THE UNIVERSITY OF CHICAGO
Booth School of Business
Business 41912, Spring Quarter 2020, Mr. Ruey S. Tsay

Lecture: Inference about sample mean

Reading material: In view of interest in big data and common use of many hypothesis testings, it is important to understand the **false discovery rate** (FDR). Read the paper *Controlling the False Discovery Rate: a Pratical and Powerful Approach to Multiple Testing* by Y. Benjamin ans Y. Hochberg, JRSSB (1995), Vol. 57, No. 1, pp. 289-300.
This is one of the most cited papers with approximately 63,326 citations as of now. You can easily download the paper from Internet.

Key concepts of this lecture note:

1. Hotelling's T^2 test
2. Likelihood ratio test
3. Various confidence regions
4. Applications
5. Missing values
6. Impact of serial correlations

1 Hypothesis test of a mean vector

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a random sample from a p-dimensional normal population with mean $\boldsymbol{\mu}$ and positive-definite variance-covariance matrix $\boldsymbol{\Sigma}$. Consider the testing problem: $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ versus $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_o$, where $\boldsymbol{\mu}_o$ is a known vector.
This is a generalization of the one-sample t test of the univariate case. When $p = 1$, the test statistic is

$$t = \frac{\bar{x} - \mu_o}{s/\sqrt{n}}, \quad \text{with} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

The statistic follows a Student-t distribution with $n-1$ degrees of freedom. One rejects H_o if the p-value of t is less than the type-I error denoted by α. For generalization, we rewrite the test as

$$t^2 = (\bar{x} - \mu_o)\left(\frac{s^2}{n}\right)^{-1}(\bar{x} - \mu_o).$$

One rejects H_o if and only if $t^2 \geq t^2_{n-1}(\alpha/2)$, the upper $100(\alpha/2)$ percentile of t-distribution with $n-1$ degrees of freedom. A natural generalization of this test statistic is

$$T^2 = (\bar{\boldsymbol{x}} - \boldsymbol{\mu}_o)' \left(\frac{\boldsymbol{S}}{n}\right)^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}_o), \tag{1}$$

where $\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$ is the sample covariance matrix.
This is the Hotelling's T^2 statistic. It is distributed as $\frac{(n-1)p}{n-p}F_{p,n-p}$, where $F_{u,v}$ denotes the F-distribution with degrees of freedom u and v. Recall that $[t_{n-1}(\alpha/2)]^2 = F_{1,n-1}(\alpha)$, where $F_{u,v}(\alpha)$ is the upper 100α percentile of the F-distribution with u and v degrees of freedom. Thus, when $p = 1$, T^2 reduces to the usual one-sample t statistic.

Example 1. Consider the monthly log returns of Boeing (BA), Abbott Labs (ABT), Motorola (MOT) and General Motors (GM) from January 1998 to December 2007. The log returns are in percentages. Let $\boldsymbol{\mu}_o = \mathbf{0}$. Test the null hypothesis that the mean vector of the log returns is zero.
Answer: Except for three possible outlying observations, the chi-square QQ-plot indicates that the normal assumption is reasonable. See the QQ-plot in Figure **??**.

```
> setwd("C:/Users/rst/teaching/ama")
> x=read.table("m-ba4c9807.txt")
> head(x)
          V1        V2         V3         V4
1 -4.7401928 -2.718622  8.1791025  4.0690786
2 18.1062431 13.282227  5.4949253 -6.8393408
3 -1.7376094 -3.995883  0.6660768  9.0107579
4 -0.5550375 -4.037210 -2.5381400 -8.5890188
5  7.1587370 -4.436577  1.4425451 -5.0585124
6 -7.3038544 -6.908645 10.0123124 -0.6009018

> source("ama.R")   # so the command qqchi2 can be used.
> qqchi2(x)
[1] "correlation coefficient:"
[1] 0.9608707

## Summary statistics
> xave=colMeans(x)
> xave
         V1          V2          V3          V4
-0.26294605  0.61075048  0.67154781  0.02679363
> sqrt(apply(x,2,var))
       V1        V2        V3        V4
11.269948  8.962623  6.383600 11.077578
```

```
> S=var(x)
> S
           V1        V2        V3         V4
V1 127.011719 28.776911  8.046536  50.919252
V2  28.776911 80.328617  8.385634  13.216236
V3   8.046536  8.385634 40.750353  -8.710395
V4  50.919252 13.216236 -8.710395 122.712745

> Si=solve(S)  <== Find the inverse of S.

> T2=nrow(x)*t(xave)%*%Si%*%xave
> T2
         [,1]
[1,] 2.107425

### Results of marginal tests
> t.test(x[,1])

        One Sample t-test

data:  x[, 1]
t = -0.2556, df = 119, p-value = 0.7987
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -2.300074  1.774182
sample estimates:
 mean of x
-0.2629461

> t.test(x[,2])

        One Sample t-test

data:  x[, 2]
t = 0.7465, df = 119, p-value = 0.4568
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -1.009311  2.230812
sample estimates:
mean of x
0.6107505
```

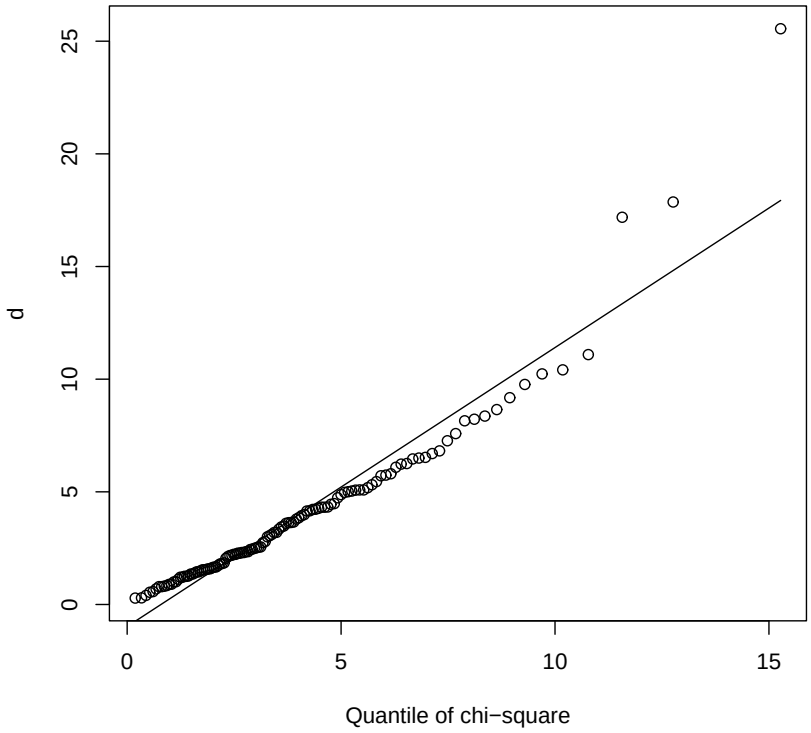


Figure 1: Chi-square QQ plot for monthly log returns of 4 stocks

```
> t.test(x[,3])

        One Sample t-test

data:  x[, 3]
t = 1.1524, df = 119, p-value = 0.2515
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.4823362  1.8254318
sample estimates:
mean of x
0.6715478
```

Example 2. Consider the Perspiration data of 20 observations in Table 5.1 of the textbook. There are three measurements, namely sweat rate(X_1), Sodium(X_2), and Potassium(X_3).

Test $H_o : \boldsymbol{\mu} = (4, 50, 10)'$ versus $H_a : \boldsymbol{\mu} \neq (4, 50, 10)'$.

```
> z=read.table("T5-1.DAT")
> dim(z)
[1] 20  3
> colnames(z) <- c("rate","Sodium","Potassium")
> qqchi2(z)
[1] "correlation coefficient:"  <== Normality seems reasonable.
[1] 0.975713

> Hotelling(z,c(4,50,10)) <== Hotelling is a command in ‘‘ama.R’’
                     [,1]
Hoteliing-T2 9.73877286
p.value      0.06492834  <== Cannot reject the null hypo at the 5% level.
```

Remark: Let $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{d}$, where $\boldsymbol{C}$ is a $p \times p$ non-singular matrix and $\boldsymbol{d}$ is a p-dimensional vector. Then, testing $\boldsymbol{H}_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ of $\boldsymbol{x}$ is equivalent to testing $H_o : \boldsymbol{\mu}_y = \boldsymbol{C}\boldsymbol{\mu}_o + \boldsymbol{d}$ of $\boldsymbol{y}$. It turns out that, as expected, the Hotelling T^2 statistic is identical.
Proof: $\bar{\boldsymbol{y}} = \boldsymbol{C}\bar{\boldsymbol{x}} + \boldsymbol{d}$ and $\boldsymbol{S}_y = \boldsymbol{C}\boldsymbol{S}\boldsymbol{C}'$.

$$\begin{aligned} T^2 &= n(\bar{\boldsymbol{y}} - \boldsymbol{\mu}_{yo})'\boldsymbol{S}_y^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\mu}_{yo}) \\ &= n(\boldsymbol{C}\bar{\boldsymbol{x}} + \boldsymbol{d} - \boldsymbol{C}\boldsymbol{\mu}_o - \boldsymbol{d})'(\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}')^{-1}(\boldsymbol{C}\bar{\boldsymbol{x}} + \boldsymbol{d} - \boldsymbol{C}\boldsymbol{\mu}_o - \boldsymbol{d}) \\ &= n[\boldsymbol{C}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_o)]'\boldsymbol{C}'^{-1}\boldsymbol{S}^{-1}\boldsymbol{C}^{-1}[\boldsymbol{C}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_o)] \\ &= n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_o)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_o). \end{aligned}$$

2 Likelihood ratio test

Key result: Hotelling T^2 test statistic is equivalent to likelihood ratio test statistic.

General theory: Let $\boldsymbol{\theta}$ be the parameter vector of a likelihood function $L(\boldsymbol{\theta})$ with observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Suppose that the parameter space is $\boldsymbol{\Theta}$. Consider the null hypothesis $\boldsymbol{\theta} \in \boldsymbol{\Theta}_o$ versus $H_a : \boldsymbol{\theta} \ni \boldsymbol{\Theta}_o$, where $\boldsymbol{\Theta}_o$ is a subspace of $\boldsymbol{\Theta}$. The likelihood ratio statistic is

$$\Lambda = \frac{\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_o} L(\boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta})}. \tag{2}$$

One rejects H_o if $\Lambda < c$, where c is some critical value depending on the type-I error. Intuitively, one rejects H_o if the maximized likelihood function over the subspace $\boldsymbol{\Theta}_o$ is much smaller than that over the parameter space $\boldsymbol{\Theta}$, indicating that it is unlikely for $\boldsymbol{\theta}$ to be in $\boldsymbol{\Theta}_o$. Specifically, under the hull hypothesis H_o, the likelihood ratio statistic

$$-2\ln(\Lambda) \sim \chi^2_{v - v_o}, \quad \text{as} \quad n \to \infty,$$

where $v = \dim(\boldsymbol{\Theta})$ and $v_o = \dim(\boldsymbol{\Theta}_o)$.

