CHAPTER 4

Martingales

4.1. Conditional expectation

What's the best way to estimate the value of a random variable X? If we want to minimize the squared error

$$\mathbb{E}\left[(X-e)^2\right]=\mathbb{E}\left[X^2-2eX+e^2\right]=\mathbb{E}\left(X^2\right)-2e\mathbb{E}(X)+e^2,$$

differentiate to obtain $2e-2\mathbb{E}(X)$, which is zero at $e=\mathbb{E}(X)$.

EXAMPLE 1. (a) Your friend throws a die and you have to estimate its value X. According to the analysis above, your best bet is to guess $\mathbb{E}(X)=3.5$.

(b) What happens if you have additional information? Suppose that your friend tells you that X is an even number. How should you modify your guess to take this new information into account?

$$\begin{aligned}\mathbb{E}(X \mid X \text{ is even }) &= \sum_{x=1}^{6} x\mathbb{P}(X=x \mid X \text{ is even }) \\ &= (1\times 0)+\left(2\times\frac{1}{3}\right)+(3\times 0)+\left(4\times\frac{1}{3}\right)+(5\times 0)+\left(6\times\frac{1}{3}\right) \\ &= 4,\end{aligned}$$

so you'd change your guess to 4 . Similar calculations show $\mathbb{E}(X \mid X \text{ is odd })=3$.

We can combine these results as follows. We define the random variable

$$Y=\begin{cases}1, & X \text{ is even,}\\ 0, & X \text{ is odd,}\end{cases}$$

or write it in a more compact way as $Y = \mathbf{1}_{\{X \text{ is even}\}}$. Then the best estimate of X, given the informaion whether it is even or odd, is the random variable $3 + Y$.

(c) In an even more extreme case, your friend may tell you the exact result X. In that case your estimate will be X itself.

Information	Best estimate of X	
none	$\mathbb{E}(X \mid \text{no info }) = \phi$	$\phi \equiv \mathbb{E}(X)$
partial	$\mathbb{E}(X \mid Y) = \phi(Y)$	$\phi(y) = 3 + y$
complete	$\mathbb{E}(X \mid X) = \phi(X)$	$\phi(x) = x$

EXAMPLE 2. Suppose you roll two fair dice and let X be the number on the first die, and Y be the total on both dice. Calculate (a) $\mathbb{E}(Y \mid X)$ and (b) $\mathbb{E}(X \mid Y)$.

(a)

$$\begin{aligned}
\mathbb{E}(Y \mid X = x) &= \sum_y y\mathbb{P}(Y = y \mid X = x) \\
&= \sum_{w=1}^{6} (x + w)\frac{1}{6} = x + 3.5,
\end{aligned}$$

so that $\mathbb{E}(Y \mid X) = X + 3.5$. The variable w in the sum above stands for the value on the second die.

(b)

$$\begin{aligned}
\mathbb{E}(X \mid Y = y) &= \sum_x x\mathbb{P}(X = x \mid Y = y) \\
&= \sum_x x\frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \\
&= \sum_x x\frac{\mathbb{P}(X = x, Y - X = y - x)}{\mathbb{P}(Y = y)} \\
&= \sum_x x\frac{\mathbb{P}(X = x)\mathbb{P}(Y - X = y - x)}{\mathbb{P}(Y = y)}
\end{aligned}$$

Now

$$\mathbb{P}(X = x) = \frac{1}{6}, \quad 1 \le x \le 6$$

$$\mathbb{P}(Y = y) = \begin{cases} \frac{y-1}{36} & 2 \le y \le 7, \\ \frac{13-y}{36} & 8 \le y \le 12. \end{cases}$$

For $2 \leq y \leq 7$ we get

$$\mathbb{E}(X \mid Y = y) = \sum_{x=1}^{y-1} x \frac{1/36}{(y-1)/36} = \frac{1}{y-1} \sum_{x=1}^{y-1} x = \frac{1}{y-1} \frac{(y-1)y}{2} = \frac{y}{2}.$$

