Azuma's Inequality

Will Perkins

March 28, 2013

Azuma's Inequality

Theorem (Azuma's Inequality)

Let X_n be a Martingale so that $|X_i - X_{i-1}| \leq d_i$ (with probability 1). Then

$$\Pr[|X_n - X_0| \geq t] \leq 2e^{-t^2/2D^2}$$

where $D^2 = \sum_{i=1}^n d_i^2$.

Azuma's Inequality

Theorem (Azuma's Inequality)

Let X_n be a Martingale so that $|X_i - X_{i-1}| \leq d_i$ (with probability 1). Then

$$\Pr[|X_n - X_0| \geq t] \leq 2e^{-t^2/2D^2}$$

where $D^2 = \sum_{i=1}^{n} d_i^2$.

If all the d_i's are 1, we get an analogue of the Chernoff Bound:

$$\Pr[|X_n - X_0| \geq t] \leq 2e^{-t^2/2n}$$

Proof: Assume for simplicty that $X_0 = 0$. We will prove one side of the inequality.

Azuma's Inequality

Proof: Assume for simplicty that $X_0 = 0$. We will prove one side of the inequality. 1. Use the exponential Markov inequality:

$$\Pr[X_n \geq t] \leq e^{-\lambda t} \mathbb{E} e^{\lambda X_n}$$

Azuma's Inequality

2. Find a bound for $\mathbb{E}e^{\lambda X_n}$.

Azuma's Inequality

2. Find a bound for $\mathbb{E}e^{\lambda X_n}$.

$$\begin{aligned}\mathbb{E}e^{\lambda X_n} &= \mathbb{E}[\mathbb{E}[e^{\lambda(X_n - X_{n-1}) + \lambda X_{n-1}} | \mathcal{F}_{n-1}]] \\ &= \mathbb{E}[e^{\lambda X_{n-1}} \mathbb{E}[e^{\lambda(X_n - X_{n-1})} | \mathcal{F}_{n-1}]]\end{aligned}$$

Azuma's Inequality

Now find a bound for the one term, $\mathbb{E}[e^{\lambda(X_n - X_{n-1})} | \mathcal{F}_{n-1}]$: Let $y = (X_n - X_{n-1})/d_n$. $-1 \leq y \leq 1$ with probability 1.
By convexity of e^x,

$$e^{d_n \lambda y} \leq \frac{1+y}{2} e^{d_n \lambda} + \frac{1-y}{2} e^{-d_n \lambda}$$

Azuma's Inequality

Now find a bound for the one term, $\mathbb{E}[e^{\lambda(X_n - X_{n-1})}|\mathcal{F}_{n-1}]$: Let $y = (X_n - X_{n-1})/d_n$. $-1 \leq y \leq 1$ with probability 1.
By convexity of e^x,

$$e^{d_n \lambda y} \leq \frac{1+y}{2} e^{d_n \lambda} + \frac{1-y}{2} e^{-d_n \lambda}$$

$$\mathbb{E}[e^{d_n \lambda y}|\mathcal{F}_{n-1}] \leq \frac{1}{2} e^{d_n \lambda} + \frac{1}{2} e^{-d_n \lambda}$$

since $\mathbb{E}[y|\mathcal{F}_{n-1}] = 0$ (Martingale Property).

$$= \cosh(d_n \lambda) \leq e^{\lambda^2 d_n^2 / 2}$$

Azuma's Inequality

3. This gives us:

$$\mathbb{E}e^{\lambda X_n} \leq e^{\lambda^2 d_n^2/2}\mathbb{E}e^{\lambda X_{n-1}}$$

and now we can repeat the same thing $n-1$ more times.

Azuma's Inequality

3. This gives us:

$$\mathbb{E}e^{\lambda X_n} \leq e^{\lambda^2 d_n^2/2}\mathbb{E}e^{\lambda X_{n-1}}$$

and now we can repeat the same thing $n-1$ more times.

$$\mathbb{E}e^{\lambda X_n} \leq e^{\lambda^2 \sum d_i^2/2} = e^{\lambda^2 D^2/2}$$

and so

$$\Pr[X_n \geq t] \leq e^{-\lambda t}e^{\lambda^2 D^2/2}$$

Azuma's Inequality

4. Now optimize over λ:

$$f(\lambda) = \lambda^2 D^2/2 - \lambda t$$

$$f'(\lambda) = \lambda D^2 - t$$

so setting $\lambda = t/D^2$ mimimizes the exponent, and gives us:

$$\Pr[X_n \geq t] \leq e^{-t^2/2D^2}$$

Azuma's Inequality

4. Now optimize over λ:

$$f(\lambda) = \lambda^2 D^2/2 - \lambda t$$

$$f'(\lambda) = \lambda D^2 - t$$

so setting $\lambda = t/D^2$ mimimizes the exponent, and gives us:

$$\Pr[X_n \geq t] \leq e^{-t^2/2D^2}$$

The same thing works to show

$$\Pr[X_n \leq -t] \leq e^{-t^2/2D^2}$$

Chromatic number of a random graph

The chromatic number of a graph, $\chi(G)$, is the smallest k so that G can be properly colored with k colors.
Examples:

1. A bipartite graph has chromatic number 2.
2. A planar graph as chromatic number at most 4 (the famous 4 color theorem)

Chromatic number of a random graph

The chromatic number of a graph, $\chi(G)$, is the smallest k so that G can be properly colored with k colors.
Examples:

1. A bipartite graph has chromatic number 2.
2. A planar graph as chromatic number at most 4 (the famous 4 color theorem)

Q: What is the chromatic number of the random graph $G(n, p)$?
This is an old and difficult problem that is not yet fully solved.

