Cramér-Type Moderate Deviation Theorems for Two-Sample Studentized (Self-normalized) U-Statistics

Wen-Xin Zhou

Department of Mathematics and Statistics
University of Melbourne

Joint work with Prof. Qi-Man Shao

Program on Self-Normalized Asymptotic Theory in Probability, Statistics and Econometrics, IMS (NUS)
21 May, 2014

Outline

1. Mann-Whitney U-test: Motivating examples
2. Studentized U-statistics
3. Cramér-type moderate deviations for Studentized U-statistics
4. Refined moderate deviations for the two-sample t-statistic
5. Applications

1. Mann-Whitney U-test: Motivating examples

Assume that we observe two independent samples $X, X_1, \ldots, X_m \overset{i.i.d.}{\sim} P$ and $Y, Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} Q$, where both P and Q are continuous distribution functions. The Mann-Whitney U-test statistic is given by

$$U_{m,n} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I\{X_i \leq Y_j\}.$$

1. Mann-Whitney U-test: Motivating examples

Assume that we observe two independent samples $X, X_1, \ldots, X_m \overset{i.i.d.}{\sim} P$ and $Y, Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} Q$, where both P and Q are continuous distribution functions. The Mann-Whitney U-test statistic is given by

$$U_{m,n} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I\{X_i \leq Y_j\}.$$

Without continuity assumption, we could simply use

$$\widetilde{U}_{m,n} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(I\{X_i \leq Y_j\} + \tfrac{1}{2} I\{X_i = Y_j\} - \tfrac{1}{2}\right),$$

which takes into account the possibility of having ties.

The Mann-Whitney (M-W) test was originally introduced to test the null hypothesis $H_0 : P = Q$. Alternatively, the M-W test has been prevalently used for testing the equality of means or medians, as a nonparametric counterpart of the t-statistic.

- Advantages: Easy to implement, good efficiency, robustness against parametric assumptions, etc.
- Disadvantages (Chung and Romano, 2011): The M-W test is only valid if the fundamental assumption of identical distributions holds; that is, $P = Q$. Why?

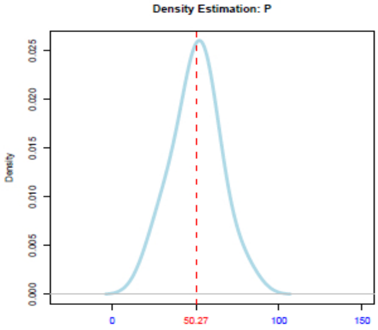
The Mann-Whitney (M-W) test was originally introduced to test the null hypothesis $H_0 : P = Q$. Alternatively, the M-W test has been prevalently used for testing the equality of means or medians, as a nonparametric counterpart of the t-statistic.

- Advantages: Easy to implement, good efficiency, robustness against parametric assumptions, etc.
- Disadvantages (Chung and Romano, 2011): The M-W test is only valid if the fundamental assumption of identical distributions holds; that is, $P = Q$. Why?
- Misapplication of the M-W test (a toy example): Assume $P \sim N(\log 2, 1)$, $Q \sim \exp(1)$, such that

$$\mathrm{Median}(P) = \mathrm{Median}(Q) = \log 2.$$

Set $\alpha = 0.05$, using the M-W test via Monte Carlo simulation shows that the rejection probability for a two-sided test is 0.2843.
In fact, the M-W test only captures the divergence from $P(X \leq Y) = \frac{1}{2}$.
However, for $X \sim N(\log 2, 1)$, $Y \sim \exp(1)$, $P(X < Y) = 0.4431$.

▶ A real data example (testing equality of means): Sutter (2009, AER) performed the M-W U-test to examine the effects of salient group membership on individual behavior, and compares them to those of team decision making. The average investments in PAY-COMM (payoff commonality) with 18 observations and MESSAGE (exchange of messages) with 24 observations are compared. Estimated densities (kernel method) for PAY-COMM and MESSAGE, denoted by P and Q, resp., are plotted in the Figure below:

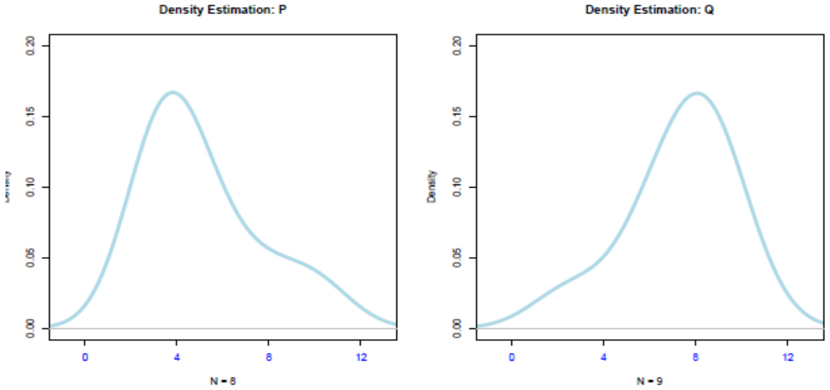


Using the M-W test, Sutter rejects the null hypothesis that the average investments in PAY-COMM and MESSAGE are the same at the 10% significance level, with a *p*-value of 0.069.

For the conventional 5% significance level, the M-W test would have failed to reject the null hypothesis.

Using the Studentized permutation *t*-test, however, yields a *p*-value of 0.042 and rejects the null hypothesis at the 5% significance level (Chung and Romano, 2011).

▶ The 3rd example (testing equality of distributions): Plott and Zeiler (2005, AER) used the M-W U-test to examine the null hypothesis that willingness to pay (WTP) and willingness to accept (WTA) are drawn from the same distribution. Estimated densities, P for WTP and Q for WTA, are plotted below:



The M-W test yields a z value of 1.738 (p-value $= 0.0821$), leading to a failure in rejecting the null hypothesis.

This is not a surprising outcome, as when testing equality of distributions, it is more recommended to use a statistic that captures the differences of the entire distributions, such as the Kolmogorov-Smirnov or the Cramér-von Mises statistic, in contrast to assessing divergence in a particular parameter of the distributions.

In the same example, the Cramér-von Mises statistic yields a p-value of 0.0546.

▶ Conclusion: The Mann-Whitney U-test has been misused in many experimental applications. As opposed to testing equality of medians, means or distributions, what the M-W test is actually testing is

$$H_0 : P(X \leq Y) = \tfrac{1}{2} \ \textit{versus} \ H_1 : P(X \leq Y) > \tfrac{1}{2}$$

or

$$H_0 : P(X \leq Y) \leq \tfrac{1}{2} \ \textit{versus} \ H_1 : P(X \leq Y) > \tfrac{1}{2}.$$

Testing H_0 arises in many applications including:

- Testing whether the physiological performance of an active drug is better than that under the control treatment;
- Testing the effects of a policy, such as unemployment insurance or a vocational training program, on the level of unemployment.

- Self-normalization: Even when testing the null $H_0 : P(X \le Y) = \frac{1}{2}$, the standard Mann-Whitney test is invalid unless it is appropriately Studentized.

- **Self-normalization:** Even when testing the null $H_0 : P(X \leq Y) = \frac{1}{2}$, the standard Mann-Whitney test is invalid unless it is appropriately Studentized.

  Recall that $X_1, \ldots, X_m \overset{i.i.d.}{\sim} P_X$ and $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} Q_Y$. Let

$$\min(m, n) \to \infty, \quad \tfrac{m}{m+n} \to \rho \in (0, 1).$$

  Then the distribution of $\sqrt{m}(U_{m,n} - \theta)$ is asymptotically normal with mean 0 and variance

$$\mathrm{Var}(Q_Y^-(X_i)) + \frac{\rho}{1-\rho}\mathrm{Var}(P_X(Y_j)),$$

  where $Q_Y^-(y) = Q_Y(-y)$.

  In other words, $\sqrt{m}(U_{m,n} - \theta)$ is not asymptotically pivotal.

