

# Chapter 9

## Martingales II

By Sarel Har-Peled, December 30, 2015<sup>①</sup>

“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..

### 9.1. Filters and Martingales

**Definition 9.1.1.** A  $\sigma$ -field  $(\Omega, \mathcal{F})$  consists of a sample space  $\Omega$  (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)  $C_1, C_2, \dots \in \mathcal{F} \Rightarrow C_1 \cup C_2 \dots \in \mathcal{F}$ .

**Definition 9.1.2.** Given a  $\sigma$ -field  $(\Omega, \mathcal{F})$ , a *probability measure*  $\mathbf{Pr} : \mathcal{F} \rightarrow \mathbb{R}^+$  is a function that satisfies the following conditions.

- (A)  $\forall A \in \mathcal{F}, 0 \leq \mathbf{Pr}[A] \leq 1$ .
- (B)  $\mathbf{Pr}[\Omega] = 1$ .
- (C) For mutually disjoint events  $C_1, C_2, \dots$ , we have  $\mathbf{Pr}[\cup_i C_i] = \sum_i \mathbf{Pr}[C_i]$ .

**Definition 9.1.3.** A *probability space*  $(\Omega, \mathcal{F}, \mathbf{Pr})$  consists of a  $\sigma$ -field  $(\Omega, \mathcal{F})$  with a probability measure  $\mathbf{Pr}$  defined on it.

**Definition 9.1.4.** Given a  $\sigma$ -field  $(\Omega, \mathcal{F})$  with  $\mathcal{F} = 2^\Omega$ , a *filter* (also *filtration*) is a nested sequence  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n$  of subsets of  $2^\Omega$ , 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  $\sigma$ -field.

---

<sup>①</sup>This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

**Definition 9.1.5.** An *elementary event* or *atomic event* is a subset of a sample space that contains only one element of  $\Omega$ .

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  $X : \Omega \rightarrow \mathbb{R}$ . Thus, each  $\mathcal{F}_i$  defines a partition of  $\Omega$  into *atomic events*. This partition is getting more and more refined as we progress down the filter.

**Example 9.1.6.** Consider an algorithm **Alg** that uses  $n$  random bits. As such, the underlying sample space is  $\Omega = \{b_1 b_2 \dots b_n \mid b_1, \dots, b_n \in \{0, 1\}\}$ ; that is, the set of all binary strings of length  $n$ . Next, let  $\mathcal{F}_i$  be the  $\sigma$ -field generated by the partition of  $\Omega$  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,

$$B_w = \{wx \mid x \in \{0, 1\}^{n-i}\},$$

and the set of atomic events in  $\mathcal{F}_i$  is  $\{B_w \mid w \in \{0, 1\}^i\}$ . The set  $\mathcal{F}_i$  is the closure of this set of atomic events under complement and union. In particular, we conclude that  $\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_n$  form a filter.

**Definition 9.1.7.** 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 9.1.8.** Let  $\mathcal{F}_0, \dots, \mathcal{F}_n$  be the filter defined in **Example 9.1.6**. 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  $2^\Sigma$ . As we move to  $\mathcal{F}_{i+1}$  the event  $\mathcal{E}$  might now be split into several atomic (and disjoint events)  $\mathcal{E}_1, \dots, \mathcal{E}_k$ . 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 9.1.9 (Conditional expectation in a filter).** Let  $(\Omega, \mathcal{F})$  be any  $\sigma$ -field, and  $Y$  any random variable that takes on distinct values on the elementary events in  $\mathcal{F}$ . Then  $\mathbf{E}[X|\mathcal{F}] = \mathbf{E}[X|Y]$ .

## 9.2. Martingales

**Definition 9.2.1.** A sequence of random variables  $Y_1, Y_2, \dots$ , is said to be a *martingale difference* sequence if for all  $i \geq 0$ , we have  $\mathbf{E}[Y_i \mid Y_1, \dots, Y_{i-1}] = 0$ .

