Martingales

Sariel Har-Peled

November 16, 2005
598shp - Randomized Algorithms

'After that he always chose out a "dog command" and sent them ahead. It had the task of informing the inhabitants in the village where we were going to stay overnight that no dog must be allowed to bark in the night otherwise it would be liquidated. I was also on one of those commands and when we came to a village in the region of Milevsko I got mixed up and told the mayor that every dog-owner whose dog barked in the night would be liquidated for strategic reasons. The mayor got frightened, immediately harnessed his horses and rode to headquarters to beg mercy for the whole village. They didn't let him in, the sentries nearly shot him and so he returned home, but before we got to the village everybody on his advice had tied rags round the dogs muzzles with the result that three of them went mad.'

– The good soldier Svejk, Jaroslav Hasek

1 Martingales

1.1 Preliminaries

Let X and 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

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

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 $\mathbf{E}\left[\mathbf{E}\left[X \,\middle|\, Y\right]\right] = E\left[Y\right]$.

Proof:

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

■

Lemma 1.3 $\mathbf{E}\Big[Y \cdot \mathbf{E}\Big[X \;\Big|\; Y\Big]\Big] = \mathbf{E}[XY]$.

Proof:

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

■

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 $i > 0$, we have $\mathbf{E}\Big[X_i \;\Big|\; X_0, \ldots, X_{i-1}\Big] = X_{i-1}$.

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

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 $\mathbf{G}_{n,p}$. Arbitrarily label the $m = n(n-1)/2$ possible edges with 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

$$X_i = \mathbf{E}\Big[f(G) \;\Big|\; I_1, \ldots, I_k\Big],$$

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*

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

Proof: Let $\alpha = \lambda/\sqrt{m}$. Let $Y_i = X_i - X_{i-1}$, so that $|Y_i| \leq 1$ and $\mathbf{E}\Big[Y_i \;\Big|\; X_0, \ldots, X_{i-1}\Big] = 0$. We are interested in bounding $E\Big[e^{\alpha Y_i} \;\Big|\; X_0, \ldots, X_{i-1}\Big]$. Note that, for $-1 \leq x \leq 1$, we have

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

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

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

Hence,

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

Therefore, by Markov's inequality, we have

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

implying the result. ■

Alternative form:

Theorem 1.8 (Azuma's Inequzality) *Let $X_0, \ldots, X_m$ be a martingale sequence such that and $|X_{i+1} - X_i| \leq 1$ for all $0 \leq i < m$. Let $\lambda > 0$ be arbitrary. Then*

$$\mathbf{Pr}\left[|X_m - X_0| > \lambda\sqrt{m}\right] < 2e^{-\lambda^2/2}.$$

Example 1.9 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: $\mathbf{Pr}[|X_n - X_0| > \lambda\sqrt{n}] \leq e^{-\lambda^2/2}$. However, $X_0 = \mathbf{E}[\chi(G)]$, and $X_n = E\left[\chi(G) \,\middle|\, G_n\right] = \chi(G)$. Thus,

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

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 *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:

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

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

1. $\forall A \in \mathcal{F}$, $0 \leq \mathbf{Pr}[A] \leq 1$.
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* $(\Omega, \mathcal{F}, \mathbf{Pr})$ consists of a σ-field $(\Omega, \mathcal{F})$ with a probability measure $\mathbf{Pr}$ defined on it.

References