Chromatic number of a random graph

It is difficult to even compute $\mathbb{E}\chi(G)$. Nevertheless, Azuma's Inequality will give us something:

Theorem

$$\Pr[|\chi(G) - \mathbb{E}\chi(G)| \geq r\sqrt{n-1}] \leq 2e^{-r^2/2}$$

Chromatic number of a random graph

It is difficult to even compute $\mathbb{E}\chi(G)$. Nevertheless, Azuma's Inequality will give us something:

Theorem

$$\Pr[|\chi(G) - \mathbb{E}\chi(G)| \geq r\sqrt{n-1}] \leq 2e^{-r^2/2}$$

This theorem states that the chromatic number is concentrated within $O(\sqrt{n})$ from its mean, whatever that is, whp.

Proof:
We are working on the probability space defined by $G(n, p)$ - $\Omega = \{0, 1\}^{\binom{n}{2}}$, $\mathcal{F}$ is all subsets, and P is the product measure in which each edge appears with probability p.

Proof:
We are working on the probability space defined by $G(n,p)$ - $\Omega = \{0,1\}^{\binom{n}{2}}$, $\mathcal{F}$ is all subsets, and P is the product measure in which each edge appears with probability p.

To define a martingale we need a filtration. There are two especially useful filtrations for a random graph: the vertex exposure filtration and the edge exposure filtration.

Edge Exposure Filtration

Let $\mathcal{F}_0 = \{\Omega, \emptyset\}$.
Let $\mathcal{F}_k = \sigma(e_1, \ldots e_k)$ where e_i is the ith edge of the $\binom{n}{2}$ possible edges.
Notice that $\mathcal{F}_{\binom{n}{2}} = \mathcal{F}$, all subsets of Ω. So the filtration has length $\binom{n}{2}$.

Vertex Exposure Filtration

Let $\mathcal{F}_1 = \{\Omega, \emptyset\}$.
Let $\mathcal{F}_k = \sigma(\{e : e \subset \{v_1, \ldots v_k\})$ where v_i is the ith vertex of the n vertices.
Here $\mathcal{F}_n = \mathcal{F}$ and the filtration has length $n - 1$.
Notice that we can order the vertices and edges so that the vertex filtration is a subsequence of the edge filtration.

The Martingale

We will use the vertex filtration.
Let $X_k = \mathbb{E}[\chi(G)|\mathcal{F}_k]$. Then

- $X_1 = \mathbb{E}\chi(G)$
- $X_n = \chi(G)$
- X_k is a (Doob's) martingale with respect to $\mathcal{F}_k$

The Martingale

We will use the vertex filtration.
Let $X_k = \mathbb{E}[\chi(G)|\mathcal{F}_k]$. Then

- $X_1 = \mathbb{E}\chi(G)$
- $X_n = \chi(G)$
- X_k is a (Doob's) martingale with respect to $\mathcal{F}_k$

Can we bound $|X_k - X_{k-1}|$?

The Martingale

We will use the vertex filtration.
Let $X_k = \mathbb{E}[\chi(G)|\mathcal{F}_k]$. Then

- $X_1 = \mathbb{E}\chi(G)$
- $X_n = \chi(G)$
- X_k is a (Doob's) martingale with respect to $\mathcal{F}_k$

Can we bound $|X_k - X_{k-1}|$?

Yes. $|X_k - X_{k-1}| \leq 1$. Why? Say G_1 and G_2 are identical except for a set of edges containing a fixed vertex v. Then $|\chi(G_1) - \chi(G_2)| \leq 1$, because v can always be given a completely new color to preserve a proper coloring. We call this the vertex Lipschitz condition.

Chromatic number of a random graph

Now we can apply Azuma's Inequality to X_k, with $D^2 = (n-1)$.

$$\Pr[|X_n - X_1| \geq t] \leq 2e^{-t^2/2(n-1)}$$

or

$$\Pr[|X_n - X_1| \geq r\sqrt{n-1}] \leq 2e^{-r^2/2}$$

Chromatic number of a random graph

Now we can apply Azuma's Inequality to X_k, with $D^2 = (n-1)$.

$$\Pr[|X_n - X_1| \geq t] \leq 2e^{-t^2/2(n-1)}$$

or

$$\Pr[|X_n - X_1| \geq r\sqrt{n-1}] \leq 2e^{-r^2/2}$$

What other graph functions satisfy either an edge or vertex Lipschitz condition?