Remark: $\boldsymbol{\Theta}_o \subset \boldsymbol{\Theta}$ indicates that the null model is nested in the alternative model. For non-nested models, likelihood ratio test does not apply. [See some recent papers on the generalized likelihood ratio test.]

Normal case: For multivariate normal distribution, the likelihood ratio statistic is relatively simple because the limiting distribution of Λ is available. Specifically, consider $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ versus $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_o$. Here $\boldsymbol{\theta} = (\boldsymbol{\mu}', \sigma_{11}, \sigma_{21}, \sigma_{22}, \ldots, \sigma_{p1}, \ldots, \sigma_{pp})'$ is of dimension $p+p(p+1)/2$. Under H_o, $\boldsymbol{\mu}$ is fixed at $\boldsymbol{\mu}_o$ so that $\boldsymbol{\Theta}_o$ consists of the space for the elements of $\boldsymbol{\Sigma}$. In this case,

$$\Lambda = \frac{\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_o, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}.$$

Recall that, under $\boldsymbol{\Theta}$,

$$\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = L(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = \frac{1}{(2\pi)^{np/2}|\widehat{\Sigma}|^{n/2}} e^{-np/2},$$

where $\widehat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i$ and $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$. Under $\boldsymbol{\Theta}_o$, the likelihood function becomes

$$L(\boldsymbol{\mu}_o, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}_o)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_o)\right].$$

By the same method as before, we can show that

$$\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_o, \boldsymbol{\Sigma}) = L(\boldsymbol{\mu}_o, \widehat{\boldsymbol{\Sigma}}_o) = \frac{1}{(2\pi)^{np/2}|\widehat{\Sigma}_o|^{n/2}} e^{-np/2},$$

where $\widehat{\boldsymbol{\Sigma}}_o = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}_o)(\boldsymbol{x}_i - \boldsymbol{\mu}_o)'$. Consequently,

$$\Lambda = \left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_o|}\right)^{n/2}.$$

The statistic $\Lambda^{2/n} = |\widehat{\boldsymbol{\Sigma}}|/|\widehat{\boldsymbol{\Sigma}}_o|$ is called the *Wilks' lambda*.

Result 5.1. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Consider the hypothesis testing: $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ versus $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_o$. Then, the likelihood ratio statistic becomes

$$\Lambda^{2/n} = \left(1 + \frac{T^2}{n-1}\right)^{-1},$$

where T^2 is the Hotelling's T^2 statistic.

Implication: Under the normality assumption, Hotelling's T^2 test is equivalent to the maximum likelihood ratio test. Rejecting H_o when T^2 is large is the same as rejecting H_o when the likelihood ratio is small.

Proof. The following equality of determinants is useful

$$\begin{aligned}
& (-1)|\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'+n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_o)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_o)'| \\
= \; & |\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'||-1-n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_o)'\left(\sum_{i=1}^{n}(\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'\right)^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_o)|.
\end{aligned}$$

This follows directly from the identity $|\boldsymbol{A}|=|\boldsymbol{A}_{22}||\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}|=|\boldsymbol{A}_{11}||\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}|$ with

$$\boldsymbol{A}=\begin{bmatrix}\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})' & \sqrt{n}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_o)\\ \sqrt{n}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_o)' & -1\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}.$$

Note also that

$$\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_o)(\boldsymbol{x}_i-\boldsymbol{\mu}_o)'=\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'+n(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_o)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_o)'.$$

Therefore,

$$(-1)|\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_o)(\boldsymbol{x}_i-\boldsymbol{\mu}_o)'|=|\sum_{i=1}^{n}(\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})|(-1)\left(1+\frac{T^2}{n-1}\right).$$

That is,

$$|n\widehat{\boldsymbol{\Sigma}}_o|=|n\widehat{\boldsymbol{\Sigma}}|\left(1+\frac{T^2}{n-1}\right).$$

Consequently,

$$\Lambda^{2/n}=\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_o|}=\left(1+\frac{T^2}{n-1}\right)^{-1}.$$

Remark: The prior result suggests that there is no need to calculate $\boldsymbol{S}^{-1}$ in computing the Hotelling's T^2 statistic. Indeed,

$$T^2=\frac{(n-1)|\widehat{\boldsymbol{\Sigma}}_o|}{|\widehat{\boldsymbol{\Sigma}}|}-(n-1),$$

which only requires calculation of determinants.

3 Confidence regions

Let $\boldsymbol{\theta}$ be the parameter vector of interest and $\boldsymbol{\Theta}$ be the paramter space. For a random sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$, let $\boldsymbol{X}$ denote the data. A $100(1-\alpha)\%$ *confidence region* $R(\boldsymbol{X})$ is defined as

$$Pr[\boldsymbol{\theta}\in R(\boldsymbol{X})]=1-\alpha,$$

where the probability is evaluated at the unknown true parameter $\boldsymbol{\theta}$. That is, the region $R(\boldsymbol{X})$ will cover $\boldsymbol{\theta}$ with probability $1-\alpha$.
For the mean vector $\boldsymbol{\mu}$ of a multivariate normal distribution, the confidence region (C.R.) can be obtained by the result

$$T^2 \sim \frac{(n-1)p}{n-p} F_{p,n-p}.$$

That is,

$$Pr\left[n(\bar{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)\right] = 1-\alpha,$$

whatever the values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, where $\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'$. In other words, if we measure the distance using the variance-covariance matrix $\frac{1}{n}\boldsymbol{S}$, then $\bar{\boldsymbol{X}}$ will be within $[(n-1)pF_{p,n-p}(\alpha)/(n-p)]^{1/2}$ of $\boldsymbol{\mu}$.

