Random Vectors Part One[1]

STA431 Winter/Spring 2013

[1]See last slide for copyright information.

Overview

1. Definitions and Basic Results
2. The Centering Rule
3. Multivariate Normal

Random Vectors and Matrices

A *random matrix* is just a matrix of random variables. Their joint probability distribution is the distribution of the random matrix. Random matrices with just one column (say, $p \times 1$) may be called *random vectors*.

Expected Value

The expected value of a matrix is defined as the matrix of expected values. Denoting the $p \times c$ random matrix $\mathbf{X}$ by $[X_{i,j}]$,

$$E(\mathbf{X}) = [E(X_{i,j})].$$

Immediately we have natural properties like

$$
\begin{aligned}
E(\mathbf{X}+\mathbf{Y}) &= E([X_{i,j}]+[Y_{i,j}]) \\
&= [E(X_{i,j}+Y_{i,j})] \\
&= [E(X_{i,j})+E(Y_{i,j})] \\
&= [E(X_{i,j})]+[E(Y_{i,j})] \\
&= E(\mathbf{X})+E(\mathbf{Y}).
\end{aligned}
$$

Moving a constant through the expected value sign

Let $\mathbf{A} = [a_{i,j}]$ be an $r \times p$ matrix of constants, while $\mathbf{X}$ is still a $p \times c$ random matrix. Then

$$\begin{aligned} E(\mathbf{AX}) &= E\left(\left[\sum_{k=1}^{p} a_{i,k}X_{k,j}\right]\right) \\ &= \left[E\left(\sum_{k=1}^{p} a_{i,k}X_{k,j}\right)\right] \\ &= \left[\sum_{k=1}^{p} a_{i,k}E(X_{k,j})\right] \\ &= \mathbf{A}E(\mathbf{X}). \end{aligned}$$

Similar calculations yield $E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$.

Variance-Covariance Matrices

Let $\mathbf{X}$ be a $p \times 1$ random vector with $E(\mathbf{X}) = \boldsymbol{\mu}$. The *variance-covariance matrix* of $\mathbf{X}$ (sometimes just called the *covariance matrix*), denoted by $V(\mathbf{X})$, is defined as

$$V(\mathbf{X}) = E\left\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\right\}.$$

$V(\mathbf{X}) = E\left\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\right\}$

$$
\begin{aligned}
V(\mathbf{X}) &= E\left\{ \left[\begin{array}{c} X_1-\mu_1 \\ X_2-\mu_2 \\ X_3-\mu_3 \end{array} \right] \left[\begin{array}{ccc} X_1-\mu_1 & X_2-\mu_2 & X_3-\mu_3 \end{array} \right] \right\} \\
&= E\left\{ \left[\begin{array}{lll} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & (X_1-\mu_1)(X_3-\mu_3) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & (X_2-\mu_2)(X_3-\mu_3) \\ (X_3-\mu_3)(X_1-\mu_1) & (X_3-\mu_3)(X_2-\mu_2) & (X_3-\mu_3)^2 \end{array} \right] \right\} \\
&= \left[\begin{array}{lll} E\{(X_1-\mu_1)^2\} & E\{(X_1-\mu_1)(X_2-\mu_2)\} & E\{(X_1-\mu_1)(X_3-\mu_3)\} \\ E\{(X_2-\mu_2)(X_1-\mu_1)\} & E\{(X_2-\mu_2)^2\} & E\{(X_2-\mu_2)(X_3-\mu_3)\} \\ E\{(X_3-\mu_3)(X_1-\mu_1)\} & E\{(X_3-\mu_3)(X_2-\mu_2)\} & E\{(X_3-\mu_3)^2\} \end{array} \right] \\
&= \left[\begin{array}{lll} V(X_1) & Cov(X_1,X_2) & Cov(X_1,X_3) \\ Cov(X_1,X_2) & V(X_2) & Cov(X_2,X_3) \\ Cov(X_1,X_3) & Cov(X_2,X_3) & V(X_3) \end{array} \right].
\end{aligned}
$$

So, the covariance matrix $V(\mathbf{X})$ is a $p \times p$ symmetric matrix with variances on the main diagonal and covariances on the off-diagonals.

