Empirical Likelihood Tests for High-dimensional Data

Ruodu Wang

Department of Statistics and Actuarial Science
University of Waterloo, Canada

ICSA - Canada Chapter 2013 Symposium
Toronto, August 2 - 3, 2013

Based on joint work with Liang Peng and Rongmao Zhang

Contents

Introduction

In this talk we discuss empirical likelihood ratio tests for high-dimensional data.

Let $X_i = (X_{i1}, \ldots, X_{ip}),\ i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \le i,j \le p}$. Here p may depend on n.

- If p is fixed, then it is a traditional statistical setting.
- If $p \to \infty$, then it is high-dimensional setting.

Typical testing questions:

- Testing $\mu = \mu_0$ (one sample mean test).
- Two sample means testing.
- Testing $\Sigma = \Sigma_0$ (covariance matrix test)
- Two sample covariance matrices testing.

We focus on testing covariance matrices.

Problem setup. Let $X_i = (X_{i1}, \dots, X_{ip}),\ i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \le i,j \le p}$.

Problem setup. Let $X_i = (X_{i1}, \ldots, X_{ip}),\ i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \le i,j \le p}$.

- Testing covariance matrix

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_1 : \Sigma \neq \Sigma_0. \tag{1}$$

Problem setup. Let $X_i = (X_{i1}, \ldots, X_{ip}),\ i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \le i,j \le p}$.

- Testing covariance matrix

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_1 : \Sigma \neq \Sigma_0. \tag{1}$$

- Testing bandedness

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau. \tag{2}$$

Problem setup. Let $X_i = (X_{i1}, \ldots, X_{ip}),\ i = 1, 2, \cdots, n$ be iid random vectors with mean $\mu = (\mu_1, \cdots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1 \le i,j \le p}$.

- Testing covariance matrix

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_1 : \Sigma \neq \Sigma_0. \tag{1}$$

- Testing bandedness

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau. \tag{2}$$

Non-parametric. No information about sparsity.

Literature.

- Testing (1) for fixed p: traditional likelihood ratio test; scaled distance measure test (John (1971, 1972) and Nagao (1973)).

Literature.

- Testing (1) for fixed p: traditional likelihood ratio test; scaled distance measure test (John (1971, 1972) and Nagao (1973)).
- Testing (1) for divergent p: Ledoit and Wolf (2002) for normal X_i and Chen, Zhang and Zhong (2010) for a factor model.
  - Specific models are imposed.
  - Restrictions are put on p.
  - p has to go to infinity as n approaches infinity.

- Testing (2) for divergent p: Cai and Jiang (2011).
  - The test statistic: the coherence converges slowly.
  - Normality are assumed.
  - Restrictions are put on p and τ.

- Testing (2) for divergent p: Cai and Jiang (2011).
  - The test statistic: the coherence converges slowly.
  - Normality are assumed.
  - Restrictions are put on p and τ.
- Testing (2) for divergent p: Qiu and Chen (2012).
  - Similar to Chen, Zhang and Zhong (2010), specific models; restrictions.

- Our goal: build up a test statistic that works for both (1) and (2); loose the condition on p; get rid of specific models.

- Our goal: build up a test statistic that works for both (1) and (2); loose the condition on p; get rid of specific models.
- First we assume μ is known. The case when μ is unknown is very similar.

Testing Covariance Matrices

Basic observations.

- $\Sigma = \Sigma_0$ is equivalent to

$$D^2 := ||\Sigma - \Sigma_0||_F^2 = \mathrm{tr}((\Sigma - \Sigma_0)^2) = 0.$$

Testing Covariance Matrices

Basic observations.

- $\Sigma = \Sigma_0$ is equivalent to

$$D^2 := ||\Sigma - \Sigma_0||_F^2 = \mathrm{tr}((\Sigma - \Sigma_0)^2) = 0.$$

- We can construct our test based on an estimator of D^2.

A natural estimator.

