Great Ideas in TCS

A Beginner's Perspective on concentration inequalities

08/25/2020 Spencer Peters

What is a concentration inequality?

Roughly

random varisble
with assumptions.

$$\Pr\left[\,|X - \mathbb{E}[X]| \geq \text{some bound}\right] \leq \text{some bound}.$$

could be
median, mode,
even 0.

Usually these
trade off against
each other.

What is a concentration inequality?

For example, the simplified Chernoff bound:

Suppose $X_1, \ldots, X_n$ are independent RVs whose values are either 0 or 1.

Let $X = \sum_{i=1}^{n} X_i$. Then for all $\delta > 0$,

$$\Pr\left[|X - \mathbb{E}[X]| \geq \delta\, \mathbb{E}[X]\right] \leq 2e^{-\frac{\delta^2 \mathbb{E}[X]}{3}}.$$

Outline

History / Background

- Law of Large Numbers
- Markov, Chebyshev, Chernoff

Azuma-Hoeffding Inequality

- Martingales
- Doob Martingales
- Application: Chromatic Number
- McDiarmid's Inequality
- Proof of Azuma-Hoeffding Inequality

Before concentration inequalities

- (Weak) Law of large numbers

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d.
and define $\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$.

Also suppose (throughout)
$\mathbb{E}[X_i]$ is finite for $i = 1, 2, \ldots n$.

Then for all $\varepsilon > 0$,

$$\lim_{n \to \infty} \Pr\left[\left| \bar{X} - \mathbb{E}[\bar{X}] \right| > \varepsilon \right] = 0.$$

- Asymptotic statement

1500s — Observed by Cardano

1713 — Variant proved by J. Bernoulli

1929 — Modern form (Markov, Khinchin, etc.)

The Original Concentration Inequalities

In the 1850s - 1880s:

Markov: For all positive RVs X,

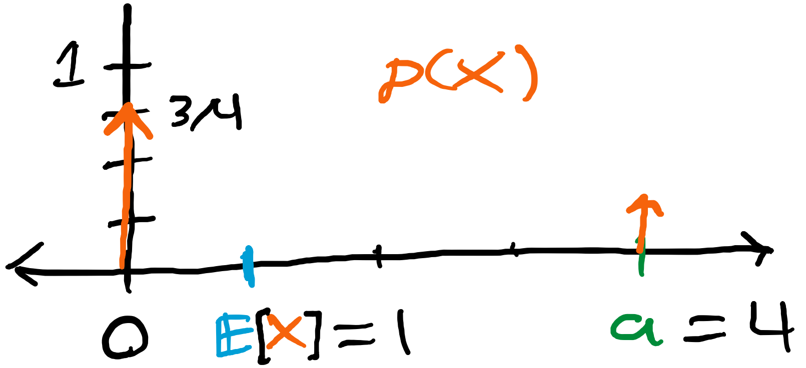
$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

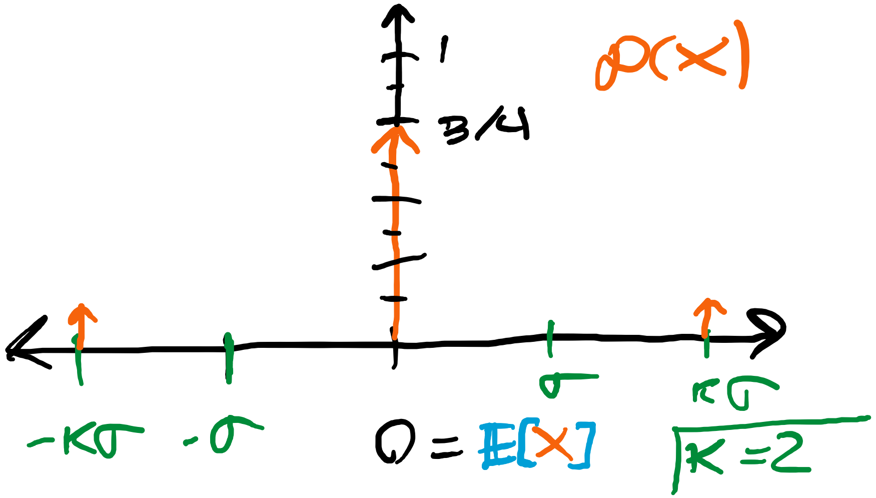
(real-valued, $\mathbb{E}[X]$ finite)

Chebyshev: For all RVs X with finite variance $\mathrm{Var}(X) = \sigma^2$,

$$\Pr[|X - \mathbb{E}[X]| \geq k\sigma] \leq \frac{1}{k^2}$$

Proofs by Picture





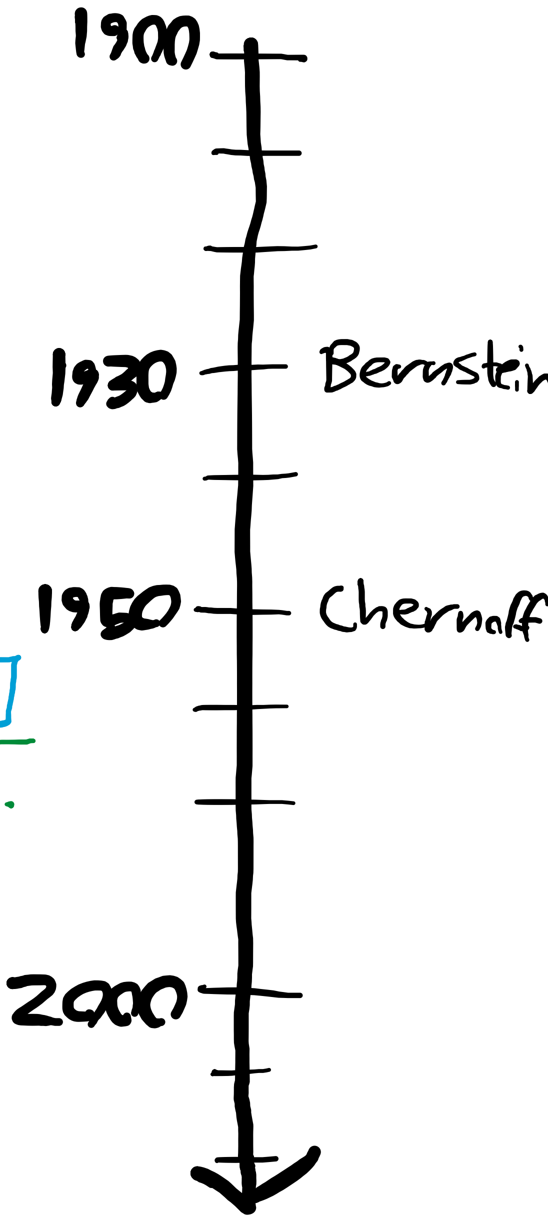
Sums of random variables

Chernoff (simplified version)

Suppose $X_1, \ldots, X_n$ are independent RVs whose values are either 0 or 1.

Let $X = \sum_{i=1}^{n} X_i$. Then for all $\delta > 0$,

$$\Pr\left[|X - \mathbb{E}[X]| \geq \delta\, \mathbb{E}[X] \right] \leq 2e^{-\frac{\delta^2 \mathbb{E}[X]}{3}}.$$



Sums of random variables (2)

Chernoff (simplified version)

Suppose $X_1, \ldots, X_n$ are independent RVs whose values are either 0 or 1.

