Stein’s Concentration Inequality for Proving the Berry-Esseen Theorem

Louis H. Y. Chen
National University of Singapore

Symposium in Memory of Charles Stein [1920 - 2016]
17 - 28 June 2019
Institute for Mathematical Sciences
Singapore
Outline

- Stein’s Proof of Berry-Esseen Theorem in the I.I.D. Case
- Proof of Berry-Esseen Theorem in the Independent but Non-identically Distributed Case
- Concentration Inequalities in Various Settings
- Without Concentration Inequality
- Photo Gallery
- Poem Dedicated to Charles Stein
Theorem 1

Let $X_1, \cdots, X_n$ be independent r.v.'s with $E X_i = 0$, $\Var(X_i) = \sigma_i^2$ and $E |X_i|^3 = \gamma_i < \infty$. Let $B^2 = \Var(\sum_{i=1}^n X_i)$, i.e. $B^2 = \sum_{i=1}^n \sigma_i^2$. Define $W = (\sum_{i=1}^n X_i)/B$ and let $\Phi$ be the d.f. of $\mathcal{N}(0, 1)$. Then

$$\sup_{x \in \mathbb{R}} |P(W \leq x) - \Phi(x)| \leq \frac{C \sum_{i=1}^n \gamma_i}{B^3}$$

where $C$ is an absolute constant.

Corollary 2

If the $X_i$ are i.i.d with $\sigma_i^2 = \sigma^2$ and $\gamma_i = \gamma$, then

$$\sup_{x \in \mathbb{R}} |P(W \leq x) - \Phi(x)| \leq \frac{C \gamma}{\sigma^3 \sqrt{n}}.$$
Stein’s Proof of the Berry-Esseen Theorem (1960’s)

- $X_1, \ldots, X_n$ i.i.d. with $\mathbb{E}X_i = 0$, $\text{Var}(X_i) = 1/n$ and $\mathbb{E}|X_i|^3 < \infty$.
- Let $W_k = \sum_{i=1}^{k} X_i$. So $\text{Var}(W_n) = 1$.
- Let $f$ be absolutely continuous and bounded with bounded $f'$.

$$
\mathbb{E}W_n f(W_n) = \sum_{i=1}^{n} \mathbb{E}X_i f(W_n) = n\mathbb{E}X_n f(W_{n-1} + X_n)
$$

$$
= n\mathbb{E}X_n [f(W_{n-1} + X_n) - f(W_{n-1})]
$$

$$
= n\mathbb{E}X_n \int_{0}^{X_n} f'(W_{n-1} + t) dt
$$

$$
= \mathbb{E} \int_{-\infty}^{\infty} f'(W_{n-1} + t) K(t) dt
$$

where $K(t) = n\mathbb{E}X_n [I(X_n > t > 0) - I(X_n \leq t \leq 0)]$.

- Call this equation a Sten identity.
Stein’s Proof of the Berry-Esseen Theorem (1960’s)

- $K$ is a probability density. Let $T \sim K$ and be independent of $X_1, \cdots, X_n$. Then $\mathbb{E}|T| = n\mathbb{E}|X_1|^3/2 = O(1/\sqrt{n})$.

- The Stein identity can be written as $\mathbb{E}W_n f(W_n) = \mathbb{E}f'(W_{n-1} + T)$.

- Let $f_x$ be the unique bounded solution of the equation $f'(w) - wf(w) = I(w \leq x) - \Phi(x)$. Then

