Basic Concentration Inequalities : Part II
Martingale Bounds

Previously, we considered concentration of $\frac{1}{n}\sum_{i=1}^{n} X_i$ around its mean :

$$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu > (b-a)\sqrt{\frac{\log(1/\delta)}{2}} \right) \leq \delta, \quad \forall \delta \in (0,1)$$

if $(X_i)_i$ is a sequence of ind't r.v.'s with $a \leq X_i \leq b$, $\forall i$.

Now, we want to understand the behavior of

$$f(X_1, X_2, \dots, X_n)$$

around its mean $\mathbb{E}[f(X_1, \dots, X_n)]$ for some function f where $(X_i)_i$ is a sequence of independent r.v.'s.

The main tool will be martingales.

Martingales

- $(\mathcal{F}_n)_n$ is a "filtration" if these σ-fields are nested : $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

  Example : $\mathcal{F}_n = \underbrace{\sigma(X_1, \dots, X_n)}_{\text{intuitively, information contained in } X_1, \dots, X_n}$ is a filtration.

- Y_n is adapted to $\mathcal{F}_n$ means each Y_n is measurable with respect to $\mathcal{F}_n$. (intuitively, Y_n can be recovered by the info. contained in $\mathcal{F}_n$.)

  Example : $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, $Y_n \triangleq X_1 + \dots + X_n$

  $\Rightarrow Y_n$ is adapted to $\mathcal{F}_n$.

☆ DEF (martingale) A sequence $(Y_n)_n$ of r.v.'s adapted to a filtration $(\mathcal{F}_n)_n$ is a martingale if, for all $n \in \mathbb{N}$,

(i) $\mathbb{E}|Y_n| < \infty$,

(ii) $\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] = Y_n$.

Example : $(X_n)_n$ independent, $\mathbb{E}X_n = 0$, $\mathbb{E}|X_n| < \infty$, $\forall n$.

$$\mathcal{F}_n = \sigma(X_1, \dots, X_n)$$
$$Y_n = X_1 + \dots + X_n.$$

Then, $\mathbb{E}|Y_n| \leq \sum_{i=1}^{n} \mathbb{E}|X_i| < \infty$ ✓

$$\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1} + Y_n \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1}] + Y_n = Y_n$$ ✓

DEF (martingale difference sequence)

A sequence $(D_n)_n$ of r.v.'s adapted to a filtration $(\mathcal{F}_n)_n$ is a martingale difference sequence if, for all n,

$$\text{(i)} \quad \mathbb{E}|D_n| < \infty,$$
$$\text{(ii)} \quad \mathbb{E}[D_{n+1} \mid \mathcal{F}_n] = 0.$$

<u>Example</u>: $(X_n)_n$ independent, $\mathbb{E}X_n = 0$, $\mathbb{E}|X_n| < \infty$

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$$

$$\Rightarrow \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1}] = 0$$

$\Rightarrow (X_n)_n$ is an m.d.s.

<u>Example</u>: More generally, if $(Y_n)_n$ is a martingale adapted to $(\mathcal{F}_n)_n$, then $D_n = Y_n - Y_{n-1}$, $n \in \mathbb{N}_+$ is a martingale difference seq.

To verify the conditions, note that

$$\text{(i)} \quad \mathbb{E}|D_n| \leq \mathbb{E}|Y_n| + \mathbb{E}|Y_{n-1}| < \infty$$
$$\text{(ii)} \quad \mathbb{E}[D_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[Y_{n+1} - Y_n \mid \mathcal{F}_n] = \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] - Y_n = 0.$$

<u>The Doob construction</u>

$$\vec{X} = (X_1, X_2, \ldots, X_n)$$
$$X_{1:i} = (X_1, X_2, \ldots, X_i)$$

Let
$$Y_0 = \mathbb{E} f(\vec{X})$$
$$Y_i = \mathbb{E}[f(\vec{X}) \mid \sigma(X_{1:i})]$$

if $|f| < \infty$, then $(Y_i)_i$ is a martingale adapted to $\sigma(X_{1:i})$

$$\mathbb{E}[Y_{i+1} \mid \sigma(X_{1:i})] = \mathbb{E}[\,\mathbb{E}[f(X) \mid X_1^{i+1}] \mid X_1^i]$$
$$= \mathbb{E}[f(X) \mid X_1^i] = Y_i, \quad \forall i.$$

Fubini's theorem

then, $D_i = Y_i - Y_{i-1}$ is an m.d.s.

$$f(X) - \mathbb{E}[f(X)] = Y_n - Y_0 = \sum_{i=1}^{n} (Y_i - Y_{i-1}) = \sum_{i=1}^{n} D_i.$$

THEOREM | Let $(D_i)_{i\in\mathbb{N}_+}$ be a martingale difference sequence adapted to the filtration $(\mathcal{F}_i)_{i\in\mathbb{N}_+}$, and

$\forall \lambda \in \mathbb{R}$, $\mathbb{E}[e^{\lambda D_i} \mid \mathcal{F}_{i-1}] \leq e^{\lambda^2 \nu_i^2/2}$ almost surely for all i.

