Applications of Probability
that the Average Student can Understand

Fu Lingyue

June 2020

Abstract

This material mainly focuses on the application of what I have learned in this semester, illustrates my visual understanding of some concepts, and chooses three comprehensive examples to apply what I have learned in this semester. The main line of the paper is organized from shallow to deep in the order of learning. It is a relatively fragmented but overall review of learning, with some fragmented ideas.

1 Introduction

In previous studies, we have studied classical generalizations, as if “randomness” were a natural concept. It’s random whether a coin flips heads or tails, but how do you say that the number of times the coin flips heads is $\frac{1}{2}$? In mathematics, there are a lot of things that are obvious that are not really easy to prove. After this semester’s study, I have a new understanding of probability.

The process of learning measure theory is tedious at first, especially when it is not combined with practical problems. When the concept of expectation was defined, I even questioned the previous understanding of “expectation”!

Figure 1: Some important theorems and concepts in probability.

It seems to me that measures are to probability theory what the completeness of the real space is to calculus. We can integrate without all that compactness, but that’s at the calculator level. As a brain that has evolved for 200,000 years or so, it shouldn’t just do these simple, rule-based calculations. Only by truly understanding probability from the whole space and the whole process can we have a better understanding of probability. Unfortunately, little is known about its beauty. I don't have a very deep understanding yet, but I hope that more people will follow me and begin to "realize" that there are many mysteries to explore. Thus in this material I will combine examples to show the important contributions of measure theory and martingale theory.

2 Probability space, our study space.

In our past, "random" was a random event. Random events are an obvious description. We even think of the probability as simply the probability of picking this group out of a bunch of combinations. When the probability space is introduced, these possibilities become points in the space.

In traditional probabilities, it's almost impossible to prove something formally that looks obvious(a comic as Fig.2). Also something we think is obvious(like an event is random, two events have some relation...), maybe needs the whole system of measure theory to support.

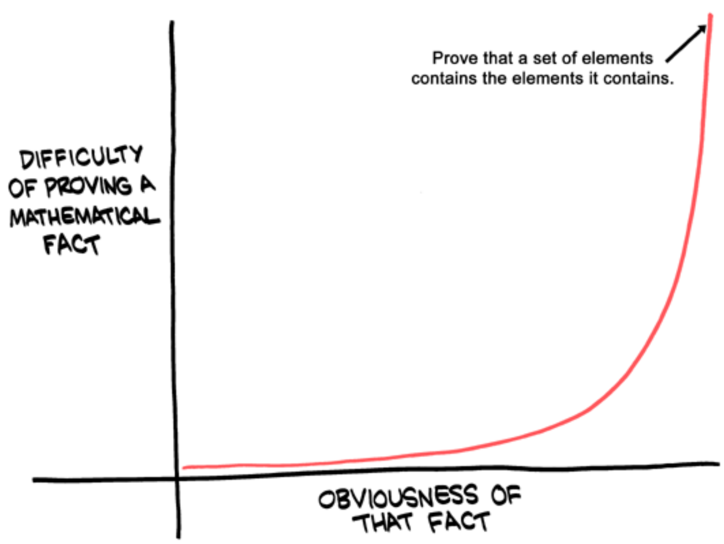

Figure 2: One of my favorite comic of proving mathemetical problem.

Here is an example, the exercise EG3 in textbook.

Example 1. *Let G be the free group with two generators a and b. Start at time 0 with the unit element 1, the empty word. At each second multiply the current word on the right by one of the four elements a , a^{-1} , b , b^{-1} , choosing each with probability $1/4$ (independently of previous choices). The choices*

$$a, a, b, a^{-1}, a, b^{-1}, a^{-1}, a, b$$

at times 1 to 9 will produce the reduced word aab of length 3 at time 9. Prove that the probability that the reduced word 1 ever occurs at a positive time is $1/3$, and explain why it *is intuitively clear that (a.s)*

$$\textit{(length of reduced word at time n)}/n \to \frac{1}{2}.$$

Remark 1. *The average student may be confused before studying probability theory and even think that it can be solved by simple numbers(especially the second statement). But actually there's something about SLLN and Markov Chain. This is not an obvious thing.*