$$P(W_n \leq x) - \Phi(x) = \mathbb{E}[f'_x(W_n) - W_n f_x(W_n)]$$

$$= \mathbb{E}[f'_x(W_n) - f'_x(W_{n-1} + T)]$$

$$= P(x - T < W_{n-1} \leq x - X_n, X_n \leq T)$$

$$- P(x - X_n < W_{n-1} \leq x - T, X_n > T)$$

$$+ \mathbb{E}[W_n f_x(W_n) - (W_{n-1} + T)f_x(W_{n-1} + T)].$$
From above

\[ |P(W_n \leq x) - \Phi(x)| \]
\[ \leq P(x - \max(X_n, T) \leq W_{n-1} \leq x - \min(X_n, T)) \]
\[ + \mathbb{E}|W_n f_x(W_n) - (W_{n-1} + T)f_x(W_{n-1} + T)| \]
\[ \leq P(x - \max(X_n, T) \leq W_{n-1} \leq x - \min(X_n, T)) \]
\[ + 2\mathbb{E}|X_n| + 2\mathbb{E}|T| \]
\[ = \mathbb{E}P(x - \max(X_n, T) \leq W_{n-1} \leq x - \min(X_n, T)|X_n, T) \]
\[ + 3n\mathbb{E}|X_1|^3. \]
Let $a < b$ and $\delta > 0$. Take $f = \ldots$

Then $|f| \leq \frac{b - a + 2\delta}{2}$ and $f'(w) = I(a - \delta \leq w \leq b + \delta)$. 

---

Stein's Concentration Inequality

Stein Memorial Symposium (1960’s)
Stein’s Proof of the Berry-Esseen Theorem (1960’s)

Stein’s concentration inequality.

Then $\mathbb{E}W_n f(W_n) = \mathbb{E} f'(W_{n-1} + T)$ yields

$$\frac{b - a + 2\delta}{2} \geq \mathbb{E}W_n f(W_n) = P(a - \delta \leq W_{n-1} + T \leq b + \delta)$$

$$\geq P(a \leq W_{n-1} \leq b, -\delta \leq T \leq \delta)$$

$$= P(a \leq W_{n-1} \leq b) P(|T| \leq \delta)$$

$$= P(a \leq W_{n-1} \leq b) (1 - P(|T| > \delta))$$

$$\geq P(a \leq W_{n-1} \leq b) (1 - \mathbb{E}|T|/\delta)$$

Letting $\delta = 2\mathbb{E}|T|$, $P(a \leq W_{n-1} \leq b)$

$$\leq \frac{b - a + 4\mathbb{E}|T|}{2} = b - a + 2n\mathbb{E}|X_1|^3.$$
Therefore

\[
|P(W_n \leq x) - \Phi(x)| = \mathbb{E} P(x - \max(X_n, T) \leq W_{n-1} \leq x - \min(X_n, T)|X_n, T) + 3n\mathbb{E}|X_1|^3 \\
\leq \mathbb{E}[\max(X_n, T) - \min(X_n, T)] + 2n\mathbb{E}|X_1|^3 \\
+ 3n\mathbb{E}|X_1|^3 \\
\leq \mathbb{E}[|X_n| + |T|] + 5n\mathbb{E}|X_1|^3 \\
\leq 6.5n\mathbb{E}|X_1|^3.
\]

It follows that

\[
\sup_{x \in \mathbb{R}} |P(W \leq x) - \Phi(x)| \leq 6.5n\mathbb{E}|X_1|^3.
\]

Independent but Non-identically Distributed

From Gothenburg to Uppsala 1979.

- $X_1, \cdots, X_n$ independent but non-identically distributed with $\mathbb{E}X_i = 0$, $\text{Var}(X_i) = \sigma_i^2$ and $\mathbb{E}|X_i|^3 = \gamma_i < \infty$.

- Let $W = \sum_{i=1}^{n} X_i$ and let $W^{(k)} = W - X_k$.

- Let $X_1', \cdots, X_n'$ be an independent copy of $X_1, \cdots, X_n$. For $i \neq l$, $\mathcal{L}(X_i, X_i', W^{(i)}|X_l) = \mathcal{L}(X_i', X_i, W^{(i)}|X_l)$.

- Denote $\mathbb{E}(\cdot | X_l)$ by $\mathbb{E}_l$.

- Then $\sum_{i=1,i \neq l}^{n} \mathbb{E}_l X_i f(W + X_i') = \sum_{i=1,i \neq l}^{n} \mathbb{E}_l X_i' f(W + X_i')$. 

