High Dimensional Testing for non-Gaussian Data

Danna Zhang

Graduate Colloquium, University of California, San Diego (via Zoom)

January 5, 2021

Hypothesis Testing

- **Toy example for statistical testing:** Administer a drug to n patients for 2 weeks. Let X_i be the reduction in blood pressure seen in the i-th patient. Model $X_1, X_2, \ldots, X_n$ as i.i.d. $N(\mu, \sigma^2)$. Problem: Test $\mu = 0$.
- If $\mu = 0$, we have $\bar{X}_n \sim N(0, \sigma^2/n)$. Intuitively, a value more than 2 standard deviations away from the mean is unusual.
  That is, if $\left|\frac{\bar{X}_n}{\sigma/\sqrt{n}}\right| \geq 2$, we reject $\mu = 0$.
- In statistics, we often set a small $\alpha \in (0, 1)$ (say, $\alpha = 0.05, 0.01$) and reject the hypothesis if

$$\left|\frac{\bar{X}_n}{\sigma/\sqrt{n}}\right| \geq z_{1-\alpha/2}$$

  where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th percentile of $N(0, 1)$.

Hypothesis Testing

- **Toy example for statistical testing:** Administer a drug to n patients for 2 weeks. Let X_i be the reduction in blood pressure seen in the i-th patient. Model $X_1, X_2, \dots, X_n$ as i.i.d. $N(\mu, \sigma^2)$. Problem: Test $\mu = 0$.
- If $\mu = 0$, we have $\bar{X}_n \sim N(0, \sigma^2/n)$. Intuitively, a value more than 2 standard deviations away from the mean is unusual.
  That is, if $\left|\frac{\bar{X}_n}{\sigma/\sqrt{n}}\right| \geq 2$, we reject $\mu = 0$.
- In statistics, we often set a small $\alpha \in (0, 1)$ (say, $\alpha = 0.05, 0.01$) and reject the hypothesis if

$$\left|\frac{\bar{X}_n}{\sigma/\sqrt{n}}\right| \geq z_{1-\alpha/2}$$

  where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th percentile of $N(0, 1)$.

Hypothesis Testing

- **Toy example for statistical testing:** Administer a drug to n patients for 2 weeks. Let X_i be the reduction in blood pressure seen in the i-th patient. Model $X_1, X_2, \ldots, X_n$ as i.i.d. $N(\mu, \sigma^2)$. Problem: Test $\mu = 0$.
- If $\mu = 0$, we have $\bar{X}_n \sim N(0, \sigma^2/n)$. Intuitively, a value more than 2 standard deviations away from the mean is unusual.
  That is, if $\left|\frac{\bar{X}_n}{\sigma/\sqrt{n}}\right| \geq 2$, we reject $\mu = 0$.
- In statistics, we often set a small $\alpha \in (0, 1)$ (say, $\alpha = 0.05, 0.01$) and reject the hypothesis if

$$\left|\frac{\bar{X}_n}{\sigma/\sqrt{n}}\right| \geq z_{1-\alpha/2}$$

  where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th percentile of $N(0, 1)$.

Inference for Covariances on High-dim Time Series Data

- Neuroimaging methods (EEG, fMRI, MEG, etc.)

  $\Downarrow$

  Time series analysis

  $\Downarrow$

  Functional brain connectivity

- Financial/economic data

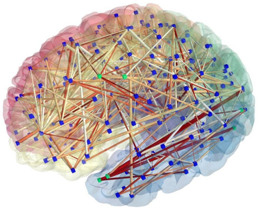
  $\Downarrow$

  High-dim time series modelling VAR, VARMA, ARCH, GARCH, etc.

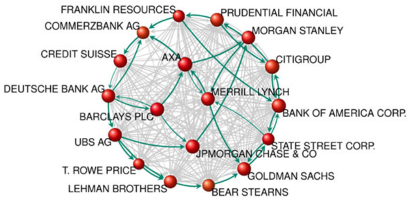
  $\Downarrow$

  Interaction and co-movement

Require simultaneous inference for covariances



Brain Network



Financial Network

Inference for Posterior Means in MCMC Experiments

- Markov chain on the state space $\mathcal{X}$

  $\boxed{Y_1} \to \boxed{Y_2} \to \cdots \to \boxed{Y_n} \to \cdots \Rightarrow \pi$

- Function $h : \mathcal{X} \to \mathbb{R}$

  e.g. means, quantiles, etc.

$$\mathbb{E}_\pi h = \int_{\mathcal{X}} h(y)\pi(dy)$$

$$\Downarrow$$

MCMC: $\hat{h} = n^{-1} \sum_{i=1}^{n} h(Y_i)$. How accurate? CLT[1]

$$\Downarrow$$

What if we have $h_1, h_2, \ldots, h_p$ with $p := p_n \to \infty$?

$$X_1 = \begin{pmatrix} h_1(Y_1) \\ h_2(Y_1) \\ \vdots \\ h_p(Y_1) \end{pmatrix}, \; X_2 = \begin{pmatrix} h_1(Y_2) \\ h_2(Y_2) \\ \vdots \\ h_p(Y_2) \end{pmatrix}, \; \ldots, \; X_n = \begin{pmatrix} h_1(Y_n) \\ h_2(Y_n) \\ \vdots \\ h_p(Y_n) \end{pmatrix}$$

[1] G.L. Jones. On the Markov chain central limit theorem. *Probability surveys*. 2004.
J.M. Flegal, G.L. Jones. Implementing Markov Chain Monte Carlo: Estimating with Confidence. *Handbook of MCMC*. 2011
Y.F. Atchadé. Markov Chain Monte Carlo confidence intervals. *Bernoulli*. 2016.

