Notes of STAT 6060

WANG Lijun

April 30, 2019

This note consists of the lecture material (the main document), four homework (indexed by "Homework") and several personal comments (indexed by "Note").

1 Inequalities

1.1 Chebyshev's Inequality

Consider a random variable X, $\forall x > 0$,

1.1.1 Basic form

$$P(X \geq x) \leq \frac{\mathrm{E}|X|}{x},$$

more precisely,

$$P(X \geq x) \leq \frac{\mathrm{E}[X\mathbb{1}(X \geq x)]}{x}.$$

It is important to know the proof idea:

$$\mathbb{1}(X \geq x) \leq \frac{X^+}{x}.$$

1.1.2 Generating function

$$P(X \geq x) = P(e^{tX} \geq e^{tx}) \leq \frac{\mathrm{E}e^{tX}}{e^{tx}} = e^{-tx}\mathrm{E}(e^{tX})$$

1.1.3 Two dimensional case

Note that

$$\mathbb{1}[(X,Y) \in A] \leq \exp[sX + tY - \inf_{(x,y)\in A}(sx+ty)],$$

then we have

$$P((X,Y) \in A) \leq \exp[-\inf_{(x,y)\in A}(sx+ty)]\mathrm{E}(e^{sX+tY})$$

1.2 Lyapunov's Inequality

If $0 \le r \le s \le t$,

$$\mathrm{E}|X|^s \le [\mathrm{E}|X|^r]^{\frac{t-s}{t-r}} [\mathrm{E}|X|^t]^{\frac{s-r}{t-r}}. \tag{1}$$

Homework 1. *Verify* (1). *Hints: Hölder's Inequality.*
If $p, q \in [1, \infty]$ with $1/p + 1/q = 1$, then

$$\mathrm{E}|XY| \le [\mathrm{E}|X|^p]^{1/p} [\mathrm{E}|Y|^q]^{1/q}.$$

Proof. Let

$$\frac{1}{p} = \frac{t-s}{t-r}, \quad \frac{1}{q} = \frac{s-r}{t-r},$$

which satisfy $1/p + 1/q = 1$, and we have $r/p + t/q = s$, then by Hölder's Inequality,

$$\begin{aligned} \mathrm{E}|X|^s &= \mathrm{E}|X|^{r/p}|X|^{t/q} \\ &\le \left[\mathrm{E}(|X|^{r/p})^p\right]^{1/p} \left[\mathrm{E}(|X|^{t/q})^q\right]^{1/q} \\ &= [\mathrm{E}|X|^r]^{\frac{t-s}{t-r}} [\mathrm{E}|X|^t]^{\frac{s-r}{t-r}}. \end{aligned}$$

□

1.3 Kimball's Inequality

Theorem 1. *Suppose $g(x)$ and $h(x)$ are monotone increasing function, then*

$$\mathrm{E}g(X)h(X) \ge \mathrm{E}g(X)\mathrm{E}h(X).$$

Proof. Let X, Y be independent random variable satisfy $X \overset{d}{=} Y$, then

$$\begin{aligned} 2\mathrm{E}g(X)\mathrm{E}h(X) &= \mathrm{E}g(X)\mathrm{E}h(Y) + \mathrm{E}g(Y)\mathrm{E}h(X) \\ &= \mathrm{E}[g(X)h(Y) + g(Y)h(X)] \\ &\le \mathrm{E}[g(X)h(X) + g(Y)h(Y)], \end{aligned}$$

where the inequality is due to

$$g(X)h(Y) + g(Y)h(X) - g(X)h(X) - g(Y)h(Y) = -(g(X) - g(Y))(h(X) - h(Y)) \le 0.$$

If $g(x)$ is monotone increasing while $h(x)$ is monotone decreasing, then

$$\mathrm{E}g(X)h(X) \le \mathrm{E}g(X)\mathrm{E}h(X).$$

□

1.4 Bennett-Hoeffding's Inequality

Theorem 2. *Let X_n be independent random variables and let $S_n = \sum_{i=1}^n X_i$. Assume that $\mathrm{E}X_i \le 0, X_i \le a (a > 0)$ for each $1 \le i \le n$, and $\sum_{i=1}^n \mathrm{E}X_i^2 \le B_n^2$. Then*

$$P(S_n \ge x) \le \exp\Big(-\frac{x^2}{2B_n^2 + ax}\Big) \tag{2}$$

$$P(S_n \ge x) \le \exp\Big(-\frac{B_n^2}{a^2}\Big[\big(1+\frac{ax}{B_n^2}\big)\log\big(1+\frac{ax}{B_n^2}\big) - \frac{ax}{B_n^2}\Big]\Big) \tag{3}$$

where the second conclusion implies the first conclusion.

Proof. Intuitively,

$$(1+y)\log(1+y) - y = (1+y)(y - \frac{y^2}{2} + \frac{y^3}{3} - \cdots) - y = y + y^2 - \frac{y^2}{2} + \cdots - y = \frac{y^2}{2}\cdots,$$

and we always have

$$(1+y)\log(1+y) - y \ge \frac{y^2}{2(1+y)}.$$

Note that for $t > 0$,

$$P(S_n \ge x) \le e^{-tx}\mathrm{E}e^{tS_n} = e^{-tx}\prod_{i=1}^n \mathrm{E}e^{tX_i}.$$

If $s \le a$, we can find C_a such that

$$e^s \le 1 + s + s^2 C_a$$

for any $s \le a$, where

$$C_a = \sup_{s \le a} \frac{e^s - (1+s)}{s^2} = \frac{e^a - (1+a)}{a^2}.$$

Then

$$\begin{aligned}
\mathrm{E}e^{tX_i} &\le \mathrm{E}\Big[1 + tX_i + t^2X_i^2\Big(\frac{e^{ta} - 1 - ta}{t^2a^2}\Big)\Big] \\
&= 1 + \mathrm{E}X_i^2\frac{e^{ta} - 1 - ta}{a^2} \\
&\le \exp\Big[\mathrm{E}X_i^2\frac{e^{ta} - 1 - ta}{a^2}\Big],
\end{aligned}$$

it follows that

$$P(S_n \ge x) \le e^{-tx}\exp\Big(\sum \mathrm{E}X_i^2\frac{e^{ta} - 1 - ta}{a^2}\Big) \le \exp\Big(-tx + B_n^2\frac{e^{ta} - 1 - ta}{a^2}\Big) \triangleq \exp(g(t)).$$

Choose t to minimize $g(t)$ by letting

$$g'(t) = -x + \frac{B_n^2}{a^2}(ae^{ta} - a) = 0,$$

and

$$t = \frac{1}{a}\log\Big(\frac{xa + B_n^2}{xa}\Big).$$

□

Corollary 1. *If* $\mathrm{E}X_i \leq 0$ *and* $\sum \mathrm{E}X_i^2 \leq B_n^2$*, then for* $p \geq 1$ *and* $x > 0$*,*

$$P(S_n \geq xB_n) \leq P\left(\max_{1\leq i\leq n} X_i \geq \frac{xB_n}{p}\right) + \left(\frac{3p}{p+x^2}\right)^p .$$

Proof. By truncation, let $Y_i = X_i \mathbb{1}(X_i \leq xB_n/p)$, then

$$\mathrm{E}Y_i = \mathrm{E}X_i - \mathrm{E}X_i \mathbb{1}\left(X_i > \frac{xB_n}{p}\right) \leq 0$$

and $Y_i \leq \frac{xB_n}{p}$. Then

$$P(S_n \geq xB_n) \leq P\left(\max X_i \geq \frac{xB_n}{p}\right) + P\left(S_n \geq xB_n, \max X_i < \frac{xB_n}{p}\right) .$$

For the second term,

$$\begin{aligned}
P\left(S_n \geq xB_n, \max X_i < \frac{xB_n}{p}\right) &\leq P(\sum Y_i \geq xB_n) \\
&\leq \exp\Big[-\frac{p^2}{x^2}\Big(\big(1+\frac{x^2}{p}\big)\log\big(1+\frac{x^2}{p}\big) - \frac{x^2}{p}\Big)\Big] \\
&\leq \exp\Big[-p\log\big(1+\frac{x^2}{p}\big) + p\Big] \\
&= \Big[\frac{1}{e}\Big(1+\frac{x^2}{p}\Big)\Big]^{-p}
\end{aligned}$$

□

1.5 Rosenthal's Inequality

If X_i are independent random variable and $\mathrm{E}X_i = C$, and $\mathrm{E}|X_i|^p < \infty$ when $p \geq 2$. Then

$$E|S_n|^p \leq C_p\left[(ES_n^2)^{p/2} + \sum_{i=1}^n E|X_i|^p\right] \tag{4}$$

$$E|S_n|^p \geq D_p\left[(ES_n^2)^{p/2} + \sum_{i=1}^n E|X_i|^p\right] \tag{5}$$

Proof of (4). For $x \geq 0$, we have

$$g(x) = g(0) + \int_0^x g'(t)dt = g(0) + \int_0^\infty g'(t)\mathbb{1}(t \leq x)dt ,$$

then

$$Eg(X) = g(0) + \int_0^\infty g'(t)P(X \geq t)dt .$$

A special case is

$$E|X|^p = \int_0^\infty px^{p-1}P(|X| \geq x)dx\,.$$

Furthermore, it can be extended to negative x by

$$g(x) = g(0) + \int_{-\infty}^\infty g'(t)\left[\mathbb{1}(0 < t \leq x) - \mathbb{1}(x \leq t < 0)\right]dt\,.$$

Note that

$$\begin{aligned}
E|S_n|^p &= \int_0^\infty p|S_n|^{p-1}P(|S_n| \geq x)dx \\
&= B_n^p \int_0^\infty px^{p-1}P(|S_n| \geq xB_n)dx \\
&\leq B_n^p \sum_{i=1}^n \int_0^\infty px^{p-1}P\left(|X_i| \geq \frac{xB_n}{p}\right)dx + B_n^p \int_0^\infty px^{p-1}\cdot 2 \cdot \left(\frac{3p}{p+x^2}\right)^p dx\,,
\end{aligned}$$

where $B_n^2 = ES_n^2$, and in the first term,

$$\begin{aligned}
B_n^p \int_0^\infty px^{p-1}P\left(|X_i| \geq \frac{xB_n}{p}\right)dx &= B_n^p \int_0^\infty p\cdot\frac{p^{p-1}}{B_n^{p-1}}P(|X_i| \geq y)\frac{p}{B_n}dy \\
&= p^p \int_0^\infty py^{p-1}P(|X_i| \geq y)dy \\
&= p^p E|X_i|^p\,,
\end{aligned}$$

and the integral in the second term is finite since $p > 2$, then

$$E|S_n|^p \leq C_p\left[(ES_n^2)^{p/2} + \sum_{i=1}^n E|X_i|^p\right]\,.$$

□

Homework 2. *Verify* (5).

