Asymptotic Theory of the Coherence of Large Sample Correlation Matrices

Wen-Xin Zhou

Department of Mathematics and Statistics
University of Melbourne

This talk is based on a joint work with Prof. Qi-Man Shao

December, 2013 Hong Kong

Outline

1. Introduction and motivation
2. Asymptotic theorems of sample correlation function
3. Asymptotic theorems of coherence of high-dimensional sample correlation matrix
4. Ultra-high dimension
5. Main idea of the proof

1. Introduction and motivation

- Motivation:
  - Consider two random variables X and Y, how to test whether X and Y are independent?
  - Consider a p-variate random vector $\boldsymbol{x} = (x_1, \cdots, x_p)$, how to test whether $x_1, \cdots, x_p$ are independent?

1. Introduction and motivation

▶ Motivation:

- Consider two random variables X and Y, how to test whether X and Y are independent?
- Consider a p-variate random vector $\boldsymbol{x} = (x_1, \cdots, x_p)$, how to test whether $x_1, \cdots, x_p$ are independent?

Observe that

- X and Y are independent $\Leftrightarrow$

$$\int_{-\infty}^{\infty} |\mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)|^2 dxdy = 0$$

- If (X, Y) follows a joint normal distribution, then X and Y are independent $\Leftrightarrow$

$$\mathrm{corr}(X, Y) = \frac{\mathbb{E}(X - \mu_X)(Y - \mu_Y)}{\sqrt{\mathbb{E}(X - \mu_X)^2 \cdot \mathbb{E}(Y - \mu_Y)^2}} = 0$$

- Test statistics:

Let $\{(X_i, Y_i)\}_{i=1}^n$ be a random sample from (X, Y).

  - Based on sample copula function:

$$\int_{-\infty}^{\infty} \left| \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x, Y_i \leq y) - \frac{1}{n^2} \sum_{i=1}^{n} I(X_i \leq x) \sum_{j=1}^{n} I(Y_j \leq y) \right|^2 dxdy$$

  - Sample correlation function:

$$r_n = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2\}^{1/2}},$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,\ \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j$.

Here we will focus on the sample correlation function.

2. Asymptotic theorems of sample correlation function

Let $(X,Y),(X_1,Y_1),\ldots,(X_n,Y_n)$ be i.i.d. Recall

$$r_n=\frac{\sum_{i=1}^n(X_i-\bar X)(Y_i-\bar Y)}{\{\sum_{i=1}^n(X_i-\bar X)^2\sum_{j=1}^n(Y_j-\bar Y)^2\}^{1/2}}.$$

Assume that $\mathbb{E}X^2<\infty$, $\mathbb{E}Y^2<\infty$ and X, Y are independent.

- Central limit theorem:

$$\sqrt{n}\,r_n\xrightarrow{d}N(0,1)$$

- Cramér type moderate deviation theorem: If $\mathbb{E}e^{t_0(X^2+Y^2)}<\infty$ for some $t_0>0$, then

$$\frac{\mathbb{P}(\sqrt{n}\,r_n>x)}{1-\Phi(x)}\to 1$$

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

3. Asymptotic theorems of coherence of high-dimensional sample correlation matrix

Let $\boldsymbol{x} = (x_1, \cdots, x_p)$ be the population random vector.

- H_0: the components of $\boldsymbol{x}$ are independent

Let $\boldsymbol{x}_k = (x_{k,1}, \cdots, x_{k,p})$, $1 \le k \le n$ be a random sample of size n from the population, and set

$$\bar{x}_i = \frac{1}{n}\sum_{k=1}^{n} x_{k,i}, \quad s_i^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_{k,i} - \bar{x}_i)^2,$$

$$r_{i,j} = \frac{1}{(n-1)s_i s_j}\sum_{k=1}^{n}(x_{k,i} - \bar{x}_i)(x_{k,j} - \bar{x}_j).$$

- High dimensionality: assume that dimension $p = p(n)$ grows as the sample size n increases in a way that $n/p \to \gamma \in (0, \infty)$.
- The coherence of the sample correlation matrix:

$$L_n = \max_{1 \le i < j \le p} |r_{i,j}|.$$

The terminology coherence is borrowed from the compressed sensing literature. The development of the compressed sensing theory also provides crucial insights into high-dimensional regression in statistics. See, for example, Candes and Tao (2007), Bickel, Ritov and Tsybakov (2009) and Candes and Plan (2009), among others.

