STAT6060: Steins Method – Final Report

Zhiling GU *

July 26, 2019

Contents

In this report, important conditions are colored red, comments and tricks are colored blue.

*Department of Statistics, The Chinese University of Hong Kong

This report is devoted to summarize the lectures of STAT 6060 and recent works of Professor Qi-Man Shao as well as related chapters of the book '*Self-Normalized Processes: Limit Theory and Statistical Applications, Probablity and Its Applications*' [7] which serves as the textbook of this course.

1 Review

1.1 Statistics

1.1.1 M-Statistics

M-estimators are a broad class of extremum estimators for which the objective function is a sample average. In 1964, Peter J. Huber proposed generalizing maximum likelihood estimation to the minimization of

$$\sum_{i=1}^{n} \rho(x_i, \theta)$$

The solutions

$$\hat{\theta} = \arg\min_{\theta} \left(\sum_{i=1}^{n} \rho(x_i, \theta) \right)$$

are called M-estimators.

1.1.2 U-Statistics

Let $X, X_1, \ldots, X_n$ be i.i.d. random variables, and let $h(x_1, x_2)$ be a real-valued sym- metric Borel measurable function such that $Eh(X_1, X_2) = \theta$. An unbiased estimate of θ is the U -statistic

$$U_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h(X_i, X_j)$$

The kernel of a U -statistic of order m is a function of m variables,

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

1.2 Widely Used Inequalities

1.2.1 Chebyshev's Inequalities

1. $\forall x \ge 0, P(X \ge x) \le E|X|/x$
2. $P(X \ge x) = P(e^{tX} \ge e^{tx}) \le E(e^{tX})/e^{tX}$
   Useful for exponential inequalities for random variables
3. $P((X, Y) \in A) = P(X \in A^1, Y \in A^2) = E(\mathbf{1}(X \in A^1, Y \in A^2)) \le E(e^{sX+tY-\inf_{(x,y)\in A}(sx+ty)})$

1.2.2 Hölder's Inequality

$E(XY) \le (E(|X|^p)^{1/p}(E(|Y|^q))^{1/q}$, where $p > 1, 1/p + 1/q = 1$

1.2.3 Lyapunov Inequality

$E(|X|^s) \le (E(|X|^r))^{\frac{t-s}{t-r}}(E(|X|^t))^{\frac{s-r}{t-r}}$, where $0 \le r \le s \le t$
Useful to produce a lower bound using higher moments

1.2.4 Kinball's Inequality

1. If $g(x)$ and $h(x)$ have the same Monotonicity,
   then $E(g(x)h(x)) \ge E(g(x))E(h(x))$
2. If $g(x)$ and $h(x)$ have different Monotonicities,
   then $E(g(x)h(x)) \le E(g(x))E(h(x))$
   Proof by introducing $Y \overset{d}{=} X, Y \perp\!\!\!\perp X$

1.3 Tail Probabilities of Partial Sums

Theorem 1.1 (2.17, Benette Hoeffding's Inequalities)**.** Assume $EX_i \le 0, X_i \le a (a > 0)$ for each $1 \le i \le n$, $\sum_{i=1}^n EX_i^2 \le B_n^2$, then

$$P(S_n \ge x) \le \exp\left(-\frac{x^2}{2(B_n^2 + ax)}\right) \quad \text{for } x > 0$$

$$Ee^{tS_n} \le \exp\left(a^{-2}\left(e^{ta} - 1 - ta\right)B_n^2\right) \quad \text{for } t > 0$$

$$P(S_n \ge x) \le \exp\left(-\frac{B_n^2}{a^2}\left\{\left(1 + \frac{ax}{B_n^2}\right)\log\left(1 + \frac{ax}{B_n^2}\right) - \frac{ax}{B_n^2}\right\}\right)$$

Uniform boundedness on the random variable and its second moment enables us to bound the tail probability and mgf from above.

Proof. 1. Consider standard normal random variables. Bound from above and below: (i) $P(Z \ge x) \le \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}$ (Chebyshev). Note that the tail probability is interesting when x is large, therefore ignore $1/x \to 1$. (ii)$P(Z \ge x) \ge \frac{1}{\sqrt{2\pi}}(\frac{1}{x} - \frac{1}{x^2})e^{-x^2/2}$ (???). Therefore $P(Z \ge x) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}$ as $x \to \infty$. Since $ax << B_n^2$, we have $P(S_n \ge x) = P(\frac{S_n}{B_n} \ge \frac{x}{B_n}) \le exp\left(-\frac{x^2}{2(B_n^2+ax)}\right)$, which is always referred to as Bernstein's Inequality.

