Cramér Type Moderate Deviations by Stein's Method

Qi-Man Shao

The Chinese University of Hong Kong

The research is partly supported by HK RGC GRF 403513

- A naive question:
  - Let $0 < a_n \leq 1,\ 0 < b_n \leq 1$. Suppose that
$$a_n - b_n \to 0 \quad \text{as} \quad n \to \infty$$
    Is it true that
$$a_n / b_n \to 1\ ?$$

- Let $0 < a_n(x) \leq 1, \;\; 0 < b(x) \leq 1$, where $b(x) \downarrow$ and continuous for $x \geq 0$. Suppose that

$$a_n(x) - b(x) \to 0$$

as $n \to \infty$ uniformly in $x \geq 0$. It is not true in general that

$$a_n(x)/b(x) \to 1 \;\; \text{as} \;\; n \to \infty$$

uniformly for $x \geq 0$.

- Let $0 < a_n(x) \leq 1, \;\; 0 < b(x) \leq 1$, where $b(x) \downarrow$ and continuous for $x \geq 0$. Suppose that

$$a_n(x) - b(x) \to 0$$

as $n \to \infty$ uniformly in $x \geq 0$. It is not true in general that

$$a_n(x)/b(x) \to 1 \;\text{ as }\; n \to \infty$$

uniformly for $x \geq 0$.
- However, it has been commonly used in practice.

▶ Cramér type moderate deviation

Let W_n be a sequence of random variables and Y be a continuous random variable. Assume that

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

Then

$$P(W_n \geq x) - P(Y \geq x) \to 0$$

uniformly in $x \geq 0$. However it may not be true that

$$P(W_n \geq x)/P(Y \geq x) \to 1$$

uniformly in $x \geq 0$.

- Our focus: Cramér type moderate deviation, that is, to find the largest possible c_n so that

$$P(W_n \geq x)/P(Y \geq x) \to 1$$

uniformly in $x \in [0, c_n]$.

1. Classical Cramér Moderate Deviation

Let $X, X_1, X_2, \cdots, X_n$ be independent identically distributed (i.i.d.) random variables with $EX = 0$ and $\text{Var}(X) = \sigma^2 < \infty$, and let

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

- Cramér (1938):
  If $Ee^{t_0|X|} < \infty$ for $t_0 > 0$, then for $x \geq 0$ and $x = o(n^{1/2})$

$$P\Big(W_n \geq x\Big)/(1 - \Phi(x)) = \exp\Big\{x^2\lambda(\frac{x}{\sqrt{n}})\Big\}\Big(1+O(\frac{1+x}{\sqrt{n}})\Big)$$

where $\lambda(t)$ is the Cramér's series, and $\Phi(x)$ is the standard normal distribution function.

Harald Cramér

- Linnik (1961):
  If $Ee^{t_0\sqrt{|X|}} < \infty$ for $t_0 > 0$, then

$$P\Big(W_n \geq x\Big)/\big(1 - \Phi(x)\big) \to 1$$

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

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

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

- Linnik (1961):
  If $Ee^{t_0\sqrt{|X|}} < \infty$ for $t_0 > 0$, then

$$P\Big(W_n \geq x\Big)/\big(1 - \Phi(x)\big) \to 1$$

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

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

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

- Linnik (1961):
  If $Ee^{t_0\sqrt{|X|}} < \infty$ for $t_0 > 0$, then

$$P\Big(W_n \geq x\Big)/(1 - \Phi(x)) \to 1$$

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

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

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

- Remark: The condition $Ee^{t_0\sqrt{|X|}} < \infty$ is necessary and the interval $(0, \ o(n^{1/6}))$ is the largest possible.

2. Self-normalized Cramér Type Moderate Deviations

Let $X, X_1, X_2, \ldots$ be i.i.d. with $E(X) = 0$ and $\sigma^2 = \mathrm{Var}(X)$. Put

$$S_n = \sum_{i=1}^{n} X_i, \quad V_n^2 = \sum_{i=1}^{n} X_i^2 .$$

Consider the self-normalized sum: S_n/V_n.

Why?

2. Self-normalized Cramér Type Moderate Deviations

Let $X, X_1, X_2, \ldots$ be i.i.d. with $E(X) = 0$ and $\sigma^2 = \mathrm{Var}(X)$. Put

$$S_n = \sum_{i=1}^{n} X_i, \quad V_n^2 = \sum_{i=1}^{n} X_i^2 \; .$$

Consider the self-normalized sum: S_n / V_n.

Why?

- σ is usually unknown, limiting results of $S_n/(\sqrt{n}\,\sigma)$ may not be used directly.
- σ needs to be estimated first!

