Stein's Method: Identifying the Limiting Distribution

Qi-Man Shao

The Chinese University of Hong Kong

qmshao@cuhk.edu.hk

The research is partially supported by Hong Kong
RGC GRF 603710 and 403513

Outline

Outline

1. Introduction

Let W_n be a random variable of interest.

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

1. Introduction

Let W_n be a random variable of interest.

- Aim: Estimate $P(W_n \geq x)$.
- Questions:
  1. What is the limiting distribution of W_n?

1. Introduction

Let W_n be a random variable 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?

1. Introduction

Let W_n be a random variable 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

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

1. Introduction

Let W_n be a random variable 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 absolute error

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

- How to identify the limiting distribution and estimate the 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 use under dependence structure.

▶ How to identify the limiting distribution and estimate the 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 use 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 provide accuracy of the 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:

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

where $z \in R$ is fixed.

Solution to the equation:

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

- 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

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

  - If h is absolutely continuous, then

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

- 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

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

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 $\hat{K}_1 = E\Big(\int_{-\infty}^{\infty} \hat{K}(t)dt \mid W \Big)$.

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 (2011). Normal Approximation by Stein's Method. Springer.
7. Chen, Fang, Shao (2013). Cramér type moderate deviations

- 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

$$E\left\{ \frac{\big(f(Y)p(Y)\big)'}{p(Y)} \right\} = \int_{-\infty}^{\infty} (f(y)p(y))' dy = 0$$

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

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

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

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

  - Stein's equation:

$$f'(y) + f(y)p'(y)/p(y) = h(y) - Eh(Y) \tag{1}$$

  - 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.

▶ 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

$$E(W - W^* \mid W) = g(W) + r_1(W)$$

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 $\Delta = W - W^*$.

Theorem (Chatterjee and Shao (2011))

Under some regular conditions on g

- *Assume that* $c_0 E|r_1(W)| \to 0$, $c_0 E|\Delta|^3 \to 0$ *and*

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

Then

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

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

Theorem (Chatterjee and Shao (2011))

Under some regular conditions on g

- *Assume that* $c_0 E|r_1(W)| \to 0$, $c_0 E|\Delta|^3 \to 0$ *and*

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

  Then

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

- *If* $|\Delta| \le \delta$, *then*

$$\begin{aligned} & |P(W \ge x) - P(Y \ge x)| \\ = \ & O(1)\Big(E|1 - (c_0/2)E(\Delta^2|W)| + c_0\delta^3 + \delta + c_0 E|r_1(W)|\Big)\ . \end{aligned}$$

▶ How was the limiting distribution identified?

Observe that for any absolutely continuous function f

$$
\begin{aligned}
0 &= E(W - W^*)(f(W^*) + f(W)) \\
&= 2Ef(W)(W - W^*) + E(W - W^*)(f(W^*) - f(W)) \\
&= 2E\{f(W)E((W - W^*)|W)\} - E(W - W^*) \int_{-\Delta}^{0} f'(W + t)dt \\
&= 2Ef(W)g(W) + 2Ef(W)r_1(W) - E \int_{-\infty}^{\infty} f'(W + t)\hat{K}(t)dt,
\end{aligned}
$$

where

$$
\hat{K}(t) = E\{\Delta(I\{-\Delta \leq t \leq 0\} - I\{0 < t \leq -\Delta\})|W\}.
$$

Thus, we have

$$Ef(W)g(W) = \frac{1}{2}E\int_{-\infty}^{\infty} f'(W+t)\hat{K}(t)dt - Ef(W)r_1(W). \qquad (3)$$

Recall the Stein equation

$$Eh(W) - Eh(Y) = Ef'(W) + Ef(W)p'(W)/p(W) \qquad (4)$$

Comparing (4) with (3), one should choose

$$p'(w)/p(w) = -c_0 g(w)$$

▶ 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 $c_1 \, e^{-y^4/12}$, where $c_1 = 2^{1/2}/(3^{1/4}\Gamma(1/4))$.

- 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

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

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

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

Let $W := W_n$ be the random variable of interest. Recall in Theorem (C-S (2011)), a key assumption is

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

Question: Can the above assumption be removed?

- Exchangeable pair approach:

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

$$E(W - W^* \mid W) = g(W) + r_1(W)$$

and

$$E((W - W^*)^2 | W) = v(W) + r_2(W)$$

Let

$$g_v(w) = 2g(w)/v(w), \;\; G_v(w) = \int_0^w g_v(t)dt.$$

Put

$$p(y) = c_1^* \exp(-G_v(y)), \quad c_1^* = \frac{1}{\int_{-\infty}^{\infty} \exp(-G_v(y))dy}$$

Let Y be a random variable with pdf $p(y)$.

Let $\Delta = W - W^*$ and $h_v(w) = h(w)/v(w)$.

Theorem (Shao (2014))

Under some regular conditions for g and v.

(i) *For absolutely continuous function h*

$$|Eh(W) - Ev(W)Eh_v(Y)| \le C\|h_v'\|(E|\Delta|^3 + E|r_1(W)| + E|r_2(W)|)$$

(ii) *If $|\Delta| \le \delta$, $v(w) \ge c_2$ and $|v'(w)/v(w)| \le c_3$, then*

$$\begin{aligned} &|P(W \le z) - E\Delta^2 E(I(Y \le z)/v(Y))| \\ &\le \frac{C\delta^3}{c_2}(1 + c_3 + E|g_v(W)|) + \frac{C}{c_2}(E|r_1(W)| + E|r_2(W)|) \end{aligned}$$

5. Application to the Curie-Weiss model at the critical temperature

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

$$A \exp\Big(\sum_{1 \le i < j \le n} \sigma_i \sigma_j \,/\, n \Big).$$

Let

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

Recall

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

where the p.d.f. of Y is given by $c_1 \exp(-y^4/12)$.

Observe that $|W - W^*| \le 2n^{-3/4}$,

$$E(W - W^*|W) = n^{-3/2}(\frac{1}{3}W^3 - n^{-1/2}W) + O(n^{-5/2})W^3,$$

$$E((W' - W)^2|W) = 2n^{-3/2}(1 - n^{-1/2}W^2) + O(n^{-5/2})(1 + W^4),$$

Set

$$g(w) = n^{-3/2}(\frac{1}{3}w^3 - n^{-1/2}w)$$

and

$$v(w) = 2n^{-3/2}(1 - n^{-1/2}w^2).$$

Applying the general result, we have

Theorem

$$|P(W \leq z) - F(z)| = O(n^{-3/4}),$$

where

$$F(z) = c_1 \int_{-\infty}^{z} (1 + n^{-1/2}c_0(t))e^{-t^4/12}\, dt,$$

$$c_0(w) = -\frac{6\sqrt{3}\,\Gamma(3/4)}{\Gamma(1/4)} + w + \frac{w^2}{2} - \frac{w^5}{15}.$$

6. Application to the random determinant?

Let $M_n = (X_{ij})_{n\times n}$ be a random matrix. Assume that X_{ij} are i.i.d. with $EX_{ij} = 0$ and $EX_{ij}^2 = 1$.

Some well-known facts:

- $E\det(M_n^2) = n!$

6. Application to the random determinant?

Let $M_n = (X_{ij})_{n\times n}$ be a random matrix. Assume that X_{ij} are i.i.d. with $EX_{ij} = 0$ and $EX_{ij}^2 = 1$.

Some well-known facts:

- $E\,\det(M_n^2) = n!$
- If $X_{ij} \sim N(0,1)$, then

$$\det(M_n^2) \stackrel{d.}{=} \prod_{j=1}^{n} \eta_j$$

where η_j are independent with χ_j^2 distribution.

- The central limit theorem
  - Girko (1997): Claimed that if $E|X_{ij}|^{4+\delta} < \infty$ for some $\delta > 0$, then

$$\frac{\log\det(M_n^2) - \log(n-1)!}{\sqrt{2\log n}} \xrightarrow{d} N(0,1)$$

- The central limit theorem

  - Girko (1997): Claimed that if $E|X_{ij}|^{4+\delta} < \infty$ for some $\delta > 0$, then
$$\frac{\log\det(M_n^2) - \log(n-1)!}{\sqrt{2\log n}} \xrightarrow{d} N(0,1)$$
  - Tao and Vu (2012): "*there are several points which are not clear in these papers*" (Girko, 1979, 1997)

  - Costello and Vu (2009): "*We believe that this statement is true, but could not understand Girko's proof.*"

- Nguyen and Vu (2012): If

$$P(|X_{ij}| > t) \leq c_2 \exp(-t^{c_1}), \;\; c_1 > 0, c_2 > 0$$

for all $t > 0$, then

$$|P\Big(\frac{\log \det(M_n)^2 - \log((n-1)!)}{\sqrt{2\log n}} \leq x\Big) - \Phi(x)| \\ \leq \;\; \log^{-1/3+o(1)} n$$

Can the general theorem be applied to prove the above conjecture?

References

- Chen, L.H.Y, Goldstein, L. and Shao (2011). 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.
- 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