Independent but Non-identically Distributed

From above

$$
\mathbb{E}_l \int f'(W + t) K(t) dt = \mathbb{E}_l W^{(l)} f(W)
$$

$$
+ \mathbb{E}_l \int f'(W + t) M^{(l)}(t) dt + \mathbb{E}_l \int f'(W + t) K_l(t) dt.
$$

where

$$
K_i(t) = \mathbb{E} X_i [I(X_i > t > 0) - I(X_i \leq t \leq 0)],
$$

$$
M_i(t) = X_i [I(X_i > t > 0) - I(X_i \leq t \leq 0)],
$$

$$
K(t) = \sum_{i=1}^{n} K_i(t),
$$

$$
M^{(l)}(t) = \sum_{i=1, i \neq l}^{n} M_i(t).
$$

$K$ is a probability density.
Letting $f$ be as before, with $\delta = \sum_{i=1}^{n} \gamma_i$.

\[
\frac{1}{2} P(a \leq W \leq b \mid X_l) \leq \int_{|t| \leq \delta} K(t)dt P(a \leq W \leq b \mid X_l)
\]

\[
\leq \mathbb{E}_l \int I(a - \delta \leq W + t \leq b + \delta)K(t)dt
\]

\[
= \mathbb{E}_l \int f'(W + t)K(t)dt
\]

\[
\leq \mathbb{E}_l W^{(l)} f(W) + \frac{1}{2} \mathbb{E}_l \int f'(W + t)^2 dt + \frac{1}{2} \mathbb{E}_l \int M^{(l)}(t)^2 dt
\]

\[
+ \frac{1}{2} \mathbb{E}_l \int f'(W + t)^2 dt + \frac{1}{2} \mathbb{E}_l \int K_l(t)^2 dt
\]

\[
\leq \frac{3(b - a)}{2} + \frac{7}{2} \delta \text{ a.s.} \quad [xy \leq \frac{1}{2} x^2 + \frac{1}{2} y^2]
\]
Independent but Non-identically Distributed

- This gives

$$P(a \leq W \leq b \mid X_t) \leq 3(b - a) + 7\delta \text{ a.s.}$$

which of course implies

$$P(a \leq W \leq b) \leq 3(b - a) + 7\delta.$$ 

- The final result

$$\sup_{x \in \mathbb{R}} |P(W \leq x) - \Phi(x)| \leq 19 \sum_{i=1}^{n} \gamma_i.$$ 

Locally Dependent Random Variables

Chen (1986), IMA Preprint Series 243, Univ. Minnesota (uniform bounds);

- $X_1, \ldots, X_n$ r.v.'s. with zero means and finite variances. Let $\mathcal{J}$ denote $\{1, \ldots, n\}$.

  - (LD1) For each $i$, there exists $i \in A_i \subset \mathcal{J}$ such that $X_i$ and $X_{A_i^c}$ are independent.
  - (LD2) For each $i$, there exists $i \in A_i \subset B_i \subset \mathcal{J}$ such that $X_i$ is independent of $X_{A_i^c}$ and $X_{A_i}$ is independent of $X_{B_i^c}$.
  - (LD3) For each $i$, there exists $i \in A_i \subset B_i \subset C_i \subset \mathcal{J}$ such that $X_i$ is independent of $X_{A_i^c}$, $X_{A_i}$ is independent of $X_{B_i^c}$ and $X_{B_i}$ is independent of $X_{C_i^c}$. 
Assume (LD3).

Let $Y_i = \sum_{j \in A_i} X_j$ and let $Z_i = \sum_{j \in B_i} X_j$.

Let $W = \sum_{i=1}^{n} X_i$ and assume $\text{Var}(W) = 1$.

Let $\xi = \xi_i = (X_i, Y_i, Z_i)$.

For Borel measurable functions $a_\xi$ and $b_\xi$ of $\xi$ such that $a_\xi \leq b_\xi$, 

$$P(a_\xi \leq W \leq b_\xi \mid \xi) \leq 0.625\sigma_i^{-1}(b_\xi - a_\xi) + 4\sigma_i^{-2}r_2 + 2.125\sigma_i^{-3}r_3 + 4\sigma_i^{-3}r_{10} \quad \text{a.s.}$$

where

$$r_2 = \sum_{i=1}^{n} \mathbb{E}|X_i Y_i| I(|Y_i| > 1),$$

$$r_3 = \sum_{i=1}^{n} \mathbb{E}|X_i| (Y_i^2 \wedge 1).$$
Non-uniform Berry-Esseen Bound