Inference for Posterior Means in MCMC Experiments

- Markov chain on the state space $\mathcal{X}$

$$\boxed{Y_1} \to \boxed{Y_2} \to \cdots \to \boxed{Y_n} \to \cdots \Rightarrow \pi$$

- Function $h : \mathcal{X} \to \mathbb{R}$

e.g. means, quantiles, etc.

$$\searrow \qquad \swarrow$$

$$\mathbb{E}_\pi h = \int_{\mathcal{X}} h(y)\pi(dy)$$

$$\Downarrow$$

MCMC: $\hat{h} = n^{-1} \sum_{i=1}^{n} h(Y_i)$. How accurate? CLT[1]

$$\Downarrow$$

What if we have $h_1, h_2, \ldots, h_p$ with $p := p_n \to \infty$?

$$X_1 = \begin{pmatrix} h_1(Y_1) \\ h_2(Y_1) \\ \vdots \\ h_p(Y_1) \end{pmatrix}, \; X_2 = \begin{pmatrix} h_1(Y_2) \\ h_2(Y_2) \\ \vdots \\ h_p(Y_2) \end{pmatrix}, \; \ldots, \; X_n = \begin{pmatrix} h_1(Y_n) \\ h_2(Y_n) \\ \vdots \\ h_p(Y_n) \end{pmatrix}$$

---

[1] G.L. Jones. On the Markov chain central limit theorem. *Probability surveys*. 2004.
J.M. Flegal, G.L. Jones. Implementing Markov Chain Monte Carlo: Estimating with Confidence. *Handbook of MCMC*. 2011
Y.F. Atchadé. Markov Chain Monte Carlo confidence intervals. *Bernoulli*. 2016.

High Dimensional Time Series

$$
\begin{matrix}
X_{11} & X_{21} & \ldots & X_{n1} \\
X_{12} & X_{22} & \ldots & X_{n2} \\
\vdots & \vdots & & \vdots \\
X_{1p} & X_{2p} & \ldots & X_{np}
\end{matrix}
$$

$$
\begin{matrix}
X_{11} & X_{21} & \cdots & X_{n1} \\
X_{12} & X_{22} & \cdots & X_{n2} \\
\vdots & \vdots & & \vdots \\
X_{1p} & X_{2p} & \cdots & X_{np} \\
\mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n
\end{matrix}
$$

High Dimensional Time Series

$$\begin{array}{cccc}
X_{11} & X_{21} & \dots & X_{n1} \\
X_{12} & X_{22} & \dots & X_{n2} \\
\vdots & \vdots & & \vdots \\
X_{1p} & X_{2p} & \dots & X_{np}
\end{array}$$

$$\begin{array}{cccc}
X_{11} & X_{21} & \dots & X_{n1} \\
X_{12} & X_{22} & \dots & X_{n2} \\
\vdots & \vdots & & \vdots \\
X_{1p} & X_{2p} & \dots & X_{np} \\
X_1 & X_2 & \dots & X_n
\end{array}$$

High Dimensional Time Series

$$
\begin{array}{cccc}
X_{11} & X_{21} & \dots & X_{n1} \\
X_{12} & X_{22} & \dots & X_{n2} \\
\vdots & \vdots & & \vdots \\
X_{1p} & X_{2p} & \dots & X_{np}
\end{array}
$$

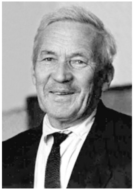
$$
\begin{array}{cccc}
X_{11} & X_{21} & \dots & X_{n1} \\
X_{12} & X_{22} & \dots & X_{n2} \\
\vdots & \vdots & & \vdots \\
X_{1p} & X_{2p} & \dots & X_{np} \\
\mathbf{X_1} & \mathbf{X_2} & \dots & \mathbf{X_n}
\end{array}
$$

Kolmogorov-Smirnov Test

Let $X_i \in \mathbb{R}$, $i \in \mathbb{Z}$, be i.i.d. random variables. Test the hypothesis that

$$H_0 : \mathbb{P}(X_i \leq x) = F(x), \quad \text{for } x \in \mathbb{R}.$$

Kolmogorov-Smirnov Test



A.N. Kolmogorov

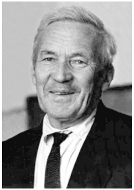
Test statistic $T_n = \sup_{x \in \mathcal{I}} \sqrt{n} \left| F_n(x) - F(x) \right|$

Asymptotic distribution[2]

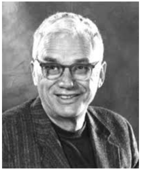
$$\mathbb{P}(T_n \geq x) \to 2 \sum_{m=1}^{\infty} (-1)^{m+1} e^{-2m^2 x^2}$$

[2] A. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuar.* 1933.
J.L. Doob. Heuristic approach to the Kolmogorov-Smirnov Theorems. *Annals of Mathematical Statistics*. 1949.
M.D. Donsker. Justification and extension of Doob's heuristic approach to the Kolmogorov- Smirnov Theorems. 1952.

