多元统计分析

第3讲 多元正态分布(1)

Johnson & Wichern Ch. 4.1 – 4.5

统计学研究中心 邓婉璐
wanludeng@tsinghua.edu.cn

2018-2019春季学期

Recall: Statistical Inference

	上课时长	作业时长	复习时长	兴趣时长	休息时长	必修均分	选修均分	文素均分	…
1	24	20	6	0	8	91	88	94	…
2	20	18	2	4	7	88	86	93	…
3	23	18	4	2	6	90	89	96	…
4	18	20	0	5	7	87	91	91	…
5	15	17	3	4	9	86	93	96	…
…	…	…	…	…	…	…	…	…	…

- Find the basic pattern from data:
  - point estimation
  - Interval estimation
- Decision based on data
- Prediction / more pattern from data:
  - Regression
  - Dimension reduction
  - etc.

Outline

- ➢Why multivariate normal distribution?
- ➢What is it? Its density and properties
- ➢What inference can we make?
  - -Parameter Estimation
  - -Sampling distribution
  - -Large-sample behavior
  - -More for next lecture…

Introduction
- Why normal distribution?

Why normal distribution?

The multivariate normal distribution plays a central role in multivariate statistics for several reasons

- It is considerably more mathematically tractable than other multivariate distributions.
- The multivariate generalization of the central limit theorem says that properly centered and scaled sums of random vectors are asymptotically multivariate normal. Since many scientifically meaningful multivariate statistics are sums of vectors, this means that many statistics have approximately a multivariate normal distribution in large samples.
- Much more general sampling models can be constructed using multivariate normal distributions as building blocks (e.g. mixtures).

What is multivariate normal distribution?
- Density and properties

Multivariate Normal Distribution

Continuous Variable

Univariate Case: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$

Multivariate Case: $f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\{-\frac{(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2}\}$

$\boldsymbol{\mu}$ is the mean of $\mathbf{X}$

$\boldsymbol{\Sigma}$ is the variance - covariance matrix of $\mathbf{X}$

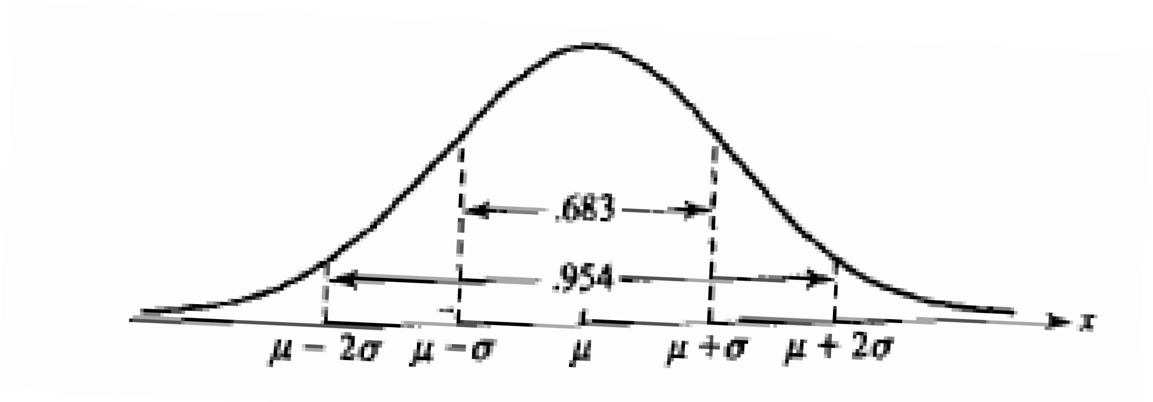
Note that $\boldsymbol{\Sigma}$ is symmetric positive definite matrix

$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

A multivariate normal distribution is completely characterized by its first two moments.
This is not typical of multivariate distributions in general.

Geometry of Multivariate Normal Distribution

1-dimension:



p-dimension: Contours of constant density for $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

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

What is the center and axes of the ellipsoids?