- For $i = 1, \ldots, n$, define the $p \times p$ matrix

$$Y_i = (X_i - \mu)(X_i - \mu)^T,$$

  and estimator

$$e(\Gamma) = \mathrm{tr}((Y_1 - \Gamma)(Y_2 - \Gamma)).$$

- $\mathbb{E}[Y_1] = \Sigma$ and $\mathbb{E}[e(\Sigma_0)] = D^2$. $\mathbb{E}[e(\Sigma_0)] = 0$ is equivalent to $\Sigma = \Sigma_0$.

We need independent copies of (Y_1, Y_2).

We need independent copies of (Y_1, Y_2).

Splitting the sample.

Let $N = [n/2]$. For $i = 1, 2, \ldots, N$, we define

$$e_i(\Sigma) = \mathrm{tr}((Y_i - \Sigma)(Y_{N+i} - \Sigma)).$$

We need independent copies of (Y_1, Y_2).

Splitting the sample.

Let $N = [n/2]$. For $i = 1, 2, \ldots, N$, we define

$$e_i(\Sigma) = \mathrm{tr}((Y_i - \Sigma)(Y_{N+i} - \Sigma)).$$

- Very difficult to estimate the variance of e_i.

We need independent copies of (Y_1, Y_2).

Splitting the sample.

Let $N = [n/2]$. For $i = 1, 2, \ldots, N$, we define

$$e_i(\Sigma) = \mathrm{tr}((Y_i - \Sigma)(Y_{N+i} - \Sigma)).$$

- Very difficult to estimate the variance of e_i.
- Empirical likelihood method automatically catches the asymptotic variance.

Define the empirical likelihood ratio function with constraint (estimating equation) $\mathbb{E}[e_1(\Sigma)] = 0$:

$$L_0(\Sigma) = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i e_i(\Sigma) = 0, p_i \geq 0\}.$$

Define the empirical likelihood ratio function with constraint (estimating equation) $\mathbb{E}[e_1(\Sigma)] = 0$:

$$L_0(\Sigma) = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i e_i(\Sigma) = 0, p_i \geq 0\}.$$

Under some regularity conditions and H_0, $-2\log L_0(\Sigma_0)$ converges weakly to χ_1^2. This seems good but....

Define the empirical likelihood ratio function with constraint (estimating equation) $\mathbb{E}[e_1(\Sigma)] = 0$:

$$L_0(\Sigma) = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i e_i(\Sigma) = 0, p_i \geq 0\}.$$

Under some regularity conditions and H_0, $-2\log L_0(\Sigma_0)$ converges weakly to χ_1^2. This seems good but....

Shortfall of the test based on L_0.

When $||\Sigma - \Sigma_0||_F^2$ is small, $\mathbb{E}[e_1(\Sigma_0)]$ will be very close to 0 (in a rate of $||\Sigma - \Sigma_0||_F^2$). In this case the test based on L_0 has a poor power. Later we will see this in power analysis.

Secondary constraint.

- We add one more constraint which is easier to break under H_1.

Secondary constraint.

- We add one more constraint which is easier to break under H_1.
- The choice of the second linear constraint can be arbitrary.

Secondary constraint.

- We add one more constraint which is easier to break under H_1.
- The choice of the second linear constraint can be arbitrary.
- With no prior information, the following statistics $v_i(\Sigma)$:

$$v_i(\Sigma) = \mathbf{1}_p^T(Y_i - Y_{N+i} - 2\Sigma)\mathbf{1}_p$$

can be used in a constraint $\mathbb{E}[v_1(\Sigma_0)] = 0$.

We define the empirical likelihood function with two constraints as

$$L_1(\Sigma_0) = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i \begin{pmatrix} e_i(\Sigma_0) \\ v_i(\Sigma_0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, p_i \geq 0\}.$$

We define the empirical likelihood function with two constraints as

$$L_1(\Sigma_0) = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i \begin{pmatrix} e_i(\Sigma_0) \\ v_i(\Sigma_0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, p_i \geq 0\}.$$

Theorem 1

Suppose $e_1(\Sigma_0)$ and $v_1(\Sigma_0)$ satisfy a regularity condition (P). Then under H_0, $-2\log L_1(\Sigma_0)$ converges in distribution to χ_2^2 as $n \to \infty$.

CLT condition (similar to the Lyapunov condition).

