Normal and Non-normal Approximation by Stein's Method

Qi-Man Shao

Hong Kong University of Science and Technology

The research is partially supported by Hong Kong RGC CERG 602608 and 603710

Outline

1. Introduction

Let W_n be a sequence of random variables of interest.

- Aim: Estimate $P(W_n \geq x)$.
- Questions:

1. What is the limiting distribution of W_n?
2. Suppose that $W_n \xrightarrow{d.} Y$. It is a common practice to use $P(Y \geq x)$ to approximate $P(W_n \geq x)$. What is the error of approximation?

   - Absolute error: Berry-Esseen type bound

$$|P(W_n \geq x) - P(Y \geq x)| = \text{error}$$

   - Relative error: Cramér type moderate deviation

$$\frac{P(W_n \geq x)}{P(Y \geq x)} = 1 + \text{error}$$

- Our focus:
  1. Identify the limiting distribution of W_n;
  2. Estimate the relative error, especially, what is the largest possible a_n such that

$$\frac{P(W_n \geq x)}{P(Y \geq x)} \to 1$$

holds uniformly in $x \in [0, a_n]$.

- In many applications, $P(Y \geq x)$ itself is very small. Only when the relative error is small, can $P(W_n \geq x)$ be approximated by $P(Y \geq x)$;
- Multiple hypothesis tests

  Consider the problem of testing simultaneously m (null) hypotheses, $H_1, H_2, \cdots, H_m$, of which m_0, are true. Let R be the number of hypotheses rejected. Table below summarizes the test results

	Declared non-significant	Declared significant	Total
True null hypotheses	U	V	m_0
Non-true null hypotheses	T	S	$m - m_0$
Total	$m - R$	R	m

- The proportion of errors committed by falsely rejecting null hypotheses: V/R

- False discovery rate (FDR): $E(V/R)$

- Benjamini-Hochberg FDR controlling procedure:
  Assume P-values are $p_1, p_2, \ldots, p_m$. Let
  $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$ be the ordered p-values, and denote by $H_{(i)}$ the null hypothesis corresponding to $p_{(i)}$. Let

  $$k = \max\{i : \; p_{(i)} \leq \frac{i}{m}\alpha\}$$

  where $0 < \alpha < 1$. Then reject all $H_{(i)}$, for $1 \leq i \leq k$.

  If the test statistics are independent, then $E(V/R) \leq \alpha$.

- P-values are usually unknown, need to be estimated.
- Question: (Fan, Hall, Yao (2007))
  How large m can be before the accuracy of the estimated P-values becomes poor?
- Korosok-Ma (2007), Fan, Hall, Yao (2007), Liu and Shao (2009), Shao (2010):
  - Let $T_{n,i}$ be the test statistic for H_i. Assume that the true P-value is $p_i = P(T_{n,i} \geq t_{n,i})$ and that there exist $a_{n,i}$ and functions f_i such that

$$\max_{1 \leq i \leq m} \sup_{0 \leq x \leq a_{n,i}} \left| \frac{P(T_{n,i} \geq x)}{f_i(x)} - 1 \right| = o(1)$$

as $n \to \infty$. If $m \leq \alpha/(2 \max_{1 \leq i \leq m} f_i(a_{n,i}))$, then the FDR is controlled at level α when it is based on the estimated P-values $\hat{p}_i = f_i(t_{n,i})$.

- How to identify the limiting distribution and estimate the relative error?

Two approaches:

- Classical and standard method: Fourier transform.

  It works well when W_n is a sum of independent random variables, however, it may be very difficult to apply for W_n under dependence structure.

- Stein's method (1972):
  A totally different approach. It works not only for independent variables but also for dependent variables. It can also give bounds for accuracy of approximation.

2. Stein's method: normal approximation

Let $Z \sim N(0,1)$, and let $\mathcal{C}_{bd}$ be the set of continuous and piecewise continuously differential functions $f : R \to R$ with $E|f'(Z)| < \infty$. Stein's method rests on the following observation.

