Skewness correction in tail probability approximations for sums of local statistics

Xiao Fang

Based on joint work with Li Luo and Qi-Man Shao

The Chinese University of Hong Kong

June 24, 2019

Central limit theorem and the Berry-Esseen bound

- Let $X_1, X_2, \ldots$ be i.i.d. with $EX_1 = 0, EX_1^2 = 1$. For each positive integer n, let

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

CLT: $W_n \to N(0,1)$ in distribution as $n \to \infty$.

Central limit theorem and the Berry-Esseen bound

- Let $X_1, X_2, \ldots$ be i.i.d. with $EX_1 = 0, EX_1^2 = 1$. For each positive integer n, let

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

CLT: $W_n \to N(0,1)$ in distribution as $n \to \infty$.

- If in addition, $E|X_1|^3 < \infty$, then

$$\sup_{x \in \mathbb{R}} |P(W_n > x) - (1 - \Phi(x))| \leqslant C \frac{E|X_1|^3}{\sqrt{n}},$$

where C is a universal constant, $\Phi(\cdot)$ denotes the standard normal distribution function.

Central limit theorem and the Berry-Esseen bound

- Let $X_1, X_2, \ldots$ be i.i.d. with $EX_1 = 0, EX_1^2 = 1$. For each positive integer n, let

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

CLT: $W_n \to N(0,1)$ in distribution as $n \to \infty$.

- If in addition, $E|X_1|^3 < \infty$, then

$$\sup_{x \in \mathbb{R}} |P(W_n > x) - (1 - \Phi(x))| \leqslant C \frac{E|X_1|^3}{\sqrt{n}},$$

where C is a universal constant, $\Phi(\cdot)$ denotes the standard normal distribution function.

- Not so useful if $x \gg \sqrt{\log n}$.

Cramér's moderate deviations

- If in addition, $Ee^{tX_1} < \infty$ for $t \in (-t_0, t_0)$ where $t_0 > 0$, then

$$\left| \frac{P(W_n > x)}{1 - \Phi(x)} - 1 \right| \leqslant C\left(\frac{1+x}{\sqrt{n}} + \frac{x^3}{\sqrt{n}}\right) \text{ for } 0 \leqslant x \leqslant n^{1/6}.$$

Cramér's moderate deviations

- If in addition, $Ee^{tX_1} < \infty$ for $t \in (-t_0, t_0)$ where $t_0 > 0$, then

$$\left| \frac{P(W_n > x)}{1 - \Phi(x)} - 1 \right| \leqslant C\Big(\frac{1+x}{\sqrt{n}} + \frac{x^3}{\sqrt{n}}\Big) \text{ for } 0 \leqslant x \leqslant n^{1/6}.$$

- Moreover, with $\gamma_n := EW_n^3$,

$$\left| \frac{P(W_n > x)}{(1 - \Phi(x))e^{\gamma_n x^3/6}} - 1 \right| \leqslant C\Big(\frac{1+x}{\sqrt{n}} + \frac{x^4}{n}\Big) \text{ for } 0 \leqslant x \leqslant n^{1/4}.$$

- The range of $x = o(n^{1/6})$ ($x = o(n^{1/4})$ resp.) is optimal for the relative error to vanish.

Cramér's moderate deviations

- If in addition, $Ee^{tX_1} < \infty$ for $t \in (-t_0, t_0)$ where $t_0 > 0$, then

$$\left| \frac{P(W_n > x)}{1 - \Phi(x)} - 1 \right| \leqslant C\left(\frac{1+x}{\sqrt{n}} + \frac{x^3}{\sqrt{n}}\right) \text{ for } 0 \leqslant x \leqslant n^{1/6}.$$

- Moreover, with $\gamma_n := EW_n^3$,

$$\left| \frac{P(W_n > x)}{(1 - \Phi(x))e^{\gamma_n x^3/6}} - 1 \right| \leqslant C\left(\frac{1+x}{\sqrt{n}} + \frac{x^4}{n}\right) \text{ for } 0 \leqslant x \leqslant n^{1/4}.$$

- The range of $x = o(n^{1/6})$ ($x = o(n^{1/4})$ resp.) is optimal for the relative error to vanish.
- We refer to the modification of the normal distribution function in red as *skewness correction*.
- Our goal is to extend the theory of skewness correction in tail probability approximation beyond sums of independent random variables.

