Concentration Inequalities

Wei-Ning Chen

April 11, 2018

Outline

Bounds on Independent Random Variables

In this section, we will introduce *sub-Gaussian* and *sub-exponential* random variables.

- sub-Gaussian property implies *Hoeffding* type inequalities
- sub-exponential property implies *Bernstein* type inequalities
- All these properties hold for bounded random variables

Markov's Inequality

Theorem (Markov's Inequality)

Let X be non-negative random variable. Then for all $t > 0$

$$\mathbb{P}\{X \geq t\} \leq \frac{\mathbb{E}X}{t}.$$

EX. Plugging-in $\tilde{X} = (X - \mathbb{E}X)^2$, we obtain *Chebyshev's inequality*.

Chernoff's Bound

Theorem (Chernoff's Bound)

Let $\mathbb{E}X = \mu$. *For any* $\lambda \geq 0$,

$$\mathbb{P}\{(X-\mu) \geq t\} \leq \frac{\mathbb{E}\left[e^{\lambda(X-\mu)}\right]}{e^{\lambda t}}.$$

Equivalently,

$$\log \mathbb{P}\{(X-\mu) \geq t\} \leq -\sup_{\lambda \geq 0}\left\{\lambda t - \log \mathbb{E}\left[e^{\lambda(X-\mu)}\right]\right\}.$$

Proof.
Apply Markov's inequality on $Y = e^{\lambda(X-\mu)}$ (given that $\mathbb{E}\left[e^{\lambda(X-\mu)}\right]$ exists) . □

Chernoff's Bound

Example (Gaussian Tail Bounds)

Let $X \sim N(\mu, \sigma)$ be a Gaussian random variable with mean μ and variance σ^2. By a straightforward calculation, we find that X has the MGF

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] = e^{\sigma^2\lambda^2/2}, \forall \lambda \in \mathbb{R}.$$

Therefore, plugging into Chernoff's Bound, we obtain

$$\sup_{\lambda \geq 0}\left\{\lambda t - \log \mathbb{E}\left[e^{\lambda(X-\mu)}\right]\right\} = \sup_{\lambda \geq 0}\left\{\lambda t - \frac{\lambda^2\sigma^2}{2}\right\} = \frac{t^2}{2\sigma^2}.$$

We conclude that

$$\mathbb{P}\left\{(X-\mu) \geq t\right\} \leq e^{-\frac{t^2}{2\sigma^2}}, \forall t \geq 0.$$

Chernoff's Bound

- For general random variable X, we want a similar tail bound:

$$\mathbb{P}\left\{X-\mu \geq t\right\} \leq e^{-ct^2}$$

- It suffices to upper bound $\mathbb{E}\left[e^{\lambda(X-\mu)}\right]$

$$\left(\text{Recall: } \log \mathbb{P}\left\{(X-\mu) \geq t\right\} \leq -\sup_{\lambda \geq 0}\left\{\lambda t - \log \mathbb{E}\left[e^{\lambda(X-\mu)}\right]\right\}.\right)$$

Sub-Gaussian R.V.s

Definition

A random variable X with mean $\mu = \mathbb{E}[X]$ is sub-Gaussian if there is a positive number σ such that

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\sigma^2\lambda^2/2}, \text{ for all } \lambda.$$

EX. $X \sim N(\mu, \sigma)$ is sub-Gaussian σ^2.

Sub-Gaussian R.V.s

Example (Bounded Random Variables)

Let X be zero-mean, and bounded on some interval $[a, b]$.

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \mathbb{E}\left[\frac{1 + X/(b-a)}{2} e^{\lambda(b-a)} + \frac{1 - X/(b-a)}{2} e^{-\lambda(b-a)}\right] = \frac{1}{2} e^{\lambda(b-a)} + \frac{1}{2} e^{-\lambda(b-a)}.$$

By Taylor expansion, we can show that

$$\frac{1}{2} e^{\lambda(b-a)} + \frac{1}{2} e^{-\lambda(b-a)} \leq e^{\frac{\lambda^2 (b-a)^2}{2}},$$

which implies X is sub-Gaussian with parameter $\sigma = (b-a)$.

