Topic 1: Deriving likelihood ratio tests for a covariance matrix under the normality assumption

Welcome to Week 4

Dr Pavel Krivitsky gives you a brief overview of topics and concepts we'll be covering in this week.

Transcript

Weekly learning outcomes

- Convert a substantive hypothesis about relationships between variables in a population to a statistical hypothesis about their covariance matrix.
- Derive a test for a given statistical hypothesis about a covariance matrix of a normal population.
- Perform a test for a given statistical hypothesis about a covariance matrix of a normal population and interpret the results in the context of the problem.
- Explain the difference between principal component analysis and factor analysis.
- Interpret results from a factor analysis.

- Identify the optimal number of factors using a variety of techniques.

Topics we will cover are:

- Topic 1: Deriving likelihood ratio tests for a covariance matrix under the normality assumption.
- Topic 2: Factor analysis concepts and interpretation.
- Topic 3: Estimating and testing factor analysis models.
- Topic 4: Overview of structural equation models.

Optional reading

An alternative presentation of the concepts for this week can be found in:

Johnson, R. A., & Wichern, D. (2008). *Applied Multivariate Statistical Analysis* (6th ed.). Pearson Prentice Hall.

- Chapter 9

Muirhead, R. (1982) *Aspects of Multivariate Statistical Theory*. Wiley, New York.

- Chapter 8

All readings are available from the course Leganto reading list. Please keep in mind that you will need to be logged into Moodle to access the Leganto reading list.

Questions about this week's topics?

This week's topics were prepared by Dr P. Krivitsky. If you have any questions or comments, please post them under Discussion or email directly: p.krivitsky@unsw.edu.au

Tests of a covariance matrix

Previously we have developed a number of techniques for decomposing and analysing covariance matrices and their properties. Here, we develop a general family of tests for their structure, which will let you specify almost arbitrary tests for the covariance structure of a multivariate normal population.

Test of $\sum = \sum_0$

We start with this simpler case since ideas are more transparent. The practically more relevant cases are about comparing covariance matrices of two or more multivariate normal populations but the derivations of the latter tests is more subtle. For these we will only formulate the final results.

Assume now that we have the sample $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from a $N_p(\boldsymbol{\mu}, \Sigma)$ distribution and we would like to test $H_0 : \Sigma = \Sigma_0$ against the alternative $H_1 : \Sigma \neq \Sigma_0$. Obviously the problem can be easily transformed into testing $\bar{H}_0 : \Sigma = I_p$ since otherwise we can consider the modified observations $\boldsymbol{Y}_i = \Sigma_0^{-\frac{1}{2}} \boldsymbol{X}_i$ which under H_0 will be multivariate normal with a covariance matrix being equal to I_p. Therefore we can assume that $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ is a sample from a $N_p(\boldsymbol{\mu}, \Sigma)$ and we want to test $H_0 : \Sigma = I_p$ versus $H_1 : \Sigma \neq I_p$.

We will derive the likelihood ratio test for this problem. The likelihood function is

$$\begin{aligned} L(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})} \\ &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2} \operatorname{tr}\left[\Sigma^{-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top\right]} \end{aligned}$$

Under the hypothesis H_0, the maximum of the likelihood function is obtained when $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$. Under the alternative we have to maximise with respect to both $\boldsymbol{\mu}$ and Σ and we know from Section "Maximum Likelihood Estimators" that the maximum of the likelihood function is obtained for $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$ and $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\tau$. Then we obtain easily the likelihood ratio

$$\Lambda = \frac{\max_{\boldsymbol{\mu}} L\left(\boldsymbol{x}; \boldsymbol{\mu}, I_p\right)}{\max_{\boldsymbol{\mu}, \Sigma} L(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)} = \frac{\mathrm{e}^{\left[-\frac{1}{2} \operatorname{tr} \boldsymbol{V}\right]}}{|\boldsymbol{V}|^{-\frac{n}{2}} n^{\frac{np}{2}} \mathrm{e}^{-\frac{np}{2}}}$$

where $\boldsymbol{V} = \Sigma_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top$. Therefore

$$-2 \log \Lambda = np \log n - n \log |\boldsymbol{V}| + \operatorname{tr} \boldsymbol{V} - np, \tag{4.1}$$

and according to the asymptotic theory the quantity in (4.1) is asymptotically distributed as $\chi^2_{p(p+1)/2}$ (the degrees of freedom being the difference of the number of free parameters under the alternative and under the hypothesis). This test would reject H_0 if the value of the $-2 \log \Lambda$ statistic is significantly large.

Sphericity test

It is more realistic to assume that the structure of the covariance matrix is only known up to some constant. Having in mind the discussion in the beginning of the previous slide, we can assume without loss of generality that $H_0 : \Sigma = \sigma^2 I_p$ against a general alternative. This test has the name "sphericity test". The likelihood ratio test can be developed in a manner similar to the previous case and the final result is that

$$-2 \log \Lambda = np \log(n\hat{\sigma}^2) - n \log|\boldsymbol{V}|.$$

Here, $\hat{\sigma}^2 = \frac{1}{np} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\tau (\boldsymbol{x}_i - \bar{\boldsymbol{x}})$. The asymptotic distribution of $np \log(n\hat{\sigma}^2) - n \log|\boldsymbol{V}|$ under the null hypothesis will be again χ^2 but the degrees of freedom are this time $\frac{p(p+1)}{2} - 1 = \frac{(p-1)(p+2)}{2}$. Again, the hypothesis will be rejected for large values.