Proof. In the Lyapunov's Inequality (1), let $r = 0$, then we have

$$E|X|^s \leq [E|X|^t]^{s/t}\,.$$

Since $p \geq 2$, then

$$ES_n^2 \leq [E|S_n|^p]^{2/p}\,,$$

that is,

$$(ES_n^2)^{p/2} \leq E|S_n|^p\,.$$

By Marcinkiewicz-Zygmund inequality, there exists A_p such that

$$E|S_n|^p \geq A_p E\left(\left[\sum X_i^2\right]^{p/2}\right)\,,$$

and since $X_i^2 \geq 0$ and $p \geq 2$, then we have

$$E\left(\left[\sum X_i^2\right]^{p/2}\right) \geq E\left(\sum (X_i^2)^{p/2}\right) = \mathrm{E}\left(\sum |X_i|^p\right) ,$$

thus take $D_p = \frac{1}{A_p+1}$,

$$E|S_n|^p \geq D_p \left[(ES_n^2)^{p/2} + \sum_{i=1}^{n} E|X_i|^p\right] .$$

□

1.6 Nonnegative Random Variables

Theorem 3. *Assume that $X_i \geq 0$ with $EX_i^2 < \infty$. Let $\mu_n = \sum_{i=1}^n EX_i$ and $B_n^2 = \sum_{i=1}^n EX_i^2$. Then for $x > 0$,*

$$P(S_n \leq \mu_n - x) \leq \exp\left(-\frac{x^2}{2B_n^2}\right) .$$

Proof. Note that

$$P(S_n \leq \mu_n - x) = P(-S_n \geq -\mu_n + x) \leq e^{-t(-\mu_n + x)} E e^{-tS_n} .$$

Since if $s \leq a$, we have $e^s \leq 1 + s + s^2 C_a$, now if $s \leq 0$, then $e^s \leq 1 + s + s^2/2$, it follows that

$$e^{-tX_i} \leq 1 - tX_i + \frac{t^2 X_i^2}{2} ,$$

and hence

$$Ee^{-tX_i} \leq 1 - EtX_i + \frac{t^2}{2} EX_i^2 \leq \exp\left(-tEX_i + t^2 \frac{EX_i^2}{2}\right) .$$

Thus,

$$e^{-t(-\mu_n + x)} E e^{-tS_n} \leq \exp\left(t\mu_n - tx - t\sum_{i=1}^{n} EX_i + t^2 \sum_{i=1}^{n} EX_i^2/2\right) = \exp(-tx + t^2 B_n^2/2) ,$$

which is maximized when $t = x/B_n^2$, so

$$P(S_n \leq \mu_n - x) \leq \exp\left(-\frac{x^2}{B_n^2}\right) .$$

□

Theorem 4 (Bernoulli Random Variables). *Assume that $P(X_i = 1) = p_i$ and $P(X_i = 0) = 1 - p_i$. Then for $x > 0$,*

$$P(S_n \geq x) \leq \left(\frac{\mu e}{x}\right)^x ,$$

where $\mu = \sum_{i=1}^n p_i$.

1.7 Symmetric Random Variables

Theorem 5. *If ε_i are independent random variables with $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = 1/2$, then for any $x > 0$,*

$$P\left(\frac{\sum_{i=1}^n a_i\varepsilon_i}{\sqrt{\sum_{i=1}^n a_i^2}} \geq x\right) \leq e^{-x^2/2}\,.$$

Proof. Without loss of generality, assume $\sum_{i=1}^n a_i^2 = 1$. Since

$$\frac{1}{2}(e^s + e^{-s}) \leq e^{s^2/2}\,,$$

which can be showed easily by Taylor expansion, then we have

$$Ee^{ta_i\varepsilon_i} = \frac{1}{2}(e^{ta_i} + e^{-ta_i}) \leq e^{\frac{1}{2}t^2a_i^2}\,.$$

It follows that

$$\begin{aligned}
P(\sum_{i=1}^n a_i\varepsilon_i \geq x) &\leq e^{-tx}Ee^{t\sum_{i=1}^n a_i\varepsilon_i} \\
&\leq e^{-tx}\prod_{i=1}^n e^{\frac{1}{2}t^2a_i^2} \\
&= \exp\left(-tx + \frac{1}{2}t^2\sum_{i=1}^n a_i^2\right) \\
&= \exp\left(-tx + \frac{1}{2}t^2\right)\,,
\end{aligned}$$

which is maximized when $t = x$, thus

$$P(\sum_{i=1}^n a_i\varepsilon_i \geq x) \leq e^{-\frac{x^2}{2}}\,.$$

□

Theorem 6. *If $X_1, \ldots, X_n$ are independent symmetric, i.e., $X_i \stackrel{d}{=} -X_i$, then for any $x > 0$,*

$$P\left(\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n X_i^2}} \geq x\right) \leq e^{-x^2/2}\,.$$

Proof. Introduce independent $\{\varepsilon_i\}$ random variables and $P(\varepsilon_i = -1) = P(\varepsilon_i = 1) = 1/2$, then $X_i \overset{d}{=} X_i\varepsilon_i$. Thus,

$$\begin{aligned} P\left(\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n X_i}} \geq x\right) &= P\left(\frac{\sum_{i=1}^n X_i\varepsilon_i}{\sqrt{\sum_{i=1}^n (X_i\varepsilon_i)^2}} \geq x\right) \\ &= P\left(\frac{\sum_{i=1}^n X_i\varepsilon_i}{\sqrt{\sum_{i=1}^n X_i^2}} \geq x\right) \\ &= E\left[P\left(\frac{\sum_{i=1}^n X_i\varepsilon_i}{\sqrt{\sum_{i=1}^n X_i^2}} \geq x\right) \Bigg| X_1, \ldots, X_n\right] \\ &= Ee^{-x^2/2} = e^{-x^2/2}. \end{aligned}$$

□

Theorem 7 (He and Shao, 2000). *Let $X_1, \ldots, X_n$ be independent random variables with $EX_i = 0$ and $\sum_{i=1}^n EX_i^2 \leq B_n^2$. Let $S_n = \sum_{i=1}^n X_i$ and $V_n^2 = \sum_{i=1}^n X_i^2$, then*

$$P(S_n \geq x(V_n + 4B_n)) \leq 2e^{-x^2/2}.$$

Proof. Introduce independent copy of $\{X_i\}$, $\{Y_i\}$, then $\{X_i - Y_i\}$ are symmetric, then

$$P\left(\sum_{i=1}^n (X_i - Y_i) \geq x\sqrt{\sum_{i=1}^n (X_i - Y_i)^2}\right) \leq e^{-x^2/2}.$$

By triangle inequality,

$$\sqrt{\sum (X_i - Y_i)^2} \leq \sqrt{\sum X_i^2} + \sqrt{\sum Y_i^2},$$

then for $x \geq 1$,

$$\begin{aligned} &\left\{\sum X_i \geq x(V_n + D_n + C_n), \left|\sum Y_i\right| \leq C_n, \sum Y_i^2 \leq D_n\right\} \\ &\subset \left\{\sum (X_i - Y_i) \geq x\left(\sqrt{\sum (X_i - Y_i)^2} - D_n + D_n + C_n\right) - C_n\right\} \\ &\subset \left\{\sum (X_i - Y_i) \geq x\sqrt{\sum (X_i - Y_i)^2}\right\}, \end{aligned}$$

then

$$\begin{aligned} &P\left(\sum X_i \geq x(V_n + D_n + C_n), |\sum Y_i| \leq C_n\right) P\left(|\sum Y_i| \leq C_n, \sum Y_i^2 \leq D_n\right) \\ &\leq P\left(\sum (X_i - Y_i) \geq x\sqrt{\sum (X_i - Y_i)^2}\right). \end{aligned}$$

Choose $C_n = 2B_n$ and $D_n = 2B_n$, then

$$\begin{aligned} P\left(|\sum Y_i| > C_n\right) &\leq \frac{E(\sum Y_i)^2}{C_n^2} = \frac{1}{4} \\ P\left(\sum Y_i^2 \geq D_n^2\right) &\leq \frac{\sum EY_i^2}{D_n^2} = \frac{1}{4}. \end{aligned}$$

By the inequality

$$P(AB) \geq 1 - P(A^c) - P(B^c)\,,$$

we have

$$P\left(|\sum Y_i| \leq C_n, \sum Y_i^2 \leq D_n\right) > \frac{1}{2}\,,$$

and hence

$$P\left(\sum X_i \geq x(V_n + 4B_n)\right) \leq 2e^{-x^2/2}\,.$$

□

Open Question 1 (Conjecture). *If $\sum a_i^2 = 1$, then*

$$P(|\sum a_i\varepsilon_i| > 1) \leq 1/2\,,$$

more generally,

$$P(|\sum a_i\varepsilon_i| > y) + P(|\sum a_i\varepsilon_i| > 1/y) \leq 1\,.$$

2 Stein's Method

Let W be a real-valued random variable. If W has a standard normal distribution, then

$$Ef'(W) = EWf(W)$$

for any absolutely continuous function f with $E|f'(W)| < \infty$. If the equation holds for any continuous and piecewise continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with $|f'(Z)| < \infty$, then W has a standard normal distribution.

The Stein's equation is

$$f'(w) - wf(w) = 1(w \leq z) - \Phi(z)\,,$$

and a more general one:

$$f'(w) - wf(w) = h(w) - Eh(Z).$$

We want to bound

$$\Delta_n = \sup_z |P(X \leq z) - P(Z \leq z)| = \sup_z |P(X \leq z) - \Phi(z)|\,.$$

The key idea is this:

$$\mathrm{E}[Yf(Y)] = \mathrm{E}[f'(Y)]$$

for all smooth f iff $Y \sim N(0,1)$. This suggests the following idea: if we can show that $\mathrm{E}[Yf(Y) - f'(Y)]$ is close to 0, then Y should be almost Normal.

The Stein function f associated with h is a function satisfying

$$f'(x) - xf(x) = h(x) - \mathrm{E}[h(Z)]\,.$$

It then follows that

$$\mathrm{E}[h(X)] - \mathrm{E}[h(Z)] = \mathrm{E}[f'(X) - Xf(X)]$$

and showing that X is close to normal amounts to showing that $\mathrm{E}[f'(X) - Xf(X)]$ is small.