Example (Bounded Random Variables (cont'd))

Note that with more carefully analysis, for example, Hoeffding's lemma (Lemma 12 in Lecture 02), we have

$$\mathbb{E}[e^{\lambda X}] \leq \exp\left((b-a)^2\lambda^2/8\right).$$

Hence X is actually sub-Gaussian with parameter $\sigma = \dfrac{(b-a)}{2}$.

Theorem (Sub-Gaussian Tail Bound)

Let X be a s sub-Gaussian with parameter σ. Then

$$\mathbb{P}\left\{|X-\mu| \geq t\right\} \leq 2e^{-\frac{t^2}{2\sigma^2}},\ \forall t > 0.$$

Proof.

$$\begin{aligned}
\log \mathbb{P}\left\{(X-\mu) \geq t\right\} &\leq -\sup_{\lambda\geq 0}\left\{\lambda t - \log \mathbb{E}\left[e^{\lambda(X-\mu)}\right]\right\} \\
&\leq -\sup_{\lambda\geq 0}\left\{\lambda t - \sigma^2\lambda^2/2\right\} = -t^2/2\sigma^2.
\end{aligned}$$

On the other hand, $-(X-\mu)$ also sub-Gaussian, so $\log \mathbb{P}\left\{-(X-\mu) \geq t\right\} \leq -t^2/2\sigma^2$.

Sub-Gaussian R.V.s

Proposition (Hoeffding Bound)

Suppose that the variables $X_i, i = 1, ..., n$ are independent, and X_i has mean μ_i and sub-Gaussian parameter σ_i. Then for all $t \geq 0$, we have

$$\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mu_i) \geq t\right] \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{n}\sigma_i^2}\right).$$

Equivalently,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) \geq \epsilon\right] \leq \exp\left(-\frac{n^2\epsilon^2}{2\sum_{i=1}^{n}\sigma_i^2}\right).$$

Proof sketch: it suffices to bound the MGF.

$$\mathbb{E}[e^{\lambda\sum(X_i-\mu_i)}] = \prod_i \mathbb{E}[e^{\lambda(X_i-\mu_i)}] \leq e^{\frac{\lambda^2\sum_i\sigma_i^2}{2}}.$$

Sub-Exponential Variables

Definition

A random variable X with mean $\mu = \mathbb{E}[X]$ is sub-Exponential if there are non-negative parameters (ν, b) such that

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\nu^2\lambda^2}{2}}, \text{ for all } |\lambda| < \frac{1}{b}.$$

Remark : It follows immediately that any sub-Gaussian variable is also sub-exponential.

Sub-Exponential Variables

Example

Let $Z \sim N(0,1)$ *and* $X = Z^2$*. Then*

$$\mathbb{E}[e^{\lambda(X-1)}] = \frac{1}{\sqrt{2\pi}} \int e^{\lambda(z^2-1)} e^{-z^2/2} dz = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}}.$$

Following some calculus, it can be verified that

$$\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq e^{4\lambda^2/2}, \text{ for } |\lambda| \leq \frac{1}{4},$$

and hence X *is sub-exponential with parameters with parameter* $(\nu, b) = (2, 4)$*.*

Sub-Exponential Tail Bound

Theorem (Sub-Exponential Tail Bound)

Suppose that X is sub-exponential with parameters (ν, b). Then

$$\mathbb{P}\{X - \mu \geq t\} \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}}, & \text{if } 0 \leq t \leq \frac{\nu^2}{b} \\ e^{-\frac{t}{2b}}, & \text{for } t > \frac{\nu^2}{b} \end{cases}$$

Moreover, the following bound also holds:

$$\mathbb{P}\{X - \mu \geq t\} \leq e^{-\frac{t^2}{2(bt+\nu^2)}}, \text{ for all } t.$$

Remark: We can apply the same trick to obtain two-sided bound.