Note that this notion of sphericity is distinct from that of the repeated measures ANOVA tested by Mauchly's Sphericity Test. (You can learn more about it in Longitudinal Data Analysis.) The specific test discussed here (with a scaled identity matrix as the null hypothesis) is called "spherical" because it implies a multivariate normal distribution that is symmetrical in every possible way---like a sphere.

General situation

Testing equality of covariance matrices from k different multivariate normal populations $N_p(\boldsymbol{\mu}_i, \Sigma_i), i = 1, 2, \ldots, k$ is a very important problem especially in discriminant analysis and multivariate analysis of variance. Let,

k be the number of populations;
p the dimension of vector;
n the total sample size $n = n_1 + n_2 + \ldots n_k$,
n_i being the sample size for each population.

The analysis of deviance test statistic that results is

$$-2\log\frac{\prod_{i=1}^{k}\left|\hat{\boldsymbol{\Sigma}}_i\right|^{\frac{n_i}{2}}}{\left|\hat{\boldsymbol{\Sigma}}_{\text{pooled}}\right|^{\frac{n}{2}}},$$

with $\hat{\boldsymbol{\Sigma}}_i$ the MLE sample variance (with denominator n_i as opposed to $n_i - 1$) of population i, and $\hat{\boldsymbol{\Sigma}}_{\text{pooled}} = \frac{1}{n}\sum_{i=1}^{k} n_i\hat{\boldsymbol{\Sigma}}_i$, asymptotically distributed $\chi^2_{(k-1)p(p+1)/2}$.

It has been noticed that this test has the defect that it is (asymptotically) biased: that is, the probability of rejecting H_0 when H_0 is false can be smaller than the probability of rejecting H_0 when H_0 is true (i.e., it may happen that in some points of the parameter space the probability of a correct decision is smaller than the probability for a wrong decision). Hence it is desirable to modify it to make it asymptotically unbiased.

Further let $N = n - k$ and $N_i = n_i - 1$. Under the null hypothesis of equality of all k covariance matrices, it holds:

$$-2\rho\log\frac{\prod_{i=1}^{k}|\boldsymbol{S}_i|^{\frac{N_i}{2}}}{|\boldsymbol{S}_{\text{pooled}}|^{\frac{N}{2}}} \tag{4.2}$$

for $\rho = 1 - \left[\left(\sum_{i=1}^{k}\frac{1}{N_i}\right) - \frac{1}{N}\right]\frac{2p^2+3p-1}{6(p+1)(k-1)}$, $\boldsymbol{S}_i$ the sample variance (with $n-1$ denominator) of population i, and $S_{\text{pooled}} = \frac{1}{N}\sum_{i=1}^{k} N_k\boldsymbol{S}_i$, is asymptotically distributed as $\chi^2_{\frac{1}{2}(k-1)p(p+1)}$. Large values of the statistic are significant and lead to the rejection of the hypothesis about equality of the k covariance matrices.

In the following, we will avoid the subtle details and refer to Chapter 8 of the monograph:

- Muirhead, R. (1982) *Aspects of Multivariate Statistical Theory*. Wiley, New York.

The *modified* LR is achieved by replacing n_i and n by N_i and N (that is, by the correct degrees of freedom). We note that indeed $\rho = 1 - [(\Sigma_{i=1}^k \frac{1}{N_i}) - \frac{1}{N}]\frac{2p^2+3p-1}{6(p+1)(k-1)}$ is close to 1 anyway if all sample sizes n_i were very large. Finally, the scaling of the test statistic by $\rho = 1 - [(\sum_{i=1}^k \frac{1}{N_i}) - \frac{1}{N}]\frac{2p^2+3p-1}{6(p+1)(k-1)}$ that is made in (4.2) serves to improve the quality of the asymptotic approximation of the statistic by the limiting $\chi^2_{\frac{1}{2}(k-1)p(p+1)}$ distribution. Such (asymptotically negligible) scalar transformations of the LR statistic that yield improved test statistic with a chi-squared null distribution of order $O(1/n)$ instead of the ordinary $O(1)$ for the standard LR, are known in the literature under the common name **Bartlett corrections**. Thus, (4.2) is a Bartlett corrected version of the modified LR statistic. This is the statistic most commonly implemented in software.

Check your understanding

Complete the below exercises to check your understanding of concepts presented so far.

1.

a) Follow the discussion about the sphericity test. Argue that if $\hat{\lambda}_i, i = 1, 2, \ldots, p$ denote the eigenvalues of the empirical covariance matrix S then

$$-2 \log \Lambda = np \log \frac{\text{arithm. mean } \hat{\lambda}_i}{\text{geom. mean } \hat{\lambda}_i}$$

Of course, the above statistic is asymptotically $\chi^2_{(p+2)(p-1)/2}$ distributed under H_0 since it only represents the sphericity test in a different form.

b) Show that the likelihood ratio test of

$$H_0 : \Sigma \text{ is the diagonal matrix}$$

rejects H_0 when $-n \log |R|$ is larger than $\chi^2_{1-\alpha, p(p-1)/2}$. (Here R is the empirical correlation matrix, p is the dimension of the multivariate normal and n is the sample size.)

Challenge: Testing Variances

If you choose to complete this task in your own RStudio, upload the following file:

Cov_Test_Example.challenge.Rmd

Click on the 'Cov_Test_Example.challenge.Rmd' in the 'Files' section to begin. Enter your response to the tasks in the 'Enter your code here' section.

The solution will also be available here on Friday of this week by clicking on the 'Solution' tab in the top right corner.

The output of the RMD file is also displayed below: