Topic Course on Probabilistic Methods
(Week 7)
Large deviation inequalities (I)

Linyuan Lu

University of South Carolina

Univeristy of South Carolina, Fall, 2012
Selected topics

- Linearity of Expectation (2 weeks)
- Alterations (1 week)
- The second moment method (1 week)
- The Local Lemma (1-2 weeks)
- Correlation Inequalities (1 week)
- Large deviation inequalities (3 weeks)
- Poisson Paradigm (1 week)
- Random graphs (2 weeks)
- Discrepancy (1 week)
- Entropy (1 week)
Subtopics

Large deviation inequality

- Chernoff inequalities
- Weighted version
- McDiarmid’s theorem
- Another generalization
- Lower tail versus upper tail
- More general versions
Chernoff inequalities: Suppose $X = \sum_{i=1}^{n} X_i$, where $X_i$ are independent 0-1 random variables with

$$\Pr(X_i = 1) = p_i, \quad \Pr(X_i = 0) = 1 - p_i.$$ 

Then we have

$$\Pr(X < E(X) - \lambda) \leq e^{-\frac{\lambda^2}{2E(X)}}$$

$$\Pr(X > E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(E(X)+\lambda/3)}}$$
A weighted version of Chernoff’s inequality:

\[ X = \sum_{i=1}^{n} a_i X_i \]
A weighted version of Chernoff’s inequality:

- \( X = \sum_{i=1}^{n} a_i X_i \)
- \( 0 \leq a_1, \ldots, a_n \leq M \)
A weighted version of Chernoff’s inequality:

- $X = \sum_{i=1}^{n} a_i X_i$
- $0 \leq a_1, \ldots, a_n \leq M$
- $X_1, \ldots, X_n$: independent, 0-1, with $Pr(X_i = 1) = p_i$
A weighted version of Chernoff’s inequality:

- \( X = \sum_{i=1}^{n} a_i X_i \)
- \( 0 \leq a_1, \ldots, a_n \leq M \)
- \( X_1, \ldots, X_n: \) independent, 0-1, with \( Pr(X_i = 1) = p_i \)
- \( E(X) = \sum_{i=1}^{n} a_i p_i \)
A weighted version of Chernoff’s inequality:

- \( X = \sum_{i=1}^{n} a_i X_i \)
- \( 0 \leq a_1, \ldots, a_n \leq M \)
- \( X_1, \ldots, X_n \): independent, 0-1, with \( Pr(X_i = 1) = p_i \)
- \( E(X) = \sum_{i=1}^{n} a_i p_i \)
- \( \nu = \sum_{i=1}^{n} a_i^2 p_i \)
A weighted version of Chernoff’s inequality:

- \( X = \sum_{i=1}^{n} a_i X_i \)
- \( 0 \leq a_1, \ldots, a_n \leq M \)
- \( X_1, \ldots, X_n \): independent, 0-1, with \( Pr(X_i = 1) = p_i \)
- \( E(X) = \sum_{i=1}^{n} a_i p_i \)
- \( \nu = \sum_{i=1}^{n} a_i^2 p_i \)

**Theorem [Chung,Lu]** We have

\[
Pr(X < E(X) - \lambda) \leq e^{-\lambda^2/2\nu} \quad (1)
\]
\[
Pr(X > E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\nu + M\lambda/3)}}. \quad (2)
\]
Theorem [McDiarmid]: Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i - E(X_i) \leq M$ for a positive constant $M$. Let $X = \sum_{i=1}^{n} X_i$. Then

$$Pr(X - E(X) > \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + M\lambda/3)}}.$$
Theorem [McDiarmid]: Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i - E(X_i) \leq M$ for a positive constant $M$. Let $X = \sum_{i=1}^{n} X_i$. Then

$$Pr(X - E(X) > \lambda) \leq e^{-\frac{\lambda^2}{2(Var(X) + M\lambda/3)}}.$$ 

Note: If $Pr(X_i = a_i) = p_i$ and $Pr(X_i = 0) = 1 - p_i$, then $Var(X) = a_i^2 p_i (1 - p_i) \leq \nu$. Thus

$$Pr(X - E(X) > \lambda) \leq e^{-\frac{\lambda^2}{2(\nu + M\lambda/3)}}.$$
Theorem [McDiarmid]: Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i - E(X_i) \leq M$ for a positive constant $M$. Let $X = \sum_{i=1}^{n} X_i$. Then

