Multivariate Normal Distribution

Properties
Sections 4.1-4.2

Estimation
Section 4.3

Generalized variance
Sections 3.1-3.4

Simple random samples
Sections 4.4-4.5 and 3.5-3.6

Evaluating the normality assumption, detecting outliers, transformations
Sections 4.6-4.8

Univariate normal distribution

If $X \sim N(\mu, \sigma^2)$ then the density function is

$$
\phi(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2\sigma^2} (x-\mu)^2}
$$

for $-\infty < x < \infty$

Note that

$$
\phi(x) = \frac{1}{(2\pi)^{\frac{1}{2}} (\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}
$$
Multivariate Normal Distribution

The notation

\[ \mathbf{x} \sim \mathcal{N}_p(\mu, \Sigma) \]

is used to indicate that a random vector \( \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \)

has a \( p \)-dimensional normal distribution with

\[ \mu = E(\mathbf{x}) \quad \text{and} \quad \Sigma = V(\mathbf{x}) \]

Density function:

\[ \phi(\mathbf{x}_1, \ldots, \mathbf{x}_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \mu)\Sigma^{-1}(\mathbf{x} - \mu)} \]

The quantity \( (\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu) \)

is called

- a squared Mahalanobis distance of \( \mathbf{x} \) from \( \mu \)
- a squared standardized distance of \( \mathbf{x} \) from \( \mu \)
- a squared statistical distance of \( \mathbf{x} \) from \( \mu \)
- a quadratic form
Note:

The density function does not exist when
- Σ is not positive definite
- \( \det(Σ) = 0 \)
- \( Σ^{-1} \) does not exist

We will assume that Σ is positive definite in most of our work. (not a restriction of practical importance).

A "degenerate" normal distribution is defined by its characteristic function

\[
ψ(t) = \exp\{it'μ - \frac{1}{2}t'Σt\}
\]

This only depends on \( μ = E(ξ) \) and \( Σ = V(ξ) \) which justifies the notation

\[ ξ \sim N(μ, Σ) \]
Multivariate normal (Gaussian) distribution.

(i) mathematical simplicity
(ii) multivariate central limit theorem
\[ \frac{1}{n} (x_1 + x_2 + \ldots + x_n) \]
(iii) many naturally occurring phenomena approximately exhibit this distribution

Samples from bivariate normal distributions:

\[ X_2 \]
\[ X_1 \]

\( p = .82 \)

\[ X_2 \]
\[ X_1 \]

\( p = -.85 \)

Inverse of a 2x2 matrix:
\[ A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \]

Determinant:
\[ |A| = a_{11} a_{22} - a_{12} a_{21} \]

Inverse:
\[ A^{-1} = \frac{1}{|A|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \]
\[ = \begin{bmatrix} \frac{a_{22}}{a_{11} a_{22} - a_{12} a_{21}} & \frac{-a_{12}}{a_{11} a_{22} - a_{12} a_{21}} \\ \frac{-a_{21}}{a_{11} a_{22} - a_{12} a_{21}} & \frac{a_{11}}{a_{11} a_{22} - a_{12} a_{21}} \end{bmatrix} \]

You can show that \( A A^{-1} = A^{-1} A = I \)
Inverse of a $3 \times 3$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} a_{22} a_{33} - a_{23} a_{32} & -a_{12} a_{33} + a_{13} a_{32} & a_{12} a_{23} - a_{13} a_{22} \\ a_{32} a_{23} - a_{33} a_{22} & a_{11} a_{33} - a_{13} a_{31} & -a_{11} a_{23} + a_{13} a_{21} \\ a_{32} a_{23} - a_{33} a_{22} & -a_{11} a_{33} + a_{13} a_{31} & a_{11} a_{23} - a_{13} a_{21} \end{bmatrix}$$

Inverse of a partitioned matrix

$$W = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

$$W^{-1} = \begin{bmatrix} A^{-1} + A^{-1} B V^{-1} C A^{-1} & -A^{-1} B V^{-1} \\ -V^{-1} C A^{-1} & V^{-1} \end{bmatrix}$$

where

$$V = D - C A^{-1} B$$
Geometry for the bivariate normal distribution: \( X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \)

\[
\phi(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)} 
\]

where \( \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \) and write \( \Sigma \) as

\[
\Sigma = \begin{bmatrix}
\sigma_{11} & \sigma_{12} \\
\sigma_{12} & \sigma_{22}
\end{bmatrix} = \begin{bmatrix}
\sigma_{11} & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_{22}
\end{bmatrix}
\]