- $X_1, \cdots, X_n$ independent with zero means and finite variances.
- Let $W = \sum_{i=1}^{n} X_i$ and assume $\text{Var}(W) = 1$.
- Let $W^{(i)} = W - X_i$. For $0 \leq a \leq b < \infty$,
  \[
  P(a \leq W^{(i)} \leq b) \leq C \left[ \frac{b - a}{(1 + a)^3} + \delta_a \right]
  \]
  where $C$ is an absolute constant,

\[
\delta_y = \frac{\alpha_y}{(1 + y)^2} + \frac{\beta_y}{(1 + y)^3}, \quad y \geq 0,
\]

\[
\alpha_y = \sum_{i=1}^{n} \mathbb{E}X_i^2 I(|X_i| > 1 + y),
\]

\[
\beta_y = \sum_{i=1}^{n} \mathbb{E}|X_i|^3 I(|X_i| \leq 1 + y).
\]
Non-uniform Berry-Esseen Bound

**Theorem 3**

Let $X_1, \cdots, X_n$ be independent with zero means and finite variances. Let $W = \sum_{i=1}^{n} X_i$ and assume $\text{Var}(W) = 1$. Then

$$
\sup_{x \in \mathbb{R}} \left| P(W \leq x) - \Phi(x) \right| 
\leq C \left[ \sum_{i=1}^{n} \frac{\mathbb{E}X_i^2 I(|X_i| > 1 + |x|)}{(1 + |x|)^2} + \sum_{i=1}^{n} \frac{\mathbb{E}|X_i|^3 I(|X_i| \leq 1 + |x|)}{(1 + |x|)^3} \right]
$$

where $C$ is an absolute constant.

**Corollary 4**

$$
\sup_{x \in \mathbb{R}} \left| P(W \leq x) - \Phi(x) \right| \leq \frac{C \sum_{i=1}^{n} \mathbb{E}|X_i|^3}{(1 + |x|)^3}
$$
Non-uniform Berry-Esseen Bound

- Non-uniform bounds were first obtained by Esseen (1945), *Acta Math.* for i.i.d r.v.’s with finite 3rd moments.

- Bikelis (1966), *Litovsk. Mat. Sb.*, obtained bound
  \[ C \sum_{i=1}^{n} \frac{\mathbb{E}|X_i|^3}{(1 + |x|^3)} \]
  for independent but non-identically r.v.’s.
Let \((S, S')\) be an exchangeable pair with finite variances such that
\[
E(S' - S | S) = \lambda S + R,
\]
where \(\lambda > 0\) and \(R\) is a r.v.

Then for \(a \leq b\),
\[
P(a \leq S \leq b) \leq \frac{E|S| + E|R|/\lambda}{ES^2 - E|SR|/\lambda - 1/2} \left( \frac{b - a}{2\delta} \right)
\]
\[
+ \left[ \text{Var} \left( E \left( \frac{1}{2\lambda} (S' - S)^2 I(|S' - S| \leq \delta) \right) | S \right) \right]^{1/2}
\]
where \(\delta = \frac{E|S' - S|^3}{\lambda}\) provided \(ES^2 - E|SR|/\lambda - 1/2 > 0\).
Theorem 5

Let \( \{X_{ij} : i, j = 1, \cdots, n\}, n \geq 2 \), be independent random variables with \( \mathbb{E}X_{ij} = c_{ij}, \ VarX_{ij} = \sigma_{ij}^2 \geq 0 \) and \( \mathbb{E}|X_{ij}|^3 < \infty \).

Assume \( c_i = c_j = 0 \) where \( c_i = \sum_{j=1}^{n} c_{ij}/n \), \( c_j = \sum_{i=1}^{n} c_{ij}/n \).

