Basic Probability Theory (基础概率论)

Yuan Yao

July 2, 2020

1 Martingales

Martingale is a really fascinating concept in the probability theory. To begin our introduction, let's take a brief view on a famous example.

Now think of you are a gambler, betting on a coin game, successively. On each successive gamble, you either win \$1 or lose \$1 , independent of the past. How will you make decisions? And what's the probability that you win a certain fortune?

Note that every time you make a decision on whether or not to take the next coin game, you have learned all the results of the previous bets.

Definition 1. *We now take a filtered space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbf{P})$. *Here,* $(\Omega, \mathcal{F}, \mathbf{P})$ *is a probability triple,* $\{\mathcal{F}_n : n \geq 0\}$ *is a* ***filtration***. *A filtration is an increasing family of sub* σ*-algebras of* $\mathcal{F}$ *such that*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$$

We define

$$\mathcal{F}_\infty := \lim \mathcal{F}_n = \sigma(\bigcup_n \mathcal{F}_n) \subseteq \mathcal{F}$$

A straight foward understanding of filtration is that, at time n, all we know about ω in Ω are exactly the values of $Z(\omega)$ for all $\mathcal{F}_n$ measurable functions Z. As time n increases, we gather more information, and no information before time n is lost.

Definition 2. *A process* $X = (X_n : n \geq 0)$ *is called* ***adapted*** *to the filtration* $\{\mathcal{F}_n\}$ *if for each* n, X_n *is* $\mathcal{F}_n$*-measurable.*

We can intuitively think of the concept as claiming the result of process X depends on the information we have up to time n, and relies on nothing aboout the future.

Now it's high time that we gain insight into our core concept – martingale. In probability theory, the concept of martingale was proposed by Paul Pierre Levy. It is said that martingale was an inspiration by a French town called Martinique, whose residents were known for their penny-pinching. Levy established the original concept of martingale method based on the mathematical concept of the stingiest principle. However, the name of martingale was proposed by Joseph Leo Doob. During the initial stage, Doob made great contributions to the development of the theory of martingale. Part of his motivation for doing this is to show that a successful betting strategy is impossible. For more information about the historical origins of the word "martingale", please refer to http://www.jehps.net/juin2009/Mansuy.pdf. Since the 1970s, martingale theory has been widely used in many fields of pure mathematics and applied mathematics, especially in mathematical physics and financial mathematics.

图 1: The French town Martinique

Definition 3. *A process X is a* ***martingale*** *relative to the filtered space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbf{P})$*, if*

1. *X is adapted (to $(\mathcal{F}_n)$).*
2. *$E(|X_n| < \infty))$.*
3. *$E[X_n|\mathcal{F}_{n-1}] = X_{n-1}$, almost surely ($n \geq 1$).*

We can intuitively think of X_n as the unit "cost" or variable in our betting process. Take the gambler's example, we can refer to $X_n - X_{n-1}$ as the gain and loss of the gambler in a single bet.

The third requirement can be regarded as a description for "fair gamble", since we can learn that

$$\begin{aligned} E[X_n - X_{n-1}|\mathcal{F}_{n-1}] &= E[X_n|\mathcal{F}_{n-1}] - E[X_{n-1}|\mathcal{F}_{n-1}] \quad \text{(By linear property)} \\ &= X_{n-1} - X_{n-1} \quad (X_{n-1} \text{ is } \mathcal{F}_{n-1}\text{-measurable}) \\ &= 0 \end{aligned}$$

By changing the third requirement of martingale, we get the definition of supermartingale and submartingale.

Definition 4. *A **supermartingale** is defined the same as martingale only to replace the third requirement by*

$$E[X_n|\mathcal{F}_{n-1}] \leq X_{n-1}$$

*A **submartingale** is defined the same as martingale only to replace the third requirement by*

$$E[X_n|\mathcal{F}_{n-1}] \geq X_{n-1}$$

Here are some observations on the properties of martingale.

1. A supermartingale decreases on average while a submartingale increases on average.
2. A process X is a supermartingale if and only if $-X$ is a submartingale.
3. A process X is a martingale if and only if X is a supermartingale and a submartingale.
4. A process X is a martingale if and only if $X - X_0$ has the same property.

