

History of Measure-Theoretic Probability and Martingales

MScFE 620 Group Project Submission 1

Avhad, Prajakta
prajakta.s.avhad@gmail.com

Hugo, Pepar Thomas Jay
peparhugo@gmail.com

Lin, Hong
plantvsbird@gmail.com

Ramsay, Ben
ramsay.ben@gmail.com

Wang, Jiao Yu
joseph_wang@live.ie

November 30, 2020

1 Introduction

The classical foundation of probability theory can be traced to the work by Abraham De Moivre and Jacob Bernoulli. This remained the canonical theory until the early twentieth century when it underwent an overhaul, along with mathematics in general at the time. This period saw a level of abstractness applied to classical calculus which gave birth to topics such as geometric probability, relative probability, Cournot's principle, and Bertrand's paradox. Andrei Kolmogorov, well known for his contributions to modern probability, established its axiomatic basis in 1933. Kolmogorov enhanced classical probability by adding countable additivity to it. Later, individuals such as Markov and Doob made further contributions to the theory to give modern probability theory as we know it today. In this report, we will describe the developments of modern probability from the perspectives of these three figures. We will begin with Kolmogorov's work, describe contributions by Markov and end with the theory of Martingales as developed by Doob.

2 Kolmogorov and axioms of probability

Modern probability theory has its roots in the definition of probability as introduced by Andrey Kolmogorov in 1933. Kolmogorov proposed three axioms of probability that generalized probability beyond existing empirical definitions to a conceptual definition. Kolmogorov axioms depend upon what is referred to as a *probability space* and it is defined as follows:

1. Ω is the set of all possible events or outcomes in the sample space.
2. \mathcal{F} is the collection of sets of events or outcomes from the sample space.
3. P is the probability measure to assign the probability to a set of events or outcomes occurring from \mathcal{F} (Meyers 2020).

A *probability space* is then defined as (Ω, \mathcal{F}, P) . It can then be stated that event A is an element of \mathcal{F} where A is a sub-set of Ω .

From the *probability space* Kolmogorov's Axioms of Probability set the baseline for mathematical probability and are as follows:

1. $0 \leq \mathbb{P}(A) \leq 1$
2. $\mathbb{P}(\Omega) = 1$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, where A and B are mutually exclusive events (Meyers 2020).

The first axiom states that the probability for any event A must be a non-negative number between 0 and 1 inclusively. It represents the fractions of occurrence for event A which is an element of \mathcal{F} .

The second axiom states that the probability of the sample space must be equal to 1. Another way to think of this is Ω contains all possible outcomes and the probability of any event happening from Ω is 1.

The third axiom states that the probability of two mutually exclusive events A and B must be equal to the sum of the individual probability of A and B .

By combining these three axioms one can deduce some interesting properties. First, it can be said the probability of any event occurring that does not exist in Ω is 0. If we know that $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, then $\mathbb{P}(\Omega \cup \Omega^c) = \mathbb{P}(\Omega) + \mathbb{P}(\Omega^c) = 1 + \mathbb{P}(\Omega^c)$ meaning $\mathbb{P}(\Omega^c) = 0$. This means the probability of any event happening outside of the sample space Ω is 0. The complement of Ω is referred to as the empty set (Taylor 2019).

The second interesting property is if $A \cup A^c = \Omega$ and $\mathbb{P}(\Omega) = 1$, then $\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1$ meaning $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. This means the probability of the complement of E occurring is 1 minus the probability of event E occurring (Taylor 2019).

The Axioms of Probabilities Kolmogorov proposed apply to both classical probabilities and frequentist probabilities. It is also important to notice that these axioms are general statements about the features of a probability measure \mathbb{P} but it does not give any details about the nature of the probability measure itself. This is where Kolmogorov separated the definition of probability from empirical study, where the nature of the probability measure is being studied, to a generalized form where the nature of the probability measure is not needed. The axioms do not “tell us where and when to apply the rules, give us guidelines or procedures for calculating probabilities, nor any insights to the nature of random processes” (Glen 2020). Thus Kolmogorov's axioms generalized probability beyond the empirical study of events and now hypothetical probability spaces could be defined without understanding how the probability measure works and allowed theoretical research of probability spaces where the nature of the probability measure is not defined.

