COMPSCI 630: Randomized Algorithms	February 14, 2018
Lecture 9	
Lecturer: Rong Ge	*Scribe: Kangning Wang*

1 Overview

In this lecture, we introduce martingales. We first present the definition in Section 2. We cover two important inequalities about them in Section 3 and 4. Finally we apply them to a few concrete problems in Section 5.

2 Martingales

Recall if $X = X_1 + X_2 + \cdots + X_n$ and X_i's are independent and identically distributed (i.i.d.), then we have strong concentration bounds for X, e.g. from the Chernoff bound or the Bernstein inequality.

One might want to prove similar concentration bounds when X_i's are *not* independent, but information about the conditional probability $X_i \mid X_1, X_2, \ldots, X_{i-1}$ is known. It turns out there is something we can do. Let's define the concept of *martingales* first.

Definition 1. A sequence $X_0, X_1, X_2, \ldots, X_n$ is a *martingale sequence* (or just a *martingale*) if

1. $\forall i$, $\mathbb{E}[|X_i|] < \infty$ (Lebesgue-integrable), and
2. $\forall i$, $\mathbb{E}[X_i \mid X_0, X_1, \ldots, X_{i-1}] = X_{i-1}$.

Here we emphasize the fact that the conditional expectation $\mathbb{E}[X_i \mid X_0, X_1, \ldots, X_{i-1}]$ is a function of the realizations of $X_0, X_1, \ldots, X_{i-1}$. Therefore the condition

$$\mathbb{E}[X_i \mid X_0, X_1, \ldots, X_{i-1}] = X_{i-1}$$

is equivalent to

$$\forall a_0, a_1, \ldots, a_{i-1}, \mathbb{E}[X_i \mid X_0 = a_0, \ldots, X_{i-1} = a_{i-1}] = a_{i-1}.$$

Also note that only requiring $\mathbb{E}[X_i \mid X_{i-1}] = X_{i-1}$ is not enough for a sequence to be a martingale.

Definition 2. A sequence $Y_1, Y_2, \ldots, Y_n$ is a *martingale difference sequence* if

1. $\forall i$, $\mathbb{E}[|Y_i|] < \infty$, and
2. $\forall i$, $\mathbb{E}[Y_i \mid Y_0, Y_1, \ldots, Y_{i-1}] = 0$.

Next we define supermartingales and submartingales, both of which are generalizations of martingales.

Definition 3. A sequence $X_0, X_1, X_2, \ldots, X_n$ is a *supermartingale* if

1. $\forall i$, $\mathbb{E}[|X_i|] < \infty$, and
2. $\forall i$, $\mathbb{E}[X_i \mid X_0, X_1, \ldots, X_{i-1}] \leq X_{i-1}$.

Definition 4. A sequence $X_0, X_1, X_2, \ldots, X_n$ is a *submartingale* if

1. $\forall i$, $\mathbb{E}[|X_i|] < \infty$, and
2. $\forall i$, $\mathbb{E}[X_i \mid X_0, X_1, \ldots, X_{i-1}] \geq X_{i-1}$.

The prefixes "super-" and "sub-" are consistent in the sense that if a supermartingale, a martingale, and a submartingale has equal expectations at a given time, the history of the supermartingale tends to be bounded below by that of the martingale, and similarly, the history of the submartingale tends to be bounded above by that of the martingale.[1]

Proposition 1. *Most concentration bounds for the sum of independent variables also work for martingale sequences and the sum of martingale difference sequences.*

3 Azuma's Inequality

In this section, we present a simplified version of Azuma's inequality.

Theorem 2 (Azuma's inequality)**.** *If $Y_1, Y_2, \ldots, Y_n$ is a martingale difference sequence and $|Y_i| \leq c$ with probability 1, then*

$$\Pr\left[\left|\sum_i Y_i\right| > \lambda \cdot c \cdot \sqrt{n}\right] \leq 2 \cdot e^{-\frac{\lambda^2}{2}}.$$

Proof. Recall the moment generating function. Let $Y = \sum_{i=1}^n Y_i$. Define $f(t) = \mathbb{E}\left[e^{tY}\right]$. Let's first prove the following claim by induction: $\mathbb{E}\left[e^{tY}\right] \leq \left(\frac{e^{ct}+e^{-ct}}{2}\right)^n (= \cosh^n(ct))$. Consider $\mathbb{E}[X + X']$ where $X = \sum_{i=1}^{n-1} Y_i$ and $X' = Y_n$. The inductive step can be written as $\mathbb{E}\left[e^{t(X+X')}\right] \leq \mathbb{E}\left[e^{tX}\right] \cdot \cosh(ct)$. We have

$$\begin{aligned}\mathbb{E}\left[e^{t(X+X')} \,\middle|\, X = x\right] &= e^{tx}\left[e^{tX'} \,\middle|\, X = x\right] \\ &\leq e^{tx}\cosh(ct)\end{aligned}$$

because of the convexity of the hyperbolic cosine function $\cosh(*)$.

