Lecture 19, 10/25/2025

Introduction to Martingales

We saw Chernoff-Hoeffding bounds for sum of independent random variables and applications.
Last lecture we saw concentration bounds also hold for negatively correlated random variables and saw an application. However there are other settings where we don't have independence or negative correlation and still concentration holds.

Martingales provide a powerful framework for such bounds are are also somewhat natural in algorithms. We saw an example of a martingale type process in the last lecture on rounding a fractional solution for the Max K-Cover problem.

<u>Background on Conditional Probability</u>

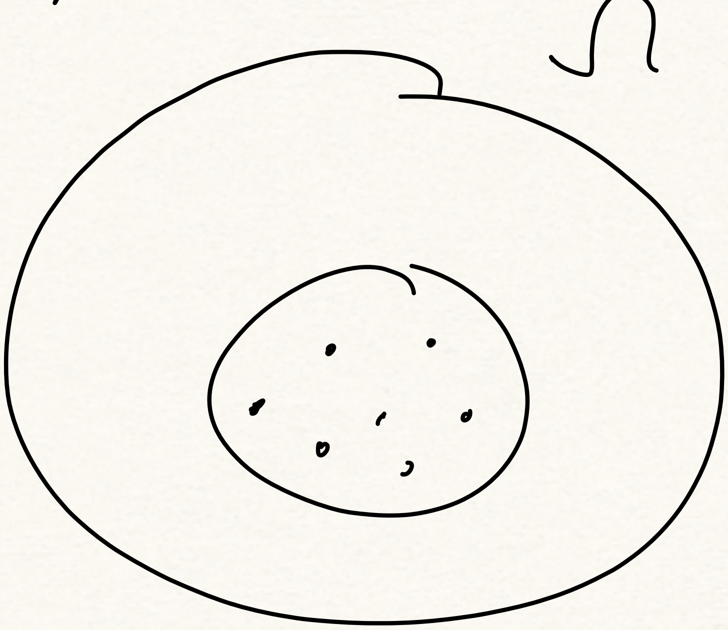
Suppose Ω is a probability space and $X : \Omega \to R$ is a real-valued random variable.

If $A \subseteq \Omega$ is an even then

$E[X \mid A]$ is a real value

which, in the setting when Ω is discrete, is defined as

$$\frac{1}{\Pr[A]} \sum_{\omega \in A} \Pr[\omega]\, X(\omega).$$



In the continuous setting it is

$$\frac{1}{\Pr[A]} \int_A f(x)\,dx$$

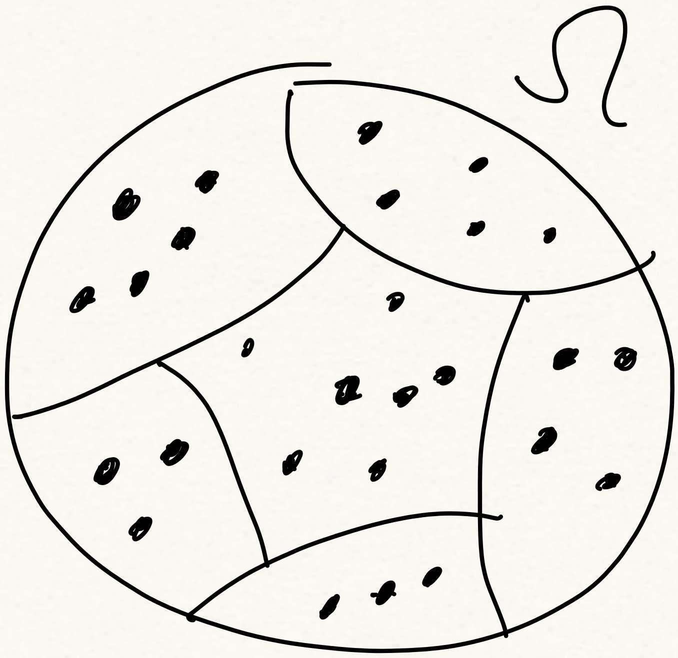
Given two random variables X and Y on Ω the random variable

$$E[X \mid Y]$$ is defined as

follows. If $Z = E[X \mid Y]$

then $Z(\omega) = E[X \mid Y = Y(\omega)]$

(this is in the discrete setting)



The meaning is the following. Y partitions Ω into parts where Y is constant in each part.