One of the main goals of compressed sensing is to construct measurement matrices $X_{n\times p}$, with the number of measurements n as small as possible relative to p, s.t. for any k-sparse signal $\beta \in \mathbb{R}^p$, one can recover β exactly from linear measurements $y = X\beta$ using a computationally efficient recovery algorithm.

Two commonly used conditions on $X_{n\times p}$ are the so called restricted isometry property (RIP) and mutual incoherence property (MIP).

- RIP: subsets of certain cardinality of the columns of X to be close to an orthonormal system.
- MIP: the pairwise correlations among the column vectors of X to be small.

It is well known that construction of large deterministic measurement matrices that satisfy either the RIP or MIP is difficult. Instead, random matrices are commonly used. Matrices generated by certain random processes have been shown to satisfy the RIP conditions with high probability.

Let $X = (x_{i,j})$ be an $n \times p$ random matrix where the entries $x_{i,j}$ are i.i.d. real random variables with mean μ and variance σ^2. Write $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)_{n\times p}$, define

$$\widetilde{L}_n = \max_{1\le i<j\le p} |\tilde{r}_{i,j}| \quad \text{with} \quad \tilde{r}_{i,j} = \frac{(\boldsymbol{x}_i - \mu)^T(\boldsymbol{x}_j - \mu)}{\|\boldsymbol{x}_i - \mu\| \cdot \|\boldsymbol{x}_j - \mu\|}.$$

The MIP condition asks that $(2k-1)\widetilde{L}_n < 1$.

In the compressed sensing literature, the quantity $\widetilde{L}_n$ is called the coherence of the matrix X. See, for example, Donoho, Elad and Temlyakov (2006).

▶ Limiting distribution of the coherence under H_0

- T. Jiang (Ann. Appl. Probab, 2004): Assume that $n/p \to \gamma \in (0,\infty)$ and that

$$\mathbb{E}|x_{i,j}|^{\theta} < \infty \quad \text{for some} \quad \theta > 30, \tag{1}$$

Then

$$\mathbb{P}(nL_n^2 - 4\log p + \log_2 p \le y) \to \exp\left(- \frac{1}{\sqrt{8\pi}} e^{-y/2}\right) \tag{2}$$

for $y \in \mathbb{R}$, where $\log x = \ln\max(x,e)$ and $\log_2 x = \log(\log x)$.

▶ Limiting distribution of the coherence under H_0

- T. Jiang (Ann. Appl. Probab, 2004): Assume that $n/p \to \gamma \in (0, \infty)$ and that

$$\mathbb{E}|x_{i,j}|^{\theta} < \infty \quad \text{for some} \quad \theta > 30, \tag{1}$$

 Then

$$\mathbb{P}(nL_n^2 - 4\log p + \log_2 p \leq y) \to \exp\left(-\frac{1}{\sqrt{8\pi}}e^{-y/2}\right) \tag{2}$$

 for $y \in \mathbb{R}$, where $\log x = \ln \max(x, e)$ and $\log_2 x = \log(\log x)$.

- W. Zhou (Trans. Amer. Math. Soc. 2007): The moment condition (1) can be replaced by

$$x^6\mathbb{P}(|x_{1,1}x_{1,2}| \geq x) \to 0 \text{ as } x \to \infty. \tag{3}$$

 If $\mathbb{E}x_{1,1}^6 < \infty$, then (3) is satisfied; conversely if (3) holds, then $\mathbb{E}|x_{1,1}|^{6-\varepsilon} < \infty$ for any $0 < \varepsilon < 1$.

- Liu, Lin and Shao (Ann. Appl. Probab. 2008): Another sufficient condition for (2) which slightly improves (3):

$$\frac{x^6}{\log^3 x}\mathbb{P}(|x_{1,1}x_{1,2}| \geq x) \to 0, \quad \text{as } x \to \infty.$$