Kolmogorov-Smirnov Test



A.N. Kolmogorov



J.L. Doob

Test statistic $T_n = \sup_{x \in \mathcal{I}} \sqrt{n} \left| F_n(x) - F(x) \right|$

Asymptotic distribution[2]

$$\mathbb{P}(T_n \geq x) \to 2 \sum_{m=1}^{\infty} (-1)^{m+1} e^{-2m^2 x^2}$$

Doob's heuristic argument:

$\mathcal{I}$: $x_1 \quad x_2 \quad \cdots \quad \cdots \quad x_L$

$$\sqrt{n} \begin{pmatrix} F_n(x_1) - F(x_1) \\ F_n(x_2) - F(x_2) \\ \vdots \\ F_n(x_L) - F(x_L) \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbb{B}(F(x_1)) \\ \mathbb{B}(F(x_2)) \\ \vdots \\ \mathbb{B}(F(x_L)) \end{pmatrix}$$

[2] A. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuar.* 1933.
J.L. Doob. Heuristic approach to the Kolmogorov-Smirnov Theorems. *Annals of Mathematical Statistics*. 1949.
M.D. Donsker. Justification and extension of Doob's heuristic approach to the Kolmogorov- Smirnov Theorems. 1952.

Kolmogorov-Smirnov Test for High-dim Time Series

Let $X_i \in \mathbb{R}^d$ be a stationary process. Test the hypothesis that

$$H_0 : \mathbb{P}(X_{ij} \le x) = F_j(x), \quad \text{for } x \in \mathbb{R}, \quad 1 \le j \le d.$$

Test statistic $\max_{1 \le j \le d} \sup_{x \in \mathcal{I}} \sqrt{n} \left| F_{nj}(x) - F_j(x) \right|$

Asymptotic distribution?

Another Look:

- Discretization $\Downarrow$ $\mathcal{I}$: $x_1 \; x_2 \; \cdots \; \cdots \; x_L$
- Higher dimension $\max_{1 \le j \le d} \max_{1 \le \ell \le L} \sqrt{n} \left| F_{nj}(x_\ell) - F_j(x_\ell) \right|$ $\quad p = dL$
- Temporal and cross-sectional dependence

High-dimensional CLT $\Rightarrow$ $\left\{ \sqrt{n} \left[F_{nj}(x_\ell) - F_j(x_\ell) \right], 1 \le j \le d, 1 \le \ell \le L \right\}$

Kolmogorov-Smirnov Test for High-dim Time Series

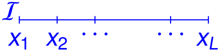
Let $X_i \in \mathbb{R}^d$ be a stationary process. Test the hypothesis that

$$H_0 : \mathbb{P}(X_{ij} \le x) = F_j(x), \quad \text{for } x \in \mathbb{R}, \quad 1 \le j \le d.$$

Test statistic $\quad \max_{1\le j\le d} \sup_{x\in\mathcal{I}} \sqrt{n}\left|F_{nj}(x) - F_j(x)\right|$

Asymptotic distribution?

Another Look:



- Discretization $\quad \Downarrow$
- Higher dimension $\quad \max_{1\le j\le d} \max_{1\le \ell\le L} \sqrt{n}\left|F_{nj}(x_\ell) - F_j(x_\ell)\right| \quad$ $p = dL$
- Temporal and cross-sectional dependence

High-dimensional CLT $\Rightarrow$ $\left\{\sqrt{n}\left[F_{nj}(x_\ell) - F_j(x_\ell)\right], 1 \le j \le d, 1 \le \ell \le L\right\}$

CLT for Time Series

- Stationary $X_i \in \mathbb{R}^p$, $\mathbb{E}X_i = \mu$, $\mathbb{E}(X_i^\top X_i) < \infty$.
  Under suitable weak dependence conditions [3], CLT for $p = 1$ or p fixed:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \Rightarrow N(0, \Sigma) \quad \text{where} \quad \Sigma = \sum_{k=-\infty}^{\infty} \mathbb{E}((X_0 - \mu)(X_k - \mu)^\top)$$

- Portnoy (1986)[4]: CLT fails for i.i.d. random vectors if $p \gg \sqrt{n}$.

[3] M. Rosenblatt. A central limit theorem and a strong mixing condition. *PNAS*. 1956.
I.A. Ibragimov, Ju. V. Linnik. Independent and Stationary Sequences of Random Variables. *Wolters-Noordhoff*. 1971.
W.B. Wu. Nolinear system theory: another look at dependence. *PNAS*. 2005.
J. Dedecker et al. Weak Dependence: With Examples and Applications. *Springer*. 2007.

[4] S. Portney. On the central limit theorem in $\mathbb{R}^p$ when $p \to \infty$. *Probability Theory and Related Fields*. 1986.

CLT for Time Series

- Stationary $X_i \in \mathbb{R}^p$, $\quad \mathbb{E}X_i = \mu$, $\quad \mathbb{E}(X_i^\top X_i) < \infty$.
  Under suitable weak dependence conditions [3], CLT for $p = 1$ or p fixed:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \Rightarrow N(0, \Sigma) \quad \text{where} \quad \Sigma = \sum_{k=-\infty}^{\infty} \mathbb{E}((X_0 - \mu)(X_k - \mu)^\top)$$

- Portnoy (1986)[4]: CLT fails for i.i.d. random vectors if $p \gg \sqrt{n}$.

[3] M. Rosenblatt. A central limit theorem and a strong mixing condition. *PNAS*. 1956.
I.A. Ibragimov, Ju. V. Linnik. Independent and Stationary Sequences of Random Variables. *Wolters-Noordhoff*. 1971.
W.B. Wu. Nolinear system theory: another look at dependence. *PNAS*. 2005.
J. Dedecker et al. Weak Dependence: With Examples and Applications. *Springer*. 2007.