Then

\[
|\Sigma| = \sigma_{11} \sigma_{22} (1-\rho^2) \quad \text{and} \quad \sigma_i = \sqrt{\sigma_{ii}} \\
\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{bmatrix}
\sigma_{22} & -\rho \sigma_1 \sigma_2 \\
-\rho \sigma_1 \sigma_2 & \sigma_{11}
\end{bmatrix}
\]

\[
\phi(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ (x_1-\mu_1)^2 - 2\rho(x_1-\mu_1)(x_2-\mu_2) + (x_2-\mu_2)^2 \right] \right\}
\]

The density function is a function of \( \mu_1, \mu_2, \sigma_1, \sigma_2, \rho \)

(i) The density is well defined if \(-1 < \rho < 1\)

(ii) If \( \rho = 0 \), then \( \phi(x_1, x_2) \) is the product of

\[
\phi(x_1) = \frac{1}{\sqrt{2\pi \sigma_1}} e^{-\frac{1}{2\sigma_1^2} (x_1-\mu_1)^2} \\
\phi(x_2) = \frac{1}{\sqrt{2\pi \sigma_2}} e^{-\frac{1}{2\sigma_2^2} (x_2-\mu_2)^2}
\]

Hence uncorrelated \( \Leftrightarrow \) independent
(iii) The density is constant for all $\mathbf{x} = (x_1, x_2)$ points for which

$$\text{constant} = C = (\mathbf{x} - \mathbf{\mu})' \Sigma^{-1} (\mathbf{x} - \mathbf{\mu})$$

$$= \left[ \frac{1}{1-\rho^2} \right] \left[ \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right]$$

This is an equation for an ellipse centered at $\mathbf{\mu} = (\mu_1, \mu_2)$.

Contour of constant density:

$$(\mathbf{x} - \mathbf{\mu})' \Sigma^{-1} (\mathbf{x} - \mathbf{\mu}) = \text{constant}$$
Spherical covariance structure

- Equal variances: $\sigma_{11} = \sigma_{22} = \sigma^2$
- Zero correlation: $\rho = 0$

If $\rho = 0$, the axes of the ellipses are parallel to the coordinate axes.

$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$

What are the lengths and positions of the ellipsoids corresponding to contours of constant density $(\rho = 0)^2$?
Solving quadratic equations:

\[ ax^2 + bx + c = 0 \]

Solution:

\[ x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \]

and

\[ x = \frac{-b \mp \sqrt{b^2 - 4ac}}{2a} \]

Eigenvalues of \( \Sigma = \begin{bmatrix} \sigma_1 & \rho \sigma_2 \\ \rho \sigma_2 & \sigma_2 \end{bmatrix} \)

are the solutions to

\[ \det(\Sigma - \lambda I) = 0 \]

\[ \begin{vmatrix} \sigma_1 - \lambda & \rho \sigma_2 \\ \rho \sigma_2 & \sigma_2 - \lambda \end{vmatrix} = 0 \]

\[ \begin{vmatrix} \sigma_1 - \lambda & \rho \sigma_2 \\ \rho \sigma_2 & \sigma_2 - \lambda \end{vmatrix} = \lambda^2 - (\sigma_1 + \sigma_2)\lambda + \sigma_1 \sigma_2 (1 - \rho^2) \]