Choose any $z \in \mathbb{R}$, and let $h(x) = I(X \le z) - \Phi(z)$. Let f_z denote the Stein function for h; thus

$$f_z'(x) - xf(x) = I(x \le z) - \Phi(z)\,.$$

Let $\mathcal{F} = \{f_z : z \in \mathbb{R}\}$. From the equation $f'(x) - xf(x) = h(x) - \mathrm{E}[h(Z)]$ it follows that

$$P(X \le z) - P(Y \le z) = \mathrm{E}[f'(X) - Xf(X)]$$

and so

$$\Delta_n = \sup_n |P(X \le z) - P(Z \le z)| \le \sup_{f \in \mathcal{F}} |E[f'(X) - Xf(X)]|\,.$$

Example 1 (Sums of Independent Random Variables). *Let $\xi_1, \dots, \xi_n$ be independent random variable such that $\mathrm{E}\xi_i = 0$ for $1 \le i \le n$ and $\sum_{i=1}^n \mathrm{E}\xi_i^2 = 1$. Let $W = \sum_{i=1}^n \xi_i$.*

Proof. Our goal is to estimate

$$\mathrm{E}h(W) - \mathrm{E}h(Z) = \mathrm{E}f'(W) - \mathrm{E}[Wf(W)]\,.$$

The main idea of Stein's method is to rewrite $\mathrm{E}[Wf(W)]$ in terms of a functional of f'.

$$\begin{aligned}
\mathrm{E}[Wf(W)] &= \sum_{i=1}^n \mathrm{E}[\xi_i f(W)] \\
&= \sum_{i=1}^n \mathrm{E}\Big[\xi_i(f(W) - f(W - \xi_i))\Big] \\
&= \sum_{i=1}^n \mathrm{E}\Big[\xi_i(f(W^{(i)} + \xi_i) - f(W^{(i)}))\Big] \quad \text{Let } W^{(i)} = W - \xi_i \\
&= \sum_{i=1}^n \mathrm{E}\Big[\xi_i \int_{-\infty}^{\infty} f'(W^{(i)} + t)[\mathbb{1}(0 < t < \xi_i) - \mathbb{1}(\xi_i < t < 0)]dt\Big] \\
&= \sum_{i=1}^n \int_{-\infty}^{\infty} \mathrm{E}f'(W^{(i)} + t)\mathrm{E}[\xi_i(\mathbb{1}(0 < t < \xi_i) - \mathbb{1}(\xi_i < t < 0))]dt \\
&= \sum_{i=1}^n \int_{-\infty}^{\infty} K_i(t)dt\,.
\end{aligned}$$

Thus, for very nice f,

$$\mathrm{E}[Wf(W)] = \sum_{i=1}^n \mathrm{E}\int_{-\infty}^{\infty} f'(W^{(i)} + t)K_i(t)dt\,.$$

From $\sum_{i=1}^n \int_{-\infty}^\infty K_i(t)dt = \sum_{i=1}^n \mathrm{E}\xi_i^2 = 1$, it follows that

$$\mathrm{E}f'(W) = \sum_{i=1} \mathrm{E} \int_{-\infty}^{\infty} f'(W)K_i(t)dt\,.$$

Thus,

$$\mathrm{E}f'(W) - \mathrm{E}[Wf(W)] = \sum_{i=1}^{n} \mathrm{E} \int_{-\infty}^{\infty} [f'(W) - f'(W^{(i)} + t)]K_i(t)dt\,.$$

By the mean value theorem,

$$|f'(W^{(i)} + t) - f'(W)| \le \|f''\|(|t| + |\xi_i|)\,.$$

Then

$$\begin{aligned}
\mathrm{E}\int_{-\infty}^{\infty} |f'(W^{(i)} + t) - f'(W)|K_i(t)dt &\le \|f''\|\mathrm{E}\int_{-\infty}^{\infty} (|t| + |\xi|)K_i(t)dt \\
&= \|f''\|\Big(\frac{1}{2}\mathrm{E}|\xi_i|^3 + \mathrm{E}|\xi_i|\mathrm{E}\xi_i^2\Big) \\
&\le \frac{3}{2}\|f''\|\mathrm{E}|\xi_i|^3 \\
&\le 3\|h''\|\mathrm{E}|\xi_i|^3\,.
\end{aligned}$$

Thus,

$$|\mathrm{E}h(W) - \mathrm{E}h(Z)| \le 3\|h''\| \sum_{i=1}^{3} \mathrm{E}|\xi_i|^3\,.$$

More sharp bound can be

$$|f'(W^{(i)} + t) - f'(W)| \le 2\|h''\| \min(|t| + |\xi_i|, 1)\,.$$

□

Theorem 8. *Assume that there exists δ such that for any h satisfying $\|h'\| \triangleq \sup_w |h'(w)| < \infty$. Then*

$$\sup_z |P(W \le z) - \Phi(z)| \le 2\delta^{1/2}\,.$$

Proof.

$$\begin{aligned}
P(W \le z) - \Phi(z) &\le \mathrm{E}h_\varepsilon(W) - \mathrm{E}h_\varepsilon(Z) + \mathrm{E}h_\varepsilon(Z) - \mathrm{E}\mathbb{1}(Z \le z) \\
&\le \frac{\delta}{\varepsilon} + \mathrm{E}\mathbb{1}(z \le Z \le z + \varepsilon) \\
&= \frac{\delta}{\varepsilon} + \varepsilon\,.
\end{aligned}$$

Take $\varepsilon = \sqrt{\delta}$. □

Theorem 9 (Berry-Esseen Bound). $\sup_z |P(W \le z) - \Phi(z)| \le 4.1 \sum_{i=1}^n E|\xi_i|^3\,.$

Proof. Note that

$$Ef'(W) - EWf(W) = \sum_{i=1}^{n} E\int_{-\infty}^{\infty}(f'(W) - f'(W^{(i)} + t)K_i(t))dt\,,$$

where

$$\begin{aligned}
&E\int(f'(W) - f'(W^{(i)} + t))K_i(t)K_i(t)dt\\
&= E\int(Wf(W) - (W^{(i)} + t)f(W^{(i)} + t))K_i(t)dt + E\int(\mathbb{1}(W \le z) - \mathbb{1}(W^{(i)} + t \le z))K_i(t)dt\\
&\triangleq A_{i,1} + A_{i,2}\,.
\end{aligned}$$

Since

$$\begin{aligned}
\left|Wf(W) - (W^{(i)} + t)f(W^{(i)} + t)\right| &= \left|(W^{(i)} + \xi_i)f(W) - (W^{(i)} + t)f(W^{(i)} + t)\right|\\
&= \left|W^{(i)}[f(W) - f(W^{(i)} + t)] + \xi_i f(W) - tf(W^{(i)} + t)\right|\\
&\le |W^{(i)}|(|t| + |\xi_i|) + |\xi_i| + |t|\\
&= (|W^{(i)}| + 1)(|t| + |\xi_i|)\,,
\end{aligned}$$

then

$$\sum E\int(W^{(i)} + 1)(|t| + |\xi_i|)K_i(t)dt \le 2\sum E|\xi_i|^3 \cdot \frac{3}{2} = 3\sum E|\xi_i|^3\,,$$

and hence $|A_{i,1}| \le 3\sum E|\xi_i|^3$. The second term $A_{i,2}$ can be written as

$$A_{i,2} = \int \left(P(W^{(i)} \le z - \xi_i) - P(W^{(i)} \le z - t)\right) K_i(t) \le \int P(z - t < W^{(i)} < z - \xi_i)K_i(t)dt\,,$$

and we claim that

$$P(a \le W^{(i)} \le b) \le b - a + 2\sum_{i=1}^{n} E|\xi_i^3|\,,$$

which is the following theorem. □

Theorem 10. *$W = \sum \xi_i$, where $\mathrm{E}\xi_i = 0$ and $\sum \mathrm{E}\xi_i^2 = 1$, then*

$$P(a \le W \le b) = b - a + 2r\,,$$

where $r = \sum \mathrm{E}|\xi_i|^3$.

Proof.

$$
\begin{aligned}
\mathrm{E}[Wf(W)] = \sum_{i=1}^n \mathrm{E}[\xi_i f(W)] &= \sum_{i=1}^n \mathrm{E}[\xi_i (f(W) - f(W-\xi_i))] \\
&= \sum_{i=1}^n \mathrm{E}\Big[\xi_i \int_{-\xi_i}^0 f'(W+t)dt\Big] \\
&= \sum_{i=1}^n \mathrm{E}\Big[\xi_i \int_{-\infty}^{\infty} f'(W+t)(\mathbb{1}(-\xi_i < t < 0) - \mathbb{1}(0 < t < -\xi_i))dt\Big] \\
&= \sum_{i=1}^n \mathrm{E}\Big[\int_{-\infty}^{\infty} f'(W+t)\xi_i(\mathbb{1}(-\xi_i < t < 0) - \mathbb{1}(0 < t < -\xi_i))dt\Big] \\
&\triangleq \sum_{i=1}^n \mathrm{E}\int_{-\infty}^{\infty} f'(W+t)\hat{K}_i(t)dt \,.
\end{aligned}
$$

Let

$$
f(w) = \begin{cases} -\frac{b-a}{2} - \delta & \text{if } w < a - \delta \\ w - \frac{a+b}{2} & \text{if } a - \delta \le w \le b + \delta \\ \frac{b-a}{2} + \delta & \text{if } w > b + \delta \end{cases} ,
$$

then

$$
\mathrm{E}[Wf(W)] \le (\frac{b-a}{2} + \delta)\mathrm{E}|W| \le \frac{b-a}{2} + \delta \,,
$$

and

$$
\begin{aligned}
\sum_{i=1}^n \mathrm{E}\int_{-\infty}^{\infty} f'(W+t)\hat{K}_i(t)dt &\ge \sum_{i=1}^n \mathrm{E}\left[\mathbb{1}(a \le W \le b)\int_{|t|\le\delta} \hat{K}_i(t)dt\right] \\
&= \sum_{i=1}^n \mathrm{E}\left[\mathbb{1}(a \le W \le b)|\xi_i| \min(|\xi_i, \delta)\right] \\
&= \mathrm{E}\left[\mathbb{1}(a \le W \le b)\sum_{i=1}^n |\xi_i| \min(|\xi_i|, \delta)\right] \\
&= \mathrm{E}\left[\mathbb{1}(a \le W \le b)\sum_{i=1}^n \eta_i\right] \\
&= \mathrm{E}\left[\mathbb{1}(a \le W \le b)\sum \mathrm{E}\eta_i\right] + \mathrm{E}\left[\mathbb{1}(a \le W \le b)\sum (\eta_i - \mathrm{E}\eta_i)\right] \\
&\ge P(a \le W \le b)\sum \mathrm{E}\eta_i - \mathrm{E}\left|\sum \eta_i - \mathrm{E}\eta_i\right| \\
&\ge P(a \le W \le b)\sum \mathrm{E}\eta_i - \delta \,,
\end{aligned}
$$

where the last inequality is follows from

$$
\mathrm{E}\left|\sum \eta_i - \mathrm{E}\eta_i\right| \le \sqrt{\sum_{i=1}^n \mathrm{E}\eta_i^2} \le \sqrt{\sum_{i=1}^n \mathrm{E}\xi_i^2\delta^2} = \delta \,.
$$

Thus,

$$\frac{b-a}{2} + \delta \geq P(a \leq W \leq b) \sum \mathrm{E}\eta_i - \delta \,.$$

Note that

$$\min(x,y) \geq x - \frac{x^2}{4y} \qquad x>0, y>0 \,.$$

Then

$$\sum \mathrm{E}\eta_i \geq \sum \mathrm{E}(\xi_i^2 - \frac{|\xi_i|^3}{4\delta}) = 1 - \frac{1}{4\delta}\mathrm{E}|\xi_i|^3 \,,$$

take $\delta = \sum \mathrm{E}|\xi_i|^3/2$, then we have

$$P(a \leq W \leq b) \leq b - a + 4\delta \,.$$

□

2.1 Randomized Concentration Inequality

Theorem 11 (Randomized Concentration Inequality). *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables satisfying $E\xi_i = 0$ and $E|\xi_i|^3 < \infty$ for each $1 \leq i \leq n$ and such that $\sum_{i=1}^n E\xi_i^2 = 1$. Let $W = \sum_{i=1}^n \xi_i$, $\Delta_1 = \Delta_1(\xi_1, \ldots, \xi_n)$, $\Delta_2 = \Delta_2(\xi_1, \ldots, \xi_n)$. Then*

$$P(\Delta_1 \leq W \leq \Delta_2) \leq 4\sum_{i=1}^n E|\xi_i|^3 + E|W(\Delta_2 - \Delta_1)| + \sum_{i=1}^n E|\xi_i(\Delta_1 - \Delta_{1,i})| + \sum_{i=1}^n E|\xi_i(\Delta_2 - \Delta_{2,i})| \,,$$

where $\Delta_{1,i}$ and $\Delta_{2,i}$ are Borel measurable functions of $(\xi_j, 1 \leq j \leq n, j \neq i)$.

