Recent Progress on Self-normalized Cram’er Type Moderate Deviations

Qi-Man Shao

(The Chinese University of Hong Kong)

qmshao@cuhk.edu.hk

The research is partly supported by HK RGC GRF 403513 and 14302515
Consider a population with mean $\mu$ and variance $\sigma^2$. We would like to test

$$H_0 : \mu = \mu_0, \quad \text{vs} \quad H_1 : \mu > \mu_0$$
Consider a population with mean $\mu$ and variance $\sigma^2$. We would like to test

$$H_0 : \mu = \mu_0, \quad \text{vs} \quad H_1 : \mu > \mu_0$$

Let $X_1, X_2, \cdots, X_n$ be a random sample from the population and let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$
Test statistics

- **z-statistic**: $\sigma$ is known

$$Z_n = \frac{\sqrt{n(\bar{X} - \mu_0)}}{\sigma}$$

- **Student’s t-statistic**: $\sigma$ is unknown

$$T_n = \frac{\sqrt{n(\bar{X} - \mu_0)}}{\hat{\sigma}},$$

where $\hat{\sigma}^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. 
More generally, let $H_n = H_n(\theta, \lambda)$ be a statistic under consideration, where $\theta$ contains parameters of interest and $\lambda$ is a vector of some unknown nuisance parameters.
More generally, let $H_n = H_n(\theta, \lambda)$ be a statistic under consideration, where $\theta$ contains parameters of interest and $\lambda$ is a vector of some unknown nuisance parameters.

- **Self-normalized or Studentized statistic:**

$$\hat{H}_n = H_n(\theta, \hat{\lambda}),$$

where $\hat{\lambda}$ is an estimator of $\lambda$. 
Examples:

- Student t-statistic
- Hotelling’s $T^2$ statistic
- Studentized U-statistics
- The largest eigenvalue of sample correlation matrices
- The Wald t-ratio statistic in the unit root test
- ...
The $p$-value of the test:

Assume that the $p$-value of the test is

$$P(\hat{H}_n \geq h_n),$$

where $h_n$ is the observed value of $\hat{H}_n$. 
The $p$-value of the test:

Assume that the $p$-value of the test is

$$ P(\hat{H}_n \geq h_n), $$

where $h_n$ is the observed value of $\hat{H}_n$.

The true $p$-value is often unknown!
Assume that $H_n$ and $\hat{H}_n$ converge to $Z$ in distribution, where $Z$ is a continuous random variable. Then

$$\sup_x |P(H_n \geq x) - P(Z \geq x)| \to 0,$$

$$\sup_x |P(\hat{H}_n \geq x) - P(Z \geq x)| \to 0.$$
How accurate is your estimated $p$-value? Are you sure you can use your estimated $p$-value?

By the weak convergence,

$$\Pr(\hat{H}_n \geq h_n) - \Pr(Z \geq h_n) \to 0$$

Is it true that

$$\Pr(\hat{H}_n \geq h_n) \Pr(Z \geq h_n) \to 1$$?
How accurate is your estimated $p$-value? Are you sure you can use your estimated $p$-value?

By the weak convergence,

$$P(\hat{H}_n \geq h_n) - P(Z \geq h_n) \rightarrow 0$$

Is it true that

$$\frac{P(\hat{H}_n \geq h_n)}{P(Z \geq h_n)} \rightarrow 1$$
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

$$a_n/b_n \to 1$$
The **key** to answer the question is in the **Cramér moderate deviation**: 

Let $W_n = H_n$ or $\hat{H}_n$. Find the largest possible $c_n$ so that 

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

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

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed (i.i.d.)
random variables with $E X_1 = 0$ and $\text{Var}(X_1) = \sigma^2$, Recall
\[ Z_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n} \sigma} \]
and
\[ T_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n} \hat{\sigma}}, \]
where $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. 

