Necessary and Sufficient Conditions for the Self-normalized Central Limit Theorem

Qi-Man Shao

The Chinese University of Hong Kong

qmshao@sta.cuhk.edu.hk

The research is partially supported by Hong Kong RGC GRF 14302515 and 14304917

▶ 1.1 Classical Central Limit Theorem

Let $\{X_i, 1 \leq i \leq n\}$ be a sequence of independent random variables. Put

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

What is the necessary and sufficient condition for the central limit theorem (CLT)?

▶ 1.1 Classical Central Limit Theorem

Let $\{X_i, 1 \le i \le n\}$ be a sequence of independent random variables. Put

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

What is the necessary and sufficient condition for the central limit theorem (CLT)?

Under what conditions, are there a_n and c_n such that

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

- $\{X_i\}$ are independent and identically distributed (i.i.d.) random variables with $E(X_1) = 0$

The following statements are equivalent

- CLT holds
- $EX_1^2 I\{|X_1| \le x\}$ is a slowly varying function, i.e., X_1 is in the domain of attraction of the normal distribution, denoted by $X_1 \in DAN$.

- $\{X_i\}$ are independent and identically distributed (i.i.d.) random variables with $E(X_1) = 0$

The following statements are equivalent

- CLT holds
- $EX_1^2 I\{|X_1| \leq x\}$ is a slowly varying function, i.e., X_1 is in the domain of attraction of the normal distribution, denoted by $X_1 \in DAN$.

Function $f(x)$ is called slowly varying if $\forall t > 0$

$$\frac{f(tx)}{f(x)} \to 1 \text{ as } x \to \infty.$$

- $\{X_i\}$ are independent and identically distributed (i.i.d.) random variables with $E(X_1) = 0$

The following statements are equivalent

- CLT holds
- $EX_1^2 I\{|X_1| \le x\}$ is a slowly varying function, i.e., X_1 is in the domain of attraction of the normal distribution, denoted by $X_1 \in DAN$.

  Function $f(x)$ is called slowly varying if $\forall t > 0$

$$\frac{f(tx)}{f(x)} \to 1 \text{ as } x \to \infty.$$

- $\max_{1\le i\le n} |X_i|/V_n \to 0$ in probability, where $V_n^2 = \sum_{i=1}^n X_i^2$

▶ Independent Random Variables

Assume that

$$\forall\, \varepsilon > 0, \quad \max_{1\le i\le n} P(|X_i| \ge \varepsilon\, a_n) \to 0.$$

Then

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

if and only if

(i) $\displaystyle\sum_{i=1}^{n} P(|X_i| \ge \varepsilon a_n) \to 0$ for any $\varepsilon > 0$;

(ii) $\displaystyle\frac{1}{a_n}\sum_{i=1}^{n} EX_i I_{\{|X_i|\le a_n\}} \to 0$;

(iii) $\displaystyle\frac{1}{a_n^2}\sum_{i=1}^{n}\left(EX_i^2 I_{\{|X_i|\le a_n\}} - (EX_i I_{\{|X_i|\le a_n\}})^2\right) \to 1$

▶ 1.2. Self-normalized central limit theorem for i.i.d. random variables

Let $\{X_i, 1 \le i \le n\}$ be independent and identically distributed random variables. Put

$$S_n = \sum_{i=1}^n X_i \ \text{ and } \ V_n^2 = \sum_{i=1}^n X_i^2.$$

- Self-normalized sum: S_n/V_n

Under what conditions, does the CLT hold for the self-normalized sum?

Assume $E(X_1) = 0$.

Logan, Mallows, Rice and Shepp (1973)'s conjecture:

$S_n/V_n \xrightarrow{d.} N(0,1)$ if and only if $X_1 \in DAN$.

Assume $E(X_1) = 0$.

Logan, Mallows, Rice and Shepp (1973)'s conjecture:

$S_n/V_n \xrightarrow{d.} N(0,1)$ if and only if $X_1 \in DAN$.

- Maller (1981):
  If $X_1 \in DAN$, then $S_n/V_n \xrightarrow{d.} N(0,1)$
- Griffin and Mason (1991):
  Assume that X_1 is symmetric. Then CLT holds for S_n/V_n if and only if $X_1 \in DAN$

Assume $E(X_1) = 0$.

Logan, Mallows, Rice and Shepp (1973)'s conjecture:

$S_n/V_n \xrightarrow{d.} N(0,1)$ if and only if $X_1 \in DAN$.

- Maller (1981):
  If $X_1 \in DAN$, then $S_n/V_n \xrightarrow{d.} N(0,1)$
- Griffin and Mason (1991):
  Assume that X_1 is symmetric. Then CLT holds for S_n/V_n if and only if $X_1 \in DAN$
- Gine, Götze and Mason (1997):
  CLT holds for S_n/V_n if and only if $X_1 \in DAN$

2. Toward to a General Self-normalized Central Limit Theorem

Let $\{X_i, 1 \le i \le n\}$ be independent random variables with $E(X_i) = 0$. Set

$$S_n = \sum_{i=1}^{n} X_i \text{ and } V_n^2 = \sum_{i=1}^{n} X_i^2.$$

- Aim: Find the necessary and sufficient condition for the self-normalized CLT, i.e.,

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

Recall the classical Lindeberg CLT:

If the Lindeberg condition is satisfied

$$\forall\, \varepsilon > 0,\ \sum_{i=1}^{n} E(X_i^2/B_n^2) I_{\{|X_i| > \varepsilon B_n\}} \to 0,$$

where $B_n^2 = \sum_{i=1}^n EX_i^2$, then

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

Also recall the self-normalized CLT for i.i.d. random variables:

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

if and only if $\max_{1\le i\le n} |X_i|/V_n \to 0$ in probability, which is equivalent to

$$\forall\, \varepsilon > 0,\ \sum_{i=1}^{n} E(X_i^2/V_n^2) I_{\{|X_i| > \varepsilon V_n\}} \to 0,$$

which looks like a Lindeberg condition.

