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 μ and variance σ^2. We would like to test

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

Consider a population with mean μ and variance σ^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: σ is known

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

  - Student's t-statistic: σ 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 θ contains parameters of interest and λ is a vector of some unknown nuisance parameters.

More generally, let $H_n = H_n(\theta, \lambda)$ be a statistic under consideration, where θ contains parameters of interest and λ is a vector of some unknown nuisance parameters.

- Self-normalized or Studenized statistic:

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

where $\widehat{\lambda}$ is an estimator of λ.

- 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

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

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

- The p-value of the test:

Assume that the p-value of the test is

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

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

The true p-value is often unknown!

Estimating p-value:

Assume that H_n and $\widehat{H}_n$ converge to Z in distribution, where Z is a continuous random variable. Then

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

$$\sup_x |P(\widehat{H}_n \geq x) - P(Z \geq x)| \to 0.$$

- Estimated p-value:

$$P(Z \geq h_n).$$

How accurate is your estimated p-value? Are you sure you can use your estimated p-value?

How accurate is your estimated p-value? Are you sure you can use your estimated p-value?

By the weak convergence,

$$P(\widehat{H}_n \geq h_n) - P(Z \geq h_n) \to 0$$

Is it true that

$$\frac{P(\widehat{H}_n \geq h_n)}{P(Z \geq h_n)} \to 1 \ ?$$

- A naive question:
  - Let $0 < a_n \le 1,\ 0 < b_n \le 1$. 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 $\widehat{H}_n$. 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 $EX_1 = 0$ 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 $\hat{\sigma}^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (X_i - \bar{X})^2.$

▶ 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 $x = o(n^{1/2})$

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

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

Harald Cramér

- Linnik (1961):

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

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

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

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

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

- Linnik (1961):

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

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

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

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

  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 $|O(1)| \leq C$.

▶ 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.

▶ 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

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

and hence

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

3. Self-normalized Cramér Moderate Deviation for Independent Random Variables

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 $(\xi_i, \eta_i), 1 \le i \le n$ be independent random vectors with $E\xi_i = 0,\ E|\xi_i|^3 < \infty, E|\eta_i|^3 < \infty, E\exp(\frac{\xi_i^2}{\eta_i^2 + c_0^2 E\eta_i^2}) < \infty$ for some constant $c_0 \ge 0$. 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):

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

where

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

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):

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

where

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

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:

$$
\begin{aligned}
L_n &= \sum_{i=1}^{n} E(|\xi_i|^3 + |\eta_i|^3), \\
\delta_{x,i} &= f((1+x)\xi_i) + f((1+x)\eta_i), \\
&\quad \text{where } f(\xi) = E|\xi|^3 I\{|\xi| > 1\} + E|\xi|^4 I\{|\xi| \le 1\} \\
r_{x,i} &= Ee^{\frac{\xi_i^2}{\eta_i^2 + c_0^2 E\eta_i^2}} I\{|(1+x)\xi_i| > 1\}, \\
\delta_x &= \sum_{i=1}^{n} \delta_{x,i}, \quad r_x = \sum_{i=1}^{n} r_{x,i}
\end{aligned}
$$

Notations:

$$
\begin{aligned}
L_n &= \sum_{i=1}^{n} E(|\xi_i|^3 + |\eta_i|^3), \\
\delta_{x,i} &= f((1+x)\xi_i) + f((1+x)\eta_i), \\
&\quad \text{where } f(\xi) = E|\xi|^3 I\{|\xi| > 1\} + E|\xi|^4 I\{|\xi| \le 1\} \\
r_{x,i} &= E e^{\frac{\xi_i^2}{\eta_i^2 + c_0^2 E\eta_i^2}} I\{|(1+x)\xi_i| > 1\}, \\
\delta_x &= \sum_{i=1}^{n} \delta_{x,i}, \quad r_x = \sum_{i=1}^{n} r_{x,i}
\end{aligned}
$$

- The result has been applied to establish self-normalized Cramér moderate deviation for dependent random variables.

4. Cramér Moderate Deviation for Self-Normalized Processes

Let $\xi_1, ..., \xi_n$ 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 $\{\xi_i, 1 \le i \le n\}$.

4. Cramér Moderate Deviation for Self-Normalized Processes

Let $\xi_1, ..., \xi_n$ 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 $\{\xi_i, 1 \le i \le n\}$.

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

$$e^{O(1)\Delta_{n,x}}\big(1 - C\,R_{n,x}\big) \le \frac{P(T_n \ge x)}{1 - \Phi(x)}$$

  and

$$\begin{aligned} P(T_n \ge x) \quad &\le \quad \big(1 - \Phi(x)\big)e^{O(1)\Delta_{n,x}}\big(1 + C\,R_{n,x}\big) \\ &\quad +P\big(|D_1|/V_n > 1/(2x)\big) + P\big(|D_2| > 1/(2x^2)\big) \end{aligned}$$

  for all $x > 1$ satisfying

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

where

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

$$R_{n,x} = I_{n,0}^{-1}\Big\{ xE(|D_1| + x|D_2|)e^{\sum_{j=1}^{n}(x\xi_j - x^2\xi_j^2/2)}$$
$$+x\sum_{i=1}^{n} E(|\xi_i(D_1 - D_1^{(i)})| + x|\xi_i(D_2 - D_2^{(i)})|)e^{\sum_{j\neq i}^{n}(x\xi_j - x^2\xi_j^2/2)}\Big\},$$

$$I_{n,0} = \prod_{i=1}^{n} Ee^{x\xi_i - x^2\xi_i^2/2},$$

and $D_1^{(i)}$ and $D_2^{(i)}$ are any random variables that don't depend on ξ_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., $h(x, y) = h(y, x)$. $\theta = Eh(X_1, X_2)$.