Exercise 5. *If X is a submartingale with finite means, show that $E(X_n) \geq E(X_0)$.*

Martingale is an important concept in probability theory, as it provides us with a tool to depict many processes. We can regard many things in the frame of martingales, and thus the properties of martingale will be of great help to us. For example, Kolmogrov's 0-1 law and the Strong Law, which we have already learned in the previous chapters, can be proved from the perspective of martingale. We will gain insight into them in the future. Now, let's turn to a famous example of martingale – Doob martingale, to see the power of martingale.

Example 6. ***Doob martingale***

Let Y be any random variable with $E[|Y|] < \infty$.
Suppose $\{\mathcal{F}_0, \mathcal{F}_1, \dots\}$ is a filtration, i.e. $\mathcal{F}_s \subset \mathcal{F}_t$ when $s < t$.
Define $Z_t = E[Y|\mathcal{F}_t]$, then $\{Z_0, Z_1, \dots\}$ is a martingale, called the doob martingale.

证明.

$$\begin{aligned}
E\left[Z_n \mid \mathcal{F}_{n-1}\right] &= E\left[E\left[Y \mid \mathcal{F}_n\right] \mid \mathcal{F}_{n-1}\right] \\
&= E\left[Y \mid \mathcal{F}_{n-1}\right] \quad \text{(Tower property)} \\
&= Z_{n-1}
\end{aligned}$$

□

With the knowledge of Doob martingale, we will then gain insights into a famous inequility – Azuma-Hoeffding inequility. The Azuma–Hoeffding inequality (named after Kazuoki Azuma and Wassily Hoeffding) gives a concentration result for the values of martingales that have bounded differences.

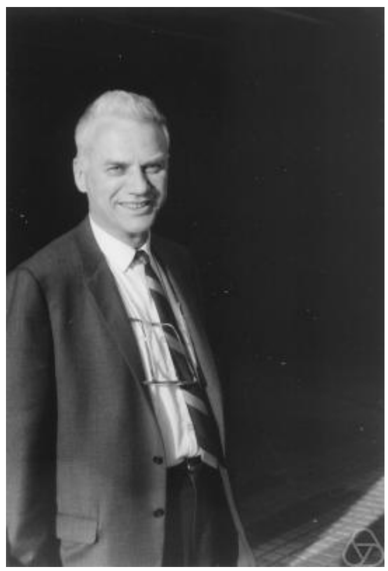

图 2: Joseph Doob

Theorem 7. ***Azuma-Hoeffding inequility***

$(X, \{\mathcal{F}_n\})$ is a martingale. Let $Y_i = X_i - X_{i-1}$ be the difference sequence. If there exist bounds $c_i > 0$ such that $|Y_i| \leq c_i$ for all i, then

$$P[|X_n - X_0| \geq \lambda] \leq \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^{n} c_i^2}\right)$$

证明. For any $t > 0$ we have

$$P\left[X_n - X_0 \geq \lambda\right] = P\left[e^{t(X_n - X_0)} \geq e^{\lambda t}\right]$$

Applying Markov's inequality and

$$P[e^{t(X_n-X_0)} \geq e^{\lambda t}] \leq e^{-\lambda t} E\left[e^{t(X_n-X_0)}\right]$$

By rewrite $X_n = X_{n-1} + Y_n$ and applying the tower property, we have

$$\begin{aligned} E\left[e^{t(X_n-X_0)}\right] &= E\left[e^{t(Y_n+X_{n-1}-X_0)}\right] \\ &= E\left[E\left[e^{t(Y_n+X_{n-1}-X_0)} \mid \mathcal{F}_{n-1}\right]\right] \end{aligned}$$

$E\left[t(X_{n-1}-X_0)\right]$ is constant under filter $\mathcal{F}_{n-1}$, so we can factor it out from the inner expectation and get

$$E\left[e^{t(Y_n+X_{n-1}-X_0)} \mid \mathcal{F}_{n-1}\right] = e^{t(X_{n-1}-X_0)} E\left[e^{tY_n} \mid \mathcal{F}_{n-1}\right]$$

Since $f(x) = e^{tx}$ is a convex function, we have

$$\forall x \in [-c, c], e^{tx} \leq \frac{(1-\frac{x}{c})e^{tc} + (1+\frac{x}{c})e^{tc}}{2}$$