Isoperimetric Inequalities

The Classic Isoperimetry Problem:

Of all 2D shapes with area 1, which has the smallest boundary?

Ans: the circle!

The Classic Isoperimetry Problem:
Of all 2D shapes with area 1, which has the smallest boundary?
Ans: the circle!

Another way of writing this is to say that if a region in the plane has area x, then its boundary must be at least $2\sqrt{\pi x}$. This is an isoperimetric inequality. [Check for a rectangle]

The Hamming Cube is the space $\{0,1\}^n$ with the Hamming metric: $d(x,y)$ is the number of coordinates in which x and y differ. Neighbors are points in the cube that differ in one coordinate. The boundary of a subset of the cube is the set of all points in the subset that neighbor a point outside the subset.

Isoperimetric Inequalities

The Hamming Cube is the space $\{0,1\}^n$ with the Hamming metric: $d(x,y)$ is the number of coordinates in which x and y differ. Neighbors are points in the cube that differ in one coordinate. The boundary of a subset of the cube is the set of all points in the subset that neighbor a point outside the subset.

A generalization of a boundary is the r-enlargement of a set A. We define

$$A_r = \{x : d(x,A) \leq r\}$$

In particular, $A \subseteq A_r$.

The Hamming Cube is the space $\{0, 1\}^n$ with the Hamming metric: $d(x, y)$ is the number of coordinates in which x and y differ. Neighbors are points in the cube that differ in one coordinate. The boundary of a subset of the cube is the set of all points in the subset that neighbor a point outside the subset.

A generalization of a boundary is the r-enlargement of a set A. We define

$$A_r = \{x : d(x, A) \leq r\}$$

In particular, $A \subseteq A_r$.

An isoperimetric inequality would show that if A is large, then A_r must be very large.

Isoperimetric Inequalities

Theorem

Let $A \subset \{0,1\}^n$. *Let* $|A| \geq \epsilon 2^n$ *and define* λ *so that* $e^{-\lambda^2/2} = \epsilon$. *Then if* $r = 2\lambda\sqrt{n}$,

$$|A_r| \geq (1-\epsilon)2^n$$

Isoperimetric Inequalities

Theorem

Let $A \subset \{0,1\}^n$. Let $|A| \geq \epsilon 2^n$ and define λ so that $e^{-\lambda^2/2} = \epsilon$. Then if $r = 2\lambda\sqrt{n}$,

$$|A_r| \geq (1-\epsilon)2^n$$

Notice that this says that if some subset has an ϵ fraction of the total volume of the Hamming cube, then almost all the hypercube is within distance $O(\sqrt{n})$ from some point in the set.

Proof: We need a random variable and a filtration. Let X be the distance of a randomly chosen point x from A. [The distance of a point x from a set is the minimum distance $d(x, y)$ over all points $y \in A$].

Proof: We need a random variable and a filtration. Let X be the distance of a randomly chosen point x from A. [The distance of a point x from a set is the minimum distance $d(x, y)$ over all points $y \in A$].

Define a filtration $\mathcal{F}_k$ by revealing one coordinate of x at a time.

Proof: We need a random variable and a filtration. Let X be the distance of a randomly chosen point x from A. [The distance of a point x from a set is the minimum distance $d(x, y)$ over all points $y \in A$].

Define a filtration $\mathcal{F}_k$ by revealing one coordinate of x at a time.

Then $X_k = \mathbb{E}[X|\mathcal{F}_k]$ is a martingale with

- $X_0 = \mathbb{E}X$
- $X_n = X$.

Proof: We need a random variable and a filtration. Let X be the distance of a randomly chosen point x from A. [The distance of a point x from a set is the minimum distance $d(x, y)$ over all points $y \in A$].

Define a filtration $\mathcal{F}_k$ by revealing one coordinate of x at a time.

Then $X_k = \mathbb{E}[X|\mathcal{F}_k]$ is a martingale with

- $X_0 = \mathbb{E}X$
- $X_n = X$.

Show that $|X_k - X_{k-1}| \leq 1$.

Azuma's Inequality tells us two things:

1. $$\Pr[X - \mathbb{E}X < -\lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon$$

2. $$\Pr[X - \mathbb{E}X > \lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon$$

Azuma's Inequality tells us two things:

1. $$\Pr[X - \mathbb{E}X < -\lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon$$

2. $$\Pr[X - \mathbb{E}X > \lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon$$

But what is $\mathbb{E}X$?

Isoperimetric Inequalities

Azuma's Inequality tells us two things:

1. $$\Pr[X - \mathbb{E}X < -\lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon$$

2. $$\Pr[X - \mathbb{E}X > \lambda\sqrt{n}] < e^{-\lambda^2/2} = \epsilon$$

But what is $\mathbb{E}X$?

Actually we know that $\Pr[X = 0] \geq \epsilon$ since $|A| \geq \epsilon 2^n$. So (1) tells us that $\mathbb{E}X \leq \lambda\sqrt{n}$. Then (2) gives:

$$\Pr[X > 2\lambda\sqrt{n}] < e^{-\lambda^2/2}$$

from which we can conclude the theorem.