- Stein's identity:
  $W \sim N(0,1)$ if and only if

$$Ef'(W) - EWf(W) = 0$$

  for any $f \in \mathcal{C}_{bd}$.

- Stein's equation:

$$f'(w) - wf(w) = I_{\{w \le z\}} - \Phi(z).$$

where $z \in R$ is fixed.

Solution to the equation:

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

- The general Stein equation:

  Let h be a real valued measurable function with $E|h(Z)| < \infty$.

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

  The solution $f = f_h$ is given by

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

- Basic properties of the Stein solution:
  - If h is bounded, then

$$\|f_h\| \leq 2\|h\|, \quad \|f_h'\| \leq 4\|h\|.$$

  - If h is absolutely continuous, then

$$\|f_h\| \leq 2\|h'\|, \quad \|f_h'\| \leq \|h'\|, \quad \|f_h''\| \leq 2\|h'\|.$$

- Main idea of Stein's approach:

Suppose that $W := W_n$ is the variable of interest and our goal is to estimate

$$Eh(W) - Eh(Z).$$

By Stein's equation, we have

$$Eh(W) - Eh(Z) = Ef'(W) - EWf(W)$$

A key step in Stein's approach is to write $EWf(W)$ as close as possible to $Ef'(W)$.

Suppose that there exist $\hat{K}(t)$ and R such that the following general Stein's identity holds

$$EWf(W) = E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt + ERf(W).$$

Then

$$\begin{aligned} Eh(W) - Eh(Z) &= Ef_h'(W) - EWf_h(W) \\ &= E\int_{-\infty}^{\infty} (f_h'(W) - f_h'(W+t))\hat{K}(t)dt \\ &\quad + Ef_h'(W)(1-\hat{K}_1) - ERf_h(W), \end{aligned}$$

where $\hat{K}_1 = E\Big(\int_{-\infty}^{\infty} \hat{K}(t)dt \mid W \Big)$. In particular, if $\|h'\| < \infty$, then

$$|Eh(W) - Eh(Z)| \le 2\|h'\| \Big(E\int |t\hat{K}(t)|dt + E|1-\hat{K}_1| + E|R| \Big).$$

- Stein's method has been applied to
  - Normal approximation:
    1. Stein (1972, 1986): Uniform Berry-Esseen inequality for i.i.d. random variables
    2. Chen and Shao (2001): Non-uniform Berry-Esseen inequality for independent random variables
    3. Chen and Shao (2004): Uniform and non-uniform Berry-Esseen inequality under local dependence
    4. Chen and Shao (2007): Uniform and non-uniform Berry-Esseen inequality for non-linear statistics
    5. Bolthausen (1984), Bolthausen and Götze (1993), Bladi and Rinott (1989), Rinott and Rotar (1997), Goldstein and Reinert (1997), Chatterjee (2008), ...
    6. Chen, L.H.Y, Goldstein, L. and Shao (2010). Normal Approximation by Stein's Method. Springer.

- Non-normal approximation:
    1. Poisson approximation: Chen (1975), Arratia, Goldstein and Gordon (1989), Barbour, Holst and Janson (1992), Chatterjee, Diaconis and Meckes (2005), ...
    2. Compound Poisson approximation: Barbour, Chen and Loh (1992), Erhardsson (2003), ...
    3. Poisson process approximation: Xia (2003), ...
    4. Peccati (2009): Malliavin calculus
    5. Chatterjee (2007, 2008, 2009): Concentration inequality, strong approximation, random matrix theory, ...

3. Stein's Method: beyond the normal approximation

Let Y be a random variable with pdf $p(y)$. Assume that $p(-\infty) = p(\infty) = 0$ and p is differentiable. Observe that

$$E\left\{\frac{\big(f(Y)p(Y)\big)'}{p(Y)}\right\} = \int (f(y)p(y))' dy = 0$$

- Stein's identity and equation (Stein, Diaconis, Holmes, Reinert (2004)):
  - Stein's identity:

$$Ef'(Y) + Ef(Y)p'(Y)/p(Y) = 0.$$

  - Stein's equation:

$$f'(y) + f(y)p'(y)/p(y) = h(y) - Eh(Y)$$

  - Stein's solution:

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

- Properties of the solution (Chatterjee and Shao (2011)):

  Let h be a measurable function and f_h be the Stein's solution. Under some regular conditions on p

$$\|f_h\| \leq C\|h\|, \quad \|f_h'\| \leq C\|h\|,$$

$$\|f_h\| \leq C\|h'\|, \quad \|f_h'\| \leq C\|h'\|, \quad \|f_h''\| \leq C\|h'\|$$

▶ Identify the limiting distribution

Let $W := W_n$ be the random variable of interest. Our goal is to identify the limiting distribution of W_n with an error of approximation.

Exchangeable pair approach:

Let (W, W^*) be an exchangeable pair. Assume that

$$E(W - W^* \mid W) = g(W) + r(W)$$

Let

$$G(t) = \int_0^t g(s)ds \text{ and } p(t) = c_1 e^{-c_0 G(t)},$$

where $c_0 > 0$ and $c_1 = 1/\int_{-\infty}^{\infty} e^{-c_0 G(t)} dt$.

Let Y have pdf $p(y)$ and $\Delta = W - W^*$.

Chatterjee and Shao (2011): Under some regular conditions on g

- Assume that $c_0 E|r| \to 0$, $c_0 E|\Delta|^3 \to 0$ and

$$c_0 E(\Delta^2|W) \xrightarrow{p.} 2.$$

  Then

$$W \xrightarrow{d.} Y .$$

- If $|\Delta| \leq \delta$, then

$$\begin{aligned} & |P(W \geq x) - P(Y \geq x)| \\ = \ & O(1)\Big(E|1 - (c_0/2)E(\Delta^2|W)| + c_0\delta^3 + \delta + c_0 E|r(W)|\Big) . \end{aligned}$$

▶ Application to the Curie-Weiss model at the critical temperature

The Curie-Weiss model of ferromagnetic interaction is a simple statistical mechanical model of spin systems.

Let $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n) \in \{-1, 1\}^n$. The joint density function of σ is given by

$$A_\beta^{-1} \exp(\beta \sum_{1 \le i < j \le n} \sigma_i \sigma_j \,/\, n),$$

where β is called the inverse of temperature.
Let $\beta = 1$ and

$$W = \frac{1}{n^{3/4}} \sum_{i=1}^{n} \sigma_i$$

- Ellis and Newman (1978):

$$W \xrightarrow{d.} Y,$$

where Y has pdf $c\ e^{-y^4/12}$.

- Chatterjee and Shao (2011):

$$|P(W \geq x) - P(Y \geq x)| = O(n^{-1/2})$$

by constructing an exchangeable pair (W, W^*) such that

$$E(W - W^*|W) = \frac{1}{3}n^{-3/2}W^3 + O(n^{-2}),$$

$$E((W - W^*)^2|W) = 2n^{-3/2} + O(n^{-2}),$$

$$|W^* - W| = O(n^{-3/4}).$$

4. Cramér type moderate deviations

Let W_n be a sequence of random variables of interest. Assume that

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

Our goal is to estimate the relative error

$$\frac{P(W_n \geq x)}{1-\Phi(x)} = 1 + \text{error}$$

▶ Classical Cramér moderate deviation

Let $X_1, X_2, \cdots, X_n$ be independent identically distributed (i.i.d.) random variables with $EX_i = 0$ and $\sigma^2 = EX_i^2 < \infty$. Let

$$W_n = \frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}}$$

- If $Ee^{t_0\sqrt{|X_1|}} < \infty$ for $t_0 > 0$, then

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

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

$$\frac{P(W_n \geq x)}{1-\Phi(x)} = 1 + O(1)\frac{(1+x^3)}{\sqrt{n}}$$

for $x \in [0, O(n^{1/6}))$.

