Chapter 20

Martingales II

By Sariel Har-Peled, March 19, 2024[1]

> "The Electric Monk was a labor-saving device, like a dishwasher or a video recorder. Dishwashers washed tedious dishes for you, thus saving you the bother of washing them yourself, video recorders watched tedious television for you, thus saving you the bother of looking at it yourself; Electric Monks believed things for you, thus saving you what was becoming an increasingly onerous task, that of believing all the things the world expected you to believe."
>
> Dirk Gently's Holistic Detective Agency, Douglas Adams

20.1. Filters and Martingales

Definition 20.1.1. A ***σ-field*** $(\Omega, \mathcal{F})$ consists of a sample space Ω (i.e., the atomic events) and a collection of subsets $\mathcal{F}$ satisfying the following conditions:
(A) $\emptyset \in \mathcal{F}$.
(B) $C \in \mathcal{F} \Rightarrow \overline{C} \in \mathcal{F}$.
(C) $C_1, C_2, \ldots \in \mathcal{F} \Rightarrow C_1 \cup C_2 \ldots \in \mathcal{F}$.

Definition 20.1.2. Given a σ-field $(\Omega, \mathcal{F})$, a ***probability measure*** $\mathbb{P} : \mathcal{F} \to \mathbb{R}^+$ is a function that satisfies the following conditions.
(A) $\forall A \in \mathcal{F}, 0 \leq \mathbb{P}[A] \leq 1$.
(B) $\mathbb{P}[\Omega] = 1$.
(C) For mutually disjoint events $C_1, C_2, \ldots$, we have $\mathbb{P}[\cup_i C_i] = \sum_i \mathbb{P}[C_i]$.

Definition 20.1.3. A ***probability space*** $(\Omega, \mathcal{F}, \mathbb{P})$ consists of a σ-field $(\Omega, \mathcal{F})$ with a probability measure $\mathbb{P}$ defined on it.

Definition 20.1.4. Given a σ-field $(\Omega, \mathcal{F})$ with $\mathcal{F} = 2^\Omega$, a ***filter*** (also ***filtration***) is a nested sequence $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$ of subsets of 2^Ω, such that:
(A) $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
(B) $\mathcal{F}_n = 2^\Omega$.
(C) For $0 \leq i \leq n$, $(\Omega, \mathcal{F}_i)$ is a σ-field.

Definition 20.1.5. An ***elementary event*** or ***atomic event*** is a subset of a sample space that contains only one element of Ω.

Intuitively, when we consider a probability space, we usually consider a random variable X. The value of X is a function of the elementary event that happens in the probability space. Formally, a random variable is a mapping $X : \Omega \to \mathbb{R}$. Thus, each $\mathcal{F}_i$ defines a partition of Ω into *atomic events*. This partition is getting more and more refined as we progress down the filter.

Example 20.1.6. Consider an algorithm **Alg** that uses n random bits. As such, the underlying sample space is $\Omega = \{b_1 b_2 \ldots b_n \mid b_1, \ldots, b_n \in \{0,1\}\}$. That is, the set of all binary strings of length n. Next, let $\mathcal{F}_i$ be the σ-field generated by the partition of Ω into the atomic events B_w, where $w \in \{0,1\}^i$; here w is the string encoding the first i random bits used by the algorithm. Specifically,

$$B_w = \left\{ wx \in \Omega \;\middle|\; x \in \{0,1\}^{n-i} \right\},$$

and the set of atomic events in $\mathcal{F}_i$ is $\mathcal{A}_i = \left\{ B_w \;\middle|\; w \in \{0,1\}^i \right\}$. The set $\mathcal{F}_i$ is the *closure* of this set of atomic events under complement and union. In particular, we conclude that $\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n$ form a filter.

As a concrete example, for $i = 3$, the set $\mathcal{A}_3$ contains $2^3 = 8$ sets, and the set $\mathcal{F}_3$ would contain all sets formed by finite unions of these sets (including the empty union). As such, the set $\mathcal{F}_3$ would have $2^{2^3} = 256$ sets.

Definition 20.1.7. A random variable X is said to be ***$\mathcal{F}_i$-measurable*** if for each $x \in \mathbb{R}$, the event $X \leq x$ is in $\mathcal{F}_i$; that is, the set $\{\omega \in \Omega \mid X(\omega) \leq x\}$ is in $\mathcal{F}_i$.

Example 20.1.8. Let $\mathcal{F}_0, \ldots, \mathcal{F}_n$ be the filter defined in Example 20.1.6. Let X be the parity of the n bits. Clearly, $X = 1$ is a valid event only in $\mathcal{F}_n$ (why?). Namely, it is only measurable in $\mathcal{F}_n$, but not in $\mathcal{F}_i$, for $i < n$.