These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes

$\pm c\sqrt{\lambda_i}\mathbf{e}_\mathbf{i}$, where $\boldsymbol{\Sigma} e_i = \boldsymbol{\lambda}_\mathbf{i}\mathbf{e}_\mathbf{i}$

Example: Bivariate Normal Density

$$\mu_1 = E(X_1), \mu_2 = E(X_2)$$

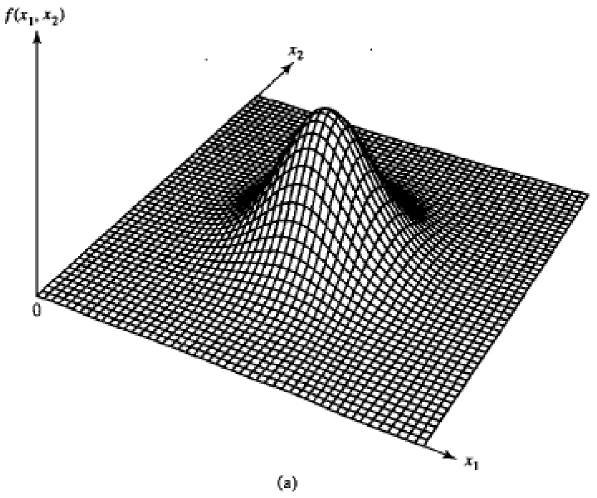
$$\sigma_{11} = Var(X_1), \sigma_{22} = Var(X_2), \rho_{12} = \sigma_{12} / \sqrt{\sigma_{11}\sigma_{22}}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{22} & \sigma_{22} \end{bmatrix}$$

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)} \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} & \sigma_{11} \end{bmatrix}$$

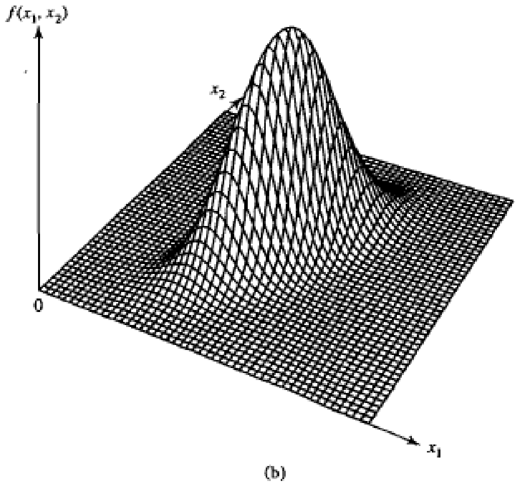
$$\boxed{f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\{-\frac{(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{2}\}}$$

Example: Bivariate Normal Density



(a)

$$\sigma_{11} = \sigma_{22}, \rho_{12} = 0$$

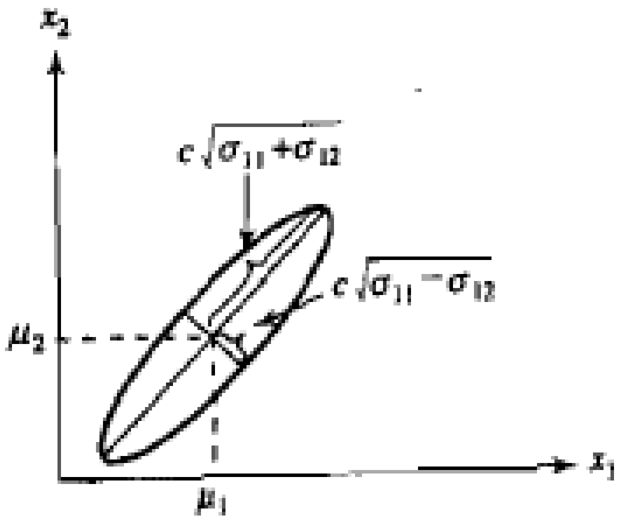


(b)

$$\sigma_{11} = \sigma_{22}, \rho_{12} = 0.75$$

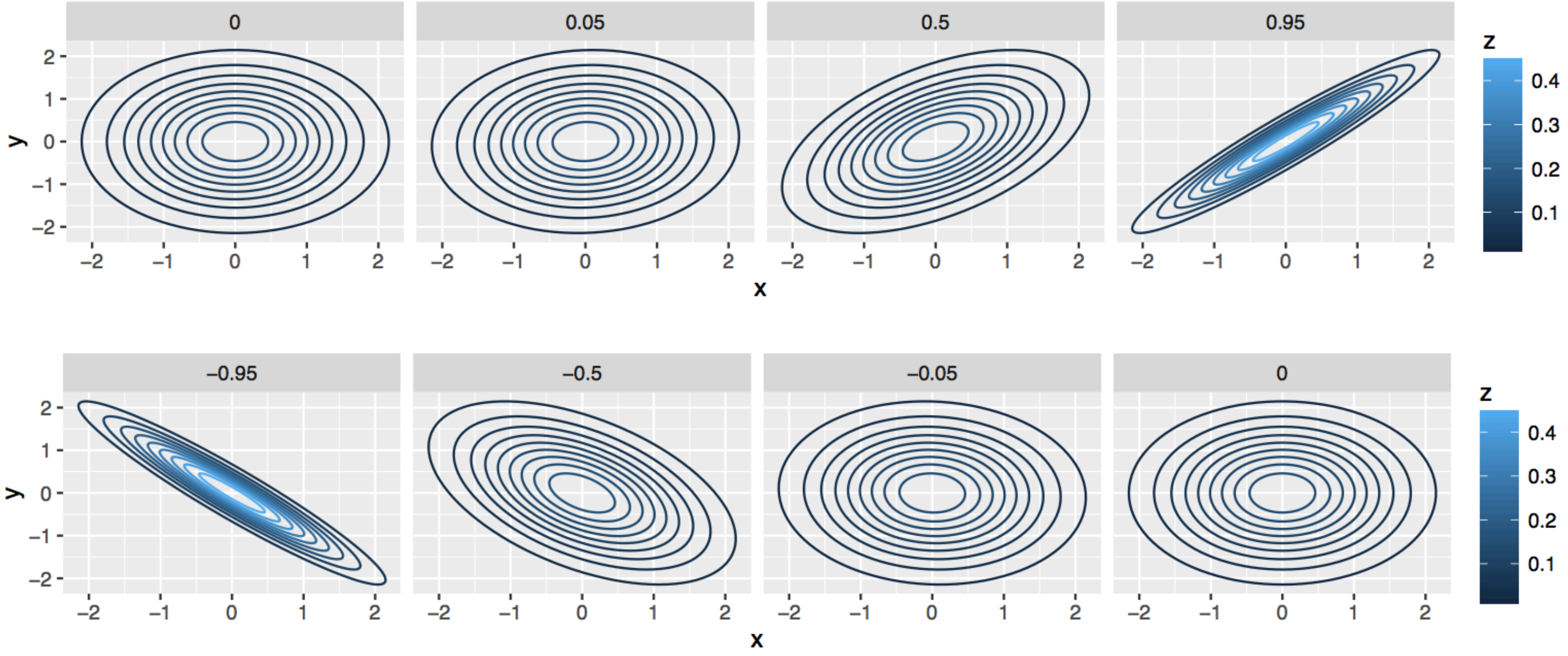
What is the eigenvalue and eigenvector Σ^{-1}?

What happens when we increase $|\rho_{12}|$?



What happens when ρ_{12}=0?

Example: Bivariate Normal Density



Properties of Multivariate Normal Distribution

Res 4.2
Equivalent definition

$\mathbf{X}$ is distributed as $N_p(\mu, \Sigma)$

$\Leftrightarrow$

$a'\mathbf{X}$ is distributed as $N(a'\mu, a'\Sigma a)$ for every $a \in R^p$.

定理 2.1

$\mathbf{X} = (X_1, X_2, ..., X_n)^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ 的充要条件 是对任何 $\mathbf{a} = (a_1, a_2, ..., a_n)^{\mathrm{T}} \in R^n$,

$$Y := \mathbf{a}^{\mathrm{T}}\mathbf{X} \sim \mathcal{N}(\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu},\ \mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a}).$$

证明. $(\Longrightarrow)$. Y 的特征函数

$$\begin{aligned}\phi_Y(t) = E\exp(\mathrm{i}tY) &= E\exp\left[\mathrm{i}(t\mathbf{a}^{\mathrm{T}})\mathbf{X}\right] \\ &= \exp\left[\mathrm{i}t(\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu}) - \tfrac{1}{2}t^2\mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a}\right].\end{aligned}$$

所以 $Y \sim \mathcal{N}(\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu},\ \mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a})$.

$(\Longleftarrow)$ 在 (1) 中取 $t = 1$ 得

$$E\exp(\mathrm{i}\mathbf{a}^{\mathrm{T}}\mathbf{X}) = \exp\left[\mathrm{i}\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a}\right].$$

故 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

Properties of Multivariate Normal Distribution

Res 4.3
Linear Combination

If $\mathbf{X}$ is distributed as $\mathrm{N_p}(\mu,\Sigma)$, the q linear combinations

$$\mathbf{AX} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix}$$

are distributed as $\mathrm{N_p}(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}')$. Also, $\mathbf{X}+\mathbf{d}$, where d is a vector of constants, is distributed as $\mathrm{N_p}(\mu+\mathbf{d},\Sigma)$

Properties of Multivariate Normal Distribution

The multivariate normal distribution is, in a fundamental sense, the distribution of affine transformations of independent Gaussians.