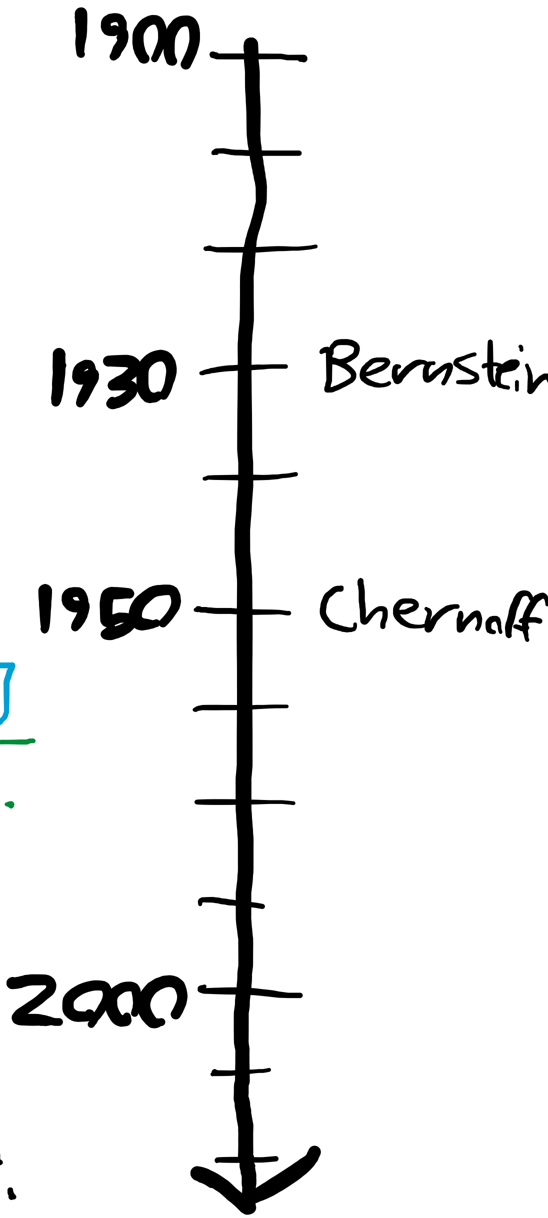
Let $X = \sum_{i=1}^{n} X_i$. Then for all $\delta > 0$,

$$\Pr\left[\,|X - E[X]| \geq \delta E[X]\,\right] \leq 2e^{-\frac{\delta^2 E[X]}{3}}.$$

Example. Suppose $X_i = 1$ with prob p_i.

Choose $\delta = \sqrt{9/E[X]}$. Then

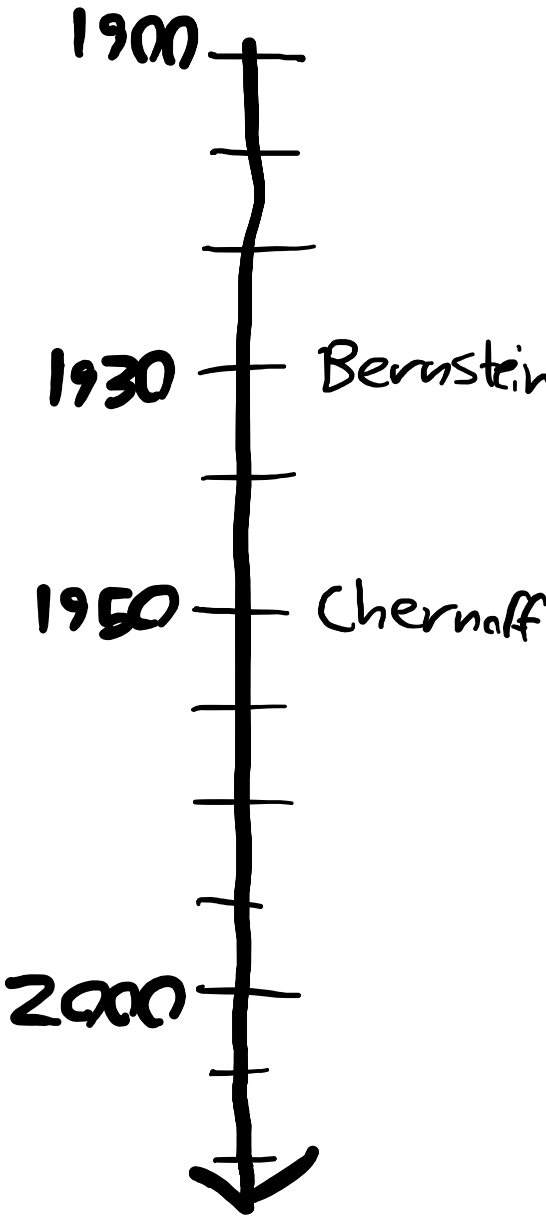
$$\Pr\left[\,|X - E[X]| \geq 3\sqrt{E[X]}\,\right] \leq 2e^{-3} = 0.099.$$



Sums of random variables (3)

$$\Pr[\cdots] \le 2e^{-\frac{\delta^2 \mathbb{E}[X]}{3}}$$

exponential bound.



Sums of random variables (3)

$\Pr[\ \cdots\ ] \leq 2e^{-\frac{\delta^2 \mathbb{E}[X]}{3}}$ exponential bound.

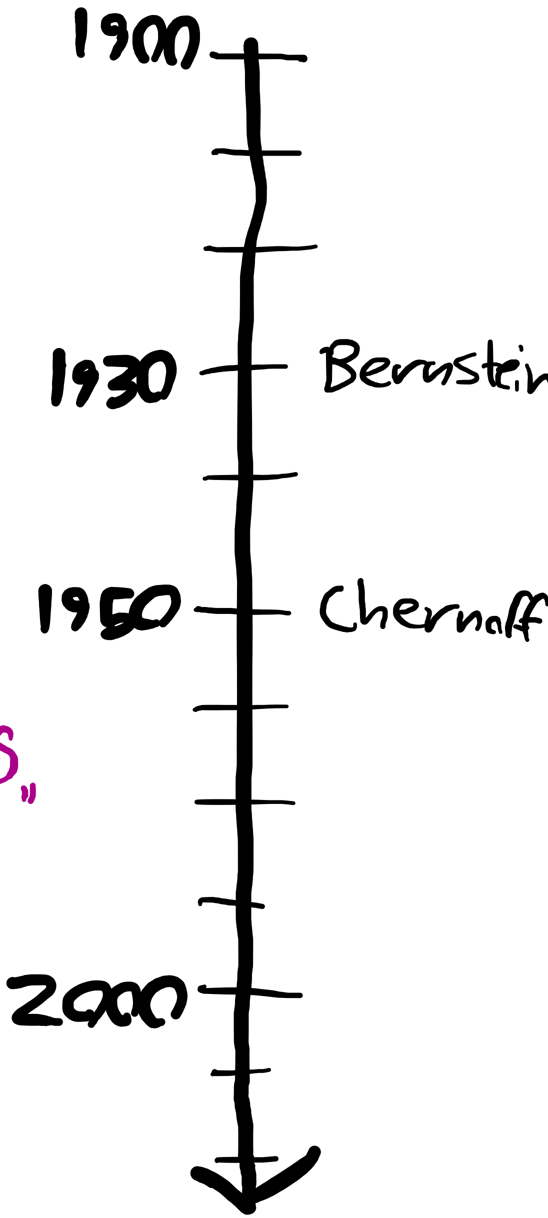
Brilliant idea: apply Markov to e^{sX}.

"moment generating function"

For all $s > 0$,

$$\Pr[X > a] = \Pr[e^{sX} > e^{sa}]$$

$$\leq \mathbb{E}[e^{sX}] / e^{sa}.$$



Sums of random variables (3)

$\Pr[\,\cdots\,] \le 2e^{-\frac{\delta^2 \mathbb{E}[X]}{3}}$ <u>exponential bound.</u>

Brilliant idea: apply Markov to e^{tX}.

"moment generating function"

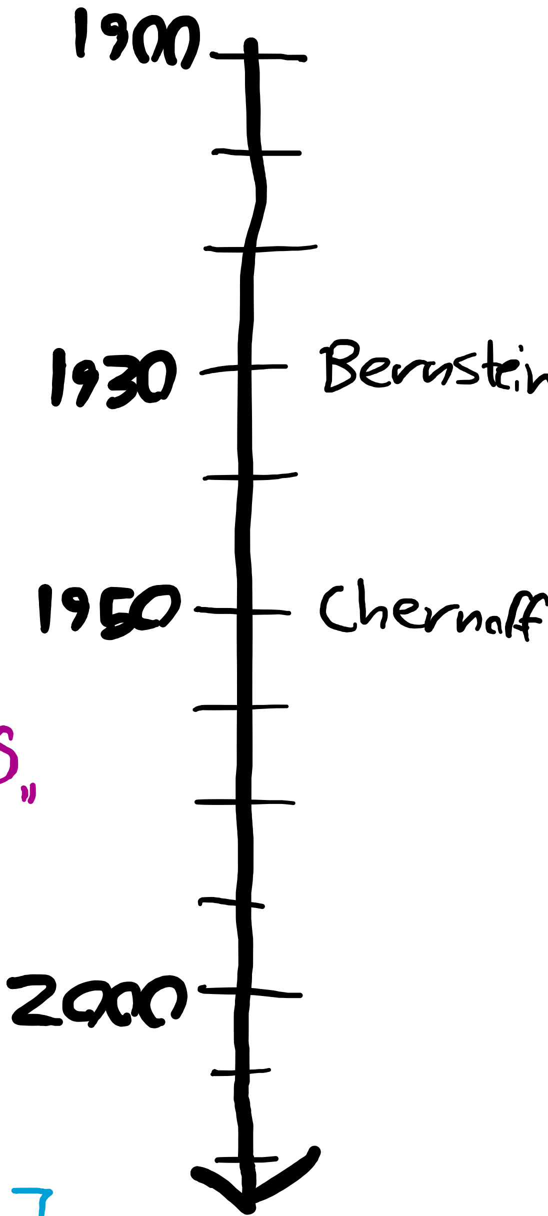
For all $s > 0$,

$\Pr[X > a] = \Pr[e^{sX} < e^{sa}]$

$\le \mathbb{E}[e^{sX}]/e^{sa}$.

When X is a sum of indep RVs $X_1, \ldots X_n$,

$\mathbb{E}[e^{sX}] = \mathbb{E}[\prod_{i=1}^n e^{sX_i}] = \prod_{i=1}^n \mathbb{E}[e^{sX_i}]$.



Beyond sums of random variables (4)

Azuma-Haeffding:

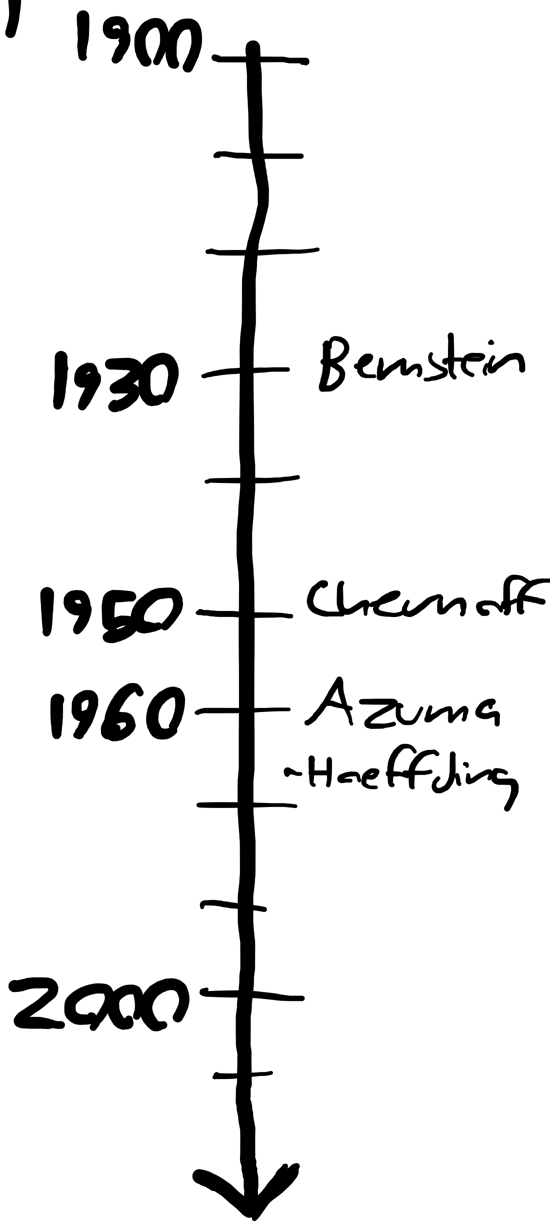
coming soon!

Suppose $X_0, X_1, \ldots, X_n$ is a martingale,

and $|X_t - X_{t-1}| \le c_t$ for $t = 1, 2, \ldots n$.

Then forall $\lambda > 0$,

$$\Pr[|X_t - X_0| > \lambda] \le 2\exp\left(\frac{-\lambda^2}{2\sum_{t=1}^{n} c_t^2}\right).$$



Beyond sums of random variables (4)

Azuma-Haeffding:

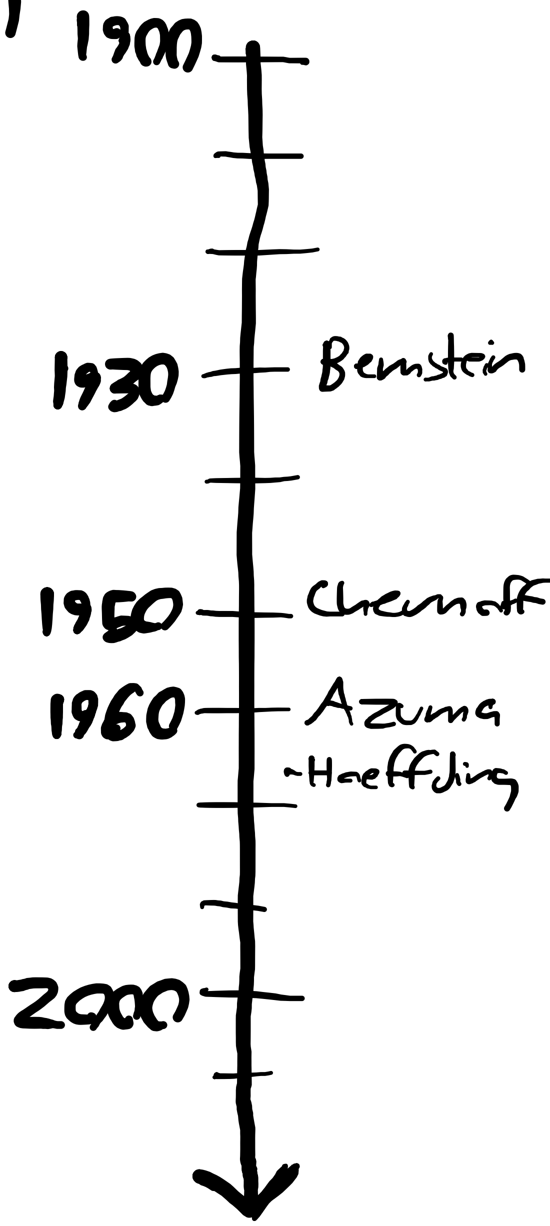
coming soon!

Suppose $X_0, X_1, \ldots, X_n$ is a martingale,
and $|X_t - X_{t-1}| \leq c_k$ for $t = 1, 2, \ldots n$.

Then forall $\lambda > 0$,

$$\Pr[|X_n - X_0| > \lambda] \leq 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

(Sums) If $Y_1, Y_2, \ldots, Y_n$ are indep,
$X_0, X_1, \ldots, X_n$ defined by $X_0 = 0$ and
for $k = 1, 2, \ldots, n$, $X_k = \sum_{i=1}^{k}(Y_k - \mathbb{E}[Y_k])$
is a martingale.



Beyond sums of random variables (4)

Azuma-Haeffding:

coming soon!

Suppose $X_0, X_1, \dots, X_n$ is a martingale,

and $|X_t - X_{t-1}| \leq c_t$ for $t = 1, 2, \dots n$.