Sums of local statistics (of independent random variables)

- Let m and n be positive integers.
- Let $\{X_1, \ldots, X_m\}$ be a sequence of independent random variables.
- Let
$$W = \sum_{i=1}^{n} \xi_i,$$
where each ξ_i is a function of $\{X_\alpha : \alpha \in \mathcal{I}_i\}$ for some $\mathcal{I}_i \subset \{1, \ldots, m\}$.
- For $1 \leqslant \alpha \leqslant m$, define $N_\alpha := \{1 \leqslant i \leqslant n : \alpha \in \mathcal{I}_i\}$.

Main result

Theorem

Under the above setting, assume that $E\xi_i = 0$ for each $1 \leqslant i \leqslant n$ and $\mathrm{Var}(W) = 1$. Assume further that

$$|\xi_i| \leqslant \delta, \quad |\mathcal{I}_i| \leqslant s, \quad |N_\alpha| \leqslant d,$$

where $|\cdot|$ denotes the cardinality when applied to a set. Denote $\gamma := EW^3$. For

$$0 \leqslant x \leqslant (mns^4 d^4 \delta^5)^{-1/2},$$

we have

$$\left| \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right| \leqslant Cmns^4 d^4 \delta^5 (1 + x^2).$$

Same bound holds for $\left| \frac{P(W < -x)}{\Phi(-x)e^{-\gamma x^3/6}} - 1 \right|$.

IID case

Corollary

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $EX_1 = 0, EX_1^2 = 1$.
- For each positive integer n, let

$$W_n = \sum_{i=1}^{n} \xi_i, \quad \xi_i = \frac{X_i}{\sqrt{n}} \text{ for } 1 \leqslant i \leqslant n.$$

- If in addition, $|X_1| \leqslant C_0$ for some $C_0 < \infty$, then the five parameters in the above theorem can be taken as

$$m = n, \ \delta = \frac{C_0}{\sqrt{n}}, \ s = 1, \ d = 1.$$

- with $\gamma_n := EW_n^3$,

$$\left| \frac{P(W_n > x)}{(1 - \Phi(x))e^{\gamma_n x^3/6}} - 1 \right| \leqslant C\left(\frac{1 + x^2}{\sqrt{n}}\right) \text{ for } 0 \leqslant x \leqslant n^{1/4}.$$

Application 1: k-runs

Let $n > k > 1$ be integers. Let $p \in (0,1)$. Let $X_1, \ldots, X_n$ be i.i.d. and $P(X_i = 1) = 1 - P(X_i = 0) = p$. Let

$$S = \sum_{i=1}^{n} X_i X_{i+1} \cdots X_{i+k-1},$$

where $X_{n+i} := X_i$ for $i \geqslant 1$. Let

$$W = \frac{S - ES}{\sqrt{\mathrm{Var}(S)}}.$$

Let $\gamma = EW^3$. If k and p are fixed, then for $0 \leqslant x \leqslant n^{1/4}$, we have,

$$\left| \frac{P(W > x)}{(1 - \Phi(x)) e^{\gamma x^3/6}} - 1 \right| \leqslant \frac{C}{\sqrt{n}} (1 + x^2).$$

Application 1: k-runs (cont.)