Let $Z_1,\cdots,Z_p$ be i.i.d. N(0,1), let $\Sigma = U\Lambda U'$ be a positive-definite symmetric matrix, and let $\mu \in R^p$ be a real vector. Then if $Z=(Z_1,\cdots,Z_p)'$ and $\Sigma^{1/2} = U\Lambda^{1/2}U'$ is the square root of Σ, we have $X=\mu+\Sigma^{1/2}Z$ be distributed as $N_p(\mu,\Sigma)$.

Proof. By independence,

$$f(z_1,\ldots,z_p) = \prod_j f_j(z_j) = (2\pi)^{-p/2}\exp\left(-z'z/2\right).$$

Make the transformation $x = \mu + \Sigma^{1/2}z$, so that $z = \Sigma^{-1/2}(x-\mu)$. Since $\frac{\partial z}{\partial x} = \frac{\partial}{\partial x}\Sigma^{-1/2}(x-\mu) = \Sigma^{-1/2}$, the Jacobian determinant is $|\Sigma^{-1/2}|$, giving

$$\begin{aligned} f(x) &= |\Sigma|^{-1/2}(2\pi)^{-p/2}\exp\left(-(\Sigma^{-1/2}(x-\mu))'(\Sigma^{-1/2}(x-\mu))/2\right) \\ &= |2\pi\Sigma|^{-1/2}\exp\left((x-\mu)'\Sigma^{-1}(x-\mu)/2\right). \end{aligned}$$

Properties of Multivariate Normal Distribution

Res 4.4
Marginal Distribution

All subsets of $\mathbf{X}$ are normally distributed.

If we respectively partition $\mathbf{X}$, its mean vector $\boldsymbol{\mu}$, and its covariance matrix $\boldsymbol{\Sigma}$ as

$$\underset{p\times 1}{\mathbf{X}} = \begin{bmatrix} \underset{q\times 1}{\mathbf{X}_1} \\ \underset{(p-q)\times 1}{\mathbf{X}_2} \end{bmatrix} \qquad \underset{p\times 1}{\boldsymbol{\mu}} = \begin{bmatrix} \underset{q\times 1}{\boldsymbol{\mu}_1} \\ \underset{(p-q)\times 1}{\boldsymbol{\mu}_2} \end{bmatrix} \qquad \underset{p\times p}{\boldsymbol{\Sigma}} = \begin{bmatrix} \underset{q\times q}{\boldsymbol{\Sigma}_{11}} & \underset{q\times(p-q)}{\boldsymbol{\Sigma}_{12}} \\ \underset{(p-q)\times q}{\boldsymbol{\Sigma}_{21}} & \underset{(p-q)\times(p-q)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}$$

then $\mathbf{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1, \Sigma_{11})$.

Properties of Multivariate Normal Distribution

Res 4.5
Independent vs Uncorrelated

(a) If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$, a $q_1 \times q_2$ matrix of zeros.

(b) If $\underset{p\times 1}{\mathbf{X}} = \begin{bmatrix} \underset{q\times 1}{\mathbf{X}_1} \\ \underset{(p-q)\times 1}{\mathbf{X}_2} \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \underset{q\times 1}{\boldsymbol{\mu}_1} \\ \underset{(p-q)\times 1}{\boldsymbol{\mu}_2} \end{bmatrix}, \begin{bmatrix} \underset{q\times q}{\boldsymbol{\Sigma}_{11}} & \underset{q\times(p-q)}{\boldsymbol{\Sigma}_{12}} \\ \underset{(p-q)\times q}{\boldsymbol{\Sigma}_{21}} & \underset{(p-q)\times(p-q)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}\right)$,

then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Suppose $\mathbf{X}_1$ and $\mathbf{X}_2$ are MVN, is $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ also MVN?

(c) If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and distributed as $N_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ and respectively.

Then, $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$

Properties of Multivariate Normal Distribution

Res 4.6
Conditional Distribution

Let $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ be distributed as $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$, and $|\boldsymbol{\Sigma}_{22}| > 0$.

