Test for Covariances

Max Turgeon

STAT 7200–Multivariate Statistics

Objectives

- Review general theory of likelihood ratio tests
- Tests for structured covariance matrices
- Test for equality of multiple covariance matrices

Likelihood ratio tests i

- We will build our tests for covariances using likelihood ratios.
    - Therefore, we quickly review the asymptotic theory for regular models.
- Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample from a density p_θ with parameter $\theta \in \mathbb{R}^d$.
- We are interested in the following hypotheses:

$$H_0 : \theta \in \Theta_0, \qquad H_1 : \theta \in \Theta_1,$$

where $\Theta_i \subseteq \mathbb{R}^d$.

Likelihood ratio tests ii

- Let $L(\theta) = \prod_{i=1}^n p_\theta(\mathbf{Y}_i)$ be the likelihood, and define the likelihood ratio

$$\Lambda = \frac{\max_{\theta\in\Theta_0} L(\theta)}{\max_{\theta\in\Theta_0\cup\Theta_1} L(\theta)}.$$

- **Recall**: we reject the null hypothesis H_0 for small values of Λ.

Likelihood ratio tests iii

Theorem (Van der Wandt, Chapter 16)
Assume Θ_0, Θ_1 are *locally linear*. Under regularity conditions on p_θ, we have

$$-2\log\Lambda \to \chi^2(k),$$

where k is the difference in the number of free parameters between the null model Θ_0 and the unrestricted model $\Theta_0 \cup \Theta_1$.

- Therefore, in practice, we need to count the number of free parameters in each model and hope the sample size n is large enough.

Tests for structured covariance matrices i

- We are going to look at several tests for structured covariance matrix.
- Throughtout, we assume $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim N_p(\mu, \Sigma)$ with Σ positive definite.
    - Like other exponential families, the multivariate normal distribution satisfies the regularity conditions of the theorem above.
    - Being positive definite implies that the unrestricted parameter space is *locally linear*, i.e. we are staying away from the boundary where Σ is singular.

Tests for structured covariance matrices ii

- A few important observations about the unrestricted model:
  - The number of free parameters is equal to the number of entries on and above the diagonal of Σ, which is $p(p+1)/2$.
  - The sample mean $\bar{\mathbf{Y}}$ maximises the likelihood **independently of the structure of Σ**.
  - The maximised likelihood for the unrestricted model is given by

$$L(\hat{\mathbf{Y}}, \hat{\Sigma}) = \frac{\exp(-np/2)}{(2\pi)^{np/2}|\hat{\Sigma}|^{n/2}}.$$

Specified covariance structure i

- We will start with the simplest hypothesis test:

$$H_0 : \Sigma_0.$$

- Note that there is no free parameter in the null model.
- Write $V = n\hat{\Sigma}$. Recall that we have

$$L(\hat{\mathbf{Y}}, \Sigma) = (2\pi)^{-np/2}|\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}V)\right).$$

Specified covariance structure ii

- Therefore, the likelihood ratio is given by

$$\begin{aligned}\Lambda &= \frac{(2\pi)^{-np/2}|\Sigma_0|^{-n/2}\exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma_0^{-1}V)\right)}{\exp(-np/2)(2\pi)^{-np/2}|\hat{\Sigma}|^{-n/2}}\\ &= \frac{|\Sigma_0|^{-n/2}\exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma_0^{-1}V)\right)}{\exp(-np/2)|n^{-1}V|^{-n/2}}\\ &= \left(\frac{e}{n}\right)^{np/2}|\Sigma_0^{-1}V|^{n/2}\exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma_0^{-1}V)\right).\end{aligned}$$

- In particular, if $\Sigma_0 = I_p$, we get

$$\Lambda = \left(\frac{e}{n}\right)^{np/2}|V|^{n/2}\exp\left(-\frac{1}{2}\mathrm{tr}(V)\right).$$

Example i

```r
library(tidyverse)
# Winnipeg avg temperature
url <- paste0("https://maxturgeon.ca/w20-stat7200/",
              "winnipeg_temp.csv")
dataset <- read.csv(url)
dataset[1:3,1:3]
```

```
##   temp_2010 temp_2011  temp_2012
## 1 -25.57500 -16.25417  -6.379167
## 2 -26.06250 -18.39583 -12.925000
## 3 -20.56667 -19.45833  -5.791667
```

Example ii

```r
n <- nrow(dataset)
p <- ncol(dataset)

V <- (n - 1)*cov(dataset)
```

```r
# Diag = 14^2
# Corr = 0.8
Sigma0 <- diag(0.8, nrow = p)
diag(Sigma0) <- 1
Sigma0 <- 14^2*Sigma0
Sigma0_invXV <- solve(Sigma0, V)
```

Example iii

```
lrt <- 0.5*n*p*(1 - log(n))
lrt <- lrt + 0.5*n*log(det(Sigma0_invXV))
lrt <- lrt - 0.5*sum(diag(Sigma0_invXV))
lrt <- -2*lrt
```

```
df <- choose(p, 2)
c(lrt, qchisq(0.95, df))
```

```
## [1] 5631.63409   61.65623
```

- *Sphericity* means the different components of $\mathbf{Y}$ are **uncorrelated** and have the **same variance**.
  - In other words, we are looking at the following null hypothesis:

$$H_0 : \Sigma = \sigma^2 I_p, \qquad \sigma^2 > 0.$$

  - Note that there is one free parameter.
- We have

$$\begin{aligned} L(\hat{\mathbf{Y}}, \sigma^2 I_p) &= (2\pi)^{-np/2} |\sigma^2 I_p|^{-n/2} \exp\left(-\frac{1}{2}\mathrm{tr}((\sigma^2 I_p)^{-1} V)\right) \\ &= (2\pi\sigma^2)^{-np/2} \exp\left(-\frac{1}{2\sigma^2}\mathrm{tr}(V)\right). \end{aligned}$$