The eigenvalues are

\[ \lambda_1 = \frac{1}{2} \left( \sigma_1 + \sigma_2 + \sqrt{(\sigma_1 + \sigma_2)^2 - 4 \sigma_1 \sigma_2 (1 - \rho^2)} \right) \]

\[ \lambda_2 = \frac{1}{2} \left( \sigma_1 + \sigma_2 - \sqrt{(\sigma_1 + \sigma_2)^2 - 4 \sigma_1 \sigma_2 (1 - \rho^2)} \right) \]
Solutions:

\[
\mathbf{e}_i = \begin{bmatrix} \mathbf{p} \\ \mathbf{d} \end{bmatrix}
\]

where

\[
b = \frac{\sigma_2 - \sigma_1}{2 \rho \sigma_2}
\]

or

\[
\mathbf{e}_1 = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{\rho^2 + 1} \\ -\frac{\sigma^2}{\rho^2 + 1} \end{bmatrix}
\]

or

\[
\mathbf{e}_2 = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{\rho^2 + 1} \\ -\frac{\sigma^2}{\rho^2 + 1} \end{bmatrix}
\]

Choose \( d \) to satisfy

\[
1 = \mathbf{e}_1^T \mathbf{e}_1 = d^2 + \frac{\sigma^2}{\rho^2 + 1}
\]

Corresponding eigenvectors are defined by

\[
(\sigma_i, \lambda_i) e_i = 0 = 0 = 0
\]

or

\[
\begin{bmatrix} \sigma_1 - \lambda_i & \rho \sigma_2 \\ \rho \sigma_2 & \sigma_2 - \lambda_i \end{bmatrix} \begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix} = 0
\]
The eigenvector corresponding to \( \lambda_2 \) is

\[
E_2 = \begin{bmatrix}
\frac{1}{\sqrt{1+c^2}} \\
\frac{c}{\sqrt{1+c^2}} \\
\frac{-c}{\sqrt{1+c^2}}
\end{bmatrix}
\]

or

\[
E_2 = \begin{bmatrix}
\frac{-1}{\sqrt{1+c^2}} \\
\frac{c}{\sqrt{1+c^2}} \\
\frac{c}{\sqrt{1+c^2}}
\end{bmatrix}
\]

where

\[
C = \frac{\sigma_{22} - \sigma_{11} - \sqrt{(\sigma_{11} + \sigma_{22})^2 - 4 \sigma_{11} \sigma_{22} (1 - \rho^2)}}{2 \rho \sigma_1 \sigma_2}
\]

Eigenvalues have the following properties:

1. \( \| E_1 \| = (E_1' E_1)^{1/2} = 1 \)
2. \( E_i' E_j = 0 \) for \( i \neq j \)

If \( X \sim N(\mu, \Sigma) \) then

\[
\tilde{X} = X - \mu \sim N(0, \Sigma)
\]

The length along this axis is proportional to \( \sqrt{\lambda_2} \)

\[
\lambda_1 \geq \lambda_2 \text{ are the eigenvalues of } \Sigma
\]

\( E_1 \) and \( E_2 \) are the corresponding eigenvectors \( (\Sigma E_i = \lambda_i E_i) \)

The length along this axis is proportional to \( \sqrt{\lambda_1} \)
The ratio of the lengths of the axes

\[ \frac{Y_1}{Y_2} \]

is the length of the major axes

length of minor axes

The actual lengths depend on the contour under consideration.

For the \((1-\alpha)\times 100\%\) contour the \(\frac{1}{2}\)-lengths are given by

\[
\sum_{\lambda_i} \frac{X(\lambda_i)}{\lambda_i^2}
\]

The spherical case:

Here \(\lambda_1 = \lambda_3\)

A 3-dimensional case: \[\hat{z}_i = \hat{z}_1 \hat{z}_2 \hat{z}_3\]

here \(\lambda_1 > \lambda_2 = \lambda_3\)
We will call this "smallest" region the central 
$(1-\alpha) \times 100\%$ of the 
multivariate normal population

The length of the interval 
containing the central 95% of the 
population is proportional to $\sigma$

Area of the region containing the central 
95% of the population is proportional to $\sqrt{\lambda_1 \lambda_2} = |\Sigma|^{1/2}$
The smallest region such that there is probability \(1 - \alpha\) that a randomly selected observation will fall in the region is a \(p\)-dimensional ellipsoid centered at \(\bar{x}\) with volume

\[
\frac{2\pi^{p/2} \Gamma(p/2)}{\Gamma(p/2)} \left[ \sum_{i=1}^{p} x_i^2 \right]^{p/2} \leq \frac{1}{2}
\]

where \(\Gamma(\cdot)\) is the gamma function.