Then, the martingale $Y_n = \sum_{i=1}^{n} D_i$ with $Y_0 = 0$ is sub-Gaussian with $\frac{1}{2}\sum_{i=1}^{n} \nu_i^2$.

Pf: $\mathbb{E}[e^{\lambda Y_n}] = \mathbb{E}[e^{\lambda Y_{n-1} + \lambda D_n}]$, $n \geq 1$.

$$\overset{\text{tower property}}{=} \mathbb{E}[\mathbb{E}[e^{\lambda D_n} \cdot e^{\lambda Y_{n-1}} \mid \mathcal{F}_{n-1}]]$$

$$\overset{Y_{n-1} \text{ is } \mathcal{F}_{n-1}\text{-meas.}}{=} \mathbb{E}[e^{\lambda Y_{n-1}} \cdot \mathbb{E}[e^{\lambda D_n} \mid \mathcal{F}_{n-1}]]$$

$$\overset{\text{conditional sub-Gaussianity}}{\leq} e^{\lambda^2 \nu_n^2/2} \cdot \mathbb{E}[e^{\lambda Y_{n-1}}]$$

By induction,

$$\mathbb{E}[e^{\lambda Y_n}] \leq e^{\frac{1}{2}\lambda^2 \sum_{i=1}^{n} \nu_i^2}.$$

Following this, we extend the Hoeffding bound from $\sum_{i=1}^{n} X_i$ to (general) martingales.

COROLLARY | (Azuma-Hoeffding)

Consider a martingale difference seq. $(D_i)_{i\in\mathbb{N}_+}$ adapted to $(\mathcal{F}_i)_{i}$, s.t. $|D_i| \leq B_i$ almost surely. Then,

$$\mathbb{P}(|Y_n| > t) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} B_i^2}\right).$$

Pf | $-B_i \leq D_i \leq B_i$ a.s.

$$\Rightarrow \quad -B_i \leq D_i \mid \mathcal{F}_{i-1} \leq B_i \quad \text{a.s.}$$

Thus, $\mathbb{E}[e^{\lambda D_i} \mid \mathcal{F}_{i-1}] \leq e^{\frac{\lambda^2}{2}\cdot\frac{(2B_i)^2}{4}} = e^{\lambda^2 \frac{B_i^2}{2}}$ a.s. by Hoeffding's lemma.

An important result:

THEOREM (McDiarmid or "Bounded Differences Inequality").

Suppose that $f: \mathbb{R}^n \to \mathbb{R}$ satisfies the following:

$$\forall i \in [n], \quad \forall x_1, x_2, \ldots, x_n, x_i' \in \mathbb{R}, \quad \exists B_i < \infty \quad \text{s.t.}$$

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

Then,

$$\mathbb{P}\left(|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| > t \right) \leq 2 \cdot \exp\left(- \frac{2t^2}{\sum_{i=1}^n B_i^2} \right),$$

for any $t > 0$.

Proof: Let $X := (X_1, \ldots, X_n)$, $\mathcal{F}_i := \sigma(X_1, \ldots, X_i)$.

Use Doob construction:

$$Y_i := \mathbb{E}[f(X) \mid \mathcal{F}_i], \; i \geq 1 \quad \text{with} \quad Y_0 = \mathbb{E}f(X).$$

$$D_i := Y_i - Y_{i-1} \to \text{martingale difference seq. adapted to } \mathcal{F}_i$$

$$\Rightarrow \quad f(X) - \mathbb{E}f(X) = \sum_{i=1}^n D_i$$

Also, let $\tilde{\mathcal{F}}_i := \sigma(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ (leave out X_i)

Then, by Fubini's theorem,

$$\mathbb{E}[f(X) \mid \mathcal{F}_{i-1}] = \mathbb{E}[\,\mathbb{E}[f(X) \mid \tilde{\mathcal{F}}_i] \mid \mathcal{F}_i].$$

$$\Rightarrow \quad D_i = \mathbb{E}[\, f(X) - \mathbb{E}[f(X) \mid \tilde{\mathcal{F}}_i] \mid \mathcal{F}_i]$$

Hence, $|D_i| \leq \mathbb{E}[\,| f(X) - \mathbb{E}[f(X) \mid \tilde{\mathcal{F}}_i]| \mid \mathcal{F}_i]$

$\leq B_i$ almost surely (a.s.)

Using Azuma-Hoeffding inequality,

$$\mathbb{P}\left(\left| \sum_{i=1}^n D_i \right| > t \right) \leq 2\exp\left(- \frac{2t^2}{\sum_{i=1}^n B_i^2} \right). \quad ∎$$