Remark: The quantity $\sqrt{(\bar{\boldsymbol{X}}-\boldsymbol{\mu})'(\boldsymbol{S}/n)^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu})}$ can be considered as the *Mahalanobis distance* of $\boldsymbol{\mu}$ from $\bar{\boldsymbol{X}}$, because the covariance matrix of $\bar{\boldsymbol{X}}$ is $\frac{1}{n}\boldsymbol{\Sigma}$, which is consistently estimated by $\frac{1}{n}\boldsymbol{S}$. Compared with the Euclidean distance, Mahalanobis distance takes into consideration the covariance structure. It is a distance measure based on correlations between variables.
For normal distribution, C.R. can be viewed as ellipsoids centered at $\boldsymbol{\mu}$ and have axes determined by the eigenvalues and eigenvectors of $\boldsymbol{S}$. For $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution, the contours of constant density are ellipsoids defined by $\boldsymbol{x}$ such that

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c^2.$$

These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\boldsymbol{e}_i$, where $\boldsymbol{\Sigma}\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i$ for $i=1,\ldots,p$. Consequently, the C.R. are centered at $\bar{\boldsymbol{x}}$ and the axes of the confidence ellipsoid are

$$\pm\sqrt{\lambda_i}\sqrt{\frac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)}\boldsymbol{e}_i,$$

where $\boldsymbol{S}\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i$ for $i=1,\ldots,p$.

Simultaneous confidence intervals: Suppose that $\boldsymbol{X} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. For any non-zero p-dimensional vector $\boldsymbol{a}$, $Z = \boldsymbol{a}'\boldsymbol{X} \sim N(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$. If $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ are a random sample from $\boldsymbol{X}$, $\{z_i = \boldsymbol{a}'\boldsymbol{x}_i\}$ is a random sample of Z. The sample mean and variance are $\bar{z} = \boldsymbol{a}'\bar{\boldsymbol{x}}$ and $s_z^2 = \boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}$, where $\boldsymbol{S}$ is the sample covariance matrix of $\boldsymbol{x}_i$.
If $\boldsymbol{a}$ is fixed, the $100(1-\alpha)\%$ confidence interval of $\boldsymbol{a}'\boldsymbol{\mu}$ is

$$\boldsymbol{a}'\bar{\boldsymbol{x}} - t_{n-1}(\alpha/2)\frac{\sqrt{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}} \leq \boldsymbol{a}'\boldsymbol{\mu} \leq \boldsymbol{a}'\bar{\boldsymbol{x}} + t_{n-1}(\alpha/2)\frac{\sqrt{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}}.$$

This interval consists of $\boldsymbol{a}'\boldsymbol{\mu}$ values for which

$$\left|\frac{\sqrt{n}(\boldsymbol{a}'\bar{\boldsymbol{x}} - \boldsymbol{a}'\boldsymbol{\mu})}{\sqrt{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}}\right| \leq t_{n-1}(\alpha/2),$$

or, equivalently,

$$\frac{n[\boldsymbol{a}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]^2}{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}} \leq t^2_{n-1}(\alpha/2).$$

If we consider all values of $\boldsymbol{a}$ for which the prior inequality holds, then we are naturally led to the determination of

$$\max_{\boldsymbol{a}} \frac{n[\boldsymbol{a}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]^2}{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}.$$

Using the Maximization Lemma (see Eq. (2.50) of the text, p. 80) with $\boldsymbol{x}=\boldsymbol{a}$, $\boldsymbol{d}=\bar{\boldsymbol{x}}-\boldsymbol{\mu}$ and $\boldsymbol{B}=\boldsymbol{S}$, we obtain

$$\max_{\boldsymbol{a}} \frac{n[\boldsymbol{a}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]^2}{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}} = n\left[\max_{\boldsymbol{a}} \frac{[\boldsymbol{a}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]^2}{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}\right] = n(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) = T^2,$$

with the maximum occurring for $\boldsymbol{a}$ proportional to $\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})$. Noting that $T^2 \sim \frac{(n-1)p}{n-p}F_{p,n-p}$, we have the following result.

Result 5.3. Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ population with $\boldsymbol{\Sigma}$ being positive definite. Then, simultaneously for all $\boldsymbol{a}$, the interval

$$\left(\boldsymbol{a}'\bar{\boldsymbol{X}} - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}, \quad \boldsymbol{a}'\bar{\boldsymbol{X}} + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}\right)$$

will contain $\boldsymbol{a}'\boldsymbol{\mu}$ with probability $1-\alpha$.