[4] S. Portney. On the central limit theorem in $\mathbb{R}^p$ when $p \to \infty$. *Probability Theory and Related Fields*. 1986.

High-dimensional CLT: GA in $\mathbb{R}^p$

Gaussian Approximation[5] (GA) for i.i.d. random vectors in $\mathbb{R}^p$:

$$\sup_{u \geq 0} \left| \mathbb{P}(\sqrt{n}|\bar{X}_n - \mu|_\infty \geq u) - \mathbb{P}(|Z|_\infty \geq u) \right| \to 0, \quad \text{as } n, p \to \infty$$

under certain conditions, where $Z = (Z_1, \ldots, Z_p)^\top \sim N(0, \mathrm{Cov}(X_i))$.

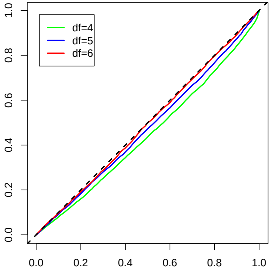
[5] V. Chernozhukov, D. Chetnerikov and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics*. 2013.

High-dimensional CLT: GA in $\mathbb{R}^p$

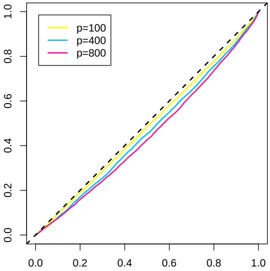
Gaussian Approximation[5] (GA) for i.i.d. random vectors in $\mathbb{R}^p$:

$$\sup_{u \geq 0} \left| \mathbb{P}(\sqrt{n}|\bar{X}_n - \mu|_\infty \geq u) - \mathbb{P}(|Z|_\infty \geq u) \right| \to 0, \quad \text{as } n, p \to \infty$$

under certain conditions, where $Z = (Z_1, \ldots, Z_p)^\top \sim N(0, \mathrm{Cov}(X_i))$.



(a) X_{ij} *i.i.d.* $\sim t(df)$, $n = 200$, $p = 400$.



(b) X_{ij} *i.i.d.* $\sim t(5)$, $n = 100$.

---

[5] V. Chernozhukov, D. Chetnerikov and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics*. 2013.

A General Framework of High-dim Stationary Processes

Stationary and causal processes of the form (nonlinear Wold representation)

$$X_i = (X_{i1}, \ldots, X_{ip})^\top = G(\ldots, \varepsilon_{i-1}, \varepsilon_i)$$

- Input: ε_i, $i \in \mathbb{Z}$ are i.i.d innovations or shocks that derive the system[6].
- Filter: $G(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))^\top$.
- Output: $X_i = G(\ldots, \varepsilon_{i-1}, \varepsilon_i)$. "How output depends on input?" [7]

---

[6] G.E.P. Box, G.M. Jenkins. Time Series Analysis: Forecasting and Control. *Springer*. 1970.

[7] N. Wiener. Nonlinear Problems in Random Theory. *Wiley*. 1958
M.B. Priestley. Nonlinear and Nonstationary Time Series Analysis. *Academic Press*. 1988
H. Tong. Non-linear Time Series: A Dynamical System Approach. *Springer*. 1990.

A General Framework of High-dim Stationary Processes

Stationary and causal processes of the form (nonlinear Wold representation)

$$X_i = (X_{i1}, \ldots, X_{ip})^\top = G(\ldots, \varepsilon_{i-1}, \varepsilon_i)$$

- Input: ε_i, $i \in \mathbb{Z}$ are i.i.d innovations or shocks that derive the system[6].
- Filter: $G(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))^\top$.
- Output: $X_i = G(\ldots, \varepsilon_{i-1}, \varepsilon_i)$. "How output depends on input?" [7]

Examples: Linear processes (e.g. VAR, ARMA) and nonlinear transforms, bilinear models, Volterra processes, Markov chain models, threshold/exponential autoregressive models (TAR/EAR), ARCH/GARCH type models, FARIMA-GARCH models, etc.

---

[6] G.E.P. Box, G.M. Jenkins. Time Series Analysis: Forecasting and Control. *Springer*. 1970.

[7] N. Wiener. Nonlinear Problems in Random Theory. *Wiley*. 1958
M.B. Priestley. Nonlinear and Nonstationary Time Series Analysis. *Academic Press*. 1988
H. Tong. Non-linear Time Series: A Dynamical System Approach. *Springer*. 1990.

A General Framework of High-dim Stationary Processes

Stationary and causal processes of the form (nonlinear Wold representation)

$$X_i = (X_{i1}, \ldots, X_{ip})^\top = G(\ldots, \varepsilon_{i-1}, \varepsilon_i)$$

- Input: ε_i, $i \in \mathbb{Z}$ are i.i.d innovations or shocks that derive the system[6].
- Filter: $G(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))^\top$.
- Output: $X_i = G(\ldots, \varepsilon_{i-1}, \varepsilon_i)$. "How output depends on input?" [7]

Example: High-dimensional linear processes

$$X_i = \sum_{k=0}^{\infty} A_k \varepsilon_{i-k}$$

where ε_i are i.i.d. with $\mathbb{E}\varepsilon_{ij} = 0$ and $\mathbb{E}\varepsilon_{ij}^2 = 1$, $A_k \in \mathbb{R}^{p \times p}$ satisfy $\sum_{k=0}^{\infty} \mathrm{tr}(A_k^\top A_k) < \infty$.