- Li, Liu and Rosalsky (PTRF, 2010), Li, Qi and Rosalsky (JMA, 2012): Assume $p \asymp n$, i.e. n/p is bounded away from 0 and ∞. When $\mathbb{E}x_{1,1}^2 < \infty$, then the following three statements are equivalent:

$$\lim_{n\to\infty} n^2 \int_{(n\log n)^{1/4}}^{\infty} \left[F^{n-1}(x) - F^{n-1}\left(\frac{\sqrt{n\log n}}{x}\right)\right] dF(x) = 0,$$

$$\left(\frac{n}{\log n}\right)^{1/2} L_n \to 2 \quad \text{in probability},$$

$$\lim_{n\to\infty} \mathbb{P}(nL_n^2 - \alpha_p \leq y) = e^{-\frac{1}{\sqrt{8\pi}}e^{-y/2}}, \quad y \in \mathbb{R}$$

where $F(x) = \mathbb{P}(|x_{1,1}| \leq x)$, $\alpha_p = 4\log p - \log_2 p$.

- The limit distribution appearing in (2) is called the extreme distribution of type I. The convergence rate to this type of extreme distribution is typically slow (P. Hall, 1979).

- The limit distribution appearing in (2) is called the extreme distribution of type I. The convergence rate to this type of extreme distribution is typically slow (P. Hall, 1979).
- Liu, Lin and Shao (2008):

  If $p = n$ and $x_{1,1} \sim N(0,1)$, then

$$\mathbb{P}(nL_n^2 - 4\log p + \log_2 p \leq y) - \exp\Big(-\frac{1}{\sqrt{8\pi}}e^{-y/2}\Big) \sim \frac{\log_2 n}{8\log n}\frac{1}{\sqrt{8\pi}}\exp\Big(-\frac{y}{2} - \frac{1}{\sqrt{8\pi}}e^{-y/2}\Big)$$

  The rate of convergence in (2) is of order $O(\log_2 n/\log n)$.

- Improving the rate of convergence:
  - Liu, Lin and Shao (2008): Assume that $c_1 n^\alpha \le p \le c_2 n^\alpha$. If $\mathbb{E}|x_{1,1}|^{3+4\alpha} < \infty$ and $\alpha > 3/4$, then

$$\sup_{y\in\mathbb{R}} \Big| \mathbb{P}(nL_n^2 - \alpha_p \le y) - \exp\big(-p(p-1)\mathbb{P}\big(\chi_1^2 \ge \alpha_p + y\big)/2\big) \Big| \le C n^{-\frac{1}{2}} (\log n)^{\frac{5}{2}},$$

where $\alpha_p = 4\log p + \log_2 p$.

- Improving the rate of convergence:
  - Liu, Lin and Shao (2008): Assume that $c_1 n^\alpha \le p \le c_2 n^\alpha$. If $\mathbb{E}|x_{1,1}|^{3+4\alpha} < \infty$ and $\alpha > 3/4$, then

$$\sup_{y\in\mathbb{R}} \Big|\mathbb{P}(nL_n^2 - \alpha_p \le y) - \exp\big(-p(p-1)\mathbb{P}(\chi_1^2 \ge \alpha_p + y)/2\big)\Big| \le C n^{-\frac{1}{2}}(\log n)^{\frac{5}{2}},$$

where $\alpha_p = 4\log p + \log_2 p$.

This result shows that the rate of convergence for L_n^2 can almost achieve order $n^{-1/2}$ but with the intermediate limit

$$\exp\Big(-\frac{p(p-1)}{2}\mathbb{P}\big(\chi_1^2 \ge 4\log p - \log_2 p + y\big)\Big)$$

instead of the final limit $\exp(-e^{-y/2}/\sqrt{8\pi})$.

- Remark: When the limiting distribution is extreme distribution of type I, it may be better to use the intermediate limit rather than the final limiting distribution.

4. Ultra-high dimension

- Ultra-high or nonpolynomial dimensionality:

$$\log p = O(n^{\beta}) \quad \text{for some} \quad \beta > 0.$$

Let $x_{k,i},\ 1 \le k \le n, 1 \le i \le p$ be i.i.d. random variables. Recall

$$L_n = \max_{1 \le i < j \le p} |r_{i,j}|,$$

where

$$\bar{x}_i = \frac{1}{n}\sum_{k=1}^{n} x_{k,i}, \quad s_i^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_{k,i} - \bar{x}_i)^2,$$

$$r_{i,j} = \frac{1}{(n-1)s_i s_j}\sum_{k=1}^{n}(x_{k,i} - \bar{x}_i)(x_{k,j} - \bar{x}_j).$$