For convenience, we refer to the simultaneous confidence intervals of Result 5.3 as T^2-intervals. In particular, the choices of $\boldsymbol{a} = (1,0,\ldots,0)'$, $(0,1,0,\ldots,0)'$, ..., $(0,\ldots,0,1)'$ allow us to conclude that

$$\bar{x}_i - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}}{n}},$$

hold simultaneously with confidence coefficient $1-\alpha$ for all $i=1,\ldots,p$.
Furthermore, by choosing $\boldsymbol{a} = (0,\ldots,0,1,0,\ldots,0,-1,0,\ldots,0)'$ with 1 at the ith position and -1 at the kth position, where $i \neq k$, we have $\boldsymbol{a}'\boldsymbol{\mu} = \mu_i - \mu_k$ and $\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a} = s_{ii} - 2s_{ik} + s_{kk}$, and the T^2-intervals

$$\bar{x}_i - \bar{x}_k - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}-2s_{ik}+s_{kk}}{n}} \leq \mu_i - \mu_k \leq$$

$$\bar{x}_i - \bar{x}_k + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}-2s_{ik}+s_{kk}}{n}}.$$

Table 1: Critical multipliers for one-at-a-time t-intervals and the T^2-intervals for selected sample size n and dimension p, where $1-\alpha = 0.95$

		$\sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(0.05)}$		
n	$t_{n-1}(0.025)$	$p=3$	$p=5$	$p=10$
15	2.145	3.495	4.825	11.51
30	2.045	3.089	3.886	5.835
50	2.010	2.961	3.631	5.044
100	1.984	2.874	3.470	4.617
200	1.972	2.834	3.396	4.438
∞	1.960	2.795	3.327	4.279

Comparison. An alternative approach to construct confidence intervals for the component means is to consider the component one at a time. This approach ignores the covariance structure of the variables and leads to the intervals

$$\bar{x}_i - t_{n-1}(\alpha/2)\sqrt{\frac{s_{ii}}{n}} \le \mu_i \le \bar{x}_i + t_{n-1}(\alpha/2)\sqrt{\frac{s_{ii}}{n}},$$

for $i = 1, \ldots, p$. Here each interval has a confidence coefficient $1-\alpha$. However, we do not know the confidence coefficient that all intervals contain their respective μ_i, except for the case that the variables X_i are independent. In the latter case, the confidence coefficient is $(1-\alpha)^p$.

Obviously, the T^2-interval is wider than the individual interval for each μ_i. Table **??** gives some comparisons of critical multipliers for one-at-a-time t-intervals and the T^2 intervals for selected n and p when the coverage probability is $1-\alpha = 0.95$. Note that the T^2 intervals hold for any linear combination $\boldsymbol{a}'\boldsymbol{\mu}$. If one is only interested in the p mean values, then T^2-intervals are likely to be too wide. On the other hand, the individual t-intervals may have simultaneous coverage probability much less than $1-\alpha$. A compromise is the Bonferroni method.

Bonferroni method for multiple comparisons. In some applications, the T^2 intervals might be too wide to be of practical value and one may only concern about a finite number of linear combinations, say $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$. In this case, we may construct m confidence intervals that are shorter than the T^2 intervals. This method to multiple comparisons is called the *Bonferroni method*.

Let C_i denote a confidence statement about the value of $\boldsymbol{a}_i'\boldsymbol{\mu}$ with $Pr[C_i \text{ true }] = 1-\alpha_i$, where $i = 1, \ldots, m$. Then,

$$\begin{aligned} Pr[\text{all } C_i \text{ true}] &= 1 - Pr[\text{at least one } C_i \text{ false}] \\ &\ge 1 - \sum_{i=1}^{m} Pr[C_i \text{ false}] = 1 - \sum_{i=1}^{m}[1 - Pr(C_i \text{ true})] \\ &= 1 - (\alpha_1 + \cdots + \alpha_m). \end{aligned}$$

Table 2: Ratio of the lengths between Bonferroni interval and T^2-interval for selected sample size n and dimension p when the probability of the interval is 0.95

	p				
n	2	3	4	5	10
15	0.877	0.778	0.693	0.617	0.289
30	0.899	0.823	0.761	0.709	0.521
50	0.906	0.837	0.783	0.738	0.583
100	0.911	0.847	0.798	0.757	0.622
200	0.913	0.852	0.804	0.766	0.640
∞	0.916	0.856	0.811	0.774	0.656

This is a special case of the Bonferroni inequality and allows us to control the overall type-I error rate $\alpha_1 + \cdots + \alpha_m$. The choices of α_i make the method flexible, depending on the prior knowledge on the importance of each interval. If there is no prior information, then $\alpha_i = \alpha/m$ is often used.
Let $z_i = \boldsymbol{a}_i'\boldsymbol{x}_i$ and $s_{zi} = \boldsymbol{a}_i'\boldsymbol{S}\boldsymbol{a}_i$. Then, each of the intervals

$$\bar{z}_i \pm t_{n-1}(\alpha/(2m))\sqrt{\frac{s_{zi}}{n}}, \quad i = 1, \cdots, m,$$

contains $\boldsymbol{a}_i'\boldsymbol{\mu}$ with probability $1 - \alpha/m$. Jointly, all $\boldsymbol{a}_i'\boldsymbol{\mu}$ are in their respective intervals with probability $p \geq 1 - m(\alpha/m) = 1 - \alpha$.
For instance, if $m = p$ and $\boldsymbol{a}_i$ is the ith unit vector, then the intervals

$$\bar{x}_i - t_{n-1}(\alpha/(2p))\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + t_{n-1}(\alpha/(2p))\sqrt{\frac{s_{ii}}{n}}$$

hold with probability at least $1 - \alpha$. We refer to these intervals as the *Bonferroni intervals*. In general, if we use $\alpha_i = \alpha/p$, then we can compare the length of Bonferroni intervals with those of the T^2 intervals. The ratio is

$$\frac{\text{Length of Bonferroni interval}}{\text{Length of } T^2 \text{ interval}} = \frac{t_{n-1}(\alpha/(2p))}{\sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)}}.$$

This ratio does not depend on the sample mean or covariance matrix. Table **??** gives the prior ratio for some selected n and p. See also, Table 5.4 of the text(page 234).