Proof. Let

$$f_{a,b}(w) = \begin{cases} -\frac{b-a}{2} - \delta & \text{if } w \leq a - \delta \\ w - \frac{a+b}{2} & \text{if } a - \delta \leq w \leq b + \delta \\ \frac{b-a}{2} + \delta & \text{if } w \geq b + \delta \end{cases} \,,$$

then

$$EWf_{\Delta_1,\Delta_2}(W) = \sum_{i=1}^n E\xi_i(f_{\Delta_1,\Delta_2}(W) - f_{\Delta_1,\Delta_2}(W - \xi_i)) + \sum_{i=1}^n E\xi_i(f_{\Delta_1,\Delta_2}(W-\xi_i) - f_{\Delta_{1i},\Delta_{2i}}(W - \xi_i)) \,.$$

It can be verified that

$$|f_{\Delta_1,\Delta_2}(w) - f_{\Delta_{1,i},\Delta_{2,i}}(w)| \leq |\Delta_1 - \Delta_{1,i}|/2 + |\Delta_2 - \Delta_{2,i}|/2 \,,$$

then yields

$$\left| \sum_{i=1}^n E\xi_i(f_{\Delta_1,\Delta_2}(W - \xi_i) - f_{\Delta_{1i},\Delta_{2i}}(W - \xi_i)) \right| \leq \frac{1}{2}\left[\sum_{i=1}^n E|\xi_i(\Delta_1 - \Delta_{1,i})| + \sum_{i=1}^n E|\xi_i(\Delta_2 - \Delta_{2,i})| \right]$$

More details can be found in Chen, Goldstein, and Shao (2010). □

2.2 U-Statistics

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables, and for some $m \geq 2$, let $h(x_1, \ldots, x_m)$ be a symmetric, real-valued function, where $m < n/2$ may depend on n, and let $\theta = Eh(X_{i_1}, \ldots, X_{i_m})$. The class of U-statistics are those random variables that can be written as

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} h(X_{i_1}, \ldots, X_{i_m}) .$$

Here we focus on $m = 2$ with symmetric kernel $h(x, y) = h(y, x)$, and $\theta = Eh(X_1, X_2)$. Without loss of generality, assume $\theta = 0$. Let $g(x) = Eh(x, X)$, then we can decompose the U-statistic as follows:

$$\begin{aligned}\sum_{1 \leq i < j \leq n} h(X_i, X_j) &= \sum_{j=2}^{n} \sum_{i=1}^{j-1} \Big(h(X_i, X_j) - g(X_i) - g(X_j)\Big) + \sum_{j=2}^{n} \sum_{i=1}^{j-1} (g(X_i) + g(X_j)) \\ &= \sum_{j=2}^{n} \sum_{i=1}^{j-1} \Big(h(X_i, X_j) - g(X_i) - g(X_j)\Big) + (n-1) \sum_{i=1}^{n} g(X_i) .\end{aligned}$$

Then

$$\begin{aligned}U_n &= \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} h(X_i, X_j) \\ &= \frac{2}{n} \sum g(X_i) + \frac{2}{n(n-1)} \sum_{j=2}^{n} \sum_{i=1}^{j-1} (h(X_i, X_j) - g(X_i) - g(X_j)) \\ &= \frac{2}{n} \sum g(X_i) + \frac{2}{n} \Delta_n .\end{aligned}$$

If $\mathrm{E}h^2(X_1, X_2) < \infty$ and $\sigma_1^2 = \mathrm{Var}(g(X_1)) > 0$, we have the central limit theorem,

$$\sup_x \Big| P\Big(\frac{\sqrt{n}}{2\sigma_1} U_n \leq x\Big) - \Phi(x) \Big| \to 0 \text{ as } n \to \infty .$$

3 Exchangeable pair

Let W_n be a sequence of random variables. Using the exchangeable pair approach of Stein's method, we can identify the limiting distribution of W_n as well as the L_1 bound of the approximation.

Write $W = W_n$ and let (W, W') be an exchangeable pair, that is, (W, W') and (W', W) have the same joint distribution. Put $\Delta = W - W'$, for the normal approximation, assume that

$$\mathrm{E}(\Delta \mid W) = \lambda(W + R_1) .$$

Note that if $h(x,y) = -h(y,x)$, then $Eh(W,W') = -Eh(W',W) = -Eh(W,W')$, and hence $Eh(W,W') = 0$, then

$$\begin{aligned}
0 &= E[(W-W')(f(W)+f(W'))] \\
&= 2E[(W-W')f(W)] - E[(W-W')(f(W)-f(W'))] \\
&= 2E\{E[(W-W')f(W) \mid W]\} - E\{\Delta[f(W)-f(W-\Delta)]\} \\
&= 2E\{f(W)E(W-W' \mid W)\} - E\{\Delta[f(W)-f(W-\Delta)]\} \\
&= 2E[f(W)\lambda(W+R_1)] - E[\Delta \int_{-\Delta}^{0} f'(W+t)dt] \\
&= 2\lambda \left[E(Wf(W)) + E(f(W)R_1) - E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt \right],
\end{aligned}$$

where

$$\hat{K}(t) = \frac{\Delta\left[\mathbb{1}(-\Delta \le t < 0) - \mathbb{1}(0 \le t \le -\Delta)\right]}{2\lambda}$$

is nonnegative, and $\int_{-\infty}^{\infty} \hat{K}(t)dt = \frac{\Delta^2}{2\lambda}$, $\int_{-\infty}^{\infty} |t|\hat{K}(t)dt = \frac{|\Delta|^3}{4\lambda}$. It follows that

$$EWf(W) = E\left(\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt\right) - E[f(W)R_1].$$

Then, for any absolutely continuous function h with $\|h'\| < \infty$,

$$|\mathrm{E}h(W) - \mathrm{E}h(Z)| \le 2\|h'\|\left(\mathrm{E}\left|1 - \frac{1}{2\lambda}\mathrm{E}(\Delta^2 \mid W)\right| + \frac{1}{\lambda}\mathrm{E}|\Delta|^3 + \mathrm{E}|R|\right).$$

Proof. Consider the Stein equation,

$$f'(w) - wf(w) = h(w) - Eh(Z).$$

Note that

$$\begin{aligned}
Eh(W) - Eh(Z) &= Ef'(W) - EWf(W) \\
&= Ef'(W) - E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt \\
&= Ef'(W) - E\int_{-\infty}^{\infty} (f'(W+t) - f'(W) + f'(W))\hat{K}(t)dt + E[R_1 f(W)] \\
&= E\left(f'(W) - f'(W)\frac{\Delta^2}{2\lambda}\right) - E\int_{-\infty}^{\infty} [f'(W+t) - f'(W)]\hat{K}(t)dt + ER_1 f(W),
\end{aligned}$$

where the first term

$$E\left[f'(W)\left(1 - \frac{\Delta^2}{2\lambda}\right)\right] = E\left[f'(W)\left(1 - \frac{E(\Delta^2 \mid W)}{2\lambda}\right)\right]$$

can be bounded by $2\|h'\|E\left|1 - \frac{E(\Delta^2|W)}{2\lambda}\right|$, and it is not proper to directly bound by $f'(W)$ in the first equation, which is not sharp.

Since $|f'(W+t)-f'(W)| \le |t|\|f'\|$, then the second term

$$E\int_{-\infty}^{\infty}[f'(W+t)-f'(W)]\hat{K}(t)dt \le 2 \le 2\|h'\|\frac{E|\Delta|^3}{4\lambda}$$

□

Now consider how to construct W' in general, denote $W = W(\xi_1,\ldots,\xi_n)$, where ξ_i are independent, let

$$W' = W(\xi_1,\ldots,\xi_I',\ldots,\xi_n)\,,$$

where I is a random index, which is independent of other random variables, $P(I=k)=\frac{1}{n}$ for $1\le k\le n$, and $\{\xi_i'\}$ are independent copy of $\{\xi_i\}$, then (W,W') is exchangeable.

Homework 3. *Verify the above property.*

Proof. Note that

$$\begin{aligned}
P(W=w,W'=w') &= E[P(W=w,W'=w' \mid I)] \\
&= \frac{1}{n}\sum_{i=1}^n P(W=w,W'=w' \mid I=i) \\
&= \frac{1}{n}\sum_{i=1}^n P(W=w',W=w \mid I=i) \\
&= E[P(W=w',W'=w \mid I)] \\
&= P(W=w',W'=w)\,,
\end{aligned}$$

which implies that (W,W') is exchangeable. □

Consider a special case, $W=\sum_1^n \xi_i$, where ξ_i are independent, $E\xi_i=0$ and $\sum_{i=1}^n E\xi_i^2 = 1$, then $W' = W-\xi_I+\xi_I'$, that is, $W-W'=\xi_I-\xi_I'$, it follows that

$$E(W-W' \mid W) = E(\xi_I-\xi_I' \mid W) = \frac{1}{n}\sum_{i=1}^n E(\xi_i-\xi_i' \mid W)\,.$$

Note that conditioning on a larger class, we have

$$\frac{1}{n}\sum_{i=1}^n E(\xi_i-\xi_i' \mid \xi_j, 1\le j\le n) = \frac{1}{n}\left(\sum_{i=1}^n(\xi_i-E\xi_i')\right) = \frac{1}{n}\sum_{i=1}^n\xi_i = \frac{1}{n}W\,,$$

thus $E(W-W' \mid W) = W/n$. And we have

$$\frac{E|\Delta|^3}{\lambda} = nE|\xi_I-\xi_I'|^3 = n\times\frac{1}{n}\sum_{i=1}^n E(\xi_i-\xi_i')^3 = \sum_{i=1}^n E(\xi_i-\xi_i')^3 \le 8\sum_{i=1}^n E|\xi_i|^3\,.$$

Next consider

$$\begin{aligned} E(\Delta^2 \mid \xi_j, 1 \le j \le n) &= \frac{1}{n}\sum_{i=1}^{n} E\left((\xi_i - \xi_i')^2 \mid \xi_j, 1 \le j \le n\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}(\xi_i^2 + E(\xi_i')^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}(\xi_i^2 + E\xi_i^2)\,, \end{aligned}$$

then

$$1-\frac{E(\Delta^2 \mid \xi_j, 1 \le j \le n)}{2\lambda} = 1-\frac{1}{2}\sum_{i=1}^{n}(\xi_i^2+E\xi_i^2) = \frac{1}{2}(1-\sum \xi_i^2) = \frac{1}{2}\sum_{i=1}^{n}(E\xi_i^2-\xi_i^2) \triangleq \frac{1}{2}\sum_{i=1}^{n}\eta_i\,,$$

where $E\eta_i = 0$ and by Jensen's Inequality,

$$E|\sum \eta_i| \le \sqrt{\mathrm{Var}(\sum \eta_i)} = \sqrt{\sum \mathrm{Var}(\eta_i)} \le \sqrt{\sum E\eta_i^2} \le \sqrt{\sum E\xi_i^4}\,.$$

Here and in the sequel, Z denotes the standard normal random variable. For the Berry-Esseen bound, we have

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le \mathrm{E}\Big|1 - \frac{1}{2\lambda}\mathrm{E}(\Delta^2 \mid W)\Big| + \mathrm{E}|R| + \Big(\frac{\mathrm{E}|\Delta|^3}{\lambda}\Big)^{1/2}.$$

It is known that it usually fails to provide an optimal bound.

