Introduction to Self-normalized Limit Theory

Qi-Man Shao

The Chinese University of Hong Kong

E-mail: *qmshao@cuhk.edu.hk*

Outline

- What is the self-normalization? Why?
- Classical limit theorems
- Self-normalized large deviations
- Self-normalized moderate deviations
- Self-normalized Cramér type moderate deviations

Main references

[1]V.H. de la Pena, T.Z. Lai and Q.M. Shao (2009). *Self-normalized processes: Limit Theory and Statistical Applications.* Spriner, Heidelberg.

[2] Q.M. Shao and W.X. Zhou (2012). Cramér type moderate deviation theorems for self-normalized processes.

1. Introduction

Let $X, X_1, X_2, \cdots, X_n$ be independent identically distributed (i.i.d.) random variables and let

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

Assume $EX = 0$ and $\sigma^2 = EX^2 < \infty$.

Standardized sum: $S_n/(\sigma\sqrt{n})$

Self-normalized sum: S_n/V_n

Why are we interested in the self-normalized sum?

- It is of interest from the probability theory alone
- It has a close relation with Studentized statistics.

  Let $H_n = H_n(\theta, \lambda)$ be a sequence of statistics under consideration, where θ contains parameters of interest and λ is a vector of some unknown nuisance parameters. It is a common practice that one needs to estimate λ first from the data, say, the estimator is $\hat{\lambda}$, and then substitute $\hat{\lambda}$ in H_n, which naturally brings a studentized statistic $\hat{H}_n = H_n(\theta, \hat{\lambda})$. Typical examples include:

  – Student's t-statistic

  – Hotelling's T^2 statistics

  – Studentized U-statistics

  – The largest eigenvalue of sample correlation matrices

2. Classical limit theorems

Let $X, X_1, X_2, \cdots, X_n$ be independent identically distributed (i.i.d.) random variables and let

$$S_n = \sum_{i=1}^{n} X_i.$$

Law of large numbers:

$$EX = \mu \iff \frac{S_n}{n} \longrightarrow \mu \quad a.s.$$

Law of the iterated logarithm:

$$\begin{aligned} & EX = 0, \quad 0 < \sigma^2 = EX^2 < \infty \\ \iff \quad & \limsup_{n\to\infty} \frac{S_n}{\sigma\sqrt{n}\,(2\ \log\log\ n)^{1/2}} = 1 \ \ a.s. \end{aligned}$$

The central limit theorem:

- If $EX = 0$ and $\sigma^2 = EX^2 < \infty$, then

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d.} N(0,1)$$

- If $EX = 0$ and $EX^2 I\{|X| \leq x\}$ is slowly varying, then there exist a_n and b_n such that

$$\frac{1}{a_n} S_n - b_n \xrightarrow{d.} N(0,1)$$

Uniform Berry-Esseen bounds:

$$\sup_x |P(\frac{S_n}{\sigma\sqrt{n}} \leq x) - \Phi(x)| \leq \frac{.7975 E|X|^3}{\sqrt{n}\sigma^3}.$$

Non-uniform Berry-Esseen bounds:

$$|P(\frac{S_n}{\sigma\sqrt{n}} \leq x) - \Phi(x)| \leq \frac{C\, E|X|^3}{(1+|x|^3)\sqrt{n}\sigma^3}$$

Cramér - Chernoff's large deviation:

If $Ee^{t_0 X} < \infty$ for some $t_0 > 0$, then $\forall\ x > EX$,

$$P\left(\frac{S_n}{n} \geq x\right)^{1/n} \to \inf_{t\geq 0} e^{-tx} Ee^{tX}.$$

Cramér's moderate deviation:

Assume $EX = 0$ and $\sigma^2 = EX^2 < \infty$.

- If $Ee^{t_0|X_1|^{1/2}} < \infty$ for $t_0 > 0$, then

$$\frac{P\left(\frac{S_n}{\sigma\sqrt{n}} \geq x\right)}{1-\Phi(x)} \to 1$$

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

- If $Ee^{t_0|X_1|} < \infty$ for $t_0 > 0$, then for $x \geq 0$ and $x = o(n^{1/2})$

$$\frac{P\left(\frac{S_n}{\sigma\sqrt{n}} \geq x\right)}{1-\Phi(x)} = \exp\left\{x^2\lambda(\frac{x}{\sqrt{n}})\right\}\left(1+O(\frac{1+x}{\sqrt{n}})\right),$$

where $\lambda(t)$ is the Cramér's series.