Test for sphericity ii

- Taking the derivative of the logarithm and setting it equal to zero, we find that $L(\hat{\mathbf{Y}}, \sigma^2 I_p)$ is maximised when

$$\widehat{\sigma^2} = \frac{\mathrm{tr} V}{np}.$$

- We then get

$$\begin{aligned} L(\hat{\mathbf{Y}}, \widehat{\sigma^2} I_p) &= (2\pi\widehat{\sigma^2})^{-np/2} \exp\left(-\frac{1}{2\widehat{\sigma^2}}\mathrm{tr}(V)\right) \\ &= (2\pi)^{-np/2} \left(\frac{\mathrm{tr} V}{np}\right)^{-np/2} \exp\left(-\frac{np}{2}\right). \end{aligned}$$

- Therefore, we have

$$
\begin{aligned}
\Lambda &= \frac{(2\pi)^{-np/2}\left(\frac{\mathrm{tr}V}{np}\right)^{-np/2}\exp\left(-\frac{np}{2}\right)}{\exp(-np/2)(2\pi)^{-np/2}|\hat{\Sigma}|^{-n/2}} \\
&= \frac{\left(\frac{\mathrm{tr}V}{np}\right)^{-np/2}}{|n^{-1}V|^{-n/2}} \\
&= \left(\frac{|V|}{(\mathrm{tr}V/p)^p}\right)^{n/2}.
\end{aligned}
$$

Example (cont'd) i

```r
lrt <- -2*0.5*n*(log(det(V)) - p*log(mean(diag(V))))
df <- choose(p, 2) - 1

c(lrt, qchisq(0.95, df))
```

```
## [1] 5630.79458   60.48089
```

- Recall that we have

$$\Lambda = \left(\frac{|V|}{(\mathrm{tr}V/p)^p}\right)^{n/2}.$$

- We can rewrite this as follows: let $l_1 \geq \cdots \geq l_p$ be the eigenvalues of V. We have

$$\begin{aligned}\Lambda^{2/n} &= \frac{|V|}{(\mathrm{tr}V/p)^p}\\ &= \frac{\prod_{j=1}^p l_j}{(\frac{1}{p}\sum_{j=1}^p l_j)^p}\\ &= \left(\frac{\prod_{j=1}^p l_j^{1/p}}{\frac{1}{p}\sum_{j=1}^p l_j}\right)^p.\end{aligned}$$

Test for sphericity (cont'd) ii

- In other words, the modified LRT $\tilde{\Lambda} = \Lambda^{2/n}$ is the ratio of the geometric to the arithmetic mean of the eigenvalues of V (all raised to the power p).
- A result of Srivastava and Khatri gives the *exact* distribution of $\tilde{\Lambda}$:

$$\tilde{\Lambda} = \prod_{j=1}^{p-1} \mathcal{B}\left(\frac{1}{2}(n-j-1), j\left(\frac{1}{2}+\frac{1}{p}\right)\right).$$

Example (cont'd) i

```r
B <- 1000
df1 <- 0.5*(n - seq_len(p-1) - 1)
df2 <- seq_len(p-1)*(0.5 + 1/p)

# Critical values
dist <- replicate(B, {
  prod(rbeta(p-1, df1, df2))
  })
```

Example (cont'd) ii

```r
# Test statistic
decomp <- eigen(V, symmetric = TRUE, only.values = TRUE)
ar_mean <- mean(decomp$values)
geo_mean <- exp(mean(log(decomp$values)))

lrt_mod <- (geo_mean/ar_mean)^p

c(lrt_mod, quantile(dist, 0.95))
```

```
##                         95%
## 1.181561e-07 8.977070e-01
```

Test for independence i

- Decompose $\mathbf{Y}_i$ into k blocks:

$$\mathbf{Y}_i = (\mathbf{Y}_{1i}, \dots, \mathbf{Y}_{ki}),$$

where $\mathbf{Y}_{1i} \sim N_{p_k}(\mu_k, \Sigma_{kk})$ and $\sum_{j=1}^k p_j = p$.

- This induces a decomposition on Σ and V:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} \\ \vdots & \ddots & \vdots \\ \Sigma_{k1} & \cdots & \Sigma_{kk} \end{pmatrix}, \qquad V = \begin{pmatrix} V_{11} & \cdots & V_{1k} \\ \vdots & \ddots & \vdots \\ V_{k1} & \cdots & V_{kk} \end{pmatrix}.$$

Test for independence ii

- We are interested in testing for independence between the different blocks $\mathbf{Y}_{1i}, \dots, \mathbf{Y}_{ki}$. This equivalent to

$$H_0 : \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_{kk} \end{pmatrix}.$$

    - Note that there are $\sum_{j=1}^{k} p_j(p_j + 1)/2$ free parameters.
- Under the null hypothesis, the likelihood can be decomposed into k likelihoods that can be maximised independently.

Test for independence iii

- This gives us

$$\max L(\hat{\mathbf{Y}}, \Sigma) = \prod_{j=1}^{k} \frac{\exp(-np_j/2)}{(2\pi)^{np_j/2}|\widehat{\Sigma_{jj}}|^{n/2}}$$
$$= \frac{\exp(-np/2)}{(2\pi)^{np/2}\prod_{j=1}^{k}|\widehat{\Sigma_{jj}}|^{n/2}}.$$

- Putting this together, we conclude that

$$\Lambda = \left(\frac{|V|}{\prod_{j=1}^{k}|V_{jj}|}\right)^{n/2}.$$