Honours Multivariate Analysis

Lecture 4 - Maximum Likelihood Estimation and Testing for Normality

Stefan S. Britz

Department of Statistical Sciences
University of Cape Town

Course Outline

1. Introduction, Examples of Multivariate Data †
2. TOOLS
    1. Visualization and Summary Statistics †
    2. Singular Value Decomposition, Eigenvalue Decomposition and Spectral Decomposition revisited †
    3. The Multivariate Normal Distribution †
    4. Multivariate Maximum Likelihood Estimation †
    5. Multivariate Inference †
3. EXPLORATORY ANALYSIS
    1. Principal Component Analysis §
    2. Factor Analysis §
    3. Correspondence Analysis §
4. CONFIRMATORY ANALYSIS
    1. For grouped Multivariate Data:
        1. Manova †
        2. Discriminant Analysis §
    2. Regression
        1. Multivariate Regression §
        2. Canonical Correlation Analysis §

† Mr Stefan Britz

§ Mr Miguel Rodo

The MVN Density and Likelihood

Consider a random sample, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from the multivariate normal population $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with the $\boldsymbol{X}_j$'s mutually independent.

The joint density of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ is given by

$$
\begin{aligned}
f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) &= \prod_{j=1}^{n} \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu})\right] \\
&= \left(\frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}}\right)^n \exp\left[-\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu})\right]
\end{aligned}
$$

The MVN Density and Likelihood

This joint density, written as a function of the variables, regards the parameters as fixed, albeit unknown, constants.

However, when observations are made, i.e. we are given values for $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, then we can consider this expression to be a function of the parameters, referred to as the *likelihood*.

We are therefore trying to ascertain how likely specific values of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (viewed as variable) are, given our fixed observations.

$$\therefore L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left(\frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \right)^n \exp \left[-\frac{1}{2} \sum_{j=1}^{n} (\boldsymbol{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu}) \right]$$

The MVN Density and Likelihood

For ease of notation to follow, define

$$\boldsymbol{A} = (n-1)\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X} - n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}' = \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$$

The MVN Density and Likelihood

For ease of notation to follow, define

$$\boldsymbol{A} = (n-1)\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X} - n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}' = \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$$

The summation in the exponent of the likelihood function can then be written as

$$\sum_{j=1}^{n}(\boldsymbol{x}_j - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu}) = tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right] + n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})$$

Make sure you go through the steps in the notes!

The MVN Density and Likelihood

For ease of notation to follow, define

$$\boldsymbol{A} = (n-1)\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X} - n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}' = \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$$

The summation in the exponent of the likelihood function can then be written as

$$\sum_{j=1}^{n}(\boldsymbol{x}_j - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu}) = tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right] + n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})$$

Make sure you go through the steps in the notes!

We can now write the joint likelihood function as

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp\left[-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right] - \frac{n}{2}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})\right]$$

MLE of $\boldsymbol{\mu}$

We consider $\boldsymbol{\Sigma}$ as fixed and maximise

$$\begin{aligned}\log[L(\boldsymbol{\mu},\boldsymbol{\Sigma})] &= l(\boldsymbol{\mu},\boldsymbol{\Sigma}) \\ &= -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}] - \frac{n}{2}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})\end{aligned}$$

with respect to $\boldsymbol{\mu}$.

By inspection:

$$\begin{aligned}l &= -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}] - \frac{n}{2}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) \\ &\leq -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}]\end{aligned}$$

with equality when $\frac{n}{2}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) = 0$, i.e., when

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}.$$

MLE of $\boldsymbol{\mu}$

We consider $\boldsymbol{\Sigma}$ as fixed and maximise

$$\begin{aligned}\log[L(\boldsymbol{\mu}, \boldsymbol{\Sigma})] &= l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}] - \frac{n}{2}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})\end{aligned}$$

with respect to $\boldsymbol{\mu}$.