As such, a random variable X is $\mathcal{F}_i$-measurable, only if it is a constant on the elementary events of $\mathcal{F}_i$. This gives us a new interpretation of what a filter is – its a sequence of refinements of the underlying probability space, that is achieved by splitting the atomic events of $\mathcal{F}_i$ into smaller atomic events in $\mathcal{F}_{i+1}$. Putting it explicitly, an atomic event $\mathcal{E}$ of $\mathcal{F}_i$, is a subset of 2^Σ. As we move to $\mathcal{F}_{i+1}$ the event $\mathcal{E}$ might now be split into several atomic (and disjoint events) $\mathcal{E}_1, \ldots, \mathcal{E}_k$. Now, naturally, the atomic event that really happens is an atomic event of $\mathcal{F}_n$. As we progress down the filter, we "zoom" into this event.

Definition 20.1.9 (Conditional expectation in a filter). Let $(\Omega, \mathcal{F})$ be any σ-field, and Y any random variable that takes on distinct values on the elementary events in $\mathcal{F}$. Then $\mathbb{E}[X \mid \mathcal{F}] = \mathbb{E}[X \mid Y]$.

20.2. Martingales

Definition 20.2.1. A sequence of random variables $Y_1, Y_2, \ldots,$ is a ***martingale difference*** sequence if for all $i \geq 0$, we have $\mathbb{E}\left[Y_i \;\middle|\; Y_1, \ldots, Y_{i-1}\right] = 0$.

Clearly, $X_1, \ldots,$ is a martingale sequence if and only if $Y_1, Y_2, \ldots,$ is a martingale difference sequence where $Y_i = X_i - X_{i-1}$.

Definition 20.2.2. A sequence of random variables $Y_1, Y_2, \ldots,$ is

$$\begin{aligned} &\text{a } \textbf{\textit{super martingale}} \text{ sequence if} \qquad \forall i \quad \mathbb{E}\left[Y_i \;\middle|\; Y_1, \ldots, Y_{i-1}\right] \leq Y_{i-1}, \\ &\text{and a } \textbf{\textit{sub martingale}} \text{ sequence if} \qquad \forall i \quad \mathbb{E}\left[Y_i \;\middle|\; Y_1, \ldots, Y_{i-1}\right] \geq Y_{i-1}. \end{aligned}$$

20.2.1. Martingales – an alternative definition

Definition 20.2.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filter $\mathcal{F}_0, \mathcal{F}_1, \ldots.$ Suppose that $X_0, X_1, \ldots,$ are random variables such that, for all $i \geq 0$, X_i is $\mathcal{F}_i$-measurable. The sequence $X_0, \ldots, X_n$ is a ***martingale*** provided that, for all $i \geq 0$, we have $\mathbb{E}[X_{i+1} \mid \mathcal{F}_i] = X_i$.

Lemma 20.2.4. *Let $(\Omega, \mathcal{F})$ and $(\Omega, \mathcal{G})$ be two σ-fields such that $\mathcal{F} \subseteq \mathcal{G}$. Then, for any random variable X, we have $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{F}] = \mathbb{E}[X \mid \mathcal{F}]$.*

Proof: $\mathbb{E}\Big[\mathbb{E}\Big[X \,\Big|\, \mathcal{G}\Big] \,\Big|\, \mathcal{F}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[X \,\Big|\, G = g\Big] \,\Big|\, F = f\Big]$

$$= \mathbb{E}\left[\frac{\sum_x x\,\mathbb{P}[X = x \cap G = g]}{\mathbb{P}[G = g]} \,\middle|\, F = f\right] = \sum_{g \in G} \frac{\frac{\sum_x x\mathbb{P}[X=x\cap G=g]}{\mathbb{P}[G=g]} \cdot \mathbb{P}[G = g \cap F = f]}{\mathbb{P}[F = f]}$$

$$= \sum_{g \in \mathcal{G}, g \subseteq f} \frac{\frac{\sum_x x\,\mathbb{P}[X=x\cap G=g]}{\mathbb{P}[G=g]} \cdot \mathbb{P}[G = g \cap F = f]}{\mathbb{P}[F = f]} = \sum_{g \in \mathcal{G}, g \subseteq f} \frac{\frac{\sum_x x\,\mathbb{P}[X=x\cap G=g]}{\mathbb{P}[G=g]} \cdot \mathbb{P}[G = g]}{\mathbb{P}[F = f]}$$

$$= \sum_{g \in \mathcal{G}, g \subseteq f} \frac{\sum_x x\,\mathbb{P}[X = x \cap G = g]}{\mathbb{P}[F = f]} = \frac{\sum_x x\left(\sum_{g \in \mathcal{G}, g \subseteq f} \mathbb{P}[X = x \cap G = g]\right)}{\mathbb{P}[F = f]}$$

$$= \frac{\sum_x x\,\mathbb{P}[X = x \cap F = f]}{\mathbb{P}[F = f]} = \mathbb{E}\Big[X \,\Big|\, \mathcal{F}\Big].$$