- Egorov (1996):

  Let $X_i, i \geq 1$ be independent and symmetric random variables. Then

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

  if and only if $\max_{1 \leq i \leq n} |X_i|/V_n \to 0$ in probability .

- Egorov (1996):

  Let $X_i, i \geq 1$ be independent and symmetric random variables. Then
  $$S_n/V_n \xrightarrow{d.} N(0,1)$$
  if and only if $\max_{1\leq i\leq n} |X_i|/V_n \to 0$ in probability .

- Question: Is $\max_{1\leq i\leq n} |X_i|/V_n \to 0$ in probability a sufficient condition in general?

- Egorov (1996):

  Let $X_i, i \geq 1$ be independent and symmetric random variables. Then

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

  if and only if $\max_{1\leq i\leq n} |X_i|/V_n \to 0$ in probability .

- Question: Is $\max_{1\leq i\leq n} |X_i|/V_n \to 0$ in probability a sufficient condition in general?

  Answer: No.

Example 1:

Let $X_i, i \geq 1$ be independent random variables satisfying

$$\begin{aligned} P(X_{2i-1} = -i^2) &= \frac{1}{1+i^2}, \; P(X_{2i-1} = 1) = \frac{i^2}{1+i^2} \\ P(X_{2i} = i^2) &= \frac{1}{1+i^2}, \; P(X_{2i} = -1) = \frac{i^2}{1+i^2} \end{aligned}$$

for $i = 1, 2, \cdots$. Then $E(X_i) = 0$ and $\max_{1 \leq i \leq n} |X_i|/V_n \to 0$ in probability, but $S_n/V_n \to 0$ in probability.

Theorem (1)

Let $\{X_n, n \geq 1\}$ *be independent non-degenerate random variables. Then*

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

if the following conditions are satisfied.

(i) $\max_{1\leq i\leq n} |X_i|/V_n \to 0$ *in probability;*

(ii) $\sum_{i=1}^{n} \left\{E(X_i/V_n)\right\}^2 \to 0;$

(iii) $E\left(\frac{S_n}{\max(V_n, a_n)}\right) \to 0$, *where* a_n *satisfies*

$$\sum_{i=1}^{n} E\left(\frac{X_i^2}{a_n^2 + X_i^2}\right) = 1. \tag{1}$$

- Remark: Theorem remains valid if (iii) is replaced by

$$(iii)^* \quad E(S_n/V_n) \to 0.$$

- Remark: Theorem remains valid if (iii) is replaced by

  $(iii)^* \quad E(S_n/V_n) \to 0.$

  However, $(iii)^*$ is not a necessary condition.

- Example 2.

  Let $Y_i := Y_{n,i}$, $1 \le i \le n$ be i.i.d. with pdf

$$P(Y_i = 1/n) = 1 - \frac{\ln n}{4n}, \; P(Y_i = 1) = P(Y_i = -1) = \frac{\ln n}{8n}.$$

  Define $X_i = Y_i - E(Y_i)$. Then we have

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

  but

$$E(S_n/V_n) \to \infty$$

Theorem (2)

If (i) is satisfied, then (ii) and (iii) are necessary for the self-normalized CLT.

Theorem (2)

If (i) is satisfied, then (ii) and (iii) are necessary for the self-normalized CLT.

Conjecture: Conditions (i), (ii) and (iii) are necessary and sufficient for the self-normalized CLT.

A key step of the proof is the following results:

Proposition (1)

Let $\{X_n, n \geq 1\}$ *be independent non-degenerate random variables satisfying*

$$\max_{i \leq n} |X_i|/V_n \to 0 \text{ in probability}$$

Let $a_n > 0$ *satisfy*

$$\sum_{i=1}^{n} E\left(\frac{X_i^2}{a_n^2 + X_i^2}\right) = 1$$

Then

$$V_n/a_n \to 1 \text{ in probability}$$

Proposition (2)

Assume that

$$\max_{1\le i\le n} |X_i|/V_n \to 0 \ \textit{in probability}$$

and let $a_n > 0$ *satisfy* $\sum_{i=1}^{n} E\left(\frac{X_i^2}{a_n^2 + X_i^2}\right) = 1$. *Then* $\sum_{i=1}^{n} \left\{E(X_i/V_n)\right\}^2 \to 0$ *is equivalent to*

$$\frac{1}{a_n^2}\sum_{i=1}^{n} (EX_i I_{\{|X_i|\le a_n\}})^2 \to 0$$

and $E\left(\frac{S_n}{\max(V_n, a_n)}\right) \to 0$ *is equivalent to*

$$\frac{1}{a_n}\sum_{i=1}^{n} EX_i I_{\{|X_i|\le a_n\}} \to 0.$$


Thank You