(P) We say a statistic T with size n satisfies condition (P) if $\mathbb{E}T^2 > 0$ and for some $\delta > 0$,

$$\frac{\mathbb{E}|T|^{2+\delta}}{(\mathbb{E}T^2)^{1+\delta/2}} = o(n^{\frac{\delta+\min(\delta,2)}{4}}).$$

For example, if $\mathbb{E}(T^4)/(\mathbb{E}(T^2))^2 = o(n)$, then T satisfies (P) with $\delta = 2$.

Remark 1

In order to prove Theorem 1, it is sufficient to prove condition (P) guarantees that the sample $t_i = \left(\frac{e_i(\Sigma_0)}{\sqrt{\mathrm{Var}(e_i(\Sigma_0))}}, \frac{v_i(\Sigma_0)}{\sqrt{\mathrm{Var}(v_i(\Sigma_0))}}\right)^T$ satisfies CLT and t_i^2 satisfies LLN, with a controlled maximum.

Remark 1

In order to prove Theorem 1, it is sufficient to prove condition (P) guarantees that the sample $t_i = \left(\frac{e_i(\Sigma_0)}{\sqrt{\mathrm{Var}(e_i(\Sigma_0))}}, \frac{v_i(\Sigma_0)}{\sqrt{\mathrm{Var}(v_i(\Sigma_0))}}\right)^T$ satisfies CLT and t_i^2 satisfies LLN, with a controlled maximum.

Remark 2

When μ is unknown, just replace μ in Y_i by the sample means and the theorem still holds with one extra moment condition.

Remark 3

In the factor model considered by Chen, Zhang and Zhong (2010), $e_1(\Sigma_0)$ and $v_1(\Sigma_0)$ satisfy (P). With this model, our test allows p to diverge arbitrarily fast or stay finite.

Testing Bandedness

The problem is testing

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau. \quad (3)$$

Here we consider μ is known.

We are interested the information of the black squares in Σ and we will ignore the stars.

$$\begin{pmatrix} * & * & \blacksquare & \blacksquare & \blacksquare \\ * & * & * & \blacksquare & \blacksquare \\ * & * & * & * & \blacksquare \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix}$$

Basic observation.

H_0 is equivalent to the black squares of Σ being 0.

- Define the τ-off-diagonal upper triangular matrix $M^{(\tau)}$ of a matrix M:

$$(M^{(\tau)})_{ij} = \begin{cases} M_{ij} & j \geq i + \tau; \\ 0 & j < i + \tau. \end{cases}$$

- H_0 is equivalent to $\mathrm{tr}((\Sigma^{(\tau)})^T \Sigma^{(\tau)}) = 0$.

- For $i = 1, \ldots, N$, Let

$$e_i' = \operatorname{tr}((Y_i^{(\tau)})^T Y_{N+i}^{(\tau)}),$$

$$v_i' = \mathbf{1}_p^T (Y_i^{(\tau)} + Y_{N+i}^{(\tau)}) \mathbf{1}_p.$$

- We define the empirical likelihood function as

$$L_2 = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i \begin{pmatrix} e_i' \\ v_i' \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, p_i \geq 0\}.$$

Here we omit the Σ_0 in e_i' and v_i'.

Theorem 2

Suppose that e_1' and v_1' satisfy (P). Then under H_0 in (3), $-2\log L_2$ converges in distribution to χ_2^2 as $n \to \infty$.

Theorem 2

Suppose that e_1' and v_1' satisfy (P). Then under H_0 in (3), $-2\log L_2$ converges in distribution to χ_2^2 as $n \to \infty$.

- The method can be used to test some other structures. One interesting application is to test the assumption or estimation of the sparsity.

Remark 4

(1) In the Gaussian model used by Cai and Jiang (2011), e_1' and v_1' satisfy (P) provided that $\tau = o\left(\frac{\sum_{1\leq i,j\leq p}\sigma_{ij}}{(\sum_{1\leq i,j\leq p}|\sigma_{ij}|)^{1/2}}\right)$.