When \(p\) is an even integer, \n
\[
\Gamma(\frac{p}{2}) = (\frac{p}{2}-1)(\frac{p}{2}-2)\cdots(1)\]

The "area" of the smallest ellipse containing \((1-\alpha)\times100\%\) of the population is

\[
\text{area} = \left( \frac{X_{(2)}; \alpha}{\pi} \right) \sqrt{\lambda} \leq \frac{1}{2}
\]

This extends to higher dimensions;

\[
\text{(volume of ellipsoid)} = \left( \text{constant} \right) \left( \sum_{i=1}^{p} x_i^2 \right)^{p/2} \leq \frac{1}{2}
\]
The "area" of the smallest ellipse containing \((1-\alpha) \times 100\%\) of the population is

\[
\text{area} = \left( \chi^2_{(2),\alpha} \right) \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}
\]

\[
= \text{(constant)} \ |\Sigma|^{\frac{1}{2}}
\]

This extends to higher dimensions

\[
\text{(volume of ellipsoid)} = \text{(constant)} \ |\Sigma|^{\frac{1}{2}}
\]

\[
P\text{-dimensional normal distribution}
\]

The smallest region such that there is probability \(1-\alpha\) that a randomly selected observation will fall in the region is a

(i) \(p\)-dimensional ellipsoid centered at \(\mu\)

(ii) with "volume"

\[
\frac{2 \pi^{p/2}}{\Gamma(p/2)} \left[ \chi^2_{(p),\alpha} \right]^{p/2} |\Sigma|^{\frac{1}{2}}
\]

where \(\Gamma(\cdot)\) is the gamma function

\[
\Gamma\left(\frac{p}{2}\right) = \left(\frac{p}{2}-1\right)\left(\frac{p}{2}-2\right) \ldots (2)(1) = \left(\frac{p}{2}-1\right)!
\]

when \(p\) is an even integer

\[
\Gamma\left(\frac{p}{2}\right) = \frac{(p-2) \ldots (5)(3)(1)}{2^{(p-1)/2} \sqrt{\pi}}
\]

when \(p\) is an odd integer
Generalized variance (Sect. 3.4)

\[ |\Sigma| = \lambda_1 \cdot \lambda_2 \cdots \lambda_p \]

= constant (squared volume of ellipsoid)

Generalized standard deviation

\[ |\Sigma|^{\frac{1}{2}} = (\lambda_1 \cdot \lambda_2 \cdots \lambda_p)^{\frac{1}{2}} \]

Consider also:

\[ |\Sigma|^{\frac{1}{2p}} = (\lambda_1 \lambda_2 \cdots \lambda_p)^{\frac{1}{2p}} \]

\[ \text{trace}(\Sigma) = \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \lambda_1 + \lambda_2 + \cdots + \lambda_p \]

(called the total variance)

Sample Data: (Chapter 3)

\[ X_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{p1} \end{bmatrix}, \quad X_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{p2} \end{bmatrix}, \quad \ldots \quad X_n = \begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{pn} \end{bmatrix} \]

are i.i.d. random vectors.

\[ E(X_j) = \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} \]

\[ \text{Var}(X_j) = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \]

If the data vectors are sampled from a multivariate normal (Gaussian) distribution

\[ X_j \sim \text{NID}(\mu, \Sigma) \quad j=1, 2, \ldots, n \]
Sample covariance matrix

\[
S = \begin{bmatrix}
S_{11} & S_{12} & \cdots & S_{1p} \\
S_{21} & S_{22} & \cdots & S_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
S_{p1} & S_{p2} & \cdots & S_{pp}
\end{bmatrix}
\]

Sample variance

\[
S_i = \frac{\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2}{n-1}
\]

Sample covariance

\[
S_{ik} = \frac{\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)}{n-1}
\]

Also

\[
S = \frac{1}{n-1} \sum_{i=1}^{p} S_i
\]

Sample mean vector:

\[
\bar{x} = \frac{1}{n} \begin{bmatrix} x_1 + x_2 + \cdots + x_n \end{bmatrix}
\]