2. Expand e^s to the second order term using Taylor expansion and find the constant $C_a = \sup_{s \le a}[\frac{e^s-(1+s)}{s^2/2}] = \frac{e^a-(1+a)}{a^2/2}$ so that $e^s \le 1 + s + C_a s^2/2$ for $s \le a$. Inequality follows by assigning $s = tX_i$ and $1 + x \le e^x$.

□

Theorem 1.2 (2.18)**.** Assume that $EX_i \le 0$ for $1 \le i \le n$ and that $\sum_{i=1}^n EX_i^2 \le B_n^2$. Then

$$P(S_n \ge x) \le P\left(\max_{1 \le i \le n} X_i \ge b\right) + \exp\left(-\frac{B_n^2}{a^2}\left\{\left(1 + \frac{ax}{B_n^2}\right)\log\left(1 + \frac{ax}{B_n^2}\right) - \frac{ax}{B_n^2}\right\}\right)$$
$$+ \sum_{i=1}^n P(a < X_i < b)\, P(S_n - X_i > x - b)$$

for $x > 0$ and $b \ge a > 0$. In particular,

$$P(S_n \ge x) \le P\left(\max_{1 \le i \le n} X_i > \delta x\right) + \left(\frac{3B_n^2}{B_n^2 + \delta x^2}\right)^{1/\delta}$$

for $x > 0$ and $\delta > 0$.

When X_i is not bounded above, use Truncation and Theorem 2.17.

Proof. Construct a one-sided truncated random variable $Y_i = X_i 1(X_i \leq xB_n/p)$. Notice (i) $E(Y_i) = E(X_i)1(X_i \geq xB_n/p) \leq 0$, (ii) $Y_i \leq xB_n/p$. Therefore,

$$\begin{aligned} P(S_n \geq xB_n) &\leq P(\max X_i \geq xB_n/p) + P(S_n \geq xB_n, \max X_i < xB_n/p) \\ &= P(\cdot) + P(\sum Y_i \geq xB_n) \end{aligned}$$

Apply the third inequality of Theorem 2.17 by letting $\tilde{a} = xB_n/p$ and $\tilde{x} = xB_n$. □

Theorem 1.3. Assume $EX_i = 0$, then

$$P(|S_n| \geq xB_n) \leq \sum P(|X_i| \geq xB_n/p) + 2\left(\frac{3p}{p+x^2}\right)$$

Theorem 1.4 (Rosenthal's Inequality)**.** Let $p \geq 2$ and let $X_1, \ldots, X_n$ be independent random variables with $EX_i = 0$ and $E|X_i|^p < \infty$ for $1 \leq i \leq n$. Then there exists a constant A_p, B_p depending only on p such that

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

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

Proof.

$$\begin{aligned} g(x) =& g(0) + \int_0^x g'(t)dt \\ =& g(0) + \int_0^\infty g'(t)1(t \leq x)dt \\ Eg(X) =& g(0) + \int_0^\infty g'(t)P(X \leq t)dt \end{aligned}$$

When x is negative, change integral to be fixed by introducing indicator function.

$$\int_0^x g'(t)dt = \int_\infty^\infty g'(t)(1(0 < t < x) - 1(x < t < 0))dt$$

□

Theorem 1.5. Assume X_i is Martingale Difference, i.e. $E(X_i|\mathbb{F}_{i-1}) = 0$, $S_n = \sum_{i=1}^n X_i$, then

$$E|S_n|^p \leq C_p\left(\sum E|X_i|^p + E(\sum E(X_i^2|\mathbb{F}_{i-1})^{p/2})\right) \tag{1}$$

$$E|S_n|^p \leq C_p^* E\left(\sum (X_i^2)^{p/2})\right) \quad \text{Lower bound also exists} \tag{2}$$

Theorem 1.6 (2.19)**.** Assume X_i is Non-negative and Independent,i.e. $X_i \geq 0$, $\mu_n = \sum EX_i$ and $B_n^2 = \sum EX_i^2$, then for any $0 < x < \mu_n$

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

Prove markov inequality

Theorem 1.7 (2.20)**.** Assume X_i is 0-1 Random Variable, $P(X_i = 1) = p_i$, $P(X_i = 0) = 1 - p_i$. Then for $x > 0$

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

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

Prove by minimizing mgf using markov inequality

1.4 Self-Normalized Sums & Symmetrization

Theorem 1.8 (2.14)**.** Assume that ε_i are independent, $P\left(\varepsilon_i = 1\right) = P\left(\varepsilon_i = -1\right) = 1/2$. Then

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

for $x > 0$ and real numbers $\{a_i\}$. $\{\varepsilon_i\}$ are called Rademacher random variables.

