Lecture 7

Inference for Covariance Matrix

Readings: Johnson & Wichern 2007, Chapter 4.4;
Rencher 2002, Chapter 7.1, 7.2

DSA 8070 Multivariate Analysis

Whitney Huang
Clemson University

Notes

Agenda

1. Covariance Matrix
2. Statistical Inference for the Covariance Matrix
3. High-Dimensional Covariance Estimation

Notes

Learning Objectives

- **Understand the sample covariance matrix and its distributional properties**: compute the sample covariance matrix and describe the Wishart distribution and its key properties
- **Conduct statistical inference for covariance structures**: perform hypothesis testing and estimation for covariance matrices in both classical and modern settings
- **Address challenges in high-dimensional covariance estimation**: recognize issues in high-dimensional data and apply shrinkage and sparse estimation techniques to obtain reliable covariance estimates

Notes

Review: Interpretation of Covariance

- $\mathrm{Cov}(X, Y) > 0$: variables tend to increase together
- $\mathrm{Cov}(X, Y) < 0$: one increases as the other decreases
- $\mathrm{Cov}(X, Y) = 0$: no linear association

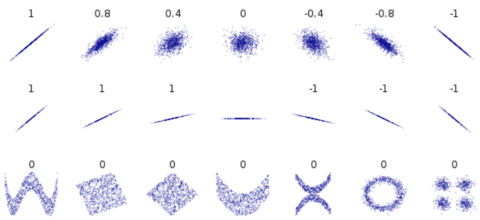


Figure: Image courtesy of Wikipedia at https://en.wikipedia.org/wiki/Correlation_and_dependence

Notes

Population Covariance Matrix and Its Properties

For a random vector $\mathbf{X} = (X_1, \ldots, X_p)^\top$ with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$, the population covariance matrix is:

$$\boldsymbol{\Sigma} = \mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top\right]$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

$\sigma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i) = \boldsymbol{\Sigma}_{ji}$ for $i, j = 1, \cdots, p \Rightarrow \boldsymbol{\Sigma}$ is a $p \times p$ symmetric matrix $\Rightarrow \boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\top$

Notes

Properties of Covariance Matrices (Cont'd)

- Positive semi-definite: A covariance matrix is positive semi-definite if, for any non-zero vector $\mathbf{a} = (a_1, \ldots, a_p)^T$,

  $$\mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a} \geq 0$$

  $\Longrightarrow$ Any linear combination of the variables $\mathbf{X}$ has non-negative variance

- Eigenvalues of $\boldsymbol{\Sigma}$ are nonnegative; eigenvectors give the principal directions

- Determinant $|\boldsymbol{\Sigma}|$ measures the generalized variance

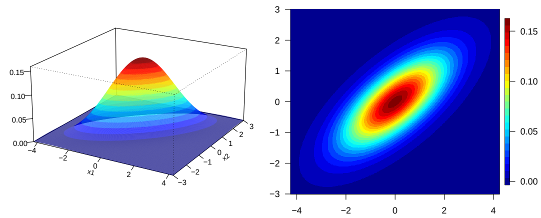
Notes

Geometric Interpretation

- Covariance matrix defines **ellipses of constant Mahalanobis distance**:

$$D^2(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}})$$

- Contours visualize variance and correlation structure



Σ typically unknown and estimated using the sample covariance matrix

Notes

Sample Covariance Matrix (SCM)

Given n observations $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$:

$$\hat{\boldsymbol{\Sigma}}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

- Symmetric and positive semi-definite
- Used in PCA, discriminant analysis, Mahalanobis distance

Notes

Review: Chi-Square and Variance Estimation

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be i.i.d. random variables.

- The sample variance is:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- Then:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\Rightarrow \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\, n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\alpha/2,\, n-1}}$$

Motivation: Just as the *chi-square* distribution governs *sample variance* in the univariate case, the *Wishart* distribution governs the *sample covariance matrix* in the multivariate case.