3 Markov and Markov processes

Any historical discussion of Markov Processes must first focus on the Russian mathematician, Andrei Markov, after whom the concept was named. In his early years, Andrei had to persevere through poor health, only being able to walk with the assistance of crutches during his childhood. Despite his handicap, during his schooling he was able to demonstrate an innate talent for math, writing his first paper in secondary school

on the integration of linear differential equations. Such precocious work got the attention of two professors at St. Petersburg university where Markov would go on to gain his doctorate and teach as a professor working alongside Pafnuty Chebyshev (O'Connor and Robertson 2006).

During his time at the university, Markov studied sequences of mutually dependent variables which led him to the topic for which he is particularly remembered, Markov chains. His work in this area gave way to a novel branch of probability theory and preempted the theory of stochastic processes. Following Markov's work in the area, Norbert Wiener would be the first to rigorously develop a continuous Markov process where Andrei Kolmogorov would go on to lay the foundations of Markov random processes as a theory.

When markov processes are examined in discrete time, they may be referred to as markov chains (Debnath and Basu 2015). Markov chains are sequences of random variables where subsequent variables are determined by their immediate predecessor, but are independent of where the predecessor originated. In its simplest form, markov chains apply to a system where one of a number of states $\{S_1, S_2, \dots, S_n\}$ can transition from one state to another. The probabilities associated with such a transition from S_i to S_j is an $n \times n$ matrix (p_{ij}) which is appropriately named the Markov transition matrix. The distinctive property where the following state depends on the current state can easily be seen in the transition matrix where any pair of states is associated with a specified probability.

When generalized to measure-theoretic perspective, markov processes can be defined as follows (Eberle 2015). Given:

- A filtration on $(\Omega, \mathcal{F}, \mathbb{P})$
- State space (S, \mathcal{B})
- A stochastic process X_t that is adapted with respect to a filtration \mathcal{F}_t

Then the stochastic process X_t on the state space is called a markov process if and only if

$$\mathbb{P}[X_t \in \mathcal{B} | \mathcal{F}_s] = \mathbb{P}[X_t \in \mathcal{B} | X_s], \quad \mathbb{P} - \text{a.s. } \forall \mathcal{B} \in \mathbb{B} \text{ and } s, t \in I \text{ and } s \leq t$$

and given the above condition, it may be deduced that:

$$\begin{aligned} \mathbb{P}[X_t \in \mathcal{B} | \mathcal{F}_s] &= \mathbb{P}[X_t \in \mathcal{B} | X_s], & \mathbb{P} - \text{a.s. } \forall \mathcal{B} \in \mathbb{B} \text{ and } s \leq t \\ E[f(X_t) | \mathcal{F}_s] &= (p_{s,t} f)(X_s), & \mathbb{P} - \text{a.s. } \forall f \in m\mathcal{F} \text{ s.t. } f : S \rightarrow \mathbb{R}_+ \text{ and } s \leq t \end{aligned}$$

Where $p_{s,t}$ is the transition probability from state s to t (Daruich 2014).

It can be seen that Markov process dynamics are entirely described through its transition function. This provides extreme convenience as new probability measures need only be conditioned on the previous state. This is in stark contrast to a general case where a new measure would have to be used for each period depending on all previous states.

Markov processes are a powerful concept and still remain an active area of research today. Common applications seen in the real-world may be found in various natural language processing applications, aspects of Google's PageRank formula, and of particular relevance to this course, quantitative finance.

4 J. L. Doob and the development of martingales

Joseph L. Doob's first paper in probability was titled "Probability and Statistics" (Athreya 2015; Doob 1934a). The paper inherited Kolmogoroff's method, and used the mathematical formulations of measure theory to describe and build the theory of probability, with the aim of providing a rigorous formulation for the study of statistics. In the paper, he achieved the last point by proving the validity of the maximum likelihood, which is frequently used in statistics today.

In the same year, Doob published another paper titled "Stochastic Processes and Statistics" which expanded upon Khintchine's results in the same topic (Doob 1934b). Two additional theorems were developed which could be used to provide an alternative proof of the maximum likelihood method.

Later in "Probability as a Measure," Doob made the observation that probability was simply a branch of measure theory. This was shown through a rigorous mathematical formulation and it was hoped that the results of this work could eventually serve as a foundation for all statistical problems (Doob 1941).