Prove using inequality $\frac{1}{2}\left(e^{-t} + e^t\right) \leq e^{t^2/2}$

Definition 1.1 (Symmetric Random Variable X)**.** $X \stackrel{d}{=} -X$

Theorem 1.9 (2.15)**.** If X_i is symmetric, then for $x > 0$

$$P\left(S_n \geq x V_n\right) \leq e^{-x^2/2}$$

where $S_n = \sum_{i=1}^n X_i$, $V_n^2 = \sum_{i=1}^n X_i^2$. And ε_i are i.i.d. Rademacher random variables independent of X_i. Here $\frac{S_n}{V_n}$ is called self-normalized or studentiazed. Remark: 70% of practical statistics are studentized.

$X_i \stackrel{d}{=} X_i \varepsilon_i$. Prove by letting $a_i = X_i$ in Theorem 2.14.

Theorem 1.10. If X_i is not symmetric, but $\{\varepsilon_i\}$ independent, $\{X_i\}$ independent,

$$\{X_i, 1 \leq i \leq n\} \stackrel{d}{=} \{X_i \varepsilon_i, 1 \leq i \leq n\}$$

we still have

$$P\left(S_n \geq x V_n\right) \leq e^{-x^2/2}$$

Proof. Prove by writing distribution function into conditional expectation.

$$\begin{aligned}
P\left(\frac{\sum X_i}{\sqrt{\sum X_i^2}} \geq x\right) &= P\left(\frac{\sum X_i \varepsilon_i}{\sqrt{\sum \left(X_i \varepsilon_i\right)^2}} \geq x\right) \text{ equivalent in distribution} \\
&= P\left(\frac{\sum X_i \varepsilon_i}{\sqrt{\sum \left(X_i\right)^2}} \geq x\right) \ \varepsilon^2 \equiv 1 \\
&= \mathrm{E}\left[P\left(\frac{\sum X_i \varepsilon_i}{\sqrt{\sum \left(X_i \varepsilon_i\right)^2}} \geq x\right) \,\middle|\, X_1, \ldots, X_n\right] \\
&= \mathrm{E}\left[P\left(\frac{\sum x_i \varepsilon_i}{\sqrt{\sum \left(x_i \varepsilon_i\right)^2}} \geq x\right) \,\middle|\, X_1 = x_1, \ldots, X_n = x_n\right] \\
&\leq \mathrm{E}\left[e^{-x^2/2} \,\middle|\, X_1 = x_1, \ldots, X_n = x_n\right] \text{ by Thm 2.14} \\
&= e^{-x^2/2}
\end{aligned}$$

□

Theorem 1.11. If $\{X_i\}$ independent, $EX_i = 0$, $\sum EX_i^2 \le B_n^2$ Then

$$P(S_n \ge x(V_n + 4B_n)) \le 2e^{-x^2/2}$$

Proof. Prove by transforming the original random variable into a symmetric one, process called 'symmetrization'.

Observe $\{X_i\} \overset{d}{=} \{Y_i\}$, then $\{X_i - Y_i\}$ is symmetric. Therefore we only need to prove

$$\left\{\sum X_i \ge x(V_n + 4B_n)\right\} \subset \left\{\sum (X_i - Y_i) \ge x\sqrt{\sum (X_i - Y_i)^2}\right\}$$

which can be separated into two steps:

1. $\{\sum X_i \ge x(V_n + 4B_n), |\sum Y_i| \le C_n\} \subset \{\sum (X_i - Y_i) \ge xV_m - C_n\}$
2. $\{\sum (X_i - Y_i) \ge xV_m - C_n\} \subset \left\{\sum (X_i - Y_i) \ge x\sqrt{\sum (X_i - Y_i)^2}\right\}$

Observe that $\sqrt{\sum (X_i - Y_i)^2} \le \sqrt{\sum X_i^2} + \sqrt{\sum Y_i^2}$, control the second term $\sqrt{\sum Y_i^2} \le D_n^2$, we have

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

Therefore when $x \ge 1$

$$\begin{aligned}
&\left\{\sum X_i \ge x(V_n + D_n + C_n), \left|\sum Y_i\right| \le C_n, \sqrt{\sum Y_i^2} \le D_n^2\right\} \\
\subset &\left\{\sum (X_i - Y_i) \ge x\sqrt{\sum (X_i - Y_i)^2}\right\} \\
&P\left(\sum X_i \ge x(V_n + D_n + C_n)\right) P\left(\left|\sum Y_i\right| \le C_n, \sum Y_i^2 \le D_n^2\right) \\
\le &P\left(\sum (X_i - Y_i) \ge x\sqrt{\sum (X_i - Y_i)^2}\right) \le e^{-x^2/2}
\end{aligned}$$

Then it only remains to control $P\left(|\sum Y_i| \le C_n, \sum Y_i^2 \le D_n^2\right)$. Note that $P(AB) \ge 1 - P(A^C) - P(B^C)$. On the other hand, by markov inequality we want $P\left(|\sum Y_i| \ge C_n\right) \le \frac{E(|\sum Y_i|)^2}{C_n^2} = \frac{1}{4}$, $P\left(|\sum Y_i^2| \ge D_n^2\right) \le \frac{E(\sum Y_i^2)}{D_n^2} = \frac{1}{4}$ where B_n in the theorem is set by $C_n = 2B_n$, $D_n = 2B_n$.