Theorem 12 (Shao and Zhang, 2019). *Let (W, W') be an exchangeable pair satisfying*

$$\mathrm{E}(\Delta \mid W) = \lambda(W + R)\,,$$

for some constant $\lambda \in (0, 1)$ and random variable R, where $\Delta = W - W'$. Then

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le \mathrm{E}\Big|1 - \frac{1}{2\lambda}\mathrm{E}(\Delta^2 \mid W)\Big| + \mathrm{E}|R| + \frac{1}{\lambda}\mathrm{E}|\mathrm{E}(\Delta\Delta^* \mid W)|\,,$$

where $\Delta^* := \Delta^*(W, W')$ is any random variable satisfying $\Delta^*(W, W') = \Delta^*(W', W)$ and $\Delta^* \ge |\Delta|$.

Proof. Note that

$$\begin{aligned} 0 &= E[(W - W')(f(W) + f(W'))] \\ &= 2\lambda\left[E(Wf(W)) + E(Rf(W)) - E\int_{-\infty}^{\infty} f'(W + t)\hat{K}(t)dt\right]\,, \end{aligned}$$

where

$$\hat{K}(t) = \frac{\Delta(\mathbb{1}(-\Delta \le t \le 0) - \mathbb{1}(0 \le t \le -\Delta))}{2\lambda}$$

and $\int \hat{K}(t) = \Delta^2/2\lambda$.

Let $f'(w) - wf(w) = \mathbb{1}(w \leq z) - \Phi(z)$, then

$$\begin{aligned}
P(W \leq z) - \Phi(z) &= Ef'(W) - EWf(W) \\
&= Ef'(W) - E\int f'(W+t)\hat{K}(t)dt + E[Rf(W)] \\
&= Ef'(W) - E\int [f'(W+t) - f'(W) + f'(W)]\hat{K}(t)dt + E[Rf(W)] \\
&= E\left[f'(W)\left(1 - \frac{\Delta^2}{2\lambda}\right)\right] - E\int (f'(W+t) - f'(W))\hat{K}(t)dt + E[Rf(W)]\,,
\end{aligned}$$

where the first term

$$E\left[f'(W)\left(1 - \frac{\Delta^2}{2\lambda}\right)\right] = E\left[f'(W)\left(1 - \frac{E(\Delta^2 \mid W)}{2\lambda}\right)\right] \leq E\left|1 - \frac{E(\Delta^2 \mid W)}{2\lambda}\right|\,,$$

and the third term

$$E[Rf(W)] \leq E|R|$$

due to $|f(W)| \leq 1$, then the main part is the second term. Note that

$$\begin{aligned}
&E\int (f'(W+t) - f'(W))\hat{K}(t)dt \\
&= E\int [(W+t)f(W+t) - Wf(W)]\hat{K}(t)dt + E\int [\mathbb{1}(W+t \leq z) - \mathbb{1}(W \leq z)]\hat{K}(t)dt\,,
\end{aligned}$$

where the second term

$$\begin{aligned}
&E\int (\mathbb{1}(W+t \leq z) - \mathbb{1}(W \leq z)\hat{K}(t))dt \\
&= \frac{1}{2\lambda}E\int \Delta \int_{-\Delta}^{0} [\mathbb{1}(W+t \leq z) - \mathbb{1}(W \leq z)]dt \\
&\leq \frac{1}{2\lambda}E\left[|\Delta|\Delta(\mathbb{1}(W' \leq z) - \mathbb{1}(W \leq z))\right] \\
&= \frac{1}{2\lambda}\left\{E[|\Delta|(-\Delta)\mathbb{1}(W \leq z)] - E[|\Delta|\Delta\mathbb{1}(W \leq z)]\right\} \\
&= -\frac{1}{\lambda}E[|\Delta|\Delta\mathbb{1}(W \leq z)] \\
&= -\frac{1}{\lambda}E[\mathbb{1}(W \leq z)E(|\Delta|\Delta \mid W)]
\end{aligned}$$

□

4 Non-normal Approximation

Let Y have density p, and letting f be an absolutely continuous function satisfied $f(a+) = f(b-) = 0$, then we have

$$\begin{aligned} E[f'(Y) + f(Y)p'(Y)/p(Y)] &= E[(f(Y)p(Y))'/p(Y)] \\ &= \int_a^b (f(y)p(y))'dy \\ &= f(b-)p(b-) - f(a+)p(a+) = 0\,. \end{aligned}$$

For any measurable function h with $E|h(Y)| < \infty$, let $f = f_h$ be the solution to the Stein equation

$$f'(w) + f(w)p'(w)/p(w) = h(w) - Eh(Y)\,,$$

which can be rewritten as

$$(f(y)p(y))' = (h(w) - Eh(Y))p(w)\,,$$

it follows that

$$\begin{aligned} f(y) &= \frac{1}{p(y)} \int_{-\infty}^{y} p(t)(h(t) - Eh(Y))dt \\ &= -\frac{1}{p(y)} \int_{y}^{\infty} p(t)(h(t) - Eh(Y))dt\,. \end{aligned}$$

Lemma 1 (Properties of the Stein Solution). *Under certain conditions:*

- $\|f\| \le C_1\|h\|$, $\|f'\| \le C_2\|h\|$
- $\|f\| \le C_3\|h'\|$, $\|f'\| \le C_4\|h'\|$, $\|f''\| \le C_5\|h'\|$.

Theorem 13. *Let (W, W') be exchangeable pair, assume that*

(i) $E(W - W' \mid W) = \lambda(g(W) + R(W))$*, actually it is not a condition, but always exists.*

(ii) $\dfrac{E(\Delta^2 \mid W)}{2\lambda} \xrightarrow{p} 1$*, where* $\Delta = W - W'$*.*

(iii) $\dfrac{E|\Delta|^3}{\lambda} \to 0$ *and* $E|R| \to 0$*.*

then $W \xrightarrow{d} Y$*, where* Y *has pdf* $p(y) = c_1 e^{-G(y)}$*, where* $G(y) = \int_0^y g(t)dt$*.*

Proof. Note that

$$\begin{aligned} 0 &= E[(W - W')(f(W') + f(W))] \\ &= 2Ef(W)(W - W') + E(W - W')(f(W') - f(W)) \\ &= 2E[f(W)E(W - W' \mid W)] - E\Delta \int_{-\Delta}^{0} f'(W + t)dt \\ &= 2\lambda \left[Ef(W)g(W) + f(W)R(W) - \frac{1}{2\lambda} E\left(\Delta \int_{-\infty}^{\infty} f'(W + t)[\mathbb{1}(-\Delta \le t \le 0) - \mathbb{1}(0 \le t \le -\Delta)]dt \right) \right] \\ &= 2\lambda \left[Ef(W)g(W) + f(W)R(W) - \frac{1}{2\lambda} E \int_{-\infty}^{\infty} f'(W + t)\hat{K}(t)dt \right]\,, \end{aligned}$$

where
$$\hat{K}(t) = E[\Delta(\mathbb{1}(-\Delta \le t \le 0) - \mathbb{1}(0 \le t \le -|Delta)) \mid W],$$
and $\int_{-\infty}^{\infty} \hat{K}(t)dt = E(\Delta^2 \mid W)$. By comparing, we should have
$$\frac{p'(y)}{p(y)} = -g(y),$$
then
$$(\log p(y))' = -g(y),$$
and it follows that
$$p(y) = c\exp(-G(y)).$$

□

4.1 Curie-Weiss Model

The Curie-Weiss model is a simple statistical mechanical model of ferromagnetic interaction, where for $n \in \mathbb{N}$, a vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ of "spin" in $\{-1, 1\}^n$ has joint probability mass function
$$p(\boldsymbol{\sigma}) = C_\beta \exp\left(\frac{\beta}{n}\sum_{i<j}\sigma_i\sigma_j\right),$$
where C_β is a normalizing constant and $\beta > 0$ is known as the inverse temperature.

Theorem 14. *The limiting distribution of $\sum_{i=1}^n \sigma_i$ is*

- *If $0 < \beta < 1$,*
$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i \xrightarrow{d} N(0, \frac{1}{1-\beta})$$

- *If $\beta = 1$,*
$$\frac{1}{n^{3/4}}\sum_{i=1}^n \sigma_i \xrightarrow{d} Y,$$
where Y has pdf $f(y) = c\exp(-y^4/12)$.

Theorem 15. *Let $W = \frac{1}{n^{3/4}}\sum_{i=1}^n \sigma_i$, and*
$$W' = W - \frac{1}{n^{3/4}}\sigma_I + \frac{1}{n^{3/4}}\sigma'_I,$$
where $\sigma'_i \mid \sigma_j, j \ne i \overset{d}{\sim} \sigma_i \mid \sigma_j, j \ne i$, then

- $|W - W'| \le 2n^{-3/4}$
- $E(W - W' \mid W) = \frac{1}{3}n^{-3/2}(W^3 + \frac{O(1)}{\sqrt{n}})$
- $E\left|1 - \frac{n^{3/2}}{2}E(\Delta^2 \mid W)\right| \le 8n^{-1/2}$

4.2 Poisson Approximation

Let $W \sim \mathcal{P}(\lambda)$, the Stein's identify is

$$EWf(W) = \lambda Ef(W+1)\,,$$

and the Stein's equation is

$$\lambda f(w+1) - wf(w) = h(w) - Eh(Y)\,,$$

where $Y \sim \mathcal{P}(\lambda)$ and $w = 0, 1, 2, \ldots$.

Theorem 16 (Possion Convergence for Independent Random Variables). *For each n let $X_{n,m}, 1 \le m \le n$ be independent random variables with $P(X_{n,m} = 1) = p_{n,m}$, $P(X_{n,m} = 0) = 1 - p_{n,m}$. Suppose*

(i) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0, \infty)$

(ii) $\max_{1 \le m \le n} p_{n,m} \to 0.$

If $S_n = X_{n,1} + \ldots + X_{n,n}$, then $S_n \xrightarrow{d} Z$ where $Z \sim \mathcal{P}(\lambda)$.