[6] G.E.P. Box, G.M. Jenkins. Time Series Analysis: Forecasting and Control. *Springer*. 1970.

[7] N. Wiener. Nonlinear Problems in Random Theory. *Wiley*. 1958
M.B. Priestley. Nonlinear and Nonstationary Time Series Analysis. *Academic Press*. 1988
H. Tong. Non-linear Time Series: A Dynamical System Approach. *Springer*. 1990.

Functional Dependence Measures

- Functional dependence measure (Wu, 2005)[8]

$$X_i = G(\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_i)$$
$$\Downarrow$$
$$X_i^* = G(\ldots, \varepsilon_{-1}, \varepsilon_0^*, \varepsilon_1, \ldots, \varepsilon_i)$$

$$q \geq 2,\ i \geq 0,\ 1 \leq j \leq p$$
$$\delta_{i,q,j} = \left\| X_{ij} - X_{ij}^* \right\|_q$$
$$\omega_{i,q} = \left\| \, |X_i - X_i^*|_\infty \right\|_q$$

- High-dimensional dependence

Cross-sectional dependence

$$\Upsilon_{q,\alpha} = \left(\sum_{j=1}^{p} \|X_{\cdot j}\|_{q,\alpha}^q \right)^{1/q}$$

Temporal dependence

$$\Psi_{q,\alpha} = \max_{1 \leq j \leq p} \|X_{\cdot j}\|_{q,\alpha}$$

[8] W.B. Wu. Nonlinear system theory: another look at dependence. *PNAS*. 2005.

High-dimensional CLT: GA in $\mathbb{R}^p$

- Assume $\mathbb{E}X_i = \mu$. The $p \times p$ long-run covariance matrix is

$$\Sigma = (\sigma_{jk}) = \sum_{k=-\infty}^{\infty} \Gamma(k), \text{ where } \Gamma(k) = \mathbb{E}(X_0 - \mu)(X_k - \mu)^\top.$$

- Let $Z = (Z_1, \ldots, Z_p)^\top \sim N(0, \Sigma)$.

Theorem

1. Assume there exists a constant $c > 0$ s.t. $\min_{1 \le j \le p} \sigma_{jj} > c$.
2. Under certain conditions on n, p and dependence measures.

$$\sup_{u \ge 0} \left| \mathbb{P}\left(\max_{1 \le j \le p} \sqrt{n} |\bar{X}_{nj} - \mu_j| / \sqrt{\sigma_{jj}} \le u \right) - \mathbb{P}\left(\max_{1 \le j \le p} |Z_j| / \sqrt{\sigma_{jj}} \le u \right) \right| \to 0$$

D. Zhang, W.B. Wu. Gaussian approximation for high dimensional time series. *Annals of Statistics*. 2017

High-dimensional CLT: GA in $\mathbb{R}^p$

- Assume $\mathbb{E}X_i = \mu$. The $p \times p$ long-run covariance matrix is

$$\Sigma = (\sigma_{jk}) = \sum_{k=-\infty}^{\infty} \Gamma(k), \text{ where } \Gamma(k) = \mathbb{E}(X_0 - \mu)(X_k - \mu)^\top.$$

- Let $Z = (Z_1, \ldots, Z_p)^\top \sim N(0, \Sigma)$.

Theorem

1. Assume there exists a constant $c > 0$ s.t. $\min_{1 \le j \le p} \sigma_{jj} > c$.
2. Under certain conditions on n, p and dependence measures.

$$\sup_{u \ge 0} \left| \mathbb{P}\left(\max_{1 \le j \le p} \sqrt{n} |\bar{X}_{nj} - \mu_j| / \sqrt{\sigma_{jj}} \le \underset{\substack{\uparrow \\ \chi_\theta}}{u} \right) - \mathbb{P}\left(\max_{1 \le j \le p} |Z_j| / \sqrt{\sigma_{jj}} \le \underset{\substack{\uparrow \\ \chi_\theta}}{u} \right) \right| \to 0$$

D. Zhang, W.B. Wu. Gaussian approximation for high dimensional time series. *Annals of Statistics*. 2017

Some applications

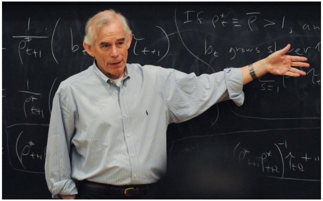
Application 1: Network connectivity detection by the inference of covariances.

Application 2: Inference for many posterior means in MCMC experiments.

Application 3: Testing the distribution of high dimensional data

$$\begin{pmatrix} \bullet & \bullet & \cdots & \bullet \\ \bullet & \bullet & \cdots & \bullet \\ \vdots & \vdots & \ddots & \vdots \\ \bullet & \bullet & \cdots & \bullet \end{pmatrix}_{p\times p} \longrightarrow \begin{pmatrix} \bullet \\ \bullet \\ \bullet \\ \vdots \\ \bullet \\ \bullet \\ \bullet \end{pmatrix}_{p^2}, \; Y_i \longrightarrow \begin{pmatrix} {\color{red}h_1}(Y_i) \\ {\color{red}h_2}(Y_i) \\ \vdots \\ {\color{red}h_p}(Y_i) \end{pmatrix}, \quad \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{id} \end{pmatrix} \longrightarrow \begin{pmatrix} 1\{X_{i1} \le x\} \\ 1\{X_{i2} \le x\} \\ \vdots \\ 1\{X_{id} \le x\} \end{pmatrix}$$