Sub-Exponential Tail Bound

Proof.
From Chernoff's bound, we have

$$\log \mathbb{P}\left\{(X-\mu) \geq t\right\} \leq -\sup_{\lambda \geq 0}\left\{\lambda t - \log \mathbb{E}\left[e^{\lambda(X-\mu)}\right]\right\}$$
$$\leq -\sup_{\lambda < 1/b}\left\{\lambda t - \lambda^2\nu^2/2\right\} \qquad (\star)$$

1. If $0 \leq t \leq \frac{\nu^2}{b}$, choose $\lambda = \frac{t}{\nu^2}(\leq \frac{1}{b})$, then $(\star) \leq -\frac{t^2}{2\nu^2}$.
2. If $t \geq \frac{\nu^2}{b}$, choose $\lambda = \frac{1}{b}$, and $(\star) \leq -\frac{1}{b} + \frac{1}{2b}\frac{\nu^2}{2b} \leq -\frac{t}{2b}$ $(\because \frac{\nu^2}{2b} \leq t)$.
3. Specially, choose $\lambda = \frac{t}{bt+\nu^2}$, $(\star) \leq -\frac{t^2}{2(bt+\nu^2)}$ □

Bernstein's Condition

Definition (Bernstein's Condition)

Given a random variable X with mean $\mu = \mathbb{E}[X]$, variance $\sigma = \mathbb{E}[X^2 - \mu^2]$, we say that *Bernstein's condition* with parameter b holds if

$$\left|\mathbb{E}[(X-\mu)^k]\right| \leq \frac{1}{2}k!\sigma^2 b^{k-2}, \text{ for all } k = 3, 4, ...$$

Theorem (Bernstein's Inequality)

For any random variable X satisfying the Bernstein condition, we have

- $\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\left(\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}\right)}$
- *X is sub-exponential with parameters* $(\sqrt{2}\sigma, 2b)$
- $\mathbb{P}\{|X-\mu| \geq t\} \leq 2e^{-\frac{t^2}{2(\sigma^2+bt)}}$

Proof.
(1) If X satisfies Bernstein's condition, then

$$\mathbb{E}[e^{\lambda(X-\mu)}] = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \lambda^k \frac{\mathbb{E}(X-\mu)^k}{k!} \overset{(i)}{\leq} 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2} \sum_{k=3}^{\infty} (|\lambda| b)^{k-2}$$

$$\overset{(ii)}{\leq} 1 + \frac{\lambda^2\sigma^2/2}{1-b|\lambda|} \leq e^{\left(\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}\right)}.$$

(2) Hence, when $|\lambda| \leq \frac{1}{2b}$, $\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{-\frac{\lambda^2(\sqrt{2}\sigma)^2}{2}}$.

(3) Choosing $\lambda = \frac{t}{bt+\sigma^2} \in [0, \frac{1}{b})$ and apply by (1) with Chernoff's bound.

Bernstein's Condition

Proposition

Let $X_k, k = 1, ..., n$ are independent sub-exponential random variables, with parameter (ν_k, b_k). Then $\sum_k (X_k - \mu_k)$ is sub-exponential with the parameter $(\nu_\star, b_\star)$, with

$$(\nu_\star, b_\star) \triangleq \left(\sqrt{\sum_k \nu_k^2 / n}, \max_k b_k\right).$$

Moreover, we have

$$\mathbb{P}\left\{\frac{1}{n}\sum_k (X_k - \mu_k) \geq t\right\} \leq \begin{cases} e^{-\frac{nt^2}{2\nu_\star^2}}, & \text{for } 0 \leq t \leq \frac{\nu_\star}{b_\star} \\ e^{-\frac{nt}{2b_\star^2}}, & \text{for } t \geq \frac{\nu_\star}{b_\star} \end{cases}$$

Summary

(1) sub-Gaussian random variable with parameter σ_i:

- Tail Bound : $\mathbb{P}\{|X-\mu| \geq t\} \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$, for all $t > 0$.
- Independent Sum : $\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_i) \geq \epsilon\right] \leq \exp\left(-\frac{n^2\epsilon^2}{2\sum_{i=1}^{n}\sigma_i^2}\right)$.

(2) sub-exponential random variable with parameter (ν_i, b_i):

- Tail Bound : $\mathbb{P}\{|X-\mu| \geq t\} \leq 2\exp\left(-\frac{t^2}{2(bt+\nu^2)}\right)$, for all $t > 0$.
- Independent Sum: $\mathbb{P}\left\{\frac{1}{n}\sum_{k}(X_k-\mu_k) \geq \epsilon\right\} \leq \exp\left(-\frac{n^2\epsilon^2}{2(nb_\star\epsilon+\nu_\star^2)}\right)$
- Bernstein's condition : relates (ν, b) with *variance* and *support*

