Large-sample Likelihood Ratio Tests[1]

STA431 Spring 2017

[1] See last slide for copyright information.

Model and null hypothesis

$$D_1, \ldots, D_n \overset{i.i.d.}{\sim} P_\theta,\ \theta \in \Theta,$$
$$H_0 : \theta \in \Theta_0 \text{ v.s. } H_A : \theta \in \Theta \cap \Theta_0^c,$$

The data have likelihood function

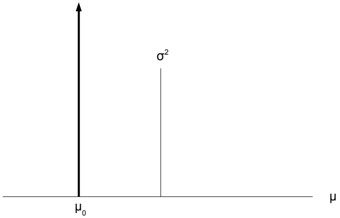
$$L(\theta) = \prod_{i=1}^{n} f(d_i; \theta),$$

where $f(d_i; \theta)$ is the density or probability mass function evaluated at d_i.

Example

$D_1, \ldots, D_n \overset{i.i.d.}{\sim} P_\theta$, $\theta \in \Theta$,
$H_0 : \theta \in \Theta_0$ v.s. $H_A : \theta \in \Theta \cap \Theta_0^c$,

$$D_1, \ldots, D_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$$
$$H_0 : \mu = \mu_0 \text{ v.s. } H_A : \mu \neq \mu_0$$
$$\Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}$$



Likelihood ratio

- Let $\widehat{\theta}$ denote the usual Maximum Likelihood Estimate (MLE).
- That is, $\widehat{\theta}$ is the parameter value for which the likelihood function is greatest, over all $\theta \in \Theta$.
- Let $\widehat{\theta}_0$ denote the *restricted* MLE. The restricted MLE is the parameter value for which the likelihood function is greatest, over all $\theta \in \Theta_0$.
- $\widehat{\theta}_0$ is *restricted* by the null hypothesis $H_0 : \theta \in \Theta_0$.
- $L(\widehat{\theta}_0) \leq L(\widehat{\theta})$, so that
- The *likelihood ratio* $\lambda = \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \leq 1$.
- If the overall MLE $\widehat{\theta}$ is located in Θ_0, the likelihood ratio will equal one. In this case, there is no reason to reject the null hypothesis.

The test statistic

It's like comparing a full to a reduced model

- We know $\lambda = \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \leq 1$.
- If it's a *lot* less than one, then the data are a lot less likely to have been observed under the null hypothesis than under the alternative hypothesis, and the null hypothesis is questionable.
- If λ is small (close to zero), then $\ln(\lambda)$ is a large negative number, and $-2\ln\lambda$ is a large positive number.

$$G^2 = -2\ln\left(\frac{\max_{\theta\in\Theta_0} L(\theta)}{\max_{\theta\in\Theta} L(\theta)}\right)$$

Difference between two -2 log likelihoods

$$
\begin{aligned}
G^2 &= -2\ln\left(\frac{\max_{\theta\in\Theta_0} L(\theta)}{\max_{\theta\in\Theta} L(\theta)}\right) \\
&= -2\ln\left(\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right) \\
&= -2\ln L(\widehat{\theta}_0) - [-2\ln L(\widehat{\theta})] \\
&= -2\ell(\widehat{\theta}_0) - [-2\ell(\widehat{\theta})].
\end{aligned}
$$

- Could minimize $-2\ell(\theta)$ twice, first over all $\theta \in \Theta$, and then over all $\theta \in \Theta_0$.
- The test statistic is the difference between the two minimum values.

Distribution of the test statistic under H_0

Approximate large sample distribution (Wilks, 1936)

Suppose the null hypothesis is that certain *linear combinations* of parameter values are equal to specified constants. Then if H_0 is true,

$$G^2 = -2 \ln \left(\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \right)$$

has an approximate chi-squared distribution for large n.

- Degrees of freedom equals number of (non-redundant, linearly independent) equalities specified by H_0.
- So count the equals signs.
- Reject when G^2 is large.