By differentiation:

$$\begin{aligned}\frac{\partial l}{\partial \boldsymbol{\mu}} &= -\frac{n}{2}2\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(-1) \qquad \text{(See Theorem A.1)} \\ n\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \hat{\boldsymbol{\mu}}) &= \mathbf{0} \\ \hat{\boldsymbol{\mu}} &= \bar{\boldsymbol{x}}\end{aligned}$$

MLE of $\boldsymbol{\mu}$

Showing that $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$ is indeed a maximum and not a minimum or saddle point:

$$\frac{\partial^2 l}{\partial \boldsymbol{\mu}\boldsymbol{\mu}'} = \frac{\partial}{\partial \boldsymbol{\mu}'}\{n\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}} - n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\} = -n\boldsymbol{\Sigma}^{-1}$$

which is negative definite. Note that $\boldsymbol{\Sigma}^{-1}$ is positive definite since $\boldsymbol{\Sigma}$ is positive definite and symmetric.

MLE of Σ

Let $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$. We first need to find the stationary point of $l(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$ through differentiation with respect to $\boldsymbol{\Sigma}^{-1}$, then show that the stationary point is a maximum. See Theorems A.2 & A.3.

$$
\begin{aligned}
l(\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}, \boldsymbol{\Sigma}) &= -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}] \\
&= -\frac{np}{2}\log(2\pi) + \frac{n}{2}\log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}] \\
\frac{\partial l}{\partial \boldsymbol{\Sigma}^{-1}} &= \frac{n}{2}\left[\left(\boldsymbol{\Sigma}^{-1}\right)^{-1}\right]' - \frac{1}{2}\boldsymbol{A}'
\end{aligned}
$$

$$
\begin{aligned}
\frac{n}{2}\hat{\boldsymbol{\Sigma}} - \frac{1}{2}\boldsymbol{A} &= \boldsymbol{0} \\
n\hat{\boldsymbol{\Sigma}} &= \boldsymbol{A} \\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{n}\boldsymbol{A} = \frac{n-1}{n}\boldsymbol{S}
\end{aligned}
$$

MLE of Σ

To show that this stationary point is a maximum, taking the second order derivative with respect to $\boldsymbol{\Sigma}^{-1}$ is complicated. Therefore, the likelihood is reparameterised. See the course notes for the gory details.

When substituting $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ into the likelihood, we end up with the following expression for the maximised likelihood:

$$\begin{aligned}
L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) &= (2\pi)^{-\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} tr\left[\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}\right] - \frac{n}{2}(\bar{\boldsymbol{x}} - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Sigma}}^{-1}(\bar{\boldsymbol{x}} - \hat{\boldsymbol{\mu}})\right] \\
&= (2\pi)^{-\frac{np}{2}} \left|\frac{1}{n}\boldsymbol{A}\right|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} tr\left[n\boldsymbol{A}^{-1}\boldsymbol{A}\right]\right] \\
&= (2\pi)^{-\frac{np}{2}} \left|\frac{1}{n}\boldsymbol{A}\right|^{-\frac{n}{2}} e^{-\frac{np}{2}}
\end{aligned}$$

Homework exercise 4.1

Johnson & Wichern exercise 4.18

Find the maximum likelihood estimates of $\boldsymbol{\mu}_{2\times 1}$ and $\boldsymbol{\Sigma}_{2\times 2}$ based on the random sample

$$\begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

from a bivariate normal population.

