Probability Lecture Notes 13

This is the lecture notes of probability class 13, which is due to be published by *Fan Yuda* in May 16, 2018. Any question is welcome to the email address *mistergalahad@gmail.com*.

Abstract

This lecture notes is expected to contain three aspects, where lay the discussion centred around the contents in our class. First and foremost, we will introduce **Martingale and Optional Stopping Theorem**. Apart from that, we will discuss **Expected value of tossing a coin until the emergence of certain patterns**. Last but not least, we will focus on the **Distribution of Probability**.

Section A Martingale

In **probability theory**, a **martingale** is a sequence of **random variable**s for which, at a particular time in the **realized** sequence, the **expected value** of the next value in the sequence is equal to the present observed value even given knowledge of all prior observed values.

Defenition

A narrow and basic definition of Martingale is a **Discrete-time Process** , in other words, a sequence of of **Random Variables** that satisfies for any time **n**

$$\mathbf{E}(|X_n|) < \infty$$

$$\mathbf{E}(X_{n+1} \mid X_1, \ldots, X_n) = X_n$$

That is, the **Conditional Expected Value** of next observation, given all the past observations, is equal to the most resent observation.

Examples

- An unbiased **random walk** (in any number of dimensions) is an example of a martingale.

- A gambler's fortune is a martingale if all the betting games which the gambler plays are fair. To be more specific: suppose X_n is a gambler's fortune after **n** tosses of a **fair coin**, where the gambler wins \$1 if the coin comes up heads and loses \$1 if it's tails. The gambler's conditional expected fortune after the next trial, given the history, is equal to his present fortune. This sequence is thus a martingale.

  A very important of the above gambl of **doubling strategy**.

- (**Abraham de Moivre or de Moivre's martingale**) Now suppose the coin is unfair, i.e., biased, with probability **p** of coming up heads and probability $q = 1 - p$ of tails. Let

$$X_{n+1} = X_n \pm 1$$

with **+** in case of **heads** and **−** in case of **tails**. Let

$$Y_n = (q/p)^{X_n}.$$

Then $\{ Y_n : n = 1, 2, 3, \dots \}$ is a martingale with respect to $\{ X_n : n = 1, 2, 3, \dots \}$. To show this

$$E[Y_{n+1} \mid X_1, \dots, X_n] = p(q/p)^{X_n+1} + q(q/p)^{X_n-1}$$

$$= p(q/p)(q/p)^{X_n} + q(p/q)(q/p)^{X_n} = q(q/p)^{X_n} + p(q/p)^{X_n} = (q/p)^{X_n} = Y_n.$$

Stopping Times

A stopping time with respect to a sequence of random variables $X_1, X_2, X_3, \dots$ is a random variable τ with the property that for each t, the occurrence or non-occurrence of the event $\tau = t$ depends only on the values of $X_1, X_2, X_3, \dots, X_t$. The intuition behind the definition is that at any particular time t, you can look at the sequence so far and tell if it is time to stop.

An example in real life might be the time at which a gambler leaves the gambling table, which might be a function of his previous winnings (for example, he might leave only when he goes broke), but he can't choose to go or stay based on the outcome of games that haven't been played yet.

Optional Stopping Theorem

Attention

According to the limited pages, we can only roughly talk about this beautiful theorem epescially the concept **Filtration**, which is an essential part of the theorem.

If you feel buried by the formula in this part you can choose skip it and only focus on our **conclusion** because the proof is somehow far-fetched from our topic.

Definition

A discrete-version theorem goes below:

Let $X = (X_t)_{t \in \mathbb{N}_0}$ be a discrete-time **Martingale** and τ with values in $\mathbb{N}_0 \cup \{\infty\}$, both with respect to a **Filtration** $(\mathcal{F}_t)_{t \in \mathbb{N}_0}$. Assume that one of the following three conditions holds:

(**a**) The stopping time τ is **almost surely** bounded, i.e., there exists a **constant c** $\in \mathbb{N}$ such that $\tau \leq c$ a.s.

(**b**) The stopping time τ has finite expectation and the conditional expectations of the absolute value of the martingale increments are almost surely bounded, more precisely, $\mathbb{E}[\tau] < \infty$ and there exists a constant c such that $\mathbb{E}\big[|X_{t+1} - X_t| \,\big|\, \mathcal{F}_t\big] \leq c$ almost surely on the event $\tau > t$ for $t \in \mathbb{N}_0$.

(**c**) There exists a constant c such that $|X_{t \wedge \tau}| \leq c$ a.s. for all $t \in \mathbb{N}_0$ where $\wedge$ denotes the **minimum operator**.