■

Theorem 20.2.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{F}_0, \ldots, \mathcal{F}_n$ be a filter with respect to it. Let X be any random variable over this probability space and define $X_i = \mathbb{E}\Big[X \,\Big|\, \mathcal{F}_i\Big]$ then, the sequence $X_0, \ldots, X_n$ is a martingale.*

Proof: We need to show that $\mathbb{E}\Big[X_{i+1} \,\Big|\, \mathcal{F}_i\Big] = X_i$. Namely,

$$\mathbb{E}[X_{i+1} \mid \mathcal{F}_i] = \mathbb{E}\Big[\mathbb{E}\Big[X \Big| \mathcal{F}_{i+1}\Big] \,\Big|\, \mathcal{F}_i\Big] = \mathbb{E}\Big[X \,\Big|\, \mathcal{F}_i\Big] = X_i,$$

by Lemma 20.2.4 and by definition of X_i. ■

Definition 20.2.6. Let $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathbb{R}$ be a real-valued function with a arguments from possibly distinct domains. The function f is said to satisfy the ***Lipschitz condition*** if for any $x_1 \in \mathcal{D}_1, \ldots, x_n \in \mathcal{D}_n$, and $i \in \{1, \ldots, n\}$ and any $y_i \in \mathcal{D}_i$, we have

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

Specifically, a function is ***c-Lipschitz***, if the inequality holds with a constant c (instead of 1).

Definition 20.2.7. Let $X_1, \ldots, X_n$ be a sequence of *independent* random variables, and a function $f = f(X_1, \ldots, X_n)$ defined over them, such that f satisfies the Lipschitz condition. The ***Doob martingale*** sequence $Y_0, \ldots, Y_m$ is defined by $Y_0 = \mathbb{E}[f(X_1, \ldots, X_n)]$ and

$$Y_i = \mathbb{E}\Big[f(X_1, \ldots, X_n) \,\Big|\, X_1, \ldots, X_i\Big], \qquad \text{for} \quad i = 1, \ldots, n.$$

Clearly, a Doob martingale $Y_0, \ldots, Y_n$ is a martingale, by Theorem 20.2.5. Furthermore, if $|X_i - X_{i-1}| \le 1$, for $i = 1, \ldots, n$, then $|Y_i - Y_{i-1}| \le 1$. and we can use Azuma's inequality on such a sequence.

20.3. Occupancy Revisited

We have m balls thrown independently and uniformly into n bins. Let Z denote the number of bins that remains empty in the end of the process. Let X_i be the bin chosen in the ith trial, and let $Z = F(X_1, \ldots, X_m)$, where F returns the number of empty bins given that m balls had thrown into bins $X_1, \ldots, X_m$. , By Azuma's inequality we have that $\mathbb{P}\left[\left|Z - \mathbb{E}[Z]\right| > \lambda\sqrt{m}\right] \leq 2\exp\left(-\lambda^2/2\right)$.

The following is an extension of Azuma's inequality shown in class. We do not provide a proof but it is similar to what we saw.

Theorem 20.3.1 (Azuma's Inequality - Stronger Form). *Let $X_0, X_1, \ldots,$ be a martingale sequence such that for each k, $|X_k - X_{k-1}| \leq c_k$, where c_k may depend on k. Then, for all $t \geq 0$, and any $\lambda > 0$, we have*

$$\mathbb{P}[|X_t - X_0| \geq \lambda] \leq 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^{t} c_k^2}\right).$$

Theorem 20.3.2. *Let $r = m/n$, and Z_{end} be the number of empty bins when m balls are thrown randomly into n bins. Then $\mu = \mathbb{E}[Z_{end}] = n\left(1 - \frac{1}{n}\right)^m \approx n\exp(-r)$, and for any $\lambda > 0$, we have*

$$\mathbb{P}\left[\left|Z_{end} - \mu\right| \geq \lambda\right] \leq 2\exp\left(-\frac{\lambda^2(n - 1/2)}{n^2 - \mu^2}\right).$$

Proof: Let $z(Y, t)$ be the expected number of empty bins in the end, if there are Y empty bins in time t. The probability of an empty bin to remain empty is $(1 - 1/n)^{m-t}$, and as such

$$z(Y, t) = Y\left(1 - \frac{1}{n}\right)^{m-t}.$$

In particular, $\mu = z(n, 0) = n(1 - 1/n)^m$.

