P}(\Omega) = \mathbf{P}\left(\bigcup_{x\in[0,1]} \{x\}\right) = \sum_{x\in[0,1]} \mathbf{P}(\{x\}) = 0$$

which is an absurd result. Defining probability within the formalism of σ-algebra and measure theory helps avoid such results. We will now define independent events.

Definition 2.5. Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. A collection of events, $X_1, X_2, \ldots,$ possibly infinite, is independent if for all $n \in \mathbb{N}$, and all possible finite combinations $X_{m_1}, X_{m_2}, \ldots, X_{m_n}$ we have

$$\mathbf{P}(X_{m_1} \cap X_{m_2} \cap \cdots \cap X_{m_n}) = \prod_{i=1}^{n} \mathbf{P}(X_{m_i})$$

It is not enough for the events to be pairwise independent. As seen in the definition, every possible finite sub collection needs to be independent for the entire collection to be independent.

3. Kolmogorov's Zero-One Law

Definition 3.1. Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. Consider infinitely many events $A_1, A_2, \ldots \in \mathcal{A}$. The event, $\{A_n \text{ i.o.}\}$ referred as "A_n infinitely often" is defined as

$$\{A_n \text{ i.o}\} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$$

Similarly, the event, $\{A_n \text{ a.a.}\}$ referred as "A_n almost always" is defined as

$$\{A_n \text{ a.a.}\} = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$$

Both $\{A_n \text{ i.o.}\}, \{A_n \text{ a.a.}\} \in \mathcal{A}$ $\{A_n \text{ i.o}\}$ can be interpreted as the set of events that are in infinitely many of the A_n. $\{A_n \text{ a.a}\}$ is the set of events that are in all but a finite number of the A_n. Since each $A_i \in \mathcal{A}$, from the closure property of σ-algebra we have that $\{A_n \text{ i.o}\}, \{An \text{ a.a.}\} \in \mathcal{A}$. Consider the example of rolling a dice. If A_i is the event that the i^{th} roll is a 6, then $\{A_n \text{ i.o}\}$ is the event that 6 is rolled infinitely many times. $\{A_n \text{ a.a}\}$ is the event that a 6 is rolled all but finitely many times. Therefore, only a finite number of 1, 2, 3, 4, 5 are rolled. We have that "almost always" is stronger than "infinitely often" as "almost always" implies "infinitely often."

Definition 3.2. Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. Let $A_1, A_2, \ldots \in \mathcal{A}$ be a sequence of events. The tail field of these events is defined as

$$\tau = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, A_{n+2}, \ldots)$$

τ is a σ-algebra and members of τ are called tail events. For example $\{A_n \text{ i.o}\}, \{A_n \text{ a.a.}\} \in \tau$.

We will also make use of the following Lemma's about independent events.

Lemma 3.3. Let $A_1, A_2, A_3, \ldots$ be independent events. Then $\sigma(A_i)$ and $\sigma(A_1, A_2, \ldots, A_{i-1}, A_{i+1}, \ldots)$ are independent classes and for all $A \in \sigma(A_1, A_2, \ldots, A_{i-1}, A_{i+1}, \ldots)$ we have $\mathbf{P}(A_i \cap X) = \mathbf{P}(A_i)\mathbf{P}(X)$.

Lemma 3.4. Let $A_1, A_2, A_3, \ldots, B_1, B_2, \ldots$ be independent. Then

i If $X \in \sigma(A_1, A_2, \ldots)$ then $X, B_1, B_2, \ldots$ are independent.

ii The σ-algebras $\sigma(A_1, A_2, \ldots)$ and $\sigma(B_1, B_2, \ldots)$ are independent classes. So, if $X \in \sigma(A_1, A_2, \ldots)$, and $Y \in \sigma(B_1, B_2, \ldots)$ then $\mathbf{P}(X \cap Y) = \mathbf{P}(X)\mathbf{P}(Y)$.

We now have the preliminaries to introduce Kolmogorov's Zero-One Law.

Theorem 3.5. Kolmogorov's Zero-One Law Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. Let $A_1, A_2, \ldots \in \mathcal{A}$ be a sequence of independent events with tail field τ. If $T \in \tau$ then $\mathbf{P}(T) \in \{0, 1\}$

Proof. Consider $n \in \mathbb{N}$. We have by definition $T \in \sigma(A_{n+1}, A_{n+2}, \ldots)$. By Lemma 3.4(i), we have that $T, A_1, A_2, \ldots, A_n$ are independent. Now consider $S \in \sigma(A_1, A_2, \ldots)$. Since $T, A_1, A_2, \ldots$ are independent we have from Lemma 3.3 that S and T are independent. We also have by definition that $T \in \tau \subseteq \sigma(A_1, A_2, \ldots)$. Therefore T is independent of itself. Therefore by definition of independent events,

$$\mathbf{P}(T) = \mathbf{P}(T \cap T) = \mathbf{P}(T)\mathbf{P}(T) = \mathbf{P}(T)^2$$

It therefore follows that $\mathbf{P}(T) \in \{0, 1\}$. □

4. Discussion

Kolmogorov's Zero-One Law is a much stronger statement than the Borel-Cantelli Lemma's which state

Lemma 4.1. Borel-Cantelli Lemma's Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. Let $A_1, A_2, \ldots \in \mathcal{A}$.

i If $\sum_n \mathbf{P}(A_n) < \infty$ then $\mathbf{P}(\{A_n \text{ i.o.}\}) = 0$

ii If $\sum_n \mathbf{P}(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent then $\mathbf{P}(\{A_n \text{ i.o.}\}) = 1$

The Borel-Cantelli Lemma applies for the specific events listed in the lemma. However, Kolmogorov's Zero-One Law applies to any tail event. Since $\{A_n \text{ i.o}\}$ is a tail event, Kolmogorov's law is more general than the Borel-Cantelli Lemma. Also, Kolmogorov's law tells us that the probability of the tail event is 0 or 1, it does not provide a means to determine what the probability is. Other means have been developed to determine this probability.

Email address, A. Agarwal: arjunagarwal010@gmail.com

Euler Circle, Palo Alto, CA, 94306