Approximating by the Normal Distribution

Elizabeth Meckes

AIM / Cornell

March 26, 2007

The central limit theorem (vanilla version)

Let $X_1, X_2, \ldots$ be independent, identically distributed (i.i.d.) random variables with $\mathbb{E}[X_1]=0$, $\mathbb{E}[X_1^2]=\sigma^2$.

Let $\tilde{S}_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i$. Then $\forall t \in \mathbb{R}$,

$$\mathbb{P}(\tilde{S}_n \le t) \xrightarrow[n\to\infty]{} \Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}\,dx.$$

Key piece of intuition:

One reason that the normal distribution comes up so much is that many random variables are "like" $\tilde{S}_n$: they are sums of a lot of little things, which don't affect each other much.

Characteristic functions (a.k.a. Fourier transforms)

Definition: If μ is a probability measure, the characteristic function φ_μ of μ is defined by $\varphi_\mu(t) = \int_{-\infty}^{\infty} e^{itx} \, d\mu(x)$.

In terms of r.v.s: if X is a random variable, its characteristic function φ_X is defined by $\varphi_X(t) = \mathbb{E}[e^{itX}]$.

The continuity theorem:

Let μ_k, μ be probability measures with characteristic functions φ_k, φ.

Then $\mu_k \xrightarrow[k\to\infty]{w^*} \mu$ iff $\varphi_k(t) \xrightarrow[k\to\infty]{} \varphi(t) \ \forall t \in \mathbb{R}$

That is, $\varphi_k(t) \to \varphi(t) \ \forall t \in \mathbb{R}$ iff for every bd. continuous $f : \mathbb{R} \to \mathbb{R}$,

$$\int f \, d\mu_k \longrightarrow \int f \, d\mu.$$

Remarks

1. Characteristic functions are one of the most commonly used tools in probability theory and applications.
2. C.f.s can be used to substantially generalize the CLT; they are particularly well suited to relaxing the assumption that all $\{X_i\}$ have the same distributions.
3. C.f.s can be used to obtain rates of convergence, since the Fourier transform is an isometry on L^2.
4. Useful for other limiting distributions, not just Gaussian.

Some basic facts about c.f.s

- $\mathbb{E}[e^{itZ}] = \varphi_Z(t) = e^{-t^2/2}$ for $Z \sim N(0,1)$

- If X_1, X_2 are independent random variables, then $\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\,\varphi_{X_2}(t)$.

- If X has finite moments, then

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\,\mathbb{E}[X^k];$$

if X has moments up to order n, then

$$\left|\varphi_X(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!}\,\mathbb{E}[X^k]\right|$$

$$\le \mathbb{E}\left[\min\left(\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right)\right]$$

Proof of CLT via c.f.s:

Have: $\{X_i\}$ i.i.d., $\mathbb{E}[X_i]=0$, $\mathbb{E}[X_i^2]=1$, $\tilde{S}_n := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$. Want: $\tilde{S}_n \xrightarrow{w^*} Z$.

By the continuity theorem, it suffices to show $\varphi_{\tilde{S}_n}(t) \to \varphi_Z(t) = e^{-t^2/2}$ $\forall t \in \mathbb{R}$.

$$\varphi_{\tilde{S}_n}(t) = \varphi_{\frac{X_1}{\sqrt{n}}+\cdots+\frac{X_n}{\sqrt{n}}}(t) = \varphi^n_{\frac{X_1}{\sqrt{n}}}(t) = \varphi^n_{X_1}\left(\frac{t}{\sqrt{n}}\right).$$

Now, $\varphi_{X_1}(0) = \mathbb{E}\left[e^{i\cdot 0\cdot X_1}\right] = 1$;

$\varphi'_{X_1}(0) = \mathbb{E}[iX] = 0$ by assumption,

$\varphi''_{X_1}(0) = -\mathbb{E}\left[\frac{1}{2}X^2\right] = -\frac{1}{2}$

$$\Rightarrow \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)$$

$$\Rightarrow \varphi^n_{X_1}\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{-t^2/2}.$$