- Up to now, we see various types of bounds on sums of independent random variables
- Many problems require bounds on more general functions of random variables
- One classical approach is based on *martingale decompositions*

Recap: Conditional Expectation

- The conditional expectation $\mathbb{E}[g(X_1, ..., X_n)|X_1, ..., X_k]$ is a function of $(X_1, ..., X_k)$

$$\mathbb{E}[g(X_1, ..., X_n)|X_1, ..., X_k] = \int g(X_1, ..., X_k, x_{k+1}, ...x_n) dP(x_{k+1}, ..., x_n|X_1, ..., X_k).$$

- The conditional expectation $\mathbb{E}[g(X_1, ..., X_n)|X_1, ..., X_k]$ is an orthogonal projection of $g(\cdot)$ to the functional space spanned by $(X_1, ..., X_k)$.

- We use the notation $\mathcal{F}_k$ to denote the space spanned by $(X_1, ..., X_k)$, i.e.

$$\mathbb{E}[g(X_1, ..., X_n)|\mathcal{F}_k] \triangleq \mathbb{E}[g(X_1, ..., X_n)|X_1, ..., X_k].$$

(Rigorously speaking , $\mathcal{F}_k \triangleq \sigma(X_1, ..., X_k)$)

Recap: Conditional Expectation

Properties of Conditional Expectation

- Pulling out known factors :

$$\mathbb{E}[g(X_1, ..., X_k)f(X_1, ..., X_n)|\mathcal{F}_k] = g(X_1, ..., X_k)\mathbb{E}[f(X_1, ..., X_n)|\mathcal{F}_k].$$

- Law of total expectation :

$$\mathbb{E}\left(\mathbb{E}(g(X_1, ..., X_n)|\mathcal{F}_k)\right) = \mathbb{E}[g(X_1, ..., X_n)]$$

- Tower property : for any $k_1 \leq k_2$, we have

$$\mathbb{E}\left(\mathbb{E}\left(g(X_1, ..., X_n)|\mathcal{F}_{k_2}\right)|\mathcal{F}_{k_1}\right) = \mathbb{E}(g(X_1, ..., X_n)|\mathcal{F}_{k_1})$$

Remark: we have $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq ... \subseteq \mathcal{F}_n$

Definition (Martingale)

Given a sequence of random variables $\{Y_k\}_{k=1}^{\infty}$, we say $\{Y_k\}$ is a martingale with respect to $\{X_k\}$, if

$$\mathbb{E}[|Y_k|] < \infty, \text{ and } \mathbb{E}[Y_k|\mathcal{F}_{k-1}] = Y_{k-1}$$

(That is, $\mathbb{E}[Y_k|X_1, ..., X_{k-1}] = Y_{k-1}$)

Example (Random Walk)

Let S_n be the one-dimensional random walk. That is,

$$S_n = \sum_{i=1}^{n} W_i,$$

where W_i takes values in $\{+1, -1\}$ with probability $1/2$ independently.
Then we have

$$\mathbb{E}[S_k|W_1, ..., W_{k-1}] = \sum_{i=1}^{k-1} W_i + \mathbb{E}[W_k|W_1, ..., W_{k-1}] = S_{k-1}.$$

Example (Dependent Increment Martingale)

Let $a > 0$*, and* W_i *takes values in* $\{+1, -1\}$ *with probability* $\frac{1}{a+1}, \frac{a}{a+1}$ *independently, and let* $S_k = \sum_{i=1}^{k} W_i$*. Then the process* $X_k = a^{S_k}$ *is a martingale:*

Check:

$$\begin{aligned}\mathbb{E}[X_k|\mathcal{F}_{k-1}] &= \mathbb{E}[a^{S_k}|\mathcal{F}_{k-1}] = a^{S_{k-1}}\mathbb{E}[a^{W_k}|\mathcal{F}_{k-1}] \\ &= a^{S_{k-1}}\left(ap + \frac{1}{a}(1-p)\right) = X_{k-1}.\end{aligned}$$

Doob's Construction

In general, we can construct a martingale by conditioning:

Definition (Doob's Martingale)

Let $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$ be a measurable function. Then

$$Y_k \triangleq \mathbb{E}[f(X_1, ..., X_n)|\mathcal{F}_k] \text{ is a martingale.}$$