Let \( \pi \) be a random permutation of \( (1, \cdots, n) \), independent of the \( X_{ij} \), and let \( W = \sum_{i=1}^{n} X_{i\pi(i)} \). Then

\[
Var(W) = \frac{1}{n} \sum_{i,j=1}^{n} \sigma_{ij}^2 + \frac{1}{n-1} \sum_{i,j=1}^{n} c_{ij}^2.
\]

Assuming \( Var(W) = 1 \), then

\[
\sup_{x \in \mathbb{R}} |P(W \leq x) - \Phi(x)| \leq 451 \gamma
\]

where \( \gamma = \frac{1}{n} \sum_{i,j=1}^{n} \mathbb{E}|X_{ij}|^3 \).
Let $X_1, \ldots, X_n$ be independent r.v.'s and let $T := T(X_1, \ldots, X_n)$ be a general sampling statistic.

In many cases, $T = W + \Delta$, where

$$W = \sum_{i=1}^{n} g_i(X_i), \quad \Delta := \Delta(X_1, \ldots, X_n) = T - W,$$

and the $g_i := g_{n,i}$ are Borel measurable functions.

Special cases include U-statistics, multisample U-statistics, L-statistics and random sums.

Assume $\mathbb{E} g_i(X_i) = 0$ for $i = 1, \ldots, n$ and $\text{Var}(W) = 1$. 

Chen and Shao (2007), *Bernoulli*.
\[ |P(T \leq x) - \Phi(x)| \leq |P(T \leq x) - P(W \leq x)| + |P(W \leq x) - \Phi(x)| \]

By Chen and Shao (2001), PTRF,
\[ |P(W \leq x) - \Phi(x)| \leq 4.1 \left[ \sum_{i=1}^{n} \mathbb{E}g_i(X_i)^2 I(|g_i(X_i)| > 1) + \sum_{i=1}^{n} \mathbb{E}|g_i(X_i)|^3 I(|g_i(X_i)| \leq 1) \right]. \]

\[ -P(x - |\Delta| \leq W \leq x) \leq P(T \leq x) - P(W \leq x) \leq P(x \leq W \leq x + |\Delta|). \]

Note that \( T = W + \Delta \).

A bound on \( |P(T \leq x) - \Phi(x)| \) follows from concentration inequalities for \( P(x - |\Delta| \leq W \leq x) \) and \( P(x \leq W \leq x + |\Delta|) \).
Let \( \{a_{ij}, b_{ij} : i, j = 1, \ldots, n\} \) be real numbers.

Let \( W = \sum_{i,j}' a_{ij} b_{\pi(i)\pi(j)} \) where \( \pi \) is a uniform random permutation of \( (1, \ldots, n) \) and \( \sum_{i,j}' \) denotes the sum over pairs \( (i, j) \) of distinct elements of \( \{1, \ldots, n\} \).

Suitably normalized \( W = V + \Delta \) where \( V = \sum_{i=1}^{n} a_i^* b_{\pi(i)}^* \) with \( \sum_{i=1}^{n} a_i^* = 0, \sum_{j=1}^{n} b_j^* = 0, \text{Cov}(V, \Delta) = 0, \) and \( \text{Var}(V) = 1 \).

(Barbour and Eagleson (1986), *Stoch. Analysis Applics*).

There is an existing Berry-Esseen bound on \( |P(V \leq x) - \Phi(x)| \).

(Bolthausen (1984), *Z. Wahrsch. verw. Geb.*).
Matrix Correlation Statistics

\[ -P(x - |\Delta| \leq V \leq x) \leq P(W \leq x) - P(V \leq x) \leq P(x \leq V \leq x + |\Delta|). \]

Note that \( W = V + \Delta \).

Construct exchangeable pair \((V, \Delta)\) and \((V', \Delta')\) and a Stein identity to prove concentration inequalities for
\[ P(x - |\Delta| \leq V \leq x) \text{ and } P(x \leq V \leq x + |\Delta|). \]

A bound on \( |P(W \leq x) - \Phi(x)| \) is then obtained.
Proposition 6

Let \((W, W')\) be an exchangeable pair of \(k\)-dimensional random vectors. Let \(D = (D_1, \ldots, D_k)^T = W' - W\). Suppose

\[
\mathbb{E}|D|^3 < \infty, \quad \inf_{\xi \in S^{k-1}} \mathbb{E}(D \cdot \xi)^2 > 0
\]

where \(S^{k-1}\) denotes the unit \((k - 1)\)-dimensional sphere. Define

\[
\delta := \frac{2\mathbb{E}|D|^3}{\inf_{\xi \in S^{k-1}} \mathbb{E}(D \cdot \xi)^2}.
\]
Proposition 7 (Continued)

