9: Martingales II

CS 598shp - Randomized Algorithms - Fall 2005

October 13, 2005

"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.

1 Filters and Martingales

Definition 1.1 Given a σ-field $(\Omega, \mathbb{F})$ with $\mathbb{F} = 2^\Omega$, a *filter* (also *filtration*) is a nested sequence $\mathbb{F}_0 \subseteq \mathbb{F}_1 \subseteq \cdots \subseteq \mathbb{F}_n$ of subsets of 2^Ω such that

1. $\mathbb{F}_0 = \{\emptyset, \Omega\}$.
2. $\mathbb{F}_n = 2^\Omega$.
3. For $0 \leq i \leq n$, $(\Omega, \mathbb{F}_i)$ is a σ-field.

Intuitively, each $\mathbb{F}_i$ define a partition of Ω into *blocks*. This partition is getting more and more refined as we progress with the filter.

Example 1.2 Consider an algorithm A that uses n random bits, and let $\mathbb{F}_i$ be the σ-field generated by the partition of Ω into the blocks B_w, where $w \in \{0,1\}^i$. Then $\mathbb{F}_0, \mathbb{F}_1, \ldots, \mathbb{F}_n$ form a filter.

Definition 1.3 A random variable X is said to be $\mathbb{F}_i$-*measurable* if for each $x \in \mathbb{R}$, the event $\{X \leq x\}$ is contained in $\mathbb{F}_i$.

Example 1.4 Let $\mathbb{F}_0, \ldots, \mathbb{F}_n$ be the filter defined in Example 1.2. Let X be the parity of the n bits. Clearly, X is a valid event only in $\mathbb{F}_n$ (why?). Namely, it is only measurable in $\mathbb{F}_n$, but not in $\mathbb{F}_i$, for $i < n$.

Namely, a random variable X is $\mathbb{F}_i$-measurable, only if it is a constant on the blocks of $\mathbb{F}_i$.

Definition 1.5 Let $(\Omega, \mathbb{F})$ be any σ-field, and Y any random variable that takes on distinct values on the elementary elements in $\mathbb{F}$. Then $\mathbf{E}\left[X \,\middle|\, \mathbb{F}\right] = \mathbf{E}\left[X \,\middle|\, Y\right]$.

2 Martingales

Definition 2.1 A sequence of random variables $Y_1, Y_2, \ldots,$ is said to be a *martingale difference* sequence if for all $i \geq 0$,

$$\mathbf{E}\Big[Y_i \,\Big|\, Y_1, \ldots, Y_{i-1}\Big] = 0.$$

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

Definition 2.2 A sequence of random variables $Y_1, Y_2, \ldots,$ is said to be a *super martingale* sequence if for all $i \geq$,

$$\mathbf{E}\Big[Y_i \,\Big|\, Y_1, \ldots, Y_{i-1}\Big] \leq Y_{i-1},$$

and a *sub martingale* sequence if

$$\mathbf{E}\Big[Y_i \,\Big|\, Y_1, \ldots, Y_{i-1}\Big] \geq Y_{i-1}.$$

Example 2.3 Let U be a urn with b black balls, and w white balls. We repeatedly select a ball and replace it by c balls having the same color. Let X_i be the fraction of black balls after the first i trials. This sequence is a martingale.

Indeed, let $n_i = b + w + i(c-1)$ be the number of balls in the urn after the ith trial. Clearly,

$$\begin{aligned}
\mathbf{E}\Big[X_i \,\Big|\, X_{i-1}, \ldots, X_0\Big] &= X_{i-1} \cdot \frac{(c-1) + X_{i-1} n_{i-1}}{n_i} + (1 - X_{i-1}) \cdot \frac{X_{i-1} n_{i-1}}{n_i} \\
&= \frac{X_{i-1}(c-1) + X_{i-1} n_{i-1}}{n_i} = X_{i-1} \frac{c - 1 + n_{i-1}}{n_i} = X_{i-1} \frac{n_i}{n_i} = X_{i-1}.
\end{aligned}$$

2.1 Martingales, an alternative definition

Definition 2.4 Let $(\Omega, \mathbb{F}, \mathbf{Pr})$ be a probability space with a filter $\mathbb{F}_0, \mathbb{F}_1, \ldots$. Suppose that $X_0, X_1, \ldots$, are random variables such that for all $i \geq 0$, X_i is $\mathbb{F}_i$-measurable. The sequence $X_0, \ldots, kX_n$ is a martingale provided, for all $i \geq 0$,

$$\mathbf{E}\Big[X_{i+1} \,\Big|\, \mathbb{F}_i\Big] = X_i.$$

Lemma 2.5 *Let $(\Omega, \mathbb{F})$ and $(\Omega, \mathbb{G})$ be two σ-fields such that $\mathbb{F} \subseteq \mathbb{G}$. Then, for any random variable X, $\mathbf{E}\Big[\mathbf{E}\Big[X \,\Big|\, \mathbb{G}\Big] \,\Big|\, \mathbb{F}\Big] = \mathbf{E}\Big[X \,\Big|\, \mathbb{F}\Big]$.*

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

$$= \mathbf{E}\left[\frac{\sum_x x \,\mathbf{Pr}[X = x \cap G = g]}{\mathbf{Pr}[G = g]} \;\middle|\; F = f\right]$$

$$= \sum_{g \in G} \frac{\frac{\sum_x x \,\mathbf{Pr}[X = x \cap G = g]}{\mathbf{Pr}[G = g]} \cdot \mathbf{Pr}[G = g \cap F = f]}{\mathbf{Pr}[F = f]}$$

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

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

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

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

$$= \frac{\sum_x x \,\mathbf{Pr}[X = x \cap F = f]}{\mathbf{Pr}[F = f]}$$

$$= \mathbf{E}\Big[X \;\Big|\; \mathbb{F}\Big].$$

■

Theorem 2.6 *Let $(\Omega, \mathbb{F}, \mathbf{Pr})$ be a probability space, and let $\mathbb{F}_0, \ldots, \mathbb{F}_n$ be a filter with respect to it. Let X be any random variable over this probability space and define $X_i = \mathbf{E}\Big[X \;\Big|\; F_i\Big]$ then, the sequence $X_0, \ldots, X_n$ is a martingale.*

Proof: We need to show that $\mathbf{E}\Big[X_{i+1} \;\Big|\; F_i\Big] = X_i$. Namely,

$$\mathbf{E}\Big[X_{i+1} \;\Big|\; F_i\Big] = \mathbf{E}\Big[\mathbf{E}\Big[X \;\Big|\; F_{i+1}\Big] \;\Big|\; F_i\Big] = \mathbf{E}\Big[X \;\Big|\; F_i\Big] = X_i,$$

by Lemma 2.5 and by definition of X_i. ■

Definition 2.7 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 $y_i \in \mathcal{D}_i$,

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

Definition 2.8 Let $X_1, \ldots, X_n$ be a sequence of random variables, and a function $f(X_1, \ldots, X_n)$ defined over them that such that f satisfies the Lipschitz condition. The *Dobb martingale* sequence $Y_0, \ldots, Y_m$ is defined by $Y_0 = \mathbf{E}[f(X_1, \ldots, X_n)]$ and $Y_i = \mathbf{E}\Big[f(X_1, \ldots, X_n) \;\Big|\; X_1, \ldots, X_i\Big]$, for $i = 1, \ldots, n$. Clearly, $Y_0, \ldots, Y_n$ is a martingale, by Theorem 2.6.

Furthermore, $|X_i - X_{i-1}| \le 1$, for $i = 1, \ldots, n$. Thus, we can use Azuma's inequality on such a sequence.

3 Occupancy Revisited

We have m balls thrown independently and uniformly into n bins. Let Z denote the number of bins that remains empty. Let X_i be the bin chosen in the ith trial, and let $Z = F(X_1, \ldots, X_m)$. Clearly, we have by Azuma's inequality that $\mathbf{Pr}[|Z - \mathbf{E}[Z]| > \lambda\sqrt{m}] \leq 2e^{-\lambda^2/2}$.

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 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 $\lambda > 0$,

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

Theorem 3.2 *Let $r = m/n$, and Z_m be the number of empty bins when m balls are thrown randomly into n bins. Then*

$$\mu = \mathbf{E}[Z_m] = n\left(1 - \frac{1}{n}\right)^m \approx ne^{-r}$$

and for $\lambda > 0$,

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

Proof: Let $z(Y, t)$ be the expected number of empty bins, i 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 $\mathbb{F}_t$ be the σ-field generated by the bins chosen in the first t steps. Let Z_m be the end of empty balls at time m, and let $Z_t = \mathbf{E}\Big[Z_m \;\Big|\; F_t\Big]$. 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:

1. 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,

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

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

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

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

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

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