(Thm) $\varepsilon_i \overset{iid}{\sim}$ $P(\varepsilon_i = 1) = P(\varepsilon_i = 0) = \frac{1}{2}$

Then $P\left(\frac{\sum_{i=1}^n a_i \varepsilon_i}{\sqrt{\sum a_i^2}} \geq x\right) \overset{①}{\leq} e^{-x^2/2}$, $\forall x > 0$

$\overset{②}{\leq} c \frac{1}{x} e^{-x^2/2}$

[Pf ①: WLOG: assume $\sum_i a_i^2 = 1$ use $\frac{1}{2}(e^s + e^{-s}) \leq e^{s^2/2}$

$E e^{t a_i \varepsilon_i} = \frac{1}{2}(e^{t a_i} + e^{-t a_i}) \leq e^{\frac{1}{2} t^2 a_i^2}$

$P(\sum a_i \varepsilon_i \geq x) \leq e^{-tx} E(e^{t \sum a_i \varepsilon_i})$

$\leq e^{-tx} \prod_{i=1}^n e^{\frac{1}{2} t^2 a_i^2}$

$= \exp(-tx + \sum_i \frac{1}{2} t^2 a_i^2)$

$= \exp(-tx + \frac{t^2}{2})$

$\forall t > 0,\ t = x$ $= e^{-x^2/2}$ ① //]

(Thm) X_i indep. symmetric $X_i \overset{d}{=} -X_i$

then $P\left(\frac{\sum X_i}{\sqrt{\sum X_i^2}} \geq x\right) \leq e^{-x^2/2}$, $\forall x > 0$

self-normalized / studentized

Note: $\frac{1}{n} \sum_i X_i \to \mu$ a.s.

$\frac{\frac{1}{n}\sum_i X_i - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$ ⋮ $\frac{\frac{1}{n}\sum_i X_i - \mu}{\hat{\sigma}/\sqrt{n}} \xrightarrow{d} N(0,1)$

↳ known assumption ⋮ ↳ t-statistic T_n. $\hat{\sigma} = \left(\frac{1}{n-1} \sum_i (X_i - \bar{X})^2\right)^{1/2}$

$$T_n = \frac{\sum X_i}{\sqrt{\frac{n}{n-1}\sum (X_i - \bar{X})^2}} \qquad S_n = \sum X_i \quad V_n^2 = \sum X_i^2$$

$$\Rightarrow T_n = \frac{S_n}{\sqrt{\frac{n}{n-1}\left(\sum X_i^2 - n\bar{X}^2\right)}} = \frac{S_n}{\sqrt{\frac{n}{n-1}\left(V_n^2 - S_n^2/n\right)}} = \frac{S_n}{\sqrt{\frac{n}{n-1} V_n^2 \left(1 - \frac{1}{n}\frac{S_n^2}{V_n^2}\right)}}$$

$$= \sqrt{\frac{n-1}{n}} \cdot \frac{S_n}{V_n} \cdot \frac{1}{\sqrt{1 - \frac{1}{n}\left(\frac{S_n}{V_n}\right)^2}} = f\left(\frac{S_n}{V_n}\right) \triangleq f(X_n) = \sqrt{\frac{n-1}{n}}\, X_n \frac{1}{\sqrt{1 - \frac{X_n^2}{n}}}$$

$$X_n \triangleq \frac{S_n}{V_n} = \frac{\sum X_i}{\left(\sum X_i^2\right)^{1/2}} \geq \sqrt{n}.$$

$$\{T_n \geq t\} = \left\{\frac{S_n}{V_n} \geq x_t\right\} \qquad t = \frac{1}{\sqrt{\frac{n}{n-1}}} \frac{x}{\sqrt{1 - \frac{1}{n}x^2}}$$

$$t^2 \cdot \frac{n-1}{n} = \frac{x^2}{1 - \frac{1}{n}x^2}$$

70% of practical statistics are studentized

(Thm) $\{\varepsilon_i\}$ indep $\{X_i\}$ indep. ☆ no symmetric condition

$$X_i \overset{d}{=} X_i \varepsilon_i$$

$$\xrightarrow[\text{independent}]{\{\varepsilon_i\},\{X_i\}} \{X_i, 1 \le i \le n\} \overset{d}{=} \{X_i \varepsilon_i, 1 \le i \le n\}$$

[Pf]: $$P\left(\frac{\sum X_i}{\sqrt{\sum X_i^2}} \ge x\right) = P\left(\frac{\sum X_i \varepsilon_i}{\sqrt{\sum (X_i \varepsilon_i)^2}} \ge x\right)$$

$$= P\left(\frac{\sum X_i \varepsilon_i}{\sqrt{\sum X_i^2}} \ge x\right)$$

(By conditional expectation)

$$= E\left(P\left(\frac{\sum X_i \varepsilon_i}{\sqrt{\sum X_i^2}} \ge x\right) \Big| X_1, \dots, X_n\right)$$

$$= E\left(P\left(\frac{\sum x_i \varepsilon_i}{\sqrt{\sum x_i^2}} \ge x\right) \Big| X_1 = x_1, \dots, X_n = x_n\right)$$

☆ $\{X_i\}, \{\varepsilon_i\}$ indep

$$\le E\left(e^{-x^2/2} \Big| X_1 = x_1, \dots, X_n = x_n\right)$$

$$= e^{-x^2/2}$$

//]

(Thm) $\{X_i\}$ indep. $EX_i=0$, $\sum EX_i^2 \le B_n^2$

$S_n=\sum X_i \qquad V_n^2=\sum X_i^2$

Then

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

He, Shao (2000)

Note: M-estimate.

a broad class of estimators, can be solved by solving an equation

Idea: Transform the original r.v. into a symmetric one.

called "symmetrization"

Consider $\{X_i\} \overset{d}{=} \{Y_i\}$ independent

$\Rightarrow \{X_i-Y_i\}$ is symmetric

only to prove

$$\{\sum X_i \ge x V_n + \cdots\} \subset \{\sum(X_i-Y_i) \ge x\sqrt{\sum(X_i-Y_i)^2}\}$$

$$\{\sum X_i \ge x V_n + \cdots, \underbrace{|\sum Y_i| \le C_n}_{①}\} \overset{①}{\subset} \{\sum(X_i-Y_i) \ge x V_n - C_n\}$$

$\cap \Leftarrow$ next to prove

$$\{\sum(X_i-Y_i) \ge x\sqrt{\sum(X_i-Y_i)^2}\}$$

Observe: $\sqrt{\sum(X_i-Y_i)^2} \le \sqrt{\sum X_i^2} + \sqrt{\sum Y_i^2}$ control this term by

② $\sum Y_i^2 \le D_n^2$

then $\{\sum(X_i-Y_i) \ge x\sqrt{\sum(X_i-Y_i)^2}, ①, \underbrace{\sum Y_i^2 \le D_n^2}_{②}\}$

②

$$\subset \left\{ \sum(X_i - Y_i) \ge x \left(\sqrt{\sum(X_i-Y_i)^2} - D_n \right) - C_n \right\}$$

$\Rightarrow \quad x \ge 1$

$$\left\{ \sum X_i \ge x(V_n + D_n + C_n) \right\}, \quad \underset{①}{|\sum Y_i| \le C_n}, \quad \underset{②}{\sum Y_i^2 \le D_n^2} \Big\}$$

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

$$\Rightarrow P\left(\sum X_i \ge x(V_n + D_n + C_n)\right) P\left(|\sum Y_i| \le C_n, \ \sum Y_i^2 \le D_n^2\right)$$

$$\le P\left(\sum(X_i - Y_i) \ge x\sqrt{\sum(X_i-Y_i)^2}\right) \le e^{-x^2/2}$$

remain to be controlled

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

$$P(|\sum Y_n| > C_n) \le \frac{E(\sum Y_i^2)}{C_n^2} = \frac{1}{4}$$

$$P(|\sum Y_i^2 \ge D_n^2) \le \frac{\sum E Y_i^2}{D_n^2} = \frac{1}{4}$$

where $C_n = 2B_n$, $D_n = 2B_n$

(Conjecture) $\sum a_i^2 = 1$

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

i.e. $x=1$, RHS $e^{-x^2/2} \to \frac{1}{2}$

$g_n(y) = P(|\sum a_i \varepsilon_i| > y) + P(|\sum a_i \varepsilon_i| > \frac{1}{y}) \le 1$