2. Studentized U-statistics

Let $X, X_1, \ldots, X_n$ be i.i.d. random variables with distribution taking values in a measurable space $(\mathbf{X}, \mathcal{X})$, consider a U-statistic of the form

$$U_n = \frac{1}{\binom{n}{d}} \sum_{1 \le i_1 < \cdots < i_d \le n} h(X_{i_1}, \ldots, X_{i_d})$$

with a symmetric kernel $h : \mathbf{X}^d \mapsto \mathbb{R}$, where $1 \le d \le n/2$. Write

$$\theta = Eh(X_1, \ldots, X_d), \;\; h_1(x) = E(h(X_1, \ldots, X_d) - \theta | X_d = x).$$

2. Studentized U-statistics

Let $X, X_1, \ldots, X_n$ be i.i.d. random variables with distribution taking values in a measurable space $(\mathbf{X}, \mathcal{X})$, consider a U-statistic of the form

$$U_n = \frac{1}{\binom{n}{d}} \sum_{1 \le i_1 < \cdots < i_d \le n} h(X_{i_1}, \ldots, X_{i_d})$$

with a symmetric kernel $h : \mathbf{X}^d \mapsto \mathbb{R}$, where $1 \le d \le n/2$. Write

$$\theta = Eh(X_1, \ldots, X_d), \;\; h_1(x) = E(h(X_1, \ldots, X_d) - \theta | X_d = x).$$

If $\sigma^2 := \mathrm{Var}(h_1(X)) > 0$, the standardized non-degenerate U-statistic is given by

$$Z_n = \frac{\sqrt{n}}{d\,\sigma}(U_n - \theta).$$

Because σ^2 is always unknown, we are interested in the following Studentized (self-normalized) U-statistic:

$$\widehat{U}_n = \frac{\sqrt{n}}{d\,\widehat{\sigma}}(U_n - \theta),$$

where $\widehat{\sigma}^2$ denotes the leave-one-out Jackknife estimator of σ^2; that is,

$$\widehat{\sigma}^2 = \frac{n-1}{(n-d)^2}\sum_{i=1}^{n}(q_i - U_n)^2,$$

where

$$q_i = \frac{1}{\binom{n-1}{d-1}} \sum_{\substack{1\le i_1<\cdots<i_{d-1}\le n,\\ i_j\ne i,\, j=1,\ldots,d-1}} h(X_i, X_{i_1}, \ldots, X_{i_{d-1}}).$$

Because σ^2 is always unknown, we are interested in the following Studentized (self-normalized) U-statistic:

$$\widehat{U}_n = \frac{\sqrt{n}}{d\,\widehat{\sigma}}(U_n - \theta),$$

where $\widehat{\sigma}^2$ denotes the leave-one-out Jackknife estimator of σ^2; that is,

$$\widehat{\sigma}^2 = \frac{n-1}{(n-d)^2}\sum_{i=1}^{n}(q_i - U_n)^2,$$

where

$$q_i = \frac{1}{\binom{n-1}{d-1}} \sum_{\substack{1\le i_1<\cdots<i_{d-1}\le n,\\ i_j\ne i,\, j=1,\ldots,d-1}} h(X_i, X_{i_1}, \ldots, X_{i_{d-1}}).$$

statistic	kernel function
t-statistic	$h(x_1,x_2) = \frac{1}{2}(x_1 + x_2)$
Sample variance	$h(x_1,x_2) = \frac{1}{2}(x_1 - x_2)^2$
Gini's mean difference	$h(x_1,x_2) = \lvert x_1 - x_2\rvert$
Wilcoxon's statistic	$h(x_1,x_2) = I\{x_1 + x_2 \le 0\}$
Kendall's τ	$h(\mathbf{x}_1,\mathbf{x}_2) = 2I\{(x_2^2 - x_1^2)(x_2^1 - x_1^1) > 0\}$

The construction of U-statistics arises most naturally from a paired comparison experiment based on independent random variables, (X_i, Y_i), $i \geq 1$. Denote by (F, G) a pair of probability measures and assume $X \sim F, Y \sim G$.

The construction of U-statistics arises most naturally from a paired comparison experiment based on independent random variables, (X_i, Y_i), $i \geq 1$. Denote by (F, G) a pair of probability measures and assume $X \sim F, Y \sim G$.

Consider independent samples, $X_1, \ldots, X_m$ from F and $Y_1, \ldots, Y_n$ from G. Let $h(x_1, \ldots, x_d, y_1, \ldots, y_s)$ be a kernel which is symmetric under independent permutations of $x_1, \ldots, x_d$ and $y_1, \ldots, y_s$. The corresponding U-statistic is

$$U_{m,n} = \frac{1}{\binom{m}{d}\binom{n}{s}} \sum_{1 \leq i_1 < \cdots < i_d \leq m} \sum_{1 \leq j_1 < \cdots < j_s \leq n} h(X_{i_1}, \ldots, X_{i_d}, Y_{j_1}, \ldots, Y_{j_s}),$$

which is an unbiased estimate of

$$\theta = Eh(X_1, \ldots, X_d, Y_1, \ldots, Y_s).$$

- Examples
  - A two-sample comparison of means: Let $h(x, y) = x - y$, a kernel of degree $(d, s) = (1, 1)$. Then $\theta = EX - EY$ and the corresponding U-statistic is

$$U_{m,n} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} (X_i - Y_j) = \bar{X}_m - \bar{Y}_n.$$

  - The Wilcoxon (1945), Mann-Whitney (1947), two-sample test: Let the kernel be $h(x, y) = I\{x \leq y\}$ with $\theta = P(X \leq Y)$.
  - The Lehmann statistic (1951), The Kochar statistic (1979), etc.

In the non-degenerate case where $\sigma_1^2 = \mathrm{Var}(h_1(X)) > 0$ and $\sigma_2^2 = \mathrm{Var}(h_2(Y)) > 0$ with

$$h_1(x) = E(h(X_1, \ldots, X_d, Y_1, \ldots, Y_s) - \theta | X_1 = x),$$
$$h_2(y) = E(h(X_1, \ldots, X_d, Y_1, \ldots, Y_s) - \theta | Y_1 = y),$$

the Studentized two-sample U-statistic is given by

$$\widehat{U}_{m,n} = \widehat{\sigma}_{m,n}^{-1}(U_{m,n} - \theta) \quad \text{with} \quad \widehat{\sigma}_{m,n}^2 = d^2\, \widehat{\sigma}_1^2/m + s^2\, \widehat{\sigma}_2^2/n,$$

where

$$\widehat{\sigma}_1^2 = \tfrac{1}{m-1} \sum_{i=1}^{m} \left(q_i - \tfrac{1}{m} \sum_{i=1}^{m} q_i\right)^2, \;\; \widehat{\sigma}_2^2 = \tfrac{1}{n-1} \sum_{j=1}^{n} \left(p_j - \tfrac{1}{n} \sum_{j=1}^{n} p_j\right)^2,$$

$$q_i = \tfrac{1}{\binom{m-1}{d-1}\binom{n}{s}} \sum \sum h(X_i, X_{i_1}, \ldots, X_{i_{d-1}}, Y_{j_1}, \ldots, Y_{j_s}),$$
$$p_j = \tfrac{1}{\binom{m}{d}\binom{n-1}{s-1}} \sum \sum h(X_1, \ldots, X_{i_d}, Y_j, Y_{j_1}, \ldots, Y_{j_{s-1}}).$$

3. Cramér-type moderate deviations for Studentized U-statistics

- One-sample case:

Theorem (Shao and Z. 2012)

Assume that $v_p := (E|h_1(X)|^p)^{1/p} < \infty$ *for some* $2 < p \leq 3$*, and that*

$$(h(x_1, \ldots, x_d) - \theta)^2 \leq c_0\Big(\kappa\sigma^2 + \sum_{j=1}^{d} h_1^2(x_j)\Big) \qquad (1)$$

for some constants $c_0 \geq 1$ *and* $\kappa \geq 0$*. Then there exists a positive constant* C *depending only on* d *such that*

$$P(\widehat{U}_n \geq x)/(1 - \Phi(x)) = 1 + O(1)\Big(\frac{v_p^p(1+x)^p}{\sigma^p\, n^{p/2-1}} + \sqrt{a_d}\,\frac{(1+x)^3}{\sqrt{n}}\Big)$$

holds uniformly for $0 \leq x \leq C^{-1}\,\min((v_p/\sigma)n^{1/2-1/p}, (n/a_d)^{1/6})$*, where* $a_d = c_0(\kappa + d)$*,* $|O(1)| \leq C$*.*

Condition (1) on the kernel function is satisfied for the class of bounded kernels, such as the one-sample Wilcoxon test statistic, Kendall's tau, Spearman's rho, etc.

More importantly, it extends the boundedness assumption to more general settings.

Condition (1) on the kernel function is satisfied for the class of bounded kernels, such as the one-sample Wilcoxon test statistic, Kendall's tau, Spearman's rho, etc.

More importantly, it extends the boundedness assumption to more general settings.

Analogously in the two-sample case, we require that

$$
\begin{aligned}
&(h(x_1, \ldots, x_d, y_1, \ldots, y_s) - \theta)^2 \\
&\qquad \le c_0\left(\kappa\sigma^2 + \sum_{i=1}^{d} h_1^2(x_i) + \sum_{j=1}^{s} h_2^2(y_j)\right) \qquad (2)
\end{aligned}
$$

where $\sigma^2 = \sigma_1^2 + \sigma_2^2 = \mathrm{Var}(h_1(X)) + \mathrm{Var}(h_2(Y))$.

- Two-sample case:

Theorem (Shao and Z. 2014)

Assume that $v_{1,p} = (E|h_1(X)|^p)^{1/p}$ *and* $v_{2,p} = (E|h_2(Y)|^p)^{1/p}$ *are finite for some* $2 < p \le 3$, *and that condition* (2) *holds. Then there exists a positive constant* C *depending only on* (d, s) *such that*

$$P(\widehat{U}_{m,n} \ge x)/(1 - \Phi(x)) = 1 + O(1)R_{n,m}(x)$$

holds uniformly for $0 \le x \le C^{-1}A_{m,n}$, *where*

$$A_{m,n} = \min\left((\sigma_1/v_{1,p})m^{1/2-1/p}, (\sigma_2/v_{2,p})n^{1/2-1/p}, a_{d,s}^{-1/6}(mn/(m+n))^{1/6}\right),$$

$$R_{m,n}(x) = (v_{1,p}/\sigma_1)^p \frac{(1+x)^p}{m^{p/2-1}} + (v_{2,p}/\sigma_2)^p \frac{(1+x)^p}{n^{p/2-1}} + \sqrt{a_{d,s}}\,(1+x)^3\sqrt{\frac{m+n}{mn}}$$

with $a_{d,s} = c_0(\kappa + d + s)$, *and* $|O(1)| \le C$.

Corollary (Moderate deviation for Studentized two-sample U-statistics with first order accuracy)

Assume w.l.o.g. that $n = \min(m, n)$, $Eh_1^2(X), Eh_2^2(Y) > 0$ *and both* $E|h_1(X)|^{2+\delta}$ *and* $E|h_2(Y)|^{2+\delta}$ *are finite for some* $0 < \delta \leq 1$. *Then*

$$P(\widehat{U}_{m,n} \geq x)/(1 - \Phi(x)) = 1 + O(1)(1 + x)^{2+\delta} n^{-\delta/2}$$

holds uniformly over $x \in [0, o(n^{\delta/(4+2\delta)}))$.

Corollary (Moderate deviation for Studentized two-sample U-statistics with first order accuracy)

Assume w.l.o.g. that $n = \min(m,n)$, $Eh_1^2(X), Eh_2^2(Y) > 0$ *and both* $E|h_1(X)|^{2+\delta}$ *and* $E|h_2(Y)|^{2+\delta}$ *are finite for some* $0 < \delta \le 1$. *Then*

$$P(\widehat{U}_{m,n} \ge x)/(1 - \Phi(x)) = 1 + O(1)(1+x)^{2+\delta} n^{-\delta/2}$$

holds uniformly over $x \in [0, o(n^{\delta/(4+2\delta)}))$.

This result addresses the dependence between the range of uniform convergence of the relative error in the central limit theorem and the required (heavy-tailed) moment conditions.

- Question: Under higher order moment conditions, say $\delta > 1$, whether it is possible to obtain a better approximation (second order accuracy) for the tail probability $P(\widehat{U}_{m,n} \ge x)$ for $c_1 n^{1/6} \le x \le c_2 n^{1/2}$.

▶ Self-normalized moderate deviations for independent r.v.'s:

Let $X_1, X_2, \ldots$ be independent random variables with $EX_i = 0$. Write

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

The self-normalized moderate deviations describe the rate of convergence of the relative error of $P(S_n/V_n \geq x)$ to $1 - \Phi(x)$.

The corresponding results with first and second order accuracies were established in Jing, Shao and Wang (2003) (first order accuracy under finite third moments) and Wang (2005, 2011) (second order accuracy under finite fourth moments).

Write

$$B_n^2 = \sum_{i=1}^{n} EX_i^2, \;\; \mathcal{L}_{kn} = B_n^{-k} \sum_{i=1}^{n} E|X_i|^k, \; k \geq 3.$$

Theorem (Jing, Shao and Wang, 2003)

If $X_1, X_2, \ldots$ are independent r.v.'s with $EX_i = 0$ and $0 < E|X_i|^3 < \infty$, then

$$P(S_n/V_n \geq x)/(1 - \Phi(x)) = 1 + O(1)(1+x)^3 \mathcal{L}_{3n}$$

for $0 \leq x \leq \mathcal{L}_{3n}^{-1/3}$, where $|O(1)|$ is bounded by an absolute constant.

Theorem (Wang, 2011)

If $X_1, X_2, \ldots$ are independent r.v.'s with $EX_i = 0$ and $0 < EX_i^4 < \infty$, then

$$P(S_n/V_n \geq x)/(1 - \Phi(x))$$
$$= \exp\left(-\frac{x^3}{3}\sum_{i=1}^{n} EX_i^3\right)\Big(1 + O(1)\big((1+x)\mathcal{L}_{3n} + (1+x)^4\mathcal{L}_{4n}\big)\Big)$$

for $0 \leq x \leq C^{-1}\mathcal{L}_{4n}^{-1/4}$, where $|O(1)|$ is bounded by an absolute constant.

Theorem (Wang, 2011)

If $X_1, X_2, \ldots$ are independent r.v.'s with $EX_i = 0$ and $0 < EX_i^4 < \infty$, then

$$P(S_n/V_n \geq x)/(1-\Phi(x))$$
$$= \exp\left(-\frac{x^3}{3}\sum_{i=1}^{n} EX_i^3\right)\left(1 + O(1)\left((1+x)\mathcal{L}_{3n} + (1+x)^4\mathcal{L}_{4n}\right)\right)$$

for $0 \leq x \leq C^{-1}\mathcal{L}_{4n}^{-1/4}$, where $|O(1)|$ is bounded by an absolute constant.

- Question: Whether a similar expansion holds for more general Studentized nonlinear statistics, such as Hoeffding's class of U-statistics after suitably Studentized.

4. Two-sample t-statistic: A first attempt

As a prototypical example of two-sample U-statistics, the two-sample t-statistic is of significant interest due to its wide applicability.

- Advantages: High degree of robustness against heavy-tailed data.
  - The robustness of the t-statistic is useful in high dimensional data analysis under the sparsity assumption on the signal of interest (Delaigle, Hall and Jin, 2011).
  - When dealing with two experimental groups, typically assumed to be independent, in scientifically controlled experiments, the two-sample t-statistic is one of the most commonly used statistics for hypothesis testing and constructing confidence intervals for the difference between the means of the two groups.

Let $X_1, \ldots, X_m$ be a random sample from a population with mean μ_1 and variance σ_1^2, and let $Y_1, \ldots, Y_n$ be a random sample from another population with mean μ_2 and variance σ_2^2. Assume that the two random samples are drawn independently.

The two-sample t-statistic is defined as

$$\widehat{T}_{m,n} = \frac{\bar{X}_m - \bar{Y}_n}{\sqrt{\widehat{\sigma}_1^2/m + \widehat{\sigma}_2^2/n}},$$

where

$$\widehat{\sigma}_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X}_m)^2, \quad \widehat{\sigma}_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y}_n)^2.$$

Corollary

Assume that $\mu_1 = \mu_2$, *and* $E|X_1|^{2+\delta} < \infty$, $E|Y_1|^{2+\delta} < \infty$ *for some* $0 < \delta \leq 1$. *Then*

$$P(\widehat{T}_{m,n} \geq x)/(1 - \Phi(x))$$
$$= 1 + O(1)(1+x)^{2+\delta}\left((v_{1,2+\delta}/\sigma_1)^{2+\delta} m^{-\delta/2} + (v_{2,2+\delta}/\sigma_2)^{2+\delta} n^{-\delta/2}\right)$$

holds uniformly for

$$0 \leq x \leq C^{-1} \min\left((\sigma_1/v_{1,2+\delta}) m^{\delta/(4+2\delta)}, (\sigma_2/v_{2,2+\delta}) n^{\delta/(4+2\delta)}\right),$$

where $v_{1,s} = \left(E|X_1 - \mu_1|^s\right)^{1/s}, v_{2,s} = \left(E|Y_1 - \mu_2|^s\right)^{1/s}$ *for all* $s > 2$ *and* $|O(1)| \leq C$.