But Chen-Stein method (Chen, 1975) can handle dependent cases, and one application of Chen-Stein method is

Theorem 17 (Arratia-Goldstein-Gordon, 1989). *Let I be an arbitrary index set, and for $\alpha \in I$, let ξ_α be a Bernoulli random variable with $p_\alpha = P(\xi_\alpha = 1) = 1 - P(\xi_\alpha = 0) > 0$. Let $W = \sum_{\alpha \in I} \xi_\alpha$ and $\lambda = \sum_{\alpha \in I} p_\alpha$, then*

$$|P(W \in A) - P(Y \in A)| \le 4(b_1 + b_2 + b_3)\,,$$

where $Y \sim \mathcal{P}(\lambda)$, and

$$\begin{aligned} b_1 &= \sum_{\alpha \in I} p_\alpha p_\beta \\ b_2 &= \sum_{\alpha \in I} \sum_{\alpha \ne \beta \in B_\alpha} E(\xi_\alpha \xi_\beta) \\ b_3 &= \sum_{\alpha \in I} E\left| E\left\{\xi_\alpha - p_\alpha \mid \sigma(\xi_\beta : \beta \notin B_\alpha)\right\}\right| \end{aligned}$$

A related open question is

Open Question 2. *Can you find a "sufficient" condition, or computable estimator for*

$$|P(W \in A) - P(Y \in A)|?$$

5 Large Deviation

Theorem 18 (Cramér-Chernoff large deviation theorem). *Let $X, X_1, \dots, X_n$ be i.i.d. random variables with $P(X \neq 0) > 0$ and let $S_n = \sum_{i=1}^n X_i$. If*

$$Ee^{\theta_0 X} < \infty \quad \text{for some } \theta_0 > 0$$

then for every $x > EX$,

$$\lim_{n\to\infty} n^{-1} \log P\left(\frac{S_n}{n} \geq x\right) = \log \rho(x)\,,$$

where $\rho(x) = \inf_{t\geq 0} e^{-tx}\mathrm{E}e^{tX}$.

Proof. Upper bound is easy. Note that

$$\begin{aligned} P\left(\frac{S_n}{n} \geq x\right) &= P\left(e^{tS_n} \geq e^{txn}\right) \\ &\leq \frac{1}{e^{txn}}\mathrm{E}e^{tS_n} \\ &= \left(e^{-tx}\mathrm{E}e^{tX_1}\right)^n\,, \end{aligned}$$

it follows that

$$P\left(\frac{S_n}{n} \geq x\right)^{1/n} \leq \inf_{t\geq 0} e^{-tx}\mathrm{E}e^{tX_1}\,.$$

For the lower bound, apply the conjugated method (change of measure).

Note 1 (Conjugated method, or change of measure). *Letting $e^{\psi(\theta)} = Ee^{\theta X}$, the basic idea is to embed P in a family of measures P_θ under which $X_1, X_2, \dots$ are i.i.d. with density function $f_\theta(x) = e^{\theta x - \psi(\theta)}$ with respect to P. Then for any event A,*

$$P(A) = \int_A \frac{dP}{dP_\theta} dP_\theta = E_\theta\left\{e^{-(\theta S_n - n\psi(\theta))} I(A)\right\}\,,$$

since the Radon-Nikodym derivative (or likelihood ratio) dP_θ/dP is equal to $\prod_{i=1}^n f_\theta(X_i) = e^{\theta S_n - n\psi(\theta)}$.

Theorem 19 (Radon-Nikodym Theorem). *Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda << \nu$, then there exists a nonnegative Borel function f on Ω such that*

$$\lambda(A) = \int_A f d\nu, \quad A \in \mathcal{F}\,.$$

Furthermore, f is unique a.e. ν, i.e., if $\lambda(A) = \int_A g d\nu$ for any $A \in \mathcal{F}$, then $f = g$ a.e. ν.

The family of density functions f_θ is an exponential family with the following properties:

$$E_\theta X = \psi'(\theta)\,, \qquad \mathrm{Var}_\theta(X) = \psi''(\theta)\,.$$

In particular, for $A = \{\bar{X}_n \geq x\}$ with $x > EX$, we choose $\theta = \theta_x$ such that $E_\theta X = x$, and therefore $x = \psi'(\theta)$. For this choice of θ, which is often called the "conjugate method",

$$\begin{aligned} & E_{\theta_x}\left\{e^{-n(\theta_x \bar{X}_n - \psi(\theta_x)) I(\bar{X}_n \geq x)}\right\} \\ & = e^{-n(\theta_x - \psi(\theta_x))} E_{\theta_x}\left\{e^{-n\theta_x(\bar{X}_n - x)} I(\bar{X}_n \geq x)\right\} \\ & = e^{-n\mathscr{I}(x)} E_{\theta_x}\left\{e^{-\sqrt{n}\theta_x(\sqrt{n}(\bar{X}_n - x))} I(\bar{X}_n \geq x)\right\}, \end{aligned}$$

where

$$\mathscr{I}(x) \triangleq \theta_x x - \psi(\theta_x) = \sup_\theta(\theta x - \psi(\theta)).$$

Introduce $\{Y_i\}$ independent,

$$P(Y_i \leq y) = \frac{\mathrm{E} e^{\lambda X_i} \mathbb{1}(X_i \leq y)}{\mathrm{E} e^{\lambda X_i}}.$$

Then

$$P(\sum X_i \geq y) = \left(\prod_{i=1}^n \mathrm{E} e^{\lambda X_i}\right) \mathrm{E}\left(e^{-\lambda \sum Y_i} \mathbb{1}(\sum Y_i \geq y)\right) \tag{6}$$

$$P((X_1, \ldots, X_n) \in A) = \left(\prod_{i=1}^n \mathrm{E} e^{\lambda X_i}\right) \mathrm{E}\left(e^{-\lambda \sum Y_i} \mathbb{1}((Y_1, \ldots, Y_n) \in A)\right). \tag{7}$$

Homework 4. *Verify* (6) *and* (7).

Note that

$$\begin{aligned} P(Y_i \leq y) = P_{Y_i}((-\infty, y]) &= \frac{E e^{\lambda X_i} \mathbb{1}(X_i \leq y)}{E e^{\lambda X_i}} \\ &= E e^{\lambda X_i - \psi_i(\lambda)} \mathbb{1}(X_i \leq y) \\ &= \int f_i(x) \mathbb{1}_{(-\infty, y]}(x) dP_{X_i} \end{aligned}$$

where $e^{\psi_i(\lambda)} = E e^{\lambda X_i}$, and $f_i(x) = e^{\lambda x - \psi_i(\lambda)}$ is the density function with respect to P_{X_i}. Then the Radon-Nikodym derivative is

$$\frac{dP_{Y_i}}{dP_{X_i}} = f_i(x),$$

then

$$\begin{aligned}
P(X_i \le y) &= \int \frac{dP_{X_i}}{dP_{Y_i}} \mathbb{1}_{(-\infty,y]}(x) dP_{Y_i} \\
&= \int e^{-(\lambda X_i - \psi_i(\lambda))} \mathbb{1}_{(-\infty,y]}(x) dP_{Y_i} \\
&= e^{\psi_i(\lambda)} \int e^{-\lambda x} \mathbb{1}_{(-\infty,y]}(x) dP_{Y_i} \\
&= Ee^{\lambda X_i} Ee^{-\lambda Y_i} \mathbb{1}(Y_i \le y) \,.
\end{aligned}$$

Since

$$\frac{d(P_{Y_1} \times P_{Y_2} \times \cdots \times P_{Y_n})}{d(P_{X_1} \times P_{X_2} \times \cdots \times P_{X_n})} = \prod_{i=1}^{n} \frac{dP_{Y_i}}{dP_{X_i}} = \prod_{i=1}^{n} f_i(x) \,,$$

then we have

$$\begin{aligned}
P((X_1, \ldots, X_n) \in A) &= \int_A \prod_{i=1}^{n} \frac{dP_{X_i}}{dP_{Y_i}} dP_{Y_1} \times \cdots \times dP_{Y_n} \\
&= \int_A e^{-\sum_{i=1}^{n}(\lambda y_i - \psi_i(\lambda))} dP_{Y_1} \times \cdots \times dP_{Y_n} \\
&= e^{\sum_{i=1}^{n} \psi_i(\lambda)} \int_A e^{-\lambda \sum_{i=1}^{n} y_i} dP_{Y_1} \times \cdots \times dP_{Y_n} \\
&= \left(\prod_{i=1}^{n} Ee^{\lambda X_i} \right) E\left(e^{-\lambda \sum Y_i} \mathbb{1}((Y_1, \ldots, Y_n) \in A) \right) \,,
\end{aligned}$$

and it follows that

$$P(\sum X_i \ge y) = \left(\prod_{i=1}^{n} Ee^{\lambda X_i} \right) E\left(e^{-\lambda \sum Y_i} \mathbb{1}(\sum Y_i \ge y) \right)$$

and

$$\begin{aligned}
\mathrm{E} g(Y_i) &= \frac{E[g(X_i) e^{\lambda X_i}]}{Ee^{\lambda X_i}} \\
\mathrm{E} Y_i &= \frac{\mathrm{E} X_i e^{\lambda X_i}}{\mathrm{E} e^{\lambda X_i}}
\end{aligned}$$

Then

$$\begin{aligned}
P\left(\frac{S_n}{n} \ge x \right) &= \left(\mathrm{E} e^{\lambda X_i} \right)^n Ee^{-\lambda \sum Y_i} \mathbb{1}(\sum Y_i / n \ge x) \\
&\ge \left(Ee^{\lambda X_1} \right)^n Ee^{-\lambda \sum Y_i} \mathbb{1}(x \le \sum Y_i / n \le x + \varepsilon) \\
&\ge \left(Ee^{\lambda X_1} \right)^n e^{-n\lambda(x+\varepsilon)} P\left(x \le \frac{\sum Y_i}{n} \le x + \varepsilon \right) \,,
\end{aligned}$$

then

$$\liminf P\left(\frac{S_n}{n} \geq x\right)^{1/n} \geq Ee^{\lambda X_1}e^{-\lambda(X+\varepsilon)} \geq \mathrm{E}e^{\lambda X_1}e^{-\lambda x} = \inf_{t\geq 0} e^{-tx}\mathrm{E}e^{tX_1}\,.$$

□

A random variable Y is said to be have a stable distribution if for every integer $n \geq 1$,

$$Y \stackrel{D}{=} \frac{X_{n1} + X_{n2} + \cdots + X_{nn} - \beta_n}{\alpha_n}\,,$$

where X_{ni} are i.i.d. and $\alpha_n > 0$ and β_n are constant.

6 Self-Normalized Large Deviation

Before considering the self-normalized sums, what if $P(S_n/V_n^2 \geq x)$? Note that

$$\begin{aligned} P\left(\frac{S_n}{V_n^2} \geq x\right) &= P(S_n - xV_n^2 \geq 0) \\ &= P(\sum_{i=1}^n (X_i - xX_i^2) \geq 0)\,, \end{aligned}$$

let $Y_i = X_i - xX_i^2$, which has a upper bound, and hence it has moment generating function, then by Cramér large deviation theorem, for any x satisfies $0 \geq E(X_1 - xX_1^2)$, i.e., $x > EX_1/EX_1^2$ and $x > 0$, we have

$$\frac{1}{n}\log P\left(\sum_{i=1}^n (X_i - xX_i)^2 \geq 0\right) \to \inf_{t\geq 0} Ee^{t(X_1 - xX_1^2)}$$

Theorem 20. *Assume that either $EX \geq 0$ or $EX^2 = \infty$. Let $V_n^2 = \sum_{i=1}^n X_i^2$. Then*

$$\lim_{n\to\infty} P(S_n \geq x\sqrt{n}V_n)^{1/n} = \sup_{c\geq 0}\inf_{t\geq 0} e^{-tx^2}Ee^{t(2cX_1 - (cX_1)^2)}$$

for $x > EX/(EX^2)^{1/2}$, where $EX/(EX^2)^{1/2} = 0$ if $EX^2 = \infty$.