Notes

Wishart Distribution and Its Properties

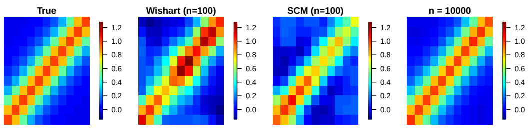
If $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then:

$$\boldsymbol{S} = \sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top \sim W_p(n-1, \boldsymbol{\Sigma})$$

- $\mathbb{E}[\boldsymbol{S}] = (n-1)\boldsymbol{\Sigma}$
- $\mathrm{Var}[\mathbf{S}_{ij}] = (n-1)[\boldsymbol{\Sigma}_{ij}^2 + \boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{jj}]$

$\hat{\boldsymbol{\Sigma}}_n = \frac{\boldsymbol{S}}{n-1} \sim \frac{W_p(n-1,\boldsymbol{\Sigma})}{n-1} \Rightarrow \mathbb{E}[\hat{\boldsymbol{\Sigma}}_n] = \boldsymbol{\Sigma}$

Fundamental for statistical inference on $\boldsymbol{\Sigma}$



Notes

Confidence Intervals for Elements of Σ

- Variance element:

$$\frac{(n-1)\hat{\sigma}_{ii}}{\sigma_{ii}} \sim \chi^2_{n-1}$$

$\Rightarrow$ CI for σ_{ii}

- For covariance or correlation: asymptotic normality or bootstrap [Efron, 1979]
- Bootstrap: resample rows of data matrix, recompute $\hat{\boldsymbol{\Sigma}}$

Notes

Statistical Inference for Σ

- Hypothesis testing: $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$
- Likelihood ratio test statistic:

$$LRT = n\left(\log|\boldsymbol{\Sigma}_0| - \log|\hat{\boldsymbol{\Sigma}}| + \mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{\Sigma}}) - p\right)$$

- Asymptotic distribution: $\chi^2_{p(p+1)/2}$
- Bootstrap methods for confidence intervals

Notes

Example: Hypothesis Test for Σ

Test $H_0 : \mathbf{\Sigma} = I$ with $p = 2$, $n = 50$:

$$LRT = n\left(\log|I| - \log|\hat{\mathbf{\Sigma}}| + \mathrm{tr}(\hat{\mathbf{\Sigma}}) - p\right)$$

Suppose $\hat{\mathbf{\Sigma}} = \begin{bmatrix} 1.5 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$

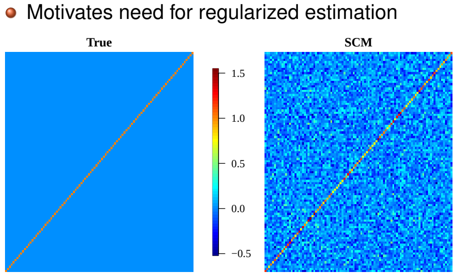
$$LRT \approx 50 \times (0 - \log(1.05) + 2.2 - 2) = 9.5$$

Compare with the χ^2_3 critical value of ≈ 7.81 at $\alpha = 0.05$. Since the test statistic (9.5) exceeds this critical value, we reject the null hypothesis that $\mathbf{\Sigma} = I$.

Notes

SCM Breakdown in High Dimensions

- When $p \approx n$ or $p > n$, $\hat{\mathbf{\Sigma}}$ becomes ill-conditioned or singular
- Eigenvalues may be unstable; inverse is unreliable
- Motivates need for regularized estimation

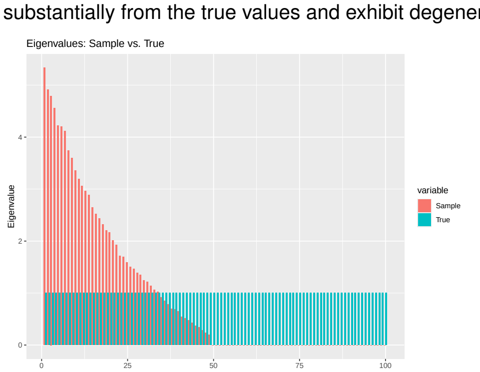


Inference for Covariance Matrix
CLEMSON UNIVERSITY
Covariance Matrix
Statistical Inference for the Covariance Matrix
High-Dimensional Covariance Estimation