Then the conditional distribution of $\mathbf{X}_1$, given $\mathbf{X}_2 = \mathbf{x}_2$, is normal and has

$$\boxed{\begin{aligned} &\text{Mean} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &\text{Covariance} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \end{aligned}}$$

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right)$$

What is the conditional distribution $X_1 | X_2$?

Notice that covariance does not depend on value of x_2

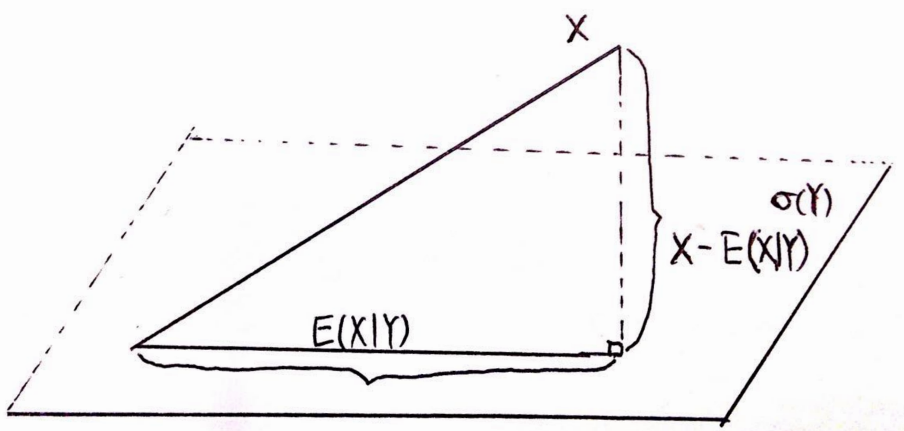
想法就是通过行变换把非对角元消为0

Related to linear model

Properties of Multivariate Normal Distribution

Res 4.6
Conditional Distribution

Recall:



例 3.6

设 X, Y 是随机变量，$h(x)$ 是实函数，若 $E(X^2) < \infty$, $E[h^2(Y)] < \infty$, 则

$$E[(X - E(X|Y))h(Y)] = 0$$

例 3.7 (最佳预测问题)

设 $E(X^2) < \infty$, $m(Y) = E(X|Y)$, 则对任何实函数 $g(y)$, 有

$$E[X - m(Y)]^2 \le E[X - g(Y)]^2,$$

其中等号成立当且仅当 $g(Y) = m(Y)$ $a.s.$

Properties of Multivariate Normal Distribution

Res 4.7
Chi-square distribution

Let $\mathbf{X}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

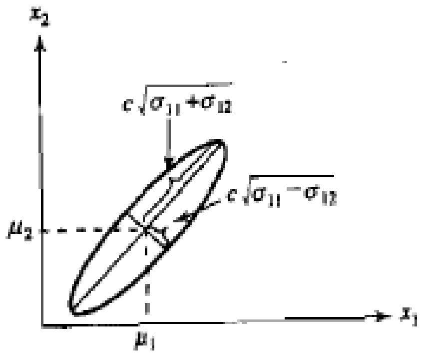
(a) $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ is distributed as χ^2_p, where χ^2_p denotes the chi-square distribution

with p degee of freedom.

(b) The distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ assigns probability $1-\alpha$ to the solid ellipsoid

$\{\mathbf{x} : (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \le \chi^2_p(\alpha)\}$,

where $\chi^2_p(\alpha)$ denotes the upper (100α) percentile of the χ^2_p distribution.



$$\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{X}-\boldsymbol{\mu}) \sim N_p(0, \mathbf{I}_p)$$

$$(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})' = [(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1/2}][\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})']$$
$$= \mathbf{Z}'\mathbf{Z} = Z_1^2 + \cdots + Z_p^2$$

Properties of Multivariate Normal Distribution

Res 4.8 Distribution of Linear Combinations of MVNs

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be mutually independent with $\mathbf{X}_j$ distributed as $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$.

(Note that each $\mathbf{X}_j$ has the same covariance matrix $\boldsymbol{\Sigma}$.) Then

$$\mathbf{V}_1 = c_1\mathbf{X}_1 + c_2\mathbf{X}_2 + \cdots + c_n\mathbf{X}_n$$

is distributed as $N_p(\sum_{j=1}^{n} c_j\boldsymbol{\mu}_j, (\sum_{j=1}^{n} c_j^2)\boldsymbol{\Sigma})$.