Remark 4

(1) In the Gaussian model used by Cai and Jiang (2011), e_1' and v_1' satisfy (P) provided that $\tau = o\left(\frac{\sum_{1\le i,j\le p}\sigma_{ij}}{(\sum_{1\le i,j\le p}|\sigma_{ij}|)^{1/2}}\right)$.

(2) With moment or boundedness conditions, the Gaussian assumption can be removed.

Power Analysis

Denote $\pi_{11} = \mathbb{E}(e_1(\Sigma)^2)$, $\pi_{22} = \mathbb{E}(v_1(\Sigma)^2)$,

$$\zeta_1 = \mathrm{tr}((\Sigma - \Sigma_0)^2)/\sqrt{\pi_{11}}$$

and

$$\zeta_2 = 2\mathbf{1}_p^T(\Sigma - \Sigma_0)\mathbf{1}_p/\sqrt{\pi_{22}}.$$

- For most models we discuss,

$$\zeta_1 = O\left(\frac{\mathrm{tr}((\Sigma - \Sigma_0)^2)}{\mathrm{tr}(\Sigma^2)}\right)$$

  and

$$\zeta_2 = O\left(\frac{\mathbf{1}_p^T(\Sigma - \Sigma_0)\mathbf{1}_p}{\mathbf{1}_p^T\Sigma^2\mathbf{1}_p}\right).$$

Theorem 3

In addition to the conditions of Theorem 1, if $H_1 : \Sigma \neq \Sigma_0$ *holds, then*

$$P\{-2\log L_1(\Sigma_0) > \xi_{1-\alpha}\} = P\{\chi^2_{2,\nu} > \xi_{1-\alpha}\} + o(1)$$

for any level α *as* $n \to \infty$, *where* $\chi^2_{2,\nu}$ *is a noncentral chi-square distribution with two degrees of freedom and noncentrality parameter* $\nu = N(\zeta_1^2 + \zeta_2^2)$,

Remark 5

- From the above power analysis, the new test rejects the null hypothesis with probability tending to one when $\max(\sqrt{n}\zeta_1, \sqrt{n}|\zeta_2|) \to \infty$.

Remark 5

- From the above power analysis, the new test rejects the null hypothesis with probability tending to one when $\max(\sqrt{n}\zeta_1, \sqrt{n}|\zeta_2|) \to \infty$.
- Note that the test given in Chen, Zhang and Zhong (2010) requires $n\zeta_1 \to \infty$.

Remark 5

- From the above power analysis, the new test rejects the null hypothesis with probability tending to one when $\max(\sqrt{n}\zeta_1, \sqrt{n}|\zeta_2|) \to \infty$.
- Note that the test given in Chen, Zhang and Zhong (2010) requires $n\zeta_1 \to \infty$.
- Our test may have a better power or a worse power in different settings.

Remark 5

- From the above power analysis, the new test rejects the null hypothesis with probability tending to one when $\max(\sqrt{n}\zeta_1, \sqrt{n}|\zeta_2|) \to \infty$.
- Note that the test given in Chen, Zhang and Zhong (2010) requires $n\zeta_1 \to \infty$.
- Our test may have a better power or a worse power in different settings.
- Same results for the test in Theorem 2.

- It is clear that our tests is powerful when $\Sigma - \Sigma_0$ is dense, and not powerful when $\Sigma - \Sigma_0$ is sparse.

- It is clear that our tests is powerful when $\Sigma - \Sigma_0$ is dense, and not powerful when $\Sigma - \Sigma_0$ is sparse.
- With only the first constraint $\mathbb{E}(e_1(\Sigma_0)) = 0$, the test power (requires $\sqrt{n}\zeta_1 \to \infty$) is worse than the test in Chen, Zhang and Zhong (2010).

- It is clear that our tests is powerful when $\Sigma - \Sigma_0$ is dense, and not powerful when $\Sigma - \Sigma_0$ is sparse.
- With only the first constraint $\mathbb{E}(e_1(\Sigma_0)) = 0$, the test power (requires $\sqrt{n}\zeta_1 \to \infty$) is worse than the test in Chen, Zhang and Zhong (2010).
- The test in Cai and Jiang (2011) is good when $\Sigma - \Sigma_0$ is sparse but is powerless when $\Sigma - \Sigma_0$ is dense, since their test power depends on $||\Sigma - \Sigma_0||_{\max}$.