Thus if a random variable Z has $E[Z] = 0$ and $|Z| \leq c$, then

$$\begin{aligned} E\left[e^{tZ}\right] &\leq E[\frac{(1-\frac{z}{c})e^{tc} + (1+\frac{z}{c})e^{tc}}{2}] \\ &= \frac{e^{tc}+e^{-tc}}{2} \\ &= \sum_{i=0}^{\infty} \frac{(tc)^{2i}}{(2i)!} \quad \text{(by Taylor's formula)} \\ &\leq \sum_{i=0}^{\infty} \frac{(tc)^{2i}}{2^i i!} \\ &= e^{\frac{(tc)^2}{2}} \end{aligned}$$

and thus we immediately apply the inequility to Y_n which has mean zero and bound c_n

$$e^{t(X_{n-1}-X_0)} E\left[e^{tY_n} \mid \mathcal{F}_{n-1}\right] \leq e^{t(X_{n-1}-X_0)} e^{t^2 c_n^2/2}$$

Now we combine all the equalities and inequalities above and derives

$$P[X_n - X_0 \geq \lambda] \leq e^{-\lambda t} \ e^{t^2 c_n^2/2} E\left[e^{t(X_{n-1}-X_0)}\right]$$

For the term $E[e^{t(X_{n-1}-X_0)}]$, we can apply the above process inductively, and this leads to

$$P[X_n - X_0 \geq \lambda] \leq exp\left(\frac{t^2}{2}\sum_{i=1}^{n} c_i^2 - \lambda t\right)$$

Now we choose t to be $\lambda / \sum c_i^2$ and get

$$P[X_n - X_0 \geq \lambda] \leq \exp(-\frac{\lambda^2}{2\sum_i c_i^2})$$

This completes our proof. $\square$

The Azuma-Hoeffding inequility has a wide range of usage. Take random walk on a line for instance. Let Y_i denote a step direction taken at time i, $Y_i \in \{-1, 1\}$, each with probability $1/2$. Let X_n denote the position at time n, *i.e.*

$$X_n = \sum_{i=1}^{n} Y_i$$

Since every step is bounded by 1, *i.e.*$|X_{k+1} - X_k| \leq 1$, then by the Azuma-Hoffding inequility we can learn that

$$P(|X_n - X_0| \geq \lambda) \leq 2e^{-\lambda^2/2n}$$

This tells us that the random walk process on a line is likely to converge.

Here are some more examples about the Azuma-Hoeffding inequility.

Exercise 8. ***Pattern Matching***

Consider a random string of characters $X = (X_1, \ldots, X_n)$. Each character I.I.D from a fixed alphabet Σ of size s. Investigate on the number of occurrence of a particular pattern $B = (b_1, \ldots, b_k)$ in the sequence X.

Exercise 9. ***Knapsack Problem***

Suppose you have n objects, the ith object having the volume V_i and worth W_i, where $V_1, \ldots, V_n$, $W_1, \ldots, W_n$ are independent non-negative random variables with finite means, and $W_i \leq M$ for all i of some fixed M. Your knapsack has a volume c, and you want to maximize the worth of things you pack in your knapsack. In mathematical language that is, find a vector $z_1, \ldots, z_n$ of 0'1 and 1's such that $\sum_{i=1}^n z_i V_i \leq c$ and maximizes $Z = \sum_{i=1}^n z_i W_i$. Show that the maximal value Z satisfies $P(|Z - E(Z)| \geq x) \leq 2exp(-x^2/(2nM^2))$ for $x > 0$.

Now we turn to a new topic about the previsible process. In a gamble, we make decisions only with the "information" of the past. From another perspective, our decisions at time n is previsible, since we have colleceted the information ahead of time n. Now we introduce the concept of previsible, and it is possible that we find out why martingale stands for a fair gamble.

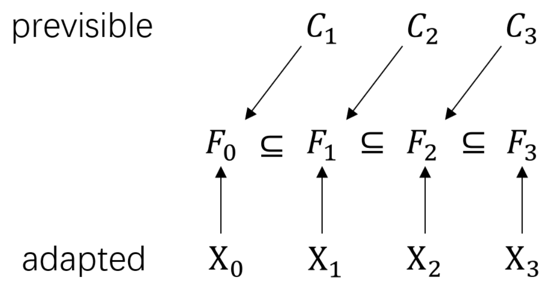

图 3: Previsible vs Adapted

Definition 10. *A process $C = (C_n : n \in \mathbb{N})$ is* ***previsible*** *if C_n is $\mathcal{F}_{n-1}$ measurable. $(n \geq 1)$*