Let

$$R_N := \frac{P(W > x)}{1 - \Phi(x)} - 1, \quad R_{skew} := \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1.$$

Table: $n = 1500, k = 2, p = 0.25, \gamma \approx 0.138$. Values of $R_N(R_{skew})$ based on 10^6 repetitions.

x	R_N	R_{skew}
2	0.262	0.050
2.5	0.344	-0.063
3	0.476	-0.208
3.5	1.201	-0.182
4	1.810	-0.358

Application 2: U-statistics

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables from a fixed distribution. Let $s \geqslant 2$ be a fixed integer. Let $h : \mathbb{R}^s \to \mathbb{R}$ be a fixed, symmetric, Borel-measurable function. We consider the U-statistic

$$S = \sum_{1 \leqslant i_1 < \cdots < i_s \leqslant m} h(X_{i_1}, \ldots, X_{i_s}).$$

- Assume that $g(X_1)$ is non-degenerate where

$$g(x) := E(h(X_1, \ldots, X_s) | X_1 = x).$$

Application 2: U-statistics (cont.)

- Assume that
$$|h(X_1, \ldots, X_s)| \leqslant C_0 < \infty.$$
- Let
$$W = \frac{S - ES}{\sqrt{\mathrm{Var}(S)}}.$$
Let $\gamma = EW^3$.
- Proposition: For $0 \leqslant x \leqslant m^{1/4}$, we have,
$$\left| \frac{P(W > x)}{(1 - \Phi(x))e^{\gamma x^3/6}} - 1 \right| \leqslant \frac{C}{\sqrt{m}}(1 + x^2).$$

Application 3: subgraph counts in the Erdös-Rényi random graph

- Let $K(N,p)$ be the Erdős-Rényi random graph with N vertices. Each pair of vertices is connected with probability p and remains disconnected with probability $1-p$, independent of all else. Let G be a given fixed graph.
- Let S be the number of copies (not necessarily induced) of G in $K(N,p)$, and let $W=(S-ES)/\sqrt{\mathrm{Var}(S)}$ be the standardized version. Let $\gamma = EW^3$.
- For fixed p, we have $n \leqslant N^v, m \leqslant N^2, \delta \leqslant CN^{1-v}, s \leqslant C, d \leqslant CN^{v-2}$.
- Proposition: For $0 \leqslant x \leqslant N^{1/2}$, we have,

$$\left|\frac{P(W>x)}{(1-\Phi(x))e^{\gamma x^3/6}} - 1\right| \leqslant \frac{C}{N}(1+x^2).$$

Literature

- Chen, Fang and Shao (2013a) developed Stein's method to prove Cramér-type moderate deviation results in normal approximation for dependent random variables under a boundedness condition.
- Chen, Fang and Shao (2013b) and Shao, Zhang and Zhang (2018) obtained Cramér-type moderate deviation results in Poisson approximation and non-normal approximations, respectively.
- Zhang (2019) refined the results by relaxing the boundedness condition.
- Braverman (2017, Chapter 4) obtained a Cramér-type moderate deviation result in a higher-order approximation for the Erlang-C queuing model. His proof relies on explicit expressions of certain conditional expectations in the model.
- To prove our general bound, we develop Stein's method for exponential concentration inequalities and for higher-order Cramér-type moderate deviations.

Recall the main theorem

- $\{X_1, \ldots, X_m\}$ is a sequence of independent random variables. $W = \sum_{i=1}^{n} \xi_i$, where each ξ_i is a function of $\{X_\alpha : \alpha \in \mathcal{I}_i\}$ for some $\mathcal{I}_i \subset \{1, \ldots, m\}$. For $1 \leqslant \alpha \leqslant m$, define $N_\alpha := \{1 \leqslant i \leqslant n : \alpha \in \mathcal{I}_i\}$.
- Assume that $E\xi_i = 0$ for each $1 \leqslant i \leqslant n$ and $\mathrm{Var}(W) = 1$. Assume further that

$$|\xi_i| \leqslant \delta, \quad |\mathcal{I}_i| \leqslant s, \quad |N_\alpha| \leqslant d.$$