The Method of Moments

Theorem: Let $Z \sim N(0,1)$ and $\{X_i\}$ a sequence of random variables such that for each i, X_i has moments of all orders. Suppose that

$$\mathbb{E}[X_n^r] \to \mathbb{E}[Z^r] = \begin{cases} (r-1)!! & , r \text{ even} \\ 0 & r \text{ odd} \end{cases}$$

for each fixed $r \in \mathbb{Z}^+$. Then

$$X_n \xrightarrow{w^*} Z \; ; \text{ i.e., } \mathbb{E} f(X_n) \to \mathbb{E} f(Z)$$

for all bounded, continuous $f : \mathbb{R} \to \mathbb{R}$.

Remarks

1. This is how CLT was first proved rigorously (by Chebychev, 1887)

2. Diaconis: "Many young probabilists now shun the method of moments as restricted and heavy-handed. There are some however who realize that moments generally get the job done without taking five years off to develop special theory".

3. It is nearly impossible to get rates of convergence from the method of moments; you can get something with actual *equality* of the first n moments to those of Z.

4. The theorem above is true for any limiting distribution which is determined by its moments. There are elegant results about when this happens, what sequences of numbers can occur as moments, etc.

5. You must have convergence of *all* moments for the method to work.

CLT via the method of moments

$\{X_i\}$ i.i.d., $\mathbb{E}[X_i]=0$, $\mathbb{E}[X_i^2]=1$,

assume that $|X_i| \le M$ almost surely.

Let $S_n = X_1 + \dots + X_n$.

$$S_n^r = \sum_{p=1}^{r} \sum_{r_1+\dots+r_p=r} \frac{r!}{r_1!\cdots r_p!} \frac{1}{p!} \Big(\overset{\text{distinct indices}\,\downarrow}{\sum_{i_1,\dots,i_p}{}'} X_{i_1}^{r_1} \cdots X_{i_p}^{r_p} \Big)$$

Define $A_n(r_1,\dots,r_p) = \sum_{i_1,\dots,i_p}{}' \left(\frac{1}{n}\right)^{r/2} \mathbb{E}[X_{i_1}^{r_1}] \cdots \mathbb{E}[X_{i_p}^{r_p}]$

Then $\mathbb{E}\left[\left(\frac{1}{\sqrt{n}} S_n\right)^r\right] = \sum_{p=1}^{r} \sum_{r_1+\dots+r_p=r} \frac{r!}{r_1!\cdots r_p!} \frac{1}{p!} A_n(r_1,\dots,r_p)$

Claim: $\lim_{n\to\infty} A_n(r_1,\dots,r_p) = \begin{cases} 1 & r_1 = \dots = r_p = 2 \\ 0 & \text{———} \end{cases}$

With the claim, get $\mathbb{E}\left[\left(\frac{1}{\sqrt{n}} S_n\right)^r\right] \to \frac{r!}{2^{r/2}(r/2)!}$

$= (r-1)!!$

(r even)

and $\mathbb{E}\left[\left(\frac{1}{\sqrt{n}} S_n\right)^r\right] \to 0$, r odd.

To prove the claim:

If any $r_j = 1$, then $\mathbb{E}[X_{i_j}^{r_j}] = \mathbb{E}[X_1] = 0.$

$\Rightarrow$ All $r_j \geq 2$. Suppose one $r_j > 2$.

Then $r > 2p$.

$$\sum_{i_1,\dots,i_p}{}' \left(\frac{1}{n}\right)^{r/2} \mathbb{E}[X_{i_1}^{r_1}]\cdots\mathbb{E}[X_{i_p}^{r_p}]$$

$$\leq \frac{M^{r-2p}}{n^{r/2}} \sum_{i_1,\dots,i_p}{}' \mathbb{E}[X_{i_1}^{2}]\cdots\mathbb{E}[X_{i_p}^{2}]$$

$$\leq M^{r-2p} \, n^{p-r/2} \to 0$$

since $p - \frac{r}{2} < 0$.