Think of C_n as your stake in game n in the gamle. You make your decisions on a new stake by your previous bets up to time $n-1$. This is an intuitive meaning of previsible. Your gains on game n should be $C_n(X_n - X_{n-1})$ and thus your total winnings up to time n should be

$$Y_n = \sum_{i=1}^{n} C_i(X_i - X_{i-1}) =: (C \cdot X)_n$$

$C \cdot X$ is called the **martingale transform** of X by C. As a matter of fact, $C \cdot X$ is the discreate analogue of the stochastic integral $\int Cdx$, in that

$$\int_o^n CdX = (C \cdot X)_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1})$$

We now take the integration by parts, that is $C_iX_i \mid_0^n = \int_0^n CdX + \int_0^n XdC$. Now we rewrite the right hand side into discreate form. $\int_0^n CdX$ is $\sum_{i=0}^n C_{i+1}(X_{i+1} - X_i)$, as we have already defined. $\int_0^n XdC$ should be $\sum_{i=0}^{n-1} X_i(C_{i+1} - C_i)$ to make it an equality (we define $C_0 = 0$ here).

Stochastic calculus is a branch of mathematics that operates on stochastic processes. It allows a consistent theory of integration to be defined for integrals of stochastic processes with respect to stochastic processes. Stochastic integration is one of the greatest acomplishments in modern probability theory. To see more about the development of stochastic integral, refer to https://projecteuclid.org/download/pdf_1/euclid.lnms/1196285381.

What is probability? Doob, one of the most important foundators of probability theory has his answer. In his book *Stochastic Processes*, a book which gives a comprehensive treatment of stochastic processes, he clearly stated his theory about probabilities.

" ... probability theory is simply a branch of measure theory, with its own special emphasis and field of application ... "

We now introduce a theorem about the martingale transform. The proof of this theorem is easy, but it demonstrates the reason why we define the intergral in this way.

Theorem 11. *If C is a bounded previsible process, and X is a martingale, then $(C \cdot X)$ is a martingale null at* 0*.*

证明.

$$\begin{aligned} E\left[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}\right] &= E\left[C_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}\right] \\ &= C_n E\left[X_n - X_{n-1} \mid \mathcal{F}_{n-1}\right] \\ &= 0 \end{aligned}$$

□

The condition C is a bounded previsible process can be switched into some other intergrability condition.

Exercise 12. *If C is a bounded nonnegative previsible process, and X is a supermartingale, then $(C \cdot X)$ is a supermartingale null at* 0*.*

We can see from this theorem that, whatever the strategy you choose to play the gamble game, you are not able to change the system in that a supermartingale should still be a supermartingale while a martingale should remain a martingale.

Actually, Theorem 11 presents a way of constructing martingale. There are many ways to construct a martingale. Let's see the following theorem.

Theorem 13. *Suppose $\{X_t\}$ is a martingale relative to $\{\mathcal{F}_t\}$ and φ is a convex function on some interval. Then $\varphi(X_t) = Y_t$ is a submartingale relative to $(\mathcal{F})$ if $\varphi(X_t)$ is intergrable.*

证明.

$$\begin{aligned} \forall s < t,\ Y_s &= \varphi(X_s) \\ &= \varphi(E(X_t \mid \mathcal{F}_s)) \\ &\le E(\varphi(X_t) \mid \mathcal{F}_s) \quad \text{(by Jensen's inequality)} \\ &= E(Y_t \mid \mathcal{F}_s) \end{aligned}$$

□

Generally speaking, if we are going to make a judgement on whether a process Y is a martingale, one way is to use integration $Y = \int C dX$. If we find such previsible process C and martingale X, then we learn that Y is a martingale. However, there are so many previsible processes. We're going to introduce a new concept called stopping time, and focus on those special previsible processes.

Definition 14. ***stopping time (optimal time)***

A map $T : \Omega \to \{0, 1, 2, \ldots, \infty\}$ is called a stopping time if

$$\{T \leq n\} = \{\omega : T(\omega) \leq n\} \in \mathcal{F}_n, \forall n \leq \infty$$

or

$$\{T = n\} = \{\omega : T(\omega) = n\} \in \mathcal{F}_n, \forall n \leq \infty$$

The two conditions above are equivalent. Here's the proof.

证明. We denote these two conditions as (a) and (b), respectively.

(a) implies (b) since $\{T = n\} = \{T \leq n\} \backslash \{T \leq n-1\} \in \mathcal{F}_n$.

(b) implies (a) since $\forall k \leq n$, $\{T = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n$, and that

$$\{T \leq n\} = \bigcup_{k=0}^{n} \{T = k\} \in \mathcal{F}_n$$

□

Example 15. *(A_n) is an adapted process, and $B \in \mathcal{B}$.*

$T = \inf\{n \geq 0 : A_n \in B\}$, $\inf(\emptyset) = \infty$, is a stopping time.

$L = \sup\{n \leq 10 : A_n \in B\}$, $\inf(\emptyset) = 0$, is not a stopping time.

Exercise 16. *If T_1 and T_2 are stopping times with respect to a filtration $\mathcal{F}$. Show that $T_1 + T_2$, $\max\{T_1, T_2\}$, $\min\{T_1, T_2\}$ are all stopping times.*

In the end we will present a theorem about stopping time. The proof of the theorem relies on a delicious construction, without using

Theorem 17. ***Optional stopping theorem***

Let M be a martingale and let σ and τ be two stopping times such that $\sigma \leq \tau \leq K$ for some constant K. Then $E[M_\tau \mid \mathcal{F}_\sigma] = M_\sigma$ a.s.

If an adapted intergrable process M is a martingale, then $EM_\tau = EM_\sigma$ for all such τ and σ.

证明. $X_n = 1_{\tau \geq n} - 1_{\sigma \geq n}$ is a random variable. Since $\sigma \leq \tau$, $X_n \in \{0, 1\}$.

$X_n^{-1}(1) = \{\tau \leq n-1\}^C \cap \{\sigma \leq n-1\} \in \mathcal{F}_{n-1}$, so $X = (X_n)$ is a previsible process.

We denote M^τ to be (M_n^τ), where $M_n^\tau = M_{\tau \wedge n}$. By definition, we immediately learn that $X \cdot M = M^\tau - M^\sigma$.

Since X is an adapted provess and M is a martingale, $X \cdot M$ should also be a martingale, and thus

$$0 = E\left[(X \cdot M)_m\right] = E[M_n^\tau] - E[M_n^\sigma]$$

We now define truncated random time for all $A \in \mathcal{F}_\sigma$

$$\sigma^A = \sigma 1_A + K 1_{A^C}$$

$$\tau^A = \tau 1_A + K 1_{A^C}$$

Since $A \in \mathcal{F}_\sigma$, it can be verified that σ^A and τ^A are stopping times, and we still have $\sigma^A \leq \tau^A \leq K$. Similarly as the process above, we can learn that

$$E[M_n^{\tau^A}] = E[M_n^{\sigma^A}]$$

Rewrite the above equility into intergral form and we get

$$\int_A M_\sigma dP + \int_{A^C} M_K dP = \int_A M_\tau dP + \int_{A^C} M_K dP$$

Thus $\int_A M_\sigma dP = \int_A M_\tau dP$. This finishes our proof. □

This proof is a constructive proof. We construt previsible process X_n and truncated stopping time σ^A in the proof. A stronger conclusion can be found at chapter A14.3 in our textbook.

2 Reference

主要参考资料

《概率和鞅》戴维 • 威廉姆斯

《概率论题解 1000 例》G. 格里梅特，D. 斯特扎克

https://courses.cs.washington.edu/courses/cse525/13sp/scribe/lec18.pdf

https://en.wikipedia.org/wiki/Azuma's_inequality

https://people.eecs.berkeley.edu/s̃inclair/cs271/n18.pdf

https://ocw.mit.edu/courses/sloan-school-of-management/15-070j-advanced-stochastic-processes-fall-2013/lecture-notes/MIT15_070JF13_Lec12.pdf

https://en.wikipedia.org/wiki/Doob_martingale
https://projecteuclid.org/download/pdf_1/euclid.lnms/1196285381
https://en.wikipedia.org/wiki/Stochastic_calculus

图片来源
https://en.wikipedia.org/wiki/Martinique
https://en.wikipedia.org/wiki/Joseph_L._Doob#Honors

历史资料推荐
https://mathshistory.st-andrews.ac.uk/Biographies/Doob/
http://www.jehps.net/juin2009/Mansuy.pdf
https://projecteuclid.org/download/pdf_1/euclid.lnms/1196285381

具体使用方式见我的报告。