THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 384 11: Theoretical Statistics

Lecture 5: Martingale inequalities

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

A bit background

- So far we have looked at sums of random variables. What if want to study properties of functions of independent random variables?

A bit background

- So far we have looked at sums of random variables. What if want to study properties of functions of independent random variables?
- Consider n independent random variables $X = (X_1, \dots X_n)$.

A bit background

- So far we have looked at sums of random variables. What if want to study properties of functions of independent random variables?
- Consider n independent random variables $X = (X_1, \ldots X_n)$.
- We want to bound $f(X_1, \ldots, X_n) - E[f(X_1, \ldots, X_n)]$

A bit background

- So far we have looked at sums of random variables. What if want to study properties of functions of independent random variables?
- Consider n independent random variables $X = (X_1, \dots X_n)$.
- We want to bound $f(X_1, \dots, X_n) - E[f(X_1, \dots, X_n)]$
- Define $Y_k = E[f(X)|X_1, \dots, X_k]$ for $k \in \{1, \dots, n-1\}$

A bit background

- So far we have looked at sums of random variables. What if want to study properties of functions of independent random variables?
- Consider n independent random variables $X = (X_1, \ldots X_n)$.
- We want to bound $f(X_1, \ldots, X_n) - E[f(X_1, \ldots, X_n)]$
- Define $Y_k = E[f(X) | X_1, \ldots, X_k]$ for $k \in \{1, \ldots, n-1\}$
- $Y_0 = E[f(X)]$ and $Y_n = f(X)$

A bit background

- So far we have looked at sums of random variables. What if want to study properties of functions of independent random variables?
- Consider n independent random variables $X = (X_1, \ldots X_n)$.
- We want to bound $f(X_1, \ldots, X_n) - E[f(X_1, \ldots, X_n)]$
- Define $Y_k = E[f(X)|X_1, \ldots, X_k]$ for $k \in \{1, \ldots, n-1\}$
- $Y_0 = E[f(X)]$ and $Y_n = f(X)$
- Now $f(X) - E[f(X)] = \sum_{i=0}^{n-1} \underbrace{(Y_{i+1} - Y_i)}_{D_i}$

A bit background

- So far we have looked at sums of random variables. What if want to study properties of functions of independent random variables?
- Consider n independent random variables $X = (X_1, \ldots X_n)$.
- We want to bound $f(X_1, \ldots, X_n) - E[f(X_1, \ldots, X_n)]$
- Define $Y_k = E[f(X)|X_1, \ldots, X_k]$ for $k \in \{1, \ldots, n-1\}$
- $Y_0 = E[f(X)]$ and $Y_n = f(X)$
- Now $f(X) - E[f(X)] = \sum_{i=0}^{n-1} \underbrace{(Y_{i+1} - Y_i)}_{D_i}$
- This forms a Martingale difference sequence.

Martingales

Definition

A sequence of random variables $\{Y_i\}$ adapted to a filtration $\mathcal{F}_i$ is a martingale if, for all i,

$$E|Y_i| < \infty \qquad E[Y_{i+1}|\mathcal{F}_i] = Y_i$$

- A filtration $\{\mathcal{F}_i\}$ is a sequence of nested $\sigma-$ fields, i.e. $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$.

Martingales

Definition

A sequence of random variables $\{Y_i\}$ adapted to a filtration $\mathcal{F}_i$ is a martingale if, for all i,

$$E|Y_i| < \infty \qquad E[Y_{i+1}|\mathcal{F}_i] = Y_i$$

- A filtration $\{\mathcal{F}_i\}$ is a sequence of nested $\sigma-$ fields, i.e. $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$.
- Y_i is adapted to $\mathcal{F}_i$ means that each Y_i is measurable w.r.t $\mathcal{F}_i$.

Martingales

Definition

A sequence of random variables $\{Y_i\}$ adapted to a filtration $\mathcal{F}_i$ is a martingale if, for all i,

$$E|Y_i| < \infty \qquad E[Y_{i+1}|\mathcal{F}_i] = Y_i$$

- A filtration $\{\mathcal{F}_i\}$ is a sequence of nested $\sigma-$ fields, i.e. $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$.
- Y_i is adapted to $\mathcal{F}_i$ means that each Y_i is measurable w.r.t $\mathcal{F}_i$.
- If $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$, then we say that $\{Y_i\}$ forms a martingale sequence w.r.t $\{X_i\}$.

Martingales

Definition

A sequence of random variables $\{Y_i\}$ adapted to a filtration $\mathcal{F}_i$ is a martingale if, for all i,

$$E|Y_i| < \infty \qquad E[Y_{i+1}|\mathcal{F}_i] = Y_i$$

- A filtration $\{\mathcal{F}_i\}$ is a sequence of nested $\sigma-$ fields, i.e. $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$.
- Y_i is adapted to $\mathcal{F}_i$ means that each Y_i is measurable w.r.t $\mathcal{F}_i$.
- If $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$, then we say that $\{Y_i\}$ forms a martingale sequence w.r.t $\{X_i\}$.
- If $\mathcal{F}_i = \sigma(Y_1, \ldots, Y_i)$, then we say that $\{Y_i\}$ forms a martingale sequence.

Working through the filtration stuff

Definition

A σ algebra (or field) (Σ) is a collection of subsets of Ω that is closed under complement and countable unions. The pair (Ω, Σ) is called a measurable space. The smallest possible σ algebra on Ω is $\{\phi, \Omega\}$, the biggest is the power set $\mathcal{P}(\Omega)$.

- A random variable $X : \Omega \to S$ is called a measurable map from (Ω, Σ) to $(S, \mathcal{S})$, if
  - Define $X^{-1}(B) = \{\omega : X(\omega) \in B\}$
  - For all $B \in \mathcal{S}$, $X^{-1}(B) \in \mathcal{F}$

Working through the filtration stuff

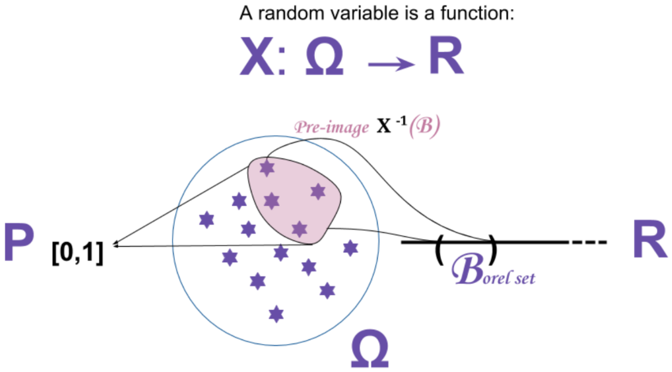

Figure 1: Courtesy: "https://rinterested.github.io/"

Working through the filtration stuff

- I am interested in two coin tosses, where $X_i = 1(\{i^{th}\text{toss is a head}\})$.
- $\Omega = \{(H, H), \ldots, (T, T)\}$
- When $t = 0$, we have not observed anything so $\mathcal{F}_0 = \{\phi, \Omega\}$
- Then $\mathcal{F}_1 = \{\phi, \Omega, \{(H, H), (H, T)\}, \{(T, H), (T, T)\}\}$ (has 4 elements)
- $\mathcal{F}_2 = \mathcal{P}(\Omega)$ has 16 elements.
- $S = \mathbb{R}$ and $\mathcal{S}$ is Borel sets, the smallest $\sigma-$algebra containing the open intervals on $\mathbb{R}$.
- Is X_1 measurable w.r.t $\mathcal{F}_1$?

Working through the filtration stuff

- Is X_1 measurable w.r.t $\mathcal{F}_1$?
  - Yes, because X_1 can take value 1 or 0.
  - For example, $X_1^{-1}((a,b)) = \{(T,H),(T,T)\}$ with $a < 0$ and $b \in (0,1)$ and this is in $\mathcal{F}_1$
  - $X_1^{-1}((a,b)) = \{(H,H),(H,T)\}$ with $a < 1$ and $b > 1$ and this is in $\mathcal{F}_1$
  - $X_1^{-1}((a,b)) = \Omega$ with $a < 0$ and $b > 1$ and this is in $\mathcal{F}_1$
  - But $X_2^{-1}((-0.1,0.1)) = \{(T,T),(H,T)\}$, this is not in $\mathcal{F}_1$, so X_2 is not measurable.

Example-partial sums of i.i.d sequences

Example

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d random variables with $E[X_1] = \mu$. $E[|X_1 - \mu|]$ is bounded. Let $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$. Then $\{Y_i = \sum_{k=1}^{i} X_k - i\mu\}$ is a martingale sequence w.r.t $\{X_i\}$.

Example-partial sums of i.i.d sequences

Example

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d random variables with $E[X_1] = \mu$. $E[|X_1 - \mu|]$ is bounded. Let $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$. Then $\{Y_i = \sum_{k=1}^{i} X_k - i\mu\}$ is a martingale sequence w.r.t $\{X_i\}$.

- Y_i is measurable w.r.t $\mathcal{F}_i$.

Example-partial sums of i.i.d sequences

Example

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d random variables with $E[X_1] = \mu$. $E[|X_1 - \mu|]$ is bounded. Let $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$. Then $\{Y_i = \sum_{k=1}^{i} X_k - i\mu\}$ is a martingale sequence w.r.t $\{X_i\}$.

- Y_i is measurable w.r.t $\mathcal{F}_i$.
- Finally,

$$
\begin{aligned}
E[Y_{i+1}|\mathcal{F}_i] &= E[X_{i+1} + \sum_{k=1}^{i} X_k - (i+1)\mu|\mathcal{F}_i] \\
&= \mu + \sum_{k=1}^{i} X_k - (i+1)\mu = Y_i
\end{aligned}
$$

Doob construction

Example

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of random variables. Let $Y_i = E[f(X)|X_1, \ldots, X_i]$ and assume that $E[|f(X)|] < \infty$. Then $\{Y_i\}_{i=0}^{n}$ is a martingale sequence w.r.t $\{X_i\}_{i=1}^{n}$.

Doob construction

Example

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of random variables. Let $Y_i = E[f(X)|X_1, \ldots, X_i]$ and assume that $E[|f(X)|] < \infty$. Then $\{Y_i\}_{i=0}^{n}$ is a martingale sequence w.r.t $\{X_i\}_{i=1}^{n}$.

- $E[|Y_i|] = E[|E[f(X)|X_1, \ldots, X_i]|] \leq E[|f(X)|] < \infty$. (Use Jensen on $|(.)|$)

Doob construction

Example

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of random variables. Let $Y_i = E[f(X)|X_1, \ldots, X_i]$ and assume that $E[|f(X)|] < \infty$. Then $\{Y_i\}_{i=0}^{n}$ is a martingale sequence w.r.t $\{X_i\}_{i=1}^{n}$.

- $E[|Y_i|] = E[|E[f(X)|X_1, \ldots, X_i]|] \leq E[|f(X)|] < \infty$. (Use Jensen on $|(.)|$)
- Furthermore,

$$\begin{aligned} E[Y_{i+1}|X_1, \ldots, X_i] &= E[E[f(X)|X_1, \ldots, X_{i+1}]|X_1, \ldots, X_i] \\ &= E[f(X)|X_1, \ldots, X_i] = Y_i \qquad \text{The tower property} \end{aligned}$$

Doob construction - examples

Example

We are throwing m balls into n bins. At step i we place ball i into a bin chosen uniformly at random. Call the index of the bin X_i. Let Z denote the number of empty bins. $E[Z|X_1, \ldots X_i]$ is a martingale.

- Whats the big deal, just write $Y_i = 1(\text{Bin } i \text{ is empty})$
- $Z = \sum_i Y_i$, and so I can compute expectation easily.
- Can we use traditional concentration arguments to say $Z - EZ$ is small?

Doob construction - examples

Example

Consider a random graph $G(n, p)$ where the edge between i, j is added with probability p, independent of any other edges. We are interested in the Chromatic number of this graph (χ), i.e. the minimum number of colors to "properly" color this graph, i.e. no two nodes connected by an edge should have the same color.

- Let the vertices be labeled as $1, \ldots, n$
- Let G_i denote the graph induced by nodes $1, \ldots i$.
- Set $E[\chi | G_i]$ is a martingale.
- This is also called a vertex exposure filtration.
- See "Sharp concentration of the chromatic number on random graphs $G_{n,p}$" by Shamir and Spencer

Likelihood ratio

Example

Let f, g be two densities such that g is absolutely continuous w.r.t f. Suppose $\{X_i\}_{i=1}^{\infty} \overset{iid}{\sim} f$ and Y_n is the likelihood ratio $\prod_{i=1}^{n} \frac{g(X_i)}{f(X_i)}$ for the first n datapoints. Then $\{Y_n\}$ forms a Martingale sequence w.r.t $\{X_n\}$.

Likelihood ratio

Example

Let f, g be two densities such that g is absolutely continuous w.r.t f. Suppose $\{X_i\}_{i=1}^{\infty} \overset{iid}{\sim} f$ and Y_n is the likelihood ratio $\prod_{i=1}^{n} \frac{g(X_i)}{f(X_i)}$ for the first n datapoints. Then $\{Y_n\}$ forms a Martingale sequence w.r.t $\{X_n\}$.

- First recall that $E[|Y_n|] = E[Y_n] = 1$

Likelihood ratio

Example

Let f, g be two densities such that g is absolutely continuous w.r.t f. Suppose $\{X_i\}_{i=1}^{\infty} \overset{iid}{\sim} f$ and Y_n is the likelihood ratio $\prod_{i=1}^{n} \frac{g(X_i)}{f(X_i)}$ for the first n datapoints. Then $\{Y_n\}$ forms a Martingale sequence w.r.t $\{X_n\}$.

- First recall that $E[|Y_n|] = E[Y_n] = 1$
- $$\begin{aligned} E[Y_{n+1}|X_1, \ldots, X_n] &= E\left[\prod_{i=1}^{n+1} \frac{g(X_i)}{f(X_i)} \middle| X_1, \ldots, X_n\right] \\ &= \prod_{i=1}^{n} \frac{g(X_i)}{f(X_i)} E\left[\frac{g(X_{n+1})}{f(X_{n+1})}\right] = Y_n \end{aligned}$$

Martingale Difference Sequence

Definition

A sequence $\{D_i\}$ of random variables adapted to a filtration $\{\mathcal{F}_i\}$ is a Martingale Difference Sequence if,

$$E[|D_i|] < \infty \qquad E[D_{i+1}|\mathcal{F}_i] = 0$$

Martingale Difference Sequence

Definition

A sequence $\{D_i\}$ of random variables adapted to a filtration $\{\mathcal{F}_i\}$ is a Martingale Difference Sequence if,

$$E[|D_i|] < \infty \qquad E[D_{i+1}|\mathcal{F}_i] = 0$$

- Let $\{Y_i\}$ be a martingale sequence.

Martingale Difference Sequence

Definition

A sequence $\{D_i\}$ of random variables adapted to a filtration $\{\mathcal{F}_i\}$ is a Martingale Difference Sequence if,

$$E[|D_i|] < \infty \qquad E[D_{i+1}|\mathcal{F}_i] = 0$$

- Let $\{Y_i\}$ be a martingale sequence.
- Then $D_{i+1} = Y_{i+1} - Y_i$ define a Martingale Difference Sequence.
- $E[D_{i+1}|\mathcal{F}_i] = E[Y_{i+1}|\mathcal{F}_i] - E[Y_i|\mathcal{F}_i] = Y_i - Y_i = 0$.

Martingale Difference Sequence

Definition

A sequence $\{D_i\}$ of random variables adapted to a filtration $\{\mathcal{F}_i\}$ is a Martingale Difference Sequence if,

$$E[|D_i|] < \infty \qquad E[D_{i+1}|\mathcal{F}_i] = 0$$

- Let $\{Y_i\}$ be a martingale sequence.
- Then $D_{i+1} = Y_{i+1} - Y_i$ define a Martingale Difference Sequence.
- $E[D_{i+1}|\mathcal{F}_i] = E[Y_{i+1}|\mathcal{F}_i] - E[Y_i|\mathcal{F}_i] = Y_i - Y_i = 0$.
  - $E[Y_{i+1}|\mathcal{F}_i] = Y_i$ because of the martingale property,
  - $E[Y_i|\mathcal{F}_i] = Y_i$ since Y_i is measurable w.r.t the filtration $\mathcal{F}_i$.

Concentration inequalities

Theorem

Consider a Martingale sequence $\{D_i\}$ (adapted to a filtration $\{\mathcal{F}_i\}$) that satisfies $E[e^{\lambda D_i}|\mathcal{F}_{i-1}] \leq e^{\lambda^2 \nu_i^2/2}$ a.s. for any $|\lambda| < 1/b_i$.

- *The sum $\sum_i D_i$ is sub-exponential with parameters $(\sqrt{\sum_k \nu_k^2}, b_*)$ where $b_* := \max_i b_i$.*
- *Hence for all $t \geq 0$,*

$$P\left[|\sum_{i=1}^n D_i| \geq t\right] \leq \begin{cases} 2e^{-\frac{t^2}{2\sum_k \nu_k^2}} & \text{If } 0 \leq t \leq \frac{\sum_k \nu_k^2}{b_*} \\ 2e^{-\frac{t}{2b_*}} & \text{If } t > \frac{\sum_k \nu_k^2}{b_*} \end{cases}$$

Proof

Proof.

Let $X := \sum_{i=1}^{n} D_i$.

$$
\begin{aligned}
E[e^{\lambda \sum_i D_i}] &= E[E[e^{\lambda \sum_i D_i} | \mathcal{F}_{n-1}]] = E[e^{\lambda \sum_{i=1}^{n-1} D_i} E[e^{\lambda D_n} | \mathcal{F}_{n-1}]] \\
&\le E[e^{\lambda \sum_{i=1}^{n-1} D_i}] e^{\lambda^2 \nu_n^2 / 2} \qquad \text{If } |\lambda| < 1/b_n \\
&\le E[e^{\lambda \sum_{i=1}^{n-2} D_i}] e^{\lambda^2 (\nu_{n-1}^2 + \nu_n^2)/2} \qquad \text{If } |\lambda| < 1/b_n, 1/b_{n-1} \\
&\le e^{\sum_i \lambda^2 \nu_i^2 / 2} \qquad \text{If } |\lambda| < \min_i 1/b_i
\end{aligned}
$$

Using our previous theorem on sub-exponential random variables, the result is proven in one direction. The other direction is identical leading to the factor of 2. □

Azuma-Hoeffding

Corollary (Azuma-Hoeffding)

Let $\{D_k\}$ be a Martingale Difference Sequence adapted to the filtration $\{\mathcal{F}_k\}$ and suppose $|D_k| \leq b_k$ a.s. for all $k \geq 1$. Then $\forall t \geq 0$,

$$P\left[|\sum_{k=1}^{n} D_k| \geq t \right] \leq 2e^{-\frac{t^2}{2\sum_k b_k^2}}$$

Proof.

- We can rework the last proof. We need $|E[e^{\lambda D_n}|\mathcal{F}_{n-1}]|$.
- This is bounded by $e^{\lambda^2 b_n^2/2}$, since D_n is mean zero sub-gaussian with $\sigma = b_n$.

□

McDiarmid's inequality

Theorem

Let $f : \mathcal{X}^n \to \mathbb{R}$ satisfy the following bounded difference condition $\forall x_1, \ldots, x_n, x_i' \in \mathcal{X}$:

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

then, $P(|f(X) - E[f(X)]| \geq t) \leq 2\exp\left(-\frac{2t^2}{\sum_i B_i^2}\right)$

- Note that this boils down to Hoeffding's when f is the sum of bounded random variables.

Proof

Proof.

- Define $Y_i = E[f(X)|\mathcal{F}_i]$ and $D_i = Y_i - Y_{i-1}$.

$\square$

Proof

Proof.

- Define $Y_i = E[f(X)|\mathcal{F}_i]$ and $D_i = Y_i - Y_{i-1}$.
- Since $\{Y_i\}$ is a Martingale sequence w.r.t $\{X_i\}$, $\{D_i\}$ is a Martingale difference sequence.

$\square$

Proof

Proof.