For $7 \leq y \leq 12$ we get

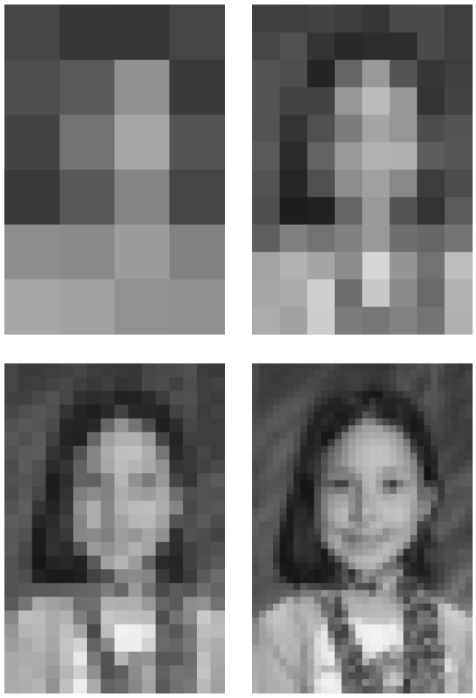
$$\mathbb{E}(X \mid Y = y) = \sum_{x=y-6}^{6} x \frac{1/36}{(13-y)/36} = \frac{1}{13-y} \sum_{x=y-6}^{6} x = \frac{y}{2}.$$

Therefore our best estimate is $\mathbb{E}(X \mid Y) = Y/2$.

If $X_1, X_2, \ldots$ is a sequence of random variables we will use $\mathcal{F}_n$ to denote "the information contained in $X_1, \ldots, X_n$ " and we will write $\mathbb{E}\left(Y \mid \mathcal{F}_n\right)$ for $\mathbb{E}\left(Y \mid X_1, \ldots, X_n\right)$. In the general case, when the $X_i s$ are not discrete, we have to rely on the following abstract definition.

Remark 3. Mathematically, the "the information contained in $X_1, \ldots, X_n$ " is identified with the σ-field generated by $X_1, \ldots, X_n$, i.e., $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. It seems hard to make the connection between these two concepts. However, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ does have the following properties:

(a) If we view the index n as the time, as time progresses, $\mathcal{F}_n$ is increasing, in the sense that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. This reflects the fact the information is accumulating as time goes forward. The increasing information flow can be mimicly explained by the following figures.



As we use more pixels to display the figure, the portrait of the girl is becoming more and more clear.

(b) Take a look at $\mathcal{F}_1 = \sigma(X_1)$. Then the events like

$$\{X_1 < 2\}, \{X_1 = 100\}, \{X_1 \geq 5\}, \cdots$$

are all contained in $\mathcal{F}_1$. Similarly, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ contains all events like

$$\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\}, \quad \text{where } A_1, \ldots, A_n \text{ are Borel sets.}$$

So $\mathcal{F}_n$ can be interpreted as the collection of all interesting events derived from $X_1, \ldots, X_n$ and thus reflects the information contained in these random variables.

Definition 4. $\mathbb{E}\,(Y \mid \mathcal{F}_n)$ is the unique random variable W satisfying the following two conditions:

(1) $W \in \mathcal{F}_n$, that is, $W = \phi\,(X_1, X_2, \ldots, X_n)$ for some function ϕ.

(2) If $A \in \mathcal{F}_n$, then $\mathbb{E}(Y \cdot \mathbf{1}_A) = \mathbb{E}(W \cdot \mathbf{1}_A)$.

4.2. Properties of conditional expectation

(1) (Law of total expectation) $\mathbb{E}\left(\mathbb{E}\left(Y \mid \mathcal{F}_n\right)\right) = \mathbb{E}(Y)$

(2) (Linearity) $\mathbb{E}\left(aY_1 + bY_2 \mid \mathcal{F}_n\right) = a\mathbb{E}\left(Y_1 \mid \mathcal{F}_n\right) + b\mathbb{E}\left(Y_2 \mid \mathcal{F}_n\right)$

(3) (Taking out what is known) If X is a function of $(X_1, X_2, \ldots, X_n)$, then $\mathbb{E}\left(XY \mid \mathcal{F}_n\right) = X\mathbb{E}\left(Y \mid \mathcal{F}_n\right)$

(4) (Tower property) For $m < n$, we have $\mathbb{E}\left(\mathbb{E}\left(Y \mid \mathcal{F}_n\right) \mid \mathcal{F}_m\right) = \mathbb{E}\left(Y \mid \mathcal{F}_m\right)$

(5) (Role of independence) If Y is independent of $\mathcal{F}_n$, then $\mathbb{E}\left(Y \mid \mathcal{F}_n\right) = \mathbb{E}(Y)$

EXAMPLE 5. Let $X_1, X_2, \ldots$ be independent random variables with mean μ and set $S_n = X_1 + X_2 + \cdots + X_n$. Let $\mathcal{F}_m = \sigma\left(X_1, X_2, \ldots, X_n\right)$ and $m < n$. Then

$$\begin{aligned}
\mathbb{E}\left(S_n \mid \mathcal{F}_m\right) &= \mathbb{E}\left(X_1 + \cdots + X_m \mid \mathcal{F}_m\right) + \mathbb{E}\left(X_{m+1} + \cdots + X_n \mid \mathcal{F}_m\right) \\
&= X_1 + \cdots + X_m + \mathbb{E}\left(X_{m+1} + \cdots + X_n\right) \\
&= S_m + (n-m)\mu.
\end{aligned}$$

Let's look at another example.