Moreover, $\mathbf{V}_1$ and $\mathbf{V}_2 = b_1\mathbf{X}_1 + b_2\mathbf{X}_2 + \cdots + b_n\mathbf{X}_n$ are jointly multivariate normal

with covariance matrix

$$\begin{bmatrix} (\sum_{j=1}^{n} c_j^2)\boldsymbol{\Sigma} & (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} \\ (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} & (\sum_{j=1}^{n} b_j^2)\boldsymbol{\Sigma} \end{bmatrix}$$

Consequently, $\mathbf{V}_1$ and $\mathbf{V}_2$ are independent if $\mathbf{b}'\mathbf{c} = \sum_{j=1}^{n} b_j c_j = 0$.

Inference

- MLE
- Sampling distribution
- Large-sample properties

Modeling with Multivariate Normal Distribution

Assume that the $p \times 1$ vectors $\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X}_n$ represents a random sample from a multivariate normal population, $N_p(\mu, \Sigma)$:

(1) $\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X}_n$ are mutually independent;

and (2) $\mathbf{X}_i$ follows $N_p(\mu, \Sigma)$, for every $i=1, \cdots, n$.

$$\text{Joint Density of } \mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X}_n = \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{\rho/2} |\mathbf{\Sigma}|^{1/2}} e^{-(\mathbf{x_j}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x_j}-\boldsymbol{\mu})/2} \right\}$$

$$= \frac{1}{(2\pi)^{n\rho/2} |\mathbf{\Sigma}|^{n/2}} e^{-\sum_{j=1}^{n} (\mathbf{x_j}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x_j}-\boldsymbol{\mu})/2}$$

Likelihood Function

Maximum Likelihood Estimation under Normality Assumption

Result 4.11 MLE

证明并不显然！！

NOT APPROVED

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be a random sample from

a normal population, $\mathrm{N_p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then,

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{X}} \text{ and } \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_\mathrm{j} - \overline{\mathbf{X}})(\mathbf{X}_\mathrm{j} - \overline{\mathbf{X}})' = \frac{n-1}{n}\mathbf{S}$$

are the maximum likelihood estimator of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

Their observed values $\bar{\mathbf{x}}$, and $\frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_\mathrm{j} - \mathbf{x})(\mathbf{x}_\mathrm{j} - \mathbf{x})'$, are called

the maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Maximum Likelihood Estimation under Normality Assumption

The maximum of the likelihood:

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

$$\propto \quad (\text{generalized variance})^{-n/2} \quad .$$

MLE of h(θ): $\boxed{h(\hat{\theta})}$ Invariance property of MLE

e.g., MLE of $\mu'\Sigma^{-1}\mu$ is $\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$,

MLE of $\sqrt{\sigma_{ii}}$ is $\sqrt{\hat{\sigma}_{ii}}$, where $\hat{\sigma}_{ii} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$.

Unbiased Estimator under Normality Assumption

Unbiased estimator:

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be a random sample from a normal population, $\mathrm{N_p}(\mu, \Sigma)$. Then,

$$\hat{\mu} = \bar{\mathbf{X}} \quad \text{and } \mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{X}_\mathrm{j} - \bar{\mathbf{X}})(\mathbf{X}_\mathrm{j} - \bar{\mathbf{X}})' = \frac{n}{n-1} \hat{\Sigma}$$

are the unbiased estimator of μ and Σ, respectively.

Sufficient Statistics

Aside:

$$\sum_{j=1}^{n}(\mathbf{x_j}-\mu)'\Sigma^{-1}(\mathbf{x_j}-\mu) = tr\Big(\sum_{j=1}^{n}(\mathbf{x_j}-\mu)'\Sigma^{-1}(\mathbf{x_j}-\mu)\Big) = tr\Big(\Sigma^{-1}\sum_{j=1}^{n}(\mathbf{x_j}-\mu)(\mathbf{x_j}-\mu)'\Big)$$

$$= tr\Big(\Sigma^{-1}\Big(\sum_{j=1}^{n}(\mathbf{x_j}-\bar{x})(\mathbf{x_j}-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)'\Big)\Big)$$

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be a random sample from a multivariate normal population, $\mathrm{N_p}(\mu, \Sigma)$. Then, $\bar{\mathbf{X}}$ and $\mathbf{S}$ are sufficient statistics.