$$Pr(X - E(X) > \lambda) \leq e^{-\frac{\lambda^2}{2(Var(X) + M\lambda/3)}}.$$  

Note: If $Pr(X_i = a_i) = p_i$ and $Pr(X_i = 0) = 1 - p_i$, then $Var(X) = a_i^2 p_i (1 - p_i) \leq \nu$. Thus

$$Pr(X - E(X) > \lambda) \leq e^{-\frac{\lambda^2}{2(\nu + M\lambda/3)}}.$$  

This theorem implies inequality of upper tail in previous Theorem.
Theorem [Chung, Lu] Suppose $X_i$ are independent random variables satisfying $X_i \leq M$, for $1 \leq i \leq n$. Let $X = \sum_{i=1}^{n} X_i$ and $\|X\| = \sqrt{\sum_{i=1}^{n} \mathbb{E}(X_i^2)}$. Then we have

$$\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\|X\|^2 + M\lambda/3)}}.$$
Theorem [Chung, Lu] Suppose $X_i$ are independent random variables satisfying $X_i \leq M$, for $1 \leq i \leq n$. Let $X = \sum_{i=1}^{n} X_i$ and $\|X\| = \sqrt{\sum_{i=1}^{n} \mathbb{E}(X_i^2)}$. Then we have

$$\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\|X\|^2 + M\lambda/3)}}.$$

This theorem implies McDiarmid’s Theorem.
Theorem [Chung, Lu] Suppose $X_i$ are independent random variables satisfying $X_i \leq M$, for $1 \leq i \leq n$. Let $X = \sum_{i=1}^{n} X_i$ and $\|X\| = \sqrt{\sum_{i=1}^{n} \mathbb{E}(X_i^2)}$. Then we have

$$\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\|X\|^2 + M\lambda/3)}}.$$ 

This theorem implies McDiarmid’s Theorem.

Let $X'_i = X_i - \mathbb{E}(X_i)$, and $X' = X - \mathbb{E}(X)$.

$$X - \mathbb{E}(X) = X' - \mathbb{E}(X')$$

$$\|X'\|^2 = \sum_{i=1}^{n} \mathbb{E}(X_i'^2) = \text{Var}(X).$$
Theorem [Chung, Lu] Suppose $X_i$ are independent random variables satisfying $X_i \geq 0$, for $1 \leq i \leq n$. Let $X = \sum_{i=1}^{n} X_i$ and $\|X\| = \sqrt{\sum_{i=1}^{n} \mathbb{E}(X_i^2)}$. Then we have

$$\Pr(X \leq E(X) - \lambda) \leq e^{-\frac{\lambda^2}{2\|X\|^2}}.$$
**Theorem [Chung, Lu]** Suppose $X_i$ are independent random variables satisfying $X_i \geq 0$, for $1 \leq i \leq n$. Let $X = \sum_{i=1}^{n} X_i$ and $\|X\| = \sqrt{\sum_{i=1}^{n} \mathbb{E}(X_i^2)}$. Then we have

$$\Pr(X \leq E(X) - \lambda) \leq e^{-\frac{\lambda^2}{2\|X\|^2}}.$$

**Proof:** Let $X'_i = -X_i$ and $X' = -X$. Applying the upper tail to $X'$ with $M = 0$, we get

$$\Pr(X \leq E(X) - \lambda) = \Pr(X' \geq E(X') + \lambda) \leq e^{-\frac{\lambda^2}{2\|X'\|^2}} = e^{-\frac{\lambda^2}{2\|X\|^2}}.$$
A special function

\[ g(y) = 2 \sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}. \]
A special function

\[ g(y) = 2 \sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}. \]

Facts:

- \( g(0) = 1. \)
A special function

\[ g(y) = 2 \sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(ey - 1 - y)}{y^2}. \]

Facts:

- \( g(0) = 1. \)
- \( g(y) \leq 1, \text{ for } y < 0. \)
A special function

\[ g(y) = 2 \sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}. \]

Facts:

- \( g(0) = 1. \)
- \( g(y) \leq 1, \) for \( y < 0. \)
- \( g(y) \) is monotone increasing, for \( y \geq 0. \)
A special function

\[ g(y) = 2 \sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}. \]

Facts:

- \( g(0) = 1 \).
- \( g(y) \leq 1, \text{ for } y < 0 \).
- \( g(y) \) is monotone increasing, for \( y \geq 0 \).
- For \( y < 3 \), we have

\[ g(y) = 2 \sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} \leq \sum_{k=2}^{\infty} \frac{y^{k-2}}{3^{k-2}} = \frac{1}{1 - y/3}. \]
Proof of upper tail

$$E(e^{tX}) = \prod_{i=1}^{n} E(e^{tX_i})$$
Proof of upper tail

\[ E(e^{tX}) = \prod_{i=1}^{n} E(e^{tX_i}) \]

\[ = \prod_{i=1}^{n} E(\sum_{k=0}^{\infty} \frac{t^k X_i^k}{k!}) \]
Proof of upper tail

\[
\begin{align*}
E(e^{tX}) &= \prod_{i=1}^{n} E(e^{tX_i}) \\
&= \prod_{i=1}^{n} E\left(\sum_{k=0}^{\infty} \frac{t^k X_i^k}{k!}\right) \\
&= \prod_{i=1}^{n} E\left(1 + tE(X_i) + \frac{1}{2} t^2 X_i^2 g(tX_i)\right)
\end{align*}
\]
Proof of upper tail

\[
E(e^{tX}) = \prod_{i=1}^{n} E(e^{tX_i}) \\
= \prod_{i=1}^{n} E\left(\sum_{k=0}^{\infty} \frac{t^k X_i^k}{k!}\right) \\
= \prod_{i=1}^{n} E(1 + tE(X_i) + \frac{1}{2} t^2 E(X_i^2) g(tX_i)) \\
\leq \prod_{i=1}^{n} (1 + tE(X_i) + \frac{1}{2} t^2 E(X_i^2) g(tM))
\]
Proof of upper tail

\[
E(e^{tX}) = \prod_{i=1}^{n} E(e^{tX_i})
\]

\[
= \prod_{i=1}^{n} E\left( \sum_{k=0}^{\infty} \frac{t^k X_i^k}{k!} \right)
\]

\[
= \prod_{i=1}^{n} E(1 + tE(X_i) + \frac{1}{2} t^2 X_i^2 g(tX_i))
\]

\[
\leq \prod_{i=1}^{n} \left( 1 + tE(X_i) + \frac{1}{2} t^2 E(X_i^2) g(tM) \right)
\]

\[
\leq \prod_{i=1}^{n} e^{tE(X_i) + \frac{1}{2} t^2 E(X_i^2) g(tM)}
\]
Proof of upper tail

\[ E(e^{tX}) = \prod_{i=1}^{n} E(e^{tX_i}) \]

\[ = \prod_{i=1}^{n} E(\sum_{k=0}^{\infty} \frac{t^k X_i^k}{k!}) \]

\[ = \prod_{i=1}^{n} E(1 + tE(X_i) + \frac{1}{2} t^2 X_i^2 g(tX_i)) \]

\[ \leq \prod_{i=1}^{n} (1 + tE(X_i) + \frac{1}{2} t^2 E(X_i^2) g(tM)) \]

\[ \leq \prod_{i=1}^{n} e^{tE(X_i) + \frac{1}{2} t^2 E(X_i^2) g(tM)} \]

\[ = e^{tE(X) + \frac{1}{2} t^2 g(tM) \|X\|^2}. \]
Hence, for $t$ satisfying $tM < 3$, we have
Hence, for $t$ satisfying $tM < 3$, we have

$$\Pr(X \geq E(X) + \lambda) = \Pr(e^{tX} \geq e^{tE(X)+t\lambda})$$
Hence, for $t$ satisfying $tM < 3$, we have

$$
\Pr(X \geq \mathbb{E}(X) + \lambda) = \Pr(e^{tX} \geq e^{t\mathbb{E}(X) + t\lambda}) \leq e^{-t\mathbb{E}(X) - t\lambda} \mathbb{E}(e^{tX})
$$
Hence, for $t$ satisfying $tM < 3$, we have

$$\Pr(X \geq E(X) + \lambda) = \Pr(e^{tX} \geq e^{tE(X) + t\lambda})$$
$$\leq e^{-tE(X) - t\lambda}E(e^{tX})$$
$$\leq e^{-t\lambda + \frac{1}{2}t^2g(tM)\|X\|^2}$$
Hence, for $t$ satisfying $tM < 3$, we have

\[
\Pr(X \geq E(X) + \lambda) = \Pr(e^{tX} \geq e^{tE(X) + t\lambda}) \\
\leq e^{-tE(X) - t\lambda} \mathbb{E}(e^{tX}) \\
\leq e^{-t\lambda + \frac{1}{2} t^2 g(tM) \|X\|^2} \\
\leq e^{-t\lambda + \frac{1}{2} t^2 \|X\|^2 \frac{1}{1-tM/3}}.
\]
Hence, for \( t \) satisfying \( tM < 3 \), we have

\[
\Pr(X \geq E(X) + \lambda) = \Pr(e^{tX} \geq e^{tE(X) + t\lambda}) \\
\leq e^{-tE(X) - t\lambda}E(e^{tX}) \\
\leq e^{-t\lambda + \frac{1}{2}t^2g(tM)\|X\|^2} \\
\leq e^{-t\lambda + \frac{1}{2}t^2\|X\|^2 \frac{1}{1-tM/3}}.
\]

Choose \( t = \frac{\lambda}{\|X\|^2 + M\lambda/3} \). We have \( 1 - \frac{Mt}{3} = \frac{\|X\|^2}{\|X\|^2 + M\lambda/3} \).
Hence, for \( t \) satisfying \( tM < 3 \), we have

\[
\Pr(X \geq E(X) + \lambda) = \Pr(e^{tX} \geq e^{tE(X)+t\lambda}) \\
\leq e^{-tE(X)-t\lambda}E(e^{tX}) \\
\leq e^{-t\lambda+\frac{1}{2}t^2g(tM)}\|X\|^2 \\
\leq e^{-t\lambda+\frac{1}{2}t^2\|X\|^2 \frac{1}{1-tM/3}}.
\]

Choose \( t = \frac{\lambda}{\|X\|^2+M\lambda/3} \). We have \( 1 - \frac{Mt}{3} = \frac{\|X\|^2}{\|X\|^2+M\lambda/3} \).

\[
\Pr(X > E(X) + \lambda) \leq e^{-t\lambda+t^2\|X\|^2 \frac{1}{2(1-Mt/3)}}
\]
Hence, for $t$ satisfying $tM < 3$, we have

\[
\Pr(X \geq E(X) + \lambda) = \Pr(e^{tX} \geq e^{tE(X)+t\lambda})
\leq e^{-tE(X)-t\lambda}E(e^{tX})
\leq e^{-t\lambda + \frac{1}{2} t^2 g(tM) \|X\|^2}
\leq e^{-t\lambda + \frac{1}{2} t^2 \|X\|^2} \frac{1}{1-tM/3}.
\]

Choose $t = \frac{\lambda}{\|X\|^2 + M\lambda/3}$. We have $1 - \frac{Mt}{3} = \frac{\|X\|^2}{\|X\|^2 + M\lambda/3}$.

\[
Pr(X > E(X) + \lambda) \leq e^{-t\lambda + t^2 \|X\|^2} \frac{1}{2(1-Mt/3)}
= e^{-\frac{\lambda^2}{\|X\|^2 + M\lambda/3}} + \frac{\lambda^2}{(\|X\|^2 + M\lambda/3)^2} \|X\|^2 \frac{\|X\|^2 + M\lambda/3}{2\|X\|^2}
\]
Hence, for $t$ satisfying $tM < 3$, we have

\[
\Pr(X \geq E(X) + \lambda) = \Pr(e^{tX} \geq e^{tE(X) + t\lambda}) \\
\leq e^{-tE(X) - t\lambda}E(e^{tX}) \\
\leq e^{-t\lambda + \frac{1}{2}t^2 g(tM) \|X\|^2} \\
\leq e^{-t\lambda + \frac{1}{2}t^2 \|X\|^2} \frac{1}{1-tM/3}.
\]

Choose $t = \frac{\lambda}{\|X\|^2 + M\lambda/3}$. We have $1 - \frac{Mt}{3} = \frac{\|X\|^2}{\|X\|^2 + M\lambda/3}$.