$Y = b$ for some fixed b constitutes a part which can be alternatively thought as an event. $A_b = \{\omega \mid Y(\omega) = b\}$.

Z assigns a value to each $\omega \in A_b$ a value equal to $E[X \mid A_b]$.

<u>Claim</u>: $E[E[X \mid Y]] = E[X]$.

<u>Proof</u>: Exercise.

When we write $E[X \mid Y_1, Y_2, .., Y_n]$ where $Y_1, \ldots, Y_n$ are several random variables then we are looking at the partition of Ω

induced by $Y_1, \dots, Y_n$ taking on different values in a joint way.
The events correspond to

$$A_{b_1 b_2 \dots b_n} = \{\omega \in \Omega \mid Y_i(\omega) = b_i\}.$$

<u>Tower property</u>

<u>Lemma</u>: $E\{E[A \mid B, C] \mid C\} = E[A \mid C]$.

<u>Proof</u>: Exercise.

Technically, a more formal way to describe this is via σ-algebras and filtrations but that requires more background.

<u>Defn:</u> A sequence of random variables $X_0, X_1, X_2, \ldots$ is a martingale sequence with respect to another sequence of random variables $Y_0, Y_1, \ldots,$ if $\forall n \geq 0$

(i) X_n is determined by $Y_0, \ldots, Y_n$.

(ii) $E[|X_n|] < \infty \quad \forall n.$

(iii) $E[X_{n+1} \mid Y_0, \ldots, Y_n] = X_n$

A sequence $X_0, X_1, \ldots,$ is a martingale if it is a martingale wrt itself. That is (i) $E[|X_n|] < \infty$ and

(ii) $E[X_{n+1} \mid X_0, \ldots, X_n] = X_n$.

<u>Example</u>: Suppose a gambler starts out with a random amount of X_0 money on day 0 and goes to the casino every day and plays some slot machine which is fair.

Let Y_i be the winnings on day i (<0 if he/she loses).

Let X_i be the total amount that player has at the end of the i th game. Because each game is fair $E[Y_i] = 0$. $\forall i$

$$E[X_{n+1} \mid Y_1, \ldots, Y_n] = X_n + E[Y_{n+1}]$$
$$= X_n.$$

The main thing that martingales allow one to capture is that the choice of how much to bet and on which slot machine to bet on can be arbitrarily dependent on all the information/choices up to the previous step.

Doob Martingale

Martingales allow one to capture a particular type of phenomenon where we are interested in a function $f : U \to \mathbb{R}$ over some objects U and we have a random variable X that takes values in U. In other words X is a random object chosen from $\mathcal{U}$ according to some process. This process can be defined by a sequence of random variables $Y_0, Y_1, \ldots, Y_n$. And X is determined by $Y_0, \ldots, Y_n$. However $Y_0, \ldots, Y_i$ for $i < n$ reveal partial information about X.

Let us define

$$X_i = E\left[f(X) \mid Y_0, \ldots, Y_i\right].$$

We will assume that $E[f(X)] < \infty$

Claim: $X_0, \ldots, X_n$ is a martingale sequence with respect to $Y_0, \ldots, Y_n$.

Proof:

$$E[X_{i+1} \mid Y_0, \ldots, Y_i]$$

$$= E\big[E[f(X) \mid Y_0, \ldots, Y_{i+1}] \mid Y_0, \ldots, Y_i\big]$$

(by defn of X_{i+1})

$$= E[f(X) \mid Y_0, \ldots, Y_i]$$

by tower rule

$$= X_i \quad \text{by defn.}$$

$\square$

Remark: In some settings it is easier to view $f(X)$ as another random variable Z and define $X_i = E[Z \mid Y_1, \ldots, Y_i]$.

Ex: Empty bins in balls and bins.