High-dimensional Inference for Parametric Time Series



Christopher A. Sims

Vector autoregressive (Sims, 1980)[9]

VAR(d): $X_i = A_1 X_{i-1} + \ldots + A_d X_{i-d} + \varepsilon_i.$

Old friend: Likelihood ratio test, F-test
Wald test, t-test

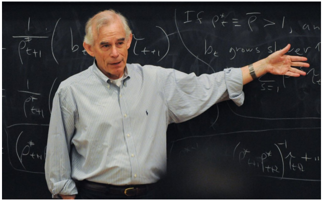
Statistical phenomenon in economic data:

high dimension $\Rightarrow$ degrees of freedom left $\searrow$

fat-tailed residuals $\Rightarrow$ false rejection of the null

[9] C.A. Sims. Macroeconomics and reality. *Econometrica*. 1980

High-dimensional Inference for Parametric Time Series



Christopher A. Sims

Vector autoregressive (Sims, 1980)[9]

VAR(d): $X_i = A_1 X_{i-1} + \ldots + A_d X_{i-d} + \varepsilon_i.$

Old friend: Likelihood ratio test, *F*-test
Wald test, *t*-test

Statistical phenomenon in economic data:

high dimension	$\Rightarrow$	degrees of freedom left $\searrow$
fat-tailed residuals	$\Rightarrow$	false rejection of the null

[9] C.A. Sims. Macroeconomics and reality. *Econometrica*. 1980

Estimation and Inference

Rewrite VAR(d) model as $\underbrace{Y}_{np\times 1} = \underbrace{Z}_{np\times dp^2}\underbrace{\beta}_{dp^2\times 1} + \underbrace{\epsilon}_{np\times 1}$.

- Moderately high-dim case $dp = o(n)$, establish asymptotic theory for

$$\hat{\beta} = \mathrm{argmin}_{\beta}|Y - Z\beta|_2^2 = (Z^\top Z)^{-1}Z^\top Y.$$

- Very high-dim case $n = o(dp)$, consider Lasso/Danzig-type estimator

$$\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^{dp^2}}(|Y - Z\beta|_2^2 + \lambda|\beta|_1),$$
$$\hat{\beta} = \arg\min|\beta|_1 \text{ subject to } |Z^\top Z\beta - Z^\top Y|_\infty \le \lambda.$$

- For heavy-tailed innovations, apply robust estimation approach

$$\hat{\beta}_H = \mathrm{argmin}_{\beta}H(Y - Z\beta),$$

where H can be Huber, regression quantile, L^q regression, etc.

Estimation and Inference

Rewrite VAR(d) model as $\underbrace{Y}_{np\times 1} = \underbrace{Z}_{np\times dp^2} \underbrace{\beta}_{dp^2\times 1} + \underbrace{\epsilon}_{np\times 1}.$

- Moderately high-dim case $dp = o(n)$, establish asymptotic theory for
$$\hat{\beta} = \mathrm{argmin}_{\beta}|Y - Z\beta|_2^2 = (Z^\top Z)^{-1} Z^\top Y.$$

- Very high-dim case $n = o(dp)$, consider Lasso/Danzig-type estimator
$$\hat{\beta} = \arg\min_{\beta\in\mathbb{R}^{dp^2}} (|Y - Z\beta|_2^2 + \lambda|\beta|_1),$$
$$\hat{\beta} = \arg\min |\beta|_1 \text{ subject to } |Z^\top Z\beta - Z^\top Y|_\infty \le \lambda.$$

- For heavy-tailed innovations, apply robust estimation approach
$$\hat{\beta}_H = \mathrm{argmin}_{\beta} H(Y - Z\beta),$$
where H can be Huber, regression quantile, L^q regression, etc.

Granger Causality Test

- Granger causality test: VAR-X model

$$X_i = AX_{i-1} + \varepsilon_i,$$
$$Y_i = BX_{i-1} + CY_{i-1} + \xi_i.$$

Let $W_i = (X_i^\top, Y_i^\top)^\top, \eta_i = (\varepsilon_i^\top, \xi_i^\top)^\top$. Then we can write the model as

$$W_i = MW_{i-1} + \eta_i, \text{ where } M = \begin{pmatrix} A & \mathbf{0} \\ B & C \end{pmatrix}.$$

Test if $B = \mathbf{0}$.

Portmanteau Test

- One-dimensional Box-Pierce, Ljung-Box portmanteau test statistic:

$$Q_{BP} = n \sum_{k=1}^{m} \hat{\gamma}_k^2 \quad \text{or} \quad Q_{LB} = n(n+2) \sum_{k=1}^{m} \hat{\gamma}_k^2/(n-k).$$

- Another natural choice of test statistic:

$$K_n = \sqrt{n} \max_{1 \le k \le m} |\hat{\gamma}_k - \gamma_k|.$$

- For high dimensional data, need to establish an asymptotic theory on

$$\mathcal{Q}_n = n \sum_{k=1}^{s_n} |\hat{\Gamma}_k - \mathbb{E}\hat{\Gamma}_k|_F^2 \quad \text{and} \quad \mathcal{K}_n = \sqrt{n} \max_{1 \le k \le s_n} |\hat{\Gamma}_k - \mathbb{E}\hat{\Gamma}_k|_\infty,$$

where $s_n \to \infty$ and $s_n = o(n)$.