Analogous to $Var(a\,X) = a^2\,Var(X)$

Let $\mathbf{X}$ be a $p \times 1$ random vector with $E(\mathbf{X}) = \boldsymbol{\mu}$ and $V(\mathbf{X}) = \boldsymbol{\Sigma}$, while $\mathbf{A} = [a_{i,j}]$ is an $r \times p$ matrix of constants. Then

$$\begin{aligned}
V(\mathbf{AX}) &= E\left\{(\mathbf{AX} - \mathbf{A}\boldsymbol{\mu})(\mathbf{AX} - \mathbf{A}\boldsymbol{\mu})'\right\} \\
&= E\left\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})\left(\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})\right)'\right\} \\
&= E\left\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}'\right\} \\
&= \mathbf{A}E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\}\mathbf{A}' \\
&= \mathbf{A}V(\mathbf{X})\mathbf{A}' \\
&= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'
\end{aligned}$$

Matrix of covariances between two random vectors

Let $\mathbf{X}$ be a $p \times 1$ random vector with $E(\mathbf{X}) = \boldsymbol{\mu}_x$ and let $\mathbf{Y}$ be a $q \times 1$ random vector with $E(\mathbf{Y}) = \boldsymbol{\mu}_y$. The $p \times q$ matrix of covariances between the elements of $\mathbf{X}$ and the elements of $\mathbf{Y}$ is

$$C(\mathbf{X}, \mathbf{Y}) = E\left\{(\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{Y} - \boldsymbol{\mu}_y)'\right\}.$$

Adding a constant has no effect

On variances and covariances

It's clear from the definitions:

- $V(\mathbf{X}) = E\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\}$
- $C(\mathbf{X},\mathbf{Y}) = E\left\{(\mathbf{X}-\boldsymbol{\mu}_x)(\mathbf{Y}-\boldsymbol{\mu}_y)'\right\}$

That

- $V(\mathbf{X}+\mathbf{a}) = V(\mathbf{X})$
- $C(\mathbf{X}+\mathbf{a},\mathbf{Y}+\mathbf{b}) = C(\mathbf{X},\mathbf{Y})$

For example, $E(\mathbf{X}+\mathbf{a}) = \boldsymbol{\mu}+\mathbf{a}$, so

$$\begin{aligned} V(\mathbf{X}+\mathbf{a}) &= E\left\{(\mathbf{X}+\mathbf{a}-(\boldsymbol{\mu}+\mathbf{a}))(\mathbf{X}+\mathbf{a}-(\boldsymbol{\mu}+\mathbf{a}))'\right\} \\ &= E\left\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\right\} \\ &= V(\mathbf{X}) \end{aligned}$$

The Centering Rule

Using $V(\mathbf{X}+\mathbf{a}) = V(\mathbf{X})$

Often, variance and covariance calculations can be simplified by subtracting off constants first. Denote the *centered* version of $\mathbf{X}$ by $\overset{c}{\mathbf{X}} = \mathbf{X} - E(\mathbf{X})$, so that

- $E(\overset{c}{\mathbf{X}}) = \mathbf{0}$ and
- $V(\overset{c}{\mathbf{X}}) = E(\overset{c}{\mathbf{X}}\overset{c}{\mathbf{X}}') = V(\mathbf{X})$

The centering rule is a general version of this.

Linear combinations

$$\begin{aligned}
\mathbf{L} &= \mathbf{A}_1\mathbf{X}_1 + \cdots + \mathbf{A}_m\mathbf{X}_m + \mathbf{b} \\
\stackrel{c}{\mathbf{L}} &= \mathbf{A}_1 \stackrel{c}{\mathbf{X}}_1 + \cdots + \mathbf{A}_m \stackrel{c}{\mathbf{X}}_m, \text{ where} \\
\stackrel{c}{\mathbf{X}}_j &= \mathbf{X}_j - E(\mathbf{X}_j) \text{ for } j = 1, \ldots, m.
\end{aligned}$$