Solution. Then length of the reduced word defines a Markov chain as follows:

$$0 \leftrightarrows^{\frac{1}{4}}_{1} 1 \leftrightarrows^{\frac{1}{4}}_{\frac{3}{4}} 2 \leftrightarrows^{\frac{1}{4}}_{\frac{3}{4}} \cdots$$

Then we just have to calculate $p := \mathbb{P}(\{\text{hitting 0 given length is 1}\})$. Here we have

$$\begin{aligned}
p &= \frac{1}{4} + \frac{3}{4}\mathbb{P}(\{\text{hitting 0 given length is 2}\}) &(1)\\
&= \frac{1}{4} + \frac{3}{4}\mathbb{P}(\{\text{hitting 0 given length is 1, hitting 1 given length is 2}\}) &(2)\\
&= \frac{1}{4} + \frac{3}{4}\mathbb{P}(\{\text{hitting 0 given length is 1}\})\mathbb{P}(\{\text{hitting 1 given length is 2}\}) &(3)\\
&= \frac{1}{4} + \frac{3}{4}p^2. &(4)
\end{aligned}$$

Step (2) to (3) uses Strong Markov Property[4] and (3) to (4) uses the symmetry in the chain. Then we get

$$p = \frac{1}{3} \text{ or } 1.$$

Let A_n^k be the event that reaching state 0 from state n in at most k steps, and p is its probability. Define $A_n := \bigcup_k A_n^k$. Thus $p_n = \lim_{k\to\infty} p_n^k$. So we just have to prove that $\forall k$ $p_1^k < 1$ (in this way $p = p_1$ can only be 1/3). In order to get this, we introduce the variable x_n:

$$x_1 = \frac{1}{4} + \frac{3}{4}x_2,$$

$$x_n = \frac{1}{4}x_{n-1} + \frac{3}{4}x_{n+1}, n \geq 2.$$

And we can solve that $x_n = (\frac{1}{3})^n, n \geq 1$. Then we have that $p_n^k \leq x_n$ obviously. So here we showed that $p = \frac{1}{3}$.

The next statement can be proved as follows.

The idea is simple: For every step(except 0 state), the length is increased 1 with $p = \frac{3}{4}$ and decreased 1 with $p = \frac{1}{4}$. Thus the average is $\frac{1}{2}$. But it truly needs some trick in probability space.

Let w_n be the word length at time n. We introduce the i.i.d. $X_n \in \{-1, +1\}$ where $\mathbb{P}(X_n = 1) = \frac{3}{4}$. Then let $Y_n = I(w_{n-1} = 0, X_n = -1)$. Here we hold an equality:

$$w_n = \Sigma_{k\leq n} X_k + \Sigma_{k\leq n} 2Y_k.$$

Use SLLN we have

$$\mathbb{P}(\frac{1}{n}\Sigma_{k\leq n}X_k \to \mathbb{E}(X_n) = \frac{1}{2}) = 1.$$

And those $\Sigma_n\mathbb{P}(w_n = 0) < \infty$ can be proved by zooming. Thus we can prove $\frac{1}{n}\Sigma_{k\leq n}Y_k \to 0$ (a.s.).

□

Through this example, measure theory shows its powerful proof ability.It is a strong backing and foundation for the rest of our content, and a suitable environment to include the martingale process.

3 Martingale, a flat line in probability space.

I think it is very important to understand abstract concepts figuratively. Examples also important to understand a new concept. Thus in this section I list some examples and understanding of martingale. In this way, I want those ordinary students like me can also feel the beauty of martingale.

For probabilists, martingales are first of all integrable processes, satisfying a particular conditional expectation property. Aside from their role in finance, they have applications to various stochastic and analytic problems and represent, with Markov processes, one of the most important types of processes depending on the past. The notion seems to arise quite directly from the idea of strategy in a game of chance. Although the intuitive understanding that no strategy in an unfavorable game always wins arose very early (in B. Bru's unpublished notes), one has to wait until the beginning of the 20th century to obtain a formalisation of the notions and of the problem. With martingale's invention, everything became clear.

3.1 Martingale and Markov Process

Martingale and Markov are both important quantities that describe processes, but they are different. And they have no containment relationship. Here is a table to list their differences.

Martingale

- It has a memory(infinite memory).
- It emphasizes the "fairness" of a gambling game.
- A classic example is the gambler gambles and his gain per inning is a martingale.

Markov Process

- It does not have a memory(finite memory).
- It emphasizes the "finite" property of a process.
- A classic example is that Q-learning in machine learning, that is the process by which a robot learns whether to turn an intersection or not.

This example shows that in the process of learning probability theory, we need to establish some knowledge of abstract concepts to help us understand.

Moreover, after some consideration of their inclusion, the following conclusion can be drawn: They are not related. Here's two examples:

Example 2. *This is Markov but not martingale process. Brownian motion with drift*

$$X(t) = \mu t + W(t).$$

The mean of X_t is getting bigger all the time, thus it is not fair, and it is not a martingale. But it is a Markov process.

Example 3. *This is a martingale but not a Markov process. Define*

$$dX(t) = \int_0^t X(s) ds dW(t).$$

Here we need to know all of the previous state to get the state at time n, thus it is not a Markov process. But this process will be a martingale, because every increment can be expressed as the sum of the path and the Brownian motion increment, and the mean of the Brownian motion is zero, so the increment will be zero, without violating the property of the martingale.

Now, some special processes have been defined, and they are the ones that we care about the most. These basic models have many nice properties and can be applied to many scenarios.

3.2 Monkeys and Shakespeare

A very classic martingale process is when monkeys knock out the complete works of Shakespeare. Here we can use martingale and optional stopping theorem to solve the problem that a monkey to type out the complete works of Shakespeare. This is a very good example that we have in the text as well.

Example 4. *At each time $t = 1, 2, 3...$ a monkey types a single capital letter (out of a choice of A-Z), chosen independently of the letters it has previously typed. We want to find out how long one expects to wait before the monkey types the word SHAKESHASHA. This is also the exercise E10.6 in textbook.*

Remark. Why we use the sequence "SHAKESHASHA"? Because it has some repeat sequence (like "SHA"), the we can better show the universality of following computation.

Solution. First, construct a process like this: before each time $n = 1, 2, \ldots,$ a new gambler arrives. And the procedure is as follows. He bets RMB 1 that **the n^{th} letter be 'S'**. If he loses, he leaves. If he wins, he receives RMB 26. Then if he wins the i^{th} game, he will use all his money to the next game. That is to say: he will bet RMB 26 that **the $n+1^{th}$ letter will be 'H'**. Do this until the sequence ends. Let W^n be the procedure of the n^{th} gambler. It is clear that $W_k^n = 0$ when $k < n$ since the n^{th} gambler even has not arrived before time

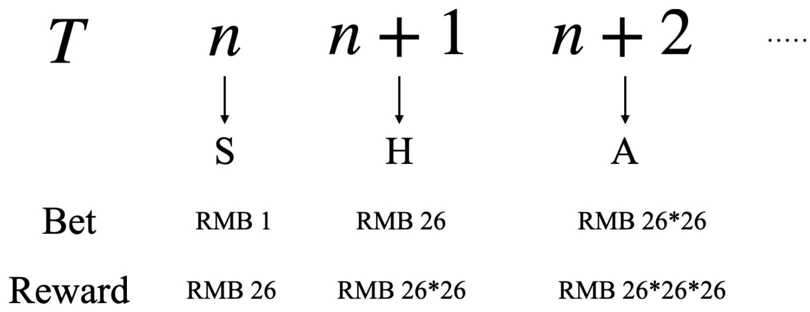

Figure 3: The betting procedure model.

stamp n. Then it is clear that every W^n is a martingale. And $W_n := \Sigma_{m=1}^{n} W_n^m$ is also a martingale. Furthermore, W has uniformly bounded increments. Let T be the first time our cute monkey has produced the consecutive target sequence, then $\mathbb{E}(T) < \infty$. At this time, "SHAKESHASHA" has just been typed. All the gamblers except $T-3$, $T-6$, $T-11$ have lost 1 RMB. Thus, according to the optional stopping theorem, we have

$$0 = \mathbb{E}(M_T) = E(\Sigma_{n=1}^{T} M_T^n) = \mathbb{E}((26^{11} - 1) + (26^6 - 1) + (26^3 - 1) + (-1)(T - 3)).$$

Solve the equation, we get $\mathbb{E}(T) = 26^{11} + 26^6 + 26$. □

Thus if we want the monkey to type the whole Shakespeare book, it have to take a very very long time. May be until it dies.

Little conclusion As the title of this chapter states, a martingale is a horizontal line in a probability space. An ascending line is a martingale, and a descending line is a martingale. These lines are measures of the process, and they have very nice properties of their own.

The word martingale itself is associated with many fields. It also has many meanings in many different dictionaries(I have read an article on the etymology of martingale [3]). Until now, machine learning, gambling, stocks and other fields have also been deeply applied martingale knowledge.

4 An Interesting Application of Central Limit Theorem

Besides, some basic theorems also happen in our daily life.

This is a comprehensive application in physics that I've seen in a physical competition. I've been scratching my head. In the course of this semester, I was surprised to find that it could be explained by the **central limit theorem**!

Drifting Speckles Shine a laser beam onto a dark surface(See Fig4). A granular pattern can be seen inside the spot. When the pattern is observed by a camera or the eye, that is moving slowly, the pattern seems to drift relative to the surface.

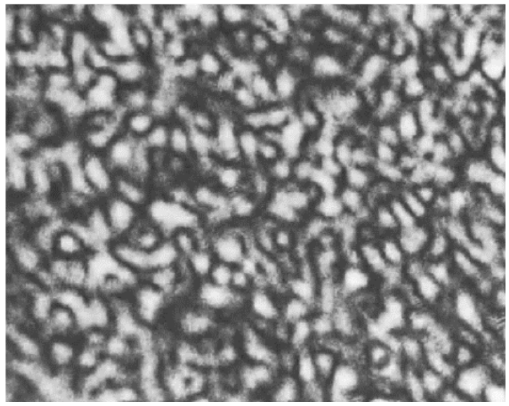

Figure 4: Spots of light on the wall.

Following sections are going to explain the phenomenon and investigate how the drift depends on relevant parameters.

4.1 Emergence of Phenomena

Laser speckle is a kind of random distribution of light intensity and its formation needs to meet the conditions of coherent light and scattering medium. And as the object moves, the light enters the eye in different places.

4.2 Introduction of Variables

- Suppose that the amplitude of light at any point in space is entirely due to the reflected light of the laser:

$$A(x, y, z_{=}|A(x, y, z)|e^{i\theta(x,y,z)} = \Sigma\frac{1}{\sqrt{N}}a_k(x, y, z) = \Sigma\frac{1}{\sqrt{N}}|a_k|e^{i\theta_k}.$$

- The light intensity is decomposed to the complex plane, and is defined as follows:

$$A^r = Re(A) = \Sigma\frac{1}{\sqrt{N}}|a_k|cos\phi_k,$$

$$A^i = Im(A) = \Sigma\frac{1}{\sqrt{N}}|a_k|sin\phi_k.$$

Their statistical rule says that

$$\langle A^r\rangle = \langle A^i\rangle = 0,$$

$$Var(A^r) = Var(A^i) = \frac{1}{N}\frac{\Sigma\langle|a_k|^2}{2}.$$

(That's the property here that leads me to use the CLT!)

4.3 Theoretical Explanation

We assume that the amplitude and phase of each beam are independent and uniformly distributed in phase. Thus we have a set of **i.i.d.**, i.e., $a_k(x, y, z)$ is independent and identically distributed.

