The CLT for Self-Normalized Sums

F. Götze  
Fakultät für Mathematik,  
Universität Bielefeld,  
Postfach 100131,  
33501 Bielefeld, Germany  
goetze@mathematik.uni-bielefeld.de

Let $X_1, X_2, \ldots$, denote independent identically distributed (i.i.d.) random variables such that $E X = 0$ and $0 < \sigma^2 = E X^2 < \infty$ and $\beta = E |X|^3$. Write

$$S_n = \sum_{j=1}^n X_j, \quad \text{and} \quad V_n^2 = \sum_{j=1}^n X_j^2,$$

Define the self-normalized sum as $S_n/V_n$, if $V_n \neq 0$, and 0 if $V_n = 0$. Self-normalized sums satisfy the CLT provided that $X$ belongs to the domain of attraction of a normal law (Maller (1981)). Griffin and Mason (1991) proved that the condition is necessary in the symmetric case. Gine, Götze and Mason (1997) proved that the condition is necessary for non-symmetric $X$ as well and showed that the tails of $S_n/V_n$ are uniformly sub-Gaussian in cases when the sequence $S_n/V_n$ is stochastically bounded. This observation suggests to consider the following explicit non-uniform bounds.

Define $\overline{\Phi}(x) = \Phi(x)$ if $x < 0$ and $\overline{\Phi}(x) = 1 - \Phi(x)$ otherwise. Consider bounds of the form

$$\left| P \left( S_n < x \sigma \sqrt{n} \, | \, \Phi(x) \right) - \overline{\Phi}(x) r_n(x) \right| \leq c \overline{\Phi}(x) r_n(x)$$
(1)

for some $r_n(x)$ such that $\sup_x \overline{\Phi}(x) r_n(x) = O(n^{-1/2})$. Hence the following results are moderate deviation results as well. Notice that for $|x| > \sqrt{n}$ we may chose $r_n(x) = 1$ since $P \left( |S_n|/V_n > |x| \right) = 0$.

The following results are joint work with G. P. Chistyakow, Institute for Low Temperature Physics and Engineering, National Academy of Sciences of Ukraine, Kharkov.

**Theorem 1.** Assume that $X$ is a symmetric random variable such that $\sigma^2 = 1$ and $\beta < \infty$. Then the relation (1) holds with

$$r_n(x) = \begin{cases}  
\beta/\sqrt{n}, & \text{for } |x| \leq 1, \\
|x|^3/\beta/\sqrt{n}, & \text{for } 1 \leq |x| \leq n^{1/6}/\beta^{1/3}, \\
1, & \text{for } |x| > n^{1/6}/\beta^{1/3}.
\end{cases}$$
(2)

Theorem 1 extends to the case of non-symmetric random variables as follows.

**Theorem 2.** Let $X$ be a random variable such that $E X = 0$, $\sigma^2 = 1$, and $\gamma = E |X|^{10/3} < \infty$. Then the relation (1) holds with

$$r_n(x) = \begin{cases}  
(\beta + \gamma \beta^{-1/3})(1 + |x|^3)/\sqrt{n}, & \text{for } |x| \leq n^{1/6}/\beta^{1/3}, \\
\exp \left\{ 65 \beta |x|^3/\sqrt{n} \right\}, & \text{for } n^{1/6}/\beta^{1/3} < |x| \leq \sqrt{n}/(2^{11} \beta), \\
|x| \exp \left\{ x^2/2 - |x|\sqrt{n}/(2^{15} \beta^2) \right\}, & \text{for } |x| > \sqrt{n}/(2^{11} \beta),
\end{cases}$$
(3)
For the case of non-i.i.d. random variables let \( X_1, \ldots, X_n \) denote independent random variables with distribution functions \( F_1, \ldots, F_n \) respectively. Assume that

\[
E X_j = 0, \quad 0 < \sigma_j^2 = E X_j^2 < \infty, \quad \text{and} \quad \beta_j = E |X_j|^3
\]

for all \( j \). Write

\[
B_n^2 = \sum_{j=1}^n \sigma_j^2, \quad A_n = \sum_{j=1}^n \beta_j, \quad L_n = A_n / B_n^2.
\]

**Theorem 3.** Let \( X_1, \ldots, X_n \) be symmetric independent random variables such that \( L_n < \infty \). Then the relations (2) and (3) hold with

\[
G(x) = \begin{cases} 
L_n, & \text{for } |x| \leq 1, \\
|x|^3 L_n, & \text{for } 1 \leq |x| \leq L_n^{-1/3}, \\
1, & \text{for } |x| > L_n^{-1/3}.
\end{cases}
\]

A somewhat analogous result to Theorem 2 holds for the non-identically distributed case as well.

Berry–Esseen bounds of increasing generality and accuracy were obtained in Hall (1987), Slavova (1985), Sharakhmetov (1995), Bentkus and Götze (1996a) (i.i.d. case), Bentkus, Bloznelis and Götze (1996b) (non-i.i.d. case).

Sub-Gaussian type behavior of tails related to self-normalized sums was observed in Logan, Mallows and Rice (1973). Shao (1997, 1998, 1999) considered large and moderate deviations. In particular, he proved \( r_n(x) = o(1) \) in the zone \( x = o(n^{1/6}) \). Wang and Jing (1999) proved a bound for symmetric non-identically distributed random variables which contains for \( |x| \geq 1 \) an extra factor \( |x| \) in bounds of type (2) and an extra factor \( \exp \left\{ \varepsilon x^2 \right\} \), \( \varepsilon > 0 \), in the non-symmetric case of (3).

**Applications**

Consider Student’s t-statistic \( T_n \) defined by

\[
T_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n T_j / \left( \frac{1}{n-1} \sum_{j=1}^n (X_j - \overline{X}_n)^2 \right)^{1/2}
\]

where \( \overline{X}_n = S_n/n \). Let \( q_n(x) := \left( \frac{n}{n + x^2 - 1} \right)^{1/2} \). It is well known that for \( x \geq 0 \)

\[
P (T_n \geq x) = P \left( \frac{S_n}{V_n} \geq x q_n(x) \right).
\]

Denote \( \Delta_{T_n}(x) = |P (T_n < x) - \Phi(x)| \). With the help of (5) we obtain the following non-uniform bounds for \( T_n \) as an immediate consequence of Theorem 2 and 3.

**Theorem 4.** Under the assumptions of Theorem 3, we have, for all positive integers \( n \) and real \( x \),

\[
\Delta_{T_n}(x) \leq c \min \left\{ (1 + |x|^3) L_n, 1 \right\} \Phi(x q_n(x)).
\]
Under the assumptions of Theorem 2, we have, for all positive integers $n$ and real $x$ such that $|x| \leq \sqrt{n}/(2^{11} \beta)$,

$$\Delta T_n(x) \leq c (\beta + \gamma \beta^{-1/3}) \frac{1 + |x|^2}{\sqrt{n}} \exp \left\{ -x^2 q_n(x)^2 \left( 1 - 65 \beta \frac{|x| q_n(x)}{\sqrt{n}} \right) \right\},$$

whereas for $|x| > \sqrt{n}/(2^{11} \beta)$, we have

$$\Delta T_n(x) \leq c \exp \left\{ - \frac{\sqrt{n} |x|}{215 \beta^2} q_n(x) \right\}.$$

REFERENCES


