Concentration of Measure 3

Jarosław Błasiok

January 14, 2020

Concentration of Measure: More general functions of independent variables

We were discussing that $\sum_i X_i$ is extremely concentrated around its expectation if all X_i are nice and sufficiently independent.

New goal

We want to show large deviation bounds

$$\mathbb{P}(|f(X_1, \dots X_n) - \mathbb{E}f(X_1, \dots X_n)| > \lambda) < \delta,$$

for wide classes of functions f and independent variables X_i.

Concentration of Measure: More general functions of independent variables

We were discussing that $\sum_i X_i$ is extremely concentrated around its expectation if all X_i are nice and sufficiently independent.

New goal

We want to show large deviation bounds

$$\mathbb{P}(|f(X_1, \dots X_n) - \mathbb{E}f(X_1, \dots X_n)| > \lambda) < \delta,$$

for wide classes of functions f and independent variables X_i.

Weird phenomenon

Often it is hard to bound $\mathbb{E}f$.
Nevertheless, we can show strong concentration around this expectation.

A non-example and an example

Non-example

If $f(X_1, \dots X_n) = X_1$ we cannot really hope to get amazing concentration.

A non-example and an example

Non-example

If $f(X_1, \ldots X_n) = X_1$ we cannot really hope to get amazing concentration.

Take-away message: We need our function f to depend mildly on each variable.

A non-example and an example

Non-example

If $f(X_1, \dots X_n) = X_1$ we cannot really hope to get amazing concentration.

Take-away message: We need our function f to depend mildly on each variable.

Example

Consider $A \in \mathbb{R}^{d \times d}$ random matrix with i.i.d. entries A_{ij} (e.g. $A_{ij} = \pm 1$). Can we show $\mathbb{P}(\|A\| - \mathbb{E}\|A\| > \lambda) < \delta$?

A non-example and an example

Non-example

If $f(X_1, \ldots X_n) = X_1$ we cannot really hope to get amazing concentration.

Take-away message: We need our function f to depend mildly on each variable.

Example

Consider $A \in \mathbb{R}^{d \times d}$ random matrix with i.i.d. entries A_{ij} (e.g. $A_{ij} = \pm 1$). Can we show $\mathbb{P}(\|A\| - \mathbb{E}\|A\| > \lambda) < \delta$?

Function $f(A_{11}, A_{12}, \ldots, A_{nn}) := \|A\|$ is very complicated (much more so than $\sum A_{ij}$), but depends mildly on each variable. There is hope.

Concentration of operator norm

What is $\mathbb{E}\|A\|$?

If A_{ij} are ± 1, there is non-zero probability that $\|A\| = n$. We could imagine that $\mathbb{E}\|A\|$ is as large as $\Theta(n)$.

Concentration of operator norm

What is $\mathbb{E}\|A\|$?

If A_{ij} are ± 1, there is non-zero probability that $\|A\| = n$. We could imagine that $\mathbb{E}\|A\|$ is as large as $\Theta(n)$.
In fact, via the net argument, we have (almost) shown during the last talk that $\mathbb{E}\|A\| = \mathcal{O}(\sqrt{n})$ (but it is non-trivial).

Concentration of operator norm

What is $\mathbb{E}\|A\|$?

If A_{ij} are ± 1, there is non-zero probability that $\|A\| = n$. We could imagine that $\mathbb{E}\|A\|$ is as large as $\Theta(n)$.
In fact, via the net argument, we have (almost) shown during the last talk that $\mathbb{E}\|A\| = \mathcal{O}(\sqrt{n})$ (but it is non-trivial).

We will show that $\|A\| - \mathbb{E}\|A\|$ is often 1-subgaussian!

Efron-Stein inequality – bounding variance

Goal

We want to show that $Var(\|A\|) \lesssim 1$.

Efron-Stein inequality – bounding variance

Goal

We want to show that $Var(\|A\|) \lesssim 1$.

Theorem (Efron-Stein)

$$Var(f) \leq \mathbb{E}_X \left[\sum_i \underset{X_i'}{Var}(f(X_1, \ldots, X_i', \ldots X_n)) \right]$$

Efron-Stein inequality – bounding variance

Goal

We want to show that $Var(\|A\|) \lesssim 1$.

Theorem (Efron-Stein)

$$Var(f) \leq \mathbb{E}_X \left[\sum_i \underset{X_i'}{Var}(f(X_1, \ldots, X_i', \ldots X_n)) \right]$$

Definition (bounded-difference function)

We say that f has bounded differences by vector $a_1, \ldots a_n$, if for any $X_1, \ldots X_n$, and X_i', we have

$$|f(X_1, \ldots X_i', \ldots X_n) - f(X_1, \ldots X_n)| \leq a_i.$$

Efron-Stein inequality – bounding variance

Goal

We want to show that $Var(\|A\|) \lesssim 1$.

Theorem (Efron-Stein)

$$Var(f) \le \mathbb{E}_X \left[\sum_i \underset{X_i'}{Var}(f(X_1, \ldots, X_i', \ldots X_n)) \right]$$

Definition (bounded-difference function)

We say that f has bounded differences by vector $a_1, \ldots a_n$, if for any $X_1, \ldots X_n$, and X_i', we have

$$|f(X_1, \ldots X_i', \ldots X_n) - f(X_1, \ldots X_n)| \le a_i.$$

By Efron-Stein inequality, we have $Var(f) \le \sum a_i^2$.
In particular $Var(\|A\|) \lesssim n$, (Std. Dev $\mathcal{O}(\sqrt{n})$).

Bit more fancy Efron-Stein

Theorem (Efron-Stein)

$$Var(f) \lesssim \mathbb{E}_X \left[\sum_i \mathbb{E}_{X_i'} [f(X_1, \ldots, X_i, \ldots X_n) - f(X_1, \ldots X_i', \ldots X_n)]_+^2 \right],$$

where $[blah]_+ = [blah]\mathbf{1}[blah > 0]$.

Bit more fancy Efron-Stein

Theorem (Efron-Stein)

$$Var(f) \lesssim \mathbb{E}_X \left[\sum_i \mathbb{E}_{X_i'} [f(X_1, \ldots, X_i, \ldots X_n) - f(X_1, \ldots X_i', \ldots X_n)]_+^2 \right],$$

where $[blah]_+ = [blah]\mathbf{1}[blah > 0]$.

Bounding $Var(\|A\|) \leq \mathcal{O}(1)$

- Take random matrix A. We have $\|A\| = \sup_{u,v \in S_2} u^T Av$.
  Fix u, v that realize this supremum.

Bit more fancy Efron-Stein

Theorem (Efron-Stein)

$$Var(f) \lesssim \mathbb{E}_X \left[\sum_i \mathbb{E}_{X_i'}[f(X_1, \ldots, X_i, \ldots X_n) - f(X_1, \ldots X_i', \ldots X_n)]_+^2 \right],$$

where $[blah]_+ = [blah]\mathbf{1}[blah > 0]$.

Bounding $Var(\|A\|) \leq \mathcal{O}(1)$

- Take random matrix A. We have $\|A\| = \sup_{u,v \in S_2} u^T A v$. Fix u, v that realize this supremum.
- Consider $\|A\| - \|\tilde{A}\|$, where $\tilde{A} = A + \alpha E_{ij}$ and E_{ij} is a matrix with non-zero entry only at the position A_{ij}, and $\alpha \leq 2$.

Bit more fancy Efron-Stein

Theorem (Efron-Stein)

$$Var(f) \lesssim \mathbb{E}_X \left[\sum_i \mathbb{E}_{X_i'} [f(X_1, \ldots, X_i, \ldots X_n) - f(X_1, \ldots X_i', \ldots X_n)]_+^2 \right],$$

where $[blah]_+ = [blah]\mathbf{1}[blah > 0]$.

Bounding $Var(\|A\|) \leq \mathcal{O}(1)$

- Take random matrix A. We have $\|A\| = \sup_{u,v \in S_2} u^T A v$. Fix u, v that realize this supremum.
- Consider $\|A\| - \|\tilde{A}\|$, where $\tilde{A} = A + \alpha E_{ij}$ and E_{ij} is a matrix with non-zero entry only at the position A_{ij}, and $\alpha \leq 2$.
- Note that $\|A\| - \|\tilde{A}\| \leq u^T A v - u^T (A + \alpha E_{ij}) v \leq 2 u_i v_j$.

Bit more fancy Efron-Stein

Theorem (Efron-Stein)

$$Var(f) \lesssim \mathbb{E}_X \left[\sum_i \mathbb{E}_{X_i'}[f(X_1, \ldots, X_i, \ldots X_n) - f(X_1, \ldots X_i', \ldots X_n)]_+^2 \right],$$

where $[blah]_+ = [blah]\mathbf{1}[blah > 0]$.

Bounding $Var(\|A\|) \leq \mathcal{O}(1)$

- Take random matrix A. We have $\|A\| = \sup_{u,v \in S_2} u^T A v$. Fix u, v that realize this supremum.
- Consider $\|A\| - \|\tilde{A}\|$, where $\tilde{A} = A + \alpha E_{ij}$ and E_{ij} is a matrix with non-zero entry only at the position A_{ij}, and $\alpha \leq 2$.
- Note that $\|A\| - \|\tilde{A}\| \leq u^T A v - u^T (A + \alpha E_{ij}) v \leq 2 u_i v_j$. Term ij, in the sum above is bounded by $4 v_i^2 v_j^2$. The entire sum is bounded by $4 \sum_{i,j} v_i^2 v_j^2 = 4\|u\|^2 \|v\|^2 = 4$.

Concentration of Lipschitz functions of gaussian random variables

Definition (Lipschitz function)

Function $f : \mathbb{R}^n \to \mathbb{R}$ is L-lipschitz if and only if for all x, y, $|f(x) - f(y)| \leq L\|x - y\|$. If f is differentiable, this is equivalent to $\|\nabla f\| \leq L$ everywhere.

Concentration of Lipschitz functions of gaussian random variables

Definition (Lipschitz function)

Function $f : \mathbb{R}^n \to \mathbb{R}$ is L-lipschitz if and only if for all x, y, $|f(x) - f(y)| \leq L\|x - y\|$. If f is differentiable, this is equivalent to $\|\nabla f\| \leq L$ everywhere.

Theorem

If $Z_1, \ldots Z_n \sim \mathcal{N}(0, 1)$, and f is L-Lipschitz
then $f(Z) - \mathbb{E}f(Z)$ is $\mathcal{O}(L)$-subgaussian.

Concentration of Lipschitz functions of gaussian random variables

Definition (Lipschitz function)

Function $f : \mathbb{R}^n \to \mathbb{R}$ is L-lipschitz if and only if for all x, y, $|f(x) - f(y)| \leq L\|x - y\|$. If f is differentiable, this is equivalent to $\|\nabla f\| \leq L$ everywhere.

Theorem

If $Z_1, \ldots Z_n \sim \mathcal{N}(0,1)$, and f is L-Lipschitz
then $f(Z) - \mathbb{E}f(Z)$ is $\mathcal{O}(L)$-subgaussian.

Example

Function $f(A_{11}, \ldots A_{nn}) := \|A\|$ is 1-Lipschitz (easy to check).
Therefore $\|A\| - \mathbb{E}\|A\|$ where entries A_{ij} are independent gaussian is 1-subgaussian.

Magic proof (Terry Tao)

Assume that f is in fact differentiable. By Jensen's inequality

$$\|f(Z) - \mathbb{E}_{Z'} f(Z')\|_p \leq \|f(Z) - f(Z')\|_p,$$

we want to show $\|f(Z) - f(Z')\|_p \lesssim \sqrt{p}L$.

Magic proof (Terry Tao)

Assume that f is in fact differentiable. By Jensen's inequality

$$\|f(Z) - \mathbb{E}_{Z'} f(Z')\|_p \le \|f(Z) - f(Z')\|_p,$$

we want to show $\|f(Z) - f(Z')\|_p \lesssim \sqrt{p}L$.
By fundamental theorem of calculus

$$\begin{aligned} f(Z) - f(Z') &= \int_0^{\pi/2} \frac{\mathrm{d}}{\mathrm{d}\theta} f(Z\cos\theta + Z'\sin\theta)\,\mathrm{d}\theta \\ &= \int_0^{\pi/2} \left\langle \nabla f(Z\cos\theta + Z'\sin\theta), Z\sin\theta - Z'\cos\theta \right\rangle \mathrm{d}\theta \end{aligned}$$

Magic proof (Terry Tao)

Assume that f is in fact differentiable. By Jensen's inequality

$$\|f(Z) - \mathbb{E}_{Z'} f(Z')\|_p \leq \|f(Z) - f(Z')\|_p,$$

we want to show $\|f(Z) - f(Z')\|_p \lesssim \sqrt{p}L$.
By fundamental theorem of calculus

$$\begin{aligned} f(Z) - f(Z') &= \int_0^{\pi/2} \frac{\mathrm{d}}{\mathrm{d}\theta} f(Z\cos\theta + Z'\sin\theta)\,\mathrm{d}\theta \\ &= \int_0^{\pi/2} \left\langle \nabla f(Z\cos\theta + Z'\sin\theta), Z\sin\theta - Z'\cos\theta \right\rangle \mathrm{d}\theta \end{aligned}$$

By convexity of ℓ_p norm

$$\|f(Z) - f(Z')\|_p \leq \int_0^{\pi/2} \| \left\langle \nabla f(Z\cos\theta + Z'\sin\theta), Z\sin\theta - Z'\cos\theta \right\rangle \|_p \,\mathrm{d}$$

Magic proof (Terry Tao)

Assume that f is in fact differentiable. By Jensen's inequality

$$\|f(Z) - \mathbb{E}_{Z'} f(Z')\|_p \leq \|f(Z) - f(Z')\|_p,$$

we want to show $\|f(Z) - f(Z')\|_p \lesssim \sqrt{p}L$.
By fundamental theorem of calculus

$$\begin{aligned} f(Z) - f(Z') &= \int_0^{\pi/2} \frac{\mathrm{d}}{\mathrm{d}\theta} f(Z\cos\theta + Z'\sin\theta)\, \mathrm{d}\theta \\ &= \int_0^{\pi/2} \left\langle \nabla f(Z\cos\theta + Z'\sin\theta), Z\sin\theta - Z'\cos\theta \right\rangle \mathrm{d}\theta \end{aligned}$$

By convexity of ℓ_p norm

$$\|f(Z) - f(Z')\|_p \leq \int_0^{\pi/2} \| \left\langle \nabla f(Z\cos\theta + Z'\sin\theta), Z\sin\theta - Z'\cos\theta \right\rangle \|_p \,\mathrm{d}$$

Note that $Z\cos\theta + Z'\sin\theta$ and $Z\sin\theta - Z'\cos\theta$ are independent gaussian vectors.

End of magic proof.

We want to bound $\| \langle \nabla f(Z\cos\theta + Z'\sin\theta), Z\sin\theta - Z'\cos\theta \rangle \|_p$. This is the same as

$$\| \left\langle \nabla f(Z), Z' \right\rangle \|_p = \left\| \|\nabla f(Z)\|_2 \right\|_p \left\| Z_1 \right\|_p \leq L\sqrt{p}.$$

End of magic proof.

We want to bound $\| \langle \nabla f(Z\cos\theta + Z'\sin\theta), Z\sin\theta - Z'\cos\theta \rangle \|_p$. This is the same as

$$\| \left\langle \nabla f(Z), Z' \right\rangle \|_p = \left\| \|\nabla f(Z)\|_2 \right\|_p \left\| Z_1 \right\|_p \le L\sqrt{p}.$$

After integrating over θ, we get

$$\begin{aligned} \|f(Z) - f(Z')\|_p &\le \int_0^{\pi/2} \| \left\langle \nabla f(Z), Z' \right\rangle \|_p \, \mathrm{d}\theta \\ &\le \frac{\pi}{2} L\sqrt{p}. \end{aligned}$$

What about subgaussian random variables?

Warning!

It is not true that a 1-Lipschitz function of subgaussian random variables is 1-subgaussian!

What about subgaussian random variables?

Warning!

It is not true that a 1-Lipschitz function of subgaussian random variables is 1-subgaussian!

Counterexample

Consider $X_i \in \pm 1$ indepndent coin flips. Take $A = \{a \in \{-1,1\}^n : \sum_i a_i \leq 0\}$, and $f(X) = \inf_{y \in A} \|X - y\|_2$.

What about subgaussian random variables?

Warning!

It is not true that a 1-Lipschitz function of subgaussian random variables is 1-subgaussian!

Counterexample

Consider $X_i \in \pm 1$ indepndent coin flips. Take $A = \{a \in \{-1, 1\}^n : \sum_i a_i \leq 0\}$, and $f(X) = \inf_{y \in A} \|X - y\|_2$. Note that f is 1-Lipschitz.

What about subgaussian random variables?

Warning!

It is not true that a 1-Lipschitz function of subgaussian random variables is 1-subgaussian!

Counterexample

Consider $X_i \in \pm 1$ indepndent coin flips. Take
$A = \{a \in \{-1, 1\}^n : \sum_i a_i \leq 0\}$, and $f(X) = \inf_{y \in A} \|X - y\|_2$.
Note that f is 1-Lipschitz.
We have $f(X)^2 = 2\left(\sum X_i\right)_+$. With probability $1/4$, $\sum X_i > \frac{1}{10\sqrt{n}}$,
hence $f(X) - M(f) > cn^{1/4}$, where $M(f)$ is median of f.

Concentration of convex, Lipschitz function of bounded variables

Definition

Function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for every a, b and every $\lambda \in [0, 1]$, we have $f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda)f(b)$.

Concentration of convex, Lipschitz function of bounded variables

Definition

Function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for every a, b and every $\lambda \in [0,1]$, we have $f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda)f(b)$.

Theorem

If $Z_1, \ldots Z_n \in [-1,1]$ are independent, and f is L-Lipschitz and convex then $f(Z) - \mathbb{E}f(Z)$ is $\mathcal{O}(L)$-subgaussian.

Concentration of convex, Lipschitz function of bounded variables

Definition

Function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for every a, b and every $\lambda \in [0,1]$, we have $f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda) f(b)$.

Theorem

If $Z_1, \ldots Z_n \in [-1,1]$ are independent, and f is L-Lipschitz and convex then $f(Z) - \mathbb{E} f(Z)$ is $\mathcal{O}(L)$-subgaussian.

Example

Norms are famously convex. If matrix A has independent entries A_{ij}, bounded by B with probability 1, then $\|A\| - \mathbb{E}\|A\|$ is B-subgaussian.

Concentration of bounded-difference functions

Definition (bounded-difference function)

We say that f has bounded differences by vector $a_1, \dots a_n$, if for any $X_1, \dots X_n$, and X_i', we have

$$|f(X_1, \dots X_i', \dots X_n) - f(X_1, \dots X_n)| \leq a_i.$$

Concentration of bounded-difference functions

Definition (bounded-difference function)

We say that f has bounded differences by vector $a_1, \dots a_n$, if for any $X_1, \dots X_n$, and X_i', we have

$$|f(X_1, \dots X_i', \dots X_n) - f(X_1, \dots X_n)| \leq a_i.$$

Theorem

If f has bounded differences by vector $a_1, \dots a_n$, and $X_1, \dots X_n$ are independent, than $f(X_1, \dots X_n) - \mathbb{E}f(X_1, \dots X_n)$ is $\|a\|$-subgaussian.

Concentration of bounded-difference functions

Definition (bounded-difference function)

We say that f has bounded differences by vector $a_1, \dots a_n$, if for any $X_1, \dots X_n$, and X_i', we have

$$|f(X_1, \dots X_i', \dots X_n) - f(X_1, \dots X_n)| \leq a_i.$$

Theorem

If f has bounded differences by vector $a_1, \dots a_n$, and $X_1, \dots X_n$ are independent, than $f(X_1, \dots X_n) - \mathbb{E}f(X_1, \dots X_n)$ is $\|a\|$-subgaussian.

Note

Hoeffding inequality is a special case where $f(X) = \sum X_i$ and X_i are bounded by a_i.

Concentration of bounded-diffrence functions
Example

Longest Common Subsequence

Let $x_1, \dots x_n, y_1, \dots y_n$ be a pair of sequences of independent bits $x_i, y_i \in \{0, 1\}$. We wish to understand $\mathrm{LCS}(x, y) := \max\{k : \exists t_1 \leq \dots \leq t_k, r_1 \leq \dots \leq r_k \text{ s.t. } X_{t_i} = Y_{r_i}\}$.

Concentration of bounded-diffrence functions
Example

Longest Common Subsequence

Let $x_1, \dots x_n, y_1, \dots y_n$ be a pair of sequences of independent bits $x_i, y_i \in \{0, 1\}$. We wish to understand $\mathrm{LCS}(x, y) := \max\{k : \exists t_1 \le \dots \le t_k, r_1 \le \dots \le r_k \text{ s.t. } X_{t_i} = Y_{r_i}\}$.

Expectation

$\mathbb{E}\mathrm{LCS}(x, y)$ is difficult to analyze. It is known that $\mathbb{E}\mathrm{LCS} = \gamma n$ for some $\gamma \in (0.75, 0.84)$.

Concentration of bounded-diffrence functions
Example

Longest Common Subsequence

Let $x_1, \ldots x_n, y_1, \ldots y_n$ be a pair of sequences of independent bits $x_i, y_i \in \{0,1\}$. We wish to understand $\mathrm{LCS}(x,y) := \max\{k : \exists t_1 \leq \ldots \leq t_k, r_1 \leq \cdots \leq r_k \text{ s.t. } X_{t_i} = Y_{r_i}\}$.