- The self-normalized sum S_n/V_n has a close relationship with the Student t-statistic, T_n , as follows:

$$T_n = \frac{S_n}{V_n}\left(\frac{n-1}{n-(S_n/V_n)^2}\right)^{1/2}$$

and

$$\{T_n \geq x\} = \left\{\frac{S_n}{V_n} \geq x\left(\frac{n}{n+x^2-1}\right)^{1/2}\right\}.$$

▶ Cramér moderate deviation theorems

- Shao (1999): If $E|X|^3 < \infty$, then

$$\frac{P(S_n/V_n \geq x)}{1 - \Phi(x)} \to 1$$

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

▶ Cramér moderate deviation theorems

- Shao (1999): If $E|X|^3 < \infty$, then

$$\frac{P(S_n/V_n \geq x)}{1-\Phi(x)} \to 1$$

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

- Jing - Shao - Wang (2003): If $E|X|^3 < \infty$, then

$$\frac{P(S_n/V_n \geq x)}{1-\Phi(x)} = 1 + O(1)\frac{(1+x)^3 E|X|^3}{\sqrt{n}\sigma^3}$$

for $0 \leq x \leq n^{1/6}\sigma/(E|X|^3)^{1/3}$, where $|O(1)| \leq C$.

- Remark: The assumption $E|X|^3 < \infty$ is necessary, which is much weaker than $Ee^{t_0\sqrt{|X|}} < \infty$ for the Cramér moderate deviation theorem of $S_n/(\sqrt{n}\sigma)$.

▶ Other related results

- Jing, Shao, Wang (2003): Self-normalized moderate deviation for independent random variables
- Jing, Shao, Zhou (2004): Saddlepoint approximation without any moment condition
- Shao - Wenxin Zhou (2012): Cramér type moderate deviation theorems for self-normalized processes

$$\frac{S_n + \Delta_1}{(V_n^2 + \Delta_2)^{1/2}}.$$

In particular, optimal results are obtained for Studentized U-statistics.

- Liu-Shao (2013): Cramér type moderate deviation for Hotelling's T^2 statistics.

3. Cramér Type Moderate Deviations under Stein's Identity

Let $W := W_n$ be a random variable of interest. Assume that there exist a random function $\hat{K}(t) \geq 0$ and a random variable R such that

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

for all nice function f. Let

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

By Stein's method, it is known that if $E|R| \to 0$,

$$E|1 - E(\hat{K}_1|W)| \to 0, \quad \text{and} \;\; E\int_{-\infty}^{\infty} |t\hat{K}(t)|dt \to 0,$$

then

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

Question:

What can we say about the moderate deviation?

Theorem (Chen, Fang, Shao (2013))

Assume that

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

for all nice function f. If there exist constants d_0, δ_1, δ_2 such that

$$|E(\hat{K}_1|W) - 1| \le \delta_1(1+|W|), \;\; |E(R \mid W)| \le \delta_2(1+|W|),$$
$$E(\hat{K}_1|W) \le d_0.$$

Then

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

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

- 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.

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

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

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

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

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

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

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

Chen, Fang and Shao (2013): For $0 < \beta < 1$

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

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

4. Cramér Type Moderate Deviations of Non-normal Approximation

Let $W := W_n$ be the random variable of interest.

Question:

What is the limiting distribution of W_n? and how to find the limit?

- Exchangeable pair approach:

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

$$E(W - W^* \mid W) = \lambda(g(W) + R(W)),$$

where $\lambda > 0$. Let

$$G(t) = \int_0^t g(s)ds$$

and let Y be a random variable with pdf

$$p(y) = c_1 e^{-G(y)}.$$

Let $\Delta = W - W^*$.

Theorem (Chatterjee and Shao (2011))

Assume that

$$E|R(W)| \to 0, \ \ \frac{1}{\lambda}E|\Delta|^3 \to 0$$

and

$$\frac{1}{2\lambda}E(\Delta^2|W) \xrightarrow{p.} 1.$$

Then, under some regular conditions on g

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

Question:

Can we have a Cramér type moderate deviation theorem?

Theorem (Shao, Mengchen Zhang, Zhuosong Zhang (2014))

Assume that $|\Delta| \le \delta$,

$$\begin{aligned}
&|1 - \frac{1}{2\lambda}E(\Delta^2|W)| \le \delta_1(1 + |g(W)|^{\tau_1}), \\
&|R(W)| \le \delta_2|g(W)|^{\tau_2} + \delta_3.
\end{aligned}$$

Then, under some regular conditions on g

$$\begin{aligned}
\frac{P(W \ge x)}{P(Y \ge x)} &= 1 + O(1)\big(\delta + \delta_1 + \delta_2 + \delta_3(1+x)\big) \\
&\quad + O(1)\big(\delta g(x)G(x) + \delta_1 G(x)g(x)^{\tau_1} + \delta_2 x g(x)^{\tau_2}\big)
\end{aligned}$$

for $x \ge 0$ *and* $\delta g(x)G(x) + \delta_1 G(x)g(x)^{\tau_1} + \delta_2 x g(x)^{\tau_2} + \delta_3 x \le 1$.