Conclusion: Classical limit theorem

- The normalizing constants are deterministic
- Moment conditions play a crucial role

3. Self-normalized limit theorems: a brief review

Self-normalized sum:

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

Do the classical limit theorems remain valid for S_n/V_n?

3.1 The central limit theorem:

- If $EX = 0$ and $EX^2 < \infty$, then

$$S_n/V_n \xrightarrow{d.} N(0,1)$$

- Gine-Götze-Mason (1995):

  $EX = 0$ and $EX^2 I\{|X| \le x\}$ is slowly varying

$$\iff S_n/V_n \xrightarrow{d.} N(0,1)$$

- Bentkus-Götze (1996):

  If $EX = 0$, $\sigma^2 = EX^2$ and $E|X|^3 < \infty$, then

$$\sup_x |P(S_n/V_n \le x) - \Phi(x)| \le C\, n^{-1/2} E|X|^3/\sigma^3$$

3.2 Self-normalized limit distribution for $X \in DASL$

- Logan-Mallows-Rice-Shepp (1973):

  If $X \in DASL(\alpha)$, then the limiting density function $p(x)$ of S_n/V_n exists and satisfies

$$p(x) \sim \frac{1}{\alpha}\left(\frac{2}{\pi}\right)^{1/2} \tau_\alpha e^{-x^2\tau_\alpha^2/2}$$

Conjecture: τ_α is the solution of

$$\begin{cases} c_1 D_\alpha(-\tau) + c_2 D_\alpha(\tau) = 0 & \text{if } \alpha \neq 1 \\ \dfrac{e^{\tau^2/2}}{\tau} - \displaystyle\int_0^\tau e^{x^2/2}\, dx = 0 & \text{if } \alpha = 1 \end{cases}$$

where $D_\alpha(x)$ is the parabolic cylinder function.

3.3 Self-normalized law of the iterated logarithm (Griffin and Kuelbs (1989))

(a) If $EX = 0$ and $EX^2 I\{|X| \le x\}$ is slowly varying, then

$$\limsup_{n\to\infty} \frac{S_n}{V_n(2\log\log n)^{1/2}} = 1 \quad a.s.$$

(b) If X is symmetric and

$$P(X \ge x) = \frac{l(x)}{x^\alpha}, \quad 0 < \alpha < 2,$$

where $l(x)$ is a slowly varying function, then there exists $0 < C_\alpha < \infty$ such that

$$\limsup_{n\to\infty} \frac{S_n}{V_n(\log\log n)^{1/2}} = C_\alpha \quad a.s.$$

However, for any $a_n \to \infty$

$$\limsup_{n\to\infty} \frac{S_n}{a_n} = 0 \;\; a.s. \text{or} \;\; \infty \;\; a.s.$$

4. Self-normalized large deviations

Question: Is there a self-normalized large deviation without assuming moment condition?

Answer: YES!

Observe that for $x > 0$

$$
\begin{aligned}
P(S_n/V_n^2 \geq x) &= P(S_n \geq x V_n^2) \\
&= P(\sum_{i=1}^{n} (X_i - x X_i^2) \geq x)
\end{aligned}
$$

Assume that $EX = 0$ or $EX^2 = \infty$. We have

$$EX - xEX^2 < 0$$

and

$$Ee^{t(X - x\,X^2)} < \infty \quad \text{for } t \geq 0.$$

Therefore, by the Chernoff large deviation

$$P(S_n/V_n^2 \geq x)^{1/n} \to \inf_{t \geq 0} Ee^{t(X - x\,X^2)}$$

Question: What if the normalizing constant is $V_n\sqrt{n}$?

It is well-know that

$$ab \le (a^2 + b^2)/2$$

Indeed, we have

$$ab = (1/2) \inf_{c>0} (c\, a^2 + b^2/c)$$

for $a, b > 0$. Thus

$$V_n\sqrt{n} = (1/2) \inf_{c>0} (V_n^2/c + c\, n)$$

and

$$
\begin{aligned}
& P(\frac{S_n}{V_n\sqrt{n}} \geq x) \\
&= P(S_n \geq xV_n\sqrt{n}) \\
&= P(S_n \geq x(1/2)\inf_{c>0}(V_n^2/c + c\,n)) \\
&= P\left(\sup_{c\geq 0} cS_n - x(1/2)(V_n^2 + c^2 n) \geq 0\right) \\
&= P\left(\sup_{c\geq 0}\sum_{i=1}^{n}\{cX_i - x(1/2)(X_i^2 + c^2)\} \geq 0\right).
\end{aligned}
$$