\[
Pr(X > E(X) + \lambda) \leq e^{-t\lambda + t^2 \|X\|^2} \frac{1}{2(1-Mt/3)} \\
= e^{-\frac{\lambda^2}{\|X\|^2 + M\lambda/3} + \frac{\lambda^2}{(\|X\|^2 + M\lambda/3)^2}} \|X\|^2 \frac{X \|\|X\|^2 + M\lambda/3}{2\|X\|^2} \\
= e^{-\frac{\lambda^2}{2(\|X\|^2 + M\lambda/3)}}. \quad \square
\]
Theorem [Chung, Lu] Let $X_i$ denote independent random variables satisfying $X_i \leq E(X_i) + a_i + M$, for $1 \leq i \leq n$. For, $X = \sum_{i=1}^{n} X_i$, we have

$$\Pr(X \geq E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=1}^{n} a_i^2 + M\lambda/3)}}.$$ 

Proof: Let $X'_i = X_i - E(X_i) - a_i$ and $X' = \sum_{i=1}^{n} X'_i$. We claim
Theorem [Chung, Lu] Let $X_i$ denote independent random variables satisfying $X_i \leq \mathbb{E}(X_i) + a_i + M$, for $1 \leq i \leq n$. For, $X = \sum_{i=1}^{n} X_i$, we have

$$\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=1}^{n} a_i^2 + M\lambda/3)}}.$$ 

Proof: Let $X'_i = X_i - \mathbb{E}(X_i) - a_i$ and $X' = \sum_{i=1}^{n} X'_i$. We claim

- $X'_i \leq M$ for $1 \leq i \leq n$. 
Theorem [Chung, Lu] Let $X_i$ denote independent random variables satisfying $X_i \leq \mathbb{E}(X_i) + a_i + M$, for $1 \leq i \leq n$. For, $X = \sum_{i=1}^{n} X_i$, we have

$$
\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=1}^{n} a_i^2 + M\lambda/3)}}.
$$

Proof: Let $X'_i = X_i - \mathbb{E}(X_i) - a_i$ and $X' = \sum_{i=1}^{n} X'_i$. We claim

- $X'_i \leq M$ for $1 \leq i \leq n$.
- $X' - \mathbb{E}(X') = X - \mathbb{E}(X)$.
Theorem [Chung, Lu] Let $X_i$ denote independent random variables satisfying $X_i \leq \mathbb{E}(X_i) + a_i + M$, for $1 \leq i \leq n$. For, $X = \sum_{i=1}^{n} X_i$, we have

$$\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=1}^{n} a_i^2 + M\lambda/3)}}.$$

Proof: Let $X'_i = X_i - \mathbb{E}(X_i) - a_i$ and $X' = \sum_{i=1}^{n} X'_i$. We claim

- $X'_i \leq M$ for $1 \leq i \leq n$.
- $X' - \mathbb{E}(X') = X - \mathbb{E}(X)$.
- $\|X'\|^2 = \text{Var}(X) + \sum_{i=1}^{n} a_i^2$. 
Theorem [Chung, Lu] Let \( X_i \) denote independent random variables satisfying \( X_i \leq E(X_i) + a_i + M \), for \( 1 \leq i \leq n \). For, \( X = \sum_{i=1}^{n} X_i \), we have

\[
Pr(X \geq E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=1}^{n} a_i^2 + M\lambda/3)}}.
\]

Proof: Let \( X'_i = X_i - E(X_i) - a_i \) and \( X' = \sum_{i=1}^{n} X'_i \). We claim

- \( X'_i \leq M \) for \( 1 \leq i \leq n \).
- \( X' - E(X') = X - E(X) \).
- \( \|X'\|^2 = \text{Var}(X) + \sum_{i=1}^{n} a_i^2 \).
\[ \Pr(X \geq E(X) + \lambda) = \Pr(X' \geq E(X') + \lambda) \]
\[ \Pr(X \geq E(X) + \lambda) = \Pr(X' \geq E(X') + \lambda) \leq e^{-\frac{\lambda^2}{2(\|X'\|^2 + M\lambda/3)}} \]
\[
\Pr(X \geq E(X) + \lambda) = \Pr(X' \geq E(X') + \lambda) \\
\leq e^{-\frac{\lambda^2}{2(\|X'\|^2 + M\lambda/3)}} \\
= e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=1}^{n} a_i^2 + M\lambda/3)}}.
\]