Then for any convex set $A$ in $\mathbb{R}^k$ and any $\epsilon > 0$, we have

$$P(W \in A^{\epsilon+\delta} \setminus A^\delta) \leq \frac{1}{\inf_{\xi \in S^{k-1}} \mathbb{E}(D \cdot \xi)^2} \left\{ \frac{16}{3} (\epsilon + 2\delta) \left[ \sum_{j=1}^{k} \text{Var}[\mathbb{E}(D_j|W)] \right]^{1/2} + 2 \left[ \sum_{j,l=1}^{k} \text{Var}[\mathbb{E}(D_jD_l I(|D| \leq \delta)|W)] \right]^{1/2} \right\}.$$
Theorem 8

Let $W$ be such that $\mathbb{E}W = 0$ and $\text{Var}(W) = 1$. Suppose there exists a random function $\hat{K}(t)$ such that

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

for all absolutely continuous functions $f$ with bounded $f'$.

Let $\hat{K}(t) = \hat{K}^{in}(t) + \hat{K}^{out}(t)$ where $\hat{K}^{in}(t) = 0$ for $|t| > 1$. Define $K(t) = \mathbb{E} \hat{K}(t)$, $K^{in}(t) = \mathbb{E} \hat{K}^{in}(t)$, and $K^{out}(t) = \mathbb{E} \hat{K}^{out}(t)$.

Then

$$d_K(W, Z) \leq 2r_1 + 11r_2 + 5r_3 + 10r_4 + 7r_5,$$

where $Z \sim \mathcal{N}(0, 1)$. 

A General Theorem

In Theorem 2,

\[
\begin{align*}
    r_1 &= \left[ \mathbb{E} \left( \int_{|t| \leq 1} (\hat{K}^{in}(t) - K^{in}(t)) dt \right)^2 \right]^{\frac{1}{2}}, \\
    r_2 &= \int_{|t| \leq 1} |tK^{in}(t)| dt, \\
    r_3 &= \mathbb{E} \int_{-\infty}^{\infty} |\hat{K}^{out}(t)| dt, \\
    r_4 &= \mathbb{E} \int_{|t| \leq 1} (\hat{K}^{in}(t) - K^{in}(t))^2 dt, \\
    r_5 &= \left[ \mathbb{E} \int_{|t| \leq 1} |t|(\hat{K}^{in}(t) - K^{in}(t))^2 dt \right]^{\frac{1}{2}}.
\end{align*}
\]
Applications

- Random measures and stochastic geometry.

- Stein couplings: \((G, W', W)\) such that
  \[\mathbb{E}[Gf(W') - Gf(W)] = \mathbb{E}Wf(W)\]
  for absolutely continuous \(f\) with \(f(x) = O(1 + |x|)\). Stein couplings include local dependence, exchangeable pairs, size-bias couplings and others.
Stein’s Method and Applications: A Program in Honor of Charles Stein, July 28 - August 31, 2003, Singapore
Charles and Margaret Stein, August 2003, Singapore
Stein’s method revisited, August 2003, Singapore
A magical moment with Stein: (from left) Louis Chen, Persi Diaconis, Charles Stein, Zhidong Bai, August 2003, Singapore
Third generation of Stein’s method: (from left) Gesine Reinert, Margaret Stein, Torkel Erhardsson, Charles Stein, Aihua Xia, August 2003, Singapore
Symposium in Probability and Statistics in honor of Charles Stein on his 90th Birthday, March 2010, Stanford University
(from left) Charles Stein, Margaret Stein, Sara Stein, March 2010, Stanford
Three generations of Stein’s method: (from left) Louis Chen, Charles Stein, Xiao Fang, March 2010, Stein’s home
With the Steins in their home, September 2013
Charles Stein holding the plaque inscribed with a poem, September 2013, Stein’s home
A thinker original and independent,
In search of perfection invariant,
Found admissible wisdom’s counterexample,
Made (fame, humility) exchangeable.
Thank You