4 Large sample case

When the sample size is large, we can apply the central limit theory so that the population may not be normal. Recall that, when n is large, $n(\bar{X} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \chi_p^2$. Thus,

$$Pr[n(\bar{X} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \chi_p^2(\alpha)] \approx 1 - \alpha,$$

where $\chi^2_p(\alpha)$ is the upper 100α percentile of a chi-square distribution with p degrees of freedom. This property can be used (a) to construct asymptotic confidence intervals for the means μ_i and (b) to perform hypothesis testing about $\boldsymbol{\mu}$.

Asymptotic confidence intervals: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and positive covariance matrix $\boldsymbol{\Sigma}$. If $n - p$ is sufficiently large,

$$\boldsymbol{a}'\bar{\boldsymbol{X}} \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}{n}}$$

will contain $\boldsymbol{a}'\boldsymbol{\mu}$, for every $\boldsymbol{a}$, with probability approximately $1 - \alpha$.

Since the chi-square distribution does not depend on $\boldsymbol{a}$, the above confidence intervals are simultaneous confidence intervals. This result can be used to compare means of different components of $\boldsymbol{X}$.

Asymptotic testing: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and positive covariance matrix $\boldsymbol{\Sigma}$. When $n-p$ is sufficiently large, the hypothesis $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ is rejected in favor of $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_o$, at the significance level α if

$$n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_o)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_o) > \chi^2_p(\alpha).$$

Remark: A simple R function `confreg` is included in `ama.R` and it provides various types of confidence intervals for the means of the components of $\boldsymbol{X}$.

Example: Consider the monthly log returns of four Midwest companies used before. The various confidence intervals are obtained as follows:

```
> setwd("C:/Users/rst/teaching/ama")
> x=read.table("m-ba4c9807.txt")
> source("ama.R")  <== so that the command confreg can be used.
> confreg(x)
[1] "C.R. based on T^2"
          [,1]     [,2]
[1,] -3.524903 2.999011
[2,] -1.983378 3.204879
[3,] -1.176112 2.519207
[4,] -3.179484 3.233071
[1] "CR based on individual t"
           [,1]     [,2]
[1,] -2.3000743 1.774182
[2,] -1.0093115 2.230812
[3,] -0.4823362 1.825432
[4,] -1.9755624 2.029150
[1] "CR based on Bonferroni"
           [,1]     [,2]
[1,] -2.8722323 2.346340
```

```
[2,] -1.4643301 2.685831
[3,] -0.8064218 2.149517
[4,] -2.5379541 2.591541
[1] "Asymp. simu. CR"
          [,1]     [,2]
[1,] -3.431874 2.905982
[2,] -1.909395 3.130896
[3,] -1.123418 2.466514
[4,] -3.088044 3.141631
```

5 Multivariate control charts

We discuss two cases. In the first case, the goal is to identify unusual observations in a sample, including the possibility of drift over time. This is called the T^2-chart, which uses the distance d_i^2 discussed before in assessing normality assumption. For the ith observation, the d_i^2 statistic is

$$d_i^2 = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})' \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \bar{\boldsymbol{x}}).$$

The upper control limit is then set by the upper quantiles of χ_p^2. Typically, 95th and 99th percentiles are used to set the upper control limit and the lower control limit is zero. Once a point is found to be outside the control limit, the individual confidence intervals for the component means can be used to identify the source of the deviation.
As an illustration, consider the monthly log returns of the four Midwest companies from 1998 to 2007. The T^2-chart is given in Figure **??**. About three data points are outside of the 99% limit, indicating a volatility period.
In the second case, we consider a control chart for future observations. The theory behind this type of control chart is the following result.

Result 5.6. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independently distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $\boldsymbol{X}$ be a future observation from the same population. Then,

$$T^2 = \frac{n}{n+1} (\boldsymbol{X} - \bar{\boldsymbol{X}})' \boldsymbol{S}^{-1} (\boldsymbol{X} - \bar{\boldsymbol{X}}) \sim \frac{p(n-1)}{n-p} F_{p,n-p},$$

and a $100(1-\alpha)\%$ p-dimensional prediction ellipsoid is given by all $\boldsymbol{x}$ satisfying

$$(\boldsymbol{x} - \bar{\boldsymbol{x}})' \boldsymbol{S}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}) \leq \frac{p(n^2-1)}{n(n-p)} F_{p,n-p}(\alpha).$$

Proof: First, $E(\boldsymbol{X} - \bar{\boldsymbol{X}}) = \boldsymbol{0}$. Since $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ are independent,

$$\text{Cov}(\boldsymbol{X} - \bar{\boldsymbol{X}}) = \boldsymbol{\Sigma} + \frac{1}{n}\boldsymbol{\Sigma} = \frac{n+1}{n}\boldsymbol{\Sigma}.$$

Figure 2: T-sq chart for the monthly log returns of four Midwest companies: 1998-2007