With this groundwork laid, Doob introduced his most important work and what he is known for today: Martingale theory (Doob 1941). The term martingale comes from a gambling practice where gamblers always bet the entire amount lost, so that each win will make the gambler even. In "What is A Martingale", Doob (1971) gives the following definition¹:

"With sample space Ω , the expectation of a random variable x can be defined as $E[x] = \sum_j x(\omega_j)p_j$, where $\omega_j \in \Omega$, $x(\omega_j)$ is the value of random variable x with outcome ω_j , and p_j refers to probability of the outcome."

Without using the term "filtration" explicitly, Doob defines $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ as a finite or infinite increasing sequence of σ -algebras generated by partitions of Ω . The sequence of $\{x_n, n \geq 1\}$ is called martingale relative to $\{\mathcal{F}_n, n \geq 1\}$ if

$$E\{x_n | \mathcal{F}_m\} = x_m$$

for $m < n$, and each x_n has an expectation. This simplistic definition can also be expanded to an uncountable sample space.

In a more concrete example, if we define x_1 as a gambler's fortune before they play a game, and x_2 as the fortune after the game, the game can be considered fair if

$$E\{x_2\} = x_1$$

and with more game played, if the fairness of the game still holds, we have

¹which is substantially shorter in length compared to an earlier definition in his book "Stochastic Processes" (Doob 1953)

$$E\{x_{n+1}|x_1, \dots, x_n\} = x_n, n = 1, 2, \dots$$

Doob then defines a mathematical model of fair game as a martingale:

$$\{x_n, \mathcal{F}_n, n \geq 1\}$$

relative to some stated Borel fields. Intuitively, the field \mathcal{F}_n summarises all the influencing variables up to time n at that point. As Doob explained in “What is a Martingale” and “Stochastic Processes” (Doob 1953; Doob 1971), the concept of a martingale is useful because it formalises the concept of a game being “fair”, meaning that a player can only ever learn more information as they play it.

5 Conclusion

In conclusion, this report describes the development in modern probability theory, with a strong focus on stochastic processes. The earliest groundwork was laid by Kolmogorov, beginning with his axioms. Later, Markov made contributions to the field through his work on random walks. This eventually paved the way for the theory of martingales which was formalised by Doob. The collective efforts of these giants now forms the foundation of modern probability and statistics, on top of which many new fields have spawned and flourished.

References

- Athreya, K. B. (2015). “Joseph L Doob and development of probability theory”. In: *Resonance* 20, pp. 286–288.
- Daruich, D. (2014). *Notes on Measure Theory and Markov Processes*. http://ddaruich.weebly.com/uploads/3/8/1/2/38127151/measure_theory_notes.pdf. Accessed: 29th November 2020.
- Debnath, L. and K. Basu (2015). “A short history of probability theory and its applications”. In: *International Journal of Mathematical Education in Science and Technology* 46, pp. 13–39.
- Doob, J. L. (1934a). “Probability and statistics”. In: *Trans. Amer. Math. Soc.* 36, pp. 759–775.
- (1934b). “Stochastic Processes and Statistics”. In: *Proceedings of the National Academy of Sciences of the United States of America* 20.6, pp. 376–379.
- (1941). “Probability as a measure”. In: *The Annals of Mathematical Statistics* 12.2, pp. 206–214.
- (1953). *Stochastic Processes*. New York, Wiley.
- (1971). “What is a Martingale?” In: *The American Mathematical Monthly* 78.5, pp. 451–463.
- Eberle, A. (2015). *Markov Processes*. http://http://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/MarkovProcesses/MPSkript1415.pdf. Accessed: 29th November 2020.
- Glen, S. (2020). *Axiomatic Probability: Definition, Kolmogorov’s Three Axioms - Statistics How To*. Accessed: 20th November 2020.
- Meyers, D (2020). *CS 547 Lecture 6: Axioms of Probability*. http://pages.cs.wisc.edu/~dsmyers/cs547/lecture_6_probability.pdf. Accessed: 28 November 2020.
- O’Connor, J. J. and E. F. Robertson (2006). *Andrei Andreyevich Markov - Biography - Maths History, University of St Andrews, Scotland*. <https://mathshistory.st-andrews.ac.uk/Biographies/Markov/>. Accessed: 29th November 2020.

Taylor, C. (2019). *What Are Probability Axioms?* <http://www.thoughtco.com/what-are-probability-axioms-3126567>. Accessed: 28th November 2020.