- Define $Y_i = E[f(X)|\mathcal{F}_i]$ and $D_i = Y_i - Y_{i-1}$.
- Since $\{Y_i\}$ is a Martingale sequence w.r.t $\{X_i\}$, $\{D_i\}$ is a Martingale difference sequence.
- We have:

$$\begin{aligned} D_i &= E[f(X)|\mathcal{F}_i] - E[f(X)|\mathcal{F}_{i-1}] \\ &= E[f(X)|X_1, \dots, X_i] - E[f(X)|X_1, \dots, X_{i-1}] \\ &\leq \sup_x (E[f(X)|X_1, \dots, x] - E[f(X)|X_1, \dots, X_{i-1}]) =: U_i \\ D_i &\geq \inf_x (E[f(X)|X_1, \dots, x] - E[f(X)|X_1, \dots, X_{i-1}]) =: L_i \end{aligned}$$

$$U_i - L_i \leq B_i$$

□

Proof

Proof.

- We also have:

$$U_i - L_i \leq B_i$$

  - How?

$$\begin{aligned} U_i - L_i &= \sup_x E[f(X)|X_1, \ldots, x] - \inf_y E[f(X)|X_1, \ldots, y] \\ &= \sup_{x,y} \left(E[f(X)|X_1, \ldots, x] - E[f(X)|X_1, \ldots, y]\right) \\ &= \sup_{x,y} \int (f(x_{1:i-1}, x, X_{i+1:n}) - f(x_{1:i-1}, y, X_{i+1:n}))dP(X_{i+1:n}) \\ &\leq \sup_{x,y} \int |f(x_{1:i-1}, x, X_{i+1:n}) - f(x_{1:i-1}, y, X_{i+1:n})|dP(X_{i+1:n}) \\ &\leq B_i \end{aligned}$$

- Now apply Azuma-Hoeffding.
- So, where is independence being used?

Example: Mean absolute deviation

Example

Consider an i.i.d random variable sequence $\{X_k\}_{k=1}^{\infty}$ with $|X_k| \le b$. Define the mean absolute deviation:

$$U = \frac{1}{\binom{n}{2}} \sum_{j<k} |X_j - X_k|$$

As we will see later, the above is a type of a pairwise U-Statistics. We want to bound $|U - E[U]|$.

- Note that the summands are not independent.

Example: Mean absolute deviation

Example

Consider an i.i.d random variable sequence $\{X_k\}_{k=1}^{\infty}$ with $|X_k| \leq b$.
Define the mean absolute deviation:

$$U = \frac{1}{\binom{n}{2}} \sum_{j<k} |X_j - X_k|$$

As we will see later, the above is a type of a pairwise U-Statistics. We want to bound $|U - E[U]|$.

- Note that the summands are not independent.
- Also note that $||X_i - X_j| - |X_i - X_j'|| \leq |X_j - X_j'| \leq 2b$

Example: Mean absolute deviation

Example

Consider an i.i.d random variable sequence $\{X_k\}_{k=1}^{\infty}$ with $|X_k| \leq b$.
Define the mean absolute deviation:

$$U = \frac{1}{\binom{n}{2}} \sum_{j<k} |X_j - X_k|$$

As we will see later, the above is a type of a pairwise U-Statistics. We want to bound $|U - E[U]|$.

- Note that the summands are not independent.
- Also note that $||X_i - X_j| - |X_i - X_j'|| \leq |X_j - X_j'| \leq 2b$
- So $|U(x_1, \ldots, x_i, \ldots, x_n) - U(x_1, \ldots, x_i', \ldots, X_n)| \leq \frac{(n-1)2b}{\binom{n}{2}} = \frac{4b}{n}$

Example: Mean absolute deviation

Example

Consider an i.i.d random variable sequence $\{X_k\}_{k=1}^{\infty}$ with $|X_k| \leq b$. Define the mean absolute deviation:

$$U = \frac{1}{\binom{n}{2}} \sum_{j<k} |X_j - X_k|$$

As we will see later, the above is a type of a pairwise U-Statistics. We want to bound $|U - E[U]|$.

- Note that the summands are not independent.
- Also note that $||X_i - X_j| - |X_i - X_j'|| \leq |X_j - X_j'| \leq 2b$
- So $|U(x_1, \ldots, x_i, \ldots, x_n) - U(x_1, \ldots, x_i', \ldots, X_n)| \leq \frac{(n-1)2b}{\binom{n}{2}} = \frac{4b}{n}$
- Use McDiarmid's inequality, $P(|U - E[U]| \geq t) \leq 2\exp\left(\frac{-nt^2}{8b^2}\right)$

Example: Number of triangles in an Erdos Renyi graph

Example

Consider an Erdős Rényi (ER(p)) random graph. What can we say about the number of triangles Δ?