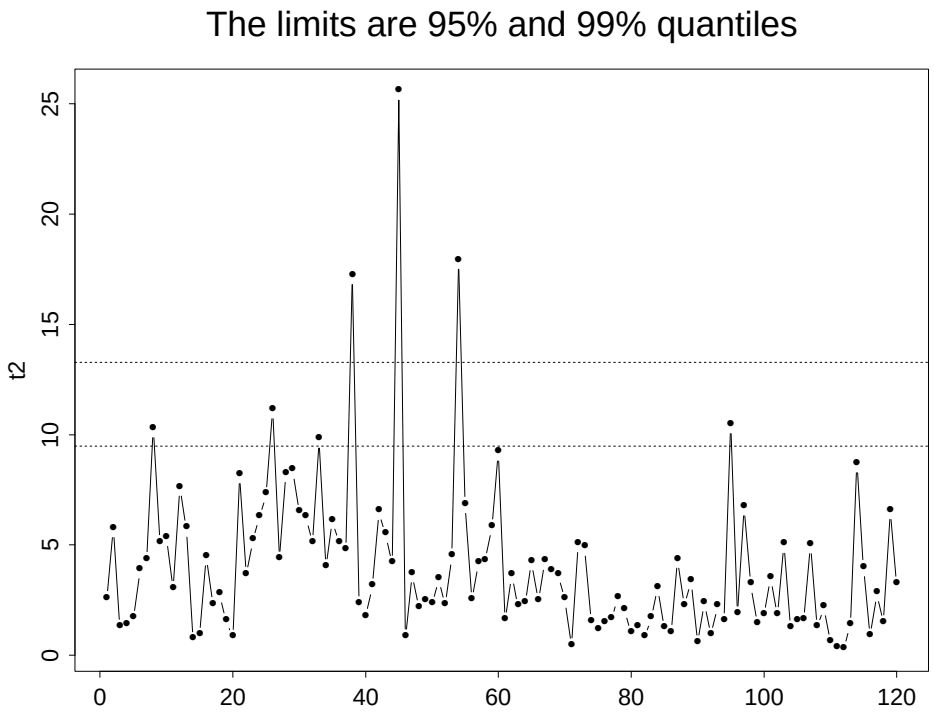
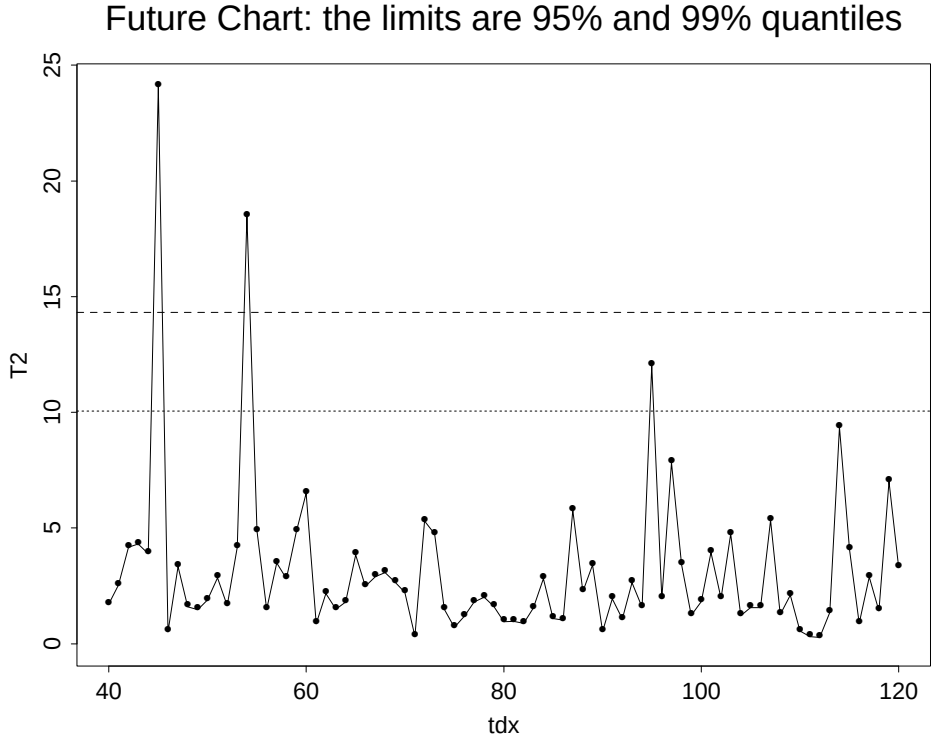


Figure 3: T^2 Future Chart for monthly log returns of four Midwest companies: 1998-2007



Thus, $\sqrt{n/(n+1)}(\boldsymbol{X} - \bar{\boldsymbol{X}}) \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Furthermore, by the result in Eq. (5.6) of the textbook,

$$\sqrt{\frac{n}{n+1}}(\boldsymbol{X} - \bar{\boldsymbol{X}})'\boldsymbol{S}^{-1}\sqrt{\frac{n}{n+1}}(\boldsymbol{X} - \bar{\boldsymbol{X}})$$

has a scaled F distribution as stated.

T^2-chart for future observations. For each new observation, plot

$$T^2 = \frac{n}{n+1}(\boldsymbol{x} - \bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}),$$

against time order. Set the lower control limit to zero and the upper control limit as

$$UCL = \frac{(n-1)p}{n-p}F_{p,n-p}(0.01).$$

For illustration, consider the monthly log return series of the four Midwest companies. We start with initial $n = 40$ observations. The T^2-chart for future observations is given in Figure **??**.

Remark: R functions for the two control charts are included in `ama.R`.