Finally,

$$A_n(2,\dots,2) = \left(\frac{1}{n}\right)^{r/2} \sum_{i_1,\dots,i_{r/2}}{}' \mathbb{E}[X_{i_1}^{2}]\cdots\mathbb{E}[X_{i_p}^{2}]$$

$$= \frac{1}{n^{r/2}} \cdot \left[n(n-1)\cdots(n-\tfrac{r}{2}+1)\right]$$

$$\longrightarrow 1.$$

The Lindeberg Method

The crucial observation here is that if $Z_1, \dots, Z_n$ are i.i.d. $N(0,1)$, then $Z = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$ is also distributed as $N(0,1)$.

Idea: to approximate $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ by Z, replace the X_i one at a time by Z_i and estimate the error.

Remarks

1. The approach is fairly bare-hands, and as such has been useful in very general settings (e.g. CLTs in infinite-dimensional settings)
2. Weakening the independence assumption is fairly straightforward here, as is the assumption that all X_i have the same dist.

CLT via the Lindeberg method

Let $\{X_i\}$ be a collection as usual of i.i.d. random variables, $\mathbb{E}[X_i]=0$, $\mathbb{E}[X_i^2]=1$.

Note that $\{X_i\}$ easily satisfy the "Lindeberg condition":

$$\frac{1}{n}\sum_{i=1}^{n}\int_{\{\frac{|x_i|}{\sqrt{n}}>\varepsilon\}} x_i^2\, d\mu(x_i) \xrightarrow[n\to\infty]{} 0$$

(here μ is the common distribution of the X_i's.)

Fix $f\in C_c^\infty$: show that

$$\mathbb{E} f\left(\tfrac{1}{\sqrt{n}}\sum X_i\right) \longrightarrow \mathbb{E} f(Z) \quad \text{by}$$

replacing the X_i one at a time by Z_i.

By Taylor's theorem, there is a K and $g(h)$ such that $g(h) \le K \min\{h^2, h^3\}$, and

$$|f(x+h_1) - f(x+h_2) - f'(x)(h_1-h_2) - \tfrac{1}{2} f''(x)(h_1^2-h_2^2)|$$

$$\le g(h_1) + g(h_2).$$

Define $T_k = X_1 + \cdots + X_{k-1} + Z_{k+1} + \cdots + Z_n$

$\Rightarrow T_n + X_n = S_n$, $T_1 + Z_1 \sim N(0, n)$.

↑ is exactly dist. as

$$|\mathbb{E}\, f(\tfrac{1}{\sqrt{n}} S_n) - f(Z)|$$

$$\le \sum_{k=1}^{n} \left| \mathbb{E}\left[f\left(\frac{T_k + X_k}{\sqrt{n}}\right) - f\left(\frac{T_k + Z_k}{\sqrt{n}}\right)\right]\right|$$

$$= \sum_{k=1}^{n} \left| \mathbb{E}\left[f\left(\frac{T_k + X_k}{\sqrt{n}}\right) - f\left(\frac{T_k + Z_k}{\sqrt{n}}\right) - f'(T_k)\left(\frac{X_k - Z_k}{\sqrt{n}}\right) - \frac{1}{2} f''(T_k)\left(\frac{X_k^2 - Z_k^2}{n}\right)\right]\right|$$

Since $\mathbb{E}[X_k] = \mathbb{E}[Z_k]$ and

$\mathbb{E}[X_k^2] = \mathbb{E}[Z_k^2]$.

By Taylor, this is

$$\leq \sum_{k=1}^{n} \left(\mathbb{E} g\left(\frac{X_k}{\sqrt{n}}\right) + \mathbb{E} g\left(\frac{Z_k}{\sqrt{n}}\right)\right)$$

$$= n \mathbb{E} g\left(\frac{X_1}{\sqrt{n}}\right) + n \mathbb{E} g\left(\frac{Z_1}{\sqrt{n}}\right).$$