It means, for normal populations, all of the information about the model parameters is contained in $\bar{\mathbf{X}}$ and $\mathbf{S}$

Sampling Distribution of Sufficient Statistics

Recall Result 3.1

$E(\overline{\mathbf{X}}) = \boldsymbol{\mu}$

$Cov(\overline{\mathbf{X}}) = \frac{1}{n}\boldsymbol{\Sigma}$

$E(S_n) = \frac{n-1}{n}\boldsymbol{\Sigma} = \boldsymbol{\Sigma} - \frac{1}{n}\boldsymbol{\Sigma}$, or $E(\frac{n}{n-1}S_n) = \boldsymbol{\Sigma}$

Now, under normality assumption
From Result 4.8

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be a random sample of size n

from $\mathrm{N_p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then,

1. $\overline{\mathbf{X}}$ is distributed as $\mathrm{N_p}(\boldsymbol{\mu}, (1/n)\boldsymbol{\Sigma})$.

2. $(n-1)\mathbf{S}$ is distributed as a Wishart random matrix with (n - 1) d. f.

3. $\overline{\mathbf{X}}$ and $\mathbf{S}$ are independent.

You can prove it by
- following the same way as you learned for 1-dim,
- or making use of properties of normal distribution under matrix operations.

Sampling Distribution of Sufficient Statistics

Wishart Distribution

$W_m(\cdot \mid \boldsymbol{\Sigma})$ = Wishart distribution with m d.f.

= distribution of $\sum_{j=1}^{m} \mathbf{Z}_j \mathbf{Z}_j'$

where $\mathbf{Z}_j$ the are each independently distributed as $N_p(\mathbf{0}, \boldsymbol{\Sigma})$

Properties of Wishart Distribution

1. If $\mathbf{A}_1$ is distrtibuted as $W_{m1}(\mathbf{A}_1 \mid \Sigma)$ independently of $\mathbf{A}_2$, which is distributed as $W_{m2}(\mathbf{A}_2 \mid \Sigma)$, then $\mathbf{A}_1 + \mathbf{A}_2$ is distributed as $W_{m1+m2}(\mathbf{A}_1 + \mathbf{A}_2 \mid \Sigma)$. That is, the degrees of freedom add.

2. If $\mathbf{A}$ is distributed as $W_m(\mathbf{A} \mid \Sigma)$, then $\mathbf{CAC}'$ is distributed as $W_m(\mathbf{CAC}' \mid \mathbf{C}\Sigma\mathbf{C}')$.

Simulation for Wishart Distribution

Wishart Distribution

$W_m(\cdot \mid \Sigma)$ or $W_p(m, \Sigma)$ Another way to represent it

We can generate random matrices from a Wishart distribution in R using the package MCMCpack, like this:

```
p <- 2
n <- 50
Sigma <- matrix(c(1,.9,.9,1),2,2)
Phi <- rwish(n-1,Sigma)
print(Phi)
```

```
##              [,1]      [,2]
## [1,] 60.07647 54.44065
## [2,] 54.44065 56.86697
```

Large-Sample Behavior of $\bar{\mathbf{X}}$ and $\mathbf{S}$

- No requirement of normality

Large Sample Behavior of $\bar{\mathbf{X}}$ and $\mathbf{S}$

Result 4.12 Law of Large Numbers

Let $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n$ be independent observations from a population with mean $E(\mathbf{Y}_i) = \mu$.

$$\text{Then, } \bar{\mathbf{Y}} = \frac{\mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_n}{n}$$

$\bar{\mathbf{Y}}$ converges in probability to μ.

A direct consequence of LLN is that

$\bar{\mathbf{X}}$ converges in probability to $\boldsymbol{\mu}$

$\mathbf{S}$ converges in probability to $\boldsymbol{\Sigma}$

Large Sample Behavior of $\bar{\mathbf{X}}$ and $\mathbf{S}$

Result 4.13 The Central Limit Theorem

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be independent observations from any population with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$. Then

$\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$ has an approximate $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution and

n$(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})$ is approximatelty χ_p^2 for large sample sizes.

Here n should be relatively large to p.

T. W. Anderson, An Introduction to Multivariate Statistical Analysis. Wiley-Interscience.

Summary

Summary

- Definition: the density and contour plots
- Properties:
  - understand the intuition: association, correlation, projection.
  - proofs, especially for conditional distribution.
- Inference
  - Parameter Estimation: MLE
  - Sampling distribution of sufficient statistics
  - Large-sample behavior: no requirement of normality