Sampling Distribution of $\bar{\boldsymbol{X}}$

Reminder of the univariate case:
If $X_1, X_2, \ldots, X_n$ are i.i.d random observation from $X \sim N(\mu, \sigma^2)$, then

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and

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

Sampling Distribution of $\bar{\boldsymbol{X}}$

Reminder of the univariate case:
If $X_1, X_2, \ldots, X_n$ are i.i.d random observation from $X \sim N(\mu, \sigma^2)$, then

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and

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

The multivariate case:
Now suppose we observe $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$, which are i.i.d random vectors drawn from $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
For the sample mean we have

$$\bar{\boldsymbol{X}} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$$

Sampling Distribution of S

For the sample variance, note that the univariate case can be expressed as

$$\begin{aligned}(n-1)s^2 &\sim \sigma^2\chi^2_{n-1} \\ &= \sigma^2\left(Z_1^2 + \ldots + Z_{n-1}^2\right) \\ &= (\sigma Z_1)^2 + \ldots + (\sigma Z_{n-1})^2\end{aligned}$$

with each $\sigma Z_i = X_i - \mu \sim N(0, \sigma^2)$.

This form is suitably generalized to the basic sampling distribution of the covariance matrix, namely the **Wishart distribution**.

This distribution is the multivariate analogue of the χ^2-distribution.

Sampling Distribution of S

If we define $\boldsymbol{Y}_i = \boldsymbol{X}_i - \boldsymbol{\mu} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ and let $\boldsymbol{Y}$ represent the $p \times n$ matrix constructed from the n independent observations of $\boldsymbol{X}_i$, then

$$\boldsymbol{Y}\boldsymbol{Y}' = \sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})' \sim W_p(n, \boldsymbol{\Sigma})$$

We say that $\boldsymbol{Y}\boldsymbol{Y}'$ is p-dimensional Wishart distributed with n degrees of freedom.

Sampling Distribution of S

If we define $\boldsymbol{Y}_i = \boldsymbol{X}_i - \boldsymbol{\mu} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ and let $\boldsymbol{Y}$ represent the $p \times n$ matrix constructed from the n independent observations of $\boldsymbol{X}_i$, then

$$\boldsymbol{Y}\boldsymbol{Y}' = \sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})' \sim W_p(n, \boldsymbol{\Sigma})$$

We say that $\boldsymbol{Y}\boldsymbol{Y}'$ is p-dimensional Wishart distributed with n degrees of freedom.

Similar to the univariate case, when $\bar{\boldsymbol{X}}$ is substituted for $\boldsymbol{\mu}$, the distribution remains Wishart, but with one less degree of freedom:

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

$$\boldsymbol{S} \sim W_p\left(n-1, \frac{1}{n-1}\boldsymbol{\Sigma}\right)$$

Properties of the Wishart distribution

Given $\boldsymbol{A} \sim W_p(n, \boldsymbol{\Sigma})$, we note the following properties:

- The Wishart distribution is a generalization of the χ^2-distribution, where $W_1(n, \sigma^2) = \sigma^2 \chi_n^2$.

Properties of the Wishart distribution

Given $\boldsymbol{A} \sim W_p(n, \boldsymbol{\Sigma})$, we note the following properties:

- The Wishart distribution is a generalization of the χ^2-distribution, where $W_1(n, \sigma^2) = \sigma^2 \chi^2_n$.
- If $\boldsymbol{c}_{p \times 1}$ is a non-zero constant vector, then $\dfrac{\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}} \sim \chi^2_n$.

Properties of the Wishart distribution

Given $\boldsymbol{A} \sim W_p(n, \boldsymbol{\Sigma})$, we note the following properties:

- The Wishart distribution is a generalization of the χ^2-distribution, where $W_1(n, \sigma^2) = \sigma^2 \chi^2_n$.
- If $\boldsymbol{c}_{p\times 1}$ is a non-zero constant vector, then $\dfrac{\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}} \sim \chi^2_n$.
- The density only exists if $n > p$.

Properties of the Wishart distribution

Given $\boldsymbol{A} \sim W_p(n, \boldsymbol{\Sigma})$, we note the following properties:

- The Wishart distribution is a generalization of the χ^2-distribution, where $W_1(n, \sigma^2) = \sigma^2 \chi^2_n$.
- If $\boldsymbol{c}_{p \times 1}$ is a non-zero constant vector, then $\dfrac{\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}} \sim \chi^2_n$.
- The density only exists if $n > p$.
- The density itself, although not of particular use to us, is given by

$$f(\boldsymbol{A}) = \frac{|\boldsymbol{A}|^{\frac{n-p-2}{2}} \exp\left[-\frac{1}{2} tr(\boldsymbol{A}\boldsymbol{\Sigma}^{-1})\right]}{2^{\frac{p(n-1)}{2}} \pi^{\frac{p(p-1)}{4}} |\boldsymbol{\Sigma}|^{\frac{n-1}{2}} \prod_{i=1}^{p} \Gamma\left(\frac{n-i}{2}\right)}$$

  for $\boldsymbol{A}$ positive definite.

Properties of the Wishart distribution

Given $\boldsymbol{A} \sim W_p(n, \boldsymbol{\Sigma})$, we note the following properties:

- The Wishart distribution is a generalization of the χ^2-distribution, where $W_1(n, \sigma^2) = \sigma^2 \chi^2_n$.

- If $\boldsymbol{c}_{p\times 1}$ is a non-zero constant vector, then $\dfrac{\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}} \sim \chi^2_n$.

- The density only exists if $n > p$.

- The density itself, although not of particular use to us, is given by

$$f(\boldsymbol{A}) = \frac{|\boldsymbol{A}|^{\frac{n-p-2}{2}} \exp\left[-\frac{1}{2} tr(\boldsymbol{A}\boldsymbol{\Sigma}^{-1})\right]}{2^{\frac{p(n-1)}{2}} \pi^{\frac{p(p-1)}{4}} |\boldsymbol{\Sigma}|^{\frac{n-1}{2}} \prod_{i=1}^{p} \Gamma\left(\frac{n-i}{2}\right)}$$

  for $\boldsymbol{A}$ positive definite.

- $tr(\boldsymbol{A}) \sim \chi^2_{np}$.

Properties of the Wishart distribution

Given $\boldsymbol{A} \sim W_p(n, \boldsymbol{\Sigma})$, we note the following properties:

- The Wishart distribution is a generalization of the χ^2-distribution, where $W_1(n, \sigma^2) = \sigma^2 \chi^2_n$.
- If $\boldsymbol{c}_{p\times 1}$ is a non-zero constant vector, then $\dfrac{\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}} \sim \chi^2_n$.
- The density only exists if $n > p$.
- The density itself, although not of particular use to us, is given by

$$f(\boldsymbol{A}) = \frac{|\boldsymbol{A}|^{\frac{n-p-2}{2}} \exp\left[-\frac{1}{2} tr(\boldsymbol{A}\boldsymbol{\Sigma}^{-1})\right]}{2^{\frac{p(n-1)}{2}} \pi^{\frac{p(p-1)}{4}} |\boldsymbol{\Sigma}|^{\frac{n-1}{2}} \prod_{i=1}^{p} \Gamma\left(\frac{n-i}{2}\right)}$$

  for $\boldsymbol{A}$ positive definite.
- $tr(\boldsymbol{A}) \sim \chi^2_{np}$.
- $\mathrm{E}(\boldsymbol{A}) = n\boldsymbol{\Sigma}$.

Summary of Sampling distribution results

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample from $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

1. $\bar{\boldsymbol{X}} \sim N_p\left(\boldsymbol{\mu}, \dfrac{1}{n}\boldsymbol{\Sigma}\right)$
2. $(n-1)\boldsymbol{S} \sim W_p(n-1, \boldsymbol{\Sigma})$
3. $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$ are independent.
4. $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$ are *sufficient statistics* for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Large Sample behaviour of $\bar{X}$ and S

If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are independent observations from some population with mean $\boldsymbol{\mu}$ and finite, nonsingular covariance $\boldsymbol{\Sigma}$, then

$$\bar{\boldsymbol{X}} \dot{\sim} N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right) \text{ or, equivalently, } \sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \dot{\sim} N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

and

$$n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})' \boldsymbol{S}^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \dot{\sim} \chi_p^2$$

for large $n - p$.

Homework exercise 4.2

Johnson & Wichern exercises 4.19 & 4.21 (combined)

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{60}$ be a random sample of size $n = 60$ from an $N_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Specify each of the distributions completely (indicate if the distribution is approximate):

1. $\bar{\boldsymbol{X}}$ and $\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$
2. $(\boldsymbol{X}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu})$
3. $(n-1)\boldsymbol{S}$
4. $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$
5. $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$

Testing the Assumption of Normality

Up until now we have been **assuming** MVN observations.

To determine whether the observations $\boldsymbol{X}_j$ appear to violate the assumption that they jointly came from a multivariate normal population, we address the following questions:

1. Do the marginal distributions of the elements of $\boldsymbol{X}$ appear to be normal?
2. Do the scatter plots of the pairs of observations on different characteristics give the elliptical appearance expected from normal populations?
3. Are there any "wild" observations that should be checked?

Univariate assessment – QQ-plots

QQ-plots, or quantile-quantile plots:

- Quick, visual way of assessing how closely the distribution of observed data matches some theoretical distribution
- Plot the sample quantile versus the quantile one would expect to observe if the observations were actually normally distributed

Univariate assessment – QQ-plots

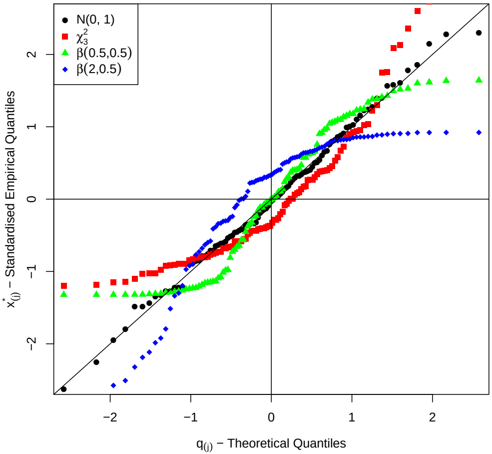
QQ-plots, or quantile-quantile plots:

- If we order the n observations such that $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$, then if the $x_{(j)}$ are distinct, exactly j observations will be less than or equal to $x_{(j)}$
- The proportion of the sample at or to the left of $x_{(j)}$, i.e. $\frac{j}{n}$, is often approximated by $\frac{j-\frac{1}{2}}{n}$ for analytical convenience
- The quantiles for a standard normal distribution are defined as those values $q_{(j)}$ such that

$$\Pr[Z \leq q_{(j)}] = \int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \, dz = p_{(j)} = \frac{j-\frac{1}{2}}{n}$$

Univariate assessment – QQ-plots

Under the assumption of normality, a plot of $(q_{(j)}, x^*_{(j)})$ for all j should be a straight line through the origin, where $x^*_{(j)}$ denotes the standardised values of $\boldsymbol{x}$



Univariate assessment – correlation test

To formally test the linearity, we can calculate the correlation coefficient for the QQ plot, defined as

$$r_q = \frac{\sum_{j=1}^{n}(x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^{n}(x_{(j)} - \bar{x})^2}\sqrt{\sum_{j=1}^{n}(q_{(j)} - \bar{q})^2}}$$

and compare it to a table of critical values, given in Appendix B.

Univariate assessment – correlation test

To formally test the linearity, we can calculate the correlation coefficient for the QQ plot, defined as

$$r_q = \frac{\sum_{j=1}^{n}(x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^{n}(x_{(j)} - \bar{x})^2}\sqrt{\sum_{j=1}^{n}(q_{(j)} - \bar{q})^2}}$$

and compare it to a table of critical values, given in Appendix B.

Many other formal tests for univariate normality exist, for example:

- Shapiro-Wilk
- Anderson-Darling
- Jarque-Bera
- Lilliefors test
- etc.