Corollary

Assume that $\mu_1 = \mu_2$, $E|X_1|^{2+\delta}, E|Y_1|^{2+\delta} < \infty$ *for some* $1 < \delta \le 2$ *and write* $\gamma_1 = E(X_1 - \mu_1)^3$, $\gamma_2 = E(Y_1 - \mu_2)^3$. *Then*

$$\begin{aligned} &P(\widehat{T}_{m,n} \ge x)/(1 - \Phi(x)) \\ &= \exp\left(- \frac{x^3\,(\gamma_1/m^2 + \gamma_2/n^2)}{3\,(\sigma_1^2/m + \sigma_2^2/n)^{3/2}} \right)(1 + O(1)\mathcal{R}_{m,n}(x)) \end{aligned}$$

holds uniformly for $0 \le x \le C^{-1} A_{m,n}$, *where*

$$\begin{aligned} A_{m,n} &= \min\left((\sigma_1/v_{1,2+\delta}) m^{\delta/(4+2\delta)}, (\sigma_2/v_{2,2+\delta}) n^{\delta/(4+2\delta)}\right), \\ \mathcal{R}_{m,n}(x) &= (v_{1,3}/\sigma_1)^3 (1+x) m^{-1/2} + (v_{1,2+\delta}/\sigma_1)^{2+\delta} (1+x)^{2+\delta} m^{-\delta/2} \\ &\quad + (v_{2,3}/\sigma_2)^3 (1+x) n^{-1/2} + (v_{2,2+\delta}/\sigma_2)^{2+\delta} (1+x)^{2+\delta} n^{-\delta/2}, \end{aligned}$$

$v_{1,s} = \left(E|X_1 - \mu_1|^s\right)^{1/s}, v_{2,s} = \left(E|Y_1 - \mu_2|^s\right)^{1/s}$ *for* $s > 2$ *and* $|O(1)| \le C$.

- Question: Whether similar moderate deviation results with second order accuracy hold for general Studentized U-statistics.

- Standardized non-degenerate (one-sample) U-statistics (Borovskikh and Weber, 2001): Recall that $X_1, X_2, \ldots$ are i.i.d. r.v.'s with dist. F. Assume that

$$E \exp(c|h(X_1, \ldots, X_d)|) < \infty \quad \text{for some } c > 0,$$

then

$$P(\sqrt{n}(U_n - \theta)/(d\,\sigma) \geq x)/(1 - \Phi(x))$$
$$= \exp\left(\frac{x^3}{\sqrt{n}}\lambda_F\left(\frac{x}{\sqrt{n}}\right)\right)\left(1 + O\left(\frac{1+x}{\sqrt{n}}\right)\right)$$

holds for $x = o(\sqrt{n})$, with defined Cramér series

$$\lambda_F(u) = \lambda_{0,F} + \lambda_{1,F}u + \lambda_{2,F}u^2 + \cdots$$

which is convergent for small u.

5. Applications

- Multiple-hypothesis testing in high dimensions: Analysis of gene expression microarray data, where the purpose is to examine whether each gene in isolation behaves differently in a control group v.s. an experimental group.

  The statistical model is

$$\begin{cases} X_{g,i} = \mu_{g,1} + \varepsilon_{g,i}, & i = 1, \ldots, m, \\ Y_{g,j} = \mu_{g,2} + \omega_{g,j}, & j = 1, \ldots, n, \end{cases} \quad \text{for } g = 1, \ldots, G,$$

  g: the gth gene; i and j: the ith and jth array; $\mu_{g,1}$ and $\mu_{g,2}$: the mean effects for the gth gene from the first and the second group.