▶ A more general result

Let W be the random variable of interest. Suppose there exist an absolutely continuous function g, a constant $\delta > 0$, a random function $\hat{K}(t) \geq 0$ and a random variable $R(W)$ such that

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

for all nice function f. Let

$$G(x) = \int_0^x g(t)dt$$

and Y be a random variable with pdf

$$p(y) = c_1 e^{-G(y)}.$$

Theorem (Shao, Zhang, Zhang)

Let

$$\hat{K}_1 = \int_{|t|\le\delta} \hat{K}(t)dt.$$

Assume there exist constants δ_1, δ_2 *and* δ_3 *such that*

$$|E(\hat{K}_1|W) - 1| \le \delta_1(1 + |g(W)|^{\tau_1}),$$
$$|R(W)| \le \delta_2|g(W)|^{\tau_2} + \delta_3.$$

Then, under some regular conditions on g

$$\begin{aligned}\frac{P(W \ge x)}{P(Y \ge x)} &= 1 + O(1)\big(\delta + \delta_1 + \delta_2 + \delta_3(1+x)\big) \\ &\quad + O(1)\big(\delta g(x)G(x) + \delta_1 G(x)g(x)^{\tau_1} + \delta_2 x g(x)^{\tau_2}\big)\end{aligned}$$

for $x \ge 0$ *and* $\delta g(x)G(x) + \delta_1 G(x)g(x)^{\tau_1} + \delta_2 x g(x)^{\tau_2} + \delta_3 x \le 1.$

Open problem:

Does a Cramér type moderate deviation theorem hold under

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

▶ An application to the Curie-Weiss model at the critical temperature

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

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

where β is called the inverse of temperature. Let $\beta = 1$, the critical temperature case, 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 $c_1 \, e^{-y^4/12}$.

- Chatterjee and Shao (2011):

$$|P(W \geq x) - P(Y \geq x)| \leq Cn^{-1/2}$$

- Chatterjee and Shao (2011):

$$|P(W \geq x) - P(Y \geq x)| \leq Cn^{-1/2}$$

- Shao, Zhang, Zhang (2014):

$$\frac{P(W \geq x)}{P(Y \geq x)} = 1 + O(1)(1 + x^6)n^{-1/2}$$

for $0 \leq x \leq n^{1/12}$.

5. Stein's Method and the Main Idea of Proofs

Let $W = W_n$ be the random variable of interest and Y be a random variable with pdf $p(y)$.

Aims:

(i) Prove $W \xrightarrow{d.} Y$

5. Stein's Method and the Main Idea of Proofs

Let $W = W_n$ be the random variable of interest and Y be a random variable with pdf $p(y)$.

Aims:

(i) Prove $W \xrightarrow{d.} Y$

(ii) For a given function h, estimate

$$Eh(W) - Eh(Y)$$

(iii) Prove a Cramér type moderate deviation theorem

5. Stein's Method and the Main Idea of Proofs

Let $W = W_n$ be the random variable of interest and Y be a random variable with pdf $p(y)$.

Aims:

(i) Prove $W \xrightarrow{d.} Y$

(ii) For a given function h, estimate

$$Eh(W) - Eh(Y)$$

(iii) Prove a Cramér type moderate deviation theorem

Main tools:

- Fourier transform, strong approximation, ...
- Stein's method

▶ Stein's method

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:

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

- Stein's equation: For a given measurable function h

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

or

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

- Properties of the solution:

Let f_h be the solution to Stein's equation. Under some regular conditions on p

$$\|f_h\| \le C\|h\|, \;\; \|f_h'\| \le C\|h\|,$$

$$\|f_h\| \le C\|h'\|, \;\; \|f_h'\| \le C\|h'\|, \;\; \|f_h''\| \le C\|h'\|$$

▶ Main idea of proofs

To prove the moderate deviation theorem, by Stein's equation

$$P(W \geq x) - P(Y \geq x) = Ef'(W) + Ef(W)p'(W)/p(W)$$

A key step:

- Let $Y \sim N(0, 1)$. Observe that

$$Ee^{tY} = e^{t^2/2}.$$

A key step is to prove that

$$Ee^{tW} = O(e^{t^2/2})$$

- For Y with pdf $p(y)$, assume that

$$p(y) = c_1 e^{-G(y)}.$$

Observe that

$$Ee^{G(Y)-G(Y-t)} = c_1 \int_{-\infty}^{\infty} e^{-G(y-t)} dy = 1 \quad \text{for all} \quad t.$$

A key step is to prove that

$$Ee^{G(W)-G(W-t)} \leq C.$$

References

- 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.
- Chen, L.H.Y, Goldstein, L. and Shao, Q.M. (2011). Normal Approximation by Stein's Method. Springer.
- Chen, L.H.Y., Fang, X. and Shao, Q.M. (2013). From Stein identities to moderate deviations. *Ann. Probab.* **41**, 262-293.
- Shao, Q.M., Zhang, M.C., Zhang, Z.S. (2014). Cramér type moderate deviations of non-normal approximation.
- Stein, C. (1986) *Approximation Computation of Expectations.* Lecture Notes 7, Inst. Math. Statist., Hayward, Calif.


Thank You