2. The Cramér Moderate Deviation
Cramér moderate deviation for z-statistic

- Cramér (1938):

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

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

where $\lambda(t)$ is the Cramér’s series, and $\Phi(x)$ is the standard normal distribution function.
Harald Cramér
Linnik (1961):

If $Ee^{t_0 \sqrt{|X_1|}} < \infty$ for $t_0 > 0$, then

$$P\left( Z_n \geq x \right) / (1 - \Phi(x)) \to 1$$

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

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

for $0 \leq x \leq n^{1/6}$.
Linnik (1961):

If $Ee^{t_0\sqrt{|X_1|}} < \infty$ for $t_0 > 0$, then

$$P\left(Z_n \geq x\right)/(1 - \Phi(x)) \to 1$$

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

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

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.
Cramér moderate deviation for t-statistic

- **Shao (1999):** If $E|X_1|^3 < \infty$, then

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

  uniformly in $x \in [0, o(n^{1/6})]$.
Cramér moderate deviation for t-statistic

- **Shao (1999):** If $E|X_1|^3 < \infty$, then

  \[
  \frac{P(T_n \geq x)}{1 - \Phi(x)} \to 1
  \]

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

- **Jing, Shao and Wang (2003):** If $E|X_1|^3 < \infty$, then

  \[
  \frac{P(T_n \geq x)}{1 - \Phi(x)} = 1 + O(1) \frac{(1 + x)^3 E|X_1 - \mu|^3}{\sqrt{n} \sigma^3}
  \]

  for $0 \leq x \leq n^{1/6} \sigma / (E|X_1 - \mu|^3)^{1/3}$, where $|O(1)| \leq C$. 

Remark. Condition $E|X_1|^3 < \infty$ is necessary.
Cramer moderate deviation for t-statistic

- **Shao (1999):** If $E|X_1|^3 < \infty$, then
  
  $$\frac{P(T_n \geq x)}{1 - \Phi(x)} \to 1$$

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

- **Jing, Shao and Wang (2003):** If $E|X_1|^3 < \infty$, then
  
  $$\frac{P(T_n \geq x)}{1 - \Phi(x)} = 1 + O(1) \frac{(1 + x)^3 E|X_1 - \mu|^3}{\sqrt{n}\sigma^3}$$

  for $0 \leq x \leq n^{1/6}\sigma/(E|X_1 - \mu|^3)^{1/3}$, where $|O(1)| \leq C$.

- **Remark.** Condition $E|X_1|^3 < \infty$ is necessary.
Relationship between t-statistic and the self-normalized sum

Without loss of generality, assume $\mu = 0$. Put

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

**Self-normalized sum:** $S_n/V_n$

It is easy to see that

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

and hence

$$\{ T_n \geq x \} = \left\{ \frac{S_n}{V_n} \geq x \left( \frac{n}{n + x^2 - 1} \right)^{1/2} \right\}.$$
Let $\xi_i, 1 \leq i \leq n$ be independent random variables with $E(\xi_i) = 0$ and $E|\xi_i|^3 < \infty$. Without loss of generality, assume

$$\sum_{i=1}^{n} E\xi_i^2 = 1.$$ 

Let

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

- **Jing-Shao-Wang (2003):**

$$\frac{P(S_n/V_n \geq x)}{1 - \Phi(x)} = 1 + O(1)(1 + x^3)L_n$$

for $0 \leq x < 1/L_n^{1/3}$, where $L_n = \sum_{i=1}^{n} E|\xi_i|^3$. 

A more general result

Let \((\xi_i, \eta_i), 1 \leq i \leq n\) be independent random vectors with 
\(E\xi_i = 0, E|\xi_i|^3 < \infty, E|\eta_i|^3 < \infty, E\exp\left(\frac{\xi_i^2}{\eta_i^2+c_0E\eta_i^2}\right) < \infty\) for some constant \(c_0 \geq 0\). Without loss of generality, assume

\[
\sum_{i=1}^{n} E\xi_i^2 = 1 = \sum_{i=1}^{n} E\eta_i^2.
\]

Let

\[
S_n = \sum_{i=1}^{n} \xi_i, \quad V_n^2 = \sum_{i=1}^{n} \eta_i^2,
\]

We are interested in the Cram’er type moderate deviation for \(S_n/V_n\).
Gao - Shao - Shi (2017):

\[
P(S_n \geq xV_n + c) \frac{1}{1 - \Phi(x + c)} = \Psi_x e^{O(1)(\delta_x + r_x)} \left( 1 + O(1)(1 + x)L_n \right),
\]

where

\[
\Psi_x = \exp \left( \frac{4}{3} \gamma^3 x^3 \sum_{i=1}^{n} E\xi_i^3 - 2\gamma^2 x^3 \sum_{i=1}^{n} E\xi_i\eta_i^2 \right)
\]

and \( \gamma = \frac{1}{2} (1 + \frac{c}{x}) \), uniformly for \( |c| \leq x/5 \) and for all \( x \geq 0 \) satisfying

\[
xL_n \leq c_1, \quad r_x \leq c_1(1 + x^2), \quad \max_i r_{x,i} \leq 1/8,
\]

\[
x \max_i (E(|\xi_i|^3 + |\eta_i|^3))^{1/3} \leq \min(1/4, 1/c_0)
\]
Gao - Shao - Shi (2017):

\[
P(S_n \geq xV_n + c) \frac{1}{1 - \Phi(x + c)} = \Psi_x e^{O(1)(\delta_x + r_x)} \left(1 + O(1)(1 + x)L_n\right),
\]

where

\[
\Psi_x = \exp \left(\frac{4}{3} \gamma^3 x^3 \sum_{i=1}^{n} E\xi_i^3 - 2\gamma^2 x^3 \sum_{i=1}^{n} E\xi_i \eta_i^2\right)
\]

and \(\gamma = \frac{1}{2}(1 + \frac{c}{x})\), uniformly for \(|c| \leq x/5\) and for all \(x \geq 0\) satisfying

\[
xL_n \leq c_1, \quad r_x \leq c_1(1 + x^2), \quad \max_i r_{x,i} \leq 1/8,
\]

\[
x \max_i (E(|\xi_i|^3 + |\eta_i|^3))^{1/3} \leq \min(1/4, 1/c_0)
\]
Notations:

\[ L_n = \sum_{i=1}^{n} E(|\xi_i|^3 + |\eta_i|^3), \]

\[ \delta_{x,i} = f((1 + x)\xi_i) + f((1 + x)\eta_i), \]

where \( f(\xi) = E|\xi|^3 I\{|\xi| > 1\} + E|\xi|^4 I\{|\xi| \leq 1\} \)

\[ r_{x,i} = Ee^{\frac{\xi_i^2}{\eta_i^2 + c_0 E\eta_i^2}} I\{(1 + x)|\xi_i| > 1\}, \]

\[ \delta_x = \sum_{i=1}^{n} \delta_{x,i}, \quad r_x = \sum_{i=1}^{n} r_{x,i} \]
Notations:

\[
L_n = \sum_{i=1}^{n} E(|\xi_i|^3 + |\eta_i|^3),
\]

\[
\delta_{x,i} = f((1 + x)\xi_i) + f((1 + x)\eta_i),
\]

where \( f(\xi) = E|\xi|^3 I\{|\xi| > 1\} + E|\xi|^4 I\{|\xi| \leq 1\} \)

\[
r_{x,i} = Ee^{\frac{\xi_i^2}{\eta_i^2 + c_0^2 E\eta_i^2}} I\{|(1 + x)\xi_i| > 1\},
\]

\[
\delta_x = \sum_{i=1}^{n} \delta_{x,i}, \quad r_x = \sum_{i=1}^{n} r_{x,i}
\]

- The result has been applied to establish self-normalized Cramér moderate deviation for dependent random variables.
Let $\xi_1, \ldots, \xi_n$ be independent random variables with $E\xi_i = 0$ and $E\xi_i^2 < \infty$ satisfying
\[
\sum_{i=1}^{n} E\xi_i^2 = 1.
\]

Let
\[
S_n = \sum_{i=1}^{n} \xi_i, \quad V_n^2 = \sum_{i=1}^{n} \xi_i^2
\]
and $D_1, D_2$ be measurable functions of $\{\xi_i, 1 \leq i \leq n\}$. 

4. Cramér Moderate Deviation for Self-Normalized Processes
4. Cramér Moderate Deviation for Self-Normalized Processes

Let $\xi_1, \ldots, \xi_n$ be independent random variables with $E\xi_i = 0$ and $E\xi_i^2 < \infty$ satisfying

$$\sum_{i=1}^{n} E\xi_i^2 = 1.$$ 

Let

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

and $D_1, D_2$ be measurable functions of $\{\xi_i, 1 \leq i \leq n\}$. Consider the self-normalized process:

$$T_n = \frac{S_n + D_1}{V_n(1 + D_2)^{1/2}}.$$
Shao and Wenxin Zhou (2016):

There is an absolute constant $C > 1$ such that

$$e^{O(1)\Delta_{n,x}}(1 - CR_{n,x}) \leq \frac{P(T_n \geq x)}{1 - \Phi(x)}$$

and

$$P(T_n \geq x) \leq (1 - \Phi(x))e^{O(1)\Delta_{n,x}}(1 + CR_{n,x})$$

$$+ P(|D_1|/V_n > 1/(2x)) + P(|D_2| > 1/(2x^2))$$

for all $x > 1$ satisfying

$$\Delta_{n,x} \leq (1 + x)^2/C, \ x^2 \max_{1 \leq i \leq n} E\xi_i^2 \leq 1,$$
where

\[ \Delta_{n,x} = x^2 \sum_{i=1}^{n} E\xi_i^2 I(x|\xi_i| > 1) + x^3 \sum_{i=1}^{n} E|\xi_i|^3 I(x|\xi_i| \leq 1), \]

\[ R_{n,x} = I_{n,0}^{-1} \left\{ xE(|D_1| + x|D_2|)e^{\sum_{j=1}^{n} (x\xi_j - x^2\xi_j^2/2)} \right. \]

\[ + x \sum_{i=1}^{n} E(|\xi_i(D_1 - D_1^{(i)})| + x|\xi_i(D_2 - D_2^{(i)})|)e^{\sum_{j \neq i} (x\xi_j - x^2\xi_j^2/2)} \left. \right\}, \]

\[ I_{n,0} = \prod_{i=1}^{n} E e^{x\xi_i - x^2\xi_i^2/2}, \]

and \( D_1^{(i)} \) and \( D_2^{(i)} \) are any random variables that don’t depend on \( \xi_i \).
Studentized U-statistics

Let $X, X_1, X_2, \ldots, X_n$ be i.i.d random variables, and let $h(x, y)$ be a symmetric kernel, i.e., $h(x, y) = h(y, x)$. $\theta = Eh(X_1, X_2)$. 

U-statistic (Hoeffding (1948)):

$$U_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j)$$

The standardized U-statistic:

$$\sqrt{n} \frac{\sigma_1}{\sigma_1} (U_n - \theta)$$

where $\sigma_1^2 := \text{Var}(g(X)) > 0$ and $g(x) = E[h(x, X)]$. 

Studentized U-statistics

Let $X, X_1, X_2, \ldots, X_n$ be i.i.d random variables, and let $h(x, y)$ be a symmetric kernel, i.e., $h(x, y) = h(y, x)$. $\theta = E h(X_1, X_2)$.

$U$-statistic (Hoeffding (1948)):

$$U_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j)$$

The standardized $U$-statistic:

$$\frac{\sqrt{n}}{2 \sigma_1} (U_n - \theta).$$

where $\sigma_1^2 := \text{Var}(g(X)) > 0$ and $g(x) = E(h(x, X))$. 
Studentized $U$-statistic:

$$T_n = \frac{\sqrt{n}}{2 s_1} (U_n - \theta),$$

where

$$s_1^2 = \frac{(n-1)}{(n-2)^2} \sum_{i=1}^{n} \left( \frac{1}{n-1} \sum_{j \neq i} h(X_i, X_j) - U_n \right)^2.$$
Studentized $U$-statistic:

$$T_n = \frac{\sqrt{n}}{2 s_1} (U_n - \theta),$$

where

$$s_1^2 = \frac{(n - 1)}{(n - 2)^2} \sum_{i=1}^{n} \left( \frac{1}{n - 1} \sum_{j \neq i} h(X_i, X_j) - U_n \right)^2.$$

Hoeffding’s decomposition: (assume $\theta = 0$)

$$T_n = \frac{S_n + D_1}{V_n(1 + D_2)^{1/2}},$$

where

$$S_n = \sum_{i=1}^{n} \xi_i, \quad \xi_i = g(X_i) / (\sigma_1 \sqrt{n}), \quad V_n^2 = \sum_{i=1}^{n} \xi_i^2,$$

$D_1$ and $D_2$ are small.
Lai, Shao and Wang (2011):

Assume that $\sigma_1 > 0$ and $E|h(X_1, X_2)|^3 < \infty$. If

$$h^2(x_1, x_2) \leq c_0(\sigma_1^2 + g^2(x_1) + g^2(x_2))$$

for some $c_0 > 0$, then

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

holds uniformly in $x \in [0, o(n^{1/6}))$. 
Lai, Shao and Wang (2011):

Assume that $\sigma_1 > 0$ and $E|h(X_1, X_2)|^3 < \infty$. If

$$h^2(x_1, x_2) \leq c_0(\sigma_1^2 + g^2(x_1) + g^2(x_2))$$

for some $c_0 > 0$, then

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

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

Shao and Wenxin Zhou (2016):