□

The theorem above is a special case of Lemma 3.1 [6] as following:

Lemma 1.12.

$$P\left(\max_{1 \le k \le n} \sum_{i=1}^{k} \|X_i\| \ge x\left(\left(\sum_{i=1}^{k} \|X_i\|^2\right)^{1/2} + B_n\right)\right) \le A\exp(-x^2/A)$$

where $B_n = (\sum_{i=1}^{n} E\|X_i\|^2)^{1/2}$

2 Stein's Method for Normal Distribution

Berry-Esseen ineuqalities are based on chracteristics function. Charles Stein in 1972 [11] proposed a different method to derive normal approximations, which works well for independent and dependent random variables.

2.1 The Stein Equation

Let Z be a standard normally distributed random variable and let $\mathscr{C}$ be the set of continuous and piecewise continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with $E|f'(Z)| < \infty$. Stein's method rests on the following observation.

Lemma 2.1 (5.1, Stein's Equation)**.** Let W be a real-valued random variable. If W has a standard normal distribution, then

$$Ef'(W) = EWf(W) \tag{3}$$

for any absolutely continuous function f with $E|f'(W)| < \infty$. If the equation above holds for any continuous and piecewise continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with $E|f'(Z)| < \infty$, then W has a standard normal distribution. That is to say, for nice functions f, $Ef'(W) = EWf(W) \iff W \sim N(0,1)$

Proof. The difficulty lies in sufficiency. Solve equation (4) given $z \in \mathbb{R}$ which is a simple ODE.

$$f'(w) - wf(w) = I(w \le z) - \Phi(z) \tag{4}$$

Consider LHS as the derivative of $e^{-w^2/2}f(w)$, the equation gives solution:

$$\begin{aligned} f_z(w) &= e^{w^2/2}\int_{-\infty}^{w}[I(x \le z) - \Phi(z)]e^{-x^2/2}dx \\ &= -e^{w^2/2}\int_{w}^{\infty}[I(x \le z) - \Phi(z)]e^{-x^2/2}dx \\ &= \begin{cases} \sqrt{2\pi}e^{w^2/2}\Phi(w)[1-\Phi(z)] & \text{if } w \le z \\ \sqrt{2\pi}e^{w^2/2}\Phi(z)[1-\Phi(w)] & \text{if } w \ge z \end{cases} \end{aligned}$$

□

Lemma 2.2 (Generalized Stein's Equation)**.** For a given real-valued measurable function h with $E|h(Z)| < \infty$, $W \sim N(0,1) \iff$

$$f'(w) - wf(w) = h(w) - Eh(Z) \tag{5}$$

Equation (3) is a special case when $h(w) = I(w \le z)$. The solution $f = f_h$ is given by

$$\begin{aligned} f_h(w) &= e^{w^2/2}\int_{-\infty}^{w}[h(x) - Eh(Z)]e^{-x^2/2}dx \\ &= -e^{w^2/2}\int_{w}^{\infty}[h(x) - Eh(Z)]e^{-x^2/2}dx \end{aligned}$$

Note that the stein's equation (3) and (5) provide tools to decide whether a random variable follows normal distribution. On the other hand, given a random variables follows normal distribution, we are able to bound any 'nice' function of the random variables by the properties of the solution of the equation.

As an example, for the function f_h defined in (5), if h is absolutely continuous, then

$$\begin{aligned} \sup|f_h(w)| &\le 2\sup_w|h'(w)| \\ \sup_w|f_h'(w)| &\le 4\sup_w|h'(w)| \\ \sup_w|f_h''(w)| &\le 2\sup_w|h'(w)| \end{aligned}$$

2.2 Stein's Method

Know:

$$W \sim N(0,1) \iff Ef'(W) - EWf(W) = 0$$
$$f'(w) - wf(w) = h(w) - Eh(Z),\ Z \sim N(0,1)$$

Objective:

$$W_n \xrightarrow{d} N(0,1) \iff Eh(W_n) - Eh(Z) \to 0 \text{ for } \forall \|h'\| \le 1$$

To achieve that, we need to show

$$Ef'(W_n) - EW_n f(W_n) = Eh(W_n) - Eh(Z) \to 0$$

Consider increments ξ_i are independent, $W_n = \sum_{i=1}^n \xi_i$. $E(\xi_i) = 0$, $\sum_{i=1}^n \xi_i^2 = 1$, $W^{(i)} \triangleq W - \xi_i$. For simplicity, $W^{(i)}$ and W below represent $W_n^{(i)}$ and W_n respectively.

