Martingale Sequences

1 Introduction

The conditional expectation is

$$\mathbb{E}[X \mid Y_1 = y_1, \dots, Y_k = y_k] = \sum_x X \cdot \Pr[X = x \mid Y_1 = y_1, \dots, Y_k = y_k].$$

With only one conditional variable,

$$\mathbb{E}[X \mid Y = y] = \frac{1}{\Pr[Y = y]} \cdot \sum_x \Pr[X = x \wedge Y = y].$$

Note that $\mathbb{E}[X \mid Y = y]$ is a random variable, mapping values y to $f(y) := \mathbb{E}[X \mid Y = y]$. We consistently use uppercase for random variables and lowercase for values they take on.

Recall that a random variable $X : \Omega \to \mathbb{R}$ is a function from points of the sample space to numerical labels. The conditional expectation is therefore a random variable (the expectation is just one numerical value, as is $\mathbb{E}[X \mid Y = y]$).

Definition 1. A (simple) martingale is a sequence $X_0, X_1, \dots$ of random variables such that for all i we have $\mathbb{E}[X_{i+1} \mid X_0, \dots, X_i] = X_i$. For a martingale, in expectation, the next value X_{i+1} is equal to X_i.

The initial motivation was to analyze sequences of bets with possibly complicated strategies, but each bet in isolation is fair. Intuitively, the betting history should not matter. If X_i is money after i bets, then $\mathbb{E}[X_{i+1} \mid X_i] = X_i$.

We would like to capture even revealed randomness that is not immediately reflected in the X_i.

Definition 2. A filter $\mathbb{F}$ is a finite sequence of $\mathbb{F}_i \subseteq 2^\Omega$ (*i.e.*, power set of Ω) such that:

1. Each $\mathbb{F}_i$ is closed under union and complement.
2. $\mathbb{F}_0 = \{\emptyset, \Omega\} \subseteq \mathbb{F}_1 \subseteq \cdots \subseteq \mathbb{F}_n = 2^\Omega$.

Intuitively, a filter gradually partitions the sample space more and more finely starting from the trivial partition and ending with the finest possible. This corresponds to gradually revealing the randomness in Ω until it has all been revealed. An event $\mathcal{E} \subseteq \Omega$, so $\mathcal{E} \in 2^\Omega$. Some events cannot be expressed with respect to some $\mathbb{F}_i$, namely when $\mathcal{E} \in \mathbb{F}_i$.

Example 3. Say we flip 2 coins and roll a 6-sided die. Then

$$\Omega = \{0,1\} \times \{0,1\} \times \{1,2,3,4,5,6\}.$$

If the die roll is revealed, then we can talk about any event that is only determined by the die roll, but we cannot express other events. In this case, the filter would be

$$\big\{\{0,1\} \times \{0,1\} \times \{i\} \mid i = 1, \dots, 6\big\}$$

and its closure under union and complement.

Definition 4. The conditional expectation $\mathbb{E}[X \mid \mathbb{F}_i]$ is equal to $\mathbb{E}[X \mid Y]$ for any random variable Y that takes distinct values on the elements of $\mathbb{F}_i$.

Here, Y is just used as an indicator for the partitions in $\mathbb{F}_i$. $\mathbb{E}[X \mid \mathbb{F}_i]$ assigns each partition as value the conditional expectation of X given that the outcome of the randomness is known to be in $\mathbb{F}_i$.

Definition 5. For a filter $\mathbb{F}_0, \mathbb{F}_1, \dots$, a sequence $X_0, X_1, \dots$ is a martingale sequence if for all i we have $\mathbb{E}[X_{i+1} \mid \mathbb{F}_i] = X_i$.

The most frequent use of martingales in computer science is the Doob martingale:

Definition 6. Let X be a random variable, and $\mathbb{F}_0, \mathbb{F}_1, \dots, \mathbb{F}_n$ a filter. Define $X_i \coloneqq \mathbb{E}[X \mid \mathbb{F}_i]$. This is a Doob martingale. (We skip the proof that it actually is a martingale).

Note that

$$X_0 = \mathbb{E}[X \mid \mathbb{F}_0] = \mathbb{E}[X \mid \text{no info revealed}] = \mathbb{E}[X].$$

Also,

$$X_n = \mathbb{E}[X \mid 2^{\Omega}] = \{\omega \mapsto X(\omega)\} = X.$$

The sequence (X_i) gradually reveals more and more information about X, starting with just the expectation, and finishing with the actual value.