It remains to verify

\[
X' - E(X') = X - E(X).
\]

\[
\|X'\|^2 = \text{Var}(X) + \sum_{i=1}^{n} a_i^2.
\]
Theorem [Chung, Lu] Suppose $X_i$ are independent random variables satisfying $X_i \leq E(X_i) + M_i$, for $0 \leq i \leq n$. We order $X_i$’s so that $M_i$ are in an increasing order. Let $X = \sum_{i=1}^{n} X_i$. Then for any $1 \leq k \leq n$, we have

$$\Pr(X \geq E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2\text{Var}(X) + \sum_{i=k}^{n} (M_i - M_k)^2 + M_k \lambda/3}}.$$
Theorem [Chung, Lu] Suppose $X_i$ are independent random variables satisfying $X_i \leq E(X_i) + M_i$, for $0 \leq i \leq n$. We order $X_i$’s so that $M_i$ are in an increasing order. Let $X = \sum_{i=1}^{n} X_i$. Then for any $1 \leq k \leq n$, we have

$$\Pr(X \geq E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=k}^{n} (M_i - M_k)^2 + M_k \lambda/3)}}.$$ 

Compared with McDiarmid’s inequality.
Theorem [Chung, Lu] Suppose $X_i$ are independent random variables satisfying $X_i \leq \mathbb{E}(X_i) + M_i$, for $0 \leq i \leq n$. We order $X_i$’s so that $M_i$ are in an increasing order. Let $X = \sum_{i=1}^{n} X_i$. Then for any $1 \leq k \leq n$, we have

$$\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2\mathbb{V}ar(X) + \sum_{i=k}^{n} (M_i - M_k)^2 + M_k \lambda/3}}.$$ 

Compared with McDiarmid’s inequality

- $M$ is replaced by $M_k$. 

Theorem [Chung, Lu] Suppose $X_i$ are independent random variables satisfying $X_i \leq \mathbb{E}(X_i) + M_i$, for $0 \leq i \leq n$. We order $X_i$’s so that $M_i$ are in an increasing order. Let $X = \sum_{i=1}^{n} X_i$. Then for any $1 \leq k \leq n$, we have

$$\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=k}^{n}(M_i - M_k)^2 + M_k \lambda/3)}}.$$ 

Compared with McDiarmid’s inequality

- $M$ is replaced by $M_k$.
- Additional cost $\sum_{i=k}^{n}(M_i - M_k)^2$. 
**Theorem [Chung, Lu]** Suppose $X_i$ are independent random variables satisfying $X_i \leq E(X_i) + M_i$, for $0 \leq i \leq n$. We order $X_i$’s so that $M_i$ are in an increasing order. Let $X = \sum_{i=1}^{n} X_i$. Then for any $1 \leq k \leq n$, we have

$$\Pr(X \geq E(X) + \lambda) \leq e^{-\frac{\lambda^2 - 2\text{Var}(X) + \sum_{i=k}^{n} (M_i - M_k)^2 + M_k \lambda / 3}}.$$ 

Compared with McDiarmid’s inequality

- $M$ is replaced by $M_k$.
- Additional cost $\sum_{i=k}^{n} (M_i - M_k)^2$.
- McDiarmid’s inequality is a special case with $k = n$. 
Proof

For fixed $k$, we choose $M = M_k$ and

$$a_i = \begin{cases} 
0 & \text{if } 1 \leq i \leq k \\
M_i - M_k & \text{if } k \leq i \leq n 
\end{cases}$$
For fixed $k$, we choose $M = M_k$ and

$$a_i = \begin{cases} 
0 & \text{if } 1 \leq i \leq k \\
M_i - M_k & \text{if } k \leq i \leq n
\end{cases}$$

$$X_i - E(X_i) \leq M_i \leq a_i + M_k. \quad \text{for } 1 \leq k \leq n.$$
Proof

For fixed $k$, we choose $M = M_k$ and

\[ a_i = \begin{cases} 
0 & \text{if } 1 \leq i \leq k \\
M_i - M_k & \text{if } k \leq i \leq n 
\end{cases} \]

\[ X_i - \mathbb{E}(X_i) \leq M_i \leq a_i + M_k. \quad \text{for } 1 \leq k \leq n. \]

\[ \sum_{i=1}^{n} a_i^2 = \sum_{i=k}^{n} (M_i - M_k)^2. \]

Apply previous theorem with these $a_i$’s. \qed
Example: Consider the sum $X = \sum_{i=1}^{n} X_i$. 
Example: Consider the sum \( X = \sum_{i=1}^{n} X_i \).