Now,

$$n \mathbb{E} g\left(\frac{X_1}{\sqrt{n}}\right) = n\Big[K \int_{\{|x| \leq \varepsilon\sqrt{n}\}} \left|\frac{x}{\sqrt{n}}\right|^3 d\mu(x)$$

$$+ K \int_{\{|x| > \varepsilon\sqrt{n}\}} \left|\frac{x}{\sqrt{n}}\right|^2 d\mu(x)\Big]$$

We've already seen that the 2nd term goes to zero for fixed ε, $n \to \infty$.

$$nK \int_{\{|x| \leq \varepsilon\sqrt{n}\}} \left|\frac{x}{\sqrt{n}}\right|^3 d\mu(x) \leq K \cdot n \cdot \varepsilon \int_{\{|x| \leq \varepsilon\sqrt{n}\}} \left|\frac{x}{\sqrt{n}}\right|^2 d\mu(x)$$

$$\leq K \cdot \varepsilon$$

The same argument works for $n \mathbb{E} g\left(\frac{Z_k}{\sqrt{n}}\right)$.

The martingale CLT

Definition: A sequence $\{X_n, \mathcal{F}_n\}$ of r.v.s and σ-algebras $(\mathcal{F}_{k-1} \subseteq \mathcal{F}_k)$ is a martingale difference if $\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = 0$ almost surely.

(Think $\mathbb{E}[X_n \mid X_1, \ldots, X_{n-1}] = 0$ - the expected change in X at the nth step is zero)

If $S_n = \sum_{i=1}^{n} X_i$, this means $\mathbb{E}[S_n \mid \mathcal{F}_{n-1}] = S_{n-1}$

Idea: drop the independence assumption from the CLT and replace it with this special kind of dependence.

<u>Theorem</u> (Brown): Let $\{X_n, \mathcal{F}_n\}$ be a martingale difference, $S_n = \sum_{i=1}^{n} X_i$. Define:

$$\sigma_n^2 = \mathbb{E}\left[X_n^2 \mid \mathcal{F}_{n-1}\right]$$

$$V_n^2 = \sum_{j=1}^{n} \sigma_j^2$$

$$s_n^2 = \mathbb{E}\left[V_n^2\right] = \mathbb{E}\left[S_n^2\right]$$

Suppose that

- $V_n^2 \, s_n^{-2} \to 1$ in probability as $n \to \infty$.
- $s_n^{-2} \sum_{j=1}^{n} \mathbb{E}\left[X_j^2 \, \mathbb{1}(|X_j| \geq \varepsilon s_n)\right] \to 0$ in probability as $n \to \infty$ ← Lindeberg condition

Then:

$$\lim_{n \to \infty} \mathbb{P}\left[S_n / s_n \leq t\right] = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} \, dx.$$

Stein's Method

Key observation: the normal distribution is characterized by the fact that $Z \sim N(0,1)$ if and only if

$$\mathbb{E}[f'(Z) - Zf(Z)] = 0$$

for all f such that $\mathbb{E}[f'(Z)]$ and $\mathbb{E}[Zf(Z)]$ exist.

In fact:

Theorem (Stein): Let $h: \mathbb{R} \to \mathbb{R}$ be bounded and piecewise C^1. Then

$$f(w) = e^{w^2/2} \int_{-\infty}^{w} [h(x) - \mathbb{E}h(Z)]\, e^{-x^2/2}\, dx$$

is a solution to

$$f'(w) - wf(w) = h(w) - \mathbb{E}h(Z).$$

Furthermore,

(i) $\|f\|_\infty \le \sqrt{2\pi}\, \|h\|_\infty$

(ii) $\|f'\|_\infty \le 2\|h\|_\infty$

(iii) $\|f''\|_\infty \le 2\|h'\|_\infty$.

Idea: Given W conjectured tobe approximately Gaussian, try to bound $\mathbb{E}[f'(W) - Wf(W)]$ for a large class of f and use the theorem to show that this implies $\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]$ is small for a large class of h.

Remarks

- This method is very useful in weakening independence assumptions.
- Convergence rates come for free: many notions of distance between distribution take the form

$$d(X,Y) = \sup |\mathbb{E}h(X) - \mathbb{E}h(Y)|$$

for some class of h.

- Any time a distribution can be characterized by a differential (or difference) operator, such an approach may be tried.

Question: How to bound $\mathbb{E}[f'(W) - Wf(W)]$?

Stein's method of exchangeable pairs:

Start with W, and make a small (random) change to get W' so that (W, W') is exchangeable; i.e., $(W, W') \sim (W', W)$.

Example: $W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$. Pick $I \in \{1, \dots, n\}$ at random, independent of everything. Let $W' = W - \frac{1}{\sqrt{n}} X_I + \frac{1}{\sqrt{n}} X_I'$, where X_I' is an independent copy of X_I.

Once you have W, W', fix f:

$$0 = \mathbb{E}\left[(W' - W)\,[f(W') + f(W)]\right]$$

$$= \mathbb{E}\left[(W' - W)\,[f(W') - f(W)] + 2f(W)(W' - W)\right]$$

$$\approx \mathbb{E}\left[(W' - W)^2 f'(W) + 2f(W)(W' - W)\right]$$

$$= \mathbb{E}\left[f'(W) \cdot \mathbb{E}[(W' - W)^2 \mid W] + 2f(W) \cdot \mathbb{E}[W' - W \mid W]\right]$$

Suppose you've constructed W' so that

$$\mathbb{E}[W'-W \mid W] \approx -\lambda W$$

$$\mathbb{E}[(W'-W)^2 \mid W] \approx 2\lambda$$

Then we have:

$$0 \approx \mathbb{E}[f'(W)\cdot 2\lambda - 2\lambda W f(W)]$$

$$= 2\lambda\, \mathbb{E}[f'(W) - W f(W)]$$

Stein's abstract normal approximation theorem:

Let (W,W') be an exchangeable pair, $\mathbb{E}W=0$, $\mathbb{E}W^2=1$, with $\lambda>0$ s.t. $\mathbb{E}[W'-W \mid W] = -\lambda W$. Then if $h\in C^1$, $\|h\|_\infty<\infty$, $\|h'\|_\infty<\infty$,

$$|\mathbb{E}h(W)-\mathbb{E}h(Z)| \leq \frac{1}{\lambda}\|h\|_\infty \sqrt{\operatorname{Var}(\mathbb{E}[(W'-W)^2 \mid W])}$$

$$+\frac{1}{4\lambda}\|h'\|_\infty\, \mathbb{E}|W'-W|^3.$$

CLT via the method of exchangeable pairs

$$W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i, \quad W' = W - \frac{1}{\sqrt{n}} X_I + \frac{1}{\sqrt{n}} X_I',$$

where $I \in \{1, \dots, n\}$ is random and X_i' is an independent copy of X_i for each i.

$$\Rightarrow W' - W = \frac{1}{\sqrt{n}} (X_I' - X_I)$$

$$\begin{aligned}
\mathbb{E}[W' - W \mid W] &= \frac{1}{\sqrt{n}} \mathbb{E}[X_I' - X_I \mid W] \\
&= \frac{1}{n} \mathbb{E}\Big[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i' - X_i \mid W\Big] \\
&= -\frac{1}{n} \mathbb{E}\Big[\frac{1}{\sqrt{n}} \sum_{i} X_i \mid W\Big] = -\underbrace{\frac{1}{n}}_{\boxed{\lambda}} W
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[(W' - W)^2 \mid W] &= \frac{1}{n} \mathbb{E}[(X_I')^2 - 2X_I' X_I + X_I^2 \mid W] \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[(X_i')^2 - 2X_i' X_i + X_i^2 \mid W] \\
&= \frac{1}{n} + \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[X_i^2 \mid W] \\
&= \frac{2}{n} + \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[X_i^2 - 1 \mid W]
\end{aligned}$$

$$\Rightarrow \mathbb{E}\left[\mathbb{E}\left[(W'-W)^2 | W\right]\right] = \frac{2}{n}$$

$$\mathrm{Var}\left(\mathbb{E}\left[(W'-W)^2 | W\right]\right)$$

$$= \mathbb{E}\left[\frac{1}{n^4}\ \mathbb{E}\left[\sum_{i=1}^{n}(X_i^2-1) \,|\, W\right]^2\right]$$