Example

Suppose $\boldsymbol{\theta} = (\theta_1, \ldots \theta_7)$, with

$$H_0: \ \theta_1 = \theta_2, \theta_6 = \theta_7, \frac{1}{3}\left(\theta_1 + \theta_2 + \theta_3\right) = \frac{1}{3}\left(\theta_4 + \theta_5 + \theta_6\right)$$

Count the equals signs or write the null hypothesis in matrix form as $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$.

$$\left(\begin{array}{rrrrrrr} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 0 \end{array}\right) \left(\begin{array}{c} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right)$$

Rows are linearly independent, so df=number of rows = 3.

Bernoulli example

- $Y_1, \ldots, Y_n \stackrel{i.i.d.}{\sim} B(1, \theta)$
- $H_0 : \theta = \theta_0$
- $\Theta = (0, 1)$
- $\Theta_0 = \{\theta_0\}$
- $L(\theta) = \theta^{\sum_{i=1}^{n} y_i}(1-\theta)^{n-\sum_{i=1}^{n} y_i}$
- $\widehat{\theta} = \overline{y}$
- $\widehat{\theta}_0 = \theta_0$

Likelihood ratio test statistic

$L(\theta) = \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i}$

$$
\begin{aligned}
G^2 &= -2\ln\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \\
&= -2\ln\frac{\theta_0^{n\overline{y}}(1-\theta_0)^{n(1-\overline{y})}}{\overline{y}^{n\overline{y}}(1-\overline{y})^{n(1-\overline{y})}} \\
&= -2\ln\left(\frac{\theta_0^{\overline{y}}(1-\theta_0)^{(1-\overline{y})}}{\overline{y}^{\overline{y}}(1-\overline{y})^{(1-\overline{y})}}\right)^n \\
&= 2n\ln\left(\frac{\theta_0^{\overline{y}}(1-\theta_0)^{(1-\overline{y})}}{\overline{y}^{\overline{y}}(1-\overline{y})^{(1-\overline{y})}}\right)^{-1} \\
&= 2n\ln\frac{\overline{y}^{\overline{y}}(1-\overline{y})^{(1-\overline{y})}}{\theta_0^{\overline{y}}(1-\theta_0)^{(1-\overline{y})}}
\end{aligned}
$$

Continued

$$
\begin{aligned}
G^2 &= 2n \ln \frac{\overline{y}^{\overline{y}}(1-\overline{y})^{(1-\overline{y})}}{\theta_0^{\overline{y}}(1-\theta_0)^{(1-\overline{y})}} \\
&= 2n \left(\ln \left(\frac{\overline{y}}{\theta_0} \right)^{\overline{y}} + \ln \left(\frac{1-\overline{y}}{1-\theta_0} \right)^{(1-\overline{y})} \right) \\
&= 2n \left(\overline{y} \ln \left(\frac{\overline{y}}{\theta_0} \right) + (1-\overline{y}) \ln \left(\frac{1-\overline{y}}{1-\theta_0} \right) \right)
\end{aligned}
$$

Coffee taste test

$n = 100,\ \theta_0 = 0.50,\ \overline{y} = 0.60$

$$\begin{aligned} G^2 &= 2n\left(\overline{y}\ln\left(\frac{\overline{y}}{\theta_0}\right) + (1-\overline{y})\ln\left(\frac{1-\overline{y}}{1-\theta_0}\right)\right) \\ &= 200\left(0.60\ln\left(\frac{0.60}{0.50}\right) + 0.40\ln\left(\frac{0.40}{0.50}\right)\right) \\ &= 4.027 \end{aligned}$$

$df = 1$, critical value $1.96^2 = 3.84$. Conclude (barely) that the new coffee blend is preferred over the old.

Univariate normal example

- $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$
- $H_0 : \mu = \mu_0$
- $\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$
- $\Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}$
- $L(\theta) = (\sigma^2)^{-n/2}(2\pi)^{-n/2} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\}$
- $\widehat{\theta} = \left(\overline{Y}, \widehat{\sigma}^2\right)$, where

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

- $\widehat{\theta}_0 = (\widehat{\mu}_0, \widehat{\sigma}_0^2) = \ldots$

Restricted MLE

For $H_0 : \mu = \mu_0$

Definitely have $\widehat{\mu}_0 = \mu_0$.

Recall that setting derivaties to zero, we obtained

$$\mu = \overline{y} \text{ and } \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2, \text{ so}$$

$$\begin{aligned}
\widehat{\mu}_0 &= \mu_0 \\
\widehat{\sigma}_0^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_0)^2
\end{aligned}$$

Likelihood ratio test statistic $G^2 = -2\ln\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}$

Have $L(\theta) = (\sigma^2)^{-n/2}(2\pi)^{-n/2}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2\}$, so

$$
\begin{aligned}
L(\widehat{\theta}) &= (\widehat{\sigma}^2)^{-n/2}(2\pi)^{-n/2}\exp\{-\frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^n (y_i-\overline{y})^2\} \\
&= (\widehat{\sigma}^2)^{-n/2}(2\pi)^{-n/2}\exp\left\{-\frac{\sum_{i=1}^n (y_i-\overline{y})^2}{2\frac{1}{n}\sum_{i=1}^n (y_i-\overline{y})^2}\right\} \\
&= (\widehat{\sigma}^2)^{-n/2}(2\pi)^{-n/2}e^{-n/2}
\end{aligned}
$$

Likelihood at restricted MLE

$L(\theta) = (\sigma^2)^{-n/2}(2\pi)^{-n/2}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\mu)^2\}$

$$\begin{aligned} L(\widehat{\theta}_0) &= (\widehat{\sigma}_0^2)^{-n/2}(2\pi)^{-n/2}\exp\{-\frac{1}{2\widehat{\sigma}_0^2}\sum_{i=1}^n(y_i-\mu_0)^2\} \\ &= (\widehat{\sigma}_0^2)^{-n/2}(2\pi)^{-n/2}\exp\left\{-\frac{\sum_{i=1}^n(y_i-\mu_0)^2}{2\frac{1}{n}\sum_{i=1}^n(y_i-\mu_0)^2}\right\} \\ &= (\widehat{\sigma}_0^2)^{-n/2}(2\pi)^{-n/2}e^{-n/2} \end{aligned}$$

Test statistic

$$
\begin{aligned}
G^2 &= -2\ln\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \\
&= -2\ln\frac{(\widehat{\sigma}_0^2)^{-n/2}(2\pi)^{-n/2}e^{-n/2}}{(\widehat{\sigma}^2)^{-n/2}(2\pi)^{-n/2}e^{-n/2}} \\
&= -2\ln\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{-n/2} \\
&= n\ln\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right) \\
&= n\ln\left(\frac{\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_0)^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})^2}\right) \\
&= n\ln\left(\frac{\sum_{i=1}^n(Y_i-\mu_0)^2}{\sum_{i=1}^n(Y_i-\overline{Y})^2}\right)
\end{aligned}
$$

Multivariate normal likelihood

SAS `proc calis` default

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}}} \exp\left\{ -\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) \right\} \\
&= \cdots \\
&= |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{y}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{y}} - \boldsymbol{\mu}) \right\},
\end{aligned}
$$

where $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^\top$ is the sample variance-covariance matrix.

Sample variance-covariance matrix

$$\mathbf{Y}_i = \left(\begin{array}{c} Y_{i,1} \\ \vdots \\ Y_{i,p} \end{array} \right) \qquad \overline{\mathbf{Y}} = \left(\begin{array}{c} \overline{Y}_1 \\ \vdots \\ \overline{Y}_p \end{array} \right)$$

$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^n (\mathbf{Y}_i - \overline{\mathbf{Y}})(\mathbf{Y}_i - \overline{\mathbf{Y}})^\top$ is a $p \times p$ matrix with (j,k) element

$$\frac{1}{n}\sum_{i=1}^n (Y_{i,j} - \overline{Y}_j)(Y_{i,k} - \overline{Y}_k)$$

This is a sample variance or covariance.

Multivariate normal likelihood at the MLE

This will be in the denominator of the likelihood ratio test.

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= |\boldsymbol{\Sigma}|^{-\frac{n}{2}} (2\pi)^{-\frac{np}{2}} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{y}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{y}} - \boldsymbol{\mu}) \right\} \\ L(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) &= |\widehat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} (2\pi)^{-\frac{np}{2}} e^{-\frac{np}{2}} \end{aligned}$$

Example: Test whether a set of normal random variables are independent

Equivalent to zero covariance

- $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \stackrel{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $H_0 : \sigma_{ij} = 0$ for $i \neq j$.
- Equivalent to independence for this multivariate normal model.
- Use $G^2 = -2 \ln \left(\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \right)$.
- $df = \binom{p}{2}$
- Have $L(\widehat{\theta})$.
- Need $L(\widehat{\theta}_0)$.

Getting the restricted MLE

For the multivariate normal, zero covariance is equivalent to independence, so under H_0,

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} f(\mathbf{y}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \prod_{i=1}^{n} \left(\prod_{j=1}^{p} f(y_{ij} | \mu_j, \sigma_j^2) \right) \\
&= \prod_{j=1}^{p} \left(\prod_{i=1}^{n} f(y_{ij} | \mu_j, \sigma_j^2) \right)
\end{aligned}
$$

Take logs and start differentiating

$$
\begin{aligned}
L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma_0}) &= \prod_{j=1}^{p} \left(\prod_{i=1}^{n} f(y_{ij}|\mu_j, \sigma_j^2) \right) \\
\ell(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}) &= \sum_{j=1}^{p} \ln \left(\prod_{i=1}^{n} f(y_{ij}|\mu_j, \sigma_j^2) \right)
\end{aligned}
$$

It's just j univariate problems, which we have already done.

Likelihood at the restricted MLE

$$
\begin{aligned}
L(\widehat{\boldsymbol{\mu}}_0, \widehat{\boldsymbol{\Sigma}}_0) &= \prod_{j=1}^{p} \left((\widehat{\sigma}_j^2)^{-n/2} (2\pi)^{-n/2} \exp\{ -\frac{1}{2\widehat{\sigma}_j^2} \sum_{i=1}^{n} (y_{ij} - \overline{y}_j)^2 \} \right) \\
&= \prod_{j=1}^{p} \left((\widehat{\sigma}_j^2)^{-n/2} (2\pi)^{-n/2} e^{-n/2} \right) \\
&= \left(\prod_{j=1}^{p} \widehat{\sigma}_j^2 \right)^{-\frac{n}{2}} (2\pi)^{-\frac{np}{2}} e^{-\frac{np}{2}},
\end{aligned}
$$

where $\widehat{\sigma}_j^2$ is a diagonal element of $\widehat{\boldsymbol{\Sigma}}$.

Test statistic

$$\begin{aligned}
G^2 &= -2\ln\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \\
&= -2\ln\frac{\left(\prod_{j=1}^p \widehat{\sigma}_j^2\right)^{-\frac{n}{2}}(2\pi)^{-\frac{np}{2}}e^{-\frac{np}{2}}}{|\widehat{\mathbf{\Sigma}}|^{-\frac{n}{2}}(2\pi)^{-\frac{np}{2}}e^{-\frac{np}{2}}} \\
&= -2\ln\left(\frac{\prod_{j=1}^p \widehat{\sigma}_j^2}{|\widehat{\mathbf{\Sigma}}|}\right)^{-\frac{n}{2}} \\
&= n\ln\left(\frac{\prod_{j=1}^p \widehat{\sigma}_j^2}{|\widehat{\mathbf{\Sigma}}|}\right) \\
&= n\left(\sum_{j=1}^p \ln\widehat{\sigma}_j^2 - \ln|\widehat{\mathbf{\Sigma}}|\right)
\end{aligned}$$