Conclusion Then X_τ is an almost surely well defined random variable and $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.

Comments

Actually in our discussion (the gambling game), we only need the first property (**a**).

Personally speaking, the next two properties are too difficult to learn about and nowhere to adopt in reality.

Proof

We only consider the proof under the first situation.

Let X^τ denote the **stopped process**, it is also a martingale. Under condition(**a**) the random variable X_τ is well defined (which is the X^τ supposed to converge to).

Write the stopped process as follow

$$X_t^\tau = X_0 + \sum_{s=0}^{\tau \wedge t - 1}(X_{s+1} - X_s), \quad t \in \mathbb{N}_0$$

gives $|X_t^\tau| \le M$ for all $t \in \mathbb{N}_0$, where

$$M := |X_0| + \sum_{s=0}^{\tau-1}|X_{s+1} - X_s| = |X_0| + \sum_{s=0}^{\infty}|X_{s+1} - X_s| \cdot \mathbf{1}_{\{\tau > s\}}$$

By the Property of the expection

$$\mathbb{E}[M] = \mathbb{E}[|X_0|] + \sum_{s=0}^{\infty}\mathbb{E}\left[|X_{s+1} - X_s| \cdot \mathbf{1}_{\{\tau > s\}}\right].$$

Notice that given that condition (**a**) holds this series only contain finite **1**, hence M is well-defined or integrable.

Therefore, the stopped process is dominated by an integrable random variable M. Since the stopped process X^τ converges almost surely to X_τ , thus implying

$$\mathbb{E}[X_\tau] = \lim_{t \to \infty} \mathbb{E}[X_t^\tau].$$

By the property of martingale we can safely draw the conclusion that

$$\mathbb{E}[X_t^\tau] = \mathbb{E}[X_0], \quad t \in \mathbb{N}_0,$$

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0].$$

Application - Double strategy in betting system

Intorduction

Consider the such a gambling game in which the gambler wins his stake if a coin comes up heads and loses it if the coin comes up tails. The strategy took by the gambler is that he would double his bet after every loss, so that the first win would recover all previous losses plus win a profit equal to the original stake.

If we take a little insight into such situation, we will find that since none of the gamblers possessed infinite wealth, and the exponential growth of the bets would eventually bankrupt unlucky gamblers who chose to use the strategy . The gambler usually wins a small net reward, thus appearing to have a sound strategy. However, the gambler's expected value does indeed remain zero(which will be prove true later).

Analysis on single round

Let one round be defined as a sequence of consecutive losses followed by either a win, or bankruptcy of the gambler. After a win, the gambler "resets" and is considered to have started a new round. A continuous sequence of martingale bets can thus be partitioned into a sequence of independent rounds.

Let q be the probability of losing, and let B be the amount of the initial bet. Let n be the finite number of bets the gambler can afford to lose.

The probability that the gambler will lose all n bets is q_n. When all bets lose, the total loss is

$$\sum_{i=1}^{n} B \cdot 2^{i-1} = B(2^n - 1)$$

The probability the gambler does not lose all n bets is $1 - q_n$. In all other cases, the gambler wins the initial bet B. Thus, the expected profit per round is

$$(1 - q^n) \cdot B - q^n \cdot B(2^n - 1) = B(1 - (2q)^n)$$

Thus we can assume the q of an even coin is $\frac{1}{2}$, leading to the profit to be 0.

conclusion

Actually, there is an interesting fact that in each single round, the gambler use such strategy can merely earn $1, which also showes that the average loss is $1 as well ! If we stand in the point of expectation, it is like a random work in one dimension(not so serious and correct).

But through the conclusion of **Optional Stopping Theorem**, we can bring out the result that the gambler's fortune over time is a martingale, and the time τ at which he decides to quit is a stopping time, and the theorem reveals that $E[X_\tau] = E[X_0]$. In other words, the gambler leaves with the same amount of money on average as when he started.

If we consider the situation that the gambler possesses n dollars and banker possesses m dollars, we can discovery the X_τ only can end in two possible value 0 and $n + m$, so we can resolve the possibility of the gambler's win is $\frac{n}{n+m}$, indicating that it grows linearly as the possession grows.

Review on the first section

I intend to dedicate the first section to *Guo XiaoXu*, one of my most respectful senior students in our college. It is he who inspired me and gived me countless tutor in this problem. I'm also fairly glad to share the genious solution with others.

I'm sorry for being too rough in English and Mathematics to convey the beauty of the theorem to the readers. To see more detail you can search for it in the $Wikipedia$ and $MathOverflow$.

Section B Tossing a Coin

Introduction

Supposing that we repeated toss an even coin until the certain sequence of **head** and **tail** emerge, what is the expected numbers of tossing of it?

To give a specific case, we assume the the pattern to be HH, which means to two consective heads.

Solution A

In this subsection, we will show a method utilizing generative function.

Let f_k denote the number of patterns that we put k coins in a line without two consecutive heads. We can get the following recursive formula

$$f_{k+2} = f_{k+1} + f_k$$

According to $f_1 = 2$ and $f_2 = 3$, we can find that the f_k coincident with the **Fibonacci Sequence** F_k

$$1, \ 1, \ 2, \ 3, \ 5, \ ...$$

Actually $f_k = F_{k+2}$. Meanwhile , the last three toss must be **THH**, so that among all 2^k sequences, there are only F_{k-1} valid sequences. And the answer is

$$\sum_{k=1}^{\infty} k * \frac{F_{k-1}}{2^k} \qquad\qquad\qquad\qquad (*)$$

We can use generative function to resolve this, let $g(x)$ be the generative function of F_x. So

$$g(x) + xg(x) = \frac{g(x)}{x} - 1$$

$$g(x) = \frac{x}{1 - x - x^2}$$

Using some algebra tricks, we define $G(x) = g(x/2) * (x/2)$, we can gain the generative function of (*),

$$G(x) = \frac{x^2}{4 - 2x - x^2}$$

Deviarative the $G(x)$ and multiply with x we arrive in our destination

$$xG'(x) = 2 * \frac{x^2}{4} + 3 * \frac{x^3}{8} + 4 * \frac{2x^4}{16} ...$$

Let $x = 1$ we can get the answer 6.

Solution B

In this subsection, we will introduce a wonderful solution that lays on the basis of **Martingale**.

Supposing that we are doing gambling in a casino, which is merely repeated tossing coins. If you bet x Dollars and win, you get $2x$ dollars back, otherwise you lose it.

And the strategy of the gamblers is very simple, they keep on betting on the head side with all their fortune. At the beginning of each toss, there will be a gambler entering the casino and he won't leave until he bankrupt.

Unfortunately, the banker of the casino is very sensitive and he will close the casino if there is one man have win for exactly n times, which means there is n consecutive **head**s finally.

So when such patterns appears, there must be n gamblers in the casino and there total money is $2^n - 2$ (summarize from 2 to 2^{n-1}). Addtionally this is a **martingale** and the expected profit is zero, that means the average loss of the banker is exactly zero! Hence, there is expected to be $2^n - 2$ gamblers entering the casino, which means the game last $2^n - 2$ rounds! Amazing!

Remark

Such method can apply to any patterns, only modifying the strategy of the gamblers.

Review on Section B

I highly recommend the book written by *Matrix67* who is absolutely intelligent in math. These two solutions both origin from its contents. I also want to thanks for the discussion with *Li WenHao* and *Jin ZhiHan* who lend me a hand in such topic.

Section C Probability Distribution

Exponetial Distribution

Exponetial distribution refers to distribution whose density function is $f(x) = \lambda e^{-\lambda x}$ for any $x > 0$

Its **cumulative function** $F(x) = \int_0^x f(t)dt = 1 - e^{-\lambda x}$

So $P(x > t) = 1 - F(t) = e^{-\lambda x}$, leading to the meroless property

$$P(x > r + s | x > r) = \frac{e^{\lambda r}}{e^{\lambda(r+s)}} = e^{-\lambda s} = P(x > s)$$

Let's imagine a kind of atom whose length of emission interval is following the **Exponetial Distribution**, let Y denote the Radom variable which counts the number of the emissions that occur in the time interval t.

Y is discrete, so

$$P(Y = n) = P(S_n \leq t, S_{n+1} > t)$$

$$f(x_i) = \lambda e^{-\lambda x} \quad \forall i$$

in which the

$$S_n = \sum_{i=1}^n x_i, \quad f_{s_n}(x) = \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}$$

The result of the density function f_{sn} can be got by the method of mathematic induction, we simply prove the formula under $n = 2$

Proof

$$P(s_2 < x) = \int_0^x \lambda e^{-\lambda t}(1 - e^{-\lambda(x-t)})dt$$

$$= \int_0^x \lambda(e^{-\lambda t} - e^{-\lambda x})dt$$

$$= 1 - e^{-\lambda x} - \lambda x e^{-\lambda x}$$

$$f_{s_2}(x) = dP/dt = \lambda^2 x e^{-\lambda x}$$

Following above discussion, it is obviously that

$$P(Y = n) = P(S_n \leq t) - P(S_{n+1} \leq t)$$

$$F_n(x) = 1 - e^{-\lambda x}(1 + \lambda x + \ldots + \frac{(\lambda x)^{n-1}}{(n-1)!})$$

$$P(Y = n) = F_n(x) - F_{n+1}(x) = e^{-\lambda x} \frac{(\lambda x)^n}{n!}$$

Geometric Distribution

If $X \sim Geomtirc(p)$, that means

$$P(x = k) = p(1-p)^k$$

Hence,

$$E[X] = \frac{1}{p} \quad Var(X) = \frac{1-p}{p^2}$$

Actually, geometric distribution can be considered as the number of trials that a **Bernoulli Experiment** came out to be success.

So the memoryless property also holds for such distribution

$$P(X > r + s | X > r) = \frac{(1-p)^{r+s}}{(1-p)^r} = (1-p)^s = P(X > s)$$

Infer from this, we can get

$$P(X = i | X + Y = n) = \frac{p^2(1-p)^{n-2}}{(n-1)p^2(1-p)^{n-2}} = \frac{1}{n-1}$$

Poisson Distribution

$$P(x = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad E[x] = \lambda \quad Var(x) = \lambda$$

A seemingly-complicated proof

We will prove that $E[x^n] = B_n$, B_n refers to the method to partition n-set.

In other words

$$B_n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\} = \sum_{k=1}^{n} S(n, k)$$

where $S(n, k)$ refers to the methods partition n-set into k disjoint non-empty sets.

We use a single equation to prove it

$$m^n = \sum_{k=0}^{n} S(n, k)(m)_k$$

The $(m)_k$ is the **falling factorial**,

$$(m)_k = \prod_{t=0}^{k-1}(m - t)$$

Hence

$$\frac{m^n}{m!} = \sum_{k=1}^{n} \frac{S(n, k)}{(m - k)!}$$

Summarize the equation, we get

$$\sum_n \frac{m^n}{m!} \lambda^m = (\sum_{k=1}^{n} S(n, k)\lambda^k)(\sum_j \frac{\lambda^j}{j!})$$

Compare the coefficient, we can find

$$\sum_{k=1}^{n} S(n,k)\lambda^k = e^{-\lambda} \sum_{n} \frac{m^n}{m!}\lambda^n$$

In conclusion,

$$B_n = e^{-1} \sum_{m} \frac{m^n}{m!} = E[X^n]$$

A seemingly-simple proof

Let's prove that Possoin diribution is a general case of Bionomial distribution

Let's consider occurences heppens at random over a period of time, its property goes follow

(1) The average number of occurances per time unit is a constant

(2) The number of occurrences in two non-overlapping time intervals are independent

Let X denot the occurences in time period t, and $E[X] = \lambda t$

Slice t into n segments each duration is t/n, it is small enough that every interval can only contain single one occurrence.

Such that

$$\lambda t = np \quad p = \frac{\lambda t}{n}$$

Let $t = 1$ then

$$P(x=0) = (1 - \frac{\lambda}{n})^n = e^{-\lambda}$$

Consider the Bionomial Distribution $B(n,p,k)$

$$\frac{B(n,p,k)}{B(n,p,k-1)} = \frac{\lambda - (k-1)p}{k(1-p)}$$

Let $p \to 0$

$$P(x=k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

The composition and synthesis of Poisson random variable

Let A and B denote disjoint event space, X and Y respectively indicate the number of A and B

Assuming that $X + Y \sim Poisson(\lambda)$, we can prove that $X \sim Poisson(p\lambda)$ and $Y \sim Poisson((1-p)\lambda)$

Consider the joint distribution

$$P(X=n, Y=m) = \binom{n+m}{n} p^n (1-p)^m e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!}$$
$$= \frac{e^{-\lambda p}(\lambda p)^n}{n!} \frac{e^{-\lambda(1-p)}(\lambda(1-p))^m}{m!}$$

Take Summation for both X and Y, we can get

$$P(X=n) = \frac{e^{-\lambda p}(\lambda p)^n}{n!}$$

$$P(Y=m) = \frac{e^{-\lambda(1-p)}(\lambda(1-p))^m}{m!}$$

Review on the Section C

In section C we concentrate on various distribution. But what we should keep in mind is that all the above distributions are similar if we take specific case to observe, thus lead to many similar and interesting properties. I want to thanks for the efforts of our professor *Wu YaoKun* who bring this alive to us.