Example 7. Consider a sum of independent random variables $Y_1, \dots, Y_n$. We care about $Y = \sum_{i=1}^n Y_i$. Define $X_i = \mathbb{E}[Y \mid Y_1, \dots, Y_i]$. The corresponding filter is

$$\mathbb{F}_i = 2^{\times_{j \le i} \{y : y \text{ is a possible outcome for } Y_j\}}$$

which reveals the Y_i one by one. Thus

$$\begin{aligned} X_0 &= \mathbb{E}[Y] = \sum_i \mathbb{E}[Y_i] \\ X_n &= Y = \sum_i Y_i. \end{aligned}$$

Let's briefly confirm that this is a martingale. We have

$$\begin{aligned} \mathbb{E}[X_{i+1} \mid \mathbb{F}_i] &= \mathbb{E}\big[\mathbb{E}[Y \mid Y_1, \dots, Y_{i+1}] \mid \mathbb{F}_i\big] \\ &= \mathbb{E}\big[\mathbb{E}[Y_1 + \dots + Y_{i+1} + \dots + Y_n \mid Y_1, \dots, Y_{i+1}] \mid \mathbb{F}_i\big] \\ &= \mathbb{E}\big[Y_1 + \dots + Y_{i+1} + \sum_{j > i+1} \mathbb{E}[Y_j] \mid \mathbb{F}_i\big] \\ &= \mathbb{E}\big[Y_1 + \dots + Y_i + \sum_{j \ge i+1} \mathbb{E}[Y_j] \mid \mathbb{F}_i\big] \\ &= \mathbb{E}\big[Y \mid Y_1, \dots, Y_i\big] \\ &= X_i. \end{aligned}$$

Example 8. [Edge Exposure Martingale]. Suppose we generate a random graph G by including each edge (u, v) independently with probability $p_{(u,v)}$. The well-known special case that $p_{(u,v)} = p$ for all (u, v) is called the Erdos-Renyi random graph $G(n, p)$.

Let $F(G)$ be some quantity we are interested in, such as largest clique size, chromatic number, number of components, etc. Consider any ordering of the $\binom{n}{2}$ candidate edges and define I_j to be the indicator for whether edge j is present.

Define $X_i = \mathbb{E}[F(G) \mid I_1, \dots, I_i]$ to be the expected value of $F(G)$ given that the first i edge candidates are known. This is a Doob martingale.

Example 9. [Vertex Exposure Martingale]. Reveal edges not one at a time, but one vertex at a time. We consider the vertices in some order $v_1, v_2, \ldots, v_n$ and at step i reveal all edges between v_i and $\{v_j : j < i\}$. Define $Y_i = \mathbb{E}[F(G) \mid \text{edges revealed on } \{v_1, \ldots, v_i\}]$. By ordering the edges appropriately (if $e = (v_i, v_j)$ with $i < j$, sort by increasing j) this forms a subsequence of the edge exposure martingale. By the following theorem, the vertex exposure martingale is a martingale.

Theorem 10. *If $X_0, X_1, \ldots$ is a martingale, and $\varphi : \mathbb{N} \to \mathbb{N}$ is any strictly increasing function, then $X_0, X_{\varphi(1)}, X_{\varphi(2)}, \ldots$ is also a martingale. In other words, any subsequence of a martingale is a martingale.*

The method of bounded differences says that martingale sequences that don't fluctuate too much from step to step have sharp tail bounds.

Theorem 11. *[Azuma's Inequality]. Let $X_0, X_1, \ldots, X_n$ be a finite martingale sequence with $|X_i - X_{i+1}| \le c_i$ for all i. Then for any $\Delta > 0$,*

$$\Pr[|X_n - X_0| \ge \Delta] \le 2 \exp\big(\frac{-\Delta^2}{2\sum_i c_i^2}\big).$$

Corollary 12. *If all $c_i = c$, then*

$$\Pr[|X_n - X_0| \ge \delta c\sqrt{n}] \le 2\exp(-\delta^2/2).$$

Example 13. Consider repeated gambling by betting on coin flips where you win \$1 if you guess right and lose \$1 if you guess wrong. The total money is a martingale. Then $X_0 = 0$, $c = 1$. So,

$$\Pr[|X_n| \ge \delta\sqrt{n}] \le 2\exp(-\delta^2/2).$$

With $\delta = 2\sqrt{\ln n}$, we get

$$\begin{aligned}\Pr[|X_n| \ge 2\sqrt{n \ln n}] &\le 2\exp(-2\ln n) \\ &= \frac{2}{n^2}.\end{aligned}$$

The most convenient way to apply Azuma's Inequality is to prove a Lipschitz condition.