Cars: Weight, length and fuel consumption

$G^2 = n\left(\sum_{j=1}^{p} \ln \widehat{\sigma}_j^2 - \ln|\widehat{\boldsymbol{\Sigma}}|\right)$

```
> kars = read.table("mcars4.data.txt"); attach(kars)
> n = length(lper100k); SigmaHat = var(cbind(weight, length, lper100k))
> SigmaHat = SigmaHat * (n-1)/n # Make it the MLE
> SigmaHat

               weight      length   lper100k
weight    129698.9859 186.4174680 984.089620
length       186.4175   0.2993794   1.472152
lper100k     984.0896   1.4721524  10.729116

> Gsq = n * ( sum(log(diag(SigmaHat))) - log(det(SigmaHat)) )
> Gsq  # df=3

[1] 347.7159
```

Numerical maximum likelihood and testing

For the multivariate normal

- Often an explicit formula for $\widehat{\theta}_0$ is out of the question.
- Maximize the log likelihood numerically.
- Equivalently, minimize $-2\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Equivalently, minimize $-2\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ plus a constant.
- Choose the constant well, and minimize

$$-2\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) - (-2\ln L(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}))$$

  over $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta_0$.
- The value of this function at the stopping place is the likelihood ratio test statistic.

Simplifying . . .

$$
\begin{aligned}
-2\ln\frac{L(\boldsymbol{\mu},\boldsymbol{\Sigma})}{L(\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\Sigma}})} &= -2\ln\frac{|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp -\frac{n}{2}\left\{tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1})+(\overline{\mathbf{y}}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu})\right\}}{|\widehat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}}e^{-\frac{np}{2}}} \\
&= -2\ln\left(|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp -\frac{n}{2}\left\{tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1})+(\overline{\mathbf{y}}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu})\right\}|\widehat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}e^{\frac{np}{2}}\right) \\
&= -2\ln\left(|\boldsymbol{\Sigma}|\exp\left\{tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1})+(\overline{\mathbf{y}}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu})\right\}|\widehat{\boldsymbol{\Sigma}}|^{-1}e^{-p}\right)^{-\frac{n}{2}} \\
&= n\left(tr\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right)-p+\ln|\boldsymbol{\Sigma}|-\ln|\widehat{\boldsymbol{\Sigma}}|+(\overline{\mathbf{y}}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu})\right)
\end{aligned}
$$

- To avoid numerical problems in minimizing the function, drop the n.
- The result is the “discrepancy function” F_{ML} on page 1247 of the Version 9.3 `proc calis` manual.
- The discrepancy function is also called the “objective function” in other parts of the manual and in the Results file.

Later in the course

Recalling $F_{ML} = tr\left(\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\right) - p + \ln|\mathbf{\Sigma}| - \ln|\widehat{\mathbf{\Sigma}}| + (\overline{\mathbf{y}} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\overline{\mathbf{y}} - \boldsymbol{\mu})$

- Model is based on systems of equations with unknown parameters $\boldsymbol{\theta} \in \Theta$.
- $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\mathbf{\Sigma}(\boldsymbol{\theta})$ are the mean and covariance matrix of the *observable* variables.
- We will give up on the parameters that appear only in $\boldsymbol{\mu}$. Estimate $\boldsymbol{\mu}$ with $\overline{\mathbf{y}}$ and it disappears from F_{ML}.
- Calculate the covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$ from the model equations.
- Minimize the objective function

$$F_{ML}(\boldsymbol{\theta}) = tr\left(\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\right) - p + \ln|\mathbf{\Sigma}(\boldsymbol{\theta})| - \ln|\widehat{\mathbf{\Sigma}}|$$

  over all $\boldsymbol{\theta} \in \Theta$.
- The result is $\widehat{\boldsymbol{\theta}}$. Can also obtain $\widehat{\boldsymbol{\theta}}_0$ by minimizing over Θ_0.

Copyright Information