Therefore,

$$\begin{aligned}\mathbb{E}\left[e^{t(X+X')}\right] &= \sum_x \Pr[X = x] \cdot \mathbb{E}\left[e^{t(X+X')} \,\middle|\, X = x\right] \\ &\leq \sum_x \Pr[X = x] \cdot e^{tx} \cdot \cosh(ct) \\ &= \mathbb{E}\left[e^{tX}\right] \cdot \cosh(ct).\end{aligned}$$

By the principle of induction, we know

$$f(t) \leq \cosh^n(ct) \leq e^{\frac{nc^2t^2}{2}}.$$

[1] `https://en.wikipedia.org/wiki/Martingale_(probability_theory)`

Here we used the fact that

$$\cosh x = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty} \frac{x^{2n}}{2^n n!} = e^{\frac{x^2}{2}}.$$

Setting $t = \frac{\lambda}{c\sqrt{n}}$, we have

$$\begin{aligned}\Pr[Y \geq \lambda c\sqrt{n}] &\leq \frac{\mathbb{E}\left[e^{tY}\right]}{e^{t\lambda c\sqrt{n}}} \\ &= \exp\left(\frac{nc^2t^2}{2} - \lambda ct\sqrt{n}\right) \\ &= \exp\left(-\frac{\lambda^2}{2}\right).\end{aligned}$$

To finish the proof, we can use a similar argument to show $\Pr[Y \leq -\lambda c\sqrt{n}] \leq \exp\left(-\frac{\lambda^2}{2}\right)$. □

4 McDiarmid's Inequality

In this section we present another important result: McDiarmid's inequality (the bounded difference inequality).

Definition 5. We say a function $f(x_1, x_2, \ldots, x_n)$ satisfies *bounded difference assumption* if for any $x_i, x_2, \ldots, x_n$ and any x_i', we have

$$|f(x_1, x_2, \ldots, x_n) - f(x_1, x_2, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c.$$

Theorem 3 (McDiarmid's inequality)**.** *If a function $f(x_1, x_2, \ldots, x_n)$ satisfies the bounded difference assumption with parameter c, and x_i's come from a product distribution (or more generally, if x_i only depend on $x_1, x_2, \ldots, x_{i-1}$), then*

$$\Pr[|f - \mathbb{E}[f]| \geq \lambda c\sqrt{n}] \leq 2e^{-\frac{\lambda^2}{2}}.$$

Proof. Define a martingale sequence: $X_0 = \mathbb{E}_{x_1,\ldots,x_n}[f(x)]$, $X_k = g(x_1, x_2, \ldots, x_k) = \mathbb{E}_{x_{k+1},\ldots,x_n}[f(x)]$, $X_n = f(x)$. Apply Azuma's inequality on this martingale. □

5 Applications

5.1 Longest Increasing Subsequence

The longest increasing subsequence problem is as follows: Given a sequence $x_1, x_2, \ldots, x_n$, find the largest k such that there exist $x_{i_1} < x_{i_2} < \cdots < x_{i_k}$ and $i_1 < i_2 < \cdots < i_k$.

Assume x_i's are independent. Let $f(x_1, x_2, \ldots, x_n)$ be the length of LIS of $(x_1, x_2, \ldots, x_n)$. McDiarmid's inequality tells us

$$\Pr[|f - \mathbb{E}[f]| \geq \lambda\sqrt{n}] \leq 2e^{-\frac{\lambda^2}{2}}$$

because f has a bounded difference of 1.

5.2 Chromatic Number of a Graph

The *chromatic number* is the minimum number of colors to color a graph such that the two endpoints of each edge is of different colors.

Given a distribution over graphs, let X_i encode edges from i to $1, 2, \ldots, i-1$. Again, McDiarmid's inequality tells us

$$\Pr[|f - \mathbb{E}[f]| \geq \lambda\sqrt{n}] \leq 2e^{-\frac{\lambda^2}{2}}$$

because f has a bounded difference of 1.