▶ 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)):

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

The standardized U-statistic:

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

where $\sigma_1^2 := \operatorname{Var}(g(X)) > 0$ 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

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

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

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

  for some $c_0 > 0$, then

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

  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

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

  for some $c_0 > 0$, then

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

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

- Shao and Wenxin Zhou (2016):

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

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

5. Self-normalized Martingales

Let $\{(\xi_i, \mathcal{F}_i),\ 1 \le i \le n\}$ 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 δ_n and ε_n such that

  $$|U_n^2 - 1| \le \delta_n^2$$

  and

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

  Then

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

  uniformly for $0 \le x \le o(1/(\varepsilon_n + \delta_n))$.

5. Self-normalized Martingales

Let $\{(\xi_i, \mathcal{F}_i),\ 1 \le i \le n\}$ 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 δ_n and ε_n such that

  $$|U_n^2 - 1| \le \delta_n^2$$

  and

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

  Then

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

  uniformly for $0 \le x \le o(1/(\varepsilon_n + \delta_n))$.

- Application to dependent random variables can be obtained.

6. Self-normalized Quantile Estimator

Let $X_1, \cdots, X_n$ 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 pth quantile ξ_p is defined by

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

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

- Asymptotic normality:

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

- 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 Γ_m denotes a Gamma distribution with m degrees of freedom, $N(0,1)$ and Γ_m are independent.

Berry-Esseen Bound

- Assumptions:
  (i) $f(\xi_p) > 0$ and f' exists in some neighborhood of ξ_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) $m < n(1-p)/2$.

Berry-Esseen Bound

- Gao - Shao - Shi (2017):

  There exists an absolute constant C such that

  $$\sup_x |P(T_n \leq x) - P(\zeta_m \leq x)| \leq C\Big(\frac{\sqrt{p(1-p)}}{\tau\sqrt{n}} + \frac{p\sqrt{m}+1}{\sqrt{np(1-p)}}\Big)$$

▶ A refined limiting distribution

Let

$$\zeta_m^* = \frac{N(pm/\sigma_n, \sqrt{1 - p^2 m/\sigma_n^2})}{\Gamma_m/m} - \frac{pm}{\sigma_n}.$$

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

- Gao-Shao-Shi (2017):

$$\begin{aligned} \sup_x \left| P(T_n \le x) - P(\zeta_m^* \le x) \right| \\ \le \quad C \left(\frac{\sigma_n + m}{\tau n} + \frac{|1-2p|}{\sigma_n} + \frac{p^2 m + 1}{\sigma_n^2} + \frac{p(1-p)}{\tau^2 n} \right). \end{aligned}$$

- When $1 \le m = O(\sqrt{n})$, the overall error is of order $n^{-1/2}$.

Cramér-Type Moderate Deviation

- Gao, Shao, Shi (2017):
  Assume that $1 \leq m = o(n^{2/5})$. Then

$$\frac{P(T_n \geq x)}{P(\zeta_m^* \geq x)} \longrightarrow 1$$

for $0 \leq x \leq o(\sqrt{n})$.

7. Cramér moderate deviation for Hotelling's T^2 statistics

Let $d \geq 2$ and $\boldsymbol{X}$ be a $d \times 1$ random vector with mean vector $\boldsymbol{\mu}$ and non-degenerate covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample of $n(n > d)$ independent observations of $\boldsymbol{X}$.

Hotelling's T^2 statistic:

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

where

$$\boldsymbol{S}_n = \sum_{i=1}^{n} \boldsymbol{X}_i, \ \bar{\boldsymbol{X}} = \boldsymbol{S}_n / n, \ \bar{\boldsymbol{V}}_n = \sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'.$$

7. Cramér moderate deviation for Hotelling's T^2 statistics

Let $d \geq 2$ and $\boldsymbol{X}$ be a $d \times 1$ random vector with mean vector $\boldsymbol{\mu}$ and non-degenerate covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample of $n(n > d)$ independent observations of $\boldsymbol{X}$.

Hotelling's T^2 statistic:

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

where

$$\boldsymbol{S}_n = \sum_{i=1}^{n} \boldsymbol{X}_i,\ \bar{\boldsymbol{X}} = \boldsymbol{S}_n/n,\ \bar{\boldsymbol{V}}_n = \sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'.$$

- T_n^2 has a limiting χ^2-distribution with d degrees of freedom.

- Dembo and Shao (2006): Assume $\boldsymbol{\mu} = 0$. 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 $\boldsymbol{\mu} = 0$. 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 $E\|\boldsymbol{X}\|^{3+\delta} < \infty$ 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 $E|\boldsymbol{X}|^3 < \infty$, then

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

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

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

- 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.


Thank You