- Weak law of large numbers

Theorem (Cai and Jiang, Ann. Statist. 2011)

1. *If* $|x_{1,1}| \leq C$ *and* $\log p = o(n)$, *then*

$$\sqrt{n/(\log p)}\, L_n \to 2 \ \textit{in probability} \tag{4}$$

2. *Assume that* $\mathbb{E}e^{t_0|x_{1,1}|^{\alpha}} < \infty$ *for some* $t_0 > 0$ *and* $\alpha > 0$, *and* $\log p = o(n^{\beta})$, *where* $\beta = \alpha/(4+\alpha)$. *Then* (4) *holds.*

- Weak law of large numbers

Theorem (Cai and Jiang, Ann. Statist. 2011)

1. *If* $|x_{1,1}| \le C$ *and* $\log p = o(n)$, *then*

$$\sqrt{n/(\log p)}\, L_n \to 2 \ \textit{in probability} \tag{4}$$

2. *Assume that* $\mathbb{E}e^{t_0|x_{1,1}|^\alpha} < \infty$ *for some* $t_0 > 0$ *and* $\alpha > 0$, *and* $\log p = o(n^\beta)$, *where* $\beta = \alpha/(4+\alpha)$. *Then* (4) *holds.*

- Remark: If, in particular, $x_{1,1} \sim N(0,1)$ with $\alpha = 2$, then $\beta = 1/2$ and (4) holds if $\log p = o(n^{1/2})$.

- Question: Whether the dependence between the moment condition and the order of dimension p is necessary and sufficient?

Theorem (Shao-Z-2013b)

1. *If* $\mathbb{E}e^{t_0|x_{1,1}|^\alpha} < \infty$ *for some* $t_0 > 0$ *and* $0 < \alpha \leq 2$, *and* $\log p = o(n^\beta)$, *where* $\beta = \alpha/(4-\alpha)$. *Then* (4) *holds, i.e.*

$$\sqrt{n/\log p}\, L_n \to 2 \ \textit{in probability}.$$

2. *Let* $0 < \beta \leq 1$. *If* (4) *holds for any* p *satisfying* $\log p = o(n^\beta)$, *then* $\mathbb{E}e^{t_0|x_{1,1}|^\alpha} < \infty$ *for some* $t_0 > 0$, *where* $\alpha/(4-\alpha) = \beta$.

Theorem (Shao-Z-2013b)

1. *If* $\mathbb{E}e^{t_0|x_{1,1}|^\alpha} < \infty$ *for some* $t_0 > 0$ *and* $0 < \alpha \leq 2$*, and* $\log p = o(n^\beta)$*, where* $\beta = \alpha/(4-\alpha)$*. Then* (4) *holds, i.e.*

$$\sqrt{n/\log p}\, L_n \to 2 \ \textit{ in probability}.$$

2. *Let* $0 < \beta \leq 1$*. If* (4) *holds for any* p *satisfying* $\log p = o(n^\beta)$*, then* $\mathbb{E}e^{t_0|x_{1,1}|^\alpha} < \infty$ *for some* $t_0 > 0$*, where* $\alpha/(4-\alpha) = \beta$*.*

- Open question:

Assume that $(\log p)/n \to \gamma \in (0, \infty)$. If $x_{1,1} \sim N(0,1)$, then by Cai and Jiang (JMA, 2012),

$$L_n \to \sqrt{1 - e^{-4\gamma}} \ \text{ in probability.}$$

What is the limit of L_n for general $x_{1,1}$? We conjecture that the limit depends on the distribution of $x_{1,1}$.

- Limiting distribution

Theorem (Cai and Jiang, 2011)

Assume that $\mathbb{E}e^{t_0|x_{1,1}|^\alpha} < \infty$ *for some* $t_0 > 0$ *and* $0 < \alpha \le 2$, *and* $\log p = o(n^\beta)$, *where* $\beta = \alpha/(4+\alpha)$. *Then*

$$\mathbb{P}(nL_n^2 - 4\log p + \log_2 p \le y) \longrightarrow \exp\Big(-\frac{1}{\sqrt{\pi}}e^{-y/2}\Big).$$

- Limiting distribution

Theorem (Cai and Jiang, 2011)

Assume that $\mathbb{E}e^{t_0|x_{1,1}|^\alpha} < \infty$ *for some* $t_0 > 0$ *and* $0 < \alpha \leq 2$, *and* $\log p = o(n^\beta)$, *where* $\beta = \alpha/(4+\alpha)$. *Then*

$$\mathbb{P}(nL_n^2 - 4\log p + \log_2 p \leq y) \longrightarrow \exp\left(-\frac{1}{\sqrt{\pi}}e^{-y/2}\right).$$

- Question: Is this moment condition optimal?