The centering rule says

$$\begin{aligned}
V(\mathbf{L}) &= V(\stackrel{c}{\mathbf{L}}) \\
C(\mathbf{L}_1, \mathbf{L}_2) &= C(\stackrel{c}{\mathbf{L}}_1, \stackrel{c}{\mathbf{L}}_2)
\end{aligned}$$

Example: $V(\mathbf{X}+\mathbf{Y})$

Using the centering rule

$$\begin{aligned}
V(\mathbf{X}+\mathbf{Y}) &= V(\overset{c}{\mathbf{X}}+\overset{c}{\mathbf{Y}}) \\
&= E(\overset{c}{\mathbf{X}}+\overset{c}{\mathbf{Y}})(\overset{c}{\mathbf{X}}+\overset{c}{\mathbf{Y}})' \\
&= E(\overset{c}{\mathbf{X}}+\overset{c}{\mathbf{Y}})(\overset{c}{\mathbf{X}}'+\overset{c}{\mathbf{Y}}') \\
&= E(\overset{c}{\mathbf{X}}\overset{c}{\mathbf{X}}')+E(\overset{c}{\mathbf{Y}}\overset{c}{\mathbf{Y}}')+E(\overset{c}{\mathbf{X}}\overset{c}{\mathbf{Y}}')+E(\overset{c}{\mathbf{Y}}\overset{c}{\mathbf{X}}') \\
&= V(\mathbf{X})+V(\mathbf{Y})+C(\mathbf{X},\mathbf{Y})+C(\mathbf{Y},\mathbf{X})
\end{aligned}$$

Example: $Cov(\overline{X}, X_j - \overline{X}) = 0$

Scalar calculation using the centering rule

Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Since $\overline{X}$ and $X_j - \overline{X}$ are both linear combinations,

$$\begin{aligned}
Cov(\overline{X}, X_j - \overline{X}) &= Cov(\stackrel{c}{\overline{X}}, \stackrel{c}{X}_j - \stackrel{c}{\overline{X}}) \\
&= E\left(\stackrel{c}{\overline{X}}\,(\stackrel{c}{X}_j - \stackrel{c}{\overline{X}})\right) \\
&= E\left(\stackrel{c}{X}_j \stackrel{c}{\overline{X}}\right) - E\left(\stackrel{c}{\overline{X}}^{\,2}\right) \\
&= E\left(\stackrel{c}{X}_j \, \frac{1}{n}\sum_{i=1}^{n} \stackrel{c}{X}_i\right) - Var\left(\stackrel{c}{\overline{X}}\right) \\
&= E\left(\frac{1}{n}\sum_{i=1}^{n} \stackrel{c}{X}_i \stackrel{c}{X}_j\right) - Var\left(\overline{X}\right)
\end{aligned}$$

Calculation continued

$$
\begin{aligned}
&= E\left(\frac{1}{n}\sum_{i=1}^{n} \overset{c}{X}_i \overset{c}{X}_j\right) - Var\left(\overline{X}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left(\overset{c}{X}_i \overset{c}{X}_j\right) - \frac{\sigma^2}{n} \\
&= \frac{1}{n} E\left(\overset{c}{X}_j^{\,2}\right) + \frac{1}{n}\sum_{i\neq j} E\left(\overset{c}{X}_i\right) E\left(\overset{c}{X}_j\right) - \frac{\sigma^2}{n} \\
&= \frac{1}{n} Var\left(\overset{c}{X}_j\right) - \frac{\sigma^2}{n} \\
&= \frac{1}{n} Var(X_j) - \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} \\
&= 0
\end{aligned}
$$

The Multivariate Normal Distribution

The $p \times 1$ random vector $\mathbf{X}$ is said to have a *multivariate normal distribution*, and we write $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if $\mathbf{X}$ has (joint) density

$$f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{\mu}$ is $p \times 1$ and $\boldsymbol{\Sigma}$ is $p \times p$ symmetric and positive definite.