6 Missing values in normal random sample

A key assumption: Missing at random.
Two methods are available:

1. The EM algorithm: Dempster, Laird and Rubin (1977, JRSSB)
2. Markov chain Monte Carlo (MCMC) method

Some references

1. *The EM Alorithm and Extensions* by G. J. McLachlan and T. Krishnan (2008), John Wiley.
2. *The EM Algorithm and Related Statistical Models* by M. Watanabe and K. Yamaguchi (2003), CRC Press.
3. *Bayesian Data Analysis*, 2nd Ed, by Gelman, Carlin, Stern and Rubin (2003), CRC Press.
4. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd Ed., by B.P. Carlin and T.A. Louis (2001), CRC Press.

EM algorithm: Iterate between *Expectation* step and *Maximization* step.

- E-step: For each data point with missing values, use the conditional distribution

$$\boldsymbol{X}_1|\boldsymbol{X}_2 = \boldsymbol{x}_2 \sim N_k(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}),$$

  where $k = \dim(\boldsymbol{X}_1)$ and the partition is based on $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$. Here $\boldsymbol{X}_1$ denotes the missing components and $\boldsymbol{X}_2$ the observed component.

- M-step: Perform MLE estimation based on the complete data.

It is helpful to make use of sufficient statistics.

MCMC method: Also makes use of the same conditional distribution. However, instead of using the expectation, one draws a random sample from the conditional distribution to fill the missing values.

Basic Result used in MCMC method: Suppose that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ form a random sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is known. Suppose that the prior distribution of $\boldsymbol{\mu}$ is $N(\boldsymbol{\mu}_o, \boldsymbol{\Sigma}_o)$. Then the posterior distribution of $\boldsymbol{\mu}$ is also multivariate normal with mean $\boldsymbol{\mu}_*$ and covariance matrix $\boldsymbol{\Sigma}_*$, where

$$\boldsymbol{\Sigma}_*^{-1} = \boldsymbol{\Sigma}_o^{-1} + n\boldsymbol{\Sigma}^{-1}, \quad \boldsymbol{\mu}_* = \boldsymbol{\Sigma}_*(\boldsymbol{\Sigma}_o^{-1}\boldsymbol{\mu}_o + n\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}),$$

where $\bar{\boldsymbol{x}} = \sum_{i=1}^n \boldsymbol{x}_i/n$.

7 Impact of serial correlations

Consider the case in which $\boldsymbol{X}_i$ is serially correlated such as it follows a vector AR(1) model

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{a}_t, \quad \boldsymbol{a}_t \sim_{iid} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

where $E(\boldsymbol{X}_t) = \boldsymbol{\mu}$ and all eigenvalues of $\boldsymbol{\Phi}$ are less than 1 in modulus. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be n consecutive observations of the model. Define the lag-ℓ autocovariance matrix of $\boldsymbol{X}_t$ as

$$\boldsymbol{\Gamma}_\ell = E(\boldsymbol{X}_t - \boldsymbol{\mu})(\boldsymbol{X}_{t-\ell} - \boldsymbol{\mu})', \quad \ell = 0, \pm 1, \pm 2, \ldots.$$

It is easy to see that $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Gamma}'_{-\ell}$. Also, by repeatedly substitution, we have

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \boldsymbol{a}_t + \boldsymbol{\Phi}\boldsymbol{a}_{t-1} + \boldsymbol{\Phi}^2\boldsymbol{a}_{t-2} + \cdots.$$

Therefore,

$$\boldsymbol{\Gamma}_0 = \sum_{i=1}^{\infty} \boldsymbol{\Phi}^i \boldsymbol{\Sigma} (\boldsymbol{\Phi}^i)'.$$

For $\ell > 0$, post-multiplying the model by $(\boldsymbol{X}_{t-\ell} - \boldsymbol{\mu})'$, taking expectation, and using the fact that $\boldsymbol{X}_{t-\ell}$ is uncorrelated with $\boldsymbol{a}_t$, we get

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}, \quad \ell = 1, 2, \ldots.$$

Taking the transpose of the above equation, we have

$$\boldsymbol{\Gamma}'_\ell = \boldsymbol{\Gamma}'_{\ell-1}\boldsymbol{\Phi}'.$$

Using $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Gamma}'_{-\ell}$, we obtain

$$\boldsymbol{\Gamma}_{-\ell} = \boldsymbol{\Gamma}_{-(\ell-1)}\boldsymbol{\Phi}', \quad \ell > 1,$$

which is equivalent to

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Gamma}_{\ell+1}\boldsymbol{\Phi}', \quad \ell = -1, -2, \ldots.$$

For the vector AR(1) model, we can show the following properties:

1. $E(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}$.
2. $\mathrm{Cov}(n^{-1/2}\sum_{t=1}^n \boldsymbol{X}_t) \to_p \boldsymbol{\Omega}$, where $\boldsymbol{\Omega} = (\boldsymbol{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_0(\boldsymbol{I} - \boldsymbol{\Phi}')^{-1} - \boldsymbol{\Gamma}_0$, as $n \to \infty$.
3. $\boldsymbol{S} = \frac{1}{n-1}\sum_{t=1}^n (\boldsymbol{X}_t - \bar{\boldsymbol{X}})(\boldsymbol{X}_t - \bar{\boldsymbol{X}})' \to_p \boldsymbol{\Gamma}_0$ as $n \to \infty$.

Using these properties, we can show that $\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$ is approximately normal with mean $\boldsymbol{0}$ and covariance $\boldsymbol{\Omega}$. This implies that $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \chi^2_p$, not the usual statistic $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$. See Table 5.10 of the text (page 257) for the difference in coverage probability due to the effect of AR(1) serial correlations.