# Martingales

497 - Randomized Algorithms

Sariel Har-Peled

September 25, 2002

*They sought it with thimbles, they sought it with care;*
*They pursued it with forks and hope;*
*They threatened its life with a railway-share;*
*They charmed it with smiles and soap.*
*— The Hunting of the Snark, Lewis Carol*

# 1 Martingales

## 1.1 Preliminaries

Let $X, Y$ be two random variables. Let $\rho(x,y) = \mathbf{Pr}[(X = x) \cap (Y = y)]$. Then,

$$\mathbf{Pr}\left[X = x \,\middle|\, Y = y\right] = \frac{\rho(x,y)}{\mathbf{Pr}[Y = y]} = \frac{\rho(x,y)}{\sum_z \rho(z,y)}$$

and

$$E\left[X \,\middle|\, Y = y\right] = \sum_x x\,\mathbf{Pr}\left[X = x \,\middle|\, Y = y\right] = \frac{\sum_x x\rho(x,y)}{\sum_z \rho(z,y)} = \frac{\sum_x x\rho(x,y)}{\mathbf{Pr}[Y = y]}.$$

**Definition 1.1** The random variable $E\left[X \,\middle|\, Y\right]$ is the random variable $f(y) = E\left[X \,\middle|\, Y = y\right]$.

**Lemma 1.2** $E\left[E\left[X \,\middle|\, Y\right]\right] = E\left[Y\right]$.

*Proof:*

$$\begin{aligned} E\left[E\left[X \,\middle|\, Y\right]\right] &= E_Y\left[E\left[X \,\middle|\, Y = y\right]\right] = \sum_y \mathbf{Pr}[Y = y]\, E\left[X \,\middle|\, Y = y\right] \\ &= \sum_y \mathbf{Pr}[Y = y] \frac{\sum_x x\,\mathbf{Pr}[X = x \cap Y = y]}{\mathbf{Pr}[Y = y]} \\ &= \sum_y \sum_x x\,\mathbf{Pr}[X = x \cap Y = y] = \sum_x x \sum_y \mathbf{Pr}[X = x \cap Y = y] \\ &= \sum_x x\,\mathbf{Pr}[X = x] = E\left[X\right]. \end{aligned}$$

■

**Lemma 1.3** $E\left[Y \cdot E\left[X \,\middle|\, Y\right]\right] = E\left[XY\right]$.

*Proof:*

$$\begin{aligned} E\left[Y \cdot E\left[X \,\middle|\, Y\right]\right] &= \sum_y \mathbf{Pr}[Y = y] \cdot y \cdot E\left[X \,\middle|\, Y = y\right] \\ &= \sum_y \mathbf{Pr}[Y = y] \cdot y \cdot \frac{\sum_x x\, \mathbf{Pr}[X = x \cap Y = y]}{\mathbf{Pr}[Y = y]} \\ &= \sum_x \sum_y xy \cdot \mathbf{Pr}[X = x \cap Y = y] = E\left[XY\right]. \end{aligned}$$

■

## 1.2 Martingales

**Definition 1.4** A sequence of random variables $X_0, X_1, \ldots,$ is said to be a *martingale sequence* if for all $i > 0$, $E\left[X_i \,\middle|\, X_0, \ldots, X_{i-1}\right] = X_{i-1}$.

**Lemma 1.5** *Let $X_0, X_1, \ldots,$ be a martingale sequence. Then, for all $i \geq 0$, $E[X_i] = E[X_0]$.*

An example for martingales is the sum of money after participating in a sequence of fair bets.

**Example 1.6** Let $G$ be a random graph on the vertex set $V = \{1, \ldots, n\}$ obtained by independently choosing to include each possible edge with probability $p$. The underlying probability space is called $\mathcal{G}_{n,p}$. Arbitrarily label the $m = n(n-1)/2$ possible edges wit the sequence $1, \ldots, m$. For $1 \leq j \leq m$, define the indicator random variable $I_j$, which takes values 1 if the edge $j$ is present in $G$, and has value 0 otherwise. These indicator variables are independent and each takes value 1 with probability $p$.

Consider any real valued function $f$ defined over the space of all graphs, e.g., the clique number, which is defined as being the size of the largest complete subgraph. The *edge exposure martingale* is defined to be the sequence of random variables $X_0, \ldots, X_m$ such that

$$X_i = E\left[f(G) \,\middle|\, I_1, \ldots, I_k\right],$$

while $X_0 = E(f(G)]$ and $X_m = f(G)$. The fact that this sequence of random variable is a martingale follows immediately from a theorem that would be described in the next lecture.

One can define similarly a *vertex exposure martingale*, where the graph $G_i$ is the graph induced on the first $i$ vertices of the random graph $G$.

**Theorem 1.7 (Azuma's Inequzality)** *Let $X_0, \ldots, X_m$ be a martingale with $X_0 = 0$, and $|X_{i+1} - X_i| \leq 1$ for all $0 \leq i < m$. Let $\lambda > 0$ be arbitrary. Then*

$$\mathbf{Pr}\left[X_m > \lambda\sqrt{m}\right] < e^{-\lambda^2/2}.$$

*Proof:* Let $\alpha = \lambda/\sqrt{m}$. Let $Y_i = X_i - X_{i-1}$, so that $|Y_i| \leq 0$ and $E[Y_i | X_0, \ldots, X_{i-1}] = 0$.