Theorem (Shao-Z-2013b)

Assume that $\mathbb{E}e^{t_0|x_{1,1}|^\alpha} < \infty$ *for some* $t_0 > 0$ *and* $0 < \alpha \le 4/3$, *and that* $\log p = o(n^\beta)$, *where* $\beta = \alpha/(4-\alpha)$.

1. *If* $0 < \alpha \le 1$, *then*

$$\mathbb{P}(nL_n^2 - 4\log p + \log_2 p \le y) \to \exp\Big(-\frac{1}{\sqrt{8\pi}}e^{-y/2}\Big).$$

2. *If* $1 < \alpha \le 4/3$, *then*

$$\mathbb{P}\{nL_n^2 - 4\log p + \log_2 p - (8\kappa^2/3)n^{-1/2}(\log p)^{3/2} \le y\} \longrightarrow \exp\Big(-\frac{1}{\sqrt{8\pi}}e^{-y/2}\Big),$$

where $\kappa = \mathbb{E}(x_{1,1}-\mu)^3/\sigma^3$ *is the skewness of* $x_{1,1}$.

▶ Remark: Results in above theorems are still valid if L_n is replace by

$$\widetilde{L}_n = \max_{1 \le i < j \le p} |\tilde{r}_{i,j}|,$$

where

$$\tilde{r}_{i,j} = \frac{\sum_{k=1}^{n}(x_{k,i} - \mu)(x_{k,j} - \mu)}{\sqrt{\sum_{k=1}^{n}(x_{k,i} - \mu)^2 \sum_{k=1}^{n}(x_{k,j} - \mu)^2}}.$$

- Extension to m-dependent case.

A variant of coherence L_n can be used to construct a test for testing bandedness of the covariance matrix.

Let $\boldsymbol{x} = (x_1, \cdots, x_p)$ be the population random vector.

- H_0: x_i and x_j are independent for all $|i - j| \geq m$.

  Here $m = m(n)$ is allowed to grow as sample size n increases.

▶ Extension to m-dependent case.

A variant of coherence L_n can be used to construct a test for testing bandedness of the covariance matrix.

Let $\boldsymbol{x} = (x_1, \cdots, x_p)$ be the population random vector.

- H_0: x_i and x_j are independent for all $|i-j| \geq m$.

  Here $m = m(n)$ is allowed to grow as sample size n increases.

- m-coherence of the data matrix $X = (x_{k,i})$ as

$$L_{n,m} = \max_{|i-j| \geq m} |r_{i,j}|.$$

- Extension to m-dependent case.

A variant of coherence L_n can be used to construct a test for testing bandedness of the covariance matrix.

Let $\boldsymbol{x} = (x_1, \cdots, x_p)$ be the population random vector.

- H_0: x_i and x_j are independent for all $|i-j| \geq m$.

  Here $m = m(n)$ is allowed to grow as sample size n increases.

- m-coherence of the data matrix $X = (x_{k,i})$ as

$$L_{n,m} = \max_{|i-j| \geq m} |r_{i,j}|.$$

Let $(\rho_{i,j})_{p \times p}$ be the correlation matrix of $\boldsymbol{x}$. For any $\delta \in (0,1)$, set

$$\Gamma_{p,\delta} = \left\{1 \leq i \leq p : |\rho_{i,j}| > 1 - \delta \text{ for some } 1 \leq j \leq p \text{ with } j \neq i\right\}.$$

The following theorem establishes the limiting distribution of $L_{n,m}$ under H_0. Recall $\alpha_p = 4\log p - \log_2 p$.