  Denote $\gamma := EW^3$.
- For

$$0 \leqslant x \leqslant (mns^4 d^4 \delta^5)^{-1/2},$$

  we have

$$\left| \frac{P(W > x)}{(1 - \Phi(x)) e^{\gamma x^3/6}} - 1 \right| \leqslant Cmns^4 d^4 \delta^5 (1 + x^2).$$

Proof: 1. Initial reduction

- Only need to consider sufficiently large x; otherwise, use an easily proved Berry-Esseen bound.
- It suffices to prove

$$\begin{aligned}&|P(W > x) - (1 - \Phi(x))e^{\gamma x^3/6}|\\ \leqslant& Cmns^4 d^4 \delta^5 (1 + x^2)(1 - \Phi(x))e^{\gamma x^3/6}\\ \leqslant& Cmns^4 d^4 \delta^5 x \exp(-\frac{x^2}{2} + \frac{\gamma x^3}{6}).\end{aligned}$$

- For $\gamma \neq 0$, $(1 - \Phi(x))e^{\gamma x^3/6}$ is a no longer a distribution function.
- Let $Z_\gamma = \gamma(Y_\gamma - \frac{1}{\gamma^2})$, where $Y_\gamma \sim Poi(\frac{1}{\gamma^2})$. We have $EZ_\gamma = 0, EZ_\gamma^2 = 1, EZ_\gamma^3 = \gamma$.

2. Use translated Poisson as an intermediate approximation

- From Cramér's expansion, we have

$$\frac{P(Z_\gamma > x)}{(1-\Phi(x))e^{\gamma x^3/6}} = 1 + O(|\gamma|)(1+x) + O(\gamma^2)x^4$$

for $0 \leqslant x \leqslant |\gamma|^{-1/2}$ and $|\gamma| \leqslant 1$.

- Therefore, it suffices to prove

$$|P(W > x) - P(Z_\gamma > x)| \leqslant Cmns^4 d^4 \delta^5 x \exp(-\frac{x^2}{2} + \frac{\gamma x^3}{6}).$$

3. Extend the solution to R (error can be controlled)

- Let $h_\alpha(\cdot)$ be a smoothed version of $I(\cdot \leqslant x)$.
- Consider the Stein equation for Z_γ:

$$\frac{1}{\gamma}(f(w+\gamma) - f(w)) - wf(w) = h_\alpha(w) - Eh_\alpha(Z_\gamma).$$

- It has a bounded solution on the support of Z_γ: $\mathcal{S} := \{\gamma\mathbb{Z}^+ - \frac{1}{\gamma}\}$.
- We extend the solution to $f : \mathbb{R} \to \mathbb{R}$ which is a fifth-order polynomial matches the discrete derivatives at $w \in \mathcal{S}$ up to the second order.
- After controlling the error,

$$Eh_\alpha(W) - Eh_\alpha(Z_\gamma) \approx E[\frac{1}{\gamma}(f(W+\gamma) - f(W)) - Wf(W)].$$

In the proof of our main result, we use a standardized Poisson distribution for an intermediate approximation. Translated Poisson distributions have been proposed as alternatives to normal distributions to approximate lattice random variables in the total variation distance. See, for example, Röllin (2005, 2007), Barbour, Luczak and Xia (2018a, b), Barbour and Xia (2018). Instead of matching the support of random variables as in these results, we use standardized Poisson distributions to correct for skewness. See Rio (2009) for a similar use of standardized Poisson distributions.

4. Bounding $E[\frac{1}{\gamma}(f(W+\gamma) - f(W)) - Wf(W)]$

- We have a local dependence structure for W.
- Following Barbour, Karoński and Ruciński (1989), Chen and Shao (2004), etc. to do Taylor's expansion up to $f^{(3)}$.
- Suffices to bound $P(W \geqslant x - \epsilon)$ (following by a moment generating function bound) and $P(x - \epsilon \leqslant W \leqslant x + \epsilon)$ (need an exponential (anti-)concentration inequality).