Further Considerations on Gaussian Approximation

Various χ^2 Tests

Kolmogorov-Smirnov Test

Mean Inference

Cramér-von Mises Test

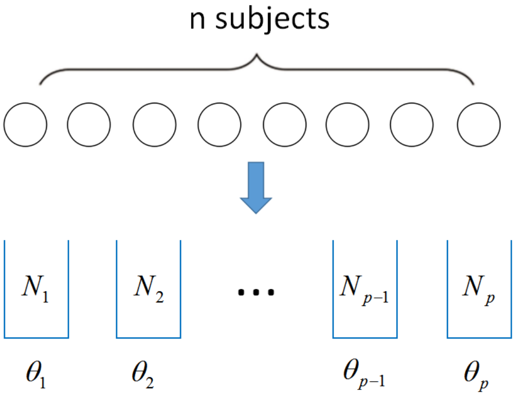
L_∞ Gaussian Approximation

L_2 Gaussian Approximation

$$f(\sqrt{n}\bar{X}_n) \quad \Leftrightarrow \quad f(Z)$$

$$f(X_1, X_2, \ldots, X_n) \quad \Leftrightarrow \quad f(Z_1, Z_2, \ldots, Z_n)$$

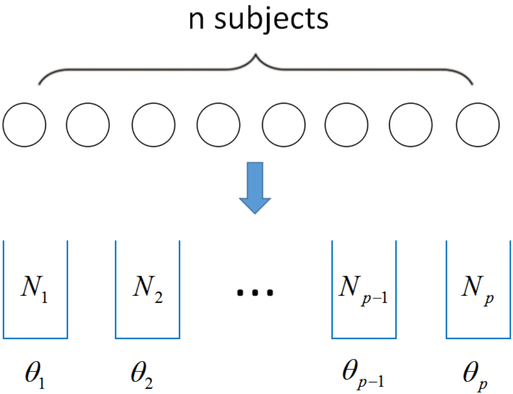
Pearson's χ^2 test



Goal: Test the hypothesis that the observations $(N_1, \ldots, N_p)$ satisfy

$$(N_1, \ldots, N_p) \sim Multi(n; \theta_1, \ldots, \theta_p), \text{ where } \theta_1 + \ldots + \theta_p = 1.$$

Pearson's χ^2 test



Goal: Test the hypothesis that the observations $(N_1, \ldots, N_p)$ satisfy

$$(N_1, \ldots, N_p) \sim Multi(n; \theta_1, \ldots, \theta_p), \text{ where } \theta_1 + \ldots + \theta_p = 1.$$

Pearson's χ^2 Statistic

Pearson's χ^2 statistic:

$$\chi^2 = \sum_{j=1}^{p} \frac{(N_j - n\theta_j)^2}{n\theta_j} = \sum_{j=1}^{p} \frac{n(\hat{\theta}_j - \theta_j)^2}{\theta_j} \text{ with } \hat{\theta}_j = \frac{N_j}{n}.$$

Asymptotic theory: when p is small and fixed, $\chi^2 \Rightarrow \chi^2_{p-1}$, by the CLT

$$\{(N_j - n\theta_j)/(n\theta_j)^{1/2}, j = 1, \ldots, p\} \Rightarrow N(0, \Sigma), \text{ with } \sigma_{jj'} = \mathbf{1}_{j=j'} - \sqrt{\theta_j \theta_{j'}}.$$

Rule of thumb in classical statistical textbooks: $n\theta_j \geq 5$ for all j.

Decision rule: For significance level $\alpha \in (0, 1)$, we reject the hypothesis if

$$\chi^2 > \chi^2_{p-1,1-\alpha}, \text{ the } (1-\alpha)\text{-th quantile of } \chi^2_{p-1}.$$

Motivating Example: Social Life Feeling Data

1490 respondents, 2^5 distinct response patterns, 5 propositions:

1. Anyone can raise his living standard if he is willing to work at it.
2. Our country has too many poor people who can do little to raise their living standard.
3. Individuals are poor because of the lack of effort on their part.
4. Poor people could improve their lot if they tried.
5. Most people have a good deal of freedom in deciding how to live.

x	00000	01000	00001	10000	00100	01001	11000	10001	01100	00101	00010	10100	11001	01101	01010	11100
O	156	174	26	8	127	35	8	2	208	26	14	4	2	65	36	19
E	162.0	174.0	22.2	5.2	122.3	31.9	8.1	1.1	208.7	30.2	16.9	8.3	2.3	65.3	37.6	19.3

x	00011	10101	10010	00110	01011	11101	11010	10011	01110	00111	10110	11011	01111	11110	10111	11111
O	9	4	1	66	13	10	5	3	195	16	18	3	129	31	9	68
E	5.8	3.0	1.8	56.5	16.2	8.8	5.3	0.9	182.7	31.8	11.2	3.4	130.9	49.9	9.6	56.7

x: response pattern

O: Observed frequency for each response pattern

E: Expected frequency after fitting the logit-probit model

Theory for Pearson's χ^2 Test Statistic

$B_i = (B_{i1}, \ldots, B_{ip}) \sim Multi(1; \theta_1, \ldots, \theta_p)$. Let $X_{ij} = (B_{ij} - \theta_j)/\sqrt{\theta_j}$. Then

$$\chi^2 = n\bar{X}_n^\top \bar{X}_n.$$

Theorem: (i) Assume that for some $0 < \delta \le 1$,

$$L_\delta = \frac{\sum_{j=1}^p \theta_j^{-\delta}}{n^\delta p^{1+\delta/2}} \to 0.$$