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

**Definition 9.2.2.** A sequence of random variables  $Y_1, Y_2, \dots$ , is

$$\begin{aligned} & \text{a } \textit{super martingale} \text{ sequence if} & \forall i & \mathbf{E}[Y_i \mid Y_1, \dots, Y_{i-1}] \leq Y_{i-1}, \\ & \text{and a } \textit{sub martingale} \text{ sequence if} & \forall i & \mathbf{E}[Y_i \mid Y_1, \dots, Y_{i-1}] \geq Y_{i-1}. \end{aligned}$$

### 9.2.1. Martingales – an alternative definition

**Definition 9.2.3.** Let  $(\Omega, \mathcal{F}, \mathbf{Pr})$  be a probability space with a filter  $\mathcal{F}_0, \mathcal{F}_1, \dots$ . Suppose that  $X_0, X_1, \dots$ , are random variables such that, for all  $i \geq 0$ ,  $X_i$  is  $\mathcal{F}_i$ -measurable. The sequence  $X_0, \dots, X_n$  is a *martingale* provided that, for all  $i \geq 0$ , we have  $\mathbf{E}[X_{i+1} | \mathcal{F}_i] = X_i$ .

**Lemma 9.2.4.** Let  $(\Omega, \mathcal{F})$  and  $(\Omega, \mathcal{G})$  be two  $\sigma$ -fields such that  $\mathcal{F} \subseteq \mathcal{G}$ . Then, for any random variable  $X$ ,  $\mathbf{E}[\mathbf{E}[X | \mathcal{G}] | \mathcal{F}] = \mathbf{E}[X | \mathcal{F}]$ .

*Proof:*  $\mathbf{E}[\mathbf{E}[X | \mathcal{G}] | \mathcal{F}] = \mathbf{E}[\mathbf{E}[X | G = g] | F = f]$

$$\begin{aligned} &= \mathbf{E}\left[\frac{\sum_x x \mathbf{Pr}[X = x \cap G = g]}{\mathbf{Pr}[G = g]} \mid F = f\right] = \sum_{g \in \mathcal{G}} \frac{\frac{\sum_x x \mathbf{Pr}[X = x \cap G = g]}{\mathbf{Pr}[G = g]} \cdot \mathbf{Pr}[G = g \cap F = f]}{\mathbf{Pr}[F = f]} \\ &= \sum_{g \in \mathcal{G}, g \subseteq f} \frac{\frac{\sum_x x \mathbf{Pr}[X = x \cap G = g]}{\mathbf{Pr}[G = g]} \cdot \mathbf{Pr}[G = g \cap F = f]}{\mathbf{Pr}[F = f]} = \sum_{g \in \mathcal{G}, g \subseteq f} \frac{\frac{\sum_x x \mathbf{Pr}[X = x \cap G = g]}{\mathbf{Pr}[G = g]} \cdot \mathbf{Pr}[G = g]}{\mathbf{Pr}[F = f]} \\ &= \sum_{g \in \mathcal{G}, g \subseteq f} \frac{\sum_x x \mathbf{Pr}[X = x \cap G = g]}{\mathbf{Pr}[F = f]} = \frac{\sum_x x (\sum_{g \in \mathcal{G}, g \subseteq f} \mathbf{Pr}[X = x \cap G = g])}{\mathbf{Pr}[F = f]} \\ &= \frac{\sum_x x \mathbf{Pr}[X = x \cap F = f]}{\mathbf{Pr}[F = f]} = \mathbf{E}[X | \mathcal{F}]. \quad \blacksquare \end{aligned}$$

**Theorem 9.2.5.** Let  $(\Omega, \mathcal{F}, \mathbf{Pr})$  be a probability space, and let  $\mathcal{F}_0, \dots, \mathcal{F}_n$  be a filter with respect to it. Let  $X$  be any random variable over this probability space and define  $X_i = \mathbf{E}[X | \mathcal{F}_i]$  then, the sequence  $X_0, \dots, X_n$  is a martingale.