  $\forall g$, $\varepsilon_{g,1}, \ldots, \varepsilon_{g,m}$ (resp. $\omega_{g,1}, \ldots, \omega_{g,n}$) are independent r.v.'s with mean zero and variance $\sigma^2_{g,1} > 0$ (resp. $\sigma^2_{g,2} > 0$). For the gth marginal test, when the population variances $\sigma^2_{g,1}$ and $\sigma^2_{g,2}$ are unequal, the two-sample t-statistic is most commonly used to carry out hypothesis testing for the null $H_{g,0} : \mu_{g,1} = \mu_{g,2}$ against the alternative $H_{g,1} : \mu_{g,1} \neq \mu_{g,2}$.

  Nonparametric alternative: Studentized Mann-Whitney U-test.

Write $\mathcal{H}_1 = \{g = 1, \ldots, G : \mu_{g,1} \neq \mu_{g,2}\}$ and assume that

$$\lim_{G\to\infty} G^{-1}|\mathcal{H}_1| = \pi_1 \in [0, 1).$$

Let $\alpha \in (0, 1)$ be fixed.

- Non-sparse setting ($0 < \pi_1 < 1$): Cao and Kosorok (2011) proposed a method to compute critical values directly for rejection regions to control FDR, for heavy tailed data (finiteness of 4th moments) and when $\log(G) = o(n^{1/3})$.

Write $\mathcal{H}_1 = \{g = 1, \ldots, G : \mu_{g,1} \neq \mu_{g,2}\}$ and assume that

$$\lim_{G\to\infty} G^{-1}|\mathcal{H}_1| = \pi_1 \in [0, 1).$$

Let $\alpha \in (0, 1)$ be fixed.

- Non-sparse setting ($0 < \pi_1 < 1$): Cao and Kosorok (2011) proposed a method to compute critical values directly for rejection regions to control FDR, for heavy tailed data (finiteness of 4th moments) and when $\log(G) = o(n^{1/3})$.

- Sparse setting ($|\mathcal{H}_1| \leq G^{\eta}, 0 < \eta < 1$): Adapting the regularized bootstrap correction method proposed by Liu and Shao (2013) to the current problem using two-sample t-statistics, the bootstrap calibration is accurate for heavy-tailed data (finiteness of 6th moments), provided that $\log(G) = o(n^{1/2})$ as $n = \min(n, m) \to \infty$.

▶ Real data example: Above procedures can be applied to the analysis of a leukemia cancer set (Golub et al., 1999) in order to identify differentially expressed genes between AML (acute lymphoblastic leukemia) and ALL (acute myeloid leukemia). The raw data consist of $G = 7129$ genes, 47 samples in class ALL and 25 in class AML.

- Testing many moment inequalities (Chernozhukov, Chetverikov and Kato, 2013):

  Let $X_1, \dots, X_m$ (resp. $Y_1, \dots, Y_n$) be i.i.d. random vectors in $\mathbb{R}^p$, where $X_i = (X_{i1}, \dots, X_{ip})^T$ (resp. $Y_j = (Y_{j1}, \dots, Y_{jp})^T$). Consider the null hypothesis

$$H_0 : EX_{1\ell} \leq EY_{1\ell}, \forall \ell \in [p] \ \ versus \ \ H_1 : \exists \ell \in [p], EX_{1\ell} > EY_{1\ell}$$

  or

$$H_0 : P(X_{1\ell} \leq Y_{1\ell}) \leq \tfrac{1}{2}, \forall \ell \in [p] \ versus \ H_1 : \exists \ell \in [p], P(X_{1\ell} \leq Y_{1\ell}) > \tfrac{1}{2},$$

  where $[p] = \{1, \dots, p\}$.

References

1. CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Testing many moment inequalities. *arXiv:1312.7614.*
2. CHUNG, E. and ROMANO, J. (2011). Asymptotically valid and exact permutation tests based on two-sample U-statistics. *Technical report.*
3. JING, B.-Y., SHAO, Q.-M. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31** 2167–2215.
4. SHAO, Q.-M. and ZHOU, W.-X. (2012). Cramér type moderate deviation theorems for self-normalized processes. *arXiv:1405.1218.*
5. SHAO, Q.-M. and ZHOU, W.-X. (2014). Cramér type moderate deviations for Studentized two-sample U-statistics with applications. *Technical report.*