Note that we have

$$\begin{aligned}\mathbb{E}[Y_k|\mathcal{F}_{k-1}] &= \mathbb{E}\left(\mathbb{E}\left(f(X_1, ..., X_n)|\mathcal{F}_k\right)|\mathcal{F}_{k-1}\right) \\ &= \mathbb{E}\left(f(X_1, ..., X_n)|\mathcal{F}_{k-1}\right) = Y_{k-1}.\end{aligned}$$

Doob's Martingale

According to Doob's martingale, $f(X_1, ..., X_n) - \mathbb{E}[f(X_1, ..., X_n)]$ can be decomposed into sum of martingale difference:

$$f(X_1, ..., X_n) - \mathbb{E}[f(X_1, ..., X_n)] = \sum_{k=1}^{n} (Y_k - Y_{k-1}),$$

with $Y_k \triangleq \mathbb{E}[f(X_1, ..., X_n) | \mathcal{F}_k]$. Note that we have

$$Y_n = \mathbb{E}[f(X_1, ..., X_n) | \mathcal{F}_n] = f(X_1, ..., X_n), \text{ and } Y_0 = \mathbb{E}[f(X_1, ..., X_n)].$$

This motivates us to study the concentration inequalities on sum of *martingale difference sequence*.

Martingale Difference Sequence

Definition (Martingale Difference)

A sequence of random variable $\{D_k\}$ is a martingale difference with respect to $\{\mathcal{F}_k\}$, if

$$\mathbb{E}[|D_k|] < \infty \text{ and } \mathbb{E}[D_k|\mathcal{F}_{k-1}] = 0.$$

Remark: If $Y_k = \sum_{i=1}^{k} D_k + Y_0$ ($\Leftrightarrow D_k = Y_k - Y_{k-1}$), then Y_k is a martingale. Also,

$$Y_n - Y_0 = \sum_{k=1}^{n} D_k.$$

Azuma-Hoeffding Inequality

Theorem (Azuma-Hoeffding)

Let $\{(D_k, \mathcal{F}_k)\}$ be a martingale difference sequence, and suppose that $|D_k| \leq b_k$ almost surely for all $k \geq 1$. Then for all $t \geq 0$,

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

Proof.
By Chernoff's bound, we have

$$\log \mathbb{P}\left\{\sum_{k=1}^{n} D_k \geq t\right\} \leq -\sup_{\lambda \geq 0}\left\{\lambda t - \log \mathbb{E}\left[e^{\lambda(\sum_{k=1}^{n} D_k)}\right]\right\}.$$

It suffices to show that $\log \mathbb{E}\left[e^{\lambda(\sum_{k=1}^{n} D_k)}\right] \leq \dfrac{\lambda^2 \sum_k b_k^2}{8}$.

Proof. (cont'd)
Notice that from the property of martingale, we have

$$\mathbb{E}\left[e^{\lambda(\sum_{k=1}^{n} D_k)}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\lambda(\sum_{k=1}^{n-1} D_k)} e^{\lambda D_n} | \mathcal{F}_{n-1}\right]\right]$$
$$= \mathbb{E}\left[e^{\lambda(\sum_{k=1}^{n-1} D_k)} \mathbb{E}\left[e^{\lambda D_n} | \mathcal{F}_{n-1}\right]\right] \qquad (\star)$$

Our goal is to show that $\mathbb{E}\left[e^{\lambda D_n} | \mathcal{F}_{n-1}\right] \leq e^{\lambda^2 b_k^2/8}$ almost surely.
By the convexity, we have

$$e^{\lambda D_n} \leq \frac{1 + D_n/b_n}{2} e^{\lambda b_n} + \frac{1 - D_n/b_n}{2} e^{-\lambda b_n},$$

and hence

$$\mathbb{E}\left[e^{\lambda D_n} | \mathcal{F}_n\right] \leq \mathbb{E}\left[\frac{1 + D_n/b_n}{2} e^{\lambda b_n} + \frac{1 - D_n/b_n}{2} e^{-\lambda b_n} | \mathcal{F}_n\right] = \frac{1}{2} e^{\lambda b_n} + \frac{1}{2} e^{-\lambda b_n}.$$

Proof. (cont'd)