Theorem (Shao-Z-2013b)

Let $\kappa = \mathbb{E}[(x_{1,1}-\mu)^3]/\sigma^3$ *and define*

$$W_{n,m} = \begin{cases} nL_{n,m}^2 - \alpha_p, & 0 < \alpha \le 1, \\ nL_{n,m}^2 - \alpha_p - (8\kappa^2/3)n^{-1/2}(\log p)^{3/2}, & 1 < \alpha \le 4/3. \end{cases}$$

Suppose $\mathbb{E}[e^{t_0|x_{1,1}|^\alpha}] < \infty$ *for some* $0 < \alpha \le 4/3$ *and* $t_0 > 0$. *Moreover, assume that, as* $n \to \infty$,

(i) $p = p_n \to \infty$, $\log p = o(n^{\beta_\alpha})$, *where* $\beta_\alpha = \alpha/(4-\alpha)$;

(ii) *there exists some* $\delta \in (0,1)$ *such that* $|\Gamma_{p,\delta}| = o(p)$ *and* $m = o(p^{\varepsilon_\delta})$, *where* $\varepsilon_\delta = (2\delta - \delta^2)/(4 - 2\delta + \delta^2)$.

Then, under H_0, $W_{n,m}$ *converges weakly to the extreme distribution appearing in* (2).

5. Main idea of the proof

Let $\{\eta_\alpha,\ \alpha \in I\}$ be random variables indexed by I. A key step in studying the limiting distribution of $\max_{\alpha\in I}\eta_\alpha$ is to show that

$$\mathbb{P}\big(\max_{\alpha\in I}\eta_\alpha \le t\big) \sim \exp\Big(-\sum_{\alpha\in I}\mathbb{P}(\eta_\alpha > t)\Big) \tag{5}$$

or approximately, treat η_α as independent.

Below are three key tools for the approach.

(i) Extreme distribution approximation via Stein-Chen's method (Arratia, Goldstein and Gordon, 1989):

Let $\{\eta_\alpha, \alpha \in I\}$ be random variables on an index set J. For each $\alpha \in I$, let B_α be a subset of I with $\alpha \in B_\alpha$. For a given $t \in R$, set $\lambda = \sum_{\alpha \in I} \mathbb{P}(\eta_\alpha > t)$. Then

$$|\mathbb{P}(\max_{\alpha \in I} \eta_\alpha \le t) - e^{-\lambda}| \le (1 \wedge \lambda^{-1})(b_1 + b_2 + b_3),$$

where

$$\begin{aligned} b_1 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(\eta_\alpha > t)\mathbb{P}(\eta_\beta > t), \\ b_2 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha, \beta \ne \alpha} \mathbb{P}(\eta_\alpha > t, \eta_\beta > t), \\ b_3 &= \sum_{\alpha \in I} \mathbb{E}\Big[\Big|\mathbb{P}(\eta_\alpha > t | \sigma\{\eta_\beta, \beta \notin B_\alpha\}) - \mathbb{P}(\eta_\alpha > t)\Big|\Big], \end{aligned}$$

and $\sigma\{\eta_\beta, \beta \notin B_\alpha\}$ is the σ-algebra generated by $\{\eta_\beta, \beta \notin B_\alpha\}$. In particular, if η_α is independent of $\{\eta_\beta, \beta \notin B_\alpha\}$ for each α, then $b_3 = 0$.

(ii) Classical and self-normalized moderate deviations
Linnik (1961), Petrov (1975), Shao (1997, 1999), etc.

(ii) Classical and self-normalized moderate deviations
Linnik (1961), Petrov (1975), Shao (1997, 1999), etc.

(iii) Randomized concentration inequality (Shao-Z-2013a).

Let $\xi_1, , ..., \xi_n$ be independent random variables, $W_n = \sum_{i=1}^n \xi_i$, $\Delta_{n,1} = \Delta_{n,1}(\xi_1, ..., \xi_n)$ and $\Delta_{n,2} = \Delta_{n,2}(\xi_1, ..., \xi_n)$ are two measurable functions of $\{\xi_1, ..., \xi_n\}$.