Further assume $\sigma_p^2 = o(np)$, where $\sigma_p^2 = \sum_{j=1}^p \theta_j^{-1} - p^2$. Then

$$\sup_t |\mathbb{P}(\chi^2 \le t) - \mathbb{P}\left(\chi^2_{p-1} \le t\right)| \to 0. \tag{1}$$

(ii) Assume $np = o(\sigma_p^2)$ and the Lindeberg condition holds for $W_i = (\sum_{j=1}^p B_{ij}/\theta_j - p)/\sigma_p$. Then we have the CLT

$$\sup_t |\mathbb{P}(\chi^2 - (p-1) \le n^{-1/2}\sigma_p t) - \Phi(t)| \to 0. \tag{2}$$

Diagonal-removed χ^2

Due to dichotomous asymptotic distributions of χ^2, a diagonal-removed version is suggested:

$$^*\chi^2 = \chi^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^\top X_i = \chi^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{(B_{ij}-\theta_j)^2}{\theta_j} = \chi^2 - \sum_{j=1}^{p}\frac{\hat{\theta}_j}{\theta_j} + 1$$

Theorem: Assume that for some $0 < \delta \leq 1$,

$$L_\delta = \frac{\sum_{j=1}^{p}\theta_j^{-\delta}}{n^\delta p^{1+\delta/2}} \to 0.$$

Then

$$\sup_t |\mathbb{P}(^*\chi^2 \leq t) - \mathbb{P}\left(\chi^2_{p-1} - (p-1) \leq t\right)| = O(L_\delta^{1/(3+\delta)}) \to 0.$$

M. Xu, D. Zhang and W.B. Wu. Pearson's chi-squared statistics: approximation theory and beyond *Biometrika*. 2019

Key Tool: High-dimensional Invariance Principle

Let $X_i \in \mathbb{R}^p$, $i \in \mathbb{N}$, be i.i.d. random vectors with $\mathbb{E}(X_i) = 0$, $\mathrm{Cov}(X_i) = \Sigma$.
Let Y_i, $i \in \mathbb{N}$, be i.i.d. $N(0, \Sigma)$ random vectors.
In the classical case with fixed dimension, due to CLT ($\sqrt{n}\bar{X}_n \Rightarrow N(0, \Sigma)$),

$$\sup_t |\mathbb{P}(n\bar{X}_n^\top \bar{X}_n \le t) - \mathbb{P}(n\bar{Y}_n^\top \bar{Y}_n \le t)| \to 0. \qquad (*)$$

Special case (Pearson's χ^2 test): $B_i = (B_{i1}, \ldots, B_{ip}) \sim Multi(1; \theta_1, \ldots, \theta_p)$.
Let $X_{ij} = (B_{ij} - \theta_j)/\sqrt{\theta_j}$. Then

$$\chi^2 = n\bar{X}_n^\top \bar{X}_n.$$

Goal: Show $(*)$ holds in the high-dimensional case where $p \to \infty$.

Social Life Feeling Date Analysis

x	00000	01000	00001	10000	00100	01001	11000	10001	01100	00101	00010	10100	11001	01101	01010	11100
O	156	174	26	8	127	35	8	2	208	26	14	4	2	65	36	19
E	162.0	174.0	22.2	5.2	122.3	31.9	8.1	1.1	208.7	30.2	16.9	8.3	2.3	65.3	37.6	19.3

x	00011	10101	10010	00110	01011	11101	11010	10011	01110	00111	10110	11011	01111	11110	10111	11111
O	9	4	1	66	13	10	5	3	195	16	18	3	129	31	9	68
E	5.8	3.0	1.8	56.5	16.2	8.8	5.3	0.9	182.7	31.8	11.2	3.4	130.9	49.9	9.6	56.7

Comparison of two methods:

- Pearson's χ^2 test statistic $\chi^2 = 38.93$ with the degrees of freedom $2^5 - 1 - 10$(number of parameters) $= 21$ and p-value is 0.01 based on the approximated distribution χ^2_{21}, suggesting a significant lack of fit.
- New test statistic ${}^*\chi^2 = 4.79$ and the p-value is 0.21 based on the approximated distribution $\chi^2_{21} - 21$, indicating the logit-probit model is satisfactory fitted to the data. Note that $L_\delta = 0.029$ for $\delta = 1$.

High-dimensional Scheme for Classical Problems

- Goodness of fit
  - Kolmogorov-Smirnov test — $\sup_{x\in\mathcal{I}} |F_n(x) - F(x)|$
  - Cramér-von Mises test — $\int_{x\in\mathcal{I}} [F_n(x) - F(x)]^2 dF(x)$
  - χ^2 test — e.g. Pearson's χ^2 test
    Freeman-Tukey test
- Joint distribution function
  - Tail-dependence in stock return pairs: $x_1, \ldots, x_p \in \mathcal{I}$, $y_1, \ldots, y_p \in \mathcal{J}$
    "Positive tail dependence": $P(X \geq x, Y \geq y) - P(X \geq x)P(Y \geq y) > 0$
  - Volatility of market index (e.g. Dow Jones Industrial Average (DJIA))
    $\mathbb{P}(X_{n+1} \geq x | X_n \geq x) - \mathbb{P}(X_{n+1} \geq x)$ for large x

Thank you for your attention!