$$\leq \frac{1}{n^4}\ \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i^2-1)(X_j^2-1)\right]$$

$$= \frac{1}{n^3}\ \mathbb{E}\left[X_1^2-1\right]^2 = \frac{\mathbb{E}\left[X_1^4-1\right]}{n^3}$$

$$\mathbb{E}\left[|W'-W|^3\right] = \mathbb{E}\left[|X_I'-X_I|^3\right] \cdot \frac{1}{n^{3/2}}$$

$$\leq \frac{\sqrt[3]{2}}{n^{3/2}}\ \mathbb{E}|X_I|^3 = \frac{\sqrt[3]{2}\ \mathbb{E}|X_1|^3}{n^{3/2}}$$

$$\Rightarrow |\mathbb{E}h(W) - \mathbb{E}h(Z)| \leq \frac{1}{\sqrt{n}}\, \|h\|_\infty\ \mathbb{E}\left[X_1^4-1\right]$$

$$+ \frac{1}{\sqrt{n}}\, \|h'\|_\infty \left(\frac{\sqrt[3]{2}\ \mathbb{E}|X_1|^3}{4}\right).$$

Other approaches to Stein's method

Zero-bias coupling:

Given W, make W^* close to W such that $\mathbb{E}[f'(W^*)] = \mathbb{E}[Wf(W)]$

Theorem (Goldstein / Reinert)

Let (W, W^*) be defined on a joint probability space with $\mathbb{E}W=0$, $\mathbb{E}W^2=1$ and W^* defined as above. Then for $h \in C^4$,

$$\left|\mathbb{E}h(W)-\mathbb{E}h(Z)\right| \le \frac{1}{3}\|h^{(3)}\|_\infty \sqrt{\mathbb{E}\left(\mathbb{E}[W^*-W|W]^2\right)} + \frac{1}{8}\|h^{(4)}\|_\infty \mathbb{E}\,(W^*-W)^2.$$

Size-bias coupling:

<u>Theorem</u> (Goldstein / Rinott):

Let $W \geq 0$ be a random variable with $\mathbb{E}W = \lambda$, $\mathrm{Var}\, W = \sigma^2$, and define W^* by $\mathbb{E}[W g(W)] = \lambda \mathbb{E} g(W^*)$ for any g s.t. the LHS exists. Then for $h \in C_c^1$,

$$\left|\mathbb{E} h\left(\frac{W-\lambda}{\sigma}\right) - \mathbb{E} h(Z)\right| \leq 2\|h\|_\infty \frac{\lambda}{\sigma^2} \sqrt{\mathrm{Var}(\mathbb{E}[W^* - W | W]}$$

$$+ \|h'\|_\infty \frac{\lambda}{\sigma^3} \mathbb{E}(W^* - W)^2.$$

Dependency graphs:

Theorem (Stein): Let $Y_1, \ldots, Y_n$ be r.v.'s, $M_i \in \{1, \ldots, n\}$ s.t.

1. $j \in M_i$ iff $i \in M_j$
2. (Y_i, Y_j) is independent of $\{Y_k\}_{k \notin M_i \cup M_j}$.

Suppose $\mathbb{E} Y_i = 0 \ \forall i$, $\mathbb{E} Y_i^2 = 1 \ \forall i$. Then for $h \in C_c^1$,

$$\left| \mathbb{E} h\left(\frac{1}{\sqrt{n}} \sum_i Y_i\right) - \mathbb{E} h(Z) \right|$$

$$\leq \frac{2}{n} \|h\|_\infty \sqrt{\mathbb{E}\left(\sum_{i=1}^n \sum_{j \in M_i} \left(Y_i Y_j - \mathbb{E}(Y_i Y_j)\right)\right)^2}$$

$$+ \frac{1}{n^{3/2}} \|h'\|_\infty \mathbb{E}\left[\sum_{i=1}^n |Y_i| \left(\sum_{j \in M_i} Y_j \right)^2 \right].$$