Theorem 1.1 [*Shao* (1997)]

If $EX = 0$ *or* $EX^2 = \infty$, *then* $\forall\ x > 0$

$$
\begin{aligned}
& P\left(S_n/V_n \geq xn^{1/2}\right)^{1/n} \\
& \to \sup_{c\geq 0}\inf_{t\geq 0} Ee^{t(cX - x(|X|^2 + c^2)/2)}
\end{aligned}
$$

5. Self-normalized moderate deviations

Theorem 1.2 [*Shao* (1997)]

(a) If $EX = 0$ *and* $EX^2 I\{|X| \le x\}$ *is slowly varying, then*

$$\ln P(S_n/V_n > x_n) \sim -x_n^2/2$$

for $x_n \to \infty$ *and* $x_n = o(\sqrt{n})$, *i.e.,* $\forall\ \varepsilon > 0$

$$e^{-(1+\varepsilon)x_n^2/2} \le P(S_n/V_n > x_n) \le e^{-(1-\varepsilon)x_n^2/2}$$

(b) Let X *be a symmetric random variable satisfying*

$$P(X \ge x) = \frac{l(x)}{x^\alpha}, \quad 0 < \alpha < 2.$$

where $l(x)$ *is a slowly varying function. Then* $\forall\ x_n \to \infty, x_n = o(\sqrt{n})$

$$\ln P\left(\frac{S_n}{V_n} \ge x_n\right) \sim -\beta_\alpha\, x_n^2$$

where β_α *is the solution of*

$$\int_0^\infty \frac{2 - e^{2x - x^2/\beta} - e^{-2x - x^2/\beta}}{x^{\alpha+1}}\, dx = 0. \qquad (1.1)$$

6. Self-normalized Cramér type moderate deviations

Theorem 1.3 (Shao (1999)) *If* $EX = 0$ *and* $E|X|^3 < \infty$, *then*

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

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

Theorem 1.4 [*Jing-Shao-Wang* (2003)].

If $EX = 0$ *and* $E|X|^3 < \infty$, *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 $x \geq 0$, *and*

$$\begin{aligned} & |P(S_n/V_n \geq x) - (1 - \Phi(x))| \\ & \leq \; A(1+x)^3 e^{-x^2/2} n^{-1/2} E|X|^3/\sigma^3 \end{aligned}$$

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

7. Self-normalized Cramér moderate deviations for independent random variables

Let $X_1, X_2, \cdots, X_n$ be independent random variables with $EX_i = 0$ and $EX_i^2 < \infty$. Put

$$S_n = \sum_{i=1}^n X_i,\ V_n^2 = \sum_{i=1}^n X_i^2,\ B_n^2 = ES_n^2$$

and

$$\begin{aligned}\Delta_{n,x} &= \frac{(1+x)^2}{B_n^2} \sum_{i=1}^n EX_i^2 I_{\{|X_i| > B_n/(1+x)\}} \\ &\quad + \frac{(1+x)^3}{B_n^3} \sum_{i=1}^n E|X_i|^3 I_{\{|X_i| \le B_n/(1+x)\}}\end{aligned}$$

for $x > 0$.

Theorem 1.5 [*Jing-Shao-Wang* (2003)].

There is an absolute constant A (> 1) *such that*

$$\frac{P(S_n \geq xV_n)}{1-\Phi(x)} = e^{O(1)\Delta_{n,x}}$$

for all $x \geq 0$ *satisfying*

$$x^2 \max_{1\leq i\leq n} EX_i^2 \leq B_n^2$$

and

$$\Delta_{n,x} \leq (1+x)^2/A,$$

where $|O(1)| \leq A$ *.*

Theorem 1.6 [*Jing-Shao-Wang* (2003)].

$$|P(S_n/V_n \geq x) - (1 - \Phi(x))| \leq A(1+x)^3 e^{-x^2/2} \frac{\sum_{i=1}^n E|X_i|^3}{B_n^3}$$

for

$$0 \leq x \leq \frac{B_n}{(\sum_{i=1}^n E|X_i|^3)^{1/3}}$$

8. Cramér moderate deviations for self-normalized processes

8.1 A general result

Let $\xi_1, ..., \xi_n$ be independent random variables satisfying

$$E\xi_i = 0, \quad 1 \le i \le n, \qquad \sum_{i=1}^n E\xi_i^2 = 1. \tag{1.2}$$

Assume the nonlinear process of interest can be decomposed as a standardized partial sum of $\{\xi_i\}_{i=1}^n$, say, W_n, plus a remainder, say, $D_{n,1}$, while its self-normalized version can be written as

$$T_n = \frac{W_n + D_{n,1}}{V_n(1 + D_{n,2})^{1/2}}, \tag{1.3}$$

where

$$W_n = \sum_{i=1}^n \xi_i, \quad V_n = \left(\sum_{i=1}^n \xi_i^2 \right)^{1/2},$$

and $D_{n,1}$, $D_{n,2}$ are measurable functions of $\{\xi_i\}_{i=1}^n$. Examples satisfying (1.3) include t-statistic, Studentized U-statistics and L-statistics.

In this section, we establish a general Cramér type moderation theorem for self-normalized process T_n in the form of (1.3). For $1 \le$

$i \le n$ and $x \ge 0$, let

$$\delta_{i,x} = (1+x)^2 E[\xi_i^2 I_{\{(1+x)|\xi_i|>1\}}] + (1+x)^3 E[|\xi_i|^3 I_{\{(1+x)|\xi_i|\le 1\}}],$$

$$\Delta_{n,x} = \sum_{i=1}^n \delta_{i,x} \quad \text{and} \quad I_{n,x} = E[e^{xW_n - x^2 V_n^2/2}] = \prod_{i=1}^n \boldsymbol{e}[e^{x\xi_i - x^2\xi_i^2/2}].$$

Let $D_{n,1}^{(i)}$ and $D_{n,2}^{(i)}$, for each $1 \le i \le n$, be arbitrary measurable functions of $\{\xi_j\}_{j=1, j\ne i}^n$, such that $\{D_{n,1}^{(i)}, D_{n,2}^{(i)}\}$ and ξ_i are independent. Set also for $x > 0$ that

$$R_{n,x} := I_{n,x}^{-1} \times \Big\{ E[(x|D_{n,1}| + x^2|D_{n,2}|) e^{\sum_{j=1}^n (x\xi_j - x^2\xi_j^2/2)}]$$

$$+ \sum_{i=1}^n E[\min(x|\xi_i|, 1)(|D_{n,1} - D_{n,1}^{(i)}| + x|D_{n,2} - D_{n,2}^{(i)}|) e^{\sum_{j\ne i}(x\xi_j - x^2\xi_j^2/2)}] \Big\},$$

where $\sum_{j\ne i} = \sum_{j=1, j\ne i}^n$.

Theorem 1.7 *(Shao and Zhou (2012)) Let T_n be defined in (1.3). Then there is an absolute constant A (>1) such that*

$$\exp\{O(1)\Delta_{n,x}\}\big(1 - A\, R_{n,x}\big) \le \frac{P(T_n \ge x)}{1-\Phi(x)} \tag{1.4}$$

and

$$P(T_n \geq x) \leq \big(1 - \Phi(x)\big) \exp\{O(1)\Delta_{n,x}\}\big(1 + A\, R_{n,x}\big) \quad (1.5)$$
$$+P\big\{|D_{n,1}|/V_n > 1/(4x)\big\} + P\big\{|D_{n,2}| > 1/(4x^2)\big\}$$

for all $x > 1$ *satisfying*

$$\max_{1 \leq i \leq n} \delta_{i,x} \leq 1 \quad (1.6)$$

and

$$\Delta_{n,x} \leq (1 + x)^2/A, \quad (1.7)$$

where $|O(1)| \leq A$*.*

8.2 Studentized U-statistics

Let $X_1, X_2, \cdots, X_n$ be a sequence of i.i.d. random variables and let $h : R^m \to R$ be a Borel measurable symmetric function of m variables, where $2 \le m < n/2$. Consider Hoeffding's U-statistic with a kernel h of degree m given by

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}), \tag{1.8}$$

which is an unbiased estimate of $\theta = \boldsymbol{e}h(X_1, \ldots, X_m)$. The U-statistic is a basic statistic and its asymptotic properties have been extensively studied in literature. Let

$$g(x) = Eh(x, X_2, \ldots, X_m),\ x \in \mathbb{R} \quad \text{and} \quad \sigma^2 = \mathrm{Var}(g(X_1)).$$

For standardized (non-degenerate) U-statistic

$$Z_n = \frac{\sqrt{n}}{m\sigma}(U_n - \theta), \tag{1.9}$$

where $\sigma > 0$ and m is fixed.

Since σ is usually unknown, consider the following Studentized U-statistic

$$T_n = \frac{\sqrt{n}}{ms_1}(U_n - \theta), \tag{1.10}$$

where s_1^2 is the leave-one-out Jackknife estimator of σ^2 given by

$$\begin{aligned} s_1^2 &= \frac{(n-1)}{(n-m)^2}\sum_{i=1}^{n}(q_i - U_n)^2 \quad \text{with} \qquad (1.11)\\ q_i &= \binom{n-1}{m-1}^{-1} \sum_{(l_1,\ldots,l_m)\in\mathcal{C}_{m-1,i}} h(X_i, X_{l_1}, ..., X_{l_{m-1}}) \end{aligned}$$

and

$$\mathcal{C}_{m-1,i} = \{(l_1, ..., l_{m-1}) : 1 \le l_1 < \cdots < l_{m-1} \le n, l_j \ne i \text{ for } 1 \le j \le m-1\}.$$

As a direct but non-trivial consequence of Theorem 1.7, we can establish a sharp Cramér type moderate deviation theorem for Studentized U-statistic T_n as follows.

Theorem 1.8 *(Shao and Zhou (2012)) Let* $2 < p \le 3$ *and assume*

$$0 < \sigma_p := (E|g(X_1) - \theta|^p)^{1/p} < \infty.$$

Suppose that there are constants $c_0 \geq 1$ *and* $\tau \geq 0$ *such that*

$$(h(x_1, ..., x_m) - \theta)^2 \leq c_0 \Big(\tau\sigma^2 + \sum_{i=1}^{m} (g(x_i) - \theta)^2 \Big). \qquad (1.12)$$

Then there exists a constant $A > 1$ *only depending on* m *such that*

$$\frac{P(T_n \geq x)}{1 - \Phi(x)} = 1 + O(1) \Big\{ \frac{\sigma_p^p (1+x)^p}{\sigma^p n^{(p-2)/2}} + c_0(1+\tau) \frac{(1+x)^3}{n^{1/2}} \Big\}, \quad (1.13)$$

for any $0 < x < \frac{1}{A} \min\{\sigma n^{(p-2)/(2p)}/\sigma_p, n^{1/6}/(c_0(1+\tau))^{1/6}\}$, *where* $|O(1)| \leq A$. *In particular, we have*

$$\frac{P(T_n \geq x)}{1 - \Phi(x)} \to 1 \qquad (1.14)$$

uniformly in $0 \leq x \leq o(n^{(p-2)/(2p)})$.

Clearly, condition (1.12) is satisfied for the t-statistic $(h(x_1, x_2) = (x_1 + x_2)/2$ with $c_0 = 2$ and $\tau = 0)$, sample variance $(h(x_1, x_2) = (x_1 - x_2)^2/2$, $c_0 = 10$, $\tau = \theta^2/\sigma^2)$, Gini's mean difference $(h(x_1, x_2) = |x_1 - x_2|$, $c_0 = 8$, $\tau = \theta^2/\sigma^2)$ and one-sample Wilcoxon's statistic $(h(x_1, x_2) = 1\{x_1 + x_2 \leq 0\}$, $c_0 = 1$, $\tau = 1/\sigma^2)$. It would be interesting to know if condition (1.12) can be weakened, but it seems impossible to remove condition (1.12) completely.

9. Miscellanies

- Horváth-Shao (1996): Large deviation for $S_n/\max_{1\le i\le n}|X_i|$
- Dembo-Shao (1998): Self-normalized moderate and large deviations in R^d
- Gine-Mason (2001): Sub-gaussian property for X in the Feller class
- Bercu-Gassiat-Rio (2002): Concentration inequalities, large and moderate deviations for self-normalized empirical processes
- De la Peña-Klass-Lai (2004): Self-normalized processes: exponential inequalities, moment and limit theorems
- Jing-Shao-Zhou (2004): Self-normalized saddle point approximations
- Jing-Shao-Wang (2003): The studentized bootstrap

- He and Shao (1996): Bahadur efficiency of studentized score tests
- He and Shao (1996, 2000): Bahadur representations for M-estimators
- Horvath and Shao (1996): Change point analysis
- Chen and Shao (1997, 2000): Monte Carlo methods in Baysian computations
- Liu and Shao (2013): Hotelling's T^2 statistics

Conclusion: Self-normalized limit theorems

- Require little or no moment assumption
- Results are more elegant and neater than many classical limit theorems
- Provide much wider applicability to other fields and to statistics in particular