Definition 14. A function $f : D_1 \times \cdots \times D_n \to \mathbb{R}$ satisfies a Lipschitz condition with c_i in its i^{th} argument if for all $x_1, \ldots x_n$ and x_i' we have

$$|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \le c_i.$$

Intuitively, changing one argument of f can change f by at most c_i. (Notice that the D_i have the trivial metric, so we can't phrase the Lipschitz condition in the usual way).

Corollary 15. *[of Azuma's Inequality]. Let f be a function that satisfies a Lipschitz condition with c_i in its i^{th} argument for each i. Let $X_1, \ldots, X_n$ be* ***independent*** *random variables. Let $Y_i = \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i]$ be the Doob martingale of f for $X_1, \ldots, X_n$. Then*

$$\Pr[|Y_n - Y_0| \ge \Delta] \le 2 \exp\big(\frac{-\Delta^2}{2\sum_i c_i^2}\big).$$

This is a strong generalization of Hoeffding bounds as it generalizes to any such f, not just the sum of independent variables as in Hoeffding.

Proof. We need to show that the Lipschitz condition implies the necessary condition for Azuma's Inequality. That is, for all i,

$$|Y_i - Y_{i-1}| \le c_i.$$

Then we have a Doob martingale with bounded differences.

$$
\begin{aligned}
|Y_i - Y_{i-1}| &= |\mathbb{E}[f(X_1,\ldots,X_n) \mid X_1,\ldots,X_i] - \mathbb{E}[f(X_1,\ldots,X_n) \mid X_1,\ldots,X_{i-1}]| \\
&= |\sum_{X_1,\ldots,X_n} f(X_1,\ldots,X_i,x_{i+1},\ldots,x_n)\Pr[X_{i+1}=x_{i+1},\ldots,X_n=x_n \mid X_1,\ldots,X_i] \\
&- \sum_{X_i,\ldots,X_n} f(X_1,\ldots,X_{i-1},x_i,\ldots,x_n)\Pr[X_i=x_i,\ldots,X_n=x_n \mid X_1,\ldots,X_{i+1}]| \\
&= |\sum_{X_1,\ldots,X_n} f(X_1,\ldots,X_i,x_{i+1},\ldots,x_n)\Pr[X_{i+1}=x_{i+1},\ldots,X_n=x_n] \\
&- \sum_{X_i,\ldots,X_n} f(X_1,\ldots,X_{i-1},x_i,\ldots,X_n)\Pr[X_{i+1}=x_{i+1},\ldots,X_n=x_n]| \\
&\le \sum_{X_1,\ldots,X_n} \Pr[X_{i+1}=x_{i+1},\ldots,X_n=x_n] \\
&\cdot |f(X_1,\ldots,X_i,x_{i+1},\ldots,x_n)\Pr[X_i=x_i]f(X_1,\ldots,X_i,x_{i+1},\ldots,x_n)| \\
&= \sum_{X_i,\ldots,X_n} \Pr[X_{i+1}=x_{i+1},\ldots,X_n=x_n] \\
&\cdot |\Pr[X_i=x_i f(X_1,\ldots,X_i,x_{i+1},\ldots,x_n) - \Pr[X_i=x_i]f(X_1,\ldots,X_{i-1},x_i,\ldots,x_n)| \\
&\le \sum_{X_i,\ldots,X_n} \Pr[X_i=x_i,\ldots,X_n=x_n]\cdot|f(X_1,\ldots,X_i,x_{i+1},\ldots,x_n) - f(X_1,\ldots,X_{i-1},x_i,\ldots,x_n)| \\
&\le \sum_{X_i,\ldots,X_n} \Pr[X_i=x_i,\ldots,X_n=x_n]\cdot c_i \\
&= c_i.
\end{aligned}
$$

$\square$