Notes

SCM Breakdown in High Dimensions

For the high-dimensional setting with $n = 50 < p = 100$ and $\mathbf{\Sigma} = I$, the estimated eigenvalues deviate substantially from the true values and exhibit degeneracy



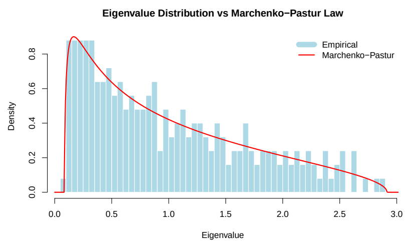
Inference for Covariance Matrix
CLEMSON UNIVERSITY
Covariance Matrix
Statistical Inference for the Covariance Matrix
High-Dimensional Covariance Estimation

Notes

Eigenvalue Distribution in High Dimensions

- When $p, n \to \infty$ with $p/n \to c \in (0, 1)$:

  Empirical eigenvalue distribution $\to$ Marchenko-Pastur law
- Bulk eigenvalues spread, largest eigenvalue deviates
- Explains instability of SCM in high-d settings



Notes

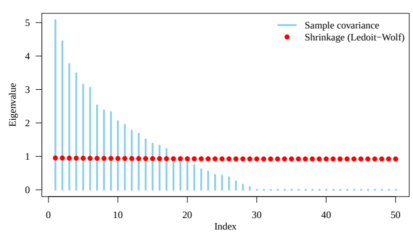
Shrinkage Estimation (Ledoit-Wolf)

$$\hat{\boldsymbol{\Sigma}}_{\text{shrink}} = (1 - \lambda)\hat{\boldsymbol{\Sigma}} + \lambda \boldsymbol{T}$$

- $\boldsymbol{T}$: target (e.g., identity, constant correlation)
- λ: shrinkage intensity (data-driven)
- Reduces variance, improves conditioning

Simulation: $p = 50$, $n = 30$

- Sample eigenvalues: unstable
- Shrinkage pulls toward identity
- Condition number improves



Notes

Sparse Estimation via Graphical Lasso

Estimate inverse covariance (precision) matrix $\Theta = \boldsymbol{\Sigma}^{-1}$:

$$\hat{\Theta} = \arg\max_{\Theta \succ 0} \log\det(\Theta) - \text{tr}(\hat{\Sigma}\Theta) - \rho\|\Theta\|_1$$

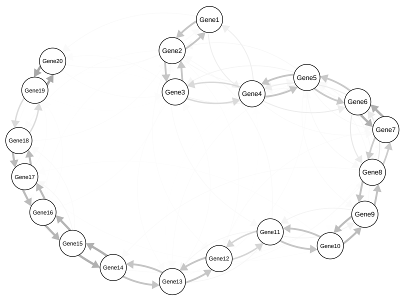
- Promotes sparsity in Θ
- Interpretable as conditional independence graph
- Useful in high-dimensional settings (e.g., genomics, finance)

Notes

Graphical Lasso Application

- Data: AR(1) covariance matrix $\Rightarrow$ sparse precision matrix
- Graphical Lasso estimates the sparse precision matrix
- Nonzero entries $\Rightarrow$ edges in the conditional independence graph

Graphical Lasso Network (nonzeros in precision)



Notes

Summary

- The sample covariance matrix (SCM) is the classical estimator, with properties governed by the Wishart distribution
- Inference tools include confidence intervals, hypothesis tests, and bootstrap methods
- In high-dimensional settings, the SCM becomes unstable or singular, motivating regularization
- Shrinkage estimators (e.g., Ledoit-Wolf) improve conditioning; sparse methods (e.g., graphical lasso) enable interpretable structure learning
- Practical implementations are available in R: `cov`, `cov.shrink`, `glasso`, `boot`

Inference for Covariance Matrix
CLEMSON UNIVERSITY
Covariance Matrix
Statistical Inference for the Covariance Matrix
High-Dimensional Covariance Estimation

Notes

Notes