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

for $x \in [0, o(n^{1/6}))$. 
Let \( \{(\xi_i, \mathcal{F}_i), 1 \leq i \leq n\} \) be a \textit{martingale difference}. Put

\[
S_n = \sum_{i=1}^{n} \xi_i, \quad V_n^2 = \sum_{i=1}^{n} \xi_i^2, \quad U_n^2 = \sum_{i=1}^{n} E(\xi_i^2 | \mathcal{F}_{i-1}).
\]

\textbf{Fan-Crama-Liu-Shao (2017+):}

Assume that there are constants \( \delta_n \) and \( \varepsilon_n \) such that

\[
|U_n^2 - 1| \leq \delta_n^2
\]

and

\[
\sum_{i=1}^{n} E(|\xi_i|^3 | \mathcal{F}_{i-1}) \leq \varepsilon_n.
\]

Then

\[
\frac{P(S_n/V_n \geq x)}{1 - \Phi(x)} \rightarrow 1
\]

uniformly for \( 0 \leq x \leq o(1/(\varepsilon_n + \delta_n)) \).
Let \(\{(\xi_i, \mathcal{F}_i), \ 1 \leq i \leq n\}\) be a martingale difference. Put
\[
S_n = \sum_{i=1}^{n} \xi_i, \quad V_n^2 = \sum_{i=1}^{n} \xi_i^2, \quad U_n^2 = \sum_{i=1}^{n} E(\xi_i^2 \mid \mathcal{F}_{i-1})
\]

Fan-Crama-Liu-Shao (2017+):
Assume that there are constants \(\delta_n\) and \(\varepsilon_n\) such that
\[
|U_n^2 - 1| \leq \delta_n^2
\]
and
\[
\sum_{i=1}^{n} E(|\xi_i|^3 \mid \mathcal{F}_{i-1}) \leq \varepsilon_n.
\]

Then
\[
\frac{P(S_n/V_n \geq x)}{1 - \Phi(x)} \rightarrow 1
\]
uniformly for \(0 \leq x \leq o(1/(\varepsilon_n + \delta_n))\).

Application to dependent random variables can be obtained.
Let $X_1, \ldots, X_n$ be a random sample from a population with distribution function $F$ and density function $f$, and let $X_{n,1} \leq X_{n,2} \leq \ldots \leq X_{n,n}$ be the order statistics. The $p$th quantile $\xi_p$ is defined by

$$F(\xi_p) = p.$$ 

**Aim:** Estimate $\xi_p$.

- **Estimator:** $\hat{\xi}_p = X_{n,k}$, with $k = [np] + 1$. 

6. Self-normalized Quantile Estimator
Asymptotic normality:

$$W_n := \frac{\sqrt{n} f(\xi_p) (X_{n,k} - \xi_p)}{\sqrt{p(1 - p)}} \overset{d.}{\to} N(0, 1).$$
• Asymptotic normality:

\[ W_n := \frac{\sqrt{n} f(\xi_p)(X_{n,k} - \xi_p)}{\sqrt{p(1 - p)}} \xrightarrow{d.} N(0, 1). \]

• Studentized quantile:

\[ T_n := \frac{X_{n,k} - \xi_p}{\sqrt{np(1 - p)(X_{n,k+m} - X_{n,k})/m}}. \]
• Asymptotic normality:

\[ W_n := \frac{\sqrt{n} f(\xi_p) (X_{n,k} - \xi_p)}{\sqrt{p(1-p)}} \xrightarrow{d.} N(0, 1). \]

• Studentized quantile:

\[ T_n := \frac{X_{n,k} - \xi_p}{\sqrt{np(1-p)(X_{n,k+m} - X_{n,k})/m}}. \]

• Limiting distribution:

\[ T_n \xrightarrow{d.} \frac{N(0, 1)}{\Gamma_m/m} := \zeta_m, \]

where \( \Gamma_m \) denotes a Gamma distribution with \( m \) degrees of freedom, \( N(0, 1) \) and \( \Gamma_m \) are independent.
Berry-Esseen Bound

- **Assumptions:**

  (i) $f(\xi_p) > 0$ and $f'$ exists in some neighborhood of $\xi_p$,

  (ii) there exists $0 < \tau < 1$ such that

  $$\sup_{q \in (p-\tau, p+\tau) \cap (0,1)} \frac{|f'(\xi_q)|}{f^3(\xi_q)} < \frac{1}{5 \tau f(\xi_p)}$$