- Let n be the number of nodes. $m = \binom{n}{2}$ be the number of ordered pairs. Call this set E.
- An $ER(p)$ graph chooses the edges randomly as iid Bernoulli r.v.s $\{X_e : e \in E\}$ with $P(X_e = 1) = p$.
- Let $\mathcal{T} \subset E^3$ be the set of 3-tuples of node pairs which can form a triangle. e.g. $\{(i,j),(j,k),(k,i)\} \in \mathcal{T}$. $|\mathcal{T}| = \binom{n}{3}$.
- We have $f(X) = \sum_{\{e_1,e_2,e_3\} \in \mathcal{T}} X_{e_1} X_{e_2} X_{e_3}$.

Example: Number of triangles in an Erdos Renyi graph–Cont.

Example

Consider an Erdős Rényi (ER(p)) random graph. What can we say about the number of triangles Δ?

- If I switch $X_e = 1$ to 0 how much can $f(X)$ change?
- It changes by all triangles incident on that edge. The maximum number of such triangles is $n-2$. So $L = n-2$.
- Hence $P(|f(X) - E[f(X)]| \geq t) \leq 2e^{-\frac{2t^2}{m(n-2)^2}}$
- $E[f(X)] = \binom{n}{3} p^3$. If we set $t = \Theta(n^2 \log n)$, then the error probability goes to zero.
- But in order for this to give concentration we need, $t/n^3 p^3 \to 0$, i.e. $np >> n^{2/3}$

Example: Number of triangles in an Erdos Renyi graph–Cont.

Example

Consider an Erdős Rényi (ER(p)) random graph. What can we say about the number of triangles Δ?

- One can however use Chen-Stein method to show that $f(X)$ is approximately $\mathit{Poisson}\left(\binom{n}{3}p^3\right)$.
- So the above should hold as long as $np \to \infty$. But McDiarmid requires a much stronger condition!
- What if we could plug in the expected value of the Lipschitz constant, i.e. np^2?
- Then the exponent would be e^{-2t^2/n^4p^4}. Taking $t = n^2p^2$, we see that concentration would amount to having $np >> \log n$ which matches with the Poisson limit argument.

Example: Number of triangles in an Erdos Renyi graph–Cont.

Example

Consider a random graph $G(n, p)$ where the edge between i, j is added with probability p, independent of any other edges. We are interested in the Chromatic number of this graph (χ), i.e. the minimum number of colors to "properly" color this graph, i.e. no two nodes connected by an edge should have the same color.

- We need independent RVs $Z_1, \ldots, Z_i$ so that we can construct a Doob martingale $E[\chi|Z_1, \ldots, Z_i]$ and apply McDiarmid's inequality.
- Let Z_i be the edges from node i to nodes $1, \ldots, i-1$.
- χ cannot decrease by more than 1, because if the graph with node i can be colored by $k-1$ colors, then the graph without node i can be colored using $\leq k-1$ colors.
- Similarly, it can't increase by more than 1, because you can just color node i with a new color, thereby increasing the chromatic number by 1.

Lipschitz functions of Gaussian random variables

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is $L-$Lipschitz w.r.t the Euclidean norm if

$$|f(x) - f(y)| \leq L\|x - y\|_2 \qquad \forall x, y \in \mathbb{R}^n$$

Theorem

Let $(X_1, \ldots, X_n)$ be a vector of iid $N(0,1)$ random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ be $L-$Lipschitz w.r.t the Euclidean norm. Then $f(X) - E[f(X)]$ is sub-gaussian with parameter at most L, i.e. $\forall t \geq 0$,

$$P\left(|f(X) - E[f(X)]| \geq t\right) \leq e^{-\frac{t^2}{2L^2}}$$

- A L Lipschitz function of a vector of i.i.d $N(0,1)$ random variables concentrate like a $N(0, L^2)$ random variable, irrespective of how long the vector is.