Proof. Main idea of its proof: change V_n to V_n^2.

Since for any positive numbers x and y,

$$xy = x\sqrt{c}\frac{y}{\sqrt{c}} \leq \frac{1}{2}\left(x^2c + \frac{y^2}{c}\right) = \frac{1}{2}\frac{c^2x^2 + y^2}{c}\,,$$

that is,

$$xy = \frac{1}{2}\inf_{c>0}\left(\frac{c^2x^2 + y^2}{c}\right)\,.$$

Thus we can write

$$x\sqrt{n}V_n = \frac{1}{2}\inf_{c>0}\left(\frac{c^2V_n^2 + x^2n}{c}\right)\,.$$

It follows that

$$\begin{aligned} P(S_n \ge xV_n\sqrt{n}) &= P\left(S_n \ge \frac{1}{2}\inf_{c>0}\frac{c^2V_n^2 + x^2n}{c}\right) \\ &= P\left(\bigcup_{c>0}\left\{2cS_n \ge c^2V_n^2 + x^2n\right\}\right) \\ &= P\left(\bigcup_{c>0}\left\{\sum_{i=1}^n (2cX_i - c^2X_i^2) \ge x^2n\right\}\right) , \end{aligned}$$

thus,

$$\begin{aligned} P(S_n \ge xV_n\sqrt{n})^{\frac{1}{n}} &\ge \sup_{c>0} P\left(\frac{\sum_{i=1}^n (2cX_i - (cX_i)^2)}{n} \ge x^2\right)^{1/n} \\ &\to \sup_{c\ge 0}\inf_{t\ge 0} e^{-tx^2} Ee^{t(2cX_1 - (cX_1)^2)} \end{aligned}$$

□

7 Self-Normalized Moderate Deviation

If $EX_n = 0$, $EX_1^2\mathbb{1}(|X_1| \le x)$ slowly varying, then $\forall x_n \to \infty$, $x_n = o(\sqrt{n})$

$$\log P(\frac{S_n}{V_n} \ge x_n) \sim -\frac{x_n^2}{2} .$$

A function $L : (0, \infty) \to \mathbb{R}$ is said to be slowly varying (at ∞) if

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1 \qquad \text{for all } c \ge 0 .$$

Proof.

Step 1 (Proof of the Upper Bound). *Note that*

$$P\left(\frac{S_n}{V_n} \ge x_n\right) \le P\left(\frac{\sum X_{i,1}}{V_n} \ge (1-\varepsilon)x_n\right) + P\left(\frac{\sum X_{i,2}}{V_n} \ge \varepsilon x_n\right) , \tag{8}$$

where the second term

$$\begin{aligned} P\left(\frac{\sum X_{i,2}}{V_n} \ge \varepsilon x_n\right) &= P\left(\frac{\sum X_i\mathbb{1}(|X_i| > z_n)}{V_n} \ge \varepsilon x_n\right) \\ &\le P\left(\left[\sum \mathbb{1}(|X_i| > z_n)\right]^{1/2} \ge \varepsilon x_n\right) \\ &\le \left(\frac{3nP(|X_i| > z_n)}{\varepsilon^2 x_n^2}\right)^{\varepsilon^2 x_n^2} , \end{aligned}$$

where the first inequality is from the Cauchy-Schwarz inequality,

$$\sum_{i=1}^{n} a_i b_i \le \sqrt{\sum_{i=1}^{n} a_i}\sqrt{\sum_{i=1}^{n} b_i^2}\,,$$

and the second inequality is based on the following lemma.

Lemma 2. *If ε_i are independent, $P(\varepsilon_i = 1) = p_i$ and $P(\varepsilon_i = 0) = 1 - p_i$, then*

$$P\left(\sum_{i=1}^{n} \varepsilon_i \ge x\right) \le \left(\frac{3\sum_{i=1}^{n} p_i}{x}\right)^{x}.$$

Since

$$P(|X_1| \ge z_n) \le \frac{EX_1^2 \mathbb{1}(|X_1| \ge z_n)}{z_n^2}\,,$$

then

$$P(|X_1| \ge z_n) = o\left(\frac{1}{z_n^2}\right).$$

It follows that

$$\left(\frac{3nP(|X_i| > z_n)}{\varepsilon^2 x_n^2}\right)^{\varepsilon^2 x_n^2} = \left(\frac{o(1)n}{\varepsilon^2 x_n^2 z_n^2}\right)^{\varepsilon^2 x_n^2} \le e^{-2x_n^2}\,,$$

which requires the assumption that

$$\frac{n}{x_n^2 z_n^2} \le c\,. \tag{9}$$

For the first term of (8)*,*

$$\begin{aligned} P\left(\sum X_{i,1} \ge (1-\varepsilon)X_n V_n\right) &\le P\left(\sum X_{i,1} \ge (1-\varepsilon)x_n V_n, V_n^2 \ge (1-\varepsilon)n\right) + P(V_n^2 < (1-\varepsilon)n) \\ &\le P(\sum X_{i,1} \ge (1-\varepsilon)^{3/2} x_n \sqrt{n}) + P(V_n^2 \le n - \varepsilon n)\,, \end{aligned} \tag{10}$$

where the first term

$$P\left(\sum X_{i,1} \ge (1-\varepsilon)^{3/2} x_n \sqrt{n}\right) \le \exp\left(-\frac{x_n^2 n(1-\varepsilon)^3}{2(n + x_n\sqrt{n} z_n)}\right),$$

then we can choose z_n such that $x_n\sqrt{n}z_n \le \varepsilon n$ and (9)*. Consider the second term of* (10)*, by the inequality for $Y_i \ge 0$,*

$$P(\sum Y_i \le \sum EY_i - x) \le \exp\left(-\frac{x^2}{2\sum EY_i^2}\right),$$

then we have

$$\begin{aligned} P\left(\sum_{i=1}^n X_i^2 \le n - \varepsilon n\right) &\le P\left(\sum_{i=1}^n X_i^2 \mathbb{1}(|X_i| \le z_n) \le n - \varepsilon n\right) \\ &\le P\left(\sum_{i=1}^n X_i^2 \mathbb{1}(|X_i| \le z_n) \le \sum E X_i^2 \mathbb{1}(|X_i| \le z_n) - \frac{\varepsilon n}{2}\right) \\ &\le \exp\left(-\frac{\varepsilon^2 n^2}{8nEX_1^4 \mathbb{1}(|X_1| \le z_n)}\right) \\ &\le \exp(-2x_n^2) \quad \textit{since } EX_1^4 \mathbb{1}(|X_1| \ge x) = o(x^2)\,, \end{aligned}$$

Step 2 (Main Idea of Proof for the Lower Bound). *Note that*

$$\begin{aligned} P(S_n \ge x_n V_n) &\ge P\left(S_n \ge \frac{1}{2}\frac{b^2 V_n^2 + x_n^2}{b}\right) \\ &= P(bS_n - \frac{1}{2}b^2 V_n^2 \ge \frac{1}{2}x_n^2) \\ &= P\left(\sum (bX_i - \frac{1}{2}(bX_i)^2) \ge \frac{x_n^2}{2}\right)\,, \end{aligned}$$

let $\xi_i = bX_i - \frac{1}{2}(bX_i)^2$. *Apply change of measure technique, introduce independent random variables* η_i *such that*

$$P(\eta_i \le y) = \frac{Ee^{\lambda \xi_i} \mathbb{1}(\xi_i \le y)}{Ee^{\lambda \xi_i}}\,,$$

then

$$P\left(\sum_{i=1}^n \xi_i \ge \frac{x_n^2}{2}\right) = [Ee^{\lambda \xi_i}]^n E e^{-\lambda \sum \eta_i} \mathbb{1}(\sum \eta_i \ge \frac{x_n^2}{2})\,.$$

More details can be found in Chapter 6 of Peña, Lai, and Shao (2008).

□

Note 2. *difference between large deviation and moderate deviation*

- *classical limit theory: the probability events of the form* $\{T_n > a\}$ *for constant* a.
- *large deviation: study events of form* $\{T_n > a\sqrt{n}\}$.
- *moderate deviation: study events of form* $\{T_n > a_n\}$ *where* $a_n \to \infty$ *but* $a_n = o(\sqrt{n})$

8 Cramér-Type Moderate Deviations

Question: Is $1 - \Phi(x_n)$ close to $P(S_n/V_n \ge x_n)$? Or formally, is

$$\frac{P\left(\frac{S_n}{V_n} \ge x_n\right)}{1 - \Phi(x_n)} \to 1?$$

The following theorems provide some results.

Theorem 21 (Cramér, 1938). *If $X_1, \dots, X_n$ be i.i.d. random variables with $E(X_i) = 0, E(X_i^2) = \sigma^2$. If $Ee^{t_0\sqrt{X_1}} < \infty$ for some $t_0 > 0$, then*

$$\frac{P\left(\frac{S_n}{\sqrt{n}\sigma} \geq x\right)}{1-\Phi(x)} \to 1$$

as $n \to \infty$ uniformly in $x \in \big(0, o(n^{1/6})\big)$.

Theorem 22 (Shao, 1999). *If $EX_1 = 0, E|X_1|^3 < \infty$, then*

$$\frac{P\left(\frac{S_n}{V_n} \geq x\right)}{1-\Phi(x)} \to 1$$

uniformly for $x \in \big(0, o(n^{1/6})\big)$.

Proof.

Step 1. *show $P(S_n/V_n \geq x_n) \geq (1-\Phi(x_n))(1+o(1))$.*

Note that

$$x_n V_n \leq \frac{b^2V_n^2 + x_n^2}{2b},$$

where the equality can be achieved if $b = x_n/V_n$. The guideline of choosing b is to let it be a constant instead of random variable, and be close to x_n/V_n, thus let $b = x_n/\sqrt{n}$ (Here we assume $EX_1^2 = 1$). It follows that

$$\begin{aligned} P\left(\frac{S_n}{V_n} \geq x_n\right) &= P(S_n \geq x_n V_n) \\ &\geq P\left(S_n \geq \frac{b^2V_n^2 + x_n^2}{2b}\right) \\ &= P\left(\sum_{i=1}^n (bX_i - \frac{1}{2}(bX_i)^2) \geq \frac{1}{2}x_n^2\right) \\ &\triangleq P\left(\sum_{i=1}^n \xi_i \geq \frac{1}{2}x_n^2\right), \end{aligned}$$

where $\xi_i = bX_i - \frac{1}{2}(bX_i)^2$.

Apply conjugate method (change of measure), let η_i be independent random variable such that

$$P(\eta_i \leq y) = \frac{Ee^{\lambda\xi_i}\mathbb{1}(\xi_i \leq y)}{Ee^{\lambda\xi_i}}, \quad \lambda > 0,$$

then

$$P\left(\sum_{i=1}^n \xi_i \geq \frac{1}{2}x_n^2\right) = \prod_{i=1}^n Ee^{\lambda\xi_i} Ee^{-\lambda\sum_{i=1}^n \eta_i}\mathbb{1}(\sum_{i=1}^n \eta_i \geq \frac{1}{2}x_n^2).$$

Choose λ such that $\sum_{i=1}^n E\eta_i$ is equal to or close to $\frac{1}{2}x_n^2$, and hence set λ such that $E\eta_1$ is close to or equal to $\frac{1}{2n}x_n^2 = \frac{E\xi_1 e^{\lambda\xi_1}}{Ee^{\lambda\xi_i}}$.