Sample mean vector:

\[
\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{bmatrix} = \frac{1}{n} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
\]
\[ A = \frac{1}{n-1} S \]

where

\[ S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

is called a corrected sum of squares and cross-products matrix. To estimate trace (E)

\[ \text{use } \text{trace } (S) = \frac{1}{n} \sum_{i=1}^{n} x_i^2 \]

use |S| to estimate

Estimated generalized variance:
Example 3.1 Air samples

\[ X_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}, \quad N_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} \]

Sample mean vector:
\[ \bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = X \]

Sample covariance matrix:
\[ S = \begin{bmatrix} 1.7 & 2.6 \\ 2.6 & 6.3 \end{bmatrix} \]

Correlation:
\[ r_{1,2} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}} = \frac{2.6}{\sqrt{1.7 \times 6.3}} = 0.7945 \]

For the previous example, the generalized variance is

\[ s_1 = 3.95 \]

The total variance is

\[ \text{trace}(S) = s_{11} + s_{22} = 17 + 6.3 = 8.0 \]
Example 3.2

Another interpretation of \( |S| \) as a generalized variance: (Section 3.4)

\[
X = \begin{bmatrix}
7 & 4 & 4 & 5 & 4
\end{bmatrix} = \begin{bmatrix}
\gamma_1'
\end{bmatrix}
\]

\[
\begin{bmatrix}
12 & 9 & 5 & 8 & 8
\end{bmatrix}
\]

**Deviation vectors:**

\[
\xi_1 = \gamma_1 - \bar{\gamma}_1
\]

\[
= \begin{bmatrix}
7 & 4 & 4 & 5 & 4
\end{bmatrix} - 4.8 \begin{bmatrix}
1
\end{bmatrix} = \begin{bmatrix}
2.2
-0.8
0.2
-0.8
\end{bmatrix}
\]

\[
\xi_2 = \gamma_2 - \bar{\gamma}_2
\]

\[
= \begin{bmatrix}
3.6
0.6
-3.4
-0.4
-0.4
\end{bmatrix}
\]
\[
\cos(\theta) = \frac{\mathbf{e}_1 \cdot \mathbf{e}_2}{\|\mathbf{e}_1\| \|\mathbf{e}_2\|} = \frac{\delta_{12}}{\sqrt{s_{11} s_{22}}}
\]

Then, the area \( A \) is:

\[
A = (n-1) \sqrt{s_{22}} \sqrt{1 - \cos^2(\theta)}
\]

Since \( s = \begin{pmatrix} s_{22} & s_{23} \\ s_{32} & s_{33} \end{pmatrix} \),

\[
A = \sqrt{s_{22} (1 - \cos^2(\theta))}
\]

Then,

\[
A^2 = \frac{\text{area}(a)^2}{(n-1)^2}
\]

The area is

\[
A = \sqrt{\frac{s_{22}}{(n-1)^2}}
\]

The area of the configuration is

\[
A = \sqrt{\frac{s_{22}}{(n-1)^2}}
\]

where

\[
L = \frac{1}{2} \sum_{i=1}^{n-1} \left(\mathbf{x}_i \cdot \mathbf{x}_{i+1} - \mathbf{x}_i \cdot \mathbf{x}_0\right)
\]

and

\[
L_{e_2} = \sqrt{\frac{\delta_{22}}{s_{22}}}
\]
For fixed $l_{e_1}$ and $l_{e_2}$,
\[
\text{area} = (n-1) \sqrt{151}
\]
decreases as $\cos(\theta) \to 1$
or $\cos(\theta) \to -1$

For fixed $\theta$ (i.e. fixed $r_{12}$)
\[
\text{area} = (n-1) \sqrt{151}
\]
decreases as either
\[
l_{e_1} = \sqrt{(n-1)s_{11}} \text{ decreases or }\]
\[
l_{e_2} = \sqrt{(n-1)s_{22}} \text{ decreases}
\]

Your book sometime writes
the collection of sample vectors
as a $p \times n$ matrix
\[
X = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1n} \\
x_{21} & x_{22} & \cdots & x_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
x_{p1} & x_{p2} & \cdots & x_{pn}
\end{bmatrix}
\]
This is useful for calculations
and examples in Chapter 3, but
it is the transpose of the
way data are entered into SAS.
It is also the transpose of
the way we will express
data for MANOVA models.