▶ A Cramér type moderate deviation under Stein's identity

Theorem (Chen, Fang, Shao (2009))

Let $W = W_n$. Assume that there exist a constant δ and random functions $\hat{K}(t) \geq 0$ and R such that

$$EWf(W) = E\int_{|t|\leq\delta} f'(W+t)\hat{K}(t)dt + E(Rf(W))$$

for all nice function f. Let $D = \int_{|t|\leq\delta}\hat{K}(t)dt$. If there exist constants d_0, δ_1, δ_2 such that

$$E(D|W) \leq d_0,$$
$$|E(D|W) - 1| \leq \delta_1(1+|W|), \quad |E(R \mid W)| \leq \delta_2(1+|W|).$$

Then

$$\frac{P(W \geq x)}{1-\Phi(x)} = 1 + O(1)d_0^3(1+x^3)\left(\delta+\delta_1+\delta_2\right)$$

for $0 \leq x \leq d_0^{-1}\min\left(\delta^{-1/3}, \delta_1^{-1/3}, \delta_2^{-1/3}\right)$.

- A special case: zero-bias approach
  - Goldstein and Reiner (1997): For any W with $EW = 0$ and $EW^2 = 1$, there exists a random variable Δ such that

$$EWf(W) = Ef'(W + \Delta).$$

    for any nice function f.

  - We can take $\delta_1 = \delta_2 = 0$ in the above general theorem. If $|\Delta| \le \delta$, then

$$\frac{P(W \ge x)}{1 - \Phi(x)} = 1 + O(1)\delta(1 + x^3)$$

    for $0 \le x \le \delta^{-1/3}$.

Applications

- Combinatorial central limit theorem

  Let $\{a_{ij}\}_{i,j=1}^n$ be an array of real numbers satisfying $\sum_{j=1}^n a_{ij} = 0$ for all i. Set $c_0 = \max_{i,j} |a_{ij}|$ and $W = \sum_{i=1}^n a_{i\pi(i)}/\sigma$, where π is a uniform random permutation of $\{1, 2, \cdots, n\}$ and $\sigma^2 = E(\sum_{i=1}^n a_{i\pi(i)})^2$.

  It is proved in Goldstein (2005) that there exists a random variable $|\Delta| \le 8c_0/\sigma$ such that $EWf(W) = Ef'(W + \Delta)$. Therefore,

$$\frac{P(W \ge x)}{1 - \Phi(x)} = 1 + O(1)(1 + x^3)c_0/\sigma$$

  for $0 \le x \le (\sigma/c_0)^{1/3}$.

- Binary expansion of a random integer

  Let X be an integer uniformly chosen from $\{0, 1, \cdots, n\}$. Let k be such that $2^{k-1} < n \leq 2^k$. Write the binary expansion of X as

$$X = \sum_{i=1}^{k} X_i 2^{k-i}$$

  and let $S = X_1 + \cdots + X_k$ be the number of ones in the binary expansion of X. Put $W = (S - k/2)/\sqrt{k/4}$. Then

$$\frac{P(W \geq x)}{1 - \Phi(x)} = 1 + O(1)(1 + x^3)/\sqrt{k}$$

  for $0 \leq x \leq k^{1/6}$.

- Cuire-Weiss model

Let $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n) \in \{-1, 1\}^n$. Recall the joint density function of σ is given by

$$A_\beta^{-1} \exp(\beta \sum_{1 \le i < j \le n} \sigma_i \sigma_j \, /n).$$

Let

$$W = \sum_{i=1}^{n} \sigma_i / B, \quad \text{where} \quad B^2 = Var(\sum_{i=1}^{n} \sigma_i).$$

Ellis and Newman (1978): the limiting distribution of W is normal when $0 < \beta < 1$.