It can be calculated that

$$\begin{aligned} Ee^{\lambda\xi_1} &= 1 + \frac{1}{2}\lambda(\lambda-1)b^2EX_1^2 + O(1)b^3E|X_1|^3 \\ E\xi_1 e^{\lambda\xi_1} &= (\lambda - \frac{1}{2})b^2EX_1^2 + O(1)b^3E|X_1|^3 \,. \end{aligned}$$

Thus, we can choose $\lambda = 1$, then $\mathrm{E}\eta_1$ is close to $\frac{1}{2}\frac{x_n^2}{n}$.

Back to

$$P(\sum_{i=1}^n \xi_i \geq \frac{1}{2}x_n^2) = \prod_{i=1}^n Ee^{\lambda\xi_i} Ee^{-\lambda\sum_{i=1}^n \eta_i}\mathbb{1}(\sum_{i=1}^n \eta_i \geq \frac{1}{2}x_n^2)\,,$$

let $W = \sum_{i=1}^n \eta_i$, then

$$\begin{aligned} & Ee^{-\lambda W}\mathbb{1}(W \geq \frac{1}{2}x_n^2) \\ =& Ee^{-\lambda(W-EW)}e^{-\lambda EW}\mathbb{1}(W - EW \geq \frac{1}{2}x_n^2 - EW) \\ =& e^{-\lambda EW}Ee^{-\lambda(W-EW)}\mathbb{1}(W - EW \geq y_n) \\ =& e^{-\lambda EW}Ee^{-\lambda\sqrt{\mathrm{Var}(W)}\frac{W-EW}{\sqrt{\mathrm{Var}(W)}}}\mathbb{1}\left(\frac{W-EW}{\sqrt{\mathrm{Var}(W)}} \geq \frac{y_n}{\sqrt{\mathrm{Var}(W)}}\right)\,. \end{aligned}$$

Claim that

$$|Ee^{-\lambda^* W^*}\mathbb{1}(W^* \geq y) - Ee^{-\lambda^* Z}\mathbb{1}(Z \geq y)| \leq e^{-\lambda^*}\sup_z |P(W^* \geq z) - (1 - \Phi(z))|\,.$$

Adopting the technique of changing expectation to expectation,

$$Eg(X) = g(0) + E\int_0^x g'(t)dt = g(0) + E\int_0^\infty g'(t)\mathbb{1}(t \leq X)dt\,,$$

we have

$$\begin{aligned} Ee^{-\lambda^* W^*}\mathbb{1}(W^* \geq y) &= E\left[\left(\lambda^*\int_{W^*}^\infty e^{-\lambda^* t}dt\right)\mathbb{1}(W^* \geq y)\right] \\ &= \lambda^* E\int_{-\infty}^\infty e^{-\lambda^* t}\mathbb{1}(t \geq W^*)\mathbb{1}(W^* \geq y)dt \\ &= \lambda^* E\int_y^\infty e^{-\lambda^* t}P(W^* \geq y, W^* \leq t)dt\,, \end{aligned}$$

similarly,

$$Ee^{-\lambda^* Z}\mathbb{1}(Z \geq y) = \lambda^*\int_y^\infty P(Z \geq y, Z \leq t)dt\,.$$

Step 2. *show* $P(S_n \geq x_n V_n) \leq (1-\Phi(x))(1+o(1))$.

Inspired by the fact of the tail probability of Normal distribution, then

$$\{S_n \geq x_n V_n\} \subset \left\{ S_n \geq \frac{1}{2} \frac{b^2 V_n^2 + x_n^2 - \varepsilon^2}{b} \right\} \cup \left\{ S_n \geq x_n V_n, S_n < \frac{b^2 V_n^2 + x_n^2 - \varepsilon^2}{2b} \right\} .$$

To show the second term is smaller than the first term, note that

$$\left\{ S_n \geq x_n V_n, S_n < \frac{b^2 V_n^2 + x_n^2 - \varepsilon^2}{2b} \right\} \subset \left\{ S_n \geq x_n V_n, x_n V_n < \frac{b^2 V_n^2 + x_n^2 - \varepsilon^2}{2b} \right\} ,$$

where

$$\begin{aligned} & x_n V_n < \frac{b^2 V_n^2 + x_n^2 - \varepsilon^2}{2b} \\ \Rightarrow & b^2 V_n^2 + x_n^2 - 2b x_n V_n \geq \varepsilon^2 \\ \Rightarrow & (bV_n - x_n)^2 \geq \varepsilon^2 \\ \Rightarrow & |bV_n - x_n| \geq \varepsilon \\ \Rightarrow & b^2 V_n^2 - x_n^2 \geq \varepsilon(bV_n + x_n) \quad \textit{or} \quad b^2 V_n^2 - x_n^2 \leq -\varepsilon(bV_n + x_n) \end{aligned}$$

We want to calculate

$$\begin{aligned} P(S_n \geq x V_n, b^2 V_n^2 \geq x_n^2 + \varepsilon x_n) &= P(bS_n \geq x(b^2 V_n^2)^{1/2}, b^2 V_n^2 \geq x_n^2 + \varepsilon x_n) \\ &= P((bS_n, b^2 V_n^2) \in A) , \end{aligned}$$

where $A = \{(s,t) : s \geq x\sqrt{t}, t \geq x_n^2 + \varepsilon x_n\}$. *Apply Chebyshev's inequality, we have*

$$P((bS_n, b^2 V_n^2) \in A) \leq E \exp(\lambda_1 b S_n - \lambda_2 b^2 V_n^2) e^{-\inf_{(s,t)\in A}(\lambda_1 s - \lambda_2 t)} = -\infty .$$

The bound is useless, so we need some modifications.

$$\{b^2 V_n^2 - x_n^2 \geq \varepsilon x_n\} \subset \{b^2 V_n^2 - x_n^2 \geq 2x_n^2\} \cup \{2x_n^2 \geq b^2 V_n^2 - x_n^2 \geq \varepsilon x_n\} ,$$

then

$$\inf_{s \geq x\sqrt{t}, 2x_n^2 \geq t \geq x_n^2 + \varepsilon x_n} (\lambda_1 s - \lambda_2 t) = \inf_{2x_n^2 \geq t \geq x_n^2 + \varepsilon x_n} (\lambda x \sqrt{t} - \lambda_2 t)$$

□

Prof. Shao shared a more general result

Theorem 23 (Jing-Shao-Wang (2003)).

$$\frac{P\left(S_n / V_n \geq x_n\right)}{1 - \Phi(x_n)} = 1 + O(1) \frac{(1 + x_n^3) \sum_{i=1}^n E|X_i|^3}{B_n^3}$$

for $0 \leq x \leq \frac{B_n}{(\sum_{i=1}^n E|X_i|^3)^{1/3}}$, *where* $B_n = (ES_n^2)^{1/2}$ *and* $|O(1)| \leq C$.

Main idea behind its proof. Note that

$$x_n V_n \leq \frac{1}{2}\frac{b^2V_n^2 + x_n^2}{b}\,,$$

where the equality can be achieved if $b = x_n/V_n$. Since $B_n^2 \sim V_n^2$, choose $b = x_n/B_n$.

Lower bound is OK due to

$$P\left(S_n \geq \frac{1}{2}\frac{b^2V_n^2 + x_n^2}{b}\right) = P(\sum_{i=1}^n (bX_i - \frac{1}{2}(bX_i)^2) \geq \frac{x_n^2}{2})\,.$$

Consider the upper bound,

$$P(S_n \geq xV_n) = P(S_n \geq xV_n, \max|X_i| \leq a) + P(S_n \geq xV_n, \max|X_i| > a)\,,$$

where the second term is bound by $\sum_{i=1}^n P(S_n \geq x_nV_n, |X_i| > a)$. Note that

$$\{S_n \geq x_nV_n, |X_i| > a\} \subset \left\{\frac{\sum_{j\neq i} X_j}{\sqrt{\sum_{j\neq i} X_j^2}} \geq (x_n^2 - 1)^{1/2}, |X_i| > a\right\}\,,$$

then

$$P(S_n \geq x_nV_n, |X_i| > a) \leq P\left(\frac{\sum_{j\neq i} X_j}{\sqrt{\sum_{j\neq i} X_j^2}} \geq \sqrt{x_n^2 - 1}\right) P(|X_i| > a)\,.$$

□

Theorem 24 (Shao and Zhou, 2016). *If $\{\xi_i\}$ are independent, $E\xi_i = 0$ and $\sum_{i=1}^n E\xi_i^2 = 1$, then*

$$\frac{P\left(\frac{\sum_{i=1}^n \xi_i + D_1}{\sqrt{\sum_{i=1}^n \xi_i^2(1+D_2)}} \geq x_n\right)}{1 - \Phi(x_n)} \to 1\,.$$

Note 3. *Prof. Shao also shared two suggestions about research with us:*

- *As long as your results are good, do not care much about the journal.*
- *The author order would be alphabetical in some fields, especially in the probability, so do not care much about the order.*

9 Self-Normalized CLT

Suppose $X_1, \ldots, X_n$ are independent, $EX_i = 0$, let $S_n = \sum_{i=1}^n X_i$ and $V_n^2 = \sum_{i=1}^n X_i^2$. The classical CLT said that

$$\frac{S_n - a_n}{b_n} \xrightarrow{d} N(0,1)\,.$$

The question is:

$$\frac{S_n}{V_n} \xrightarrow{d} N(0,1)\,?$$

Consider two special cases:

- If $X_1, \ldots, X_n$ are i.i.d., then

$$\begin{aligned}\frac{S_n}{V_n} \xrightarrow{d} N(0,1) &\Longleftrightarrow EX_1^2 \mathbb{1}(|X_1| \le x) \text{ is slowly varying} \\ &\Longleftrightarrow \frac{\max_{1\le i\le n} |X_i|}{V_n} \xrightarrow{p} 0\,.\end{aligned}$$

- If $X_1, \ldots, X_n$ are independent but symmetric, then

$$\frac{S_n}{V_n} \xrightarrow{d} \Longleftrightarrow \frac{\max_{1\le i\le n} |X_i|}{V_n} \xrightarrow{p} 0\,.$$

But $\max_{1\le i\le n} |X_i|/V_n \xrightarrow{p} 0$ is not a sufficient for general case.

Theorem 25 (Shao, 2018). *If*

(i) $\max \frac{|X_i|}{V_n} \xrightarrow{p} 0$

(ii) $\sum_{i=1}^{n} \left(E\left(\frac{X_i}{V_n}\right)^2 \right) \to 0$

(iii) $\sum_{i=1}^{n} \left(\frac{S_n}{\max(V_n, a_n)}\right) \to 0$ *where* $\sum_{i=1}^{n} E\left(\frac{X_i^2}{X_i^2 + a_n^2}\right) = 1$*,*

then

$$\frac{S_n}{V_n} \xrightarrow{d} N(0,1)\,.$$

If (i) is satisfied, then (ii) and (iii) are necessary for self-normalized CLT.

Open Question 3 (Conjecture). *(i)(ii)(iii) are necessary and sufficient for the self-normalized CLT.*