Notice that by Taylor expansion, we can show that

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

(the constant can be improved by carefully apply Jensen's inequality).
Iteratively,

$$(\star) \leq \mathbb{E}\left[e^{\lambda(\sum_{k=1}^{n-1} D_k)}\right] e^{\lambda^2 b_n^2/8} \leq \mathbb{E}\left[e^{\lambda(\sum_{k=1}^{n-2} D_k)}\right] e^{\lambda^2 b_n^2/8} e^{\lambda^2 b_{n-1}^2/8} \leq \ldots \leq e^{\frac{\lambda^2 \sum_k b_k^2}{8}},$$

with probability 1.
Therefore, from the Chernoff's bound

$$\begin{aligned}
\log \mathbb{P}\left\{\sum_{k=1}^{n} D_k \geq t\right\} &\leq -\sup_{\lambda \geq 0}\left\{\lambda t - \log \mathbb{E}\left[e^{\lambda(\sum_{k=1}^{n} D_k)}\right]\right\} \\
&\leq -\sup_{\lambda \geq 0}\left\{\lambda t - \frac{\lambda^2 \sum_k b_k^2}{8}\right\}.
\end{aligned}$$

Choosing the optimal λ, the proof is complete. □

Bounded Differences Inequality

We say that $f : \mathbb{R}^n \to \mathbb{R}$ satisfies the bounded difference property with parameters $(L_1, ..., L_n)$, if for each $k = 1, 2, ..., n$,

$$|f(x_1, ..., x_{k-1}, x_k, x_{k+1}, ..., x_n) - f(x_1, ..., x_{k-1}, x_k', x_{k+1}, ..., x_n)| \leq L_k, \ \forall x_k, x_k'.$$

Theorem (Bounded differences inequality)

Suppose that f satisfies the bounded difference property and that the random vector $(X_1, X_2, ..., X_n)$ has independent components. Then

$$\mathbb{P}\left\{|f(X_1, ..., X_n) - \mathbb{E}[f(X_1, ..., X_n)]| \geq t\right\} \leq 2e^{-\frac{2t^2}{\sum_{k=1}^n L_k^2}}, \ \textit{for all } t \geq 0.$$

Remark : In Doob's martingale, $X_1, ..., X_n$ don't have to be independent!

Bounded Differences Inequality

Proof.
Recalling the Doob's martingale, and according to Azuma's inequality, it suffices to show that the difference is bounded almost surely:

$$D_k = \mathbb{E}[f(X_1, ..., X_n)|X_1, ..., X_k] - \mathbb{E}[f(X_1, ..., X_n)|X_1, ..., X_{k-1}].$$

Define the function $g : \mathbb{R}^k \to \mathbb{R}$ as $g(x_1, ..., x_k) \triangleq \mathbb{E}[f(X_1, ..., X_n)|x_1, ..., x_k]$, then we have

$$D_k = g(X_1, ..., X_k) - \mathbb{E}_{X_k'}[g(X_1, ..., X_{k-1}, X_k')].$$

check:

$$\begin{aligned}
\mathbb{E}[f(X_1, ..., X_n)|X_1, ..., X_{k-1}] &= \mathbb{E}[g(X_1, ..., X_k)|X_1, ..., X_{k-1}] \\
&= \int g(X_1, ..., X_{k-1}, x_k) dP(x_k|X_1, ..., X_k) \\
(\because \text{ independence assumption !}) \quad &= \int g(X_1, ..., X_{k-1}, x_k) dP(x_k) = \mathbb{E}_{X_k'}[g(X_1, ..., X_{k-1}, X_k')]
\end{aligned}$$

Bounded Differences Inequality

Proof (cont'd)

Therefore,

$$D_k = g(X_1, ..., X_k) - \mathbb{E}_{X_k'}[g(X_1, ..., X_{k-1}, X_k')] = \mathbb{E}_{X_k'}[g(X_1, ..., X_k) - g(X_1, ..., X_k')].$$

Notice that we have

$$\begin{aligned} |g(x_1, ..., x_k) - g(x_1, ..., x_k)| &= |\mathbb{E}_{X_{k+1}^n}[f(x_1, ..., x_k, X_{k+1}, ..., X_n) - f(x_1, ..., x_k', X_{k+1}, ..., X_n)]| \\ &\leq L_k, \end{aligned}$$

showing $|D_k| \leq L_k$ almost surely, and hence establish the theorem. □