We are interested in bounding $E\left[e^{\alpha Y_i} \,\middle|\, X_0, \ldots, X_{i-1}\right]$. Note that, for $-1 \le x \le 1$, we have

$$e^{\alpha x} \le h(x) = \frac{e^{\alpha} + e^{-\alpha}}{2} + \frac{e^{\alpha} - e^{-\alpha}}{2} x,$$

as $e^{\alpha x}$ is a convex function, $h(-1) = e^{-\alpha}$, $h(1) = e^{\alpha}$, and $h(x)$ is a linear function. Thus,

$$\begin{aligned}
E\left[e^{\alpha Y_i} \,\middle|\, X_0, \ldots, X_{i-1}\right] &\le E\left[h(Y_i) \,\middle|\, X_0, \ldots, X_{i-1}\right] = h\Big(E\left[Y_i \,\middle|\, X_0, \ldots, X_{i-1}\right]\Big) \\
&= h(0) = \frac{e^{\alpha} + e^{-\alpha}}{2} \\
&= \frac{(1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \cdots) + (1 - \alpha + \frac{\alpha^2}{2!} - \frac{\alpha^3}{3!} + \cdots)}{2} \\
&= 1 + \frac{\alpha^2}{2} + \frac{\alpha^4}{4!} + \frac{\alpha^6}{6!} + \cdots \\
&\le 1 + \frac{1}{1!}\left(\frac{\alpha^2}{2}\right) + \frac{1}{2!}\left(\frac{\alpha^2}{2}\right)^2 + \frac{1}{3!}\left(\frac{\alpha^2}{2}\right)^3 + \cdots = e^{\alpha^2/2}
\end{aligned}$$

Hence,

$$\begin{aligned}
E\left[e^{\alpha X_m}\right] &= E\left[\prod_{i=1}^{m} e^{\alpha Y_i}\right] = E\left[\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right) e^{\alpha Y_m}\right] \\
&= E\left[\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right) E\left[e^{\alpha Y_m} \,\middle|\, X_0, \ldots, X_{m-1}\right]\right] \le e^{\alpha^2/2} E\left[\prod_{i=1}^{m-1} e^{\alpha Y_i}\right] \\
&\le e^{m\alpha^2/2}
\end{aligned}$$

Therefore, by Markov's inequality, we have

$$\begin{aligned}
\mathbf{Pr}\big[X_m > \lambda\sqrt{m}\big] &= \mathbf{Pr}\Big[e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}}\Big] = \frac{E\left[e^{\alpha X_m}\right]}{e^{\alpha\lambda\sqrt{m}}} = e^{m\alpha^2/2 - \alpha\lambda\sqrt{m}} \\
&= \exp\big(m(\lambda/\sqrt{m})^2/2 - (\lambda/\sqrt{m})\lambda\sqrt{m}\big) = e^{-\lambda^2/2},
\end{aligned}$$

implying the result. ■

**Example 1.8** Let $\chi(H)$ be the chromatic number of a graph $H$. What is chromatic number of a random graph? How does this random variable behaves?

Consider the vertex exposure martingale, and let $X_i = E\left[\chi(G) \,\middle|\, G_i\right]$. Again, without proving it, we claim that $X_0, \ldots, X_n = X$ is a martingale, and as such, we have: $\mathbf{Pr}[|X_n - X_0| > \lambda\sqrt{n}] \le e^{-\lambda^2/2}$. However, $X_0 = E[\chi(G)]$, and $X_n = E\left[\chi(G) \,\middle|\, G_n\right] = \chi(G)$. Thus,

$$\mathbf{Pr}\Big[\Big|\chi(G) - E\left[\chi(G)\right]\Big| > \lambda\sqrt{n}\Big] \le e^{-\lambda^2/2}.$$

Namely, the chromatic number of a random graph is high concentrated! And we do not even know, what is the expectation of this variable!

## 2 Even more probability

**Definition 2.1** A *$\sigma$-field* $(\Omega, \mathbb{F})$ consists of a sample space $\Omega$ (i.e., the atomic events) and a collection of subsets $\mathbb{F}$ satisfying the following conditions:

1. $\emptyset \in \mathbb{F}$.
2. $C \in \mathbb{F} \Rightarrow \overline{C} \in \mathbb{F}$.
3. $C_1, C_2, \ldots \in \mathbb{F} \Rightarrow C_1 \cup C_2 \ldots \in \mathbb{F}$.

**Definition 2.2** Given a $\sigma$-field $(\Omega, \mathbb{F})$, a *probability measure* $\mathbf{Pr} : \mathbb{F} \to \mathbb{R}^+$ is a function that satisfies the following conditions.

1. $\forall A \in \mathbb{F}$, $0 \leq \mathbf{Pr}[A] \leq 1$.
2. $\mathbf{Pr}[\Omega] = 1$.
3. For mutually disjoint events $C_1, C_2, \ldots$, we have $\mathbf{Pr}[\cup_i C_i] = \sum_i \mathbf{Pr}[C_i]$.

**Definition 2.3** A *probability space* $(\Omega, \mathbb{F}, \mathbf{Pr})$ consists of a $\sigma$-field $(\Omega, \mathbb{F})$ with a probability measure $\mathbf{Pr}$ defined on it.