$\boldsymbol{\Sigma}$ positive definite

- Positive definite means that for any non-zero $p \times 1$ vector $\mathbf{a}$, we have $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} > 0$.
- Since the one-dimensional random variable $Y = \sum_{i=1}^{p} a_i X_i$ may be written as $Y = \mathbf{a}'\mathbf{X}$ and $Var(Y) = V(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$, it is natural to require that $\boldsymbol{\Sigma}$ be positive definite.
- All it means is that every non-zero linear combination of $\mathbf{X}$ values has a positive variance.
- And recall $\boldsymbol{\Sigma}$ positive definite is equivalent to $\boldsymbol{\Sigma}^{-1}$ positive definite.

Analogies

(Multivariate normal reduces to the univariate normal when $p = 1$)

- Univariate Normal
  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$
  - $E(X) = \mu, V(X) = \sigma^2$
  - $\frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$
- Multivariate Normal
  - $f(\mathbf{x}) = \frac{1}{|\mathbf{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$
  - $E(\mathbf{X}) = \boldsymbol{\mu},\ V(\mathbf{X}) = \mathbf{\Sigma}$
  - $(\mathbf{X}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(p)$

More properties of the multivariate normal

- If $\mathbf{c}$ is a vector of constants, $\mathbf{X} + \mathbf{c} \sim N(\mathbf{c} + \boldsymbol{\mu}, \boldsymbol{\Sigma})$
- If $\mathbf{A}$ is a matrix of constants, $\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$
- Linear combinations of multivariate normals are multivariate normal.
- All the marginals (dimension less than p) of $\mathbf{X}$ are (multivariate) normal, but it is possible in theory to have a collection of univariate normals whose joint distribution is not multivariate normal.
- For the multivariate normal, zero covariance implies independence. The multivariate normal is the only continuous distribution with this property.

An easy example

If you do it the easy way

Let $\mathbf{X} = (X_1, X_2, X_3)'$ be multivariate normal with

$$\boldsymbol{\mu} = \left[\begin{array}{c} 1 \\ 0 \\ 6 \end{array} \right] \text{ and } \boldsymbol{\Sigma} = \left[\begin{array}{ccc} 2 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 2 \end{array} \right].$$

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$. Find the joint distribution of Y_1 and Y_2.

In matrix terms

$Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$ means $\mathbf{Y} = \mathbf{AX}$

$$\left[\begin{array}{c} Y_1 \\ Y_2 \end{array}\right] = \left[\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 1 & 1 \end{array}\right] \left[\begin{array}{c} X_1 \\ X_2 \\ X_3 \end{array}\right]$$

$\mathbf{Y} = \mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}')$

You could do it by hand, but

```
> mu = cbind(c(1,0,6))
> Sigma = rbind( c(2,1,0),
+                c(1,4,0),
+                c(0,0,2) )
> A = rbind( c(1,1,0),
+            c(0,1,1) ); A
> A %*% mu                  # E(Y)
     [,1]
[1,]    1
[2,]    6
> A %*% Sigma %*% t(A)   # V(Y)
     [,1] [,2]
[1,]    8    5
[2,]    5    6
```

Multivariate normal likelihood

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

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

Showing the details

For the multivariate normal likelihood

Adding and subtracting $\overline{\mathbf{x}}$ in $\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})$, we get

$$
\begin{aligned}
& \sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}} + \overline{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \overline{\mathbf{x}} + \overline{\mathbf{x}} - \boldsymbol{\mu}) \\
= & \sum_{i=1}^n (\mathbf{a}_i + \mathbf{b})'\boldsymbol{\Sigma}^{-1}(\mathbf{a}_i + \mathbf{b}) \\
= & \sum_{i=1}^n \left(\mathbf{a}_i'\boldsymbol{\Sigma}^{-1}\mathbf{a}_i + \mathbf{a}_i'\boldsymbol{\Sigma}^{-1}\mathbf{b} + \mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{a}_i + \mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b}\right) \\
= & \left(\sum_{i=1}^n \mathbf{a}_i'\boldsymbol{\Sigma}^{-1}\mathbf{a}_i\right) + \mathbf{0} + \mathbf{0} + n\,\mathbf{b}'\boldsymbol{\Sigma}^{-1}\mathbf{b} \\
= & \sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}) \; + \; n\,(\overline{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})
\end{aligned}
$$