  (iii) $m < n(1 - p)/2$. 
Gao - Shao - Shi (2017):

There exists an absolute constant $C$ such that

$$\sup_x |P(T_n \leq x) - P(\zeta_m \leq x)| \leq C \left( \frac{\sqrt{p(1-p)}}{\tau \sqrt{n}} + \frac{p \sqrt{m} + 1}{\sqrt{np(1-p)}} \right)$$
A refined limiting distribution

Let
\[
\tilde{\zeta}_m^* = \frac{N(pm/\sigma_n, \sqrt{1 - p^2m/\sigma_n^2})}{\Gamma_m/m} - \frac{pm}{\sigma_n}.
\]

where \( \sigma_n^2 = np(1 - p) \).

- Gao-Shao-Shi (2017):

\[
\sup_x \left| P(T_n \leq x) - P(\tilde{\zeta}_m^* \leq x) \right| \\
\leq C \left( \frac{\sigma_n + m}{\tau n} + \frac{|1 - 2p|}{\sigma_n} + \frac{p^2m + 1}{\sigma_n^2} + \frac{p(1 - p)}{\tau^2 n} \right).
\]

- When \( 1 \leq m = O(\sqrt{n}) \), the overall error is of order \( n^{-1/2} \).
Gao, Shao, Shi (2017):
Assume that $1 \leq m = o(n^{2/5})$. Then

$$\frac{P(T_n \geq x)}{P(\zeta^*_m \geq x)} \rightarrow 1$$

for $0 \leq x \leq o(\sqrt{n})$. 
Let $d \geq 2$ and $X$ be a $d \times 1$ random vector with mean vector $\mu$ and non-degenerate covariance matrix $\Sigma$. Let $X_1, X_2, \ldots, X_n$ be a random sample of $n(n > d)$ independent observations of $X$.

**Hotelling’s $T^2$ statistic:**

$$T_n^2 = (S_n - n\mu)'\bar{V}_n^{-1}(S_n - n\mu),$$

where

$$S_n = \sum_{i=1}^{n} X_i, \quad \bar{X} = S_n/n, \quad \bar{V}_n = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'.$$
Let $d \geq 2$ and $X$ be a $d \times 1$ random vector with mean vector $\mu$ and non-degenerate covariance matrix $\Sigma$. Let $X_1, X_2, \ldots, X_n$ be a random sample of $n(n > d)$ independent observations of $X$.

**Hotelling’s $T^2$ statistic:**

$$T_n^2 = (S_n - n\mu)'\bar{V}_n^{-1}(S_n - n\mu),$$

where

$$S_n = \sum_{i=1}^{n} X_i, \quad \bar{X} = S_n/n, \quad \bar{V}_n = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'.$$

- $T_n^2$ has a limiting $\chi^2$-distribution with $d$ degrees of freedom.
Dembo and Shao (2006): Assume $\mu = 0$. For any $x_n \to \infty$ and $x_n = o(n)$, we have

$$\ln P\left(T_n^2 \geq x_n\right) \sim -\frac{1}{2}x_n$$
Dembo and Shao (2006): Assume $\mu = 0$. For any $x_n \to \infty$ and $x_n = o(n)$, we have

$$\ln P\left(T_n^2 \geq x_n\right) \sim -\frac{1}{2}x_n$$

Liu and Shao (2013):

- Suppose that $E\|X\|^{3+\delta} < \infty$ for some $\delta > 0$. Then

$$\frac{P\left(T_n^2 \geq x\right)}{P\left(\chi^2(d) \geq x\right)} \to 1$$

uniformly for $x \in [0, o(n^{1/3})]$.  
- Similar result holds for two-sample Hotelling $T^2$ statistic
Conjecture: If $E|X|^3 < \infty$, then

$$\lim_{n \to \infty} \frac{P(T_n^2 \geq x)}{P(\chi_d^2 \geq x)} = 1$$

holds uniformly in $0 \leq x \leq o(n^{1/3})$; Moreover

$$P(T_n^2 \geq x)/P(\chi_d^2 \geq x) = 1 + O(1) \frac{(1 + x)^{3/2} E|X|^3}{n^{1/2} |\Sigma|^{3/2}}.$$
Conclusion:

Many limit theorems hold under self-normalization, which require no moment assumptions or much less moment assumptions than those under the regular standardization. Self-normalized limit theory can provide theoretical justifications for the use of many commonly used statistics in practice.