2.2.1 Case $E|\xi|^3 < \infty$

Observe:

$$W^{(i)} \perp\!\!\!\perp \xi_i$$
$$f(a+b) - f(a) = \int_0^b f'(a+t)dt = \int_{-\infty}^{\infty} f'(a+t)(1(0<t<b) - 1(b<t<0))dt$$

Therefore

$$f(W^{(i)} + \xi_i) - f(W^{(i)}) = \int_{-\infty}^{\infty} f'(W^{(i)} + t)(1(0<t<\xi_i) - 1(\xi_i < t < 0))dt$$

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

Along with the properties of $K_i(t) \triangleq E(\xi_i(1(0<t<\xi_i) - 1(\xi_i < t < 0)))$

- $K_i(t) \ge 0$
- $\int_{-\infty}^{\infty} K_i(t)dt = E(\xi_i^2)$
- $\int_{-\infty}^{\infty} |t| K_i(t)dt = \frac{1}{2}E|\xi_i|^3$

2.2.2 Case $E|\xi|^3 < \infty$ not given

In this case Lindeberg's condition is equivalent to $\beta_2 + \beta_3 = 0$ where

$$\beta_2 = \sum_{i=1}^n E\xi_i^2 I(|\xi_i| > 1) \quad \text{and} \quad \beta_3 = \sum_{i=1}^n E|\xi_i|^3 I(|\xi_i| \le 1)$$

2.2.3 When h is an indicator function

In this case, by stein's equation, f'' does not exist.

Theorem 2.3. If $\exists \delta$ such that for h, $|Eh(W) - Eh(Z)| \le \|h'\|\delta$, then

$$\sup_z |P(W \le z) - \Phi(z)| \le 2\sqrt{\delta}$$

Proof. Introduce an absolute continuous function $h_\varepsilon(w) \ge$ Indicator function $1(w \le Z)$ □

2.3 Self-Normalized Berry-Esseen Inequality

Let $X_1, \ldots, X_n$ be independent random variables with $EX_i = 0$ and $EX_i^2 < \infty$

$$S_n = \sum_{i=1}^{n} X_i, \quad V_n^2 = \sum_{i=1}^{n} X_i^2, \quad B_n^2 = \sum_{i=1}^{n} EX_i^2$$

Bentkus et al in 1996 poved that [2]

$$\sup_z |P(S_n/V_n \le z) - \Phi(z)| \le C(\beta_2 + \beta_3)$$

where C is an absolute constant and

$$\beta_2 = B_n^{-2} \sum_{i=1}^{n} EX_i^2 I(|X_i| > B_n), \quad \beta_3 = B_n^{-3} \sum_{i=1}^{n} E|X_i|^3 I(|X_i| \le B_n)$$

Proof. Prove using Stein's Method. Let h in equation (5) be an indicator function.

From $f'(w) - wf(w) = 1(x \le z) - \Phi(z)$ we observe

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

In addition,

$$f'(W) - f'(W^{(i)} + t) = [Wf(W) - (W^{(i)} + t) f(W^{(i)} + t)] + [1(W \le Z) - 1(W^{(i)} + t) \le z)]$$

Take expectation on both sides, the second part is bounded using following theorem. □

2.4 Concentration Inequality

Theorem 2.4 (5.8, Deterministic Concentration Inequality)**.** $W = \sum_i \xi_i$. If $E\xi_i = 0$, $\sum_i E\xi_i^2 = 1$, then

$$P(a \le W \le b) \le b - a + 2 \sum_i E|\xi_i|^3$$

Proof. Again use $EWf(W) = \sum_{i=1}^{n} E \int_{-\infty}^{\infty} f'(W^{(i)} + t) K_i(t) dt$, the LHS can be written into

$$\begin{aligned} E(Wf(W)) &= \sum E(\xi f(W)) = \sum E(\xi_i (f(W) - f(W - \xi_i)) \quad W^{(i)} \perp\!\!\!\perp \xi_i \\ &= \sum E \int_{-\infty}^{\infty} f'(W + t)(\xi_i (1(0 < t < \xi_i) - 1(\xi_i < t < 0))) dt \\ &= \sum E \int_{-\infty}^{\infty} f'(W + t) \hat{K}_i(t) dt \end{aligned}$$

where $\hat{K}_i(t) \triangleq \xi_i (1(0 < t < \xi_i) - 1(\xi_i < t < 0))$.