Then forall $\lambda > 0$,

$$\Pr[|X_n - X_0| > \lambda] \leq 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_t^2}\right).$$

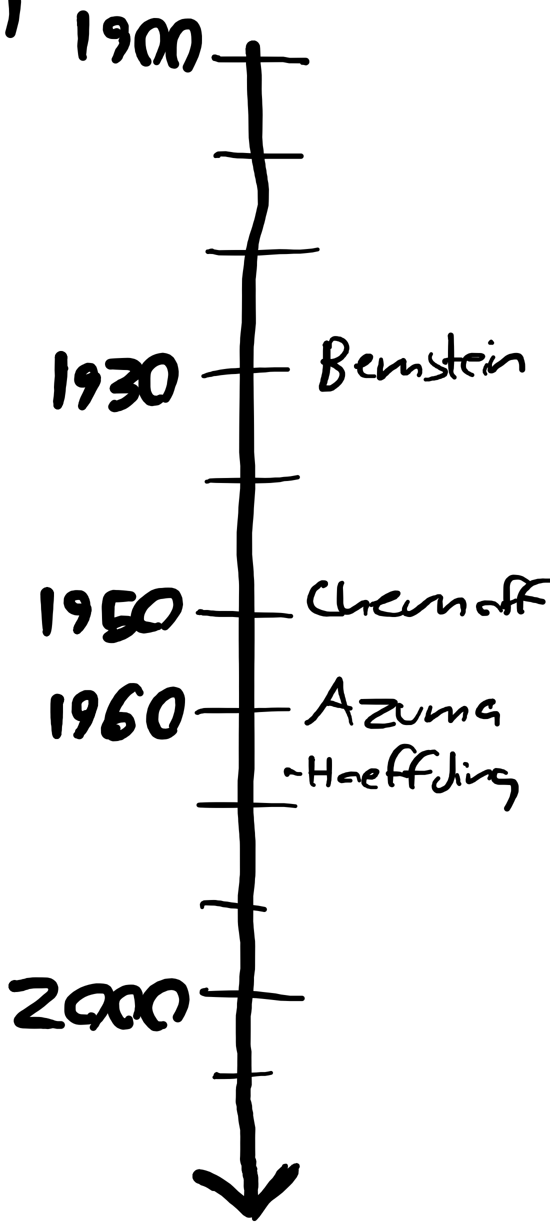
(Sums) If $Y_1, Y_2, \dots, Y_n$ are indep, then

$X_0, X_1, \dots, X_n$ defined by $X_0 = 0$ and

for $t = 1, 2, \dots, n$, $X_t = \sum_{i=1}^{t}(Y_t - \mathbb{E}[Y_t])$

is a martingale.

Hence applies to $\Pr[|\sum_i Y_i - \mathbb{E}[\sum_i Y_i]| > \lambda]$.



...and many more!

Martingales

Suppose $X_0, X_1, \ldots, X_n$ is a martingale ...

what's that?

It means for all $t = 1, 2, \ldots, n$,

$$\mathbb{E}[X_t \mid X_0, \ldots X_{t-1}] = X_{t-1}.$$

(still assuming $\mathbb{E}[X_t]$ is finite.)

Martingales

Suppose $X_0, X_1, \ldots, X_n$ is a martingale ...
what's that?

It means for all $t = 1, 2, \ldots, n$,

$$\mathbb{E}[X_t \mid X_0, \ldots X_{t-1}] = X_{t-1}.$$

(still assuming $\mathbb{E}[X_t]$ is finite.)

Example. Gambler bets Z_t dollars on each of n fair games of "double or nothing".
X_t is the gambler's wealth at time t.
Still a martingale if Z_t depends on X_{t-1}!

Common efficient-market assumption in finance.

Martingales

$$\mathbb{E}[X_t \mid X_0, \ldots X_{t-1}] = X_{t-1}.$$

What exactly is $\mathbb{E}[X_t \mid X_0, \ldots X_{t-1}]$?

Martingales

$$\mathbb{E}[X_t \mid X_0, \dots X_{t-1}] = X_{t-1}.$$

What exactly is $\mathbb{E}[X_t \mid X_0, \dots X_{t-1}]$?

It's not a number like $\mathbb{E}[X_t]$ is.

Martingales

$$\mathbb{E}[X_t \mid X_0, \dots X_{t-1}] = X_{t-1}.$$

What exactly is $\mathbb{E}[X_t \mid X_0, \dots X_{t-1}]$?

It's not a number like $\mathbb{E}[X_t]$ is.

It's actually a <u>random variable</u>

$$\mathbb{E}[X_t \mid X_0, \dots X_{t-1}] : \Omega \longrightarrow \mathbb{R}$$

Its value in the event $X_0 = x_0, \dots X_{t-1} = x_{t-1}$

is the number

$$\mathbb{E}[X_t \mid X_0 = x_0, \dots X_{t-1} = x_{t-1}].$$

Doob martingales

Suppose $X_1, \dots, X_n, Y$ are RVs.

Then the sequence $Y_0, Y_1, \dots Y_n$ defined by

$Y_t = \mathbb{E}[Y \mid X_1, \dots, X_t]$ is a *martingale*.

Doob martingales

Suppose $X_1, \ldots, X_n, Y$ are RVs.

Then the sequence $Y_0, Y_1, \ldots Y_n$ defined by

$Y_t = \mathbb{E}[Y \mid X_1, \ldots, X_t]$ is a martingale.

Proof. It suffices to show

not hard, but involves low level probability theory.

$\mathbb{E}[Y_t \mid X_1, \ldots, X_{t-1}] = Y_{t-1}$. But

$\mathbb{E}[Y_t \mid X_1, \ldots, X_{t-1}] = \mathbb{E}[\mathbb{E}[Y \mid X_1, \ldots, X_t] \mid X_1, \ldots X_{t-1}]$

$= \mathbb{E}[Y \mid X_1, \ldots, X_{t-1}] = Y_{t-1}$. $\square$

$(\mathbb{E}[\mathbb{E}[A \mid B, C] \mid B] = \mathbb{E}[A \mid B])$

Doob martingales

Suppose $X_1, \ldots, X_n, Y$ are RVs.

Then the sequence $Y_0, Y_1, \ldots Y_n$ defined by

$Y_t = \mathbb{E}[Y \mid X_1, \ldots, X_t]$ is a martingale.

Notice

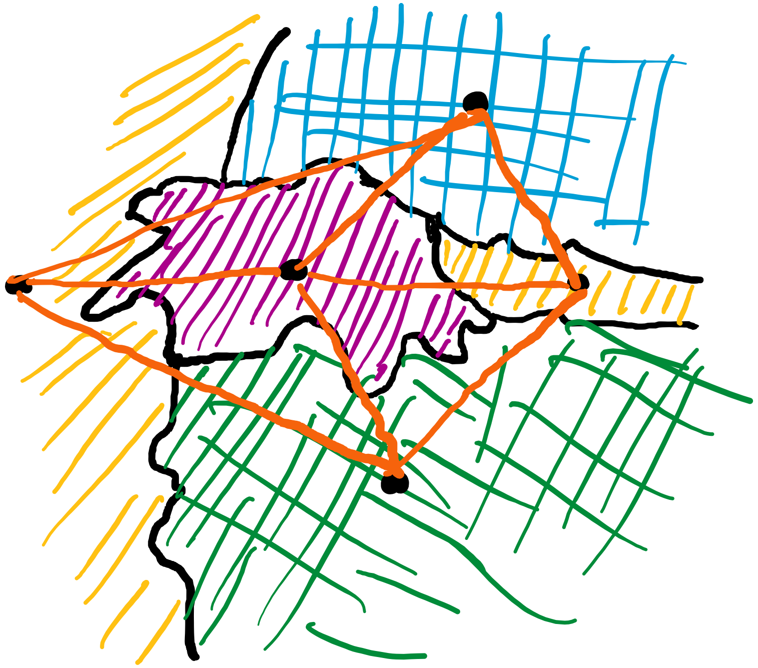
$Y_0 = \mathbb{E}[Y]$, and if $Y = f(X_1, \ldots X_n)$

then $Y_n = Y$.

$Y_0, Y_1, \ldots Y_n$ can be thought of as a series of increasingly good estimates of Y.