Chen, Fang and Shao (2009): For $0 < \beta < 1$

$$\frac{P(W \ge x)}{1 - \Phi(x)} = 1 + O(1)(1 + x^3)/\sqrt{n}$$

for $0 \le x \le n^{1/6}$.

5. Cramér type moderate deviation for Studentized U-statistics

Let $X, X_1, X_2, \dots, X_n$ be i.i.d random variables, and let $h(x, y)$ be a symmetric kernel, i.e., $h(x, y) = h(y, x)$. $\theta = Eh(X_1, X_2)$.

U-statistic (Hoeffding (1948)):

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

The standardized U-statistic:

$$\frac{\sqrt{n}}{2\,\sigma_1}(U_n - \theta).$$

where $\sigma_1^2 := \mathrm{Var}(g(X)) > 0$ and $g(x) = E(h(x, X))$.

Studentized U-statistic:

$$T_n = \frac{\sqrt{n}}{2\,s_1}(U_n - \theta),$$

where

$$s_1^2 = \frac{(n-1)}{(n-2)^2}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i} h(X_i, X_j) - U_n\right)^2.$$

Hoeffding's decomposition: (assume $\theta = 0$)

$$T_n = \frac{W_n + D_1}{V_n(1+D_2)^{1/2}},$$

where

$$W_n = \sum_{i=1}^{n}\xi_i,\ \xi_i = g(X_i)/(\sigma_1\sqrt{n}),\ V_n^2 = \sum_{i=1}^{n}\xi_i^2,$$

D_1 and D_2 are small.

- Berry-Esseen bounds: Callaert and Veraverbeke (1981), Zhao (1983), Wang, Jing and Zhao (2000), ...
- Cramér type moderate deviations: Vandemaele and Veraverbeke (1985), Wang (1998), Lai, Shao and Wang (2009)

- Lai, Shao and Wang (2009):
  Assume that $\sigma_1 > 0$ and $E|h(X_1, X_2)|^3 < \infty$. If

$$h^2(x_1, x_2) \le c_0(\sigma_1^2 + g^2(x_1) + g^2(x_2))$$

for some $c_0 > 0$, then

$$\frac{P(T_n \ge x)}{1 - \Phi(x)} \to 1$$

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

  - Conjecture:

$$\frac{P(T_n \ge x)}{1 - \Phi(x)} = 1 + O(1)\frac{(1+x)^3}{\sqrt{n}}$$

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

- Shao and Zhou (2011): The conjecture is true. Similar result holds for $h(x_1, x_2, \cdots, x_m)$.

6. Cramér type deviations for Studentized non-linear statistics

Let $\xi_1, ..., \xi_n$ be independent random variables with $E\xi_i = 0$ and $E\xi_i^2 < \infty$ satisfying

$$\sum_{i=1}^{n} E\xi_i^2 = 1.$$

Let

$$W_n = \sum_{i=1}^{n} \xi_i, \quad V_n^2 = \sum_{i=1}^{n} \xi_i^2$$

and D_1, D_2 be measurable functions of $\{\xi_i, 1 \le i \le n\}$.

Assume

$$T_n = \frac{W_n + D_1}{V_n(1 + D_2)^{1/2}}.$$

Theorem (Shao and Zhou (2011))

There is an absolute constant $A > 1$ such that

$$e^{O(1)\Delta_{n,x}}\big(1 - AR_{n,x}\big) \le \frac{P(T_n \ge x)}{1-\Phi(x)}$$

and

$$\begin{aligned} P(T_n \ge x) &\le \big(1-\Phi(x)\big)e^{O(1)\Delta_{n,x}}\big(1 + AR_{n,x}\big) \\ &\quad + P\big(|D_1|/V_n > 1/(2x)\big) + P\big(|D_2| > 1/(2x^2)\big) \end{aligned}$$