Multivariate assessment

We can now compare the contours of constant density from observed data with the ellipsoid as defined in chapter 3.3, where a set of bivariate outcomes $\boldsymbol{x}$ such that

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \leq \chi_p^2(\alpha)$$

has probability $1-\alpha$.

- Typically we calculate the above for $\alpha = 0.5$ and calculate the proportion of points for which the squared distance is less than $\chi_p^2(0.5)$.
- If this deviates from 50%, it is evidence against the assumption of normality.

Multivariate assessment

We can now compare the contours of constant density from observed data with the ellipsoid as defined in chapter 3.3, where a set of bivariate outcomes $\boldsymbol{x}$ such that

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \leq \chi_p^2(\alpha)$$

has probability $1-\alpha$.

- Typically we calculate the above for $\alpha = 0.5$ and calculate the proportion of points for which the squared distance is less than $\chi_p^2(0.5)$.
- If this deviates from 50%, it is evidence against the assumption of normality.

Again, many other formal test can be applied:

- Mardia
- Henze-Zirkler
- Royston
- etc.

Multivariate assessment

We can also construct a chi-square plot based on the assumption that, given underlying normality, the squared distances, $d_j^2 = (\boldsymbol{x}_j - \bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})$ should behave like chi-square random variables.

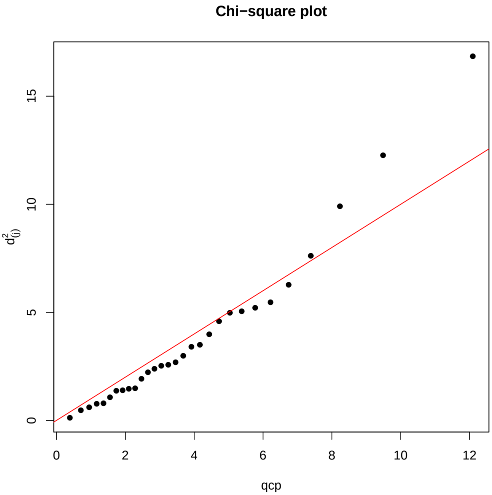
To construct these plots,

1. Calculate $d_j^2 = (\boldsymbol{x}_j - \bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})$.
2. Order the squared distances from smallest to largest as $d_{(1)}^2 \leq d_{(2)}^2 \leq \ldots \leq d_{(n)}^2$.
3. Graph the pairs $\left(q_{c,p}\left(\frac{j-\frac{1}{2}}{n}\right), d_{(j)}^2\right)$ where $q_{c,p}\left(\frac{j-\frac{1}{2}}{n}\right) = \chi_p^2\left(\frac{n-j+\frac{1}{2}}{n}\right)$.

If the variables are multivariate normal distributed, the plot should be a straight line through the origin.

Example

Example 4.14 on page 186 of Johnson & Wichern. Solution in `Lecture4.R`.



Caution!

It is important to note the following crucial drawbacks of all measures of fit:

With small samples, only severe deviations will indicate lack of fit, whilst very large samples will invariably produce statistically significant lack of fit.

Homework exercise 4.3

Johnson & Wichern exercises 4.28 & 4.29 (combined)

Consider the air pollution data

1. Construct a QQ-plot for the solar radiation measurements and carry out a test for normality based on the correlation coefficient r_q. You are not prescribed to use a specific α-value; what can you report based on Table B.1?

Now examine the pairs $X_5 = \text{NO}_2$ and $X_6 = \text{O}_3$ for bivariate normality.

2. Calculate the distances $(\boldsymbol{x}_j - \bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}),\ j = 1, 2, \ldots, 42$, where $\boldsymbol{x}_j' = [x_{j5}, x_{j6}]$.

3. Determine the proportion of observations $\boldsymbol{x}_j' = [x_{j5}, x_{j6}],\ j = 1, 2, \ldots, 42$ falling within the approximate 50% probability contour of a bivariate normal distribution.

4. Construct a chi-square plot for the ordered distances in Part 3.