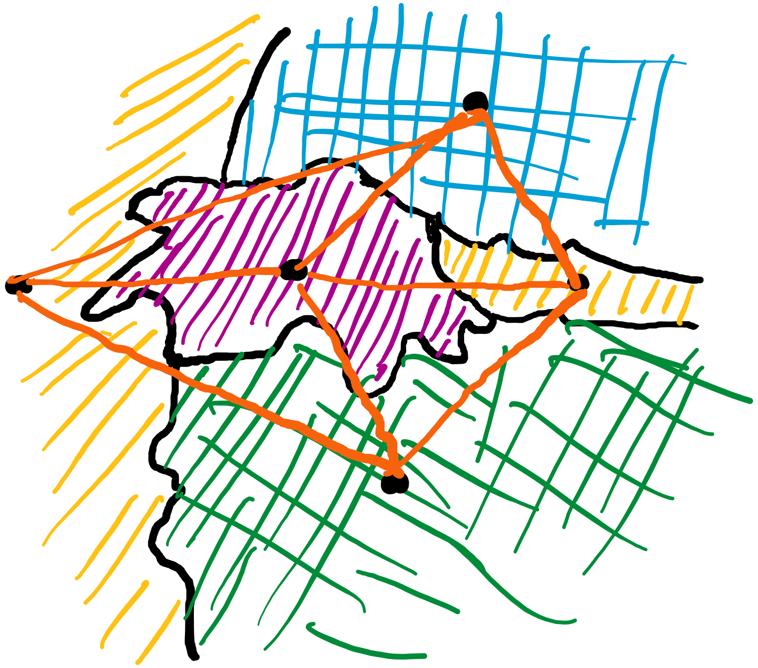
Application: Chromatic Number (from [2])

Recall that given a graph G,
the chromatic number $\chi(G)$
is the fewest number of colors
required to color all vertices
so that no two adjacent vertices
have the same color.



Application: Chromatic Number

Recall that given a graph G,
the chromatic number $\chi(G)$
is the fewest number of colors required to color all vertices so that no two adjacent vertices have the same color.



Apparently, if you make the random graph $G = G_{n,p}$ on n vertices by adding each possible edge independently with probability p, finding $\mathbb{E}[\chi(G)]$ is hard.

Application: Chromatic Number

However, we can show $\chi(G_{n,p})$ is close to $\mathbb{E}[\chi(G_{n,p})]$.

Application: Chromatic Number

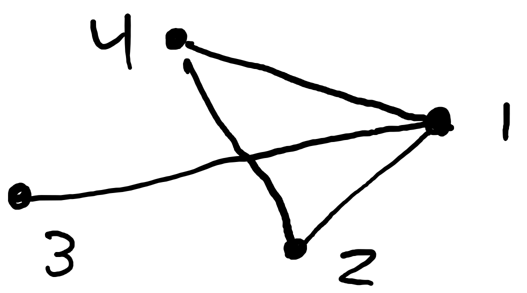
However, we can show $\chi(G_{n,p})$ is close to $\mathbb{E}[\chi(G_{n,p})]$.

For $t = 1, 2, \dots n$ let G_t be the subgraph of $G_{n,p}$ induced by the vertices $1, 2, \dots, t$.

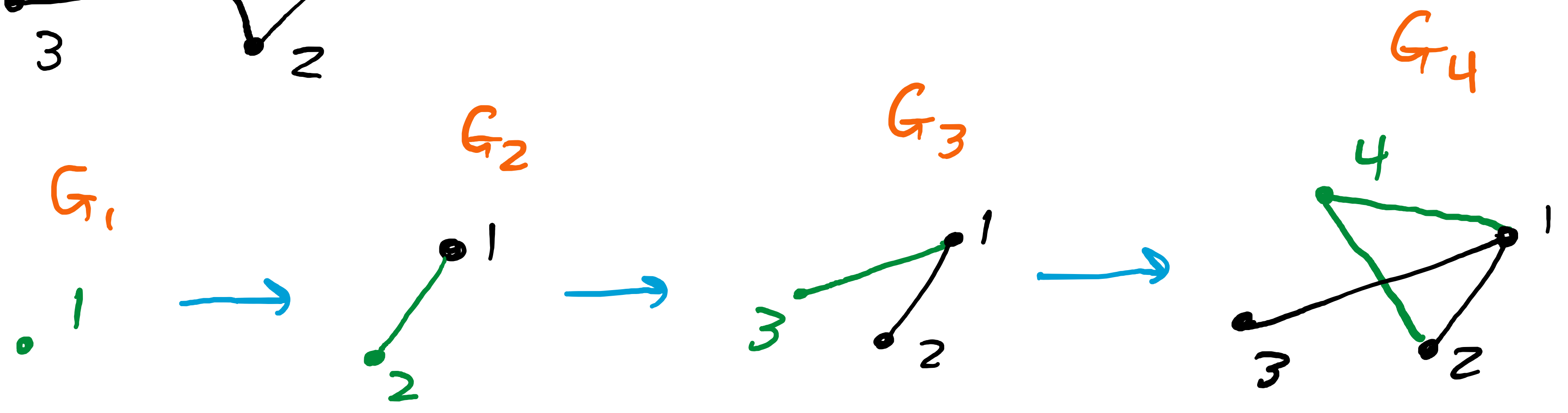
Application: Chromatic Number

However, we can show $\chi(G_{n,p})$ is close to $\mathbb{E}[\chi(G_{n,p})]$.

For $t = 1, 2, \dots n$ let G_t be the subgraph of $G_{n,p}$ induced by the vertices $1, 2, \dots, t$.



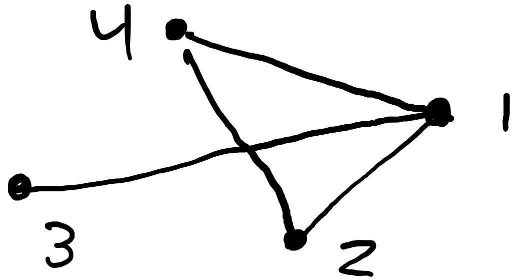
Example $G_{4,p}$



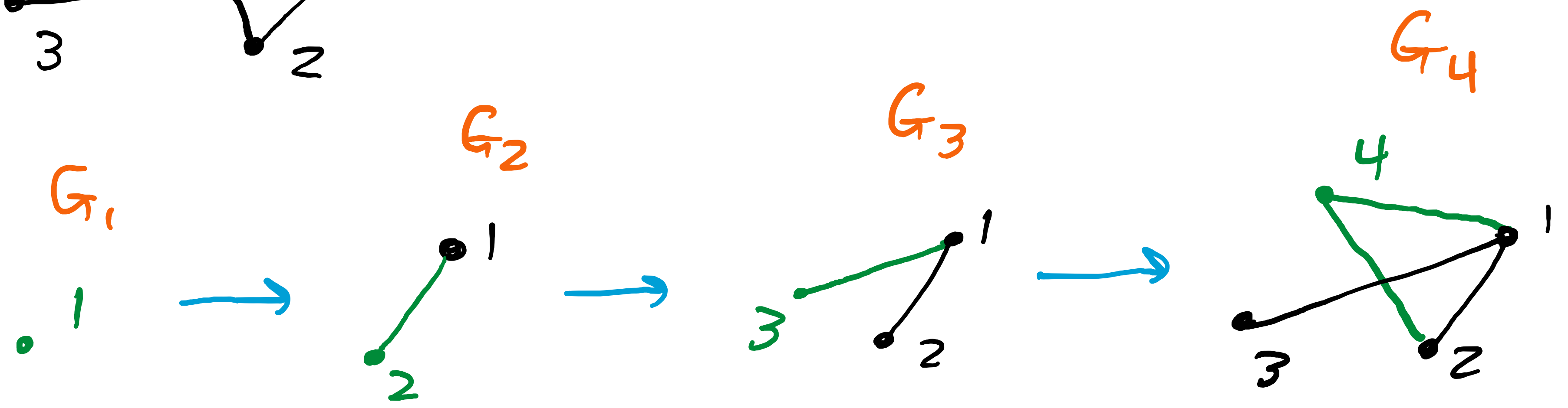
Application: Chromatic Number

However, we can show $\chi(G_{n,p})$ is close to $\mathbb{E}[\chi(G_{n,p})]$.

For $t = 1, 2, \dots n$ let G_t be the subgraph of $G_{n,p}$ induced by the vertices $1, 2, \dots, t$.

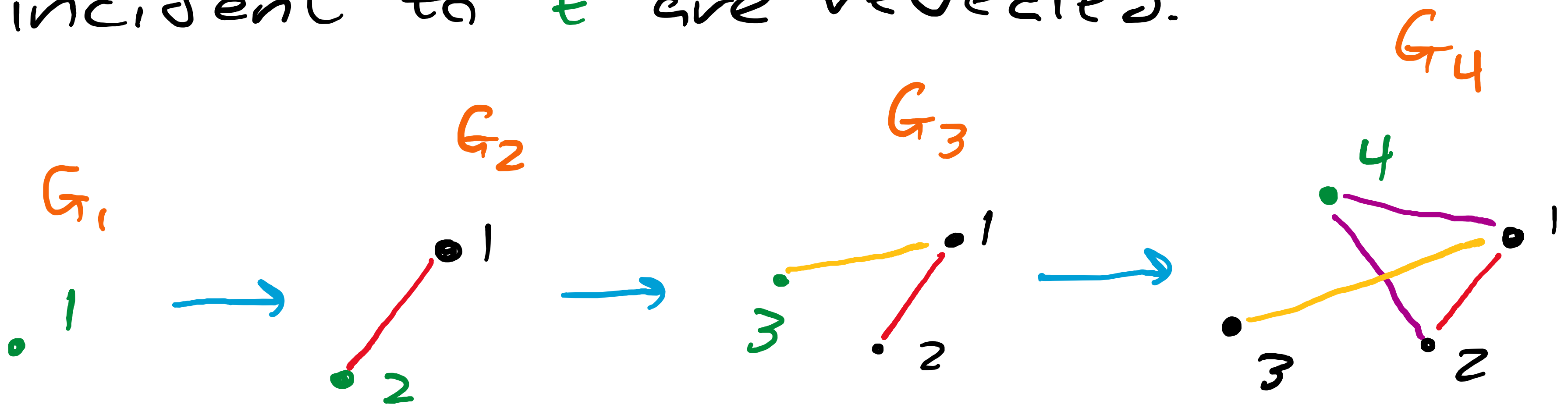


Example $G_{4,p}$



Application: Chromatic Number

When we add vertex t, some random edges incident to t are revealed.



Any pair of graphs differing in only the revealed edges can differ in chromatic number by at most 1, since all revealed edges can be assigned a new color.

Application: Chromatic Number

Thus if we define, for $t = 0, 1, \dots, n$,

$$X_t = \mathbb{E}[\chi(G) \mid G_1, \dots, G_t]$$

the sequence $X_0, \dots, X_n$ is a Doob martingale,

and for all $t = 1, 2, \dots n$,

$$|X_t - X_{t-1}| \leq 1.$$

Application: Chromatic Number

Thus if we define, for $t = 0, 1, \ldots, n$,

$$X_t = \mathbb{E}[\chi(G) \mid G_1, \ldots, G_t]$$

the sequence $X_0, \ldots, X_t$ is a Doob martingale, and for all $t = 1, 2, \ldots n$,

$$|X_t - X_{t-1}| \le 1.$$

Thus by Azuma-Hoeffding,

$$\Pr[|\chi(G) - \mathbb{E}[\chi(G)]| \ge \lambda] \le 2\exp\left(\frac{-\lambda^2}{2\sum_{i=1}^{n} 1^2}\right)$$

$$\Rightarrow \Pr[|\chi(G) - \mathbb{E}[\chi(G)]| \ge k\sqrt{n}] \le 2\exp\left(-\frac{1}{2}k^2\right).$$

McDiarmid's Inequality

Recall the Doob martingale $Y_0, Y_1, \ldots Y_n$

defined by $Y_t = \mathbb{E}[Y \mid X_1, \ldots X_t]$.

If we define $Y = f(X_1, \ldots X_n)$ and

for all $i = 1, \ldots, k$ and values $x_i, x_i', \vec{x}_{-i}$,

$$|f(x_i, \vec{x}_{-i}) - f(x_i', \vec{x}_{-i})| \leq c_i,$$

then for all $t = 1, \ldots, n$,

$$|Y_t - Y_{t-1}| \leq c_t.$$

McDiarmid's Inequality

Thus by Azuma-Hoeffding,

letting $f = f(X_1, \dots X_n)$,

$$\mathbb{E}[\,|f - \mathbb{E}[f]| \geq \lambda\,] \leq 2\exp\left(\frac{-\lambda^2}{2\sum_{i=1}^{n} c_i^2}\right).$$

Proof of Azuma-Hoeffding (following [2])

For $t = 1, 2, \ldots n$, define

$$\Delta X_t = X_t - X_{t-1}.$$

Suppose $X_0, X_1, \ldots, X_n$ is a <u>martingale</u>, and $|X_k - X_{k-1}| \leq c_k$ for $k = 1, 2, \ldots n$.

Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \geq \lambda] \leq 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Proof of Azuma-Hoeffding (following [2])

For $t = 1, 2, \ldots n$, define

$$\Delta X_t = X_t - X_{t-1}.$$

We have $|\Delta X_t| \le c_t$,

and since $X_0, X_1, \ldots, X_n$

is a martingale, $\mathbb{E}[\Delta X_t \mid X_1, X_2, \ldots X_{t-1}] = 0.$

Suppose $X_0, X_1, \ldots, X_n$ is a martingale, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \ldots n$. Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \ge \lambda] \le 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Proof of Azuma-Hoeffding

For $t = 1, 2, \ldots n$, define

$$\Delta X_t = X_t - X_{t-1}.$$

We have $|\Delta X_t| \le c_t$,

and since $X_0, X_1, \ldots, X_n$ is a martingale, $\mathbb{E}[\Delta X_t \mid X_1, X_2, \ldots X_{t-1}] = 0$.

We will prove

$$\Pr[X_n - X_0 \ge \varepsilon] \le \exp\left(\frac{-\varepsilon^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Suppose $X_0, X_1, \ldots, X_n$ is a <u>martingale</u>, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \ldots n$.

Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \ge \lambda] \le 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Proof of Azuma-Hoeffding

For $t = 1, 2, \ldots n$, define

$$\Delta X_t = X_t - X_{t-1}.$$

We have $|\Delta X_t| \le c_t$,

and since $X_0, X_1, \ldots, X_n$ is a martingale, $\mathbb{E}[\Delta X_t \mid X_1, X_2, \ldots X_{t-1}] = 0.$

Suppose $X_0, X_1, \ldots, X_n$ is a martingale, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \ldots n$.

Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \ge \lambda] \le 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^n c_k^2}\right).$$

We will prove

$$\Pr[X_n - X_0 \ge \varepsilon] \le \exp\left(\frac{-\varepsilon^2}{2\sum_{k=1}^n c_k^2}\right).$$

The other direction is symmetric, and the result follows from union bound.

Proof of Azuma-Hoeffding

Suppose $X_0, X_1, \dots, X_n$ is a martingale, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \dots n$.

Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \ge \lambda] \le 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Use the brilliant idea from Chernoff bounds.

For all $s > 0$,

$$\Pr[X_n - X_0 > \lambda]$$
$$= \Pr[\exp(s(X_n - X_0)) > e^{s\lambda}]$$
$$\le \mathbb{E}[\exp(s(X_n - X_0))] / e^{s\lambda} \qquad \text{(Markov)}$$

Proof of Azuma-Hoeffding

Use the brilliant idea from Chernoff bounds.

For all $s > 0$,

$$\Pr[X_n - X_0 > \lambda]$$
$$= \Pr[\exp(s(X_n - X_0)) > e^{s\lambda}]$$
$$\le \mathbb{E}[\exp(s(X_n - X_0))] / e^{s\lambda} \quad \text{(Markov)}$$
$$= \mathbb{E}[\exp(s \textstyle\sum_{i=1}^n \Delta X_i)] / e^{s\lambda}$$

Suppose $X_0, X_1, \ldots, X_n$ is a martingale, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \ldots n$.

Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \ge \lambda] \_ \; 2 \exp\left(\frac{-\lambda^2}{2\sum_{k=1}^n c_k^2}\right).$$

Proof of Azuma-Hoeffding

Use the brilliant idea from Chernoff bounds.

For all $s > 0$,

$$\Pr[X_n - X_0 > \lambda]$$
$$= \Pr[\exp(s(X_n - X_0)) > e^{s\lambda}]$$
$$\le \mathbb{E}[\exp(s(X_n - X_0))] / e^{s\lambda} \qquad \text{(Markov)}$$
$$= \mathbb{E}[\exp(s \textstyle\sum_{i=1}^{n} \Delta X_i)] / e^{s\lambda}$$
$$= \mathbb{E}[\textstyle\prod_{i=1}^{n} \exp(s \Delta X_i)] / e^{s\lambda}.$$

Suppose $X_0, X_1, \ldots, X_n$ is a martingale, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \ldots n$.

Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \ge \lambda] \le 2 \exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Proof of Azuma-Hoeffding

Suppose $X_0, X_1, \dots, X_n$ is a martingale, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \dots n$.

Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \ge \lambda] \le 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Use the brilliant idea from Chernoff bounds.

For all $s > 0$,

$$\Pr[X_n - X_0 > \lambda]$$
$$= \Pr[\exp(s(X_n - X_0)) > e^{s\lambda}]$$
$$\le \mathbb{E}[\exp(s(X_n - X_0))] / e^{s\lambda} \quad \text{(Markov)}$$
$$= \mathbb{E}[\exp(s \textstyle\sum_{i=1}^{n} \Delta X_i)] / e^{s\lambda}$$
$$= \mathbb{E}[\textstyle\prod_{i=1}^{n} \exp(s\, \Delta X_i)] / e^{s\lambda}$$
$$= \mathbb{E}[\textstyle\prod_{i=1}^{n-1} \exp(s\, \Delta X_i)\, \mathbb{E}[\exp(s\, \Delta X_n) \mid X_0, \dots, X_{n-1}]] / e^{s\lambda}$$

$$(\mathbb{E}[AB] = \mathbb{E}[A\, \mathbb{E}[B \mid A]])$$

Proof of Azuma-Hoeffding

$$\dots = \mathbb{E}\Big[\textstyle\prod_{i=1}^{n-1} \exp(s \; \Delta X_i \;) \; \mathbb{E}[\exp(s \, \Delta X_n) \; | X_0, \dots, X_{n-1}]\Big] / e^{s\lambda}$$

Suppose $X_0, X_1, \dots, X_n$ is a martingale, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \dots n$.

Then for all $\lambda > 0$,

$$\Pr[|X_n - X_0| \ge \lambda] \_ \; 2 \exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Repeating this last step gives

$$\Pr[X_n - X_0 > \lambda] \le \mathbb{E}\Big[\textstyle\prod_{i=1}^{n} \mathbb{E}[\exp(s \, \Delta x_i) \, | X_0, \dots, X_{i-1}]\Big]$$

Proof of Azuma-Hoeffding

$$\ldots = \mathbb{E}\Big[\prod_{i=1}^{n-1} \exp(s\ \Delta X_i\ )$$
$$\mathbb{E}\big[\exp(s\,\Delta X_n)$$
$$|X_0, \ldots, X_{n-1}\big]\Big] / e^{s\lambda}$$

Suppose $X_0, X_1, \ldots, X_n$ is a martingale, and $|X_k - X_{k-1}| \le c_k$ for $k = 1, 2, \ldots n$.
Then for all $\lambda > 0$,
$$\Pr[|X_n - X_0| \ge \lambda] \le 2\exp\left(\frac{-\lambda^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

Repeating this last step gives

$$\Pr[X_n - X_0 > \lambda] \le \mathbb{E}\Big[\prod_{i=1}^{n} \mathbb{E}\big[\exp(s\,\Delta X_i)\,|\,X_0, \ldots, X_{i-1}\big]\Big]$$

To finish the proof, we give a bound on

$$\mathbb{E}\big[\exp(s\ \Delta X_i)\ |\ X_0 = x_0, \ldots, X_{i-1} = x_{i-1}\big]$$

independent of $x_0, \ldots, x_{i-1}$.

Proof of Azuma-Hoeffding

Bound $\mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}]$

Recall (1) $\mathbb{E}[\Delta X_i \mid X_0, \ldots, X_{i-1}] = 0$,

(2) $\exp(s\,\Delta X_i)$ is convex in ΔX_i,

(3) $|\Delta X_i| \leq c_i$.

Proof of Azuma-Hoeffding

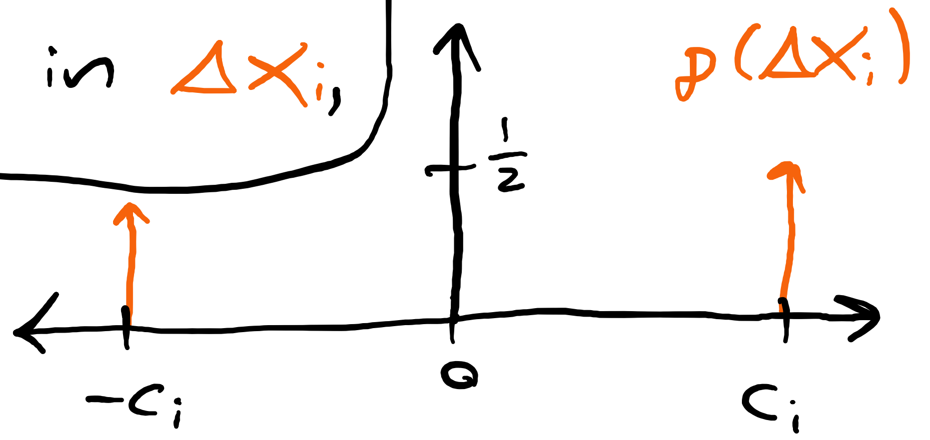
Bound $\mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}]$

Recall (1) $\mathbb{E}[\Delta X_i \mid X_0, \ldots, X_{i-1}] = 0$,

(2) $\exp(s\,\Delta X_i)$ is convex in ΔX_i,

(3) $|\Delta X_i| \le c_i$.

All these imply worst case is:



Proof of Azuma-Hoeffding

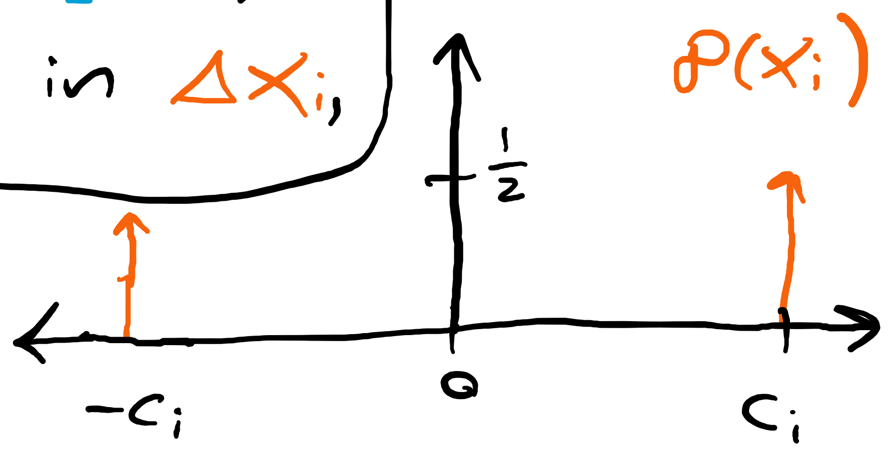
Bound $\mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}]$

Recall (1) $\mathbb{E}[\Delta X_i \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] = 0$,

(2) $\exp(s\,\Delta X_i)$ is convex in ΔX_i,

(3) $|\Delta X_i| \leq c_i$.