5. Bounding $P(x - \epsilon \leqslant W \leqslant x + \epsilon)$.

- Let x be sufficiently large but $= o(\sqrt{n})$; $\epsilon = O(\frac{1}{\sqrt{n}})$.
- $P(x - \epsilon \leqslant W \leqslant x + \epsilon) \leqslant ?$

[Shao 2010] for sums of independent random variables:

- Let $W = \sum_{i=1}^{n} \xi_i$, where $\xi_1, \ldots, \xi_n$ are independent, $E\xi_i = 0$, $\sum_{i=1}^{n} E\xi_i^2 = 1$. Suppose $|\xi_i| \leqslant \delta = O(\frac{1}{\sqrt{n}})$.
- Define

$$f(w) = \begin{cases} 0, & w \leqslant x - \epsilon - 2\delta \\ e^{xw}(w - (x - \epsilon - 2\delta)), & x - \epsilon - 2\delta < w \leqslant x + \epsilon + 2\delta \\ e^{x(x+\epsilon+2\delta)}(2\epsilon + 4\delta), & w > x + \epsilon + 2\delta. \end{cases}$$

- Facts: $|f(w)| \leqslant \frac{C}{\sqrt{n}} e^{x^2}$, f is non-decreasing and $f'(w) \geqslant ce^{x^2} I(x - \epsilon - 2\delta \leqslant w \leqslant x + \epsilon + 2\delta)$.

- Let $W' = W - \xi_I + \xi_I'$, $I \sim Unif\{1, \ldots, n\}$, ξ_i' is an independent copy of ξ_i.
- We have $\mathcal{L}(W, W') = \mathcal{L}(W', W)$, and hence

$$E(W - W')(f(W) + f(W')) = 0, \text{ and after rearranging,}$$

$$LHS := 2E(W - W')f(W) = E(W - W')(f(W) - f(W')) =: RHS$$

- It can be checked that

$$LHS = \frac{2}{n} EWf(W) \preceq C \frac{x}{n^{3/2}} e^{x^2} P(W \geqslant x - \epsilon - 2\delta).$$

- $$\begin{aligned} RHS =& E(W-W')(f(W)-f(W')) \\ \geqslant & E(W-W')(f(W)-f(W'))I(x-\epsilon \leqslant W \leqslant x+\epsilon) \\ \geqslant & ce^{x^2}E(W-W')^2 I(x-\epsilon \leqslant W \leqslant x+\epsilon) \\ \succeq & \frac{c}{n}e^{x^2}P(x-\epsilon \leqslant W \leqslant x+\epsilon). \end{aligned}$$

- We have

$$\begin{aligned} & P(x-\epsilon \leqslant W \leqslant x+\epsilon) \\ \leqslant & \frac{Cx}{\sqrt{n}}P(W \geqslant x-\epsilon) \leqslant \frac{Cx}{\sqrt{n}}e^{-x^2}Ee^{xW} \\ \leqslant & \frac{Cx}{\sqrt{n}}e^{-\frac{x^2}{2}} \leqslant \frac{Cx^2}{\sqrt{n}}(1-\Phi(x)). \end{aligned}$$

- For $Z \sim N(0,1)$,

$$P(x-\epsilon \leqslant Z \leqslant x+\epsilon) \leqslant C\epsilon e^{-\frac{(x-\epsilon)^2}{2}} \leqslant \frac{Cx}{\sqrt{n}}(1-\Phi(x)).$$

Conclusions

- We suggest using $(1 - \Phi(x))e^{\gamma x^3/6}$ to approximate tail probabilities in the central limit theorem.
- We justify the suggestion for sums of local statistics under boundedness conditions, which has some applicability.
- How to: improve the rate, consider local dependence, even higher-order expansions, etc.?

Thank you!