? if $g_n(y) \uparrow$, the inequality is proved, since $y=0$, $y=+\infty$ satisfies

Stein's Method.

Stein (1972)

Recall: CLT proof converge in c.f. $\Rightarrow$ converge in dist.

Stein: $W \sim N(0,1)$

$$\Leftrightarrow \quad Ef'(W) = E(Wf(W)) \quad \forall \text{"nice" } f$$

($\Rightarrow$) [Pf.

$$E(wf(w)) = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} w f(w)\, dw$$

$$= -\int \frac{1}{\sqrt{2\pi}} f(w)\, de^{-\frac{w^2}{2}}$$

$$= \frac{1}{\sqrt{2\pi}}\left(f(w) e^{-\frac{w^2}{2}} \Big|_{\Omega} - \int e^{-\frac{w^2}{2}} f'(w)\, dw \right)$$

$$= E(f'(w)) \quad //]$$

$(\Leftarrow)$ [Pf: For given z, let $f = f_z$

Stein's Equation: $f'(w) - wf(w) = 1(w \le z) - \Phi(z)$ $(*)$ if $(*)$ holds

$$f'(W) - Wf(W) = 1_{(W \le z)} - \Phi(z)$$

$$E\left(f'(W) - Wf(W)\right) = P(W \le z) - \Phi(z) = 0$$

Stands for $\forall z \Rightarrow W \sim N(0,1)$

Now need to solve $(*)$.

$$e^{-w^2/2}\left(f'(w) - wf(w)\right) = e^{-\frac{w^2}{2}}\left(1(w \le z) - \Phi(z)\right)$$

$$\frac{d}{dw}\left(e^{-w^2/2} f(w)\right) = e^{-w^2/2}(1(w \le z)) - e^{-\frac{w^2}{2}}\Phi(z).$$

$$f(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{-\frac{t^2}{2}}\left(1(t \le z) - \Phi(z)\right)dt$$

$$= -e^{\frac{w^2}{2}} \int_{w}^{\infty} e^{-\frac{t^2}{2}}\left(1(t \le z) - \Phi(z)\right)dt$$

$$= \begin{cases} \sqrt{2\pi}\, e^{\frac{w^2}{2}} \Phi(w)(1 - \Phi(z)), & w \le z \\ \sqrt{2\pi}\, e^{\frac{w^2}{2}} \Phi(z)(1 - \Phi(w)), & w \ge z \end{cases}$$

Properties: 1° $0 \le f(w) \le \sqrt{2\pi}/4$

2° $|f'(w)| \le 1$

3° $wf(w) \uparrow$

(General Steins Equation)

given h, $Z \sim N(0,1)$

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

$$f_h(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{-\frac{t^2}{2}} (h(t) - Eh(Z))\,dt$$

$$= -e^{\frac{w^2}{2}} \int_{w}^{\infty} e^{-\frac{t^2}{2}} (h(t) - Eh(Z))\,dt$$

Properties:

① $\|f\| = \sup_t |f(t)| \leq 2\|h\|$

② $\|f'\| \leq 4\|h\|$

③ $\|f\| \leq 2\|h'\|$, $\|f'\| \leq \|h'\|$, $\|f''\| \leq 2\|h'\|$

most of the time we only need to study the properties

Aim: $W_n \xrightarrow{d} N(0,1)$

$Eh(W_n) - Eh(Z)$ for $\forall h$ with bdd der.

by stein eqn

$= E f'(W_n) - E W_n f(W_n)$

idea of prove $E W_n f(W_n) \to E f'(W_n)$