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

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

by [Lemma 9.2.4](#) and by definition of  $X_i$ . \blacksquare

**Definition 9.2.6.** Let  $f : \mathcal{D}_1 \times \dots \times \mathcal{D}_n \rightarrow \mathbb{R}$  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  $x_1 \in \mathcal{D}_1, \dots, x_n \in \mathcal{D}_n$ , and  $i \in \{1, \dots, n\}$  and any  $y_i \in \mathcal{D}_i$ ,

$$\left| f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n) \right| \leq 1.$$

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

**Definition 9.2.7.** Let  $X_1, \dots, X_n$  be a sequence of *independent* random variables, and a function  $f(X_1, \dots, X_n)$  defined over them that such that  $f$  satisfies the Lipschitz condition. The *Doob martingale* sequence  $Y_0, \dots, Y_n$  is defined by  $Y_0 = \mathbf{E}[f(X_1, \dots, X_n)]$  and  $Y_i = \mathbf{E}[f(X_1, \dots, X_n) | X_1, \dots, X_i]$ , for  $i = 1, \dots, n$ .

Clearly, a Doob martingale  $Y_0, \dots, Y_n$  is a martingale, by [Theorem 9.2.5](#). Furthermore, if  $|X_i - X_{i-1}| \leq 1$ , for  $i = 1, \dots, n$ , then  $|Y_i - Y_{i-1}| \leq 1$ . and we can use Azuma's inequality on such a sequence.

### 9.3. 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  $i$ th trial, and let  $Z = F(X_1, \dots, X_m)$ , where  $F$  returns the number of empty bins given that  $m$  balls had thrown into bins  $X_1, \dots, X_m$ . Clearly, we have by Azuma's inequality that  $\Pr[|Z - \mathbf{E}[Z]| > \lambda \sqrt{m}] \leq 2e^{-\lambda^2/2}$ .

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 9.3.1 (Azuma's Inequality - Stronger Form).** *Let  $X_0, X_1, \dots$ , 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  $\lambda > 0$ ,*

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

**Theorem 9.3.2.** *Let  $r = m/n$ , and  $Z_{\text{end}}$  be the number of empty bins when  $m$  balls are thrown randomly into  $n$  bins. Then  $\mu = \mathbf{E}[Z_{\text{end}}] = n\left(1 - \frac{1}{n}\right)^m \approx ne^{-r}$ , and for any  $\lambda > 0$ , we have*

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

*Proof:* Let  $z(Y, t)$  be the expected number of empty bins, if there are  $Y$  empty bins in time  $t$ . Clearly,

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

In particular,  $\mu = z(n, 0) = n\left(1 - \frac{1}{n}\right)^m$ .

Let  $\mathcal{F}_t$  be the  $\sigma$ -field generated by the bins chosen in the first  $t$  steps. Let  $Z_{\text{end}}$  be the number of empty bins at time  $m$ , and let  $Z_t = \mathbf{E}[Z_{\text{end}} | \mathcal{F}_t]$ . Namely,  $Z_t$  is the expected number of empty bins after we know where the first  $t$  balls had been placed. The random variables  $Z_0, Z_1, \dots, Z_m$  form a martingale. Let  $Y_t$  be the number of empty bins after  $t$  balls were 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,

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

(B) Otherwise, with probability  $Y_{t-1}/n$  the ball falls into an empty bin, and  $Y_t = Y_{t-1} - 1$ . Namely,  $Z_t = z(Y_{t-1} - 1, t)$ . 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,  $Z_0, \dots, Z_m$  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

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

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

### 9.3.1. 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

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

which is interesting only if  $\lambda > \sqrt{2 \ln n}$ . On the other hand, [Theorem 9.3.2](#) implies that

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

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

## 9.4. Some useful estimates

**Lemma 9.4.1.** *For any  $n \geq 2$ , and  $m \geq 1$ , we have that  $(1 - 1/n)^m \geq 1 - m/n$ .*

*Proof:* Follows by induction. Indeed, for  $m = 1$  the claim is immediate. For  $m \geq 2$ , we have

$$\left(1 - \frac{1}{n}\right)^m = \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n}\right)^{m-1} \geq \left(1 - \frac{1}{n}\right) \left(1 - \frac{m-1}{n}\right) \geq 1 - \frac{m}{n}. \quad \blacksquare$$

This implies the following.

**Lemma 9.4.2.** *For any  $m \leq n$ , we have that  $1 - m/n \leq (1 - 1/n)^m \leq \exp(-m/n)$ .*