for all $x > 1$ satisfying

$$\Delta_{n,x} \le (1+x)^2/A,\ x^2 \max_{1\le i\le n} E\xi_i^2 \le 1,$$

where

$$\begin{aligned}
\Delta_{n,x} &= x^2 \sum_{i=1}^{n} E\xi_i^2 I(x|\xi_i| > 1) + x^3 \sum_{i=1}^{n} E|\xi_i|^3 I(x|\xi_i| \le 1), \\
R_{n,x} &= I_{n,0}^{-1} \Big\{ xE(|D_1| + x|D_2|) e^{\sum_{j=1}^{n}(x\xi_j - x^2\xi_j^2/2)} \\
&\quad + x \sum_{i=1}^{n} E(|\xi_i(D_1 - D_1^{(i)})| + x|\xi_i(D_2 - D_2^{(i)})|) e^{\sum_{j\ne i}^{n}(x\xi_j - x^2\xi_j^2/2)} \Big\}, \\
I_{n,0} &= \prod_{i=1}^{n} E e^{x\xi_i - x^2\xi_i^2/2},
\end{aligned}$$

and $D_1^{(i)}$ and $D_2^{(i)}$ are any random variables that don't depend on ξ_i.

▶ Main idea of the proof

Observe that

$$1 + s/2 - s^2/2 \le (1+s)^{1/2} \le 1 + s/2, \quad s \ge -1,$$

$$\begin{aligned} V_n(1+D_2)^{1/2} &\ge (1 + V_n^2 - 1)^{1/2}(1 - \min(1, |D_2|)) \\ &\ge V_n^2/2 + 1/2 - (V_n^2 - 1)^2 - 2|D_2| \end{aligned}$$

and

$$V_n(1+D_2)^{1/2} \le (1+D_2)/2 + V_n^2/2 = V_n^2/2 + 1/2 + D_2/2.$$

Therefore for any $x \ge 0$,

$$\{T_n \ge x\} \quad \subset \quad \{xW_n - x^2V_n^2/2 \ge x^2/2 - x(x(V_n^2-1)^2 + D_1 + 2x|D_2|)\}$$

$$\{T_n \ge x\} \supset \{xW_n - x^2V_n^2/2 \ge x^2/2 + x(xD_2/2 - D_1)\}.$$

Then, use the conjugate method and apply the following randomized concentration inequality.

Theorem (Shao and Zhou (2011))

Let Δ_1 *and* Δ_2 *be any measurable functions of* $\{\xi_i, 1 \le i \le n\}$. *Then*

$$\begin{aligned} & P(\Delta_1 \le W_n \le \Delta_2) \\ \le \; & 21(\beta_2 + \beta_3) + 6E|\Delta_2 - \Delta_1| \\ & +4\sum_{i=1}^{n}\{E|\xi_i(\Delta_1 - \Delta_1^{(i)})| + E|\xi_i(\Delta - \Delta_2^{(i)})|\} \end{aligned}$$

where

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

$\Delta_1^{(i)}$ *and* $\Delta_2^{(i)}$ *are any random variables that don't depend on* ξ_i.

Conclusion

Stein's method is a powerful tool for normal and non-normal approximation. It can be applied to obtain Berry-Esseen type bounds as well as Cramér type moderate deviations. The method is of unlimited usefulness.

References

- Chen, L.H.Y, Goldstein, L. and Shao (2010). Normal Approximation by Stein's Method. Springer.
- Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Stat. Prob.* **2,** 583-602, Univ. California Press. Berkeley, Calif.
- Stein, C. (1986) *Approximation Computation of Expectations.* Lecture Notes 7, Inst. Math. Statist., Hayward, Calif.
- Chen, L.H.Y. and Shao, Q.M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13**, 581-599.
- Chatterjee, S. and Shao, Q.M. (2011). Non-normal approximation by Stein's method of exchangeable pairs with application to the Curie-Weiss model. *Ann. Appl. Probab.* **21**, 464-483.

THANK YOU!