Since the reflections occur on rough walls, we can assume that the number of superimposed beams N is very large. According to the Central Limit Theorem, we have

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} \frac{\Sigma\langle|a_k|^2}{2},$$

which means

$$P_{A^r} = \frac{\sqrt{2\pi}\sigma}{1} e^{-\frac{[A^r]^2}{2\sigma^2}},$$

$$P_{A^i} = \frac{\sqrt{2\pi}\sigma}{1} e^{-\frac{[A^i]^2}{2\sigma^2}}.$$

Because A^r, A^i are complex plane factorizations, they are independent to each other. Thus we can infer

$$P(A^r, A^i) = \frac{\sqrt{2\pi}\sigma}{1} e^{-\frac{[A^r]^2+[A^i]^2}{2\sigma^2}}.$$

In order to get the relation between light intensity I and phase θ, let's set up the their equations:

$$A^r = \sqrt{I} cos\theta,$$

$$A^i = \sqrt{I} sin\theta,$$

$$P(I, \theta) = P(A^r, A^i)|K|.$$

Finally, we get the probability distribution of light:

$$P(I, \theta) = \begin{cases} \frac{1}{4\pi\sigma^2} e^{-\frac{I}{2\sigma^2}}, & I \geq 0, -\pi \leq \theta < \pi \\ 0, & \text{otherwise.} \end{cases}$$

And because of the independent of P_I and P_θ, we can get

$$P_I = \frac{1}{2\sigma^2} e^{-\frac{I}{2\sigma^2}}, I \geq 0.$$

$$P_\theta = \frac{1}{2\pi}, -\pi \leq \theta \leq \pi.$$

So here, these two things explain why the position of the light plate is changing. And in the probability distribution of the known intensity P_I cases, we can use the computer simulation of speckle phenomenon(See Fig.5).

By this example, we can feel that these mathematical principles are actually lurking in our lives. These mathematical foundations will also lead to new inventions in different fields.

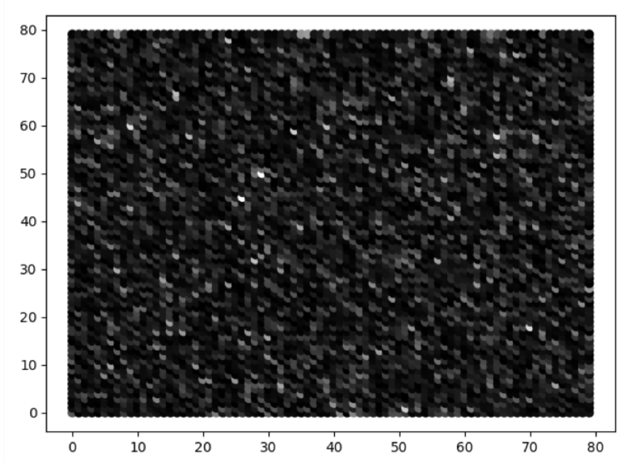

Figure 5: The computer simulation result of our solution.

5 Conclusion

In the future, the integration and crossover between disciplines will be more and more attractive. The mathematical basis of science and engineering is very important. The point of this course for us is to train our minds, but also to understand some of the basic structures of mathematics.

In a word, this material is intended to support an important point: probability theory has a very broad and profound application space. The power of probability theory is demonstrated by examples that are easily understood by ordinary students. “Simple is not always easy to prove,” which is also something to keep in mind. Keep enjoying the math!

References

[1] What is a martingale and why is it important? quora.com. URL: https://www.quora.com/What-is-a-martingale-and-why-is-it-important.

[2] MUB (https://math.stackexchange.com/users/426654/mub). Exercises eg.3 and eg.4 in probability with martingales, by david williams. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/2474780 (version: 2017-10-16).

[3] Roger MANSUY. The origins of the word “martingale”. jehps.net. URL:http://www.jehps.net/juin2009/Mansuy.pdf.

[4] J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

[5] user13247. Why are stochastic processes with decreasing expected value called supermartingales? Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/858817 (version: 2020-06-12).