Let $\mathcal{F}_t$ be the σ-field generated by the bins chosen in the first t steps. Let Z_{end} be the number of empty bins at time m, and let $Z_t = \mathbb{E}\left[Z_{\text{end}} \,\middle|\, \mathcal{F}_t\right]$. Namely, Z_t is the expected number of empty bins after we know where the first t balls had been placed. The random variables $Z_0, Z_1, \ldots, Z_m$ form a martingale. Let Y_t be the number of empty bins after t balls where thrown. We have $Z_{t-1} = z(Y_{t-1}, t - 1)$. Consider the ball thrown in the t-step. Clearly:

(A) With probability $1 - Y_{t-1}/n$ the ball falls into a non-empty bin. Then $Y_t = Y_{t-1}$, and $Z_t = z(Y_{t-1}, t)$. Thus,

$$\begin{aligned}\Delta_t = Z_t - Z_{t-1} &= z(Y_{t-1}, t) - z(Y_{t-1}, t-1) = Y_{t-1}\left(\left(1 - \frac{1}{n}\right)^{m-t} - \left(1 - \frac{1}{n}\right)^{m-t+1}\right)\\ &= \frac{Y_{t-1}}{n}\left(1 - \frac{1}{n}\right)^{m-t} \leq \left(1 - \frac{1}{n}\right)^{m-t}.\end{aligned}$$

(B) Otherwise, with probability Y_{t-1}/n the ball falls into an empty bin, and $Y_t = Y_{t-1} - 1$. Namely, $Z_t = z(Y_t - 1, t)$. And we have that

$$\begin{aligned}\Delta_t = Z_t - Z_{t-1} &= z(Y_{t-1} - 1, t) - z(Y_{t-1}, t-1) = (Y_{t-1} - 1)\left(1 - \frac{1}{n}\right)^{m-t} - Y_{t-1}\left(1 - \frac{1}{n}\right)^{m-t+1}\\ &= \left(1 - \frac{1}{n}\right)^{m-t}\left(Y_{t-1} - 1 - Y_{t-1}\left(1 - \frac{1}{n}\right)\right) = \left(1 - \frac{1}{n}\right)^{m-t}\left(-1 + \frac{Y_{t-1}}{n}\right) = -\left(1 - \frac{1}{n}\right)^{m-t}\left(1 - \frac{Y_{t-1}}{n}\right)\\ &\geq -\left(1 - \frac{1}{n}\right)^{m-t}.\end{aligned}$$

Thus, $Z_0, \ldots, Z_m$ is a martingale sequence, where $|Z_t - Z_{t-1}| \leq |\Delta_t| \leq c_t$, where $c_t = \left(1 - \frac{1}{n}\right)^{m-t}$. We have

$$\sum_{t=1}^{m} c_t^2 = \sum_{t=1}^{m} \left(1 - \frac{1}{n}\right)^{2(m-t)} = \sum_{t=0}^{m-1} \left(1 - \frac{1}{n}\right)^{2t} = \frac{1 - (1 - 1/n)^{2m}}{1 - (1 - 1/n)^2} = \frac{n^2\left(1 - (1 - 1/n)^{2m}\right)}{2n - 1} = \frac{n^2 - \mu^2}{2n - 1}.$$

Now, deploying Azuma's inequality, yield the result. ■

20.3.1. Lets verify this is indeed an improvement

Consider the case where $m = n \ln n$. Then, $\mu = n\left(1 - \frac{1}{n}\right)^m \leq 1$. And using the "weak" Azuma's inequality implies that

$$\mathbb{P}\left[\left|Z_{\mathrm{end}} - \mu\right| \geq \lambda \sqrt{n}\right] = \mathbb{P}\left[\left|Z_{\mathrm{end}} - \mu\right| \geq \lambda \sqrt{\frac{n}{m}} \sqrt{m}\right] \leq 2 \exp\left(-\frac{\lambda^2 n}{2m}\right) = 2 \exp\left(-\frac{\lambda^2}{2 \ln n}\right),$$

which is interesting only if $\lambda > \sqrt{2 \ln n}$. On the other hand, Theorem 20.3.2 implies that

$$\mathbb{P}\left[|Z_{\mathrm{end}} - \mu| \geq \lambda \sqrt{n}\right] \leq 2 \exp\left(-\frac{\lambda^2 n(n - 1/2)}{n^2 - \mu^2}\right) \leq 2 \exp\left(-\lambda^2\right),$$

which is interesting for any $\lambda \geq 1$ (say).