Then construct a absolutely continuous function f such that $f'(w) = 1(a \le w \le b)$. This piecewisely linear function has property $|f(w)| \le (b-a)/2 + \delta$, $f'(w) \ge 0$ LHS has a upper bound, and the RHS has a lower bound by truncation. □

Theorem 2.5 (Randomized Concentration Inequality I)**.** Refer to [3] and Chapter 10, Page 108 of [7].

$$\begin{aligned} P(\Delta_1 \le W \le \Delta_2) \le & E|W(\Delta_2 - \Delta_1)| + 4\sum E|\xi_i|^3 \\ & + E|\xi_i W(\Delta_1 - \Delta_{1,i})| + E|\xi_i W(\Delta_2 - \Delta_{2,i})| \end{aligned}$$

where

$$\begin{aligned} \Delta_{1,i} &= \Delta_{1,i}(\xi_j, j \neq i) \\ \Delta_{2,i} &= \Delta_{2,i}(\xi_j, j \neq i) \end{aligned}$$

Theorem 2.6 (Randomized Concentration Inequality II)**.**

$$|P(W + \Delta \le z) - \Phi(z)| \le 8(\sum E|\xi_i|^2 + E|\Delta| + \sum_{i=1}^{n} E|\xi_i(\Delta - \Delta_i)|$$

where $W = \sum \xi_i$, $\Delta = \Delta(\xi_1, \dots, \xi_n)$, $\Delta_i = \Delta_i(\xi_j, j \neq i)$.

There are two possible choices for Δ_i:

- $\Delta_i = \Delta(\xi_1, \dots, \xi_{i-1}, 0, \xi_{i+1}, \dots)$
- $\Delta_i = \Delta(\xi_1, \dots, \xi_{i-1}, \xi_i^*, \xi_{i+1}, \dots)$ where $\{\xi_i^*\}$ is independent copies of $\{\xi_i\}$

2.5 Exchangeable Pair Approach

In order to find the limiting distribution and the error of approximation of W $(= W_n)$, we apply exchangeable pair approach.

Definition 2.1 (Exchangeable Pair)**.** (W, W') is said to be an exchangeable pair if

$$(W, W') \overset{d}{=} (W', W)$$

Some properties of exchangeable pair:

- Trivial cases include: $W \perp\!\!\!\perp W'$, $W = W'$
- For asymmetric $h(x, y)$, $Eh(W, W') \overset{\text{Joint Dist.}}{=} Eh(W', W) \overset{\text{Change Var.}}{=} -Eh(W, W') = 0$
- If $W = W(\xi_1, \dots, \xi_n)$ where ξ_i are independent, then $W' = W(\xi_1, \dots, \xi_I', \dots, \xi_n)$ and W are exchangeable, where I is a random index independent from other r.v.s with $P(I = k) = 1/n, 1 \le k \le n$ and $\{\xi_I'\}$ being independent copy of $\{\xi_I\}$.
- If $W = W(\xi_1, \dots, \xi_n)$ where ξ_i are independent, then $W' = W(\xi_1, \dots, \xi_I', \dots, \xi_n)$ and W are exchangeable, where I is a random index independent from other r.v.s with $\xi_i'|\xi_j, j \neq i \overset{d}{\sim} \xi_i|\xi_j, j \neq i$. By this construction, $W = \sum_i \xi_i$ satisties the condition of Theorem 2.7.

Theorem 2.7. Assume $E(W - W'|W) = \lambda(W + R_1)$, $\lambda \ge 0$ (Normal random variable satisfies this condition, related to linear regression), then for any absolute continuous function f

$$EWf(W) = E(\int_{-\infty}^{\infty} f'(w + t)\hat{K}(t)dt) - Ef(W)R_1 \tag{6}$$

Proof. Consider $h(w, w') = (w - w')(f(w) + f(w'))$ which is a asymmtric function,

$$\begin{aligned}
0 &= E(W - W')(f(W) + f(W')) \\
&= 2E(W - W')f(W) - E(W - W')(f(W) - f(W')) \text{ Trick} \\
&\text{Define } \Delta = W - W' \\
&= 2E(E(W - W')f(W)|W) - E\Delta(f(W) - f(W - \Delta)) \\
&= 2E(f(W)\lambda(W + R_1)) - E(\Delta \int_{-\Delta}^{0} f'(w + t)dt) \\
&\text{Define } \hat{K}(t) = \Delta(1(-\Delta \le t < 0) - 1(0 \le t < -\Delta))/2\lambda \\
&= 2\lambda[(E(Wf(W)) + E(R_1 f(W)) - E(\Delta \int_{-\infty}^{\infty} f'(w + t)\hat{K}(t)dt))]
\end{aligned}$$

□

Theorem 2.8. Assume $\|h'\| < \infty$

$$|Eh(W) - Eh(Z))| \le 2\|h'\|(E|1 - \frac{E(\Delta^2(w))}{2\lambda}| + \frac{|\Delta|^3}{\lambda} + E|R_1|)$$