Expectation

$\mathbb{E}\mathrm{LCS}(x,y)$ is difficult to analyze. It is known that $\mathbb{E}\mathrm{LCS} = \gamma n$ for some $\gamma \in (0.75, 0.84)$.

Concentration

It is immediate that $\mathrm{LCS}(X,Y)$ is $\sqrt{n}$-subgaussian around its expectation!
Changing any single bit can change value of the function by at most one.

Martingales

Definition

Martingale A sequence of random variables $Y_1, \ldots Y_n$ is a martingale with respect to $X_1, \ldots X_n$ if and only if $\mathbb{E}[Y_k | X_1, \ldots X_{k-1}] = Y_{k-1}$.

Martingales

Definition

Martingale A sequence of random variables $Y_1, \dots Y_n$ is a martingale with respect to $X_1, \dots X_n$ if and only if $\mathbb{E}[Y_k | X_1, \dots X_{k-1}] = Y_{k-1}$.

Motivating example: a random walk

Consider a sequence of independent random variables $X_1, \dots X_n$, with $\mathbb{E}X_i = 0$. Then $Y_i = \sum_{j \leq i} X_i$ is a martingale.

Martingales

Definition

Martingale A sequence of random variables $Y_1, \dots Y_n$ is a martingale with respect to $X_1, \dots X_n$ if and only if $\mathbb{E}[Y_k | X_1, \dots X_{k-1}] = Y_{k-1}$.

Motivating example: a random walk

Consider a sequence of independent random variables $X_1, \dots X_n$, with $\mathbb{E}X_i = 0$. Then $Y_i = \sum_{j \leq i} X_i$ is a martingale.

Theorem (Azumas inequality)

If with probability 1 with respect to $X_1, \dots X_{k-1}$, variable $Y_k - Y_{k-1}$ is a_k-subgaussian,
then Y_N is $\sqrt{\sum a_i^2}$-subgaussian.

Martingales

Definition

Martingale A sequence of random variables $Y_1, \dots Y_n$ is a <u>martingale</u> with respect to $X_1, \dots X_n$ if and only if $\mathbb{E}[Y_k | X_1, \dots X_{k-1}] = Y_{k-1}$.

Motivating example: a random walk

Consider a sequence of independent random variables $X_1, \dots X_n$, with $\mathbb{E} X_i = 0$. Then $Y_i = \sum_{j \le i} X_i$ is a martingale.

Theorem (Azumas inequality)

If with probability 1 with respect to $X_1, \dots X_{k-1}$, variable $Y_k - Y_{k-1}$ is a_k-subgaussian,
then Y_N is $\sqrt{\sum a_i^2}$-subgaussian.

Proof?

The same as in the first talk — with careful conditioning.

Azumas inequality implies bounded-difference concentration

- Take $X_1, \ldots X_n$ independent, and $Y_i = \mathbb{E}_{X'_{i+1},\ldots X'_n} f(X_1, \ldots X_i, X'_{i+1}, \ldots X'_n)$.

Azumas inequality implies bounded-difference concentration

- Take $X_1, \ldots X_n$ independent, and $Y_i = \mathbb{E}_{X'_{i+1},\ldots X'_n} f(X_1, \ldots X_i, X'_{i+1}, \ldots X'_n)$.
- Note that Y_i is a martingale with respect to X_i. (Doobs martingale.)

Azumas inequality implies bounded-difference concentration

- Take $X_1, \dots X_n$ independent, and $Y_i = \mathbb{E}_{X'_{i+1}, \dots X'_n} f(X_1, \dots X_i, X'_{i+1}, \dots X'_n)$.
- Note that Y_i is a martingale with respect to X_i. (Doobs martingale.)
- Since f has bounded difference $Y_{i+1} - Y_i$ are bounded by a_i with probability 1. In particular, it is a_i-subgaussian.

Azumas inequality implies bounded-difference concentration

- Take $X_1, \dots X_n$ independent, and $Y_i = \mathbb{E}_{X'_{i+1}, \dots X'_n} f(X_1, \dots X_i, X'_{i+1}, \dots X'_n)$.
- Note that Y_i is a martingale with respect to X_i. (Doobs martingale.)
- Since f has bounded difference $Y_{i+1} - Y_i$ are bounded by a_i with probability 1. In particular, it is a_i-subgaussian.
- Azumas inequality implies that Y_n is $\|a\|$-subgaussian. But by construction $Y_n = f(X_1, \dots X_n)$.