Simulation

Testing covariance matrices

- We assume a dense model and a local alternative.
- We compare with Chen, Zhang and Zhong (2010) for testing covariance matrices and Cai and Jiang (2011) for testing bandedness.
- The ELT has biased size for small n, so we also give a bootstrap calibrated version of ELT.

Table : Testing covariance matrices

(n, p)	$EL(0.05)$ $\delta = 0$	$BCEL(0.05)$ $\delta = 0$	$CZZ(0.05)$ $\delta = 0$	$EL(0.05)$ $\delta = 1$	$BCEL(0.05)$ $\delta = 1$	$CZZ(0.05)$ $\delta = 1$
(50, 25)	0.127	0.054	0.053	0.296	0.118	0.219
(50, 50)	0.148	0.065	0.067	0.324	0.136	0.216
(50, 100)	0.138	0.068	0.038	0.317	0.125	0.212
(50, 200)	0.168	0.081	0.041	0.310	0.113	0.221
(50, 400)	0.151	0.071	0.045	0.342	0.145	0.242
(50, 800)	0.154	0.064	0.041	0.337	0.137	0.219
(200, 25)	0.065	0.048	0.052	0.348	0.305	0.179
(200, 50)	0.058	0.052	0.041	0.336	0.298	0.162
(200, 100)	0.068	0.054	0.059	0.353	0.319	0.179
(200, 200)	0.056	0.051	0.058	0.358	0.322	0.155
(200, 400)	0.069	0.064	0.051	0.374	0.343	0.180
(200, 800)	0.058	0.047	0.050	0.366	0.338	0.182

Table : Testing bandedness

(n, p)	$EL(0.05)$ $\delta = 0$	$BCEL(0.05)$ $\delta = 0$	$CJ(0.05)$ $\delta = 0$	$EL(0.05)$ $\delta = 1$	$BCEL(0.05)$ $\delta = 1$	$CJ(0.05)$ $\delta = 1$
(50, 25)	0.118	0.036	0.015	0.272	0.093	0.017
(50, 50)	0.124	0.049	0.010	0.266	0.097	0.018
(50, 100)	0.126	0.057	0.005	0.268	0.099	0.004
(50, 200)	0.128	0.058	0.003	0.268	0.100	0.001
(50, 400)	0.113	0.053	0.002	0.282	0.121	0.001
(50, 800)	0.128	0.062	0.001	0.281	0.109	0.000
(200, 25)	0.078	0.062	0.019	0.288	0.253	0.034
(200, 50)	0.074	0.059	0.033	0.323	0.286	0.020
(200, 100)	0.057	0.053	0.019	0.332	0.304	0.044
(200, 200)	0.066	0.046	0.024	0.293	0.263	0.032
(200, 400)	0.061	0.052	0.020	0.336	0.304	0.016
(200, 800)	0.053	0.046	0.026	0.317	0.297	0.025

Conclusion.

The new technique

- works for non-parametric models;
- allows arbitrary p; requires only moment conditions;
- avoids to estimate asymptotic variance; the limiting distribution is always χ^2_2;
- can be applied to testing sample mean, two-sample means, and two-sample covariance matrices under the HD framework.

Shortfalls:

- the number of observations is reduced by half;
- the power is good in the dense setting but not in the sparse setting.
- The optimal choice of the second constraint is unknown.

References I

- Cai, 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.
- Chen, S., Zhang, L. and Zhong, P. (2010). Tests for High-Dimensional Covariance Matrices. *J. Amer. Statist. Assoc.* **105**, 810-819.
- Ledoit, O., and Wolf, M. (2002). Some Hypothesis Tests for the Covariance Matrix When the Dimension Is Large Compare to the Sample Size. *Ann. Statist.*, **30**, 1081–1102.

References II

- Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.*, **1**, 700–709.
- Qin, J. and Lawless, J.F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–325.
- Qiu, Y. and Chen, S. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.* **40**, 1285-1314.

Thank you!