- \( X_1, X_2, \ldots, X_n \): independent random variables.
An application

Example: Consider the sum $X = \sum_{i=1}^{n} X_i$.

- $X_1, X_2, \ldots, X_n$: independent random variables.
- For $1 \leq i \leq n - 1$, we have

$$
\Pr(X_i = 0) = 1 - p \quad \text{and} \quad \Pr(X_i = 1) = p.
$$
Example: Consider the sum $X = \sum_{i=1}^{n} X_i$.

- $X_1, X_2, \ldots, X_n$: independent random variables.
- For $1 \leq i \leq n - 1$, we have
  \[
  \Pr(X_i = 0) = 1 - p \quad \text{and} \quad \Pr(X_i = 1) = p.
  \]

- $X_n$ is special.
  \[
  \Pr(X_n = 0) = 1 - p \quad \text{and} \quad \Pr(X_n = \sqrt{n}) = p.
  \]
**Example:** Consider the sum \( X = \sum_{i=1}^{n} X_i \).

- \( X_1, X_2, \ldots, X_n \): independent random variables.
- For \( 1 \leq i \leq n - 1 \), we have
  \[
  \Pr(X_i = 0) = 1 - p \quad \text{and} \quad \Pr(X_i = 1) = p.
  \]
- \( X_n \) is special.
  \[
  \Pr(X_n = 0) = 1 - p \quad \text{and} \quad \Pr(X_n = \sqrt{n}) = p.
  \]
Expectation and Variance

\[
E(X) = \sum_{i=1}^{n} E(X_i)
\]
\[
= (n - 1)p + \sqrt{np}.
\]
Expectation and Variance

\[
E(X) = \sum_{i=1}^{n} E(X_i)
= (n - 1)p + \sqrt{np}.
\]

\[
\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i)
= (n - 1)p(1 - p) + np(1 - p)
= (2n - 1)p(1 - p).
\]
Applying McDiarmid’s Theorem
Applying McDiarmid’s Theorem

\[ M = (1 - p) \sqrt{n} \]
Applying McDiarmid’s Theorem

- \( M = (1 - p)\sqrt{n} \)

We have

\[
\Pr(X \geq E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2((2n-1)p(1-p)+(1-p)\sqrt{n}\lambda/3)}}.
\]
Applying McDiarmid’s Theorem

- \( M = (1 - p)\sqrt{n} \)

We have

\[
\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2((2n-1)p(1-p)+(1-p)\sqrt{n}\lambda/3)}}.
\]

In particular, for constant \( p \in (0, 1) \) and \( \lambda = \Theta\left(n^{\frac{1}{2}} + \epsilon\right) \), we have

\[
\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\Theta(n^\epsilon)}.
\]
Applying last Theorem
Applying last Theorem

- \( M_1 = \ldots = M_{n-1} = (1 - p) \)
- \( M_n = \sqrt{n}(1 - p) \)
Applying last Theorem

- \( M_1 = \ldots = M_{n-1} = (1 - p) \)
- \( M_n = \sqrt{n}(1 - p) \)

We choose \( k = n - 1 \),

\[
\text{Var}(X) + (M_n - M_{n-1})^2 \leq (1 - p^2)n.
\]
Applying last Theorem

- \( M_1 = \ldots = M_{n-1} = (1 - p) \)
- \( M_n = \sqrt{n}(1 - p) \)

We choose \( k = n - 1 \),

\[
\text{Var}(X) + (M_n - M_{n-1})^2 \leq (1 - p^2)n.
\]

\[
\Pr(X_i \geq E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2((1-p^2)n+(1-p)^2\lambda/3)}}.
\]
Applying last Theorem

- $M_1 = \ldots = M_{n-1} = (1 - p)$
- $M_n = \sqrt{n}(1 - p)$

We choose $k = n - 1$, 

$$\text{Var}(X) + (M_n - M_{n-1})^2 \leq (1 - p^2)n.$$ 

$$\Pr(X_i \geq E(X) + \lambda) \leq e^{-\frac{\lambda^2}{2((1-p^2)n+(1-p)^2\lambda/3)}}.$$ 

For constant $p \in (0, 1)$ and $\lambda = \Theta(n^{\frac{1}{2}+\epsilon})$, we have

$$\Pr(X \geq E(X) + \lambda) \leq e^{-\Theta(n^{2\epsilon})}.$$
Reference