EXAMPLE 6. Let $X_1, X_2, \ldots$ be independent random variables with mean $\mu = 0$ and variance σ^2. Set $S_n = X_1 + X_2 + \cdots + X_n$. Let $\mathcal{F}_m = \sigma\left(X_1, \ldots, X_m\right)$ and $m < n$. Then

$$\begin{aligned}
\mathbb{E}\left(S_n^2 \mid \mathcal{F}_m\right) &= \mathbb{E}\left(\left(S_m + \left(S_n - S_m\right)\right)^2 \mid \mathcal{F}_m\right) \\
&= \mathbb{E}\left(S_m^2 + 2S_m\left(S_n - S_m\right) + \left(S_n - S_m\right)^2 \mid \mathcal{F}_m\right) \\
&= S_m^2 + 2S_m\mathbb{E}\left(S_n - S_m \mid \mathcal{F}_m\right) + \mathbb{E}\left(\left(S_n - S_m\right)^2\right) \\
&= S_m^2 + 0 + \operatorname{Var}\left(S_n - S_m\right) \\
&= S_m^2 + (n-m)\sigma^2.
\end{aligned}$$

4.3. Martingales

EXAMPLE 7. (Fair and unfair game) Suppose that we play the same game repetitively in a casino, and let ζ_j be our winning in the *j*-th round. Assume that (ζ_j) is an i.i.d sequence with $\mathbb{E}|\zeta_j| < \infty$. So

$$S_n := \sum_{j=1}^{n} \zeta_j$$

is our total wining after the *n*-th round. Set $S_0 := 0$, and let $\mathcal{F}_n := \sigma(\zeta_1, \ldots, \zeta_n)$ be the total information up to the *n*-th round. Then $\mathbb{E}(S_{n+k} \mid \mathcal{F}_n)$ is the predicted future average winning after the $(n+k)$-th round, based on current informaiton. Since

$$\begin{aligned}
\mathbb{E}(S_{n+k} \mid \mathcal{F}_n) &= \mathbb{E}(S_n + \zeta_{n+1} + \ldots + \zeta_{n+k} \mid \mathcal{F}_n) \\
&= \mathbb{E}(S_n \mid \mathcal{F}_n) + \mathbb{E}(\zeta_{n+1} + \ldots + \zeta_{n+k} \mid \mathcal{F}_n) \\
&= S_n + \mathbb{E}\zeta_{n+1} + \ldots + \mathbb{E}\zeta_{n+k} \\
&= S_n + k \cdot \mathbb{E}\zeta_1,
\end{aligned}$$

so

$$\mathbb{E}(S_{n+k} \mid \mathcal{F}_n) = S_n + k \cdot \mathbb{E}\zeta_1 \begin{cases} = S_n, & \text{if } \mathbb{E}\zeta_1 = 0 \qquad \text{(fair)} \\ \geq S_n, & \text{if } \mathbb{E}\zeta_1 \geq 0 \qquad \text{(advange to us)} \\ \leq S_n, & \text{if } \mathbb{E}\zeta_1 \leq 0 \qquad \text{(disadvange to us)} \end{cases}$$

A martingale is the model of a fair game. Let $X_0, X_1, \ldots$ be random variables and define $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ to be the "information in $X_0, X_1, \ldots, X_n$ ". The family $(\mathcal{F}_n)$ is then the “information flow” as time progresses, which is called the filtration generated by $X_0, X_1, \ldots$ A sequence $Y_0, Y_1, \ldots$ is said to be adapted to the filtration if $Y_n \in \mathcal{F}_n$ for every $n \geq 0$, i.e., $Y_n = \phi_n(X_0, \ldots, X_n)$ for some function ϕ_n.

Definition 8. The sequence $M_0, M_1, \ldots$ of random variables is called a martingale (with respect to $(\mathcal{F}_n)$) if

(a) $\mathbb{E}(|M_n|) < \infty$ for $n \geq 0$.

(b) (M_n) is adapted to $(\mathcal{F}_n)$.

(c) $\mathbb{E}(M_m \mid \mathcal{F}_n) = M_n$ for all $m \geq n \geq 0$.