Proof. (i) Use stein's method and the indentity (6) in the previous theorem. or (ii) Construct exchangeable pair $W = \sum_i \xi_i$. Exchangeable pair approach can give us optimal rate but condition may not be optimal, in this case we need to assume $E\xi_i^4 < \infty$. □

Theorem 2.9 (Berry–Esseen Inequality for exchangeable pairs [10])**.** Assume $E(W - W'|W) = \lambda(W + R_1), E((W - W')^2|W) = 2\lambda + R_2$, then

$$|P(W \le z) - \Phi(z)| \le E|R_1| + E|R_2| + \frac{1}{\lambda}E\,|E(\Delta\Delta^*\,|w)|$$

where $\Delta = W - W'$, $\Delta^* \ge |\Delta|$ Consider$\Delta = \Delta^+ - \Delta^-$

3 Stein's Method for Nonnormal Distribution

3.1 General Distribution

Consider $Y \overset{pdf}{\sim} p(y)$, $p(y) > 0$, we have

$$E\left(\frac{(p(y)f(y))'}{p(y)}\right) = p(y)f(y)|_{-\infty}^{\infty} = 0$$

Theorem 3.1 (Stein's Identity for nonnormal distribution)**.**

$$E\left(\frac{(p(y)f(y))'}{p(y)}\right) = 0 \tag{7}$$

$$Ef'(y) + E\left(\frac{(p(y)f'(y))}{p(y)}\right) = 0 \tag{8}$$

Correspond to stein's indentity for normal distribtion (3)

Theorem 3.2 (Stein's Equation for nonnormal distribution)**.**

$$\frac{(p(y)f(y))'}{p(y)} = h(y) - Eh(y) \tag{9}$$

Correspond to stein's equation for normal distribtion (5)

Theorem 3.3. Let (W, W') be exchangeable. Assume:

- $E(W - W'|W) = \lambda(g(W) + R(W))$
- $\Delta = W - W'$, $\frac{E(\Delta^2|W)}{2\lambda} \xrightarrow{p} 1$ $\left(or\ \frac{E(\Delta^2|W)}{2\lambda} = v(w)\right)$
- $\frac{E|\Delta|^3}{\lambda} \to 0$, $E|R| \to 0$

then

$$W \xrightarrow{d} Y$$

where Y has pdf: $p(y) = Ce^{-G(y)}$, $G(y) = \int_0^y g(t)dt$ $\left(or\, p(y) = C/v(y)e^{-G(y)},\ G(y) = \int_0^y g(t)/v(t)dt\right)$

Proof. Similar to the proof of Theorem 2.7. The key is to guess the form of $g(w) = \frac{p(y)}{p'(y)} = (ln(p(y))')$ by observing similar terms in stein's equation.

$$Eh(W) - Eh(Y) = E(f'(W) + \frac{p'(W)}{p(W)}f(W)) \tag{10}$$

□

This theorem provide the tool to determine the limiting distribution of W under the framework of exchangeable pair.

3.2 Poisson Distribution

L.H.Y. Chen [4] in 1975 first proposed Poisson Approxiamtion using stein's method, later stein's equation is derived in 1995 [1].

4 Large Deviations for Self-Normalized Sums

Theorem 4.1 (Cramér-Chernoff large deviation theorem)**.** Let $X, X_1, \ldots, X_n$ be i.i.d. random variables with $P(X \neq 0) > 0$ and let $S_n = \sum_{i=1}^n X_i$. If

$$Ee^{\theta_0 X} < \infty$$

then for every $x > EX$

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

or equivalently,

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

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

Definition 4.1 (Self-nomalized)**.** $\{X_i\}$ i.i.d., $S_n = \sum_{i=1}^n X_i$, $V_n^2 = \sum_{i=1}^n X_i^2$, then S_n/V_n is called self-normalized t-statistics.

Definition 4.2 (Slowly varying)**.** $l(x)$ is called slowly varying if

$$\forall t > 0,\ \frac{l(tx)}{l(x)} \to 1$$

Theorem 4.2 (Self-normalized law of the iterated logarithm[5]). If $EX_1 = 0$, $l(x) \triangleq EX_1^2 1(|X_1| \leq x)$ is slowly varying, then

$$\limsup \frac{S_n}{V_n\sqrt{2\log\log n}} \overset{a.s.}{=} 1$$

Theorem 4.3 (Self-normalized large deviations [9]). Assume that either $EX \geq 0$ or $EX^2 = \infty$. Let $V_n^2 = \sum_{i=1}^n X_i^2$. Then

$$lim_{n\to\infty} P\left(S_n \geq x\sqrt{n}V_n\right)^{1/n} = \sup_{b\geq 0}\inf_{b\geq 0} Ee^{t\left(bX - x\left(X^2+b^2\right)/2\right)}$$