Continuing the calculation

$$
\begin{aligned}
\sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}) &= \sum_{i=1}^n tr \left\{ (\mathbf{x}_i - \overline{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}) \right\} \\
&= \sum_{i=1}^n tr \left\{ \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}) (\mathbf{x}_i - \overline{\mathbf{x}})' \right\} \\
&= tr \left\{ \sum_{i=1}^n \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}) (\mathbf{x}_i - \overline{\mathbf{x}})' \right\} \\
&= tr \left\{ \boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}}) (\mathbf{x}_i - \overline{\mathbf{x}})' \right\} \\
&= n \, tr \left\{ \boldsymbol{\Sigma}^{-1} \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}}) (\mathbf{x}_i - \overline{\mathbf{x}})' \right\} \\
&= n \, tr \left(\boldsymbol{\Sigma}^{-1} \widehat{\boldsymbol{\Sigma}} \right)
\end{aligned}
$$

Substituting . . .

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

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

Maximizing the likelihood over $\boldsymbol{\mu}$ for any positive definite $\boldsymbol{\Sigma}$ without calculus

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}) \right\}$$

- Take log, maximize $-(\overline{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu})$.
- That is, minimize $(\overline{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu})$.
- Because $\boldsymbol{\Sigma}$ is positive definite, so is $\boldsymbol{\Sigma}^{-1}$.
- Thus $(\overline{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}) > 0$ for $\overline{\mathbf{x}} - \boldsymbol{\mu} \neq 0$
- And equal to zero only when $\boldsymbol{\mu} = \overline{\mathbf{x}}$.
- So that's where the likelihood has its maximum, for each $\boldsymbol{\Sigma}$.

Showing $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(p)$

$$
\begin{aligned}
\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu} &\sim N(\mathbf{0},\ \boldsymbol{\Sigma}) \\
\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y} &\sim N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right) \\
&= N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\ \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right) \\
&= N(\mathbf{0}, \mathbf{I})
\end{aligned}
$$

So $\mathbf{Z}$ is a vector of p independent standard normals, and

$$
\mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} = \mathbf{Z}'\mathbf{Z} = \sum_{j=1}^{p} Z_i^2 \sim \chi^2(p) \qquad \blacksquare
$$

$\overline{X}$ and S^2 independent

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N\left(\mu\mathbf{1}, \sigma^2\mathbf{I}\right) \qquad \mathbf{Y} = \begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_{n-1} - \overline{X} \\ \\ \overline{X} \end{pmatrix} = \mathbf{AX}$$

$\mathbf{Y} = \mathbf{AX}$

In more detail

$$\begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} & -\frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \\ X_n \end{pmatrix} = \begin{pmatrix} X_1 - \overline{X} \\ X_2 - \overline{X} \\ \vdots \\ X_{n-1} - \overline{X} \\ \overline{X} \end{pmatrix}$$

The argument

$$\mathbf{Y} = \mathbf{A}\mathbf{X} = \begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_{n-1} - \overline{X} \\ \\ \overline{X} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_2 \\ \hline \\ \overline{X} \end{pmatrix}$$

- $\mathbf{Y}$ is multivariate normal.
- $Cov\left(\overline{X}, (X_j - \overline{X})\right) = 0$ (Exercise)
- So $\overline{X}$ and $\mathbf{Y}_2$ are independent.
- So $\overline{X}$ and $S^2 = g(\mathbf{Y}_2)$ are independent. ■

Leads to the t distribution

If

- $Z \sim N(0,1)$ and
- $Y \sim \chi^2(\nu)$ and
- Z and Y are independent, then

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t(\nu)$$

Random sample from a normal distribution

Let $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then

- $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim N(0,1)$ and
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ and
- These quantities are independent, so

$$\begin{aligned} T &= \frac{\sqrt{n}(\overline{X}-\mu)/\sigma}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} \\ &= \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t(n-1) \end{aligned}$$

Copyright Information