Remark 9. In discrete-time, the condition (c) in Definition 6 can be replaced by

(c’) $\mathbb{E}(M_{n+1} \mid \mathcal{F}_n)$=$M_n$ for all $n \geq 0.$

In fact, if (c') holds, then $\mathbb{E}(M_{n+1} - M_n \mid \mathcal{F}_n) = 0$ for all n. Therefore, if $m \geq n$,

$$\begin{aligned}\mathbb{E}(M_m - M_n \mid \mathcal{F}_n) &= \mathbb{E}\left(\sum_{j=n}^{m-1} M_{j+1} - M_j \mid \mathcal{F}_n\right) \\ &= \mathbb{E}\left(\sum_{j=n}^{m-1} \mathbb{E}(M_{j+1} - M_j \mid \mathcal{F}_j) \mid \mathcal{F}_n\right) \\ &= 0\end{aligned}$$

so that $\mathbb{E}(M_m \mid \mathcal{F}_n) = M_n$.

Example 10. Let $X_1, X_2, \ldots$ be independent random variables with mean μ. Put $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. Then $M_n := S_n - n\mu$ is an $(\mathcal{F}_n)$-martingale.

Proof. It holds that

$$\mathbb{E}(M_{n+1} - M_n \mid \mathcal{F}_n) = \mathbb{E}(X_{n+1} - \mu \mid \mathcal{F}_n) = \mathbb{E}(X_{n+1} - \mu) = 0.$$

By Remark 9, (M_n) is a martingale. □

Example 11. Polya's urn.

Begin with an urn that holds two balls: one red and the other green. Draw a ball at random, then return it with another of the same colour.

Define X_n to be the number of red balls in the urn after n draws. Then

$$\begin{aligned}\mathbb{P}(X_{n+1} = k+1 \mid X_n = k) &= \frac{k}{n+2} \\ \mathbb{P}(X_{n+1} = k \mid X_n = k) &= 1 - \frac{k}{n+2}.\end{aligned}$$

This gives

$$\mathbb{E}(X_{n+1} \mid X_n = k) = (k+1)\frac{k}{n+2} + k\left(1 - \frac{k}{n+2}\right) = k\left(\frac{n+3}{n+2}\right),$$

so that $\mathbb{E}(X_{n+1} \mid X_n) = X_n(n+3)/(n+2)$. Note that (X_n) has the Markov property, that is, the bahavior of X_{n+1} depends only on X_n and

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n \left(\frac{n+3}{n+2}\right).$$

Therefore,

$$\mathbb{E}\left(\frac{X_{n+1}}{(n+1)+2} \mid \mathcal{F}_n\right) = \frac{X_n}{n+2}.$$

If we define $M_n = X_n/(n+2)$, then (M_n) is a martingale. Here M_n stands for the proportion of red balls in the urn after the *n*th draw.

4.4. Optional sampling theorem

Suppose that $(\mathcal{F}_n)$ is a filtration, which describes some information flow.

Definition 12. (Stopping time). A random time τ is called a stopping time if for any $n \geq 0$, whether τ has already happened (i.e., the event $\{\tau \leq n\}$) can be judged based on the information up to time n.

Example 13. Suppose some motorists are told to drive north on Highway 99 in Seattle and stop at the rst motorcycle shop past the second realtor after the city limits. So they drive north, pass the city limits, pass two realtors, and come to the next motorcycle shop, and stop. That is a stopping time. If they are instead told to stop at the third stop light before the city limits (and they had not been there before), they would need to drive to the city limits, then turn around and return past three stop lights. That is not a stopping time, because they have to go ahead of where they wanted to stop to know to stop there.

Theorem 14. *(Optional sampling theorem). If (M_n) is a martingale and τ a bounded stopping time, then*

$$\mathbb{E}(M_0) = \mathbb{E}(M_\tau).$$

Proof. Let k be the bound, i.e., $0 \leq \tau \leq k$. We prove the result by induction on k. If $k = 0$, then obviously the result is true.

Now suppose the result is true for $k-1$. Write

$$\mathbb{E}\left(X_0 - X_\tau\right) = \mathbb{E}\left(X_0 - X_{\tau\wedge(k-1)}\right) - \mathbb{E}\left(\left(X_k - X_{k-1}\right)\mathbf{1}_{\{\tau=k\}}\right).$$

The first term on the right is zero by the induction hypothesis. As for the second term, note that

$$\mathbf{1}_{\{\tau=k\}} = \mathbf{1}_{\{\tau\le k-1\}^c} \in \mathcal{F}_{k-1}.$$

Therefore

$$\mathbb{E}\left(\left(X_k - X_{k-1}\right)\mathbf{1}_{\{\tau=k\}}\right) = \mathbb{E}\left(\mathbb{E}\left(X_k - X_{k-1} \mid \mathcal{F}_{k-1}\right)\mathbf{1}_{\{\tau=k\}}\right) = 0,$$

which gives the result.

□

REMARK 15. The simple symmetric random walk S_n is a martingale, and $T := \inf(n \ge 1 : S_n = 1)$ is a stopping time with $\mathbb{P}(T < \infty) = 1$. However $\mathbb{E}\left(S_0\right) = 0 \neq \mathbb{E}\left(S_T\right) = 1$, so the optional sampling theorem may fail when T is not bounded.

In practice, we often encounter stopping time T that is not bounded. Don't worry. We have a standard trick to tackle this. We consider the truncated stopping time

$$T_n := T \wedge n = \min(T, n).$$

Then T_n is bounded by n and we can apply the optional sampling theorem for T_n. Then we let $n \to \infty$ and make use of $T_n \uparrow T$ to obtain the desired results for T. The next result deals with the typical situation where this trick works.

THEOREM 16. *Suppose M_n is a martingale and T a stopping time with $P(T < \infty) = 1$ and $|M_{T\wedge n}| \le K$ for some constant K, where $n \in \mathbb{N}$ is arbitrary. Then $\mathbb{E}M_T = \mathbb{E}M_0$.*

PROOF. See Theorem 5.11 in "Essentials of Stochastic Processes" by R. Durrett. □

4.5. Analysis of random walk using martingales

Let $X_1, X_2, \ldots$ be independent with $\mathbb{P}(X_i = -1) = q, \mathbb{P}(X = 1) = p$, where $p + q = 1$. Note that the mean and variance of X_i are given respectively by $\mu = p - q$ and $\sigma^2 = 1 - (p-q)^2$.

Let $S_0 = j$ and $S_n = S_0 + X_1 + \cdots + X_n$ and define the stopping time $T := \inf (n \geq 0 : S_n = 0$ or $S_n = N)$, where we assume that $0 \leq j \leq N$.

Case $(p = q)$. Since (S_n) is a martingale and $|S_{T \wedge n}| \leq N$ for all n, by Theorem 16, one gets

$$\begin{aligned} \mathbb{E}\left(S_0\right) &= \mathbb{E}\left(S_T\right) \\ j &= 0 \times \mathbb{P}\left(S_T = 0\right) + N \times \mathbb{P}\left(S_T = N\right), \end{aligned}$$

which implies that $\mathbb{P}\left(S_T = N\right) = j/N$.

Case $(p \neq q)$. We now use the same idea to treat the case where $p \neq q$.We introduce a martingale: $M_n = (q/p)^{S_n}$ (check that this really is a martingale!). We have seen earlier that $P(T < \infty) = 1$. Note also that

$$S_{T \wedge n} \leq \max\left\{(q/p)^0, (q/p)^N\right\}.$$

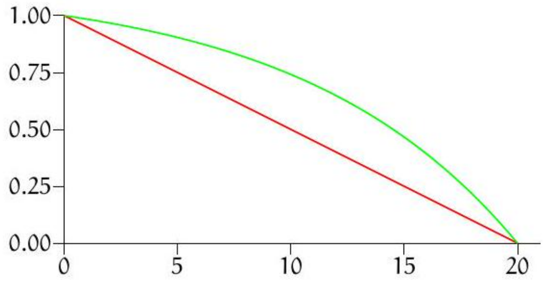
By Theorem 16,

$$\begin{aligned} \mathbb{E}\left(\left(\frac{q}{p}\right)^{S_0}\right) &= \mathbb{E}\left(\left(\frac{q}{p}\right)^{S_T}\right) \\ \left(\frac{q}{p}\right)^j &= \left(\frac{q}{p}\right)^0 \mathbb{P}\left(S_T = 0\right) + \left(\frac{q}{p}\right)^N \mathbb{P}\left(S_T = N\right) \\ \left(\frac{q}{p}\right)^j &= 1 - \mathbb{P}\left(S_T = N\right) + \left(\frac{q}{p}\right)^N \mathbb{P}\left(S_T = N\right). \end{aligned}$$

A little algebra now shows that

$$\mathbb{P}\left(S_T = N\right) = \frac{1 - (q/p)^j}{1 - (q/p)^N}.$$

Probability of ruin as a function of j, where $N = 20$: $p = 1/2$ (red) $\quad p = 9/19$ (green)


1.00
0.75
0.50
0.25
0.00
0
5
10
15
20