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

Another equivalent presentation of this theorem: If $EX = 0$, $EX_1^2 1(|X_1| \leq x)$ is slowly varying, then $\forall x_n \to \infty$, $x_n = o(\sqrt{n})$, we have

$$\ln P(\frac{S_n}{V_n} \geq x) \sim -\frac{x_n^2}{2}$$

i.e. $\forall 0 < \varepsilon < 1$,

$$e^{-\frac{(1+\varepsilon)x_n^2}{2}} < P(\frac{S_n}{V_n} \geq x_n) \leq e^{-\frac{(1-\varepsilon)x_n^2}{2}}$$

$$P(\frac{S_n}{\sigma\sqrt{n}} \geq x) \to 0 \iff Ee^{t_0 X_1} < \infty$$

Proof. First truncate

$$S_n = \sum X_i 1(|X_i| \leq z_n) + \sum X_i 1(|X_i| > z_n) \triangleq \sum X_{i,1} + \sum X_{i,2}$$

. By $P(X + Y \geq x) \leq P(X \geq (1-\varepsilon)x) + P(Y \geq \varepsilon x)$

$$P\left(\frac{S_n}{V_n} \geq x_n\right) \leq P\left(\frac{\sum X_{i,1}}{V_n} \geq (1-\varepsilon)x_n\right) + P\left(\frac{\sum X_{i,2}}{V_n} \geq \varepsilon x_n\right) \triangleq \mathrm{I} + \mathrm{II}$$

Bound (I) using Benette Hoeffding's Inequality, bound (II) using Cauchy's Inequality. □

Theorem 4.4 (Cramér-Type Large Deviations for student's t-statistics [8]). If $EX_i = 0$, $E|X_i|^3 < \infty$, then

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

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

Proof. Trick: use the relationship between V_n and V_n^2 □

5 Summary

This course unveils how Stein's method can be applied to determine the limiting distributions of sample sums when samples are (i) independent or not; (ii) follows normal distrbutions or not. When samples are not necesserily normal random variables, adaptions to Stein's equation are also available, one typical example being Stein-Chen's method for poisson approximation. In addition, large deviation problems, which study the tail behaviour of random variables, are also introduced, where Self-normalizing plays an important role to eliminate the moment boundedness constraints. When applied to the t-statistics, large deviations also facilitates research of hypothesis testing. Tricks such as truncation, exchangeable pairs approach, together with the inequalities stated at the start of the report, are widely used in the proofs of the theorems.

References

[1] A. D. Barbour, L. H. Y. Chen, and K. P. Choi, *Poisson approximation for unbounded functions, i: Independent summands*, Statistica Sinica, pp. 749–766, http://www.jstor.org/stable/24305068.

[2] V. Bentkus and F. Götze, *The berry-esseen bound for student's statistic*, The Annals of Probability, 24 (1996), pp. 491–503.

[3] B. Bercu, E. Gassiat, E. Rio, et al., *Concentration inequalities, large and moderate deviations for self-normalized empirical processes*, The Annals of Probability, 30 (2002), pp. 1576–1604.

[4] L. H. Chen, *Poisson approximation for dependent trials*, The Annals of Probability, (1975), pp. 534–545.

[5] P. S. Griffin and J. D. Kuelbs, *Self-normalized laws of the iterated logarithm*, Ann. Probab., 17 (1989), pp. 1571–1601, https://doi.org/10.1214/aop/1176991175, https://doi.org/10.1214/aop/1176991175.

[6] X. He and Q.-M. Shao, *On parameters of increasing dimensions*, Journal of Multivariate Analysis, 73 (2000), pp. 120–135, https://EconPapers.repec.org/RePEc:eee:jmvana:v:73:y:2000:i:1:p:120-135.

[7] T. L. Lai and Q.-M. Shao, *Self-Normalized Processes: Limit Theory and Statistical Applications. Probability and Its Applications*, Springer, 2009.

[8] Q.-M. Shao, *A cramér type large deviation result for student's t-statistic*, Journal of Theoretical Probability, 12 (1999), pp. 385–398, https://doi.org/10.1023/A:1021626127372.

[9] Q.-M. Shao et al., *Self-normalized large deviations*, The Annals of Probability, 25 (1997), pp. 285–328.

[10] Q.-M. Shao, Z.-S. Zhang, et al., *Berry–esseen bounds of normal and nonnormal approximation for unbounded exchangeable pairs*, The Annals of Probability, 47 (2019), pp. 61–108.

[11] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, in Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory, Berkeley, Calif., 1972, University of California Press, pp. 583–602, https://projecteuclid.org/euclid.bsmsp/1200514239.