All these imply worst case is:



Hence

$$\mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] \leq \frac{\exp(-s\,c_i) + \exp(s\,c_i)}{2}.$$

Proof of Azuma-Hoeffding

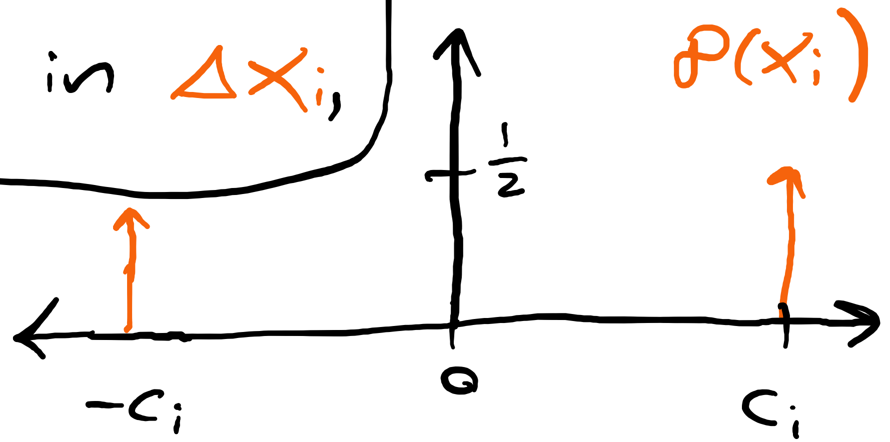
Bound $r_i = \mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}]$

Recall (1) $\mathbb{E}[\Delta X_i \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] = 0$,

(2) $\exp(s\,\Delta X_i)$ is convex in ΔX_i,

(3) $|\Delta X_i| \le c_i$.

All these imply worst case is:



Hence

$$\mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] \le \frac{\exp(-s\,c_i) + \exp(s\,c_i)}{2}$$

$$\le \exp((s\,c_i)^2/2) \quad \text{(Taylor series calculation).}$$

Proof of Azuma–Hoeffding

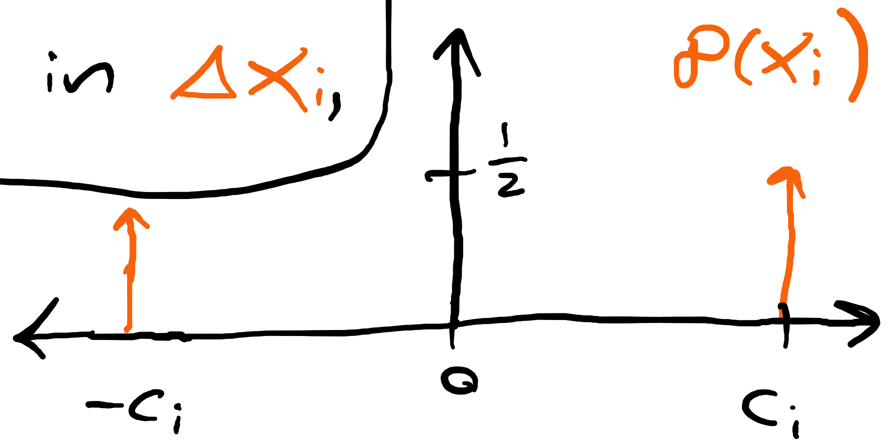
Bound $r_i = \mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}]$

Recall (1) $\mathbb{E}[\Delta X_i \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] = 0$,

(2) $\exp(s\,\Delta X_i)$ is convex in ΔX_i,

(3) $|\Delta X_i| \le c_i$.

All these imply worst case is:



Hence

$$\mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] \le \frac{\exp(-s\,c_i) + \exp(s\,c_i)}{2}$$

$$\le \exp((s\,c_i)^2/2) \quad \left(\text{Taylor series calculation}\right).$$

Note we really need this quadratic dependence.

Proof of Azuma-Hoeffding

We found

$$\mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] \leq \exp((sc_i)^2/2).$$
$$= b_i.$$

Proof of Azuma-Hoeffding

We found

$r_i = \mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] \le \exp((s c_i)^2/2)$.

$= b_i$.

From earlier, we had

$$\Pr[X_n - X_0 > \lambda] \le \frac{\mathbb{E}[\prod_{i=1}^n \mathbb{E}[\exp(s\,\Delta X_i) \mid X_0, \ldots, X_{i-1}]]}{\exp(s\lambda)}$$

$$\le \mathbb{E}[\textstyle\prod_{i=1}^n b_i]/\exp(s\lambda) = \prod_{i=1}^n b_i/\exp(s\lambda).$$

Proof of Azuma-Hoeffding

We found

$$r_i = \mathbb{E}[\exp(s\,\Delta X_i) \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}] \le \exp((s c_i)^2/2)$$
$$= b_i.$$

$$\Pr[X_n - X_0 > \lambda] \le \frac{\mathbb{E}[\prod_{i=1}^n \mathbb{E}[\exp(s\,\Delta X_i) \mid X_0, \ldots, X_{i-1}]]}{\exp(s\lambda)}$$
$$\le \mathbb{E}[\textstyle\prod_{i=1}^n b_i]/\exp(s\lambda) = \prod_{i=1}^n b_i/\exp(s\lambda).$$

Putting these together gives

$$\Pr[X_n - X_0 > \lambda] \le \exp\left(\frac{s^2}{2}\sum_{i=1}^n c_i^2 - s\lambda\right).$$

Proof of Azuma-Hoeffding

Putting these together gives

$$\Pr[X_n - X_0 > \lambda] \leq \exp\left(\frac{s^2}{2}\sum_{i=1}^{n} c_i^2 - s\lambda\right).$$

Finally, minimize over $s > 0$.
The minimizer is $s = \lambda / \sum_{i=1}^{n} c_i^2$.

Proof of Azuma-Hoeffding

Putting these together gives

$$\Pr[X_n - X_0 > \lambda] \leq \exp\left(\frac{s^2}{2}\sum_{i=1}^{n} c_i^2 - s\lambda\right).$$

Finally, minimize over $s > 0$.

The minimizer is $s = \lambda / \sum_{i=1}^{n} c_i^2$.

Plugging this in yields

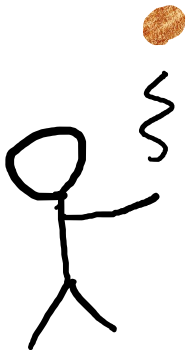
$$\Pr[X_n - X_0 > \lambda] \leq \exp\left(\frac{-s^2}{2\sum_{i=1}^{n} c_i^2}\right)$$

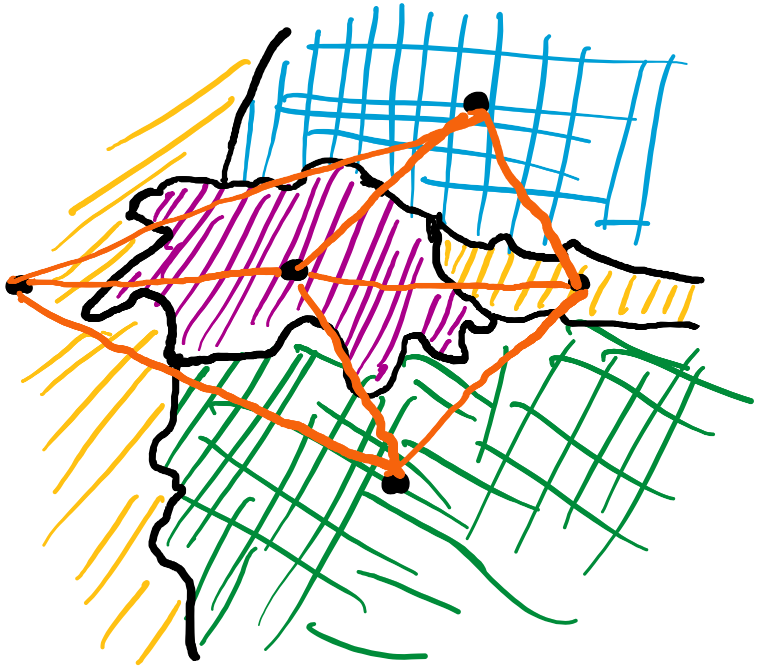
as desired.

Summary

Concentration inequalities bound the difference between random variables and their expectations.
(medians, 0)

They extend past sums of independent random variables to martingales and functions of random variables, enabling fascinating applications!





Thanks for listening!

Extra thanks to Michela Meister for helpful advice and feedback on practice talks (all errors mine).

References

[1] Wikipedia,

[2] Probability & Computing, Mitzenmacher and Upfal, 2nd ed.,

[3] Lecture notes, randomized algorithms course at ETH Zürich, Fall 2019 (A. Steger & E. Welzl).