1. Assume that

$$\mathbb{E}\xi_i = 0, \quad i = 1, 2, \ldots, n \quad \text{and} \quad \sum_{i=1}^n \mathbb{E}\xi_i^2 = 1;$$

2. Define

$$\beta_2 = \sum_{i=1}^n \mathbb{E}[\xi_i^2 I(|\xi_i| > 1)], \quad \beta_3 = \sum_{i=1}^n \mathbb{E}[|\xi_i|^3 I(|\xi_i| \le 1)].$$

Theorem (Randomized concentration inequality, Shao-Z-2013a)

For each $1 \le i \le n$, *let* $\Delta_{n,1}^{(i)}$ *and* $\Delta_{n,2}^{(i)}$ *by random variables such that* ξ_i *and* $(\Delta_{n,1}^{(i)}, \Delta_{n,2}^{(i)}, W_n - \xi_i)$ *are independent. Then*

$$\begin{aligned} & \mathbb{P}(\Delta_{n,1} \le W_n \le \Delta_{n,2}) \\ \le\ & 21(\beta_2 + \beta_3) + 6\,\mathbb{E}|\Delta_{n,2} - \Delta_{n,1}| \\ & \quad + 2\sum_{i=1}^{n} \{\mathbb{E}|\xi_i(\Delta_{n,1} - \Delta_{n,1}^{(i)})| + \mathbb{E}|\xi_i(\Delta_{n,2} - \Delta_{n,2}^{(i)})|\}. \end{aligned}$$

- Remark:
  A similar result was obtained by Chen and Shao (Bernoulli, 2007) with $\mathbb{E}|W_n(\Delta_{n,2} - \Delta_{n,1})|$ instead of $\mathbb{E}|\Delta_{n,2} - \Delta_{n,1}|$. It turns out that using the term $\mathbb{E}|W_n(\Delta_{n,2} - \Delta_{n,1})|$ will not yield the optimal results in the current case.

- Why the randomized concentration inequality is important?

Assume that we already have a good approximation

$$\mathbb{P}(W_n \geq x) \sim A(x) \quad \text{as } n \to \infty$$

uniformly in $0 \leq x \leq o(a_n)$, for some $a_n \uparrow \infty$.

- Why the randomized concentration inequality is important?

Assume that we already have a good approximation

$$\mathbb{P}(W_n \geq x) \sim A(x) \quad \text{as } n \to \infty$$

uniformly in $0 \leq x \leq o(a_n)$, for some $a_n \uparrow \infty$.

If there is a random perturbation around x, say, $\mathbb{P}(W_n \geq x + \Delta_n)$, it is easy to see that

$$\begin{aligned}
\mathbb{P}(W_n \geq x + \Delta_n) &\leq \mathbb{P}(W_n \geq x - \delta_n) + \mathbb{P}(\Delta_n < -\delta_n), \\
\mathbb{P}(W_n \geq x + \Delta_n) &\geq \mathbb{P}(W_n \geq x + \delta_n) - \mathbb{P}(\Delta_n > \delta_n).
\end{aligned}$$

Therefore, we need to choose δ_n big enough so that $\mathbb{P}(|\Delta_n| > \delta_n)$ decays rapidly, and small enough so that as $n \to \infty$,

$$A(x) \sim A(x \pm \delta_n) \quad \text{uniformly in } 0 \leq x \leq o(a_n).$$

Unfortunately, the above strategy does not always work. Assume $\Delta_n \geq 0$ for simplicity, so instead we try to work directly on

$$\mathbb{P}(W_n \geq x) - \mathbb{P}(W_n \geq x + \Delta_n) = \mathbb{P}(x \leq W_n < x + \Delta_n),$$

where $\Delta_n = \Delta_n(\xi_1, \ldots, \xi_n)$.

When $\Delta_n = \delta_n \geq 0$ is nonrandom, we have

$$\mathbb{P}(x \leq W_n < x + \delta_n) \leq C\Big\{[\Phi(x + \delta_n) - \Phi(x)] + (\beta_2 + \beta_3)\Big\},$$

where $\beta_2 + \beta_3$ comes from the Berry-Esseen inequality.

- Main references

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximation: The Chen-Stein method. *Ann. Probab.* **17** 9-25.

CAI, T. T. and JIANG, T. (2011). Limiting Laws of Coherence of Random Matrices with Applications to Testing Covariance Structure and Construction of Compressed Sensing Matrices. *Ann. Statist.* **39** 1496-1525.

DE LA PEÑA, V.H., LAI, T. L. and SHAO, Q.-M. (2009). *Self-Normalized Processes: Theory and Statistical Applications*. Springer, Berlin.

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865-880.

SHAO, Q.-M. and ZHOU, W.-X. (2013a). Cramér type moderate deviation theorems for self-normalized processes. *Submitted.*

SHAO, Q.-M. and ZHOU, W.-X. (2013b). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab.* To appear.

THANK YOU !!!