Suppose we throw m balls into n bins independently. We think of this as a sequential process where we place ball i in the ith step and Y_i is the random choice of ith ball. $Y_i \in [n]$. Let X be the # of empty bins after all m balls are thrown. Then we know

$$X_i = E[X \mid Y_1, \ldots, Y_i].$$

$$X_0 = E[X] = n\left(1 - \frac{1}{n}\right)^m.$$

Example: Let $G(n,p)$ be a random graph on n vertices where each potential edge in the graph is chosen to be added independently with probability p. Let X be the chromatic number of the random graph. Technically we should write it as $X_{n,p}$ to show dependence on n and p. It is somewhat complicated to estimate. But we can define a Doob martingale called the edge exposure martingale where $Y_1, Y_2, \ldots, Y_{\binom{n}{2}}$ correspond to the binary random variables for picking the edges in some fixed order.

Then $X_0 = E[X(G_{n,p})]$

and $X_i = E[X \mid Y_1, \ldots, Y_i]$.

We can define another Doob martingale called the vertex exposure martingale where we $Y_1, Y_2, \ldots, Y_n$ is an ordering of vertices and Y_i reveals information about the edges of vertex i to vertices 1 to $i-1$ in the random process.

Azuma-Hoeffding Inequality

Recall the additive Chernoff inequality:

Let $X = \sum_{i=1}^{n} X_i$ where

(i) X_i are independent

(ii) $X_i \in [a_i, b_i] \quad \forall i \in [n]$

Then,

$$\Pr\left[X - E[X] > \lambda\right] \leq e^{-\frac{\lambda^2}{2\sum_i (b_i - a_i)^2}}$$

and

$$\Pr\left[X - E[X] < \lambda\right] \quad e^{-\frac{\lambda^2}{2\sum_i (b_i - a_i)^2}}$$

A simpler form is when

$[a_i, b_i] = [-c_i, c_i]$ in which case we have

$$\Pr\left[X - E[X] > \lambda\right] \leq e^{\frac{-\lambda^2}{2\sum_{i=1}^{n} c_i^2}}$$

Similarly the lower tail.

Azuma-Hoeffding bound extends this to the martingale setting.

<u>Theorem:</u> Let $X_0, X_1, X_2, \ldots$, be a martingale sequence where

$$|X_i - X_{i-1}| \leq c_i \quad \forall i \geq 1$$

deterministically. Then

$$\Pr\left[X_n \geq X_0 + \lambda\right] \leq e^{\frac{-\lambda^2}{2\sum_{i=1}^{n} c_i^2}}$$

Proof: We need an auxiliary Lemma

Lemma: Let X be a random variable in $[-1, 1]$ with $E[X]=0$. Then $E\left[e^{aZ}\right] \leq e^{a^2/2}$.

Proof: The function e^{ax} is convex in $[-1,1]$. For any $x \in [-1,1]$ we can write $x = \frac{1+x}{2}(+1) + \frac{1-x}{2}(-1)$

Hence by convexity of e^{ax}

$$e^{ax} \leq \frac{1+x}{2} e^{a} + \frac{1-x}{2} e^{-a}$$

$$\leq \frac{e^{a}+e^{-a}}{2} + \frac{e^{a}-e^{-a}}{2} x$$

If $X \in [-1,1]$ with $E[X]=0$ taking expectation on both sides

$$E\left[e^{aX}\right] \leq \frac{e^{a}+e^{-a}}{2}$$

By Taylor expansion

$$\leq \left(1 + \frac{a^2}{2!} + \frac{a^4}{4!} + \cdots\right)$$

$$\leq e^{\frac{a^2}{2}}$$

$\square$.

<u>Corollary</u>: $\mathbb{E}\{X\} = 0 \quad X \in [-c, c]$.

$$\mathbb{E}\left[e^{aX}\right] \leq e^{\frac{a^2c^2}{2}}$$

<u>Proof</u>: Consider $X' = \dfrac{X}{c}$ and apply previous Lemma.

$\square$.

Now we do the main proof.

Let $t>0$ be a parameter to be chosen later.

$$\Pr\left[X_n - X_0 \geq \lambda\right] \leq \Pr\left[e^{t(X_n - X_0)} \geq e^{t\lambda}\right]$$

$$\leq \frac{\mathbb{E}\left[e^{t(X_n - X_0)}\right.}{e^{t\lambda}}.$$

So it boils down to estimating

$$\mathbb{E}\left[e^{t(X_n - X_0)}\right]$$

For this we consider

$Z_i = X_i - X_{i-1}$. Recall

$|Z_i| \leq c_i$ and

$E[Z_i \mid X_0, \dots, X_{i-1}] = 0$ since

$E[X_i - X_{i-1} \mid X_0 \dots X_{i-1}]$

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

$= X_{i-1} - X_{i-1} = 0.$

Consider $E[e^{tZ_i} \mid X_0, \dots, X_{i-1}]$

By Lemma

$$\leq e^{\frac{-t^2}{2c_i^2}}.$$

Now

$$E\left[e^{t(X_n - X_0)}\right] = E\left[e^{t(Z_n + Z_{n-1} + \cdots + Z_1)}\right]$$

$$= E\left[e^{t(Z_{n-1} + \cdots + Z_1)} \cdot e^{tZ_n}\right]$$

$$= E\left[e^{t(Z_{n-1} + \cdots + Z_1)}\right] E\left[e^{tZ_n} \mid X_0, \ldots, X_{n-1}\right]$$

$$\leq E\left[e^{t(Z_{n-1} + \cdots + Z_1)}\right] e^{\frac{t^2 c_n^2}{2}}$$

$$\leq E\left[e^{t(Z_{n-2} + \cdots + Z_1)}\right] \cdot e^{\frac{t^2 c_{n-1}^2}{2}} \cdot e^{\frac{t^2 c_n^2}{2}}$$

$$\leq e^{\left(t^2 \sum_{i=1}^{n} c_i^2\right)/2}.$$

Therefore $E\left[e^{t(X_n - X_0)}\right] \leq e^{\frac{\sum c_i^2}{2} \cdot t^2}$.

Putting together

$$\frac{E\left[e^{t(X_n - X_0)}\right]}{e^{t\lambda}} \le \frac{e^{t^2 \frac{\sum c_i^2}{2}}}{e^{t\lambda}}.$$

$$\le e^{t^2\left(\frac{\sum c_i^2}{2}\right) - t\lambda}.$$

Choosing $t = \frac{\lambda}{\sum c_i^2}$.

$$\le e^{-\frac{\lambda^2}{2\sum_i c_i^2}}$$

□.

<u>Corollary</u>: If $c_i \le c \ \forall i$ then

$$\Pr\left[|X_n - X_0| \ge c\cdot\alpha\sqrt{n}\right] \le 2e^{-\frac{\alpha^2}{2}}.$$

McDiarmid's Inequality and Applications

Chernoff-Hoeffding bound says that if $X = X_1 + X_2 + \cdots + X_n$ where $X_1, \ldots, X_n$ are independent and in a bounded range then X has concentration. Now consider the setting when we have an arbitrary function $f : U_1 \times U_2 \cdots \times U_n \to \mathbb{R}$

In other words f is a function of n "variables" each of which has domain U_i.

Defn: $f : U_1 \times U_2 \cdots \times U_n \to \mathbb{R}$ is c-Lipschitz for some $c > 0$ if

$$|f(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)| \le c$$

$\forall i,\ y, z \in U_i,\ x_j \in U_j\ j \neq i$.

In other words changing one coordinate does not change the value of the function by more than c in absolute value.

A more refined definition is that f is $(c_1, c_2, \dots, c_n)$-Lipschitz if

$$|f(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)| \le c_i$$

$\forall i,\ y, z \in \mathcal{U}_i,\ x_j \in \mathcal{U}_j\ j \neq i.$

Theorem: Suppose f is $(c_1, c_2, \ldots, c_n)$-Lipschitz.
Let $X_1, X_2, \ldots, X_n$ be independent random variables where $X_i \in \mathcal{U}_i$.
Then

$$\Pr\left[\left|f(X_1, \ldots, X_n) - E[f(X_1, \ldots, X_n)]\right| \geq \lambda\right] \leq 2 e^{\frac{-\lambda^2}{2 \sum_i c_i^2}}.$$

Note that independence is required.
In some settings it can be relaxed.

We reduce this to the Azuma's inequality via the Doob martingale. Main thing to note is where independence is used.

Let $\bar{X}$ denote $(X_1, \dots, X_n)$.

and $Z_i = E[f(\bar{X}) \mid X_1, \dots, X_i]$

$Z_0 = E[f(\bar{X})]$.

Recall that $Z_0, \dots, Z_n$ is a Doob martingale sequence. We wish to apply Azuma's inequality to $Z_0, \dots, Z_n$. For this we need to bound $Z_i - Z_{i-1}$.

$$Z_i - Z_{i-1} = E[f(\bar{X}) \mid X_1, \ldots, X_i] - E[f(\bar{X}) \mid X_1, \ldots, X_{i-1}]$$

Recall that $X_1, \ldots, X_n$ are independent

$\Rightarrow$ $X_{i+1}, \ldots, X_n$ are independent of

$X_1, \ldots, X_i$.

We have

$$Z_i - Z_{i-1} \leq \sup_{\substack{u_1, \ldots, u_{i-1} \\ a, b}} \Big| E[f(\bar{X}) \mid X_1 = u_1, \ldots, X_{i-1} = u_{i-1}, X_i = a] - E[f(\bar{X}) \mid X_1 = u_1, \ldots, X_{i-1} = u_{i-1}, X_i = b] \Big|$$

$$\leq \sup_{\substack{u_1, u_2, \ldots, u_{i-1} \\ a, b}} \Big| E[f(u_1, \ldots, u_{i-1}, a, X_{i+1}, \ldots, X_n)] - E[f(u_1, \ldots, u_{i-1}, b, X_{i+1}, \ldots, X_n)] \Big|$$

$$\leq c_i$$

where we used c_i Lipschitzness of f and independence of $X_{i+1}, \dots, X_n$ from $X_1, \dots, X_i$.

Now we can apply Azuma's inequality to $Z_0, \dots, Z_n$ and conclude that

$$\Pr\left[Z_n - Z_0 > \lambda\right] \leq e^{-\frac{\lambda^2}{2\sum c_i^2}}.$$

We have $Z_0 = E[f(\bar{X})]$

and $Z_n = f(\bar{X})$.

$\square$.

Applications

Balls and Bins

We throw m balls into n bins independently. Let $X_1, \ldots, X_m$ be the random choices of the m balls. $X_i \in [n]$. $\forall i \in [m]$.

Define $f(X_1, X_2, \ldots, X_m)$ to be the # of empty bins.

$E[f(X_1, \ldots, X_m)]$ is easy to calculate and is equal to $n(1-\frac{1}{n})^m$.

We claim f is 1-Lipschitz. This

is easy to verify. Changing the assignment of one ball can change # of empty bins by at most 1.

Thus we get concentration:

$$\Pr\left[f(X_1,\dots,X_m) - n\left(1-\frac{1}{n}\right)^m > \lambda\right]$$

$$\leq e^{-\frac{\lambda^2}{2n}}.$$

Hence if $\lambda \geq \alpha\sqrt{n}$ the prob

$$\leq e^{-\frac{\alpha^2}{2}}.$$

<u>Chromatic number of random graph.</u>

Consider random graph $G(n,p)$.

Let $\chi(G)$ be the chromatic number of G. For instance it is known that $\mathbb{E}[\chi(G(n,\frac{1}{2}))] = \frac{n}{2\log n}$.

What about concentration?

We consider the vertex exposure martingale. Fix ordering of vertices and let Y_i for vertex i denote the random vector of edges to vertices 1 to $i-1$.

Let $f(Y_1, \ldots, Y_n)$ be the chromatic number of $G(n,p)$.

We claim f is 1-Lipschitz.

Changing the edges incident to a

single vertex can change chromatic number